THE UNIVERSITY
of EDINBURGH

Structural and Biochemical insights into the ATP-dependent Chromatin Remodeler LSH

Simon Varzandeh

Doctor of Philosophy – University of Edinburgh – 2016

Abstract

Chromatin remodelling proteins support a variety of cellular functions and utilise the energy from ATP hydrolysis to either reposition or evict nucleosomes. One such protein, Lymphoid specific helicase (LSH), regulates DNA methylation in mammalian cells cooperatively with DNA Methyltransferase 3B (DNMT3B) through binding of the N-terminal domain of LSH. The correct functioning of LSH is essential for heterochromatin formation, with a knockout of LSH causing perinatal lethality or severe developmental abnormalities. There is little biochemical data and no structural data on LSH. Therefore, we aim to determine the structural characteristics and regulatory mechanism of LSH *in vitro.*

LSH was expressed in an optimised insect cell system which increased protein yield 25-fold with greater than 95% purity. LSH is monomeric with increased thermal stability upon ATP or ADP binding. Full length LSH could not be crystallised therefore a core ATPase region of LSH missing the N-terminal domain was identified through limited proteolysis. This also provided evidence the N-terminal domain of LSH is disordered, which was proven through biophysical characterisation of LSH^{1-176}. Expression of the LSH ATPase region was weak and the protein was unstable; suggesting the N-terminal domain of LSH is required for LSH stability. Therefore, complementary structural methods were used to study LSH.

Crosslinking mass-spectrometry revealed the N and C termini are in close proximity, suggesting flexible linking regions, which was supported by limited proteolysis experiments. Negative staining Electron Microscopy defined LSH as a tri-lobal and elongated structure which could harbour the ATPase region in the two spherical lobes. 3D modelling of SAXS data obtained of LSH was in agreement with EM data.

To understand molecular mechanisms of LSH, functional studies investigating LSH:DNA and LSH:DNMT3B interactions were performed. LSH had a K_D for dsDNA of 0.4 μM in solution. LSH does not bind ssDNA nor does it have a greater affinity for methylated dsDNA. LSH was found to bind the dsDNA overhangs of nucleosomes but not to core nucleosomes, suggesting LSH solely interacts with DNA in chromatin and not histones. A stable complex of LSH:DNMT3B could not be achieved *in vitro*, however, other components for complex formation may have been missing.

This study has improved our understanding of LSH structure, biophysical properties and its biochemical interaction with DNA and nucleosomes. This study has laid the foundations for the structural investigations of a LSH:nucleosome and potentially a LSH:DNMT3B complex *in vitro* to gain a greater understanding of how functional domains of LSH regulates its enzymatic function.

Lay summary

In each cell, DNA makes up all our genetic material which if stretched out is two metres long. Therefore DNA has to be condensed in order to physically fit inside the nucleus of a cell. This is done so by DNA wrapping around histone proteins into 'bead on a string' like structures, termed nucleosomes which are further condensed into compact chromatin. In order for genes to be accessed by regulatory proteins, a family of chromatin remodelling enzymes reposition or evict nucleosomes. Therefore chromatin remodelers are critical for cell regulation, with mutations or non-functional enzymes resulting in disease or lethality. One such chromatin remodelling enzyme named Lymphoid Specific Helicase (LSH) regulates genes through cooperation with an enzyme known as DNA methyltransferase 3B (DNMT3B) which mediates gene activation and deactivation through the addition of methyl groups to DNA. This study has investigated the regulatory mechanisms of LSH and its interactions with DNA, nucleosomes and DNMT3B by using structural and biochemical methods.

Declaration

The work presented in this thesis is the original work of the author. This thesis has been composed by the author and has not been submitted in whole or in part for any other degree.

Simon Varzandeh

August, 2016

Acknowledgements

A vast number of people have been both helpful and inspirational throughout my PhD, something I hope has made me both a better scientist and teacher, things which I will not forget for the rest of my life.

I would firstly like to thank my supervisors; Dr Julia Richardson and Dr Irina Stancheva, both of you have provided support, useful suggestions and have been important in making me a better scientist.

I would secondly like to thank and show my gratitude to the following people:

Old and present members of the Stancheva lab (Burak Ozkan, Dr Katrina Gordon, Dr Tuo Zhang, Natalia Torrea, Ilaria Amendola, Christian Belton, Dr Ausma Termanis, Becky Tillotson, Sadie Kemp and Dani Wicaksono) and the Richardson lab (Dr Maryia Trubitsyna, Heather Grey, Bavanthi Navarathne and Dr Liz Morris). You have all addressed useful points for my project and been kind enough to pick me up when I was down, I hope I gave the same back to you.

The Cook lab (Dr Atlanta Cook, Dr Ula McCaughan, Dr Uma Jayachandran and Valdeko Kruuzvee) and Spagnolo labs (Dr Laura Spagnolo, Dr Giuseppe Cannone and Dr James Parker) and for useful suggestions at lab meetings. Laura and Giuseppe also a big thank you for the EM work on LSH.

Dr Philipp Voigt, Kim Webb and Elana Bryan were all very helpful in providing useful help and materials for nucleosome based work. Dr Su-ling Leong, Dr Harish Thakur and Dr Eric Lynch from the Sawin lab and Dr Onur Sen from the Hardwick lab for aid with all things insect cell expression based and for good banter!

Phu le Thanh for setting up Phucked up! Brewing for the TGIFs, we made some good beers!

The EPPF staff – Dr Martin Wear, Dr Matt Nowicki, Dr Liz Blackburn and Dr Janice Bramham. You have all been a tremendous resource of knowledge and help with all things protein expression, purification and characterisation. Martin, your random banter kept spirits up.

Dr Li-Hsuan Yen, Dr Meng Yuan, Jia Ning, James Kinkaed, Peter Fernandes, Andromachi Xypnitou, Guilia Romanelli, Dr Ula McCaughan, Valdeko Kruuzvee, Laura Tuck and Didi He for making office life a lot of fun.

Adam Irvine for the ERI intership who provided me additional knowledge on patents and technology transfer which would be beneficial in whatever I decide to do next.

Dave Kelly for writing a small script to count fluorescent cells for the titration assay for protein expression in insect cells.

Dr Juan Zou for everything relating to crosslinking mass spectrometry.

Media kitchen staff for aiding in the smooth running of the lab.

Dr Andy Cronshaw and Dr Logan MacKay for mass spectrometry and Dr Juraj Bella for NMR of proteins.

My internal (Dr Jayaprakash Arulanandam) and external (Prof. Tom Owen-Hughes) examiners for the useful suggestions and insights into how to improve the some of the decisions I made during my PhD.

I had the pleasure of supervising two bright students Mari Eltermann and Laura Fitzpatrick during my time as a PhD student, whom I hope I have aided towards enjoying science.
A special thankyou to C^3 (Tanmay Gupta) and Capt. T (Sanju Ashraf) it was easier writing the thesis with you guys doing the same, we got through it and ate like kings doing it!
A very big thankyou to Emilija Ruksenaite, you've helped so much in keeping my spirits high whilst writing. Thank you for reading my thesis and having discussions with me about it.
Lastly, I thank my Brother, Alex, my Mum, Michelle and my Dad, Hossein for their continued moral support throughout my PhD and for believing in me when I couldn't. I dedicate this thesis to all those here that have contributed and for those who may contribute further. I am humbled by any benefit this work may have on gaining a deeper understanding of biology.

Abbreviations

A	Amps
Amp^R	Ampicillin resistance
Å	Angstrom
ACF	ATP-dependent Chromatin assembly and remodelling Factor
ADD	ATRX-DNMT3A-DNMT3L
ADP	Adenosine diphosphate
AGE	Agarose Gel Electrophoresis
AMP	Adenosine monophosphate
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine triphosphate
BAF1	Brg1 Aassociated Factors
BCA	Bicinchoninic acid
BEV	Baculovirus Expression Vector
bp	Base pairs
BME	Beta-MercaptoEthanol
BS3	Bis (sulfosuccinimidyl) suberate
53BP1	P53 Binding Protein 1
CAPS	*N*-cyclohexyl-3-aminopropanesulfonic acid
°C	degrees Celsius
CC	Coiled Coil
CD	Circular Dichroism
CDCA7	Cell Division Cycle Associated 7
CenpA	Centromere protein A
CH_3	Methyl group
CHCA	α-cyano-4-hydroxyconnamic acid
CHD	Chromodomain-Helicase-DNA binding
CHD1	Chromodomain helicase DNA binding protein 1
CHIP-seq	Chromatin Immunoprecipitation-Sequencing
Chl^R	Chloramphenicol resistance
CHRAC	CHRomatin Accessibility Complex
C-terminus	Carboxy terminus
CTD	C-Terminal Domain
DBP	DNA binding protein
DBD	DNA binding domain
DDM1	Defective in DNA Methylation 1
DDR	DNA Damage Response
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic acid
dsDNA	Double stranded Deoxyribonucleic acid
ssDNA	Single stranded Deoxyribonucleic acid
DNMT	DNA Methyltransferase
DNMT1	DNA Methyltransferase 1
DNMT3A	DNA Methyltransferase 3A

DNMT3B	DNA Methyltransferase 3B
DSBR	Double strand break repair
DTT	Dithiothreitol
EB	Elution buffer
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
EHMT1	Euchromatic histone-lysine N-methyltransferase 1 (aka G9A)
EHMT2	Euchromatic histone-lysine N-methyltransferase 2 (aka GLP)
EM	Electron Microscopy
EOM	Ensemble Optimisation Model
EMSA	Electrophoretic Mobility Shift Assay
EDTA	EthyleneDiamineTetraAcetic acid
esBAF	ES cell specific chromatin remodelling complex
EZH2	Enhancer of Zeste Homolog 2
FL	Full length
FP	Fluorescence polarisation
FT	Flow through
$Gent^R$	Gentamicin resistance
HAT	Histone AcetylTransferase
HDAC	Histone Deacetylase
HDAC1	Histone Deacetylase 1
HELLS	HELicase Lymphoid Specific
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
6His	Hexahistidine tag
HPLC	High-Performance Liquid Chromatography
HMT	Home-made Taq Polymerase
HNSCC	Head and neck squamous cell carcinoma
HSA	Helicase SANT associated domain
HSS	HAND-SANT-SLIDE domain
IAA	IodoAcetic acid
ICF	Immunodeficiency, Centromeric instability Facial anomalies syndrome
IMAC	Immobilised metal affinity chromatography
INO80	INOsitol requiring 80
IR	Ionising Radiation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISWI	Imitation SWItch
ITC	Isothermal Titration Calorimetry
Kan^R	Kanamycin resistance
Kd	Dissociation constant
KD	Knockdown
kDa	Kilodalton
kpsi	Thousand pounds per square inch
LB	Luria Broth
LINES	Long Interspersed Nuclear Elements
lncRNA	Long non-coding RNA
LSH	Lymphoid Specific Helicase
M	Milli (10^{-3})

MALDI-ToF	Matrix-assisted Laser Deabsorption Ionisation – Time of Flight
MBD	methyl-CpG-binding-domain
MBP	methyl-CpG-binding-domain protein
5mC	5-methyl Cytosine
MDa	MegaDalton
MDC1	Mediator of DNA Damage Checkpoint 1
µ	Micro (10^{-6})
MeCP2	methyl-CpG-binding protein 2
MEFs	Mouse Embryonic Fibroblasts
MES	2-(*N*-morpholino) ethanesulfonic acid
Mg^{2+}	Magnesium ion
MLL1	Mixed Lineage Leukaemia 1
MRE11	Meiotic recombination 11 homolog A
MT	Catalytic Methyltransferase domain
MW	Molecular Weight
MWCO	Molecular Weight Cut Off
n	Nano (10^{-9})
NBS1	Nijmegen Breaking Syndrome Protein 1
NEB	New England Biolabs Inc.
Ni-NTA	Nickel Nitrilotriacetic acid
NLS	Nuclear Localisation Signal
NMR	Nuclear magnetic resonance
N-terminus	Amino terminus
NTD	N-Terminal Domain
NURF	NUcleosome Remodelling Factor
NuRD	Nucleosome Remodeling and Deactylase
□	Ohms
OD	Optical density
ORF	Open Reading Frame
p55	55-kDa polypeptide
PASG	Proliferation associated SNF2-like gene
PBAF	PolyBromo-Associated BAF
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PcrA	Plasmid copy reduced A
PEG	Polyethylene glycol
PDB	Protein databank
PHD	Plant Homeo-Domain
PRC2	Polycomb Repressor Complex 2
PTM	Post Translational Modification
pI	Isoelectric point
PVDF	Polyvinylidene fluoride
PWWP	Pro-Tryp-Tryp-Pro
RE	Restriction enzyme
RNAPII	RNA Polymerase II
rpm	Rotations per minute

R_g	Radius of gyration
R_h	Hydrodynamic radius
RMSD	Root Mean Squared Deviation
RSC	Remodels the Structure of Chromatin
Rvb1/Rvb2	RuVB-like 1/RuvB-like 2
SAM	S-Adenosyl methionine
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium dodecyl Sulphate polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
SEC-MALS	SEC coupled to Mutli-angle Light Scattering
SF2	Helicase Superfamily 2
SNF2	Sucrose Non-Fermentable Family 2
*Sf*9	*Spodoptera frugiperda* isolate 9
SWR1	Sick With Rat8
SOC	Super Optimal Broth medium
SINES	Short interspersed nuclear elements
SMARCA6	Matrix associated Actin dependent Regulator of Chromatin, subfamily A, member 6
SRCAP	SNF2 Related CREB-Activator Protein
Sulfo-NHS	N-hydroxysulphosuccinimide
SWI/SNF	SWItching defective/Sucrose Non-Fermenting
TAE	Tris-Acetate EDTA
TBE	Tris Boric acid EDTA
TBS	Tris-buffered saline
TCEP-HCl	Tris (2-CarboxyEthyl) Phosphine Hydrochloride
TDA	Thermal Denaturation Assay
TE	Tris-EDTA
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
TGE	Tris-Glycine-EDTA
TGS	Tris-glycine SDS
TSS	Transcriptional Start Sites
T_m	Melting Temperature
UV	Ultra-violet
v/v	Volume per volume
VDJ	Variable, Diverse and Joining
w/v	Weight per volume
WT	Wildtype
XIST	X-Inactive Specific Transcript
XL-MS	Crosslinking Mass Spectrometry
YFP	Yellow Fluorescent Protein
ZBTB24	Zinc-finger and BTB domain containing 24

Amino Acids

A	Ala	alanine
C	Cys	cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

Nucleotides

A	adenine
C	cytosine
G	guanine
T	thymine

Contents

Table of Contents

Table of Figures

Table of Tables

Chapter 1| Introduction

1.1| Chromatin organisation and structure

1.1.1| Chromatin organisation and structure

Chromatin is the topological solution for the packaging of 1.7-2 metres of DNA into a eukaryotic nucleus (Kornberg, 1974). The repeating unit of chromatin is a nucleosome core particle, which consists of a histone octamer wrapped in 1.65 left-handed super-helical turns of 147 bp dsDNA, forming a disc-shaped complex conserved across eukaryotes (Figure 1.1) (Kornberg 1974, Luger et al. 1997; Luger & Collins 2001). The negatively charged DNA is wound around the basic surface of the histone octamer (Figure 1.1). Interactions between the histone octamer and DNA of a nucleosome core particle is through hydrogen bonding between the histone octamers main chain amide atoms and the phosphate oxygen atoms of the DNA backbone (Luger & Collins 2001).

The histone octamer comprises two H2A/H2B dimers and a $(H3/H4)_2$ tetramer (Luger *et al.*, 1997; Luger and Collins, 2001). All core histones share a conserved eukaryotic structural motif: the histone fold. (Luger *et al.*, 1997; Luger and Collins, 2001). The histone fold is necessary for the formation of a tight homo or heterodimer with another histone fold and is required for binding ~120 bp of the dsDNA (Luger *et al.*, 1997; Luger and Collins, 2001).

Structured regions outside of the histone fold, known as histone fold extensions, are necessary for protein-protein interactions within the histone octamer and occur between the H2A-H2B dimer and one half of the H3-H4 tetramer resulting in DNA binding of ~10 bp from each H3 histone fold extension (for a total of 20 bp) (Luger *et al.*, 1997; Luger and Collins, 2001).

The nucleosome has a symmetry axis at the dyad where the $(H3/H4)_2$ tetramer is located and is flanked by the DNA entry and exit site (Figure 1.1) (Kornberg 1974, Luger et al. 1997; Luger & Collins 2001, Saha et al, 2006). Flexible histone tails are defined as those parts of the histone proteins that extend from the confines of the DNA superhelix and do not form an integral part of the histone octamer nor make contact with DNA (Luger *et al.*, 1997; Luger and Collins, 2001). Histone tails are the locations of post-translational modifications (PTMs) (discussed later).

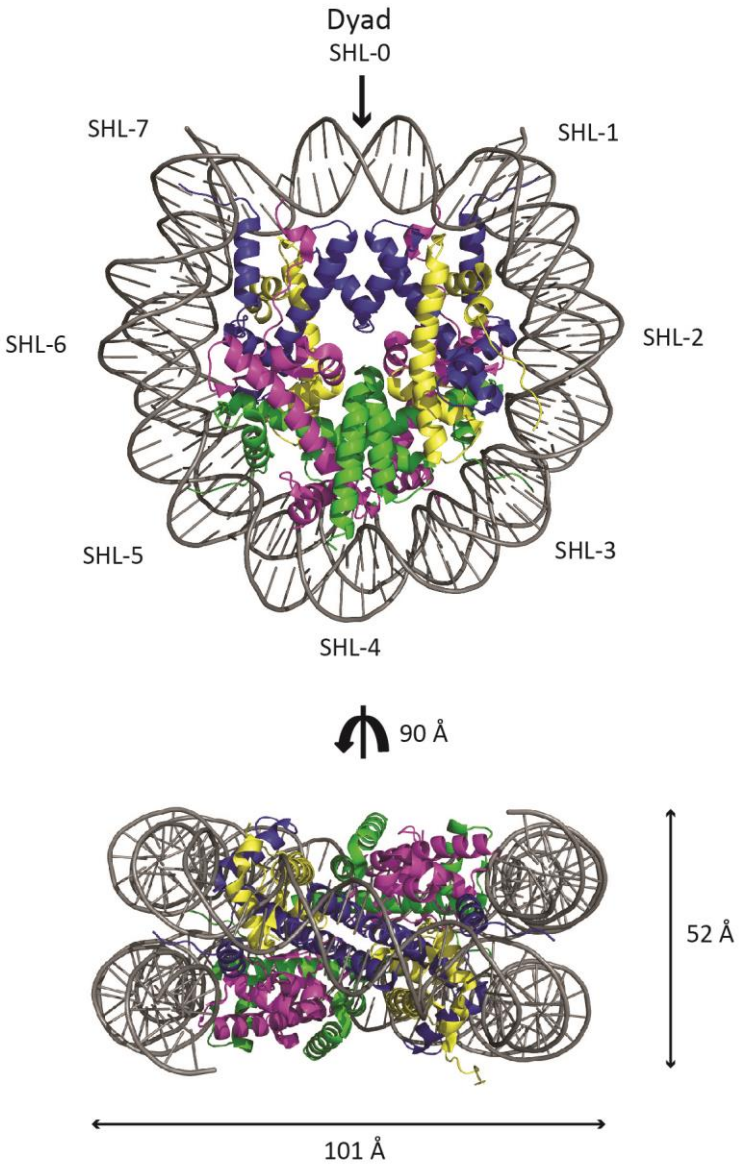

Figure 1.1| The basic structural unit of chromatin - the nucleosome core particle.
Crystal structure of a nucleosome core particle with xenopus laevis histones asssembled with a palindromic widom '601' derivative at 2.2 Å. (PDB - 3UT9) (Chua et al. 2012)
H3 is blue, H4 is yellow, H2A is green and H2B is purple. The 147 bp dsDNA is grey. There are 14 DNA-histone contacts with each 10 bp turn of a the DNA helix (Kornberg 1974; Luger et al. 1997; Luger & Collins 2001). The dyad is the position of the super-helix location 0 (SHL-0), with each 10 bp turn clockwise with +1 added .

Variants of the core histones have been identified in chromatin and exist predominantly in the histone families H2A and H3 (Talbert and Henikoff, 2010a; Weber and Henikoff, 2014). The H2A and H3 variants alter the biophysical properties of the nucleosome and confer specialised functions regulating transcription (Talbert and Henikoff, 2010a; Weber and Henikoff, 2014). For example, the H3 variant Centromere protein A (CenpA) is found at active centromeres in mammals and is important for the formation of active centromeres (Niikura et al. 2015, Yoda et al. 2000; Warburton et al. 1997). As another example, the H2A variant H2A.X is phosphorylated at regions flanking a DNA double-strand break (DSB) to recruit DNA repair proteins (Iacovoni et al., 2010, Polo and Jackson, 2011, Bakkenist and Kastan, 2003).

Binding of the linker histone H1 to the nucleosomal dyad and the linker DNA entering and exiting the nucleosome core particle forms the chromatosome, a more stable nucleosomal structure (Bednar et al. 1998; Hizume et al. 2005; Harshman et al. 2013, Thoma et al. 1979).. Linked chromatosomes form 10-11 nm fibres which have been assumed to fold into 30 nm fibres which subsequently form interphase nuclei and mitotic nucleosomes (Finch and Klug, 1976; Maeshima, Hihara and Eltsov, 2010). 30 nm fibres have been identified and studied extensively *in vitro* using several structural techniques including EM (Finch and Klug, 1976), cryo EM (Robinson *et al.*, 2006) and SAXS (Paulson and Langmore, 1983). However, 30 nm fibres were not identified in studies which used cryo electron microscopy of vitreous sections (CEMOVIS), a high resolution imaging technique of cell structures close to the native-state which uses unmodified thin sections of vitrified mammalian mitotic cells (McDowall, Smith and Dubochet, 1986; Eltsov *et al.*, 2008). Instead, chromosomes had homogeneous grainy texture with ~11 nm spacing, which has led to the hypothesis that chromosomes are made up of a 'polymer melt' in which 10 nm nucleosome fibres form dynamic polymer chains which rearrange constantly (McDowall, Smith and Dubochet, 1986; Eltsov *et al.*, 2008; Maeshima *et al.*, 2014). Chromatin with an extended and loose arrangement of nucleosome was also detected in the nuclei of transcriptionally yeast using polymer modelling and chromosome conformation capture (3C) (Dekker, 2008). However *In vivo* 30 nm fibres have been identfied in terminally differentated cells such as chicken erythrocytes in which transcription is almost completely silenced (Woodcock, 1994). These studies therefore indicate 30 nm fibres are rare inside transcriptionally active cells. How nucleosome fibres behave dynamically in cells is an ongoing research topic.

The compaction of DNA into chromatin restricts regulatory processes such as replication, recombination, transcription and repair by blocking DNA binding sites or because the bent DNA is too distorted for binding. Three well-defined processes regulate and organise the chromatin environment: histone post-translational modification, DNA cytosine methylation and ATP-dependent chromatin remodelling (Ho and Crabtree, 2010). These changes alter the recruitment and action of regulatory proteins and are dependent on one another for proper regulation of the chromatin environment and therefore the genetic regulation of the cell.

1.1.2| Histone tail modifications

The core histones have flexible N- and C-terminal tails which extend from the DNA superhelix and are sites of PTMs (Luger *et al.*, 1997; Luger and Collins, 2001). A number of PTMs have also been found on the histone folds and domains required for histone:DNA and histone:histone contacts (Cosgrove, Boeke and Wolberger, 2004; Cosgrove and Wolberger, 2005).

The PTMs that have been identified on histones include phosphorylation of serine, threonine, tyrosine side-chain oxygens, lysine ubiquitination, lysine sumoylation, lysine acetylation, lysine and arginine methylation and ADP-ribosylation (Bannister and Kouzarides, 2011; Rothbart and Strahl, 2014).

PTMs are regulated through different proteins and protein complexes. These include 'writers' to establish the histone PTM and 'erasers' to remove the histone PTM. The first example of a writer was 55-kDa polypeptide (p55) Histone AcetylTransferase (HAT) and the eraser was Histone Deacetylase 1 (HDAC), both of which are transcription-associated proteins (Brownell *et al.*, 1996; Taunton, Hassig and Schreiber, 1996).

Histone PTMs influence gene regulation through the recruitment of 'readers' which recognise and bind the modification. Certain chromatin-associated proteins possess domains to bind to specific modifications: for example, the bromodomain binds preferentially to acetylated lysine (Dhalluin *et al.*, 1999; Bannister and Kouzarides, 2011).

Histone PTMs are epigenetic marks and are involved in numerous processes from regulating transcription to DNA repair (Rothbart and Strahl, 2014). For example H3 lysine 4 tri-methylation (H3K4me3) enrichment co-occurs with active transcription, whilst H3K27me3 is a dominant mark of repressed transcription (Voigt, Tee and Reinberg, 2013; Rothbart and Strahl, 2014). Chromatin acetylation by HATs generates the transcriptionally active euchromatin, whilst chromatin deacetylation by HDACs generates transcriptionally repressed heterochromatin (Eberharter and Becker, 2002).

1.1.3| DNA cytosine methylation

DNA cytosine methylation is the formation of 5-methylcytosine (5mC) through the addition of a methyl (CH_3) group to the C5 position of the cytosine pyrimidine ring and typically occurs at CpG dinucleotides (Razin and Riggs, 1980; Bird, 1986). 60-80% of CpG dinucleotides are methylated, however, clusters of the dinucleotides (CpG islands) are typically unmethylated (Ehrlich *et al.*, 1982). In mammals DNA methylation is performed by three active DNA cytosine methyltransferases (DNMTs). DNMT1 is a maintenance methyltransferase and preserves pre-existing methylation patterns through methylation of hemimethylated DNA (Stein *et al.*, 1982; Bestor *et al.*, 1988). DNMT3A and DNMT3B are *de-novo* methyltransferases and create new methylation patterns (Okano *et al.*, 1999). DNMTs are essential for normal development (Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999).

DNA methylation of CpG dinucleotides at promoter regions often results in gene repression through the recruitment of methyl-CpG-binding-domain proteins (MBPs). For example Methyl-CpG-binding protein 2 (MeCP2) recruits the SIN3 histone deacetylase complex resulting in the compaction of chromatin (Jones *et al.*, 1998; Tajima and Suetake, 1998; Bird and Wolffe, 1999; Lim and Maher, 2010). DNA methylation regulation is important in genomic imprinting (Jaenisch 1997, Bird & Wolffe 1999), X-linked inactivation (Bird, 2002), transposable element silencing (Walsh, Chaillet and Bestor, 1998) and cancer (Robertson et al. 1999, Hansen et al. 2011).

The domain architecture and their functions of the *de novo* DNA methyltransferases DNMT3A and DNMT3B are shown and explained in Figure 1.2.

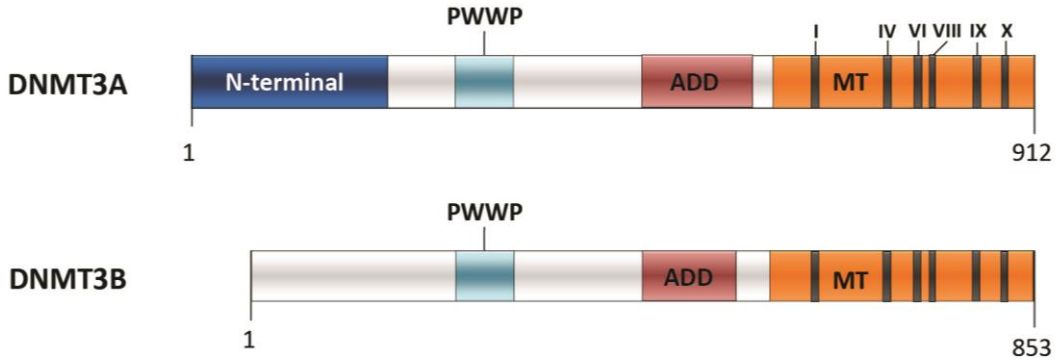

Figure 1.2| Domain architecture of the mammalian *de novo* DNA methyltransferases DNMT3A and DNMT3B

The N-terminal domain of DNMT3A exhibits DNA binding capability (Suetake et al. 2011). It also interacts with Euchromatic histone-lysine N-methyltransferase 2 (EHMT2 – also known as G9A) and/or EHMT1 (also known as GLP) and is necessary for DNA methylation at transcription factor promoter to regulate embryonic stem cell pluripotency (Chang et al. 2011; Epsztejn-Litman et al. 2008).

The Pro-Tryp-Tryp-Pro (PWWP) domain has numerous functions including heterochromatin binding (Chen et al. 2004), DNA binding (Purdy et al. 2010) and binding to the H3 modifications – H3K9me3 and H3K36me3 (Dhayalan et al. 2010). MT is the catalytic methyltransferase domain; and I, IV, VI, VIII, IX and X are motifs in the catalytic domain. Motifs I and X are required for bindng the methyl group donor S-adenosyl methionine (SAM). Motifs VIII and IX are required for DNA binding. Motifs IV, VI and VIII are for the catalysis of DNA methylation (Jurkowska et al. 2011).

The ATRX-DNMT3-DNMTL (ADD) domain binds has been shown to bind the MT domain inhibiting activity and is released upon binding to unmodified H3 tails with preferential DNA methylation expected at these site due to enhanced DNA methyltranferase activity (Zhang et al. 2010; Guo et al. 2015). The ADD domain also binds to enhancer of zeste homolog 2 (EZH2) a histone-lysine N-methyltranserase part of the polycomb repressor complex 2 (PRC2) which is required for H3K27 methylation (Rush et al. 2009; Viré et al. 2005).

Adapted from (Yang et al 2015). This figure is to scale.

DNA cytosine methylation has been shown to occur at the DNA accessible linker regions in chromatin as nucleosomes prevent DNA cytosine methylation (Felle *et al.*, 2011). Therefore, other enzymes which organise chromatin architecture are essential for the ability of DNMTs to methylate specific regions.

1.1.4| Long non-coding RNAs and Histone chaperones

Other, additional factors involved in regulating chromatin structure will be described briefly. Long non-coding RNAs (lncRNAs) are longer than 200 nucleotides, lack an ORF and are often polyadenylated (Fatica and Bozzoni, 2014). Several lncRNAs have been shown to escort chromatin modifiers such as the DNA methyltransferase DNMT3A and histone methyltransferase complex PRC2 to genomic loci on chromatin (Rinn *et al.*, 2007; Zhao *et al.*, 2008; Guttman and Rinn, 2012; Fatica and Bozzoni, 2014). The lncRNA X-inactive specific transcript (XIST) has been found to be directly involved in repressive chromatin formation (Fatica and Bozzoni, 2014).

Histone chaperones are proteins that bind to histones to regulate the assembly of nucleosomes (Ransom, Dennehey and Tyler, 2010; Burgess and Zhang, 2014). The majority of histone chaperones bind to H2A-H2B or H3-H4 and have been categorised into these two classes. However, histone chaperones for H3 and H2A variants likely also exist (Tagami *et al.*, 2004; Burgess and Zhang, 2014). Histone chaperones can participate in nucleosome assembly by shuttling synthesized histones from the cytoplasm to the nucleus (Mosammaparast *et al.*, 2002; Campos *et al.*, 2010); acting as a histone reservoir and regulating histone supply (Groth *et al.*, 2007; Cook *et al.*, 2011); bridging histones and histone acetyltransferases (Parthun, Widom and Gottschling, 1996; Han *et al.*, 2007); or directing histones onto DNA for nucleosome assembly (Tagami *et al.*, 2004).

Chromatin architecture is organised predominantly by ATP-dependent chromatin remodelers which are typically recruited to chromatin through binding directly to epigenetic marks or through recruitment of epigenetic modification binding proteins, which will be discussed further in section 1.2.6.

1.1.5| ATP-dependent chromatin remodelers

ATP-dependent chromatin remodelers either individually or as part of a complex hydrolyse ATP to alter the structure of chromatin. Chromatin remodelers or remodelling complexes can expose DNA through nucleosome sliding, nucleosome eviction or nucleosome unwrapping for the access of chromatin modifying proteins. Certain chromatin remodelers are able to alter the histone octamer through histone dimer eviction or histone dimer exchange (discussed in 1.2.6). These processes are outlined in Figure 1.3 (Clapier & Cairns, 2009).

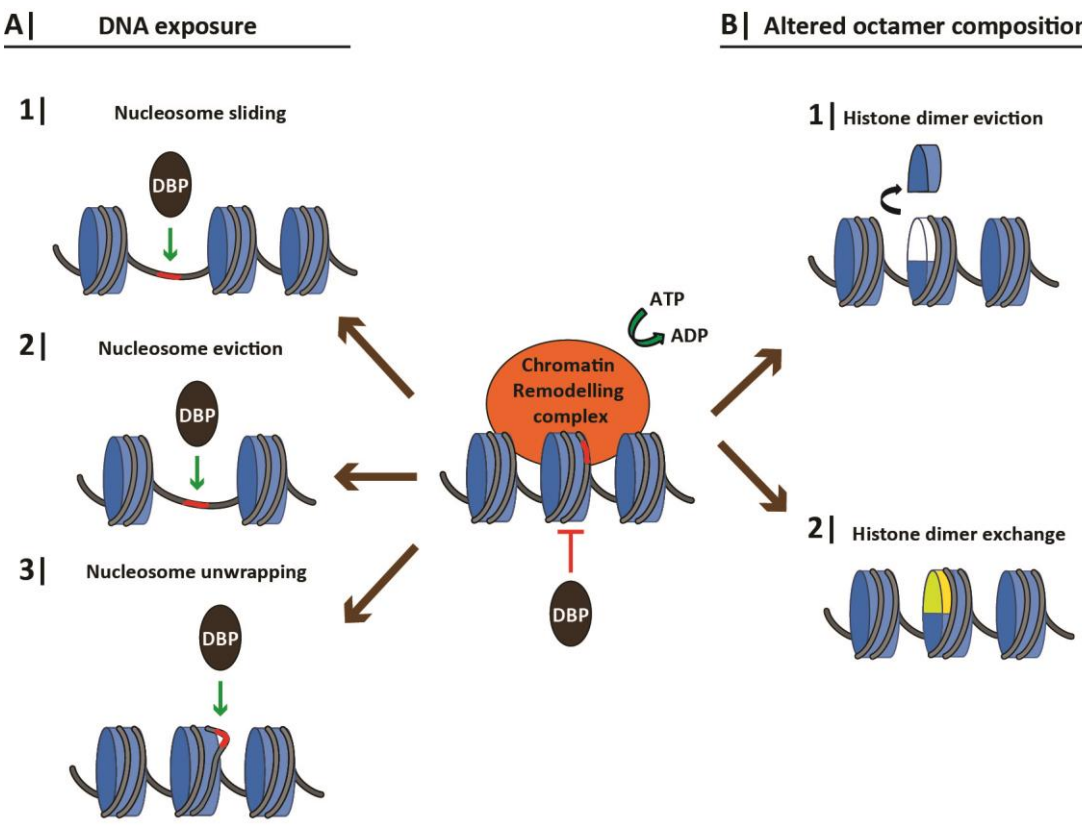

Figure 1.3| The alterations chromatin remodelers or remodelling complexes make to chromatin

A| The DNA on the nucleosome (red segment) is inaccessible to DNA binding proteins (DBP - brown oval), therefore three possible actions expose DNA for binding.
1| Nucleosome sliding makes DNA accessible through individual nucleosome repositioning.
2| Nucleosome eviction makes previously histone octamer DNA accessible for the DBP.
3| Nucleosome unwrapping involves destabilises certain histone-DNA contacts making DNA accessible to the DBP.
B| ATP-dependent chromatin remodelling can also alter histone octamer composition altering chromatin dynamics by:
1| Histone dimer eviction.
2| Histone dimer exchange, for example H2A-H2B dimer with a histone variant H2A.X-H2B (yellow).
Adapted from (Clapier & Cairns 2009).

1.2| Structure and function of ATP-dependent chromatin remodelers

1.2.1| Groupings of ATP-dependent chromatin remodelers

The majority of chromatin remodelers are genetically non-redundant, with severe consequences for embryo development when the encoding genes are mutated (Ho & Crabtree 2010). All chromatin remodelers have shared architectures and functions which include: 1) A conserved Sucrose Non Fermentable 2 (SNF2) family ATPase domain which couples DNA binding to ATP hydrolysis for DNA translocation. 2) Recognising and coupling to nucleosomes. 3) Domains that regulate the ATPase domain. 4) domains that interact with other chromatin regulatory proteins (Clapier and Cairns, 2009).

All chromatin remodelers contain a conserved SNF2 family ATPase region, which can be located at any site along the protein. The unique domains of ATPase chromatin remodelers are at flanking regions or between the SNF2 family ATPase region and confer specialised functions with certain PTMs, transcription factors, chromatin modifiers and nucleosomal DNA (Clapier & Cairns 2009, Flaus & Owen-Hughes 2011, Yodh, 2013).

Chromatin remodelers have been empirically classified into 4 groups based on the similarity of domains flanking the conserved ATPase region, which are necessary for conferring specific specialisations (Figure 1.4). The four groups include: SWItching defective/Sucrose Non-Fermenting (SWI/SNF), Imitation SWItch (ISWI), Chromodomain-Helicase-DNA binding (CHD) and INOsitol requiring 80 (INO80). However the current empirical categorisation of ATP-dependent chromatin remodelers does not take into account all the previous subfamilies recognised from a phylogenetic comparison of the SNF2 family ATPase region (Figure 1.4) (Flaus and Owen-Hughes, 2011).

The subfamilies LSH, ATRX and ALC1 cannot be categorised due to large areas of their flanking regions unable to be assigned a function using protein-domain finding tools. The flanking regions are unique in sequence and potentially in function therefore the categorisation in Figure 1.4 is an oversimplification. (Flaus and Owen-Hughes, 2011). Figure 1.4 is used to illustrate the empirical groupings for the illustration of the major characterised chromatin remodelers and the position of their unique flanking domains.

The application of Improved and specific protein-domain finding tools and structural through to biochemical investigations of unique domains are needed for the empirical characterisation of chromatin remodelers based on their full-length sequences.

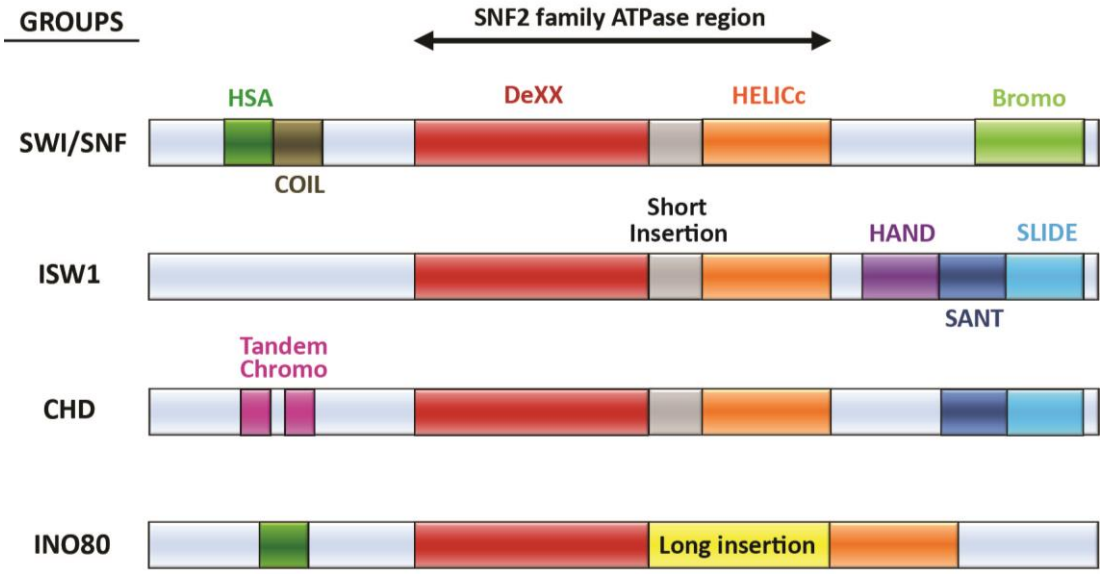

Figure 1.4| Empirical classifications of the 4 major groups of chromatin remodelers
The SNF2 family ATPase region contains an N-terminal DExx domain and C-terminal HELICc domain spaced by a linker region. The ATPase region is required for ATP hydrolysis, DNA binding and DNA translocation (Clapier & Cairns 2009).
The Helicase SANT associated (HSA) domain is a binding site for actin related proteins (Szerlong et al. 2008). Coiled coil domains are required for homo-dimerisation or binding to other coiled coils for protein-protein interactions (Tang et al. 2010). The bromo domain is a binding site for acetylated lysine on histone tails (Zeng & Zhou 2002). The HAND domain binds to DNA/nucleosomes. The SANT domain binds to histone tails and DNA (Boyer et al. 2004). The SLIDE is a protein binding domains necessary for the formation of several remodelling complexes and binds DNA (Dong et al. 2013). The HAND-SANT-SLIDE (HSS) domains are crucial for chromatin sliding activity (Yamada et al. 2011). The tandem chromo domain binds lysine methylated H3 tails (Flanagan et al. 2005).
This figure is not to scale.

1.2.2| SWI/SNF remodelers

A SWI/SNF remodelling complex was first identified as a MDa complex in yeast (Côté *et al.*, 1994) and each catalytic remodelling ATPase forms complexes with 8-14 additional proteins (Clapier and Cairns, 2009). The catalytic remodelling ATPases of the complexes possess an N-terminal coiled coil, a central SNF2 ATPase region and a C-terminal bromodomain (Figure 1.4). Two SWI-SNF remodelling complexes exist in eukaryotes, SWI/SNF and Remodels the Structure of Chromatin (RSC) exist in yeast. Brg1 Associated Factors (BAF1) and PolyBromo-Associated BAF (PBAF) exist in humans (Clapier and Cairns, 2009). SWI/SNF remodelling complexes are able to eject or slide nucleosomes at numerous loci to facilitate gene regulation, replication, DNA repair and tumour suppression (Mohrmann & Verrijzer 2005, Clapier & Cairns 2009). SWI/SNF remodelling complexes have not been found to be involved in chromatin assembly (Clapier and Cairns, 2009).

1.2.3| ISWI remodelers

The ISWI remodelling complexes contain 2-4 proteins, with Nucleosome Remodelling Factor (NURF), CHRomatin Accessibility Complex (CHRAC) and ATP-dependent Chromatin assembly and remodelling Factor (ACF) complexes initially identified in *Drosophila melanogaster* (Tsukiyama and Wu, 1995; Ito *et al.*, 1997; Varga-Weisz *et al.*, 1997).The catalytic remodelling ATPase of the complex possesses a central SNF2 ATPase region and a C-terminal HSS domain necessary for histone tail and DNA binding (Figure 1.4) The additional subunits of the complexes include bromodomains and DNA binding motifs, which contribute to the diversity of each complex (Corona and Tamkun, 2004; Clapier and Cairns, 2009). For example, the ACF and CHRAC complexes regulate nucleosome spacing using sliding to assemble chromatin for transcription repression. In contrast, the NURF complex aids in activating RNA Polymerase II (RNAPII) through randomising nucleosome spacing (Stopka and Skoultchi, 2003; Corona and Tamkun, 2004; Clapier and Cairns, 2009). ISWI are involved in the regular spacing of nucleosomes (Gkikopoulos *et al.*, 2011; Pointner *et al.*, 2012). The ISWI remodelling complexes are involved in the replication of heterochromatin (Collins *et al.*, 2002), the initiation of replication (Vincent, Kwong and Tsukiyama, 2008) and DNA repair (Erdel and Rippe, 2011).

1.2.4| CHD remodelers

The CHD family remodelers can exist as individual monomers or in complex with 1-10 additional proteins. In higher eukaryotes, there are three CHD subfamilies comprising of CHD1-2, CHD3-4 and CHD5-9 (Marfella and Imbalzano 2007). CHD1-9 all possess an N-terminal tandem chromo domain and a central SNF2 ATPase region (Clapier and Cairns, 2009, Ryan et al., 2011) (Figure 1.4). Each subfamily harbours specific flanking domains or motifs. CHD1-2 possess a C-terminal DNA binding domain (DBD) (Marfella and Imbalzano 2007). CHD3-4 possess a C-terminal Plant Homeo-Domain (PHD), which interacts with methyl-lysine on histone tails (Clapier and Cairns, 2009, Marfella and Imbalzano 2007). CHD3-4 form common human protein complex containing a CHD is the Nucleosome Remodeling and Deacetylase (NuRD) complex (Marfella and Imbalzano, 2007; Clapier and Cairns, 2009). Proteins of CHD5-9 contain an C-terminal SANT-SLIDE domain, a DNA binding domain and a paired Brahma and Kismet domain (BRK), the function of which is specific to higher eukaryotes (Marfella and Imbalzano). Each CHD subfamily member is required for a specific cellular function ranging from transcription activation or repression to development and differentiation (Marfella and Imbalzano 2007. Like ISWI, CHD remodelers slide nucleosomes for the regular positioning of nucleosomes (Gkikopoulos *et al.*, 2011; Pointner *et al.*, 2012) and eject nucleosomes, which is important for promoting transcription (Marfella and Imbalzano, 2007; Clapier and Cairns, 2009). CHD complexes such as NuRD (Nucleosome Remodeling Deacetylase) contain additional subunits with DNA binding domains and histone tail binding domains and methyl-CpG-binding-domain (MBD) and HDAC1/2 necessary for its repressive role in gene expression (Marfella & Imbalzano 2007; Clapier & Cairns 2009; Reynolds et al. 2012).

1.2.5| INO80 remodelers

The INO80 remodelling complexes comprise 10 proteins, and includes the Sick With Rat8 (SWR1) complexes, first identified in *S. cerevisiae* (Bao and Shen, 2007; Clapier and Cairns, 2009). The human complexes include INO80, SNF2 Related CREB-Activator Protein (SRCAP) and the transactivator of transcription interactive protein 60 kDa (TIP60) in complex with p400. TIP60 is a HAT, whereas p400 is the remodelling ATPase and histone variant exchanger (Clapier & Cairns 2009, Gévry *et al.*, 2007; Courilleau *et al.*, 2012). The INO80 catalytic remodelling ATPases have an N-terminal HSA domain and a central ATPase domain which has a large insertion splitting the ATPase region.

This is necessary for binding of the AAA+ATPase helicases RuVB-like 1 and 2 (Rvb1/Rvb2) and ARP5 (Jónsson *et al.*, 2004; Bao and Shen, 2007; Clapier and Cairns, 2009). The INO80 remodelling complex removes H2A.Z:H2B dimers from nucleosomes and replaces with H2A:H2B dimers, whereas the SWR1 remodelling complex performs the reverse (Mizuguchi *et al.*, 2004; Papamichos-Chronakis *et al.*, 2011). The INO80 remodelling complex is the only remodeler possessing ATP-dependent 3'-5' helicase activity. INO80 can slide, evict and regularly space nucleosomes (Shen *et al.*, 2000; Udugama, Sabri and Bartholomew, 2011). INO80 is involved in transcription (Cai *et al.*, 2007) and replication to ensure proper chromosome segregation (Papamichos-Chronakis and Peterson, 2008; Hur *et al.*, 2010) and DNA repair, through binding to the DNA damage signal γH2A.X (Van Attikum *et al.*, 2004).

1.2.6| Recruitment of Chromatin remodelers to chromatin

An *in vitro* study compared 7 chromatin remodelers (ACF, ISWI, Snf2H, Chd1, Mi-2, Brg1 and NURF) on different nucleosomal templates and found that each remodeler repositioned the nucleosomes at distinct positions (Rippe *et al.*, 2007). An additional genome wide chromatin immunoprecipitation (ChIP) study found the drosophila chromatin remodelers ISWI, NuRD, INO80 and (P)BAP each targeted a unique set of genomic targets at distinct positions on chromatin (Moshkin *et al.*, 2012).

The DNA binding motifs on certain chromatin remodelers such as the HSS domain in ISWI or the DBD in CHD can influence the outcome of the remodelling reaction (Stockdale *et al.*, 2006; Partensky and Narlikar, 2009; van Vugt *et al.*, 2009). This was shown by switching domains between remodelers to alter the repositioning of nucleosomes (Stockdale *et al.*, 2006; Partensky and Narlikar, 2009; van Vugt *et al.*, 2009). This is likely due to the different affinities of these motifs for DNA before and after repositioning of the nucleosome (Manelyte and Längst, 2013). These DNA binding domains also influence internucleosomal distances with the remodelers ACF and CHD1 being shown to regularly space nucleosomes *in vitro* (Yang *et al.*, 2006; Jeffrey N McKnight *et al.*, 2011). DNA G-rich repeats which are found in telomeres are recognised by the chromatin remodeler ATRX, which preferentially binds G-quadruplexes *in vitro* (Law *et al.*, 2010) which could influence recruitment to these regions.

Chromatin remodelers on their own do not have domains specific for methylated CpG DNA (Manelyte and Längst, 2013). However, the NuRD chromatin remodelling complex contains a methyl binding domain protein (MBD2) which recruits NuRD to methylated promoters (Zhang *et al.*, 1999). Therefore, epigenetic marks influence recruitment of chromatin remodelers.

Chromatin remodelers and remodelling complexes contain histone tail modification interacting domains (discussed in section 1.1.2), allowing them to target specific chromatin environments (Manelyte and Längst, 2013). For example, the bromodomains of SWI/SNF and RSC interact specifically with tetraacetylated-lysines on H3 tails and possess increased affinity for nucleosomes containing these histone modifications (Chatterjee *et al.*, 2011). In comparison, the SANT domains of the ISWI family of remodelers interact specifically with unmodified H4 tails. Removal of the H4 tail or acetylation of H4K16 has been shown to reduce the remodelling activity of ISWI *in vitro* (Dang, Kagalwala and Bartholomew, 2006).

As a further example the PHD domain of CHD4 preferentially binds to H3K4 and H3K9me3 (Mansfield *et al.*, 2011). Alternative isoforms of NURF with and without its C-terminal bromodomain and PHD finger have been found in cells suggesting alternative splicing can alter localisation of remodelling complexes in the genome (Kwon *et al.*, 2009).

Apart from possessing domains which recognise histone modifications, some chromatin remodelers also possess specific affinity and activity to histone variants and can exchange histone variants with action from DNA and RNA polymerase for histone displacement from DNA (Talbert and Henikoff, 2010b; Manelyte and Längst, 2013). A well-studied example is how the SWR1 complex exchanges H2A.Z for H2A nucleosomes. ATPase activity is stimulated in the presence of nucleosomes containing only H2A, without H2A exchange. The ATPase activity of SWR1 is increased upon the presence of H2AZ/H2B dimers, alongside which there is exchange of one nucleosomal H2A/H2B dimer for one H2A.X/H2B dimer. A second exchange step follows to complete a homotypic H2A.Z nucleosome (Luk *et al.*, 2010). Other notable examples of histone variant exchange include ATRX which has been shown to recruit death domain associated protein (DAXX) to telomeres where both subunits deposit H3.3 at telomeric chromatin (Shechter *et al.*, 2009). H3.3 incorporation in the *drosophila* male pronucleus by CHD1 which enables the paternal genome to participate in zygotic mitosis (Konev *et al.*, 2007).

The transient interactions of chromatin remodelers with sequence specific DNA binding proteins can also provide a method of recruitment which does not require epigenetic binding motifs of the remodelers themselves (Manelyte and Längst, 2013). For example the NuRD complex is directly recruited to the promoters of target genes by the transcription factors NAB2, BCL11B and Ikaros (Murawska and Brehm, 2011). Whilst complexes containing BRG1 are recruited to two target genes in Schwann cells via the transcription factor Sox10 (Weider *et al.*, 2012). *In vitro* experiments have shown certain transcription factors can also halt the remodelling activity of one enzyme (ISWIa) whilst leave another unimpeded (SWI/SNF) (Li *et al.*, 2015).

1.2.7| ATP-dependent chromatin groupings and subfamilies based on phylogenetic analysis of the SNF2 family ATPase region

A clearer categorisation of all the known ATP-dependent chromatin remodelers has been performed through the characterisation of the conserved SNF2 family ATPase region enabling classification into phylogenetic groupings and subfamilies of all known chromatin remodelers, which are shown in Figure 1.5.

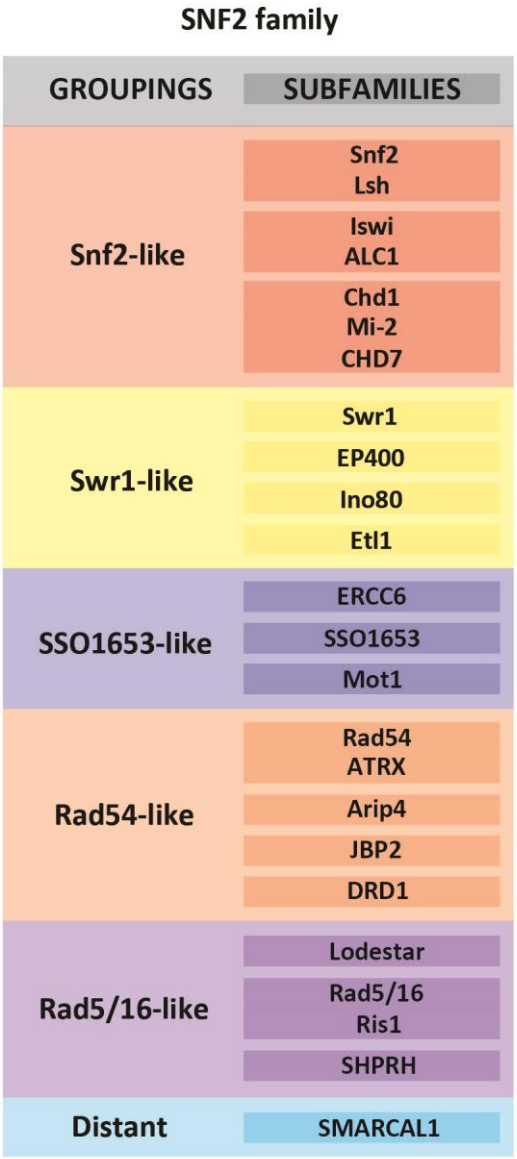

Figure 1.5| Groupings and subfamilies of ATP dependent chromatin remodelers defined by phylogenomic comparison of the SNF2 family ATPase region
Remodelers closer together and without gaps are phylogenetically closer than others in the same subfamily (Flaus et al. 2006). Adapted from (Flaus & Owen-Hughes 2011).

ATP-dependent chromatin remodelers function through the use of the SNF2 ATPase region for DNA/nucleosome binding, ATP binding and hydrolysis and DNA translocation as well as additional function/s dependent on flanking domains. The conserved structural features of the SNF2 family ATPase region will be explained in the next section.

1.3| Functional and structural features of chromatin remodelers

1.3.1| Conserved functional and structural features of ATP-dependent chromatin remodelers

Within the Snf2 family ATPase region are the SNF2_N and Helicase_C domains which both contain RecA domain lobes 1 and 2 respectively and are defined by 7 helicase-related sequence conserved motifs (Figure 1.6). Although all chromatin remodelers possess these motifs they are not legitimate helicases; they are only required for DNA binding and translocation, but not for unwinding double-stranded DNA (Dürr et al., 2006). The RecA domain lobes are responsible for ATP hydrolysis to produce the energy required for DNA translocation and for DNA binding (Singleton, Dillingham and Wigley, 2007).

The Snf2 family ATPase region of zebrafish RAD54 was the first eukaryotic chromatin remodeler to have its crystal structure solved (Thomä *et al.*, 2005). The structure of the helical motifs is conserved to the crystal structures of the Plasmid copy reduced A (PcrA) SF1 helicase with a single strand tailed DNA duplex captured in the transition states of ATP hydrolysis (Velankar *et al.*, 1999). Helical motifs are also conserved in the structure of Drosophila melanogaster Vasa SF2 helicase in complex with single-stranded RNA and AMPPNP (Sengoku *et al.*, 2006). The PcrA and Vasa structures have been used to suggest how ATP hydrolysis is linked to DNA translocation in helicases (Mazin *et al.*, 2010).

ATP (or AMPPNP) is bound by motif I (Walker A), II (Walker B) and VI (Velankar et al. 1999, Sengoku et al. 2006). The magnesium binding site is occupied by the lysine residue of motif I in the apo-enzyme. This lysine residue is displaced upon binding of an ATP-Mg^{2+} complex and forms a hydrogen bond with the ATP-β-phosphate. The magnesium ion is coordinated by the threonine of motif I and aspartate of motif II, enabling a water molecule that hydrolyses the bond between the β and γ- phosphates to be activated by the glutamate of motif II (Velankar *et al.*, 1999). This hydrolysis event is sensed by the two arginines of motif VI through binding of the phosphate tail of ATP in the ATPase active site and transmits the hydrolysis induced conformational changes to the DNA binding site resulting in DNA translocation (Velankar et al. 1999, Sengoku et al. 2006).

How dsDNA binds to the ATPase region has been elucidated from a crystal structure of the SNF2 ATPase region of the *Sulfolobus solfactaricus* RAD54 homolog in complex with dsDNA (Dürr *et al.*, 2005). Firstly dsDNA binds along the cleft between the RecA domain lobes where structural changes through ATP hydrolysis could result in translocase activity (Dürr *et al.*, 2005). Predominant binding of dsDNA occurs on RecA domain lobe 1 through recognition of the two phosphate chains along the minor groove (Dürr *et al.*, 2005). The dsDNA has minimal contacts with RecA domain lobe 2 which is analogous to other SF2 helicases (Dürr *et al.*, 2005). Instead of strand separation, a translocation of the dsDNA along the ATPase region was proposed, of which the forces generated break histone:DNA contacts which would be necessary when sliding nucleosomes (Dürr *et al.*, 2005). This has been supported by biochemical data (Ristic *et al.*, 2001) and in other remodelers including RSC (Saha et al. 2002, Lia et al. 2006) and ISWI (Whitehouse *et al.*, 2003a).

The SNF2 family ATPase region is also defined by 15 conserved boxes which are short amino acid sequences (2-19 aa) necessary for stabilising structural features (Figure 1.6) (Flaus *et al.*, 2006). The major difference between SNF2 family members and SF2 members is the extended sequence between the Rec A domain lobes (Flaus and Owen-Hughes, 2001). There are between 153-626 residues for SNF2 family members (subfamily dependent) instead of the 38-78 residues for SF2 family members (Flaus and Owen-Hughes 2001, Flaus et al. 2006) The extended sequence between the spherical RecA domain lobes make up the helical protrusions separated by a flexible linker (Figure 1.6A) which are shown on the zebrafish RAD54 SNF2 ATPase region crystal structure (Figure 1.6B). The sequences at the bases of the protrusions are conserved in boxes H, B, J, C and K and attach the protrusions to the RecA domain lobes (Flaus et al. 2006, Flaus & Owen-Hughes 2011). Box B is within the linker region which passes between the RecA domain lobes and contains a pair of conserved arginines across all subfamilies and are essential for function (Richmond & Peterson 1996, Flaus et al. 2006, Flaus & Owen-Hughes 2011). Behind the second RecA domain lobe is the major insertion region which can be relatively short with 34 residues in RAD54 or long with 280 residues in INO80 with the long insertion providing a protein binding platform (Flaus et al. 2006, Flaus & Owen-Hughes 2011) (see section 1.2.5). Boxes M and N are highly conserved charged regions which make up the C-terminal brace anchored at the base of protrusion 2 (Figure 1.6A and Figure 1.6B) (Flaus et al. 2006, Flaus & Owen-Hughes 2011).

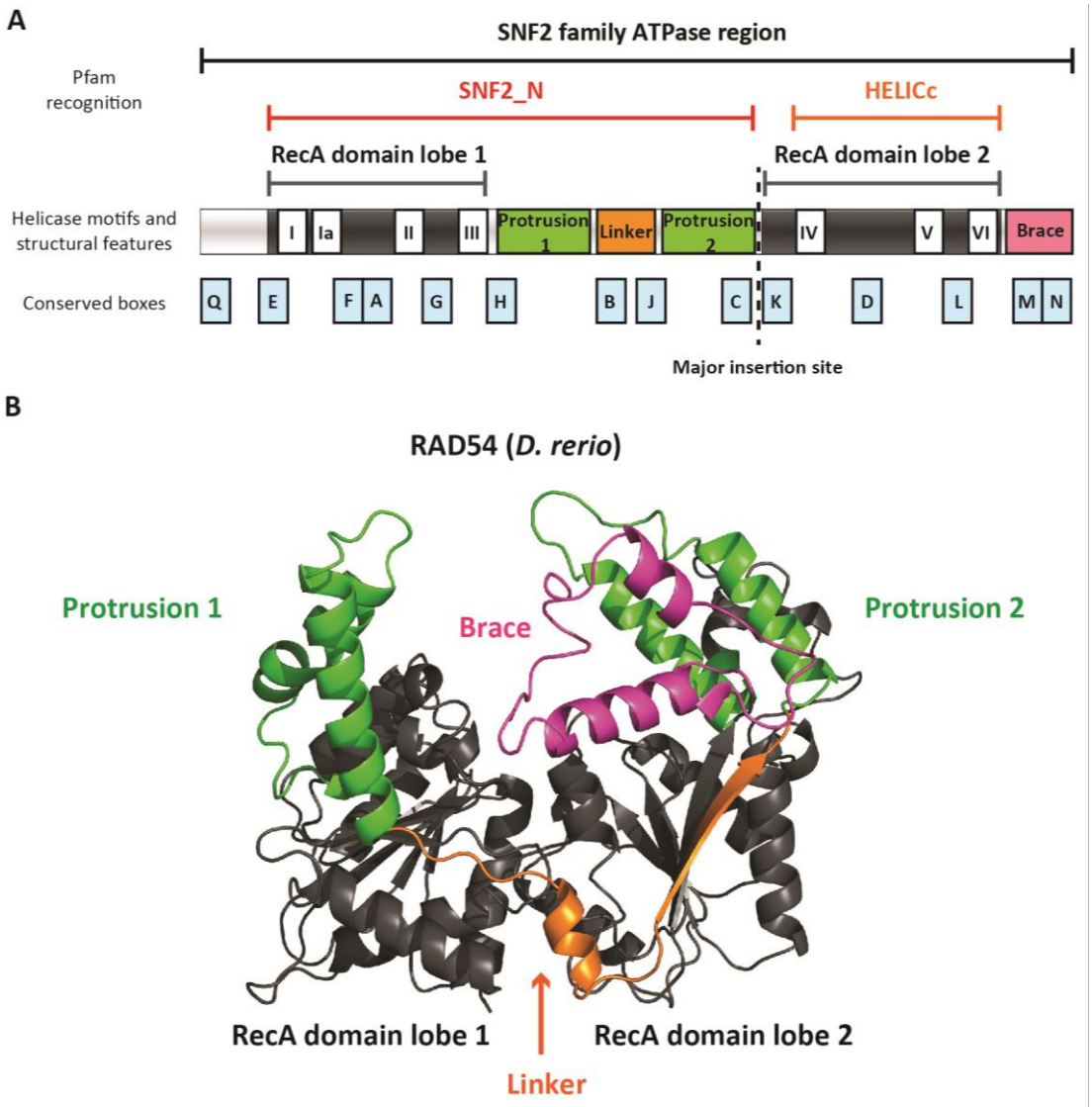

Figure 1.6| SNF2 family ATPase region architecture and structural features represented on the Zebrafish RAD54 ATPase region.
A| Helicase motifs, structural features and conserved boxes are shown relative to RecA domain lobes 1 and 2 (Flaus et al. 2006). The figure is not to scale. Adapted from (Flaus & Owen-Hughes 2011).
B| The SNF2 family ATPase region of Zebrafish RAD54 (PDB - 1Z3I) (Thomä et al. 2005) is shown with structural features including RecA domain lobes 1 and 2 (grey), Protrusion 1 and 2 (green), the linker between the protrusions (orange) and the brace (purple).

Protrusions 1 and 2 did not show a function for DNA binding through modelling DNA to the RAD54 structure (Flaus and Owen-Hughes, 2011). However the structure of the tandem chromodomains and SNF2 family ATPase region of Chd1 from *S. cerevisiae* (Hauk *et al.*, 2010) has provided structural and biochemical evidence that the tandem chromodomains contact the protrusions blocking access of the DNA to its binding site on the ATPase region (Figure 1.7B) (Hauk et al. 2010, Flaus & Owen-Hughes 2011). This finding provided evidence that structurally independent domains flanking the ATPase region are inhibitors of DNA-dependent ATP hydrolysis, thereby preventing DNA translocation. Therefore the allosteric movement of the flanking domains is required for activating DNA-dependent ATP hydrolysis, resulting in DNA translocation (Hauk & Bowman 2011, Flaus & Owen-Hughes 2011).

1.3.2| Structures of the ATPase region of eukaryotic ATP-dependent chromatin remodelers

As shown earlier in section 1.2.1 and 1.3.1 the regions flanking the SNF2 family ATPase region are important for regulating ATPase activity and DNA translocation or binding regulatory proteins, DNA or modified histone tails. The structural diversity of the regions directly flanking the SNF2 family ATPase region are very different as shown in the structures of the SNF2 ATPase region of RAD54 from *D. rerio* (Thomä *et al.*, 2005), the chromodomains - SNF2 ATPase region of Chd1 from *S. cerevisiae* (Hauk *et al.*, 2010), and the HSA - SNF2 ATPase region from *T. thermituga* (Xia *et al.*, 2016) (Figure 1.7). For simplification, I have represented the two major SNF2 ATPase region lobes as 1 and 2 (1.7A). The flanking N and C-terminal domains are also shown (Figure 1.7).

ATPase lobes 1 and 2 are conserved between the three crystal structures (Figure 1.7B). The only difference is the orientation of ATPase lobe 2, due to differences in packing as a result of the flanking domains. (Figure 1.7B). The N-terminal domain (NTD) of RAD54 is unique among its orthologs and is mostly unstructured, which has been shown to contain a region necessary for binding RAD51 (Jiang *et al.*, 1996; Golub *et al.*, 1997; Raschle *et al.*, 2004). The C-terminal domain (CTD) of RAD54 is unique to other subfamilies of chromatin remodelers as it contains a zinc-coordinating motif which is speculated to stabilise the assembly of RAD54 (Thomä et al. 2005, Mazin et al. 2010). The NTD packs with ATPase lobe 1 and the CTD packs with ATPase lobe 2.

The recent structure of SNF2 from *M. thermophila* shows the NTD (a portion of HSA) stacked against ATPase lobe 1 only. HSA binds actin-related proteins but has also been shown to bind DNA (Dechassa *et al.*, 2012; Schubert *et al.*, 2013). Complementary biochemical studies showed that the HSA domain reduces the remodelling activity more than 30 fold, which indicates that it decouples ATP hydrolysis from remodelling (Xia et al. 2016). The CTD of SNF2 is a specific SNF2 ATP coupling (SnAC) domain. The SnAC domain is seen close to the surface of ATPase lobe 2, with no contact to ATPase lobe 1 (Xia *et al.*, 2016). The SnAC domain is required for coupling ATP hydrolysis to nucleosome remodelling through action as a histone anchor (Sen *et al.*, 2011, 2013). SNF2 is kept in an inactive conformation through direct interaction with ATPase lobe 1 and lobe 2 which blocks the two arginines of motif VI from the ATP binding site (Xia et al. 2016). It is suggested the NTD and CTD of SNF2 negatively and positively couple ATP hydrolysis to remodelling respectively, and likely allosterically alter the orientation and interface between ATPase 1 and 2 (Xia *et al.*, 2016).

In contrast, the NTD (Tandem chromodomain) and CTD (C-term bridge) of CHD1 separate ATPase lobe 1 and lobe 2 (Figure 1.7B). The ATPase lobes are too far apart from one another for efficient ATP hydrolysis, as the two arginines of motif VI are not close enough to bind the phosphate tail of the ATPγS nucleotide (Figure 1.7B) (Hauk et al., 2010). As explained earlier, the tandem chromodomains obstruct DNA from binding to the ATPase region (Hauk et al., 2010, Flaus & Owen-Hughes, 2011). A C-term bridge crosses the ATPase lobe 2 and packs against the ATPase lobe 1 (Figure 1.7B) and may influence motions of the ATPase lobes during DNA binding and translocation, as ATPase lobe 2 would need to swivel 52° to close the ATPase cleft for efficient ATP hydrolysis (Hauk *et al.*, 2010).

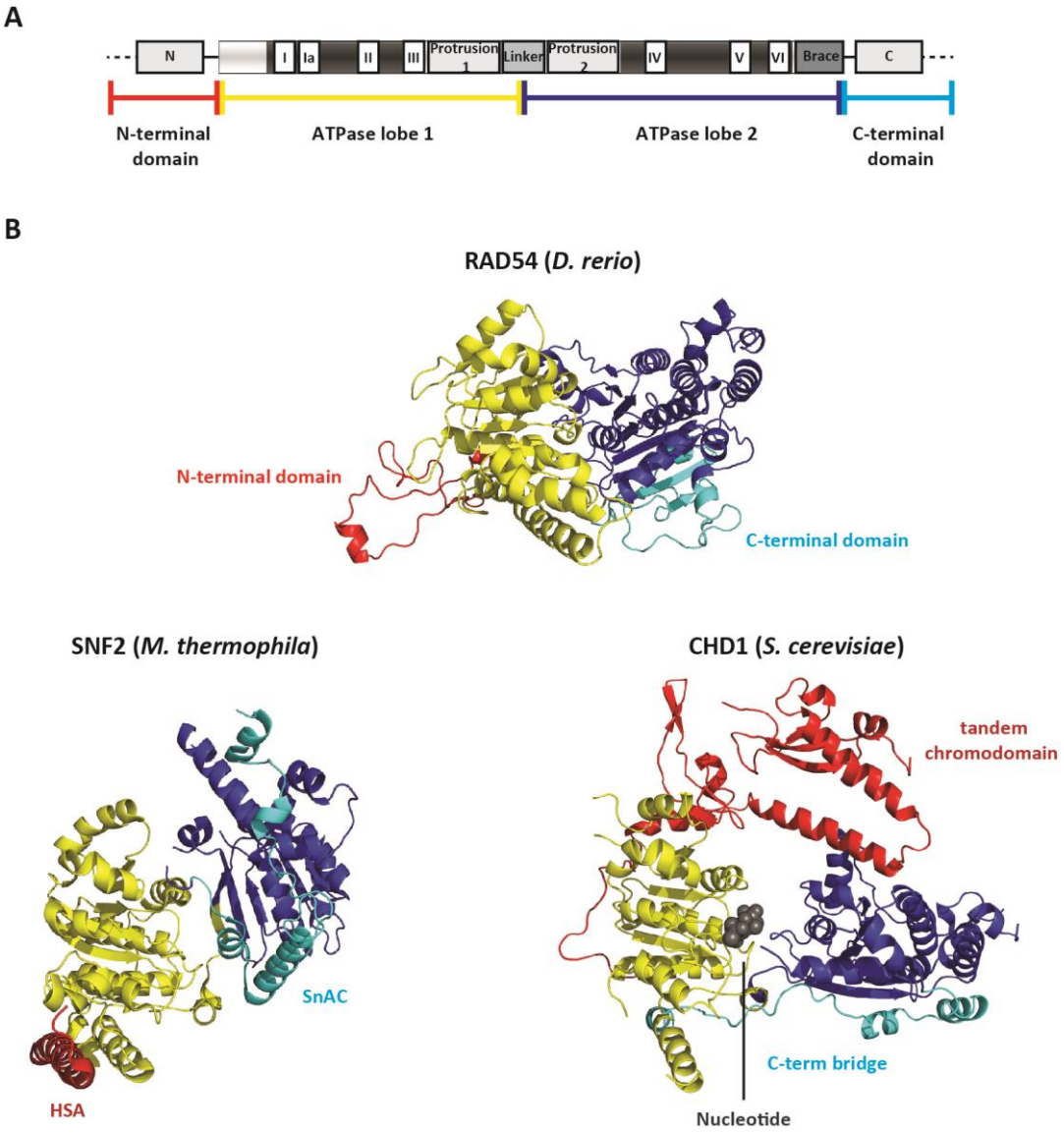

Figure 1.7| Structures of the ATPase region and flanking domains of the eukaryotic ATP-dependent chromatin remodelers RAD54, SNF2 and CHD1

A| Helicase motifs and structural features with the simplified structural domains defined as the N-terminal domain, ATPase lobes 1 and 2 and C-terminal domain shown below.

B| The N-terminal domain for RAD54 (are shown in red (HSA for SNF2 and the Tandem chromodomain for CHD1). ATPase lobe 1 is yellow and ATPase lobe 2 is blue for all structures. The C-terminal domain for RAD54 (SnaC for SNF2 and C-term bridge for CHD1) are shown in cyan. The nucleotide is ATPγS.

RAD54 (PDB - 1Z3I) (Thomä et al. 2005).

SNF2 (PDB-5HZR) (Xia et al. 2016).

CHD1 (PDB-3MWY) (Hauk et al. 2010).

All chromatin remodelers possess functional domains closely flanking the ATPase region which allosterically regulate and coordinate ATP hydrolysis and DNA translocation activity (Szerlong et al., 2008, Hauk et al., 2010, Flaus & Owen-Hughes, 2011).

Two structurally-independent flanking domains AutoN and NegC have been identified in *S. cerevisiae* ISW1 and flank the ATPase region at the N and C-termini respectively (Clapier & Cairns 2012, Manning & Peterson 2013, Mueller-Planitz et al. 2013).
The AutoN domain acts as a brake preventing DNA-stimulated ATP hydrolysis without the presence of the H4 tail. (Clapier & Cairns 2012, Manning & Peterson 2013). Therefore, the AutoN domain likely acts as a negative regulator of DNA-stimulated ATPase activity. (Manning & Peterson 2013)

The second allosteric domain, NegC is equivalent to the C-term bridge in the CHD1 structure (Figure 1.7) and is required to couple ATP hydrolysis to DNA translocation. Interestingly, a 'core' ATPase region not containing the AutoN, NegC or HSA domains is able to translocate DNA and remodel nucleosomes without the H4 tail, extranucleosomal DNA or binding to the SHL7 position (Mueller Planitz 2013). These are biochemical characteristics similar to the catalytic ATPases of SWI/SNF chromatin remodelers (Clapier & Cairns 2012, Manning & Peterson 2013, Mueller-Planitz et al. 2013).

The chromo-wedge in Chd1 (Figure 1.7B) has also been shown to possess an AutoN-like region as it contacts the 2nd ATPase lobe preventing DNA binding, thereby repressing ATPase activity (Manning & Peterson, 2013, Hauk et al. 2010, Hauk and Bowman, 2011). When the AutoN-like region in Chd1 is removed, the H4 tail is no longer necessary for DNA-stimulated ATPase activity (Hauk & Bowman 2011, Hauk et al. 2010, Manning & Peterson 2013). The AutoN and NegC domains are not conserved in the SWI-SNF family (Clapier and Cairns, 2012), making it specific to a set of remodelers and not all of them. The AutoN, NegC and SnAC domains may not be present in other chromatin remodelling families (Clapier & Cairns, 2009, Manning & Peterson, 2013).

Further investigations are required to biochemically and structurally define how the independent structural regions of ATP-dependent chromatin remodelers regulate their particular function *in vivo*. Further work is also required to characterise flanking domains which do not have any predicted or conserved function.

1.4| Lymphoid-specific helicase

This investigation focuses on the chromatin remodeler lymphoid-specific helicase (LSH), also known as; Helicase lymphoid-specific (HELLS), Proliferation Associated SNF2-like Gene (PASG) or SWI/SNF related, Matrix associated Actin-dependent Regulator of Chromatin, subfamily A, member 6 (SMARCA6) (Briones and Muegge, 2012). Human and mouse LSH are well conserved with >95% identical amino acid sequences. The only recognisable difference being an additional 16 amino acids presented on the N-terminus of human LSH.

1.4.1| *In vivo* functions of LSH

Murine LSH was first cloned from murine fetal thymus tissue using a degenerative PCR technique when searching for a helicase involved in Variable, Diverse and Joining (VDJ) recombination in early lymphocyte development (Jarvis *et al.*, 1996). LSH is prominently expressed in early thymocytes, T and B cells, hence the name lymphoid-specific helicase (Jarvis *et al.*, 1996; Geiman and Muegge, 2000).

Since the initial finding of LSH, the meaning of the name has been shown to be unrelated to initially perceived function. LSH is not lymphoid-specific as mRNA has been detected in many tissues at low levels. LSH is highly expressed in the embryo and all highly proliferative adult tissues (Raabe et al. 2001, Briones & Muegge 2012). LSH is unlikely to possess helicase activity as the helicase motifs conserved across the chromatin remodelling families do not enable helicase activity (Dürr et al., 2006).

Studies from the Muegge lab have shown that $LSH^{-/-}$ mice die after birth, with multiple defects identified including kidney necrosis, early senescence of fibroblasts, premature aging and reduced embryonal growth and aberrant gene expression of various genes (Dennis *et al.*, 2001; Fan *et al.*, 2003; Sun *et al.*, 2004; Xi *et al.*, 2007).

LSH is a nuclear protein associated with chromatin, specifically heterochromatin (Fan *et al.*, 2003). Microscopy experiments have shown that the recruitment of catalytically inactive ATPase LSH^{K237Q} to heterochromatin is not altered when compared to LSH (Lungu *et al.*, 2015). However, ATPase activity is required for the release of LSH from heterochromatin (Lungu *et al.*, 2015). A further study has shown nucleosome density in $LSH^{-/-}$ embryonic stem cells can only be recovered to wild-type levels by the re-expression of LSH with a functional ATPase, indicating that chromatin remodelling is a primary function of LSH (Ren *et al.*, 2015).

$LSH^{-/-}$ cells show 50-70 % reduction in cytosine DNA methylation identified at repeat elements; including both major and minor satellites, as well the retroviral elements LINEs (Long interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements). This results in increased transcription of these elements by transcription activating histone acetylation and H3K4me3 modifications creating genomic instability (Dennis et al. 2001; Fan et al. 2003; Sun et al. 2004; Tao, et al. 2011; Myant et al. 2011).

$LSH^{-/-}$ in mouse embryonic fibroblasts (MEFs) show loss of cytosine methylation at distinct genomic sites demonstrating that LSH is necessary for *de novo* DNA methylation (Myant et al., 2011, Tao et al., 2011). Recent studies have shown the ATPase activity of LSH is essential for *de novo* methylation at repeat sequences (Ren *et al.*, 2015; Termanis *et al.*, 2016).
These findings are supported by the interaction of LSH with DNMT3B from co-immunoprecipitation and pulldown studies (Zhu et al. 2006, Myant & Stancheva 2008). No primary interaction with DNMT1 is seen, suggesting LSH is important for the establishment of DNA methylation at genomic sites (Briones & Muegge 2012, Myant & Stancheva 2008). Changes in the histone modifications H3K27me3 and H2AK116 ubiquitylation are identified in $LSH^{-/-}$ cells suggesting LSH is involved in histone modifier regulation (Xi et al., 2007). This has been in part recognised with LSH recruiting DNMTs and HDACs in transcription silencing events (Myant & Stancheva, 2008).
Both LSH and DNMT3B have both been found to interact with the ES cell-specific chromatin remodelling complex (esBAF) (Ho *et al.*, 2009), therefore, more work on defining the complexes which contain LSH is required to characterise its numerous functions which likely change at different stages of development (Briones and Muegge, 2012).

1.4.2| LSH in cancer

LSH has been found to be upregulated and overexpressed in a variety of cancers including: head and neck squamous cell carcinoma (HNSCC) (Waseem *et al.*, 2010), melanoma (Ryu *et al.*, 2007), leukaemia (Lee et al. 2000), and lung adenocarcinoma cancer (Yano *et al.*, 2004; Wang *et al.*, 2015)

LSH was found as an important cofactor in Ras-induced senescence bypass and cell proliferation, as LSH knockdown in RΔN keratinocytes caused increased senescence and lack of maintainable cell proliferation (Keyes *et al.*, 2011). This study also provided evidence that LSH is necessary for senescence bypass in p63 proliferating cells and therefore necessary for tumourigenesis in keratinocytes (Keyes et al. 2011).

E2 factor transcription factor 3 (E2F3) is dependent on an interaction with the pRB (retinoblastoma) tumour suppressor protein for cell cycle arrest at the G1 checkpoint (Burkhart and Sage, 2008; van den Heuvel and Dyson, 2008). Interestingly, LSH expression increased in parallel with E2F3 in tumourigenesis and bound to 93% of the transcriptional start sites (TSS) bound by E2F3. 86% of these contained H3K4me3, a mark associated with active gene transcription, a result of LSH-E2F3 cooperation increasing the production of the H3K4 methyltransferase mixed lineage leukaemia 1 (MLL1) production in cancer cells. Thereby identifying LSH as an important regulator of E2F3 necessary for cell proliferation in cancer, with recent work showing upregulation of LSH is essential for retinoblastoma tumour progression through E2F signalling. (von Eyss et al. 2011, Benavente et al. 2014)

1.3.3| LSH in DNA repair

Aside from LSH interacting with cancer proliferation proteins, work by Burrage *et al* (2012) has shown LSH to be necessary for the repair of DNA double-strand breaks (DSBs) independent of DNA methylation. LSH KD cells failed to repair ~50% more DSB lesions than WT cells 8 hrs after irradiation, showing LSH is important for DSBR (Burrage et al. 2012).

DNA damage response (DDR) signalling is initiated by Ataxia Telangiectasia Mutated (ATM) a serine protein kinase autophosphorylated at serine 329 in response to DNA damage (Bakkenist and Kastan, 2003). ATM binds with the MRN (MRE11, RAD50, NBS1) complex which phosphorylates Serine 139 of histone H2AX to γH2AX (Lee and Paull, 2004). Phosphorylated H2A.X accumulates at flanking regions megabases away from the DSB producing a platform for the recruitment of DNA damage repair and mediator proteins and coordinates cell cycle arrest during double-strand break repair (Iacovoni et al. 2010, Polo and Jackson, 2011, Bakkenist and Kastan, 2003).

Both WT and LSH KD cells which underwent Ionising Radiation (IR) expressed equal levels of ATM, however, phosphorylation of the variant histone H2A.X was reduced by <50% in LSH knockdown cells compared to WT cells. This caused early and reduced recruitment of DNA damage response proteins Mediator of DNA Damage Checkpoint 1 (MDC1) and p53 Binding Protein 1 (53BP1) which localise at IR foci to initiate DNA repair. This reinforces the inefficient DNA repair and apoptosis in LSH KD cells (Burrage et al. 2012).

Interestingly, rescuing $LSH^{-/-}$ MEFs with WT LSH produced γH2AX levels and 53BPI to IR foci equivalent to WT cells, however, LSH^{K237Q} showed no WT phenotype suggesting that LSH ATPase activity is crucial for efficient H2A.X phosphorylation and DSBR (Burrage et al. 2012).

1.4.4| LSH in disease

Point mutations in LSH have recently been identified as being causative for Immunodeficiency, Centromeric instability Facial anomalies (ICF) syndrome (Thijssen *et al.*, 2015). ICF is characterised by often fatal chronic gastrointestinal and respiratory infections and the facial anomalies of hypotelorism, a flat epicanthus and nasal bridge (Hagleitner *et al.*, 2008). CpG hypomethylation of juxtacentromeric satellite types II and III cause centromeric instability and is used to diagnose ICF syndrome (Jiang *et al.*, 2005; De Greef *et al.*, 2011). 50% of cases are caused by mutations in the MT domain of DNMT3B and 30% of cases are caused by mutations in the Zinc-finger and BTB domain containing 24 (ZBTB24) (Jiang *et al.*, 2005; De Greef *et al.*, 2011). Four point mutant variants of Cell Division Cycle Associated 7 (CDCA7) and 5 point mutant variants of LSH have been identified in unexplained cases of ICF syndrome (Thijssen *et al.*, 2015). LSH^{K128*}, LSH^{K204*}, LSH^{Q699R}, LSH^{S762R} (causes frameshift deletion) and LSH^{801del} were identified as mutations in patients with ICF (Thijssen *et al.*, 2015). Therefore, this study provides evidence of how LSH is required for the establishment and potentially the maintenance of CpG methylation at centromeric repeat regions, a common process likely shared with DNMT3B, ZBTB24 and CDCA7 (Zhu et al. 2006, Thijssen et al. 2015).

1.4.5| LSH domain architecture and function

A schematic showing the structural and functional domains of *M. musculus* LSH is shown in Figure 1.8.

The N-terminal domain of LSH has a coiled-coil (CC) region (LSH^{14-96}) but does not share any sequence conservation with other chromatin remodelers and no specific functional domains are predicted. The region around LSH^{112} contains a nuclear localisation signal (NLS) (Lee et al. 2000).

The CC region may be necessary for protein:protein interactions. The N-terminal domain of LSH has been shown to be necessary for protein binding as pulldown experiments and co-immunoprecipitation experiments identified that the CC domain of LSH binds to E2F3 *in vitro* and *in vivo* (von Eyss et al. 2012). Co-immunoprecipitation experiments have shown LSH binds to DNMT3B (Zhu et al. 2006, Myant & Stancheva 2008) and a pulldown experiment showed recombinant GST-LSH^{1-503} bound to DNMT3B (Myant and Stancheva, 2008).

$LSH^{200-736}$ is the Snf2 ATPase region, with features equivalent to all chromatin remodelers in the SNF2-like grouping (Figure 1.5) apart from the distance between the conserved boxes B-C (see figure 1.6). LSH has 110 amino acids and others members in the Snf2-like grouping have between 61-65 amino acids, showing protrusion 2 of LSH is ~45 residues longer than in most other remodelers, however, the functional influence of the additional ~45 residues in LSH is currently unknown (Flaus *et al.*, 2006).

The mutant LSH^{K237Q} is a catalytically inactive ATPase. LSH^{K237} is the conserved lysine of helicase motif I which has been shown in the SF2 helicase Vasa to occupy the Mg^{2+} binding site and is displaced upon ATP-Mg^{2+} to form a hydrogen bond with the ATP β-phosphate enabling ATP hydrolysis through the 2 arginine residues in motif VI (see section 1.3.1 for more details). Therefore LSH^{K237Q} can no longer form a contact with ATP β-phosphate preventing ATP hydrolysis and therefore remodelling activity.

The C-terminal domain ($LSH^{760-810}$) also has an unknown function, however, it may bridge the ATPase lobes as in CHD1 (Figure 1.7B) or localise at ATPase lobe 2 as in SNF2 (Figure 1.7B).

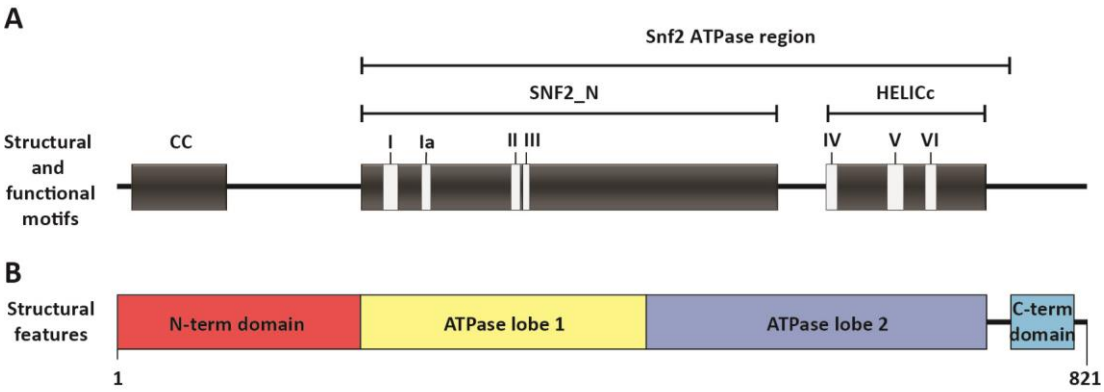

Figure 1.8| Murine LSH domain architecture
A| Functional motifs of LSH comprising the coiled coil (CC), SNF2_N and helicase motifs (I, Ia, II and III), HELICc and helicase motifs (IV, V and VI).
B| The simplified structural domains of LSH in relation to the RAD54 structure.
This figure is to scale.

Defective in DNA Methylation 1 (DDM1) the plant homologue of LSH, possess similar functions *in vivo* with a direct role in DNA methylation and DNA repair (Jeddeloh, Stokes and Richards, 1999; Saze and Kakutani, 2007; Yao *et al.*, 2012).

DDM1 has in part been biochemically characterised with the ability to hydrolyse ATP, a function enhanced 2-3 fold with free or nucleosomal DNA (Brzeski and Jerzmanowski, 2003). DDM1 was also shown to remodel nucleosomes via sliding the nucleosome along a stretch of DNA using ATP hydrolysis (Brzeski and Jerzmanowski, 2003). These functions of DDM1 provide some amenable functions for LSH, notably ATPase activity, DNA binding and nucleosome sliding.

Recombinant LSH expressed in a baculovirus expression system and purified has been shown to bind DNA and have ATPase activity which increases 10 fold upon the binding to DNA or nucleosomes (Burrage et al 2012).

1.5| Aims of the project

From a review of the literature, it is clear LSH has a critical and unrelated functions in cellular processes including DNA repair, cancer progression and regulating *de novo* methylation at repeat elements to adjust transcription. However, how the flanking domains of LSH influence the structure of the ATPase domain and the biological function of LSH function is currently unknown due to no published structural or biophysical investigations and few results describing the biochemical activities of LSH *in vitro.* Therefore, one of the aims was to determine the structural characteristics and regulatory mechanism of LSH *in vitro* using biochemical, biophysical and structural methods.

LSH has been shown to bind the *de-novo* methyltransferase DNMT3B through co-immunoprecipitation and pulldown experiments (Zhu et al. 2006, Myant & Stancheva 2008). Therefore, another aim was to test if LSH and DNMT3B interact directly *in vitro* and what the functional role of such how such an interaction may have on LSH and DNMT3B.

Chapter 2| Materials & Methods

2.1| Common Buffers and Reagents

Buffers and reagents were kept at room temperature unless otherwise stated.

PBS

140 mM NaCl, 3 mM KCl, 2 mM KH_2PO_4, 10 mM Na_2HPO_4.

TBE

89 mM Tris-HCl pH 8, 89 mM Boric acid, 2 mM EDTA.

TGE

25 mM Tris-HCl pH 8.3, 190 mM Glycine, 1 mM EDTA.

TGS

25 mM Tris, 250 mM Glycine, 0.1% (w/v) SDS.

TBS

50 mM Tris-HCl pH 8.0, 150 mM NaCl.

TAE

40 mM Tris, 20 mM glacial Acetic Acid, 1 mM EDTA and pH adjusted to 8.0.

TE

10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0.

SOC media

20 g/L Difco Bacto tryptone, 5 g/L Difco Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

LB media

10 g/L Difco Bacto tryptone, 5 g/L Difco Bacto yeast extract, 5 g/L NaCl, pH 7.2.

LB Agar

LB media + 2% Difco Bacto agar.

2.1.1|Reagents and buffers for the manipulation of DNA

6x Sucrose loading buffer

60 mM Tris-HCl (pH 7.6), 0.25 % (w/v) Bromophenol Blue, 40% (w/v) sucrose, 60 mM EDTA.

Methylene blue staining solution

0.0002% (w/v) methylene blue (Sigma) in 0.1x TAE buffer.

10x IV buffer

750 mM Tris-HCl pH 8.8, 200 mM Ammonium sulphate, 0.1% Tween 20.

EB
10 mM Tris-HCl pH 8.5 (20°C).

2.1.2| Reagents and buffers for electrophoresis/blotting

4x SDS-PAGE sample buffer
200 mM Tris-HCL pH 7, 8% (w/v) SDS, 20% (v/v) Glycerol, a 0.04% (w/v) bromophenol blue. The final DTT concentration was adjusted to 100 mM prior to use.

SDS-PAGE resolving gel
8-20% (w/v) 29:1 acrylamide:bis-acrylamide, 0.1% (w/v) SDS, 375 mM Tris-HCl pH 8.8, 0.1% (v/v) APS, 0.02% (v/v) TEMED.

SDS-PAGE stacking gel
4% (w/v) 29:1 acrylamide:bis-acrylamide, 0.1% (w/v) SDS, 125 mM Tris-HCl pH 6.8, 0.1% APS, 0.02% (v/v) TEMED.

SDS-PAGE running buffer
25 mM Tris, 250 mM Glycine, 0.1 % (w/v) SDS.

Nitrocellulose Western blot transfer buffer
25 mM Tris, 250 mM Glycine.

PVDF Western blot transfer buffer
25 mM Tris, 250 mM Glycine, 20% methanol, 0.1% SDS.

Ponceau S staining solution
1% (v/v) glacial acetic acid, 0.5% (w/v) Ponceau S.

Coomassie blue staining solution
40 % (v/v) ethanol, 10% (v/v) glacial acetic acid, 0.05% (w/v) Coomassie brilliant blue R-250.

Coomassie blue de-staining solution
10% (v/v) glacial acetic acid, 40% ethanol (v/v).

Native TBE acrylamide gel
89 mM Tris-HCl pH 7.6, 89 mM Boric acid, 2 mM EDTA pH 8, 5-20% (w/v) 29:1 acrylamide:bis-acrylamide, 0.1% (w/v) APS, 0.1% (v/v) TEMED.

1x Protease buffer
20 mM HEPES pH 7.5, 50 mM NaCl and 10 mM $MgSO_4$.

2.2| Bacterial cell work

2.2.1| Growth of *Escherichia coli* strains

All *Escherichia coli* (*E.coli*) strains were grown at 37°C in liquid LB media or on solid LB agar plates. *E. coli* strains DH10B or DH5α were used for all cloning and amplification of plasmids. The DH10 EMBacY *E.coli* cells were used for all cloning and amplification of baculoviral bacmids. Antibiotics were added to both liquid LB media and solid LB agar at the following working concentrations:

Table 2.1| The working concentrations of antibiotics used in this study

Antibiotic	Working concentration
Carbenicillin	100 µg/ml
Gentamicin	10 µg/ml
Kanamycin	50 µg/ml
Streptomycin	25 µg/ml
Tetracycline	10 µg/ml

2.2.2| Generation of chemically competent *E. coli* cells

TFB1 buffer

30 mM KAc, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% (v/v) Glycerol, pH adjusted to 5.8 with HOAc and filter sterilised.

TFB2 buffer

10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% (v/v) Glycerol, pH adjusted to 6.5 with KOH and filter sterilised.

A single *E. coli* colony was inoculated into 5 ml liquid LB and grown overnight (~16 hrs). The culture was diluted 1:200 into pre-warmed 100 ml liquid LB media containing 20 mM $MgSO_4$ and grown at 37°C until an OD_{600} of 0.6 was reached. Cells were then incubated on ice for 10 mins and pelleted by centrifugation for 5 mins at 5000 rpm and 4°C. The supernatant was decanted and cells were re-suspended in 40 ml TFB1 buffer and incubated on ice for 5 mins. Cells were pelleted for 10 mins at 3000rpm and 4°C. The supernatant was decanted and cells were re-suspended in 4 ml of cold TFB2 buffer and incubated on ice for 30 mins. 100 µl aliquots of competent cells were then flash frozen in liquid nitrogen and transferred to -80°C for long-term storage.

2.2.3| *E.coli* transformation protocol

A 100 µl aliquot of chemically competent *E. coli* cells were thawed on ice and 2-100 ng DNA was added, followed by incubation on ice for 30 mins. Cells were heat shocked for 45 secs (all strains) at 42°C and returned to ice for 2-3 mins. 900 µl SOC media was added and cells were allowed to recover at 37°C and 220 rpm for 1 hr (DH5α or DH10B strains) or 4-5 hrs (EMBacY strain). Recovered cells were plated onto LB agar containing the necessary antibiotic/s (and IPTG and X-gal for EMBacY strains) and incubated at 37°C for 16-20 hrs (DH5α or DH10B strains) or 24-48 hours (EMBacY strain) to form detectable colonies.

2.3| DNA manipulation experiments

2.3.1| PCR amplification of DNA

Polymerase chain reaction (PCR) amplification of DNA was carried out using the Biometra TProfessional Gradient 96 Thermocycler. For cloning purposes, Phusion® High-Fidelity DNA Polymerase (NEB) was used with High-Fidelity (HF) buffer according to the manufacturer's instructions. For all other PCRs, homemade Taq (HMT) polymerase was used with the following reaction composition:

Table 2.2| HMT PCR components and volumes for a single reaction

HMT PCR components	Volume
25mM Mg^{2+}	6 µl
10x IV buffer	5 µl
dNTP's (10 µM)	1 µl
DMSO	3.5 µl
Primers (10 µM)	0.625 µl each
DNA template (10 ng/µl)	1 µl
250U/µl HMT polymerase	1 µl
ddH_2O	31 µl

PCR cycling conditions were determined empirically for each primer pair.

2.3.2| Agarose Gel Electrophoresis (AGE)

A 1% or 2% (w/v) agarose gel was prepared using UltraPure™ Agarose (Invitrogen) with TAE buffer and the addition of 5 µl SafeView Nucleic acid stain (NBS Biologicals) per 100 ml agarose. A 1 kb DNA ladder (NEB) or 50 bp DNA ladder (NEB) were used as molecular size markers. Gels were typically run at 100 V for 60 mins and visualised by UV illumination using auto exposure for intense bands on the GELDOC EZ system (Bio-Rad).

2.3.3| Plasmid DNA preparation

Small scale plasmid DNA extraction and purification was carried out using 5 ml confluent *E. coli* following the QIAprep Spin Miniprep kit (Qiagen) manual. Where more plasmid DNA was needed, cultures were scaled up to 200 ml confluent *E. coli* and a plasmid midiprep kit (Qiagen) was used.

2.3.4| EMBacY blue/white screening

Agar plates containing carbenicillin, gentamicin and tetracycline were set at room temperature and allowed to dry at 37°C for 1 hr. 100 µl of IPTG (Sigma) at 0.1 M in ddH_2O and 100 µl X-gal (Sigma) at 20 mg/ml in DMSO were spread on the set agar and allowed to dry at 37°C for 30 mins before use. Plates were stored at 4°C and protected from light for up to 4 weeks.

2.3.5| Isolating EMBacY Bacmid DNA from *E. coli* cells

The buffers used in the QIAprep Spin Miniprep Kit (Qiagen) are also used for isolating and purifying EMBacY DNA from *E. coli* cells. The buffer compositions are written below.

P1

50 mM Tris-HCl pH 8, 10 mM EDTA, 100 ug/ml RNase.

P2

200 mM NaOH, 1% SDS (w/v).

N3

4.2 M Gu-HCl, 0.9 M Potassium acetate pH 4.8.

A single *E. coli* (EMBacY strain) colony was inoculated into 5 ml LB liquid media containing relevant antibiotics (carbenicillin, gentamicin and tetracycline) and grown at 37°C and 220 rpm for 20-22 hrs. All culture was pelleted at 13000 rpm at 4°C for 1 min and supernatant discarded. The cells were thoroughly resuspended in 250 µl P1. 250 µl P2 was added and mixed by inverting for no more than 5 mins. 350 µl N3 was added and mixed by inverting. Cell debris was pelleted by at 13000 rpm for 10 mins. The supernatant was centrifuged for a further 10 mins in a new Eppendorf tube. The bacmid DNA containing supernatant was mixed with 800 µl isopropanol by inverting. DNA was pelleted at 13000 rpm for 15 mins. All supernatant was discarded and the DNA pellet was washed 2 times with 500 µl Ethanol (70%) and centrifuged at 13000 rpm for 5 mins. The pellet was air dried for 10 mins and gently re-suspended in 100 µl EB buffer.

The quantity (ng/µl) and the purity of bacmid DNA was measured using UV spectroscopy (section 2.4.4). The bacmid DNA was diluted to the transfection concentration of 1 µg/µl. Bacmid DNA was verified to contain the ORF of the gene of interest using specific primer combinations in the gene as the forward primer and M13R or M13F positioned on the outside of the Tn7 integration site as the reverse primer. PCR products corresponding to correct DNA lengths confirmed integration.

2.3.6| PCR product purification

PCR products were purified using the QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions. The quantity (ng/µl) and the purity of DNA was measured using UV spectroscopy (section 2.4.4).

2.3.7| RE digestion

RE's used in this study were from NEB or Fermentas. Reactions were prepared using the appropriate reaction buffer following manufacturer's instructions. 2 units of RE were used per 1 µg of DNA and incubated for 30 mins at 37°C. The products of RE digestion were analysed by agarose gel electrophoresis.

2.3.8| DNA extraction from agarose gels

Digested DNA was size sorted using AGE. DNA fragments were visualised using a Safe Imager™ 2.0 Blue Light Transilluminator (Invitrogen) and excised with a clean scalpel. DNA was extracted from the excised gel pieces using a QIAquick Gel Extraction Kit (Qiagen-28704) according to manufacturer's instructions.

2.3.9| DNA Ligation

Ligation of the gene and plasmid was performed using T4 DNA ligase (NEB) in a 20 µl reaction volume following manufacturer's instructions. A vector:insert ratio of 1:3 was used for all ligation reactions using the following formula:

$$\text{Insert mass in ng} = 3 \times \left[\frac{\text{insert length (bp)}}{\text{vector length (bp)}}\right] \times \text{vector mass in ng} \qquad \text{Equation (2.1)}$$

The total mass of DNA used was between 70 ng and 100 ng (vector and insert combined). The mixture was incubated for 30 mins at room temperature and 10 µl was transformed into 100 µl chemically competent DH5α *E.coli* cells following the transformation protocol.

2.3.10| Colony PCR

Colony PCR was used to verify the presence of the transformed plasmid. Amplification was carried out using an insert-specific primer and a vector-specific primer close to the site of integration. HMT polymerase PCR reaction mixtures were used (see section 2.3.1) and DNA was made available from mixing a selected colony in the PCR reaction mixture.

2.3.11| DNA Sequencing

For all sequencing reactions, plasmid DNA was sequenced directly. Sequencing reactions were prepared as follows:

Table 2.3| Sequencing reagents and volumes to use for a single reaction

Component	Volume
5x sequencing buffer (Applied Biosystems)	2 µl
5x BIG DYE 3.1 Terminator mix (Applied Biosystems)	2 µl
3.2 µM sequencing primer (in H_2O or TE buffer)	1 µl
Plasmid DNA (200-500 ng)	1-5 µl
ddH_2O	Up to 10 µl total volume

DNA was amplified as follows:

Table 2.4| PCR cycling conditions

Step	Temperature	Time	Cycles
1	95°C	30 secs	25
2	50°C	20 secs	
3	60°C	4 mins	
4	60°C	1 min 15 secs	
5	4°C	HOLD	

All DNA sequencing was carried out by the GenePool service at Edinburgh University, using ABI 3730 capillary sequencers. Sequences were analysed using ClustalW2/Clustal Omega software or MEGAlign (DNAstar) software.

2.3.12| Cloning of LSH gene variants for EMBacY bacmid construction

DNA sequences encoding full-length and/or truncated LSH and/or full-length DNMT3B coding sequences were cloned into the MultiBac pFL plasmid before being transposed into the EMBacY bacmid via Tn7 transposition (Figure 2.1) (Fitzgerald *et al.*, 2006)(Trowitzsch *et al.*, 2010). The pFL plasmid contains multiple cloning sites (MCS) downstream of the p10 and polH (polyhedrin) promoters (Figure 2.1). Protein tagging and truncations were achieved by PCR-based methods. A plasmid containing LSH or DNMT3B was PCR amplified using primers with overhangs coding for the tag of interest (6His, StrepII, 6His-TEV, StrepII-TEV or 6His-StrepII-TEV) and restriction sites to enable restriction-based cloning into the pFL plasmid.

BamHI and SalI restriction enzymes were used and for all LSH constructs. BamHI and EcoRI or NheI and SphI were used for all DNMT3B constructs. Products of the PCR reaction performed with primer pairs were analysed by agarose gel electrophoresis and the DNA was purified using the PCR product purification protocol (section 2.3.6). The purified PCR product (insert) and pFL plasmid (vector) were digested with appropriate restriction enzymes followed by PCR product purification protocol, ligation and transformation into DH5α cells followed by growth on selective carbenicillin agar plates. Positive colonies were selected by colony PCR and the plasmid DNA was purified. After verification of the DNA sequence, the plasmid was transformed into DH10 EMBacY *E. coli* cells, containing the EMBacY bacmid (baculovirus genome) and a Tn7 'helper plasmid'. The helper plasmid encodes for the Tn7 transposition complex necessary for integration of the pFL plasmid into the EMBacY bacmid via homologous recombination at attTn7 sites. Successful transposition causes disruption of the LacZ gene in the EMBacY bacmid (Figure 2.1) allowing positive colonies to be selected by blue/white colony screening. Re-streaking of white colonies and PCR were used to confirm positive clones. An LSH or DNMT3B sequencing primer and a bacmid specific M13 primer were used to test successful T7 transposition of the pFL plasmid into the EMBacY bacmid (Figure 2.1).

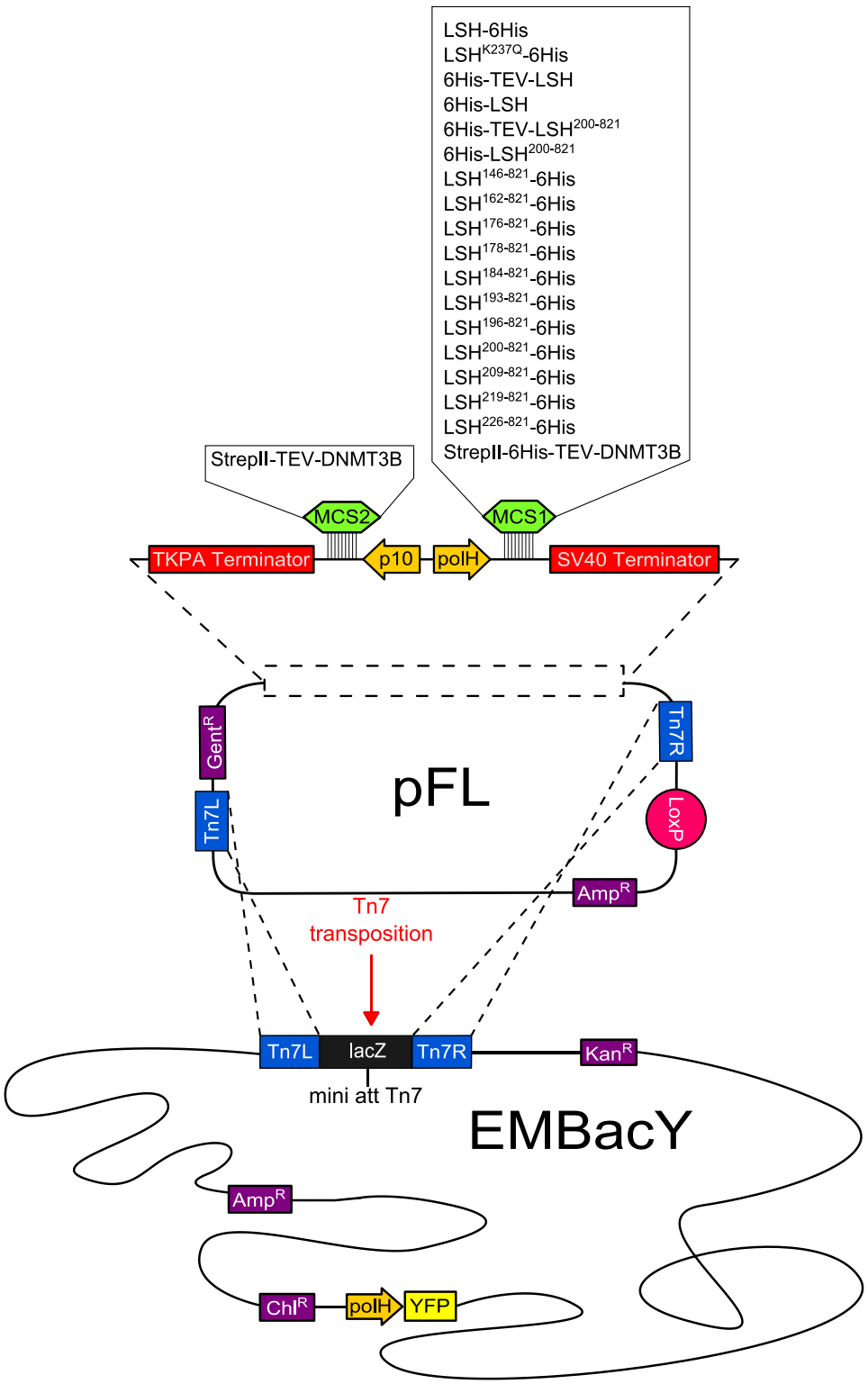

Figure 2.1|LSH and DNMT3B pFL constructs made and used in this study
Affinity tagged LSH or DNMT3B sequences were cloned into the pFL plasmid using the multiple cloning sites 1 and 2 (MCS1 and MCS2) under control of promoter polH and p10 respectively. The pFL plasmid was integrated into the transfective EMBacY bacmid following Tn7 transposition. The LoxP site could be used for Cre-Lox recombination of another pFL plasmid using Cre recombinase.

2.4| Common protein biochemistry methods

2.4.1| SDS-PAGE

Gels were cast using the BioRad mini PROTEAN® Tetra Handcast systems. The resolving gel was prepared, poured between glass plates (0.75-1.5 mm) and layered with 100% Isopropanol. Once set the isopropanol was removed and washed with ddH_2O before adding the stacking gel.

SDS-PAGE gradient gel:

Equal volumes of 4% stacking and 20% resolving mixtures were pipetted slowly into a 10 ml plastic pipette respectively. Three bubbles were drawn up to form a gradient and the mixture was pipetted slowly into the gel caster.

Hand-cast gels were stored at 4°C up to one week.

Protein samples were prepared in SDS sample buffer and boiled at 90°C for 5 mins before loading alongside 3.5 µl of a protein ladder (PageRuler Plus 10-250 kDa - Pierce). Gels were run at 250V for 30 to 45 mins or until dye front was near the bottom. For visualisation of proteins, the gel was stained with Coomassie stain for 20 minutes with agitation and detained in Coomassie de-stain. The gels were scanned on A GELDOC EZ system (Bio-Rad) using the Coomassie setting.

2.4.2| Western blotting

Proteins were separated on an SDS-PAGE gel. To transfer large proteins, nitrocellulose membrane and the wet transfer method with nitrocellulose transfer buffer were used. The transfer was performed at 4°C and 400 mA for 1 hr. To transfer histones a PVDF membrane was used. PVDF was charged in 100% methanol and the transfer was performed in PVDF transfer buffer. Membranes were blocked with 2% (w/v) milk with 0.1% (v/v) Tween20. Primary antibody binding was performed overnight at 4°C or for 3 hrs at room temperature with the desired antibody. Membranes were washed three times (10 mins) in PBS + 0.1% (v/v) Tween20 and blocked again for 30 mins. A secondary antibody was added and membranes were incubated at room temperature for at least 1 hr. Membranes were washed three times as above (final wash was done with PBS only). An Odyssey scanner (Li-Cor) and Image Studio Lite (v3.1) software were used to visualise and quantify bands respectively.

Table 2.5| Primary antibodies used for Western blotting

Name	Type	Source, catalog number	Western blot dilution factor
α-DNMT3b	Mouse monoclonal	Abcam, ab13604	1:1000
α-HDAC1	Rabbit polyclonal	Santa Cruz, sc-7872/D171	1:500
α-H4	Rabbit polyclonal	Milipore, 07-108	1:500
α-6His	Mouse monoclonal	A7058 SIGMA	1:1000
α-C-LSH	Rabbit polyclonal Recognises residues around 503aa	HELLS antibody #7998 Cell signaling technology	1:1000
α-N-LSH	Mouse monoclonal Recognises 1-223aa	LSH antibody (H4) Santa Cruz, sc-46665	1:1000
α-Strep2	Mouse monoclonal	IBA, 2-1507-001	1:1000
α-Tubulin	Mouse monoclonal	CRUK	1:1000

Table 2.6| Secondary antibodies used for Western blotting

Name	Type	Catalog number	dilution factor
IRDye® 680LT	Donkey anti-Rabbit IgG (H + L)	LI-COR, 926-68023	1:10000
IRDye® 800CW	Donkey anti-Mouse IgG (H + L)	LI-COR, 926-32212	1:10000

2.4.3| Protein quantification by UV spectroscopy

Protein concentration was measured on a nanodrop 2000™ using the molecular mass (Mw) and molar extinction coefficient (ε) for the protein predicted using the ProtParam tool: http://web.expasy.org/protparam/ (Gasteiger *et al.*, 2005). The Mw, ε (assumed all cysteines are reduced) and isoelectric point (pI) for each protein tested is below:

Table 2.7|Extinction coefficient, Mw and pI of proteins purified

Protein	ε (M^{-1} cm^{-1})	Mw (Da)	pI
LSH-6His	76320	95948.5	8.2
$LSH^{176-826}$-6His	66350	75403.2	8.43
LSH^{1-176}	9970	20563.3	6.27
StrepII-6His-TEV-	126740	101466.4	8.35
BSA	43824	66400	4.7

2.4.4| DNA quantification by UV spectroscopy

DNA concentration was measured on a nanodrop 2000™ using the default nucleic acid setting.

2.4.5| Protein quantification by BCA assay

A working solution was prepared by mixing BCA solution (Sigma - B9643) and copper (II) sulphate solution (Sigma - C2284) in a 50:1 ratio. 1 µL of the total protein was mixed with 1 mL of working solution. Samples were incubated at 65°C for 15 mins shaking at 200 rpm. After the incubation, samples were chilled on ice briefly and the absorbance at 562 nm was measured using a spectrophotometer. The absorbance values were divided by 0.04 to calculate the protein concentration.

2.4.6| Protein quantification by Bradford assay

Protein Assay Dye Reagent Concentrate (Bio-Rad) or Coomassie protein assay reagent (Pierce) were used according to manufacturer's guidelines.

2.5| Insect cell work

2.5.1| Insect cell culture

All baculovirus production was carried out using *Sf*9 insect cells (an ovarian clonal isolate of *Spodoptera frugiperda* – Thermo Fisher). All *Sf*9 cell culture was done at 27°C and in Sf-900 II serum-free medium (Sf-900 II SFM; Gibco). High Five™ cells (a clonal isolate derived from the parental *Trichopulsia ni* cell line – Thermo Fisher) were used only for protein expression. All High Five™ cell culture was performed at 27°C and in ExpressFive® SFM (ThermoFisher) supplemented with 16 mM L-Glutamine (G8540-Sigma) and 100 µg/L 5000 Mr dextran sulphate (31404-Sigma).

Cells were maintained as adherent cultures in T-75 flasks and split 1:5 (*Sf*9) or 1:10 (High Five™) when ~90% confluent (every 4-6 days). Suspension cultures were only initiated after passage 3 from T-75 flasks and grown in glass baffled flasks with membrane screw cap (Duran) or round plastic easy grip bottles (Corning). Suspension cultures were scaled up into 1 L glass baffled flasks (Duran). A maximum of 400 ml of cell culture was grown in 1 L baffled flasks. Suspension cultures were maintained at cell densities between 1 - 10x10^6 cells/ml (*Sf*9) and 1 - 6x10^6 cells/ml (High Five™).

Cell densities were measured using a Countess® Automated Cell Counter (Thermo-Fisher) using Countess® Cell Counting Chamber Slides (Thermo-Fisher). Cells were not passaged more than 30 times and fresh stocks were reanimated if viability dropped below 90%.

For production of frozen cell stocks, a suspension culture was grown to a cell density of 1x 10^6 cells/ml. Cells were pelleted by spinning at 1500 rpm for 3 mins, and re-suspended in filter sterilised freezing medium (46.5% conditioned media, 46.5% fresh *Sf*-900 II SFM, 7% DMSO) to a final density of 2 x 10^7 cells/ml. 1 ml aliquots in 1.8 ml cryotubes (Nunc) were frozen slowly by placing them in a chilled polystyrene box followed by a -80°C freezer for 48-72 hrs before finally placing in liquid nitrogen storage.

Insect cells were reanimated by warming to 37°C from frozen and adding a 1 ml aliquot into a T-75 flask containing 10 ml Sf-900 II SFM media at 27°C, which was replaced with fresh media, 1 hr after attachment.

2.5.2| Baculovirus transfection and virus generation

Bacmid DNA was transfected into *Sf*9 cells to produce an initial virus stock (V_0). Transfections were performed using X-tremeGENE HP DNA transfection reagent (Roche). 50 µl of bacmid DNA (1 µg/µl) was added to 500 µl Sf-900 II SFM and gently mixed. 20 µl X-tremeGENE HP was added directly into the DNA/SFM mixture and incubated at room temperature for 30 mins before 110 µl was added dropwise to 1-2x10^6 *Sf*9 cells (70-90% confluency) in 5 wells of a 6 well plate. The last well contained non-transfected *Sf*9 cells as a negative control. 72 hrs after transfection, YFP expression was monitored using fluorescence microscopy via a UV fluorescent lamp connected to a Nikon Eclipse TS100 microscope attached with Q-imaging recorder using QCapturePro software.

96 hrs after transfection, media from each well was harvested and pooled as V_0. Fresh *Sf*-900 II SFM was then added to transfected cells, and protein expression was confirmed by western blot following an additional 48 hrs growth in fresh medium. V_0 was then used to produce a more infective and higher volume V_1 virus stock as well as baculovirus infected insect cells (BIICs).

2.5.3| BIICs, V1 and V2 virus generation

5-10 ml of V_0 was added to 200 mL *Sf*9 suspension culture at 1x10^6 cells/ml. After 24 hrs 100 ml of culture was pelleted at 1500 rpm for 3 mins and frozen at a density of 1x10^7 cells/ml using the cell freezing method in section 2.5.1. The growth of the remaining 100 ml was continued and fluorescence monitored via microscopy with peak fluorescence typically seen by day 3 or 4. Cells were spun down as before and the cell pellet frozen at -80°C for later analysis. V1 containing media was filtered (0.22 µm), protected from light, stored at 4°C and used within 2 weeks. For V2 generation 1 ml of BIICs was thawed into 200 ml of *Sf*9 cells at a cell density of 1x10^6 cells/ml. After 3-4 days peak fluorescence was seen, V2 containing media was treated as above.

2.5.4| 96-well plate BIIC titer assay

To establish the optimal BIIC:*Sf*9 cell volume for YFP expression, a 96-well plate BIIC titer assay was developed. A black clear bottomed cell culture 96 well plate was used (Thermo-Fisher 165305.) *Sf*9 or High Five™ cells were diluted to 0.75x10^6 cells/ml from suspension cultures. BIICs were diluted as indicated below.

Table 2.8| Cell concentrations and volume for the 96-well plate BIIC titer assay

BIIC:*Sf*9 cells ratio	BIIC concentration	BIIC volume to add	Volume of *Sf*9 or High Five™ cells ($7.5x10^5$/ml)	Total volume
1:5	$7.5x10^5$ cells/ml	20 µl	80 µl	100 µl
1:10	$7.5x10^5$ cells/ml	10 µl	90 µl	100 µl
1:20	$7.5x10^5$ cells/ml	5 µl	95 µl	100 µl
1:50	$7.5x10^5$ cells/ml	2 µl	98 µl	100 µl
1:100	$3.75x10^4$ cells/ml	20 µl	80 µl	100 µl
1:200	$3.75x10^4$ cells/ml	10 µl	90 µl	100 µl
1:500	$3.75x10^4$ cells/ml	5 µl	95 µl	100 µl
1:1000	$3.75x10^4$ cells/ml	2 µl	98 µl	100 µl
1:0	0 cells/ml	0 µl	100 µl	100 µl

Cells were incubated at 27°C for 72 hrs before visualisation. Images were taken as described in section 2.5.2. Images were analysed using a macro created by Dr David Kelly in Image-Pro plus software (Media Cybernetics) to measure the area of fluorescent vs non-fluorescent cells.

Cell fluorescence was also analysed on a SpectraMax M5 multiplate reader (Molecular Devices). The settings used were excitation at 495 nm and emission at 515 nm (cut off 515 nm) with end point assay and 3 point scanning. All results were analysed in Graphpad Prism (Graphpad software Inc.).

2.5.5| Characterisation of V1 or V2 virus

Various ratios of baculovirus to *Sf*9 cells were tested to determine the ratio resulting in the best recombinant protein expression. *Sf*9 cells at $1x10^6$ cells/ml in suspension culture were infected with varying volumes of V2 virus. For every 12 hrs following the 1st day after inoculation 1.5 ml of cells were pelleted and resuspended in 300 µl of lysis buffer (50 mM HEPES pH 7, 500 mM NaCl, 10% (v/v) Glycerol, 1 mM DTT, 0.1% (v/v) Triton X-100) and sonicated (3 x 20 secs on/off) on ice using a soniprep 150. Samples were centrifuged at 13000 rpm for 15 mins and soluble lysate protein concentration was measured using the BCA assay. 50 µg of each sample were analysed by SDS-PAGE, followed by western blotting with α-LSH antibody and α-Tubulin (loading control) with Li-Cor secondary antibodies used for quantitative analysis using an Odyssey scanner (Li-Cor) and image studio lite (3.1) software.

2.5.6| Large-scale protein production in Insect cells

Large scale LSH protein expression|

40 ml of V1 or V2 virus was inoculated into 400 ml of High Five™ cells growing at a density of 3-4x10^6 cells/ml, which had been freshly split the day before to 2x10^6 cells/ml from a stock at 4x10^6 cells/ml. 1 day after inoculation, 1 L of cells was supplemented with 50 ml of filter sterilised supplement solution containing 5 g Glucose (Sigma-G7021) 200 mM L-Glutamine (Sigma-G8540) and 5 g Hypep Wheat Hydrolysate (Sigma-H6784) mixed in 18.2□ ddH_2O. The maximum fluorescence was detected 3-4 days after inoculation. Cells were centrifuged at 4000 rpm for 10 mins and cell pellets flash frozen in liquid nitrogen for long term storage at -80°C.

Large scale DNMT3B protein expression|

20 ml of V1 or V2 virus was inoculated into 400 ml of *Sf*9 cells growing at a density of 3-4x10^6 cells/ml, which had been freshly split from a 6-8x10^6 cells/ml stock. The maximum fluorescence was detected 3 days after inoculation. Cells were pelleted and frozen as above.

2.6| Protein purification

2.6.1| Concentrating proteins

All proteins were concentrated with 500 µl to 20 ml vivaspin spin concentrators (GE Healthcare) with a 10 kDa MWCO. Vivaspin spin concentrators were equilibrated with the protein buffer and centrifuged at rpm recommended by the manufacturer.

2.6.2| SEC of proteins and/or DNA

All SEC of purified recombinant proteins and DNA were performed on a ÄKTA Purifier 100 HPLC unit with sample collector. For small volumes (50 µl to 500 µl) a Superdex 200 10/300GL 24 ml column (GE Healthcare) was used. For large volumes (500 µl to 5 ml) a Hi-load Superdex 200pg 16/60 120ml column (GE Healthcare) was used. Columns were equilibrated with 2 CV of SEC buffer (50 mM HEPES pH 7, 500 mM NaCl, 10% (v/v) Glycerol, 1 mM DTT) before manual injection of sample in a 500 µl sample loop (200 10/300GL) or 5 ml sample loop (200pg 16/60). Fractions were collected in 0.25 ml or 0.5 ml volumes with a 0.25 CV delayed fractionation setting. On-line UV 260 nm and 280 nm was monitored.

2.6.3| Storage of proteins

Purified recombinant proteins in SEC buffer were aliquoted into PCR tubes in a maximum volume of 50 µl before being flash frozen in liquid nitrogen and stored at -80°C.

2.6.4| LSH-6His and LSH^{K237Q}-6His purification

Unless otherwise stated all protein purifications were performed on ice or at 4-10°C.

Buffers used for purification:

Lysis buffer
50 mM HEPES pH 7, 500 mM NaCl, 10% (v/v) Glycerol, 0.5 mM DTT, 2 mM $MgCl_2$, 0.2 mM PMSF, 1x cOmplete™ EDTA-free protease inhibitors (Roche).

Ni-NTA wash buffer
50 mM HEPES pH 7, 500 mM NaCl, 50 mM Imidazole pH 7, 0.5 mM DTT, 0.2 mM PMSF.

Ni-NTA high salt wash buffer
50 mM HEPES pH 7, 1 M NaCl, 0.5 mM DTT, 0.2 mM PMSF.

Ni-NTA low salt wash buffer
50 mM HEPES pH 7, 100 mM NaCl, 0.5 mM DTT, 0.2 mM PMSF.

Ni-NTA elution buffer
50 mM HEPES pH 7, 100 mM NaCl, 500 mM Imidazole pH 7, 10% (v/v) Glycerol, 0.5 mM DTT, 0.2 mM PMSF.

P11 buffer
250 mM KPO_4 pH 7.4, 1 mM EDTA, 1 mM DTT.

PC wash buffer
50 mM HEPES pH 7, 100 mM NaCl, 10% (v/v) Glycerol, 1 mM DTT, 0.2 mM PMSF.

PC elution buffer
50 mM HEPES pH 7, 500 mM NaCl, 10% (v/v) Glycerol, 1 mM DTT, 0.2 mM PMSF.

A 10 g cell pellet was resuspended in 50 ml lysis buffer and sonicated with a soniprep 150 (MSE) on ice (4x for 30 secs on/off) before being passed once through a cell disruptor (Constant Systems^{LTD}) at a pressure of 18 kpsi and 4°C. DNase I (Sigma) was added at a final concentration of 1 U/ml lysate and incubated on ice for 30 mins before being centrifuged for 1 hr at 25000 rpm and 4°C. The clarified lysate was incubated with Ni-NTA resin (GE Healthcare) pre-equilibrated with lysis buffer (- 0.5 mM DTT) for 16 hrs (1 ml NiNTA resin per 100ml lysate). The Ni-NTA resin containing lysate was pelleted at 1500 rpm for 2 mins before being manually loaded into a 10 ml polyprep column. The packed resin was washed with 40 column volumes (CV) of Ni-NTA wash buffer followed by 20 CV of Ni-NTA high salt buffer and a final 20 CV of Ni-NTA wash buffer. 10 CV of Ni-NTA low salt wash buffer was applied before adding Ni-NTA elution buffer and collecting 1 ml fractions. Fractions were analysed by SDS-PAGE and Coomassie staining and the elution fractions were pooled.

P11 dry fibrous cellulose phosphate (Whatman) was used for phosphocellulose cation exchange. P11 cellulose was rehydrated to a ~50% slurry in P11 buffer. To prepare the phosphocellulose slurry, 50 g of Whatman P11 was resuspended in 1 L of 0.5 M NaOH and stirred for 5 mins using a magnetic stir bar. The mixture was poured into 4 x 250 ml Nalgene bottles and centrifuged for 5 mins at 2000 rpm. The supernatant was removed and the phosphocellulose matrix was suspended in ddH_2O (250 ml in each bottle) and mixed by shaking for 5 mins. The bottles were centrifuged for 5 mins at 2000 rpm and ddH_2O was removed. After five repeated washes the phosphocellulose matrix was suspended in 0.5 M HCl (250 ml in each bottle) and mixed by shaking for 5 mins. The bottles were centrifuged for 5 mins at 2000 rpm and 0.5 M HCl was removed. After a repeat wash in 0.5 M HCl, the phosphocellulose matrix was washed 5 times in ddH_2O as above.

The phosphocellulose matrix was washed five times in P11 buffer, all matrix was pooled and suspended in 500 ml P11 buffer to make the phosphocellulose slurry. The phosphocellulose slurry was degassed for 1 hr before being stored at 4°C.

Pooled IMAC elutions were applied to ~4 ml hydrated P11 phosphocellulose packed into a polyprep 10 ml column which had been equilibrated with 10 CV of PC wash buffer. 40 CV of PC wash buffer was applied before eluting 1 ml fractions with PC elution buffer. Fractions were analysed using SDS-PAGE and the fractions were pooled and concentrated to 5-10 mg/ml. SEC was carried out as in section 2.6.2 and protein was stored as in section 2.6.3.

2.6.5| 6His-TEV-LSH$^{1-176}$ recombinant protein expression and purification

(Cloning and optimised expression of 6His-TEV-LSH$^{1-176}$ was performed by Mari Eltermann, a MChem student undertaking her research project in the Richardson lab under my co-supervision)

6His-TEV-LSH$^{1-176}$ was amplified using PCR and cloned into a pET-28a plasmid using restriction digestion and ligation into chemically competent DH5α *E.coli*. The insert was verified by sequencing and the plasmid transformed into chemically competent BL21-DE3 *E.coli*. Cells were grown in LB medium containing with 50 µg/ml Kanamycin at 37°C, 250 rpm shaking in 1/5th the maximum volume of the flask to an OD of 0.8 before being induced with 1mM IPTG. Cells were harvested 3 hrs after induction using centrifugation at 6000 rpm for 15 mins, media was discarded. Cell pellets were flash frozen in liquid nitrogen and kept at -80°C.

Buffers used for purification:

Lysis buffer
50 mM HEPES pH 7, 250 mM NaCl, 10% (v/v) Glycerol, 2 mM $MgCl_2$, 0.2 mM PMSF, 1x cOmplete™ EDTA-free protease inhibitors (Roche), 0.5mM DTT.

Ni-NTA buffer A
50 mM HEPES pH 7, 250 mM NaCl, 0.5 mM DTT, 0.2 mM PMSF.

Ni-NTA buffer B
50 mM HEPES pH 7, 250 mM NaCl, 500 mM Imidazole pH 7, 0.5 mM DTT, 0.2 mM PMSF.

Anion buffer A
50 mM HEPES pH 7, 100 mM NaCl, 1 mM DTT, 0.2 mM PMSF.

Anion buffer B
50 mM HEPES pH 7, 1 M NaCl, 1 mM DTT, 0.2 mM PMSF.

A 5g cell pellet was resuspended in a total volume of 50 ml lysis buffer and lysozyme was added to a final concentration of 0.2 mg/ml. The lysate was passed once through a cell disruptor (Constant Systems^{LTD}) at 25 kpsi and 4°C. DNase I (Sigma) was added to a final concentration of 1 U/ml and incubated on ice for 30 mins before centrifugation at 20000 rpm for 1 hr at 4°C.

The supernatant was directly loaded onto a 5 ml Histrap IMAC crude column (GE healthcare) pre-equilibrated with 10 CV Ni-NTA buffer A on an ÄKTA Purifier 100 HPLC system. The column was washed with 10 CV 15% Ni-NTA buffer B followed by a 30 CV gradient of 0-100% Ni-NTA buffer B. 1 ml fractions were collected and analysed by SDS-PAGE and Coomassie staining. Elution fractions were pooled and concentration measured using UV spectroscopy (section 2.4.3). For 6His cleavage, 1 mg of 6His-TEV protease (gifted by the Sawin and Arulanandam labs) was mixed with 25 mg of 6His-TEV-LSH$^{1-176}$. The mixture was dialysed overnight in snakeskin dialysis tubing with a 10 kDa MWCO (Thermo-Fisher) in 1.5 L Ni-NTA buffer A. The dialysed material was re-purified following the IMAC purification as above to purify LSH$^{1-176}$. Fractions were analysed by SDS-PAGE and Coomassie staining. Fractions containing LSH$^{1-176}$ were pooled and dialysed into 1.5 L Anion buffer A.

LSH^{1-176} was applied to a 1 ml MonoQ column (GE Healthcare) pre-equilibrated with 10CV of Anion buffer A on an ÄKTA HPLC system. A 20 CV gradient of 0-100% Anion Buffer B was applied with collection of 0.5 ml fractions. Fractions were analysed via SDS-PAGE and Coomassie staining. Fractions containing pure LSH^{1-176} were pooled and concentrated to 10 mg/ml. SEC was carried out as in section 2.6.2 and protein stored as in section 2.6.3.

2.6.6| StrepII-6His-TEV-DNMT3B purification

Buffers used for the purification:

Lysis buffer A
50 mM HEPES pH 7, 10% (v/v) Glycerol, 1 mM DTT, 0.2 mM PMSF.

Lysis buffer B
50 mM HEPES pH 7, 150 mM NaCl, 10% (v/v) Glycerol, 1 mM DTT, 0.2 mM PMSF.

Lysis buffer C
50 mM HEPES pH 7, 1 M NaCl, 10% (v/v) Glycerol, 0.5 mM DTT, 0.2 mM PMSF, 1x cOmplete™ EDTA-free protease inhibitors (Roche).

Ni-NTA buffer A
50 mM HEPES pH 7, 1 M NaCl, 0.5 mM DTT, 0.2 mM PMSF.

Ni-NTA buffer B
50 mM HEPES pH 7, 1 M NaCl, 500 mM Imidazole pH 7, 0.5 mM DTT, 0.2 mM PMSF.

Ni-NTA wash buffer

50 mM HEPES pH 7, 1 M NaCl, 60 mM Imidazole pH 7, 0.5 mM DTT, 0.2 mM PMSF.

Ni-NTA high salt buffer
50 mM HEPES pH 7, 2 M NaCl, 0.5 mM DTT, 0.2 mM PMSF.

A 10 g cell pellet was resuspended in 40 ml lysis buffer A and sonicated with a soniprep 150 (MSE) on ice (4x for 30 s on/off) before centrifugation at 5000 rpm for 10 mins. The supernatant was discarded and the pelleted material was resuspended in 40 ml lysis buffer B and treated as above. The pelleted material was resuspended in lysis buffer C and sonicated as above before centrifugation at 20000 rpm for 1 hr. The supernatant was loaded onto a 5 ml Histrap IMAC crude column (GE Healthcare) pre-equilibrated with 10 CV Lysis buffer A (without DTT) on an ÄKTA HPLC system. The loaded column was washed with 10 CV 15% Ni-NTA buffer B followed by a 30 CV gradient of 0-100% Ni-NTA buffer B. 1 ml fractions were collected and analysed by SDS-PAGE and Coomassie staining.

In an alternative procedure, the supernatant was incubated with Ni-NTA (GE healthcare) resin for 16 hrs (1 ml Ni-NTA resin per 100 ml lysate). The Ni-NTA resin was pelleted at 1500 rpm for 2 mins before being manually loaded into a polyprep column. The packed Ni-NTA resin was washed with 40 CV of Ni-NTA wash buffer followed by 20 CV of Ni-NTA high salt buffer and a final 20 CV of Ni-NTA wash buffer before eluting 1 ml fractions with Ni-NTA buffer B. Fractions were and analysed by SDS-PAGE and Coomassie staining.

Pooled elutions were buffer exchanged overnight in 1.5 L SEC buffer using 10 kDa MWCO snakeskin dialysis tubing (Thermo-fisher) before concentrating to 4 mg/ml. SEC was carried out as in section 2.6.2 and protein stored as in section 2.6.3.

2.6.7| Limited Proteolysis

The proteases used were trypsin, elastase, chymotrypsin and subtilisin (Sigma). 1 mg/ml stocks of the proteases were made in 1x protease buffer, flash frozen and stored at -80°C for long-term use. Three working dilutions were made (0.1, 0.01, 0.001 mg/ml) in 1x protease buffer. For initial screening experiments, 10 µl of the protein (0.6 mg/ml) was mixed with 3 µl of diluted protease. Protein-protease reactions were incubated on ice for 30 mins and stopped by adding 5 µl 4x SDS-loading buffer and boiling immediately. Samples were analysed on an SDS-PAGE gel. For time course experiments subtilisin (0.001mg/ml) or Trypsin (0.01mg/ml) were used as above however every 10 mins for 90 mins samples were taken and analysed using SDS-PAGE.

2.7| MALDI-ToF Mass Spectrometry

MALDI-ToF Buffers:

Buffer A

200 mM ammonium bicarbonate and 50% (v/v) acetonitrile.

Buffer B

200 mM ammonium bicarbonate, 50% (v/v) acetonitrile, 20 mM DTT.

Buffer C

200 mM ammonium bicarbonate, 50% (v/v) acetonitrile, 50 mM iodoacetamide.

Buffer D

20 mM ammonium bicarbonate, 50% (v/v) acetonitrile.

Trypsin digestion buffer

50 mM ammonium bicarbonate, 13.3 ug/ml trypsin (Sigma).

Samples were separated by SDS-PAGE and stained with Coomassie stain. After de-staining, bands were cut out with a clean scalpel and then cut into 1 mm cubes and kept in ddH_2O overnight. The gel pieces were incubated twice with 500 µl Buffer A for 30 mins at room temperature to remove SDS. The gel pieces were incubated with 500 µl Buffer B at 30°C to reduce the protein. The gel pieces were washed 3 times in 500 µl buffer A. The gel pieces were incubated with 500 µl Buffer C for 20 minutes in the dark to alkylate the cysteines. The gel pieces were washed 3 times with buffer D and centrifuged at 13000 rpm for 2 mins. The gel pieces were washed 3 times in 300 µl acetonitrile. The acetonitrile was removed and gel pieces were air dried. 30 µl of trypsin digestion buffer was added for 30 minutes at 4°C until the gel pieces swelled. More trypsin digest buffer was added if the gel pieces were not completely submerged. Tubes were sealed with parafilm and incubated at 30°C for 16 hrs. Samples were sonicated in a sonicator bath for 3 mins and used immediately or kept at -80°C for later use.

2.7.1| Matrix preparation and analysis

10 mg CHCA was mixed with 400 µl ddH_2O, 100 µl 3% (v/v) TFA and 500 µl acetonitrile. The mixture was centrifuged at 5000 rpm for 1 min and 0.5ul supernatant and 0.5ul peptide sample were spotted onto a stainless steel 100 sample plate (Applied Biosystems) and allowed to air dry. The plate was inserted into the Voyager DE-STR (applied Biosystems) or a Bruker UltraflexExtreme MALDI ToF/ToF and data collected as advised by the manufacturer. Mass peaks were analysed in Data Explorer software (Applied Biosystems) or DataAnalysis software (Bruker). The mass of the fragments was calibrated according to known trypsin peaks detected in the sample or from a peptide calibration standard (Bruker). The Mascot server (http://www.matrixscience.com/server.html) or the ProteinProspector tool MS-FiT (protein identification using MS-FiT) was used to identify the protein fragments and to assign a protein based on coverage and unique peptides.

2.8| Biophysical Methods

2.8.1| Size Exclusion Chromatography – Multi-Angle Light Scattering (SEC-MALS):

Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) was performed using an ÄKTA micro HPLC with a Superdex 200 10/300 GL column coupled with an on-line UV detector, Viscotek MALS-20 scattering detector (Malvern) and refractive index detector Viscotek VE3580 (Malvern). The Superdex column was equilibrated with GF buffer overnight (0.5 ml/min flow rate). 100 µl BSA at 1 mg/ml was injected and the monomer peak (66 kDa) used as a mass calibrant. Samples were injected in 100 µl volumes at concentrations varying from 0.5 mg/ml to 10 mg/ml. Analysis of the mass was performed using the OmniSEC software package (Malvern). Sample runs and analysis were performed with Dr Martin Wear.

2.8.2| Thermal Denaturation Assay (TDA)

A Biorad iQ5 Multicolor Real-Time PCR detection system was used for all thermal denaturation assays. The filters were set for excitation at 475 nm and emission at 575 nm, optimal for the SYPRO® orange protein gel stain (Life Technologies). The assay was performed in polypropylene, clear 96 well iQ Real-Time PCR plates (Bio-Rad) with Microseal® 'B' optical adhesive seals (Bio-Rad). The parameters in the iQ5 optimal system software version 2.1 was as follows:

Table 2.9| Bio-Rad iQ5 optimal system software 2.1 parameters for each TDA run

Parameter	Setting
Sample volume	50 µl
Fluorophore	SYBR
Probe/Primer	SYBR
Units	copy number
Seal type	film
Whole plate loading	on
Temperature range	20-80°C with 0.5°C increments and a 30 second dwell time

The concentrations of protein used were optimised to identify the best signal to noise ratio. Each well contained: testing buffer, protein, 5x SYPRO Orange to a final volume of 50 µl. The different testing buffers used in the study are detailed in table 2.10.

Table 2.10| Concentration of protein used and buffers tested in the TDAs

Protein	Concentration (µM)	Variables tested
LSH	1	pH (3-10), Glycerol (0-20% v/v), NaCl (0-500 mM), ATP (0-2 mM), ADP (0-2 mM), AMP (0-2 mM), $MgCl_2$ (0-20 mM), $ZnCl_2$ (0-20 mM), $CaCl_2$ (0-20 mM) DTT (1 mM)
LSH^{K237Q}	1	pH (7), Glycerol (0-10% v/v), NaCl (0-500 mM), ATP (0-1 mM), ADP (0-1 mM, AMP (0-2 mM), DTT (1mM)
LSH^{1-176}	1-20	pH (7), Glycerol (10%), NaCl (500 mM), DTT (1 mM)

Results were analysed and interpreted using Microsoft excel and Graphpad Prism (Graphpad software Inc.).

2.8.3| Dynamic light scattering (DLS)

Dynamic light scattering (DLS) was performed using a Zetasizer Auto Plate Sampler (Malvern) using 384 polypropylene plates (Corning). Protein concentration was typically 1 mg/ml and the sample was centrifuged down for 30 mins at 13000 rpm before 60 µl was loaded in a well. Protein buffer (50 mM HEPES pH 7, 10% (v/v) Glycerol, 100-500 mM NaCl and 1 mM DTT) was used as a blank. Each sample was measured three times following standard operating procedures as part of the Zetasizer software. All analysis was performed using Zetasizer software.

2.9| Structural biology methods

2.9.1| Protein sequence alignment, structural prediction and analysis software

Protein sequence alignment was performed on protein sequences taken from UNIPROT (find these sequences). Clustal Omega was used for sequence alignment with the parameters set to Clustal w/o numbers and default settings.

LSH from *Homo sapiens*:	Q9NRZ9
LSH from *Mus musculus*:	Q60848
LSH from *Danio rerio*:	B7ZD98
LSH from *Xenopus laevis*:	Q6DD35
SNF2 from *S. cerevisiae*:	P22082
CHD1 from *M. musculus*:	P40201
ISWI from *Drosophila melanogaster*:	Q24368

LSH disorder prediction was performed using the IUPred server (http://iupred.enzim.hu/). The long disorder setting was chosen and raw data was made into a graph using Graphpad Prism (Graphpad software Inc.).

LSH *ab-initio* structural predictions were performed using the I-TASSER structural prediction server http://zhanglab.ccmb.med.umich.edu/I-TASSER/ (Yang *et al.*, 2015), the Phyre server (Kelley *et al.*, 2015), or the SAXSTER server (dos Reis, Aparicio and Zhang, 2011). Chd1 – PDB ID: 3MWY. Rad54 – PDB ID: 1Z3I. MtSNF2 – PDB ID: 5hzr. All 3D structural images were generated using the PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC or the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311)(Pettersen *et al.*, 2004).

2.9.2| Chemical crosslinking – Mass Spectrometry (XL-MS)

Crosslinking experiments were carried out using the amine-to-amine crosslinker BS^3 (Thermo Fisher Scientific). Zero-length crosslinking was carried out with EDC (Thermo Fisher Scientific) in the presence of Sulfo-NHS (Thermo Fisher Scientific).

BS^3 intramolecular crosslinking experiments:

BS^3 was made at 5 mg/ml concentrations in the reaction buffer: 50mM HEPES pH 7, 10% Glycerol (v/v), 500 mM NaCl, 1 mM DTT. LSH was diluted to 1 mg/ml in reaction buffer.

BS^3 was diluted with reaction buffer into a series of concentrations. 10 µl of LSH at 1 mg/ml was mixed with 10 µl of BS^3 at the different dilutions defined in table 2.11 and incubated on ice for 2 hrs.

Table 2.11| BS^3 concentrations for incubation with LSH at different weight:weight ratios

Final BS^3 concentration	Volume of BS3 (5 mg/ml)	Volume of reaction buffer	To use for the weight ratio LSH:BS3
0.1 mg/ml	5 µl	245 µl	10:1
0.25 mg/ml	5 µl	95 µl	4:1
0.5 mg/ml	5 µl	45 µl	2:1
1 mg/ml	5 µl	20 µl	1:1
2 mg/ml	10 µl	15 µl	1:2
3 mg/ml	10 µl	6.67 µl	1:3
4 mg/ml	10 µl	2.5 µl	1:4
5 mg/ml	0 µl	0 µl	1:5

After incubation 6 µl of 4x SDS-PAGE loading buffer was added and samples were immediately boiled. 5.2 µl of the 26 µl reactions were analysed by SDS-PAGE and Coomassie staining. For scale up for mass spectrometry up to 60 µg of LSH was crosslinked with BS3 before undergoing trypsin digestion (section 2.7).

EDC-NHS crosslinking experiments:

EDC and NHS were made to 1.5 mg/ml and 3 mg/ml concentrations respectively in the reaction buffer: 50 mM MES pH 5.5, 500 mM NaCl, 10% (v/v) Glycerol, 1 mM DTT. LSH was used at 1.5 mg/ml. The volumes required for a series of LSH:EDC:Sulfo-NHS weight ratios is defined in table 2.12.

Table 2.12| Volumes for LSH:EDC:Sulfo-NHS weight ratios

LSH/EDC/Sulfo-NHS ratio	2:1:2	1:1:2	1:2:4	1:3:6	1:5:10	LSH control
LSH (1.5 µg/µl)	2 µl	2 µl	2 µl	2 µl	2 µl	2 µl
EDC (1.5 µg/µl)	1 µl	2 µl	4 µl	6 µl	10 µl	0 µl
Sulfo-NHS (3.0 µg/µl)	1 µl	2 µl	4 µl	6 µl	10 µl	0 µl
MES reaction buffer	28 µl	26 µl	22 µl	18 µl	10 µl	30 µl

Ammonium bicarbonate (2 µl of 2.7 M) was added to quench the reaction, then 10 µl of 4x SDS-PAGE sample buffer was added. Samples were boiled and analysed by SDS-PAGE and Coomassie staining. Experiments using the ratios of 1:1:2 and 1:2:4 were scaled up to use 30 µg of LSH each, before undergoing trypsin digestion (section 2.7).

Preparation of samples for mass spectrometry:

The digested peptides were desalted using C18-StageTips (Rappsilber, Ishihama and Mann, 2003) and analysed on a LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) by Dr Juan Zou (Rappsilber lab). The raw mass spectrometric data files were processed into peak lists using MaxQuant (Cox and Mann, 2008a) with default parameters, except for “FTMS top peaks per 100 Da” which was set to 100 and “FTMS de-isotoping” was disabled. The peak lists were searched against the sequences of the protein using Xi software (Giese, Fischer and Rappsilber, 2015) for identification of cross-linked peptides and non-cross-linked linear peptides. Search parameters were as follows: MS accuracy - 6 ppm; MS2 accuracy - 20 ppm; enzyme - trypsin; specificity - fully tryptic; allowed number of missed cleavages - four; fixed modifications - carbamidomethylation on cysteine; variable modifications - oxidation on methionine. 46 auto-validated spectra matches with score cut-off of 7.0 to support 33 link pairs (Table 5.1).

2.9.3| Negative-staining EM

LSH negative staining EM grids and imaging was performed by Dr Giuseppe Cannone. The analysis was co-performed with Dr Giuseppe Cannone.

LSH at a concentration of 1 mg/ml (9.7 µM) was diluted 1:50 and 1:100 in SEC buffer for negative staining on EM grids. Each copper 400 mesh 5 nm carbon grid (AGAR) was charged in a SC7620 Sputter Coater (Quorum) for 60 secs. 2 µl of sample was spotted onto the grid and left to settle for 2 mins. Filter paper (Whatman 50) was used to remove excess sample. The grid was desalted by washing 3 times in ddH_2O. The grid was washed 3 times in 2% (w/v) Uranyl Acetate, with 5 mins incubation for the last wash. After removal, the grid was air dried for 10 mins and stored in the carbon grid box at room temperature. Grids were imaged on a FEI F20 or a JEOL-1400 transmission electron microscope. Images were saved as 8-bit images using imageJ (Schneider, Rasband and Eliceiri, 2012) and boxes of particles were made using EMAN boxer software before 2D class averaging being made in EMAN 2.1 software (Tang *et al.*, 2007).

2.9.4| Small Angle X-ray Scattering (SAXS)

SAXS data was collected at the European Synchrotron Radiation Facility (ESRF) at the BM29 BioSAXS beamline (λ = 0.931 Å) on a Pilatus 1M detector (Pernot *et al.*, 2013). The sample to detector distance was 2.867 m and scattering data were collected within the momentum transfer range (q) of 0.0025-0.5 Å-1. All samples in SEC buffer (50 mM HEPES pH 7, 10% (v/v) Glycerol, 500 mM NaCl, 1 mM DTT) were centrifuged down at 13000 rpm for 30 mins. To check for concentration dependent effect during sampling serial dilutions of the 10 mg/ml stock were made (5, 2.5, 1.25, 0.625 mg/ml) and concentrations were measured using UV spectroscopy (section 2.4.3). Ligands were made in SEC buffer at 50 mM concentrations and diluted 1-10 mM into protein samples.

For LSH:DNA complexes – LSH was mixed with annealed 20 bp dsDNA (SAXS 20 bp dsDNA) in SEC buffer (with 100 mM NaCl instead of 500 mM NaCl) to equimolar concentrations of 20.61 µM. BSA (A7638-Sigma) made fresh in 50 mM HEPES pH 7.5 at 4.25 mg/ml using UV spectroscopy (section 2.4.3) was used as the mass standard. Each 50 µl sample was flowed (0.1 ml/min) during detection through a 1 mm diameter quartz glass capillary and kept at 4°C to avoid radiation damage. The molecular masses of the different LSH proteins and LSH:DNA complex were calculated from the extrapolated intensity at zero angle I(0), obtained with GNOM, by comparison with the I(0) of BSA (molecular mass 66 kDa). Data were collected in multiple 30 sec frames and auto-processed in EDNA (inspected and averaged, normalised to the incident beam intensity and the scattering of the buffer subtracted).

2.9.5| SAXS data analysis and modelling

SAXS data was analysed using the ATSAS 2.1 suite (Petoukhov *et al.*, 2007, 2012). The SASPlot analysis tool was used for Guinier analysis and to determine the P(r) distribution. Low-resolution structure models were constructed by a*b initio* modelling by simulated annealing using Dammin (Svergun, 1999), Dammif (Franke and Svergun, 2009), GASBOR (Dmitri I Svergun, Petoukhov and Koch, 2001) and EOM (Bernado *et al.*, 2007). A representative gasbor model was used for representation.

2.9.5| Protein crystallisation trials

Manual crystallisation screening was performed using 24 well sitting/hanging drop (Molecular Dimensions) or 48 well sitting drop plates (Hampton). Automated crystallisation screening was performed in 96-well MRC plates (Molecular Dimensions) using the Gryphon robot (Art Robbins) following manufacturer suggestions. Protein concentrations between 4-10 mg/ml were used for screening. ADP was added to a final concentration of 1 mM from a stock of 50 mM in 250 mM HEPES pH 7. Preparations contained 0.1 µl of protein and 0.1 µl of well buffer and the well volume was 60 µl. Once pipetting had finished, plates were quickly sealed using a XCS-384 pressure sealer (FluidX) and stored at 6°C or 16°C.The commercial screens used in this study were as follows: Cations, Nucleix (Qiagen), Index, Natrix, PEG/Ion (Hampton), ProPlex, Midas, Structure 1 and 2, JCSG+ and Morpheus (Molecular Dimensions). Automated imaging was performed on a CrysCam XY stage imaging system (Art Robbins). Manual imaging was performed using a Nikon Stereomicroscope (Nikon).

2.10| Preparation of nucleosomes

2.10.1| Annealing oligonucleotides

All oligonucleotides were ordered from Integrated DNA Technologies and resuspended in ddH_2O to a 1 mM stock before being checked using UV spectroscopy (section 2.4.4) and the concentration was adjusted to 0.5 mM. Equal volumes of complementary oligonucleotides were mixed gently with equal volumes of 2x TEN buffer (20 mM Tris pH 7.5, 100 mM NaCl, 2 mM EDTA). The mixture was heated to 90°C for 10 mins in a heat block, with a hot block covering the Eppendorf lids to prevent condensation. The heat block was turned off and blocks were removed with slow cooling to room temperature for ~4 hrs. The annealed oligonucleotides were analysed using native PAGE and stained with methylene blue stain for 30 mins followed by destaining in ddH_2O until background stain was no longer visible. Scanning was performed on A GELDOC EZ system (Bio-Rad) under the Coomassie setting. If the DNA was IR700 labelled the gel was scanned on an Odyssey scanner (Li-Cor).

2.10.2| Histone octamer formation

Buffers used for histone octamer formation:

Unfolding buffer

7 M guanidine-HCl, 20 mM Tris pH 7.5, 10 mM DTT.

Refolding Buffer

10 mM Tris pH 8, 2 M NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol.

Reconstitution buffer

10 mM Tris, 0.5 mM EDTA.

Core histone proteins (H2A, H2B, H3 and H4) were expressed and purified from *E.coli* (gifted by the Voigt lab). The histones were purified from inclusion bodies under denaturing conditions by ion exchange chromatography. Purified core histone proteins were mixed in equimolar amounts in unfolding buffer on ice, before dialysis against 1 L refolding buffer at 4°C overnight. After centrifugation for 20 mins at 13000 rpm and 4°C, 250 µl of 2 mg mixed histones was injected into a pre-equilibrated Superdex 200 10/300GL column (GE Healthcare) attached to a ÄKTA HPLC system. 0.5 ml fractions were collected and analysed using SDS-PAGE and Coomassie staining. Protein concentration was determined by Bradford assay and UV spectroscopy.

2.10.3| Nucleosome DNA preparation

DNA fragments were amplified from a PBS601 plasmid (gifted from the Voigt lab) using primers to make DNA fragments with differing overhangs from the 147 bp core widom histone octamer binding sequence. The forward primer contained a 5` IR700 fluorophore for EMSA experiments. DNA was amplified using Phusion® High-Fidelity DNA Polymerase (NEB) to amplify a small amount of DNA according to manufacturer's protocol using the following PCR cycling conditions:

Table 2.13| PCR cycling conditions for amplifying nucleosome DNA

Step	Temperature	Time	
1	98°C	30 secs	
2	98°C	20 secs	x35
3	48°C	30 secs	
4	70°C	30 secs	
5	70°C	3 mins	
6	4°C	Pause	

Product size was estimated with 2% AGE by comparison with a 50 bp DNA ladder. The PCR product was used a template for scale-up with HMT polymerase typically in 100 µl reaction volumes for each PCR reaction (see 2.3.2). 96-well plates were used and typically generated 150-200 µg of DNA after purification. The PCR products were purified using the PCR product purification protocol (section 2.3.6). One spin column was used for 0.5 ml of PCR reaction because the QIAGEN spin columns only have a 10 µg binding capacity.

2.10.4| Nucleosome reconstitution

1 µg of DNA was mixed with varying amounts of histone octamers (1, 2, 3 or 4 µg) in a total volume of 100 µl refolding buffer. 5 M NaCl was added to keep the ionic strength of the mixture at 2 M NaCl. Dialysis was performed using slide-a-lyzer 10 kDa MWCO units (Life Technologies) floating on 200 ml of refolding buffer with 800 ml reconstitution buffer added slowly (1 ml/min) overnight at 4°C. The next morning fresh dialysis in 200 ml reconstitution buffer was performed for 2 hrs. Samples were analysed on a 6% TBE gel acrylamide gel and either visualised using an Odyssey scanner (Li-Cor) or the gel was incubated for 15 mins in SYBR Safe (Life Technologies) following the manufacturer's suggestions and scanned on the GELDOC EZ (Biorad) using the SYBR Safe setting.

2.10.5| DNA EMSA

Buffers used for the EMSA:

Binding buffer

50 mM HEPES pH 7, 100 mM NaCl, 10% Glycerol, 1 mM DTT.

1x Tris-CAPS buffer

60 mM Tris, 40 mM CAPS.

DNA or Nucleosomes with a 5' IR700 dye were used for all EMSA experiments. Reaction volumes were 20 µl. DNA or reconstituted nucleosomes (5 nM-1 µM final concentration) were mixed in binding buffer, before the addition of LSH at increasing concentrations (5 nM-10 µM) for 30 mins in ice. For analysis of DNA binding by LSH a 2% agarose gel made in 1x Tris-CAPS buffer was used. For analysis of nucleosome binding by LSH a 0.5% agarose gel made in 1x Tris-CAPS buffer was used. All 20 µl of sample was loaded and gels were run at 80 V for 60 mins at 6°C. Gels were immediately visualised on an Odyssey scanner (Li-Cor) at a wavelength of 700 nm and intensity necessary to prevent saturation of signal (varied between 0.5-7). Quantitative analysis was performed in Image studio lite (3.1) and graphs made in Graphpad Prism. Nucleosome competition experiments with LSH and DNMT3B were carried out using the conditions described above. Reconstituted nucleosomes (5 nM) were incubated primarily with 160 nM LSH at 4°C, before incubation with increasing concentrations of DNMT3B (5-160 nM) for 30 mins at 4°C or *vice versa* before being analysed by AGE as above.

2.10.6| Multiprotein AGE

This method is adapted from Kim, 2011. This method uses the same buffers from section 2.10.5. For analysis of LSH-6His and StrepII-6His-DNMT3B or LSH^{1-176} and StrepII-6His-DNMT3B complex formation a 0.5% agarose gel made in 1x Tris-CAPS buffer was used. Proteins were mixed in binding buffer, each to a final concentration of 1 µM in 10 µl and incubated for 30 minutes on ice. All 20 µl of the sample were loaded and the gels were run at 80 V for 60 mins at 6°C. For visualisation of proteins, gels were Coomassie stained and scanned on A GELDOC EZ system (Bio-Rad) using the Coomassie setting.

2.10.7| Fluorescence polarisation (FP) DNA binding assay

All samples were analysed using a SpectraMax M5 multiplate reader (Molecular Devices). Sample volumes of 100 µl were used in black opaque polystyrene 96-well plates (Corning). 50 nM 5'6FAM labelled 12 bp dsDNA was used as this gave good signal without saturating the detector. Excitation was set at 493 nm and emission at 521 nm wavelengths for 6FAM DNA. A titration of LSH was performed from 50 nM-10 µM in binding buffer (50 mM HEPES, 10-50 mM NaCl, 10% Glycerol, 1 mM $MgCl_2$, 1 mM DTT). Mixtures were incubated on ice for 15 mins before being scanned. Buffer only readings were used to subtract background. Fluorescence anisotropy was measured. For competition assays, non-labelled 36 bp dsDNA was added into the reaction at a concentration of 10 µM until anisotropy decreased back to DNA only levels. The analysis was performed using Graphpad Prism (Graphpad software Inc.).

The anisotropy is calculated using the following set of equations:

$$K_D = \frac{LF \text{ x } (RT - B)}{B} \qquad \text{Equation (2.2)}$$

K_D = Dissociation constant
LF = unbound ligand concentration
B = Receptor: Ligand complex concentration
RT = Total receptor concentration
RT-B = Free receptor concentration

$$B = \frac{RT \text{ x } (LT - B)}{K_D + (LT - B)} \qquad \text{Equation (2.3)}$$

The quadratic is used to solve B in the following equation:

$$B = \frac{(LT + K_D + RT) - \sqrt{(RT + K_D + LT)^2 - 4LT \text{ x } RT}}{2} \qquad \text{Equation (2.4)}$$

When the anisotropy values are used directly the following equation is used:

$$B = LT \text{ x } \frac{A - Af}{Ab - Af} \qquad \text{Equation (2.5)}$$

LT = Total free ligand
A = Anisotropy
Af = Anisotropy of the free ligand
Ab = Anisotropy of the bound ligand

When calculating anisotropy the formula to rearranged to:

$$A = A_0 + (A_b - A_0)(K_D + [L]_t + [P]_t) - \frac{\sqrt{(K_D+[L]_t+[P]_t)^2-4[L]_t[P]_t}}{2[L]_t} \quad \text{Equation (2.6)}$$

A_0 = Anisotropy of free 12 bp 5'^{6FAM} dsDNA
A_b = Anisotropy of 12 bp 5'^{6FAM} dsDNA bound to LSH-6His
$[L]_t$ = Total 12 bp 5'^{6FAM} dsDNA concentration
$[P]_t$ = Total LSH-6His concentration
The **K_D** is defined as the concentration of the protein at which 50% is bound to the ligand

Equation 2.5 is used by the plate reader software when the plate reader measures the anisotropy.

2.11| Primers, Plasmids and oligonucleotides

Table 2.14| A List of primers used in this study. IRD700 stands for IR700 dye

PRIMER (5'- 3')	SEQUENCE
OSV-F-BamHI-6His-TEV LSH	TTTAGGATCCATGCACCATCACCATCACCACGAGAACCT ATATTTCCAAGGTATGGCCGAACAAACGGAGCCTGCG
OSV-F-BamHI-6His LSH	TTTAGGATCCATGCACCATCACCATCACCACATGGCCGAA CAAACGGAGCCTGCG
OSV-F-BamHI-6His-TEV LSH$^{200-821}$	TTTAGGATCCATGCACCATCACCATCACCACGAGAACCTA TATTTCCAAGGTAAGCATTTCACAGGAGGAGTAATG
OSV-F-BamHI-6His LSH$^{200-821}$	TTTAGGATCCATGCACCATCACCATCACCACAAGCATTTC ACAGGAGGAGTAATG
OSV-F-BamHI-LSH$^{200-821}$	CGTTGGATCCATGTTCACAGGAGGAGTAATGAGG
OSV-R-SalI-6His-LSH$^{200-821}$	CGTTGTCGACTCAGTGGTGATGGTGATGGTGAAATAAAC ATTCAGCACTGGAATC
OSV-F-BamHI-LSH	CGTTGGATCCATGGCCGAACAAACGG
OSV-F-BamHI-LSH$^{146-821}$	CGTTGGATCCATGTCTTCCACTACGAGTCTTTGTG
OSV-F-BamHI-LSH$^{162-821}$	CGTTGGATCCATGTCAAATAGTATGATTAAAGATAGATT GTC
OSV-F-BamHI-LSH$^{176-821}$	CGTTGGATCCATGCAGAACTCTAAATTCTTTTTTGACCCA
OSV-F-BamHI-LSH$^{178-821}$	CGTTGGATCCATGTCTAAATTCTTTTTTGACCCAG
OSV-F-BamHI-LSH$^{184-821}$	CGTTGGATCCATGCCAGTTCGGAAATGTAACGG
OSV-F-BamHI-LSH$^{193-821}$	CTTAGGATCCATGGTACCCTTTCAACAACCAAAGC
OSV-F-BamHI-LSH$^{196-821}$	GTTAGGATCCATGCAACAACCAAAGCATTTCACAGG
OSV-F-BamHI-LSH$^{209-821}$	GCTTGGATCCATGCATTTCACAGGAGGAGTAATG
OSV-F-BamHI-LSH$^{219-821}$	GCTTGGATCCATGCTTTGGGAAAATGGAATT
OSV-F-BamHI-LSH$^{226-821}$	GTTAGGATCCATGAATGGCATTTTAGCAGATGAAATGGG
OSV-R-SalI-LSH	GCTCGTCGACCTACTGCCTAACAGTTTGAGAC
OSV-R-SalI-6His-LSH	GCTCGTCGACCTAATGGTGATGATGGTGATGCTGCCTAA CAGTTTGAGAC
OSV-F-seq-LSH-0	GTATCGATTCGCGACCTACTC
OSV-F-seq-LSH-1	GGATTATTCATACCGTCCCACC

OSV-F-seq-LSH-2	GACCCAGTTCGGAAATGTAAC
OSV-F-seq-LSH-3	CTTGAAAAGCTTTGAGTCTTGG
OSV-F-seq-LSH-4	GGACATCTTGATGGATTATTGCC
OSV-R-seq-LSH	CTCTAGATTCGAAAGCGGCC
M13 reverse	CAGGAAACAGCTATGAC
M13 forward	CCCAGTCACGACGTTGTAAAACG
OSV-F-BamHI-StrepII-GAGA-6His-TEV-DNMT3B-1 (1st PCR step)	GGAGCTCATCACCACCATCACCACGAAAACGTCGAGCTG CAGGGAATGAAGGGAGACAGCAGACATC
OSV-F-BamHI-StrepII-GAGA-6His-TEV-DNMT3B-2 (2nd PCR step)	GCATGGATCCATATGTGGTCCCACCCACAATTTGAGAAG GGAGCTGGAGCTCATCACCACCATCAC
OSV-R-EcoRI-DNMT3B	GCTAGAATTCCTAAGCGTAATCTGGTACGTCG
OSV-F-Nhe1-StrepII-TEV-DNMT3B	GCTTGCTAGCATGTGGAGCCATCCGCAATTTGAAAAAGA AAACCTGTACTTCCAGGGAAAGGGAGACAGCAGACATCT G
OSV-R-SphI-DNMT3B	TACGACGTACCAGATTACGCTTAGGCATGCTTGC
OSV-F-seq-DNMT3B-1	CACTCGACGAAGACTTGATCAC
OSV-F-seq-DNMT3B-2	CATCAGTTGACTTGAGCCAGG
OSV-F-seq-DNMT3B-3	GCTTCTGAAGTCACCAACAAC
OSV-F-seq-DNMT3B-4	GATCTCTCTAACGTCAATCCTGC
OSV-F-IR700-25-Nuc	/5IRD700/GACCCAAGCGAACACCGG
OSV-F-IR700-12-Nuc	/5IRD700/ACCGGCACTGGGACAGGATG
OSV-F-IR700-6-Nuc	/5IRD700/ACTGGGACAGGATGTATATATGTGAC
OSV-F-IR700-0-Nuc	/5IRD700/ACAGGATGTATATATGTGACACG
OSV-F-IR700-26-Nuc	GCTTCACCTCGTGACCC
OSV-R-Nuc-26	CGATCTAGACCATGATGC
OSV-R-Nuc-0	CTGGAGAATCCCGGTGCC

Table 2.15| A list of plasmids used in this study

VECTOR	PLASMID NAME
MSCV	MSCVh-DNMT3B
pAM	pAMm-LSH (mouse synthetic LSH orf)
pAM	pAMm-LSHK237Q
pFL	pFL-LSH-6His
pFL	pFL-LSH^{K237Q}-6His
pFL	pFL-6His-TEV-LSH
pFL	pFL-6His-LSH
pFL	pFL-6His-TEV-LSH$^{200-821}$
pFL	pFL-6His-LSH$^{200-821}$
pFL	pFL-LSH$^{146-821}$-6His
pFL	pFL-LSH$^{162-821}$-6His
pFL	pFL-LSH$^{176-821}$-6His
pFL	pFL-LSH$^{178-821}$-6His
pFL	pFL-LSH$^{184-821}$-6His
pFL	pFL-LSH$^{193-821}$-6His
pFL	pFL-LSH$^{196-821}$-6His

pFL	pFL-LSH$^{200-821}$-6His
pFL	pFL-LSH$^{209-821}$-6His
pFL	pFL-LSH$^{219-821}$-6His
pFL	pFL-LSH$^{226-821}$-6His
pFL	pFL-polH-StrepII-6His-GAGA-TEV-DNMT3B
pFL	pFL-p10-StrepII-TEV-DNMT3B
pFL	pFL-p10-StrepII-DNMT3B-polH-LSH-6His
pET-28a	6His-TEV-LSH$^{1-176}$
pUC	pUC-601

Table 2.16| A list of oligonucleotides used in this study. IRD700 represents IR700 dye. 6-FAM represents Fluorescein. ^{Me} represents a methyl group attached to C.

Oligonucleotide	**Sequence**
12mer-R	AACGCACGCAGC
12mer-F	GCTGCGTGCGTT
12mer-F-IR700	/5IRD700/GCTGCGTGCGTT
12mer-F-6FAM	/56-FAM/GCTGCGTGCGTT
12mer-non-hairpin-IR700	/5IRD700/GTTGCGTGCTTT
Methylated-12mer-R	AAC^{Me}GCAC^{Me}GCAGC
Methylated-12mer-F-6FAM	/56-FAM/GCTGC^{Me}GTGC^{Me}GTT
Methylated-12mer-F-IR700	/5IRD700/GCTGC^{Me}GTGC^{Me}GTT
20mer IR700 F	/5IRD700/GTGGACTGCGTGCGTTAGTG
16mer IR700 F	/5IRD700/GGACTGCGTGCGTTAG
20mer R	CACTAACGCACGCAGTCCAC
16mer R	CTAACGCACGCAGTCC
X-LSH 20 mer R (CGend)	CGACGAACGCACGCAGCTCA
X-LSH 20 mer F (CGend)	CGTGAGCTGCGTGCGTTCGT
X-LSH 16 mer R (CGend)	CGTAACGCACGCAGCT
X-LSH 16 mer F (CGend)	CGAGCTGCGTGCGTTA

Chapter 3| Cloning, expression and purification of LSH

3.1| Introduction

In order to study the mechanism by which LSH function is regulated and what role its N-terminal domain has in regulating enzymatic function a large amount (several mg) of folded protein that is more than 95% pure is required for structural and biochemical investigation (Graslund, 2008; Almo *et al.*, 2013). LSH has been overexpressed in bacteria (personal communication), however, the protein was insoluble and refolding attempts were unsuccessful, potentially due to LSH being a multi-domain 96 kDa protein. Soluble, active LSH was successfully expressed using an insect cell system (Burrage *et al.*, 2012). Recombinant protein expression in insect cells involves a Baculovirus Expression Vector system (BEVs). Recombinant baculovirus is acquired from *Autographa californica* multinuclear polyhedrovirus (AcMNPV) to infect lepidopteron cells derived from *Spodoptera frugiperda* (isolates *Sf*9/*Sf*21) or *Tricloplusia ni* (HighFive™) (Almo *et al.*, 2013). LSH was cloned using the Bac-to-Bac insect cell expression system, expressed and purified on a small scale yielding less than 80% pure LSH (Burrage *et al.*, 2012). A higher homogeneity of protein sample would be required for biophysical characterisation, structural studies and biochemical assays.

I decided to use the MultiBac baculoviral expression system, developed by Imre Berger and colleagues (Fitzgerald *et al.*, 2006) instead of the Bac-to-Bac® baculovirus expression system adopted by Burrage et al. 2012. The MultiBac baculoviral expression system has the added advantage over the Bac-to-Bac system of expressing a yellow fluorescent protein (YFP) gene upstream of a polH promoter, allowing baculovirus infected insect cells to be monitored using fluorescent microscopy enabling visualisation of positively infected *Sf*9 cells (Fitzgerald *et al.*, 2006; Trowitzsch *et al.*, 2010). The MultiBac system also allows for cloning of binding partner proteins (e.g. DNMT3B) for multiprotein complex expression within a single transfective baculovirus through integratable transfer vectors which cannot be done with the Bac-to-Bac® system. Multiple genes of interest are cloned into transfer vectors, followed by integration into the baculoviral EMBacY Bacmid via Tn7 transposition using *E.coli* (Figure 3.1) (Fitzgerald *et al.*, 2006; Trowitzsch *et al.*, 2010). Transfection of the EMBacY bacmid in *Sf*9 insect cells causes virus production and recombinant protein expression under the polyhedron promoters p10 and polH (Figure 3.1) (Fitzgerald *et al.*, 2006). Figure 3.1 explains how the MultiBac system works.

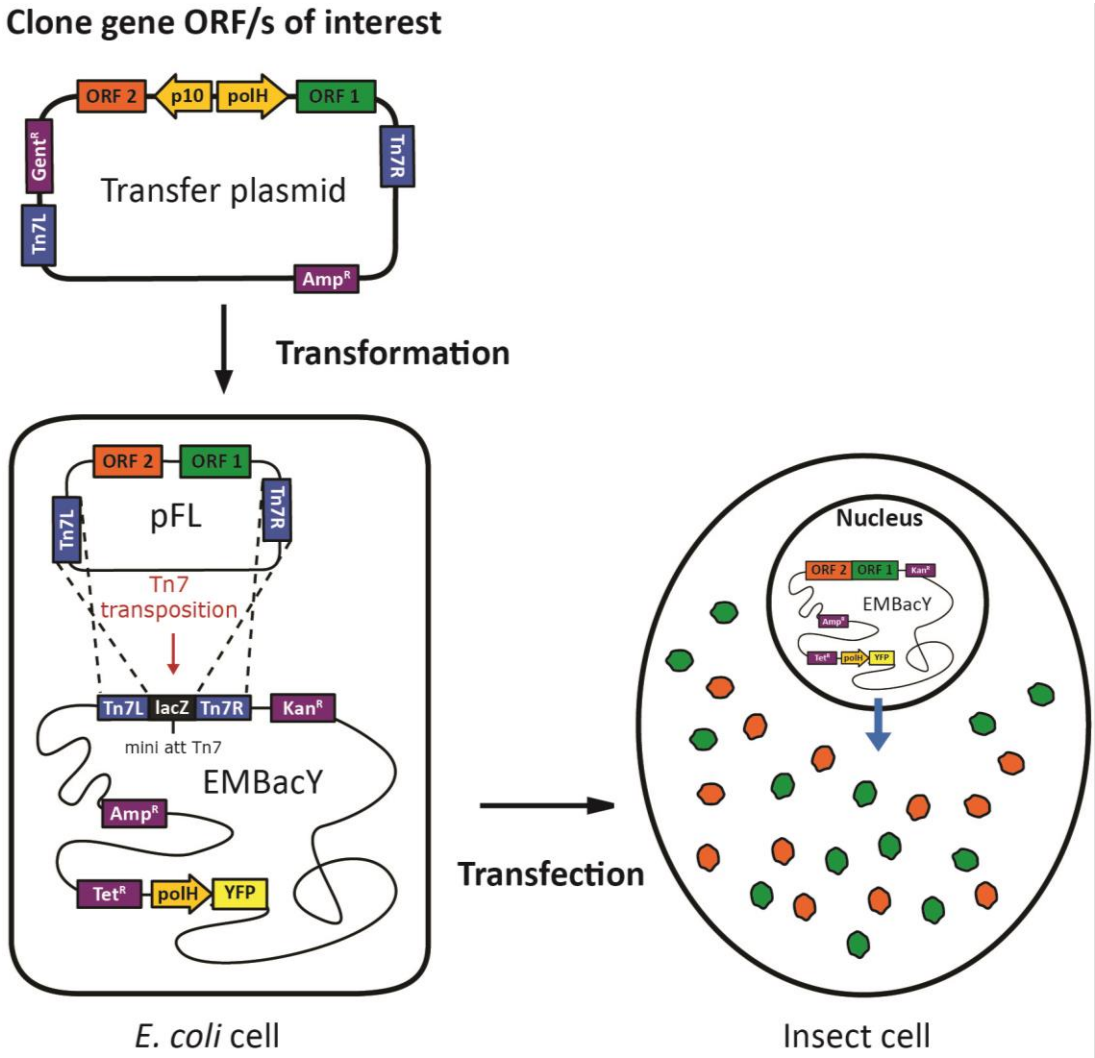

Figure 3.1| The MultiBac system for recombinant protein expression in insect cells. The ORF/s of interest is cloned into the transfer plasmid under control of a polH or p10 promoter in-between two flanking Tn7 sites and downstream of the polH promoter (not shown). The polH and p10 promoters cause overexpression of the ORF/s late on in insect cell infection. The transfer plasmid contains antibiotic resistance genes as selectable markers. The transfer plasmid containing the ORF/s of interest is then transformed into EMBacY bacmid containing *E. coli* cells. The EMbacY bacmid contains a mini-attTn7 site within the lacZ gene enabling Tn7 transposition of the flanking Tn7 recombination sites. Successful Tn7 transposition will knockdown the lacZ preventing β-galactosidase generation necessary to hydrolyse lactose causing white colonies to form in the presence of X-gal. Selectable antibiotic markers present on the transfer plasmid and EMBacY genome are used as further selective agents to the blue/white colony selection. The successfully transposed EMBacY bacmid is isolated from the bacterial host and used to transfect insect cell cultures for baculovirus and recombinant protein production of the protein of interest. The EMBacY bacmid also possesses a YFP ORF downstream of a polH promoter allowing polH expression to be monitored using fluorescence microscopy.

First, I decided to use the MultiBac system to express LSH alone. I decided to clone full-length LSH and the ATPase region LSH$^{200-821}$ as this region has been crystallised successfully for RAD54, CHD1 and SNF2. It would enable structural and biochemical functional studies to be tested with and without the N-terminal domain which would shed light on how this domain regulates the ATPase regions of LSH.

3.2| Results and Discussion

3.2.1| Cloning and Expression of LSH and LSH$^{200-821}$

Initially, full-length LSH and a truncated form LSH$^{200-821}$ comprising the core ATPase domain were cloned separately into a pFL plasmid (using restriction-based cloning) also encoding a 6His tag with or without a TEV cleavage site upstream of the LSH gene (under the control of the polH promoter) (Figure 2.1). The four variants of LSH bacmids were then transfected into *Sf*9 cells for generation of recombinant viruses and virus amplification. The EMBacY bacmid contains a YFP ORF upstream of a polH promoter (Figure 2.1). Although LSH and YFP are expressed under separate polH promoters, expression from both is presumed to be equivalent (Fitzgerald *et al.*, 2006). YFP expression can be monitored in live cells by fluorescence microscopy, enabling successful transfections to be identified and the extent of viral infection to be measured over time (Figure 3.2A).

Following transfection of all LSH bacmid variants into *Sf*9 cells, YFP expression from a small number of cells was seen after 2 days. After 3 and 4 days post-infection YFP signal increased significantly, indicating that the majority of cells were infected with baculovirus and overexpressing proteins from the viral polH promoters (Figure 3.2A). The baculovirus after 4 days defined as V0 was collected and the growth of infected cells was continued with fresh media. The majority of cells showed YFP overexpression 5 days post-transfection and were used to verify LSH expression in *Sf*9 cells by Western blot analysis of extracted proteins.

LSH has an NLS in the N-terminal domain in the region around LSH^{112} to allow transport into the nucleus (Lee et al. 2000, Yan et al. 2003). The NLS is not present in the two LSH$^{200-821}$ variants. Therefore nuclear and cytoplasmic extracts were collected from all four LSH variants expressed in *Sf*9 cells to determine if the NLS domain was functional.

Western blots were performed using primary antibodies against LSH (Figure 3.2B) and the 6Histag (Figure 3.2C). Secondary antibodies coupled with an infrared dye were used for visualisation. A greater accumulation of LSH in the nuclear extract (lanes 2 and 3) compared to the cytoplasmic extract (lanes 7 and 8) of the two LSH variants was expected and can be seen on the anti-LSH and anti-6His Western blots (Figure 3.2B and C). The signal on the anti-His Western blot was weaker than that on the anti-LSH western blot, likely due to weaker antibody binding.

As the $LSH^{200-821}$ variants do not contain the NLS the band intensity appeared equivalent in nuclear and cytoplasmic fractions on the anti-LSH Western blot (Figure 3.2B). The expression of both $LSH^{200-821}$ variants was considerably weaker than full-length variants on the anti-LSH Western blot (Figure 3.2B), with no detection seen on the anti-6His Western blot (Figure 3.2C).

To establish if the 6Histag was accessible for binding to Ni-NTA lysates containing the 6HisTEV-LSH and 6HisTEV-$LSH^{200-821}$ proteins underwent IMAC purification (Figure 3.2D). Unbound material was washed away with 10mM imidazole, and bound material was eluted with 1 M Imidazole. The presence of 6HisTEV-LSH in the eluted fractions confirmed that the 6histag was available for binding (Figure 3.2D).

The absence of a band at ~70 kDa in elutions from the 6HisTEV-$LSH^{200-821}$ IMAC purification suggested that the 6Histag is inaccessible or not present (Figure 3.2D). $LSH^{200-821}$ was temporarily abandoned due to weak expression and no evidence of a 6Histag whereas 6HisTEV-LSH was carried forward and provides the potential production of a native LSH by cleavage of the 6Histag.

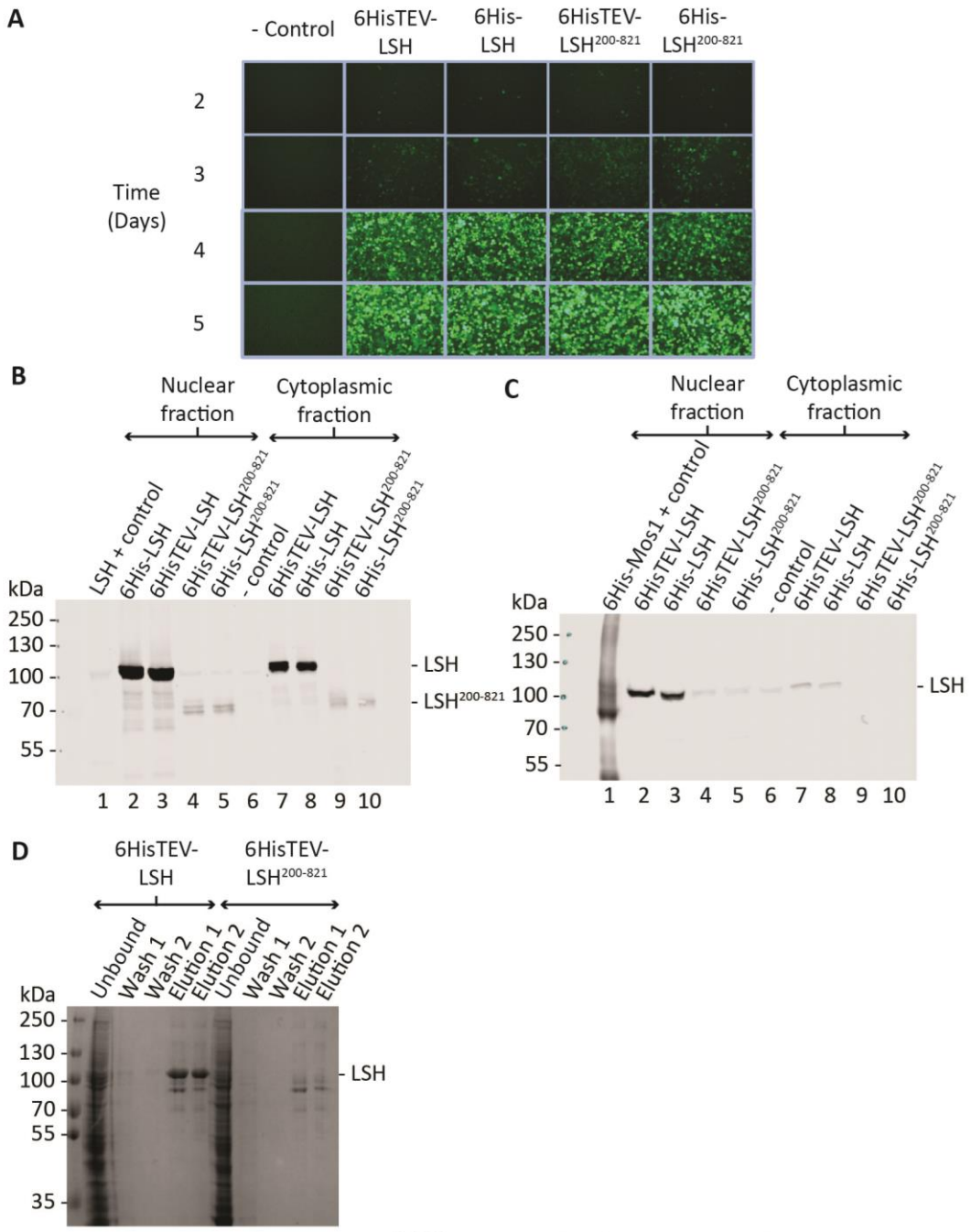

Figure 3.2|Expression of LSH and LSH$^{200-821}$ N-terminal 6His tagged proteins.
A| Fluorescence microscopy images of *Sf9* cells at x400 magnification. Transfection of *Sf9* cells with the four LSH bacmid constructs. YFP co-expression under a downstream polH promoter was used to monitor transfection. The control was untransfected *Sf9* cells.
B|Anti-LSH Western blot of proteins extracted from *Sf9* cells transfected with the four LSH bacmid constructs. The LSH+ control was nuclear extract from mouse embryonic fibroblasts (lane 1). The negative control (lane 6) is protein extract from untransfected *Sf9* cells.
C|Anti-6His Western blot of nuclear and cytoplasmic fractions of all four variants of LSH. The 6His+ control was the 6His-Mos1 protein (lane 1). The negative control (lane 6) is protein extract from untransfected *Sf9* cells.
D| Coomassie stained SDS-PAGE of fractions from an IMAC purification of 6HisTEV-LSH and 6HisTEV-LSH$^{200-821}$ lysate. The bands just above 100 kDa in elutions 1 and 2 correspond to 6HisTEV-LSH. No LSH$^{200-821}$ protein was observed in the elution fractions.

3.2.2| Amplification of baculovirus and V2 characterisation

In order to scale up 6HisTEV-LSH expression, the V0 baculovirus had to be amplified to larger volumes through sequential infection of larger volumes of *Sf*9 cells. Initial V0 virus stock of 6HisTEV-LSH was used to generate a V1 stock by infecting 50ml of *Sf*9 cells with V0. After 3 days of incubation, supernatant containing virus was used to generate V2 stock by infecting 200 ml of *Sf*9 cells with V1 baculovirus. Supernatant containing virus was collected after 3 days of incubation. The V2 stock was to be used for large scale 6HisTEV-LSH expression; therefore, the volume of V2 virus needed and length of incubation with *Sf*9 cells necessary for optimal 6HisTEV-LSH expression needed to be determined. To characterise the V2 stock, four baculovirus dilutions were used in separate 60 ml *Sf*9 cell cultures at V2:*Sf*9 cell volume ratios: 1:50, 1:100, 1:150 and 1:200. For each ratio, 1.5 ml of cells were taken at 1, 2, 2.5, 3, 3.5, 4, 4.5 and 5 days post-infection and imaged using fluorescence microscopy to identify YFP expression (a marker for polH overexpression) before being pelleted and frozen.

YFP expression peaks at ~3-3.5 days post-infection for all V2:*Sf*9 cell volume ratios tested (Figure 3.3A). However, as the density of cells between images is variable, it is difficult to determine the optimal expression of YFP (Figure 3.3A). Therefore, I quantified LSH expression by Western blot analysis from each frozen cell pellet (Figure 3.2B). The anti-LSH antibody was used to detect LSH expression and anti-Tubulin antibody was used as a loading control for normalisation of each sample. The intensity of LSH protein peaked at ~3-4 days post-infection, with a modest decrease in intensity thereafter until day 5, establishing that protein expression is stable for at least ~1.5 days (Figure 3.3B and Figure 3.3C). Normalisation was relative as each protein was measured at different wavelengths (Figure 3.3B). From Figure 3.3C the LSH band intensity begins to plateau for all V2: *Sf*9 cell volume ratios at day 3. Optimal expression appears between 4-4.5 days post-infection for all V2:*Sf*9 cell volume ratios. Therefore I concluded that the optimum conditions for 6HisTEV-LSH expression were using a V2:*Sf*9 cell volume of 1:100 or 1:150 and harvesting the *Sf*9 cells 4 or 4.5 days post infection.

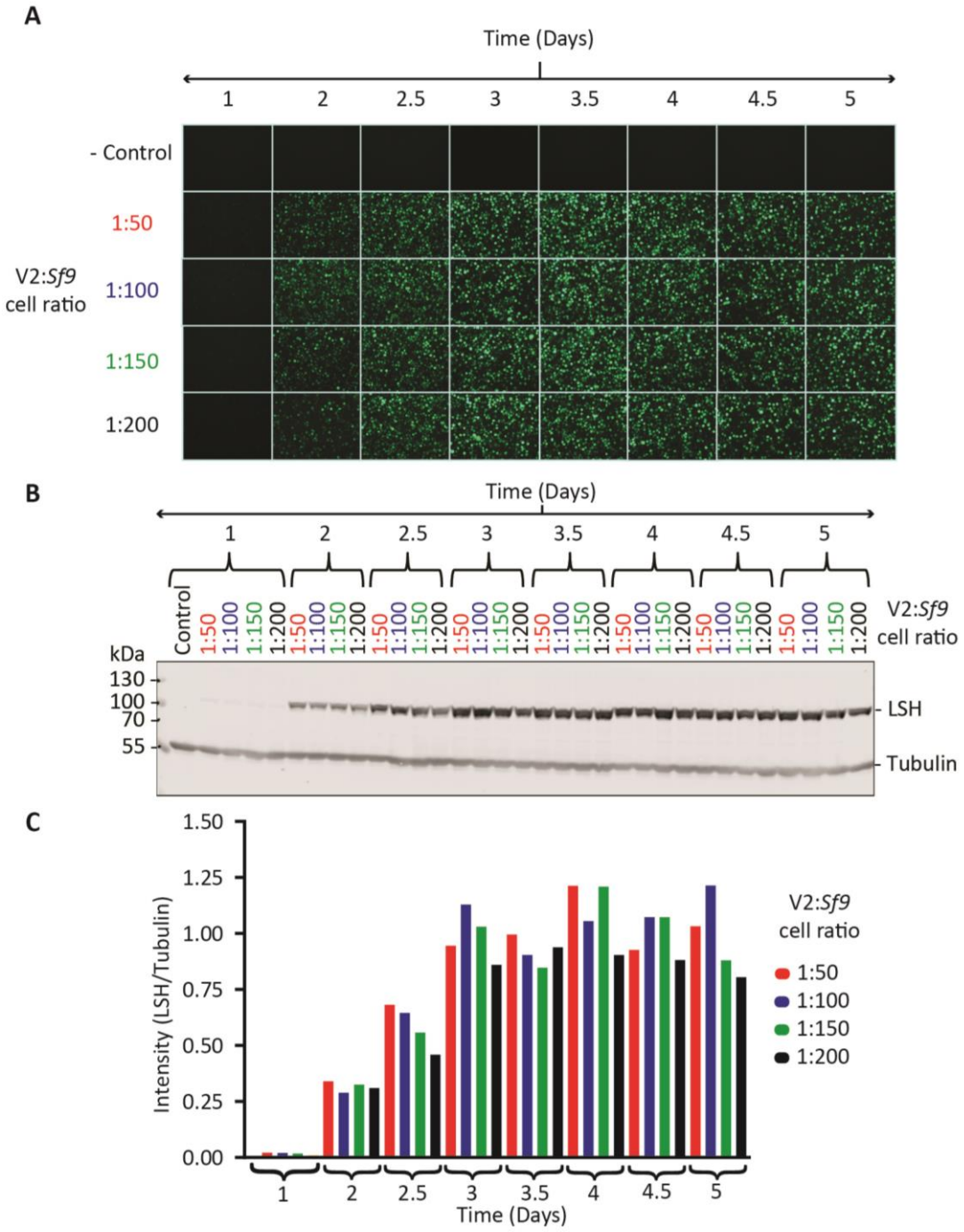

Figure 3.3| Characterisation of 6HisTEV-LSH V2 baculovirus.
A| Fluorescence microscopy images of infected *Sf9* cells at x400 magnification taken over 1 to 5 days post infection. Infection of *Sf9* cells at 1x10^6/ ml with varying volumes of 6HisTEV-LSH V2 baculovirus. YFP co-expression under control of a downstream polH promoter was used to monitor transfection. The negative control was untransfected *Sf9* cells.
B| Western blot of LSH and Tubulin, from whole cell lysate of each sample imaged in **A**. The band at ~100 kDa is LSH and the band at ~55 kDa (Tubulin) was used as a loading control.
C| Graph showing the relative normalised intensities of LSH/Tubulin at 4 different infection ratios over the time course.

3.2.3| Comparing purification of LSH protein constructs with an N or C-terminal 6Histag.

To purify LSH a primary step of phosphocellulose cation exchange chromatography followed by a secondary IMAC step was used. The phosphocellulose cation exchange step has been used to purify DNA binding proteins from DNA and was also used due to its large protein binding capacity (50-100 mg/ml phosphocellulose slurry) (Rossi & Taylor 2011). Therefore, phosphocellulose was chosen as a primary purification step before the IMAC step to purify LSH from DNA and proteins with each step respectively.

Purification of 6HisTEV-LSH proved problematic due to weak binding to IMAC resin and elution of the protein in the wash steps (data not shown). One potential reason for this could be cleavage of the 6Histag via proteases. To test this hypothesis a Western blot of fractions from the IMAC purification was performed (Figure 3.4A). Primary antibodies for anti-LSH and anti-6His were used. LSH was present in the flow through and early washes of the anti-LSH Western blot, whilst the 6Histag was detected only in the elution's in the anti-6Histag Western blot (Figure 3.4A). This suggests degradation of the 6His-tag by proteases or that an untagged and tagged LSH proteins were being expressed. More protease inhibitors were added (2 times more) and during the dialysis 0.2 mM PMSF was added to the buffer to prevent proteolysis of LSH, however this did not solve the problem (data not shown).

A new construct of LSH was cloned with a C-terminal non-cleavable 6Histag (LSH-6His). $LSH^{200-821}$ was also cloned with a C-terminal 6Histag ($LSH^{200-821}$-6his) to test if the expression could be improved. Expression of LSH-6His was comparable to 6HisTEV-LSH (Figure 3.3B). However expression of $LSH^{200-821}$-6His was reduced compared to LSH-6His (Figure 3.4B) as seen previously (Figure 3.2B), suggesting instability of this construct, therefore purification of this protein was not continued in this study.

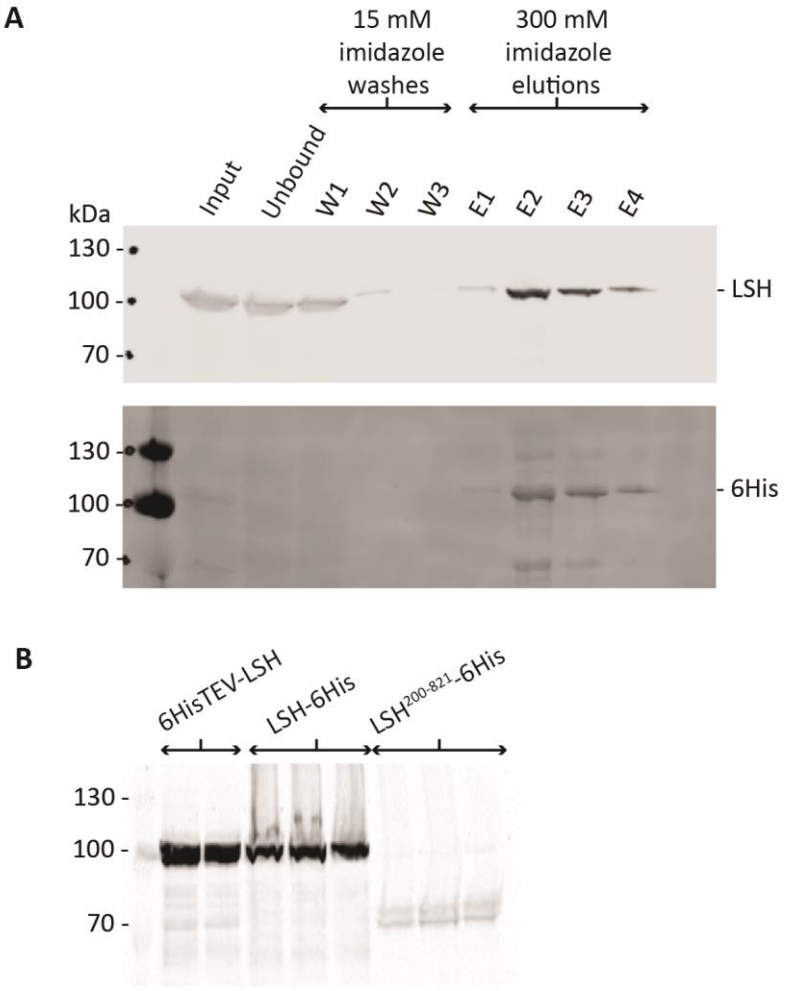

Figure 3.4| Western blotting of 6HisTEV-LSH
A| Western blots of IMAC fractions using 6HisTEV-LSH lysate, blotted with anti-LSH and anti-6His (0.1% of the total volume loaded per well for the Input and Unbound fractions). Washes performed with 15 mM imidazole (1% of the total volume loaded per well for the W1-W3 fractions). Elutions performed with 300 mM imidazole (5% of the total volume loaded per well for the E1-E4 fractions).
B| Anti-LSH Western blot 6HisTEV-LSH, LSH-6His and $LSH^{200-821}$-6His, 0.5 µg total protein

A comparison of the purification of 6HisTEV-LSH and LSH-6His was performed by purifying separately using phosphocellulose cation exchange chromatography followed by IMAC (Figure 3.5). LSH-6His eluted with greater yield and purity from the phosphocellulose cation exchange column than 6HisTEV-LSH (Figure 3.4C and 3.4A respectively). Furthermore, LSH-6His was purer after the IMAC step, in comparison to 6HisTEV-LSH (Figure 3.5D and 3.5B respectively). Importantly, more LSH-6His bound to the IMAC resin, as less was seen in the unbound fraction in comparison to 6HisTEV-LSH (Figure 3.5D and 3.5B respectively). As LSH-6His appeared less prone to degradation and was purer after two rounds of purification, it was selected as a better candidate to pursue for a large-scale expression and purification.

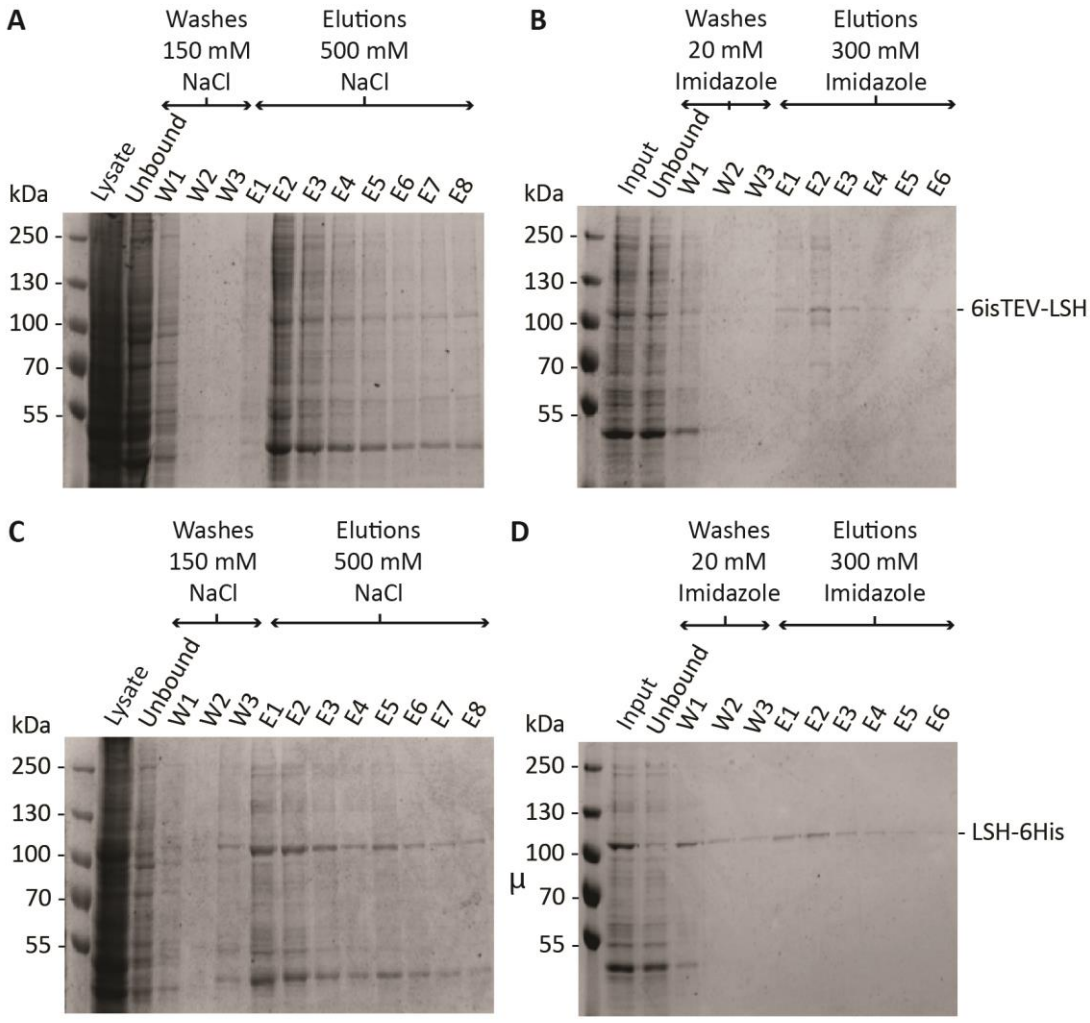

Figure 3.5|Coomassie stained SDS-PAGE gels of the purification of 6HisTEV-LSH or LSH-6His using phosphocellulose cation exchange followed by IMAC

A| Phosphocellulose purification of 6HisTEV-LSH (0.01% of the total volume loaded per well for the Input and Unbound fractions). 20 ml washes contained 150 mM NaCl (0.1% of the total volume loaded per well for the W1-W3 fractions). 1 ml elutions contained 500 mM NaCl (1.5% of the total volume loaded per well for the E1-E8 fractions).

B| IMAC purification of 6HisTEV-LSH (1% of the total volume loaded per well for the Input and Unbound fractions). 10 CV washes contained 20 mM Imidazole (0.15% of the total volume loaded per well for the W1-W3 fractions). 1 ml elutions contained 300 mM Imidazole (1.5% of the total volume loaded per well for the E1-E6 fractions).

C| Phosphocellulose purification of LSH-6His (0.01% of the total volume loaded per well for the Input and Unbound fractions). 20 ml washes contained 150 mM NaCl (0.1% of the total volume loaded per well for the W1-W3 fractions). 1 ml elutions contained 500 mM NaCl (1.5% of the total volume loaded per well for the E1-E8 fractions).

D| IMAC purification of LSH-6His. 10 ml washes contained 20 mM Imidazole (0.15% of the total volume loaded per well for the W1-W3 fractions). 1 ml elutions contained 300 mM Imidazole (1.5% of the total volume loaded per well for the E1-E6 fractions).

3.2.4| Large scale LSH expression and purification scale up

Expression of LSH-6His was scaled up to 4 litres of *Sf*9 culture at $1x10^6$ cells/ml and infected with 40 ml V2 virus. Purification was performed in three steps: phosphocellulose cation exchange chromatography, IMAC and SEC. The phosphocellulose step was minimally successful at purifying LSH from major contaminants (Figure 3.6A). The IMAC purification of the pooled phosphocellulose elutions was successful in removing major contaminants during washes with 15 mM imidazole, with the major protein being LSH-6His (Figure 3.6B). The pooled IMAC elutions were buffer exchanged for imidazole removal and underwent SEC using an ÄKTA HPLC with online UV detection at 260 nm and 280 nm. A broad peak with two shoulders eluting at 7-12 ml was determined to be unwanted contaminants through visualisation with silver staining SDS-PAGE (Figure 3.6D). A symmetrical peak eluting at 13 ml was identified as LSH-6His by mass using silver staining SDS-PAGE (Figure 3.6D). The UV ratio (260/280 nm) was ~0.6 indicative of negligible DNA contamination. Therefore, LSH had been purified to a quality sufficient for structural biology. MALDI-ToF was used to confirm if the 100 kDa protein was LSH (Figure S3.1). A coverage of 40% LSH was measured with MALDI-ToF and a Mascot score of 227 (Figure S3.1). A protein with a Mascot score of 70 is recognised as having a 95% probability that the match is not random (Figure S3.1). Considering that LSH had a score nearly 4 times that of 70 it is highly unlikely that the protein band analysed is anything other than LSH.

However, the total yield of purified LSH expressed from 4 litres of *Sf*9 cells was 200 µg, precluding structural biology experiments. Therefore. methods to increase the yield of LSH-6His reproducibly using the baculoviral/insect cell system were sought.

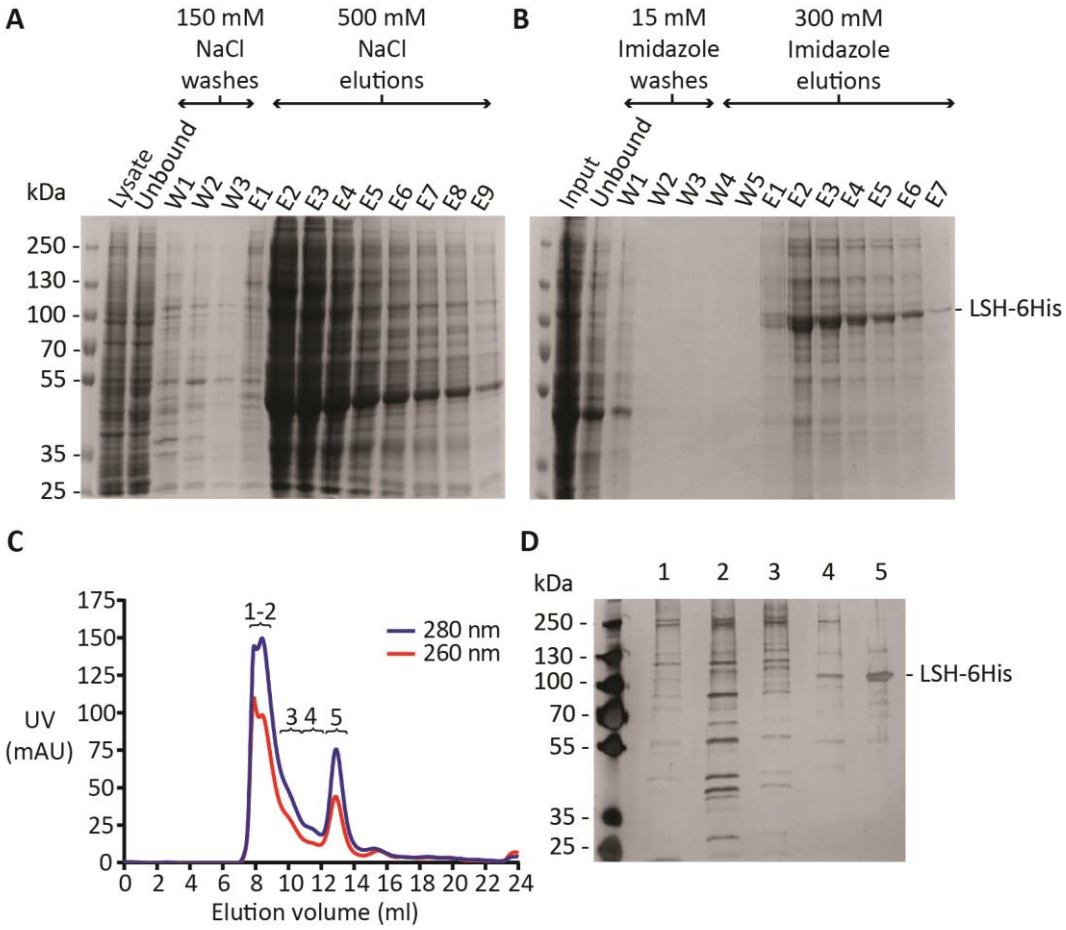

Figure 3.6| Purifcation of LSH-6His using phosphocellulose cation exchange, IMAC and SEC.
A| Coomassie stained SDS-PAGE gel of LSH-6His phosphocellulose purification (0.001% of the total volume loaded per well for the Lysate and Unbound fractions). 50 ml washes contained 150 mM NaCl (0.02% of the total volume loaded per well for the W1-W3 fractions). 1 ml elutions contained 500 mM NaCl (1% of the total volume loaded per well for the E1-E9 fractions).
B| Coomassie stained SDS-PAGE gel of LSH-6His IMAC purification. 10 CV washes contained 15 mM Imidazole (0.15% of the total volume loaded per well for the W1-W5 fractions). 1 ml elutions contained 300 mM Imidazole (1% of the total volume loaded per well for the E1-E9 fractions).
C| Chromatogram of LSH-6His SEC with UV detection (280 nm and 260 nm) (1% of the total volume loaded per well for the 1-5 fractions). .
D| Silver stained SDS-PAGE gel of SEC peak elutions 1-5, 2% total volume loaded per well.

3.2.5| Improving LSH expression in insect cells

Quick and cost-effective solutions to enhance protein expression in insect cells were investigated initially. First, the insect cell line used for protein expression was changed to HighFive™ which possesses superior expression for various proteins when compared to *Sf*9 insect cells (Wilde *et al.*, 2014). LSH expression in *Sf*9 and HighFive™ cells was quantified through Western blotting with anti-LSH and anti-Tubulin antibodies (Figure 3.7A) as described previously (Section 3.2). LSH expression was ~3 fold higher in HighFive™ cells than in Sf9 cells (Figure 3.7B).

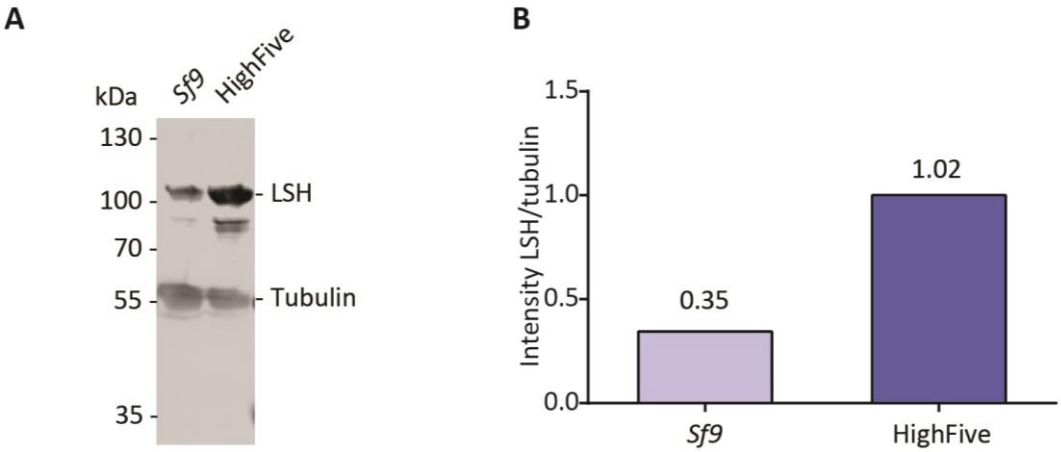

Figure 3.7|Comparison of LSH expression in HighFive and *Sf9* cells.
A| Anti-LSH and anti-Tubulin western blot of *Sf9* and HighFive lysates, four days post infection.
B|Graph showing the relative normalised intensities of LSH/Tubulin.

Therefore, HighFive cells were adopted as the cell line for LSH-6His expression in all subsequent experiments. The HighFive cell line also displays a fast doubling rate (18-24 hrs) compared to the *Sf*9 cell line (48-72 hrs), allowing faster cell culture scale up.

However, I observed aggressive cell clumping in HighFive suspension cell cultures which could not be remedied using the suppliers (Life Technologies) recommendation of heparin addition. Previously dextran sulphate (Mr 5000) was reported to be a rapid and effective method to prevent cell clumping without diminishing baculoviral infection (Dee, Shuler and Wood, 1997). This method minimised cell clumping during LSH expression and was adopted. A second disadvantage is HighFive cells have a 100 fold lower baculoviral generation than in *Sf*9 cells (Wilde *et al.*, 2014). Therefore, *Sf*9 cells were used for baculovirus propagation and HighFive cells used for protein expression.

The designers of the Multibac system suggest that the optimal cell density is 1-2x10^6 cells/ml due to higher cell densities exhausting nutrients (Reuveny *et al.*, 1993; Fitzgerald *et al.*, 2006). HighFive cells at 2-3x10^6 cells/ml in Express Five medium deplete glucose and glutamine 72 hrs post infection, limiting these major components in such media (Rhiel et al. 1997). Yeastolate hydrolysate has also been shown to be rapidly depleted in high-density insect cells (Drews et al. 1995).

Supplementing high-density insect cell cultures such as *Sf*9 or HighFive with glucose, glutamine, hydrolysates and lipids was proposed as a cost effective solution to improving overall volumetric yield (Schlaeger 1996). Supplementation and/or media exchange allows recombinant protein yield per cell in high-density cell cultures to be equivalent or higher than in low-density cell cultures were nutrients are not a limiting factor in *Sf*9 cells (Reuveny *et al.*, 1993; Bédard *et al.*, 1994; Ikonomou, Schneider and Agathos, 2003; Ohki *et al.*, 2012) and HighFive cells (Ikonomou *et al.*, 2004; Rausch, Pörtner and Knäblein, 2013).

Therefore I designed a 10x supplementation cocktail based on insect cell medium design and supplementation (Bédard *et al.*, 1994; Schlaeger, 1996; Ikonomou, Schneider and Agathos, 2003) The cocktail contained glucose (50 g/L), glutamine (100 mM), yeastolate ultrafiltrate (40 g/L) and 1000x lipid mixture (diluted to 10x) added to a final 1x concentration in insect cell culture. I decided to test HighFive cell cultures at 3-4x10^6 cells/ml as they were still in exponential growth phase. HighFive cultures with or without media exchange were infected with 1:25 V2 baculovirus:HighFive cell volume. The ratio was decreased 4x from 1:100 as determined in section 3.2.2 due to a 4 times increase in cell density. I tested the supplementation cocktail in culture with or without media exchange 1 day after V2 baculoviral infection.

The cell density for media-exchanged cultures was 5-5.5x10^6 cells/ml, nearly 2 times greater than those without media exchange at 3x10^6 cells/ml 2 days post infection (Figure 3.8A). The lack of growth arrest in media-exchanged cultures suggests incomplete baculoviral infection, however, the variability of cell density in media-exchanged and supplemented cultures suggest anomalies in cell counts (Figure 3.8A).

The cell viability was measured with trypan blue staining. A drop by 10-30% in cell viability 5 days post infection between conditions did not vary enough with any statistical significance based on overlapping standard deviation values (figure 3.8B). The cell viability typically dropped due to viral cell lysis (data not shown). Overall viability for all conditions 5 days post infection was high, with the lowest mean viability of 82% for the + supplement cultures (Figure 3.8B).

LSH expression was quantified through Western blotting with anti-LSH and anti-tubulin antibodies as performed previously (Figure 3.3). When the media was not changed the largest amount of LSH-6His was produced 4 and 5 days post-infection, with + supplementation producing the highest yield at day 4 which was statistically significant compared without supplementation (Figure 3.8C and 3.8E). However, at day 5 the LSH-6His yield of cultures without supplementation had increased within the error of + supplementation at days 4 and 5 (Figure 3.8C and 3.8E). For media exchanged cultures, LSH-6His expression was also greatest at days 4 and 5 post-infection (Figure 3.8D and 3.8F), but the difference in LSH expression with or without supplementation was not statistically significant due to the large error between repeats (Figure 3.8F).

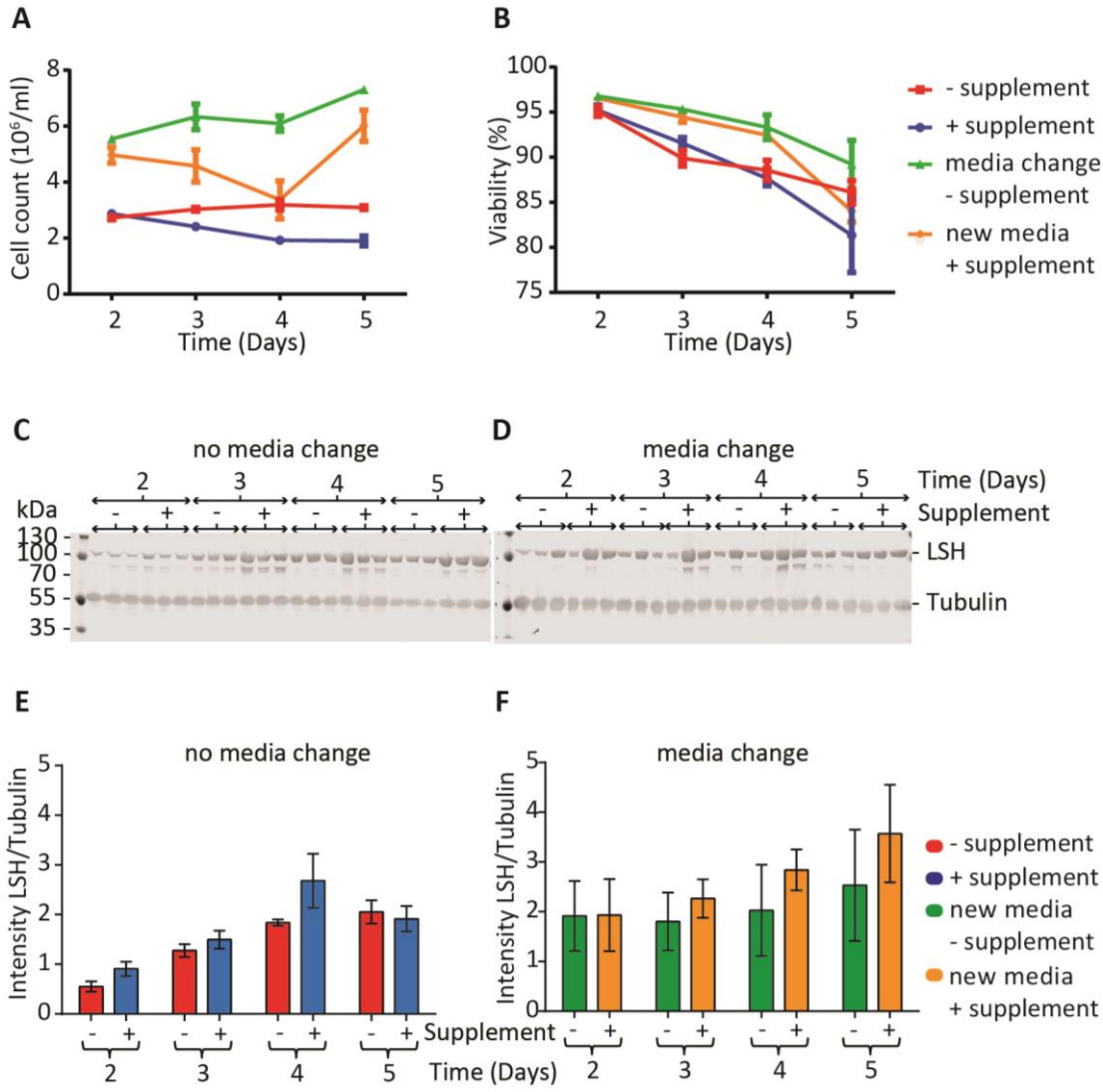

Figure 3.8|Testing and optimising LSH-6His expression in HighFive cells through media change before V2 baculovirus infection and/or supplementation 2 days post infection.
A| Cell counts from day 2-5 after infection of the four variables tested (no supplement, supplement at day 2, media change with no supplement at day 2, media change with supplement at day 2).
B| Cell viability from day 2-5 after infection of the four variables tested (same as A).
C| Anti-LSH and anti-Tubulin western blot for no supplement and supplement at day 2.
D| Anti-LSH and anti-Tubulin western blot for media change with no supplement at day 2 and media change at day 2 with supplement.
E| Graph showing the mean and standard deviation of normalised intensity for LSH/Tubulin signal from blot C.
F| Graph showing the mean and standard deviation of normalised intensity for LSH/Tubulin signal from blot D.

The lack of a noticeably increased yield from supplementation in media exchanged cultures could have been due to media not being a limiting factor for cell growth and protein expression. However, the experiment did provide evidence that expression was stable 5 days post infection with overall cell viability >80% (Figure 3.8B, 3.8E and 3.8F).

Accumulation of lactate and glutamate, the by-products of glucose and glutamine hydrolysis respectively are known to be detrimental to cell viability (Gorfien et al. 2003; Drugmand et al. 2005). However, my data suggests supplementation was not toxic to the HighFive™ cells.

My results also provide evidence that media exchange can allow the HighFive cell density to reach expected stationary phase levels of $6x10^6$ cells/ml and still give expression levels equivalent to lower cell densities, thereby increasing volumetric yield (Figure 3.8A, E and F). However, whether this is economically worthwhile is difficult to determine as protein expression in non-media exchanged cells would have to be compared at equal cell densities. Given that the cost of supplementation is ~10% that of media exchange, it is worth using even if the enhancement is negligible. At least it is a safeguard to prevent nutrient deprivation at cell densities greater than $3x10^6$cells/ml.

A further adaption to the expression protocol was the use of glass baffled flasks with membrane screw caps (Duran) instead of the 2L roller bottles used upright (Corning). This was more economical and the conical shape and baffles were expected to improve oxygenation. However, protein expression was not compared between these flasks. Respiration increases in insect cells upon baculoviral infection, with a decrease in oxygen rate indicative of exhausted glucose and glutamine in the media (Wong *et al.*, 1994; Kamen *et al.*, 1996). Therefore inadequate oxygen supply hinders recombinant protein production in *Sf*9 cells (Wang, Kwong and Bentley, 1993; Wong *et al.*, 1994; Kamen *et al.*, 1996; Palomares, López and Ramírez, 2004) and HighFive cells (Rhiel, Mitchell-Logean and Murhammer, 1997).

A final influencing factor was the reproducibility of baculovirus infection. Reduced infection of insect cells was observed with V2 baculovirus more than 2 weeks old, possibly due to baculovirus aggregation (Jorio, Tran and Kamen, 2006). Cryo-freezing V1 baculovirus infected *Sf*9 cells 24 hrs post infection at cell densities of $1x10^7$cells/ml has been shown to maintain the baculovirus inside infected *Sf*9 cells (Wasilko *et al.*, 2009) and is termed baculovirus infected insect cells (BIICs). Thawing BIIC's into *Sf*9 cells ($1x10^6$cells/ml) produces a V2 baculovirus stock by viral spread to fresh uninfected cells (Wasilko *et al.*, 2009). This allows for V2 virus to be made as and when needed, with reproducible titer (Wasilko *et al.*, 2009). However, the volume of BIICs required for complete infection of *Sf*9 cells needed to be assayed. Two common methods for determining baculoviral titer are the plaque assay (Brown and Faulkner, 1977; King, 2012) and quantitative real-time PCR (Hitchman *et al.*, 2007). However, the plaque assay is labour intensive and time consuming (> 1 week) and tests infectivity of cells on an agarose plate rather than in media, making an accurate comparison difficult. The qPCR method determines the titer through total viral RNA, and not the infectivity of the baculovirus, which needs to be determined via the plaque assay alongside.

Therefore I designed a 96-well plate baculovirus titer assay which took the advantage of YFP expression which can be monitored by fluorescence microscopy based on a protocol described previously (Hopkins and Esposito, 2009). As LSH and YFP are expressed under separate polH promoters, expression of both is presumed to be equivalent (Fitzgerald et al. 2006). YFP expression could be monitored by fluorescence microscopy as done previously (Figure 3.2A and Figure 3.3A).

Serial dilutions of BIICs to a constant *Sf*9 cell density were made in a black transparent bottomed 96 well plate. 3 days post infection wells were imaged at x400 magnification using a fluorescent microscope with a camera attached (figure 3.9A). By eye the 1:10 ratio BIIC:*Sf*9 samples contained the most fluorescence 3 days post infection. The 1:5 ratio displayed weaker fluorescence than 1:10 or 1:20 ratios, likely due to cell death from over infection. For quantitative comparison, images were analysed using a macro from Image Pro plus software written by Dr David Kelly to determine the area of fluorescence in each image which was represented in graph form (Figure 3.9B). The 1:10 condition had the greatest fluorescence at day 3 post infection, with 1:20 having approximately half the fluorescence of 1:10 (figure 3.9B). To speed up quantitative analysis time I used a multiplate reader with emission and excitation in the range of YFP – 495 nm and 515 nm respectively (Figure 3.9C). These data show intensities very similar to the total area of fluorescence measured by the microscope (Figure 3.9B), making the plate reader method robust enough for repeated use. Taken together the data in Figure 3.9 shows that a 1 ml BIIC ($1x10^7$ cells/ml) was necessary to completely infect 100ml of *Sf*9 cells ($1x10^6$ cells/ml) 3 days post-addition to create V2 baculovirus.

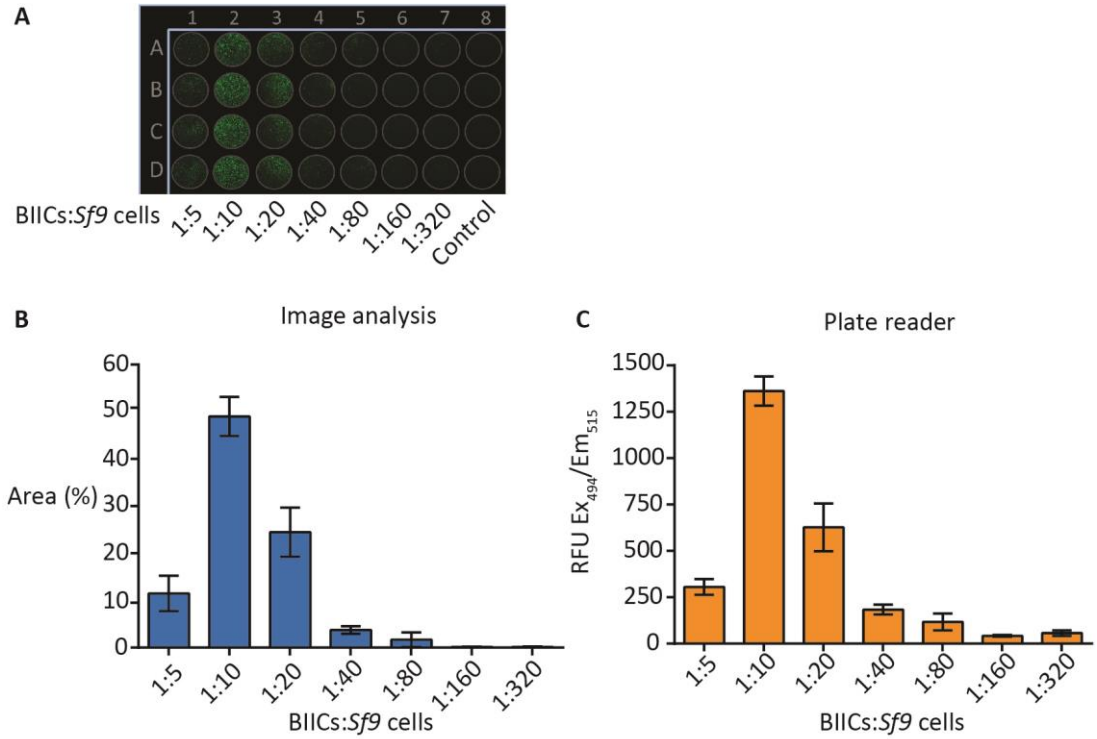

Figure 3.9|Testing the infectivity of BIICs on *Sf9* cells using a 96 well titration assay.
A| Fluorescent microscopy images of *Sf9* cells at x400 magnification. Images representative of infection after 3 days for each ratio of BIICs:*Sf9* cells.
B| Total area of fluorescent cells per image measured for each BIICs:*Sf9* cells in a 96 well plate after 3 days infection. Averages from a single image from 4 replicate wells were taken for mean and standard deviation.
C| Relative fluorescence units (RFU Ex_{495}/Em_{515})measured for each ratio of BIICs:*Sf9* cells in a 96-well plate after 3 days infection using an automated plate reader. Averages from a single reading from 4 replicate wells were taken for mean and standard deviation.

To summarise the optimised method involved making fresh V2 baculovirus with BIICs which was then used within one week. HighFive™ cells were freshly split to a density of $4x10^6$cells/ml and infected 1:10 (V2 baculovirus:HighFive™ cells) with V2 virus. Two days after infection, HighFive™ cells were supplemented with the glucose, glutamine, hydrolysate and lipid cocktail. Pellets were collected between 3 and 4 days post-infection.

3.2.6| Optimising LSH purification

LSH-6his was expressed in 1L HighFive™ cells. The purification strategy was changed to recover more protein. HighFive cells were lysed in a stronger ionic buffer (500 mM NaCl instead of 100 mM NaCl) to allow lysis of both the cytoplasmic and nuclear cell fractions using a cell disruptor. The order of chromatography steps was reversed. Thus, IMAC was used as the primary purification method, with an overnight incubation of lysate with Ni-NTA resin, to enhance binding. IMAC wash steps included a 1M NaCl wash to remove DNA by breaking electrostatic protein:DNA contacts. The IMAC elutions contained 100 mM NaCl for direct loading onto phosphocellulose cation exchange resin. Buffer exchange to remove imidazole was performed on the phosphocellulose column during the 100 mM NaCl wash steps. Phosphocellulose elutions were pooled, concentrated and a final size exclusion chromatography step used to remove contaminants and aggregates.

8 mg of ~75% pure LSH-6His eluted from the primary IMAC step (Figure 3.10A). This was likely due to increased LSH-6His expression, with LSH-6His being clearly visible in the lysate fraction (Figure 3.10A). The phosphocellulose step produced an estimated 90% pure LSH-6His, particularly in the later elution fractions (E4 onwards) (Figure 3.10B). However, some LSH-6His was detected in the unbound fraction suggestive of incomplete binding or because it was bound to DNA tightly (Figure 3.10B). The unbound LSH-6His sample was used in a new phosphocellulose purification with more LSH-6His binding and eluting (data not shown). However, not all LSH-6His bound suggestive of protein denaturation and/or protein aggregation. The chromatogram of the final SEC step separated LSH-6His into a void peak (labelled 1), an oligomerisation peak (labelled 2 - 11.4 ml) and a symmetrical peak (labelled 3 - 13 ml) (Figure 3.10C and 3.10D). Most LSH-6His eluted between 12-14 ml as seen previously (Figure 3.6C and 3.10D). LSH-6His was estimated to be greater than 95% pure with some degradation products seen in overloaded lanes (Figure 3.10D).

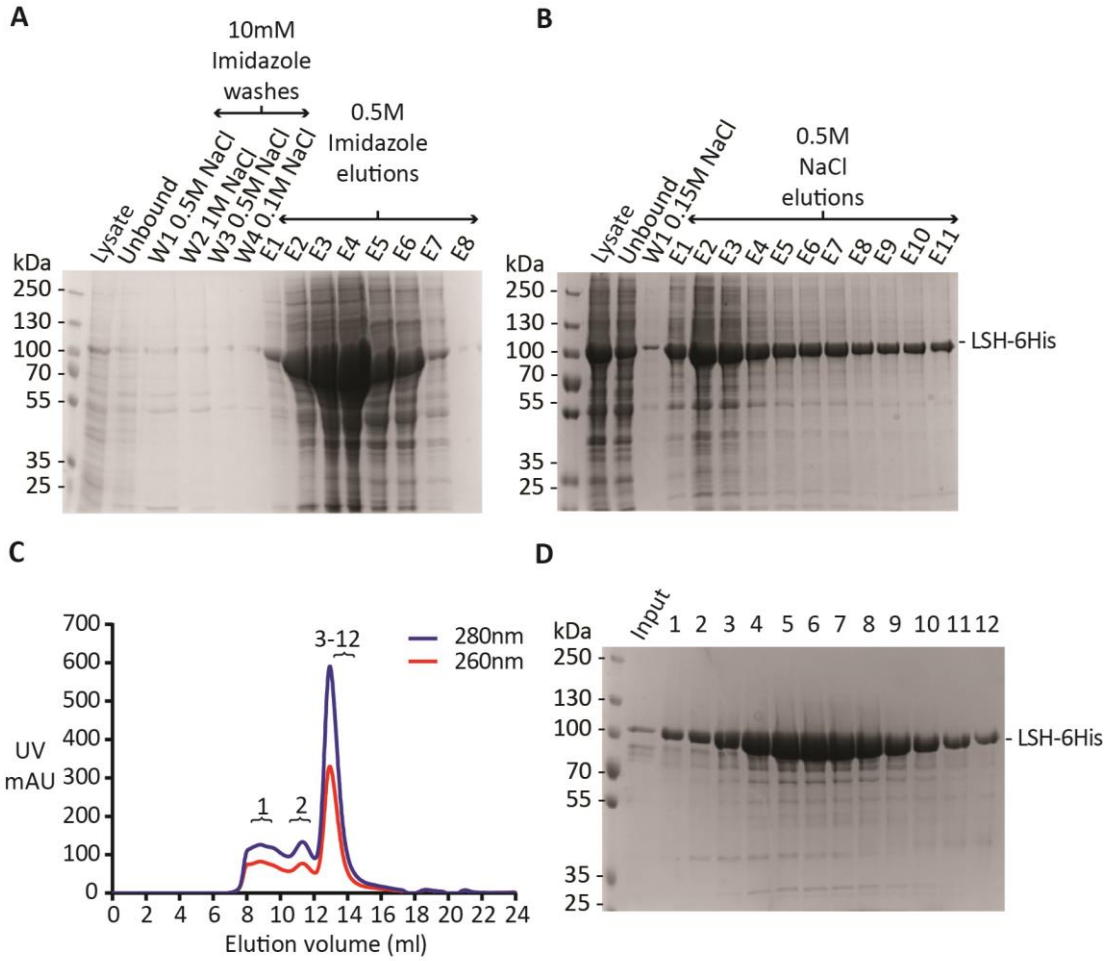

Figure 3.10|LSH-6His purification using IMAC, phosphocellulose cation exchange and SEC
A| Coomassie stained SDS-PAGE gel of LSH-6His IMAC purification (0.001% of the total volume loaded per well for the Lysate and Unbound fractions). 50 ml washes contained 10 mM Imidazole and varying NaCl concentrations (0.02% of the total volume loaded per well for the W1-W4 fractions). 1 ml elutons contained 0.5M Imidazole and 0.1 M NaCl (0.5% of the total volume loaded per well for the E1-E8 fractions).
B| Coomassie stained SDS-PAGE gel of LSH-6His phosphocellulose purification. 30 ml wash contained 0.15 M NaCl (0.0375% of the total volume loaded per well for the W1 fraction). 1 ml elutions contained 0.5 M NaCl (0.5% of the total volume loaded per well for the E1-E8 fractions).
C| Chromatogram of LSH-6His SEC with UV detection (280 nm and 260 nm).
D| Coomassie stained SDS-PAGE gel of SEC peak elutions 1-12, 0.5% total volume loaded per well.

The overall yield of purified LSH from 1L of HighFive™ culture was ~5 mg. This was a 25 fold improvement in yield, solely aided by expression and purification optimisation. Given the large yield of LSH, the alternative methods for improving expression of codon optimisation and leader sequence addition were not pursued further. Expression and purification was reproducible on more than 3 occasions throughout this study.

The expression and purification method was also used to express the LSH ATPase inactive mutant LSH^{K237Q}, which had been cloned with a C-terminal 6Histag. A symmetrical peak of LSH^{K237Q}-6His eluting at a peak elution volume of 13 ml which was equivalent to LSH-6His was observed during SEC purification (3.11).

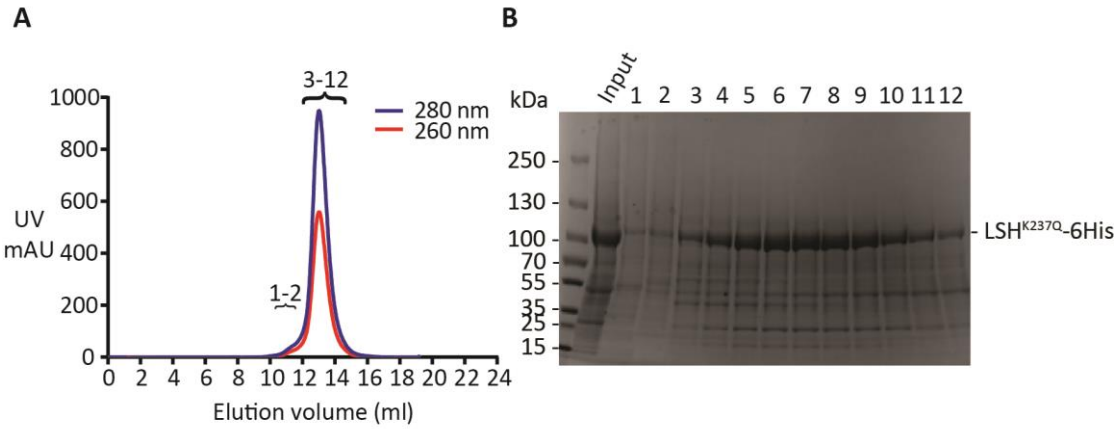

Figure 3.11| LSH^{K237Q}-6His purification with SEC.
A| Chromatogram of LSH^{K237Q}-6His SEC with UV detection (280 nm and 260 nm).
B| Coomassie stained SDS-PAGE gel of SEC fractions 1-12, 0.5% total volume loaded per well.

LSH-6His could now be reproducibly expressed with 5 mg of greater than 95% purity from 1L of HighFive cell culture. However, if protein expression had not increased through the use of HIghFive cells, high-density culture supplementation and reproducible infection using BIICs, the other optimisation methods of codon optimisation and enhancer sequences were contemplated as secondary options.

The DNA code is degenerate and dependent on the organism, certain synonymous codons are preferred. Using synonymous codons suited to the expression organism, without changing the protein sequence has been shown to increase the yield for heterologous protein expression (Zolotukhin 1996). One study observed up to a 7-fold increase in protein expression in *Sf*9 cells with codon optimisation but for some constructs there was minimal change (Fath et al., 2011). Insertion of a 21 bp lobster tropomyosin cDNA leader sequence upstream of the polH promoter can increase luciferase expression 7-fold (Sano et al., 2002). Such a leader sequence could have been cloned into a pFL plasmid with a codon optimised LSH ORF with the hope of enhanced expression from both methods.

However, given both methods would require LSH to be re-cloned; this would only be used if the already tested optimisation methods failed to provide an adequate improvement in LSH yield.

3.3| Conclusion

I used the MultiBac system to express full-length LSH and $LSH^{200-821}$ separately. LSH and $LSH^{200-821}$ were initially cloned with a hexahistidine (6His) tag linked to the N-terminus via a Tobacco etch virus (TEV) cleavage site. Successful expression of both LSH and $LSH^{200-821}$ was observed, however, the $LSH^{200-821}$ variant had reduced expression compared to LSH indicating instability of this protein. Furthermore, degradation of the N-terminal sequence was observed, so LSH was cloned with a C-terminal 6His tag. C-terminal tagged LSH had reduced degradation and better binding to IMAC resin instead of the N-terminal tagged LSH. Low yields of pure LSH were obtained therefore improvements in expression were sought. Using a HighFive™ insect cell line, enhanced LSH expression was achieved through infection of cells at high density and supplementing with nutrients. A 96-well plate viral titer assay was designed to determine accurately the amount of virus needed for reproducible infection using BIICs. Moreover, a purification strategy was designed to optimise LSH-6His and the ATPase inactive mutant LSH^{K237Q}-6His recovery and purity to achieve 5 mg from 1L of HighFive™ cell culture with greater than 95% purity enabling the biochemical, biophysical and structural investigations of LSH.

Chapter 4| Biophysical characterisation of LSH for structural biology, crystallisation trials and defining stable domains

4.1| Introduction

There are relatively few results describing the biochemical activities of LSH *in vitro* and no structural data. The major aim of this project was to biochemically characterise and solve the structure of LSH to study how its flanking domains regulate enzymatic function. The reproducible expression and high-yield purification of LSH-6His made this feasible. Biochemical and biophysical protein characterisation aids in determining if a protein is a good crystallisation target and can provide information to improve the chances of success. To this end I used a thermal denaturation assay (TDA), dynamic light scattering (DLS) and SEC coupled to multi-angle lights scattering (SEC-MALS) to determine the homogeneity, oligomeric state and buffer conditions for the optimal stability of LSH-6His.

4.2| Results

4.2.1| Thermal stability of LSH-6His

Protein stability over extended periods and after freeze/thawing is required for structural and biochemical investigations. The addition of functional ligands may improve protein stability and optimal buffers can reduce aggregation and denaturation of the protein. To determine the optimal buffer conditions for LSH-6His I used a thermal denaturation assay (TDA) to measure the thermal stability of LSH-6His. The TDA uses the environmentally sensitive dye SYPRO orange, which becomes highly fluorescent upon binding to hydrophobic patches on unfolding proteins (Niesen, Berglund and Vedadi, 2007). As a protein unfolds due to increasing temperature, more hydrophobic amino acids become exposed, thereby increasing dye binding and fluorescence at 575 nm (Niesen, Berglund and Vedadi, 2007). A quantitative PCR machine is used to both increase temperature and measure fluorescence (Niesen, Berglund and Vedadi, 2007). The melting temperature (T_m) of the protein is calculated as the minimum of the negative derivative of fluorescence change using the Bio-Rad QPCR software.

First, I used the TDA to measure the T_m of LSH-6His at three different concentrations: 0.5, 1 and 2 µM (Figure 4.1). I wanted to establish the lowest concentration necessary for good signal-to-noise where greater signal provides greater accuracy of the T_m (Figure 4.1). A single transition curve was measured for each concentration; however, the initial fluorescence (at 20 °C) for all concentrations of LSH-6His was greater than the buffer only sample, suggesting partial unfolding or aggregation of LSH-6His or that LSH has hydrophobic patches (Figure 4.1A). Higher LSH-6His concentrations coincided with increased thermal stability of 42.8°C for 2 µM LSH-6His in comparison to 38.7°C for 0.5 µM LSH-6His (Figure 4.1B and Figure 4.1C). An increase of 4.1°C in T_m from 0.5 µM to 2 µM of LSH-6His was measured. However, at 0.5µM LSH-6His there is a poor signal-to-noise ratio whereas at 1 and 2 µM signal-to-noise ratio is higher. Therefore, using LSH-6His at higher concentrations is more accurate and further TDA experiments were performed using 1 µM LSH.

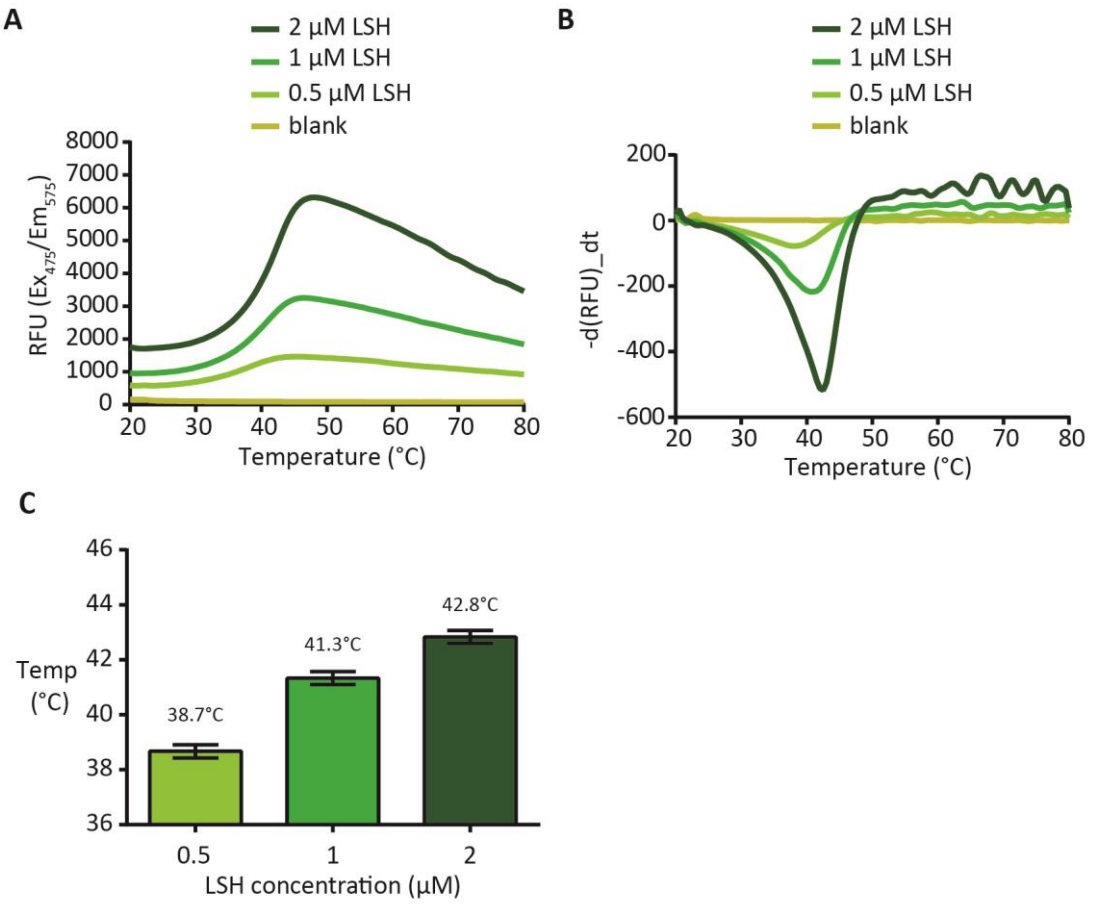

Figure 4.1| TDA of LSH-6His at three concentrations
A| Fluorescence intensity of 0.5, 1 and 2 µM LSH-6His thermal denaturation.
B| Derivitised fluorescence intensity of 0.5, 1 and 2 µM LSH-6His thermal denaturation from **A|**.
C| Bar graph of the Tm of LSH-6His at 0.5, 1 and 2 µM with standard error from 3 replicates.

I decided to test a range of buffers in the pH range 3-10 to determine the pH at which LSH-6His was optimally thermally stable. Thermal denaturation of LSH-6His in buffers ranging from pH 3-5 produced a maximum fluorescence at 20°C indicating aggregated LSH-6His. The T_m of LSH-6His in each buffer ranging from pH 5-10 was subtracted from the T_m of LSH-6His in 50 mM Tris pH 7 buffer (Figure 4.2A). The T_m of LSH-6His increased by +0.5 – 3.5°C in buffers in the pH range of 7-7.5 and 8.5-9. The T_m of LSH-6His was greater by +1.5°C in sodium citrate pH 5.5, however the curve had high initial fluorescence, suggesting a partially unfolded protein. LSH-6His in buffers in the pH range 6-7 and 9-10 had a lower T_m of -0.5 - 8°C when compared to the Tris pH 7 buffer, however the melt curves were similar in profile suggesting the protein was still properly folded but less stable.

To find the optimal buffer, the buffers in the pH range 7-7.5 and 8.5-9, in which LSH-6His was most stable, were re-tested in triplicate (figure 4.2B). Of the buffers tested sodium malonate pH 7 (43.3°C) and HEPES pH 7.25 (42.7°C) produced the highest T_m for LSH-6His, however the large experimental error made it difficult to discern which buffer gave LSH-6His the highest T_m. The HEPES pH 7.25 buffer was chosen because it is in the middle of the pH buffering range of HEPES (6.8-8.2) whilst sodium malonate pH 7 buffer is at the upper limit of the pH buffering range of sodium malonate (4-7). This makes the sodium malonate pH 7 buffer less suitable for buffering additional components which are greater than pH 7.

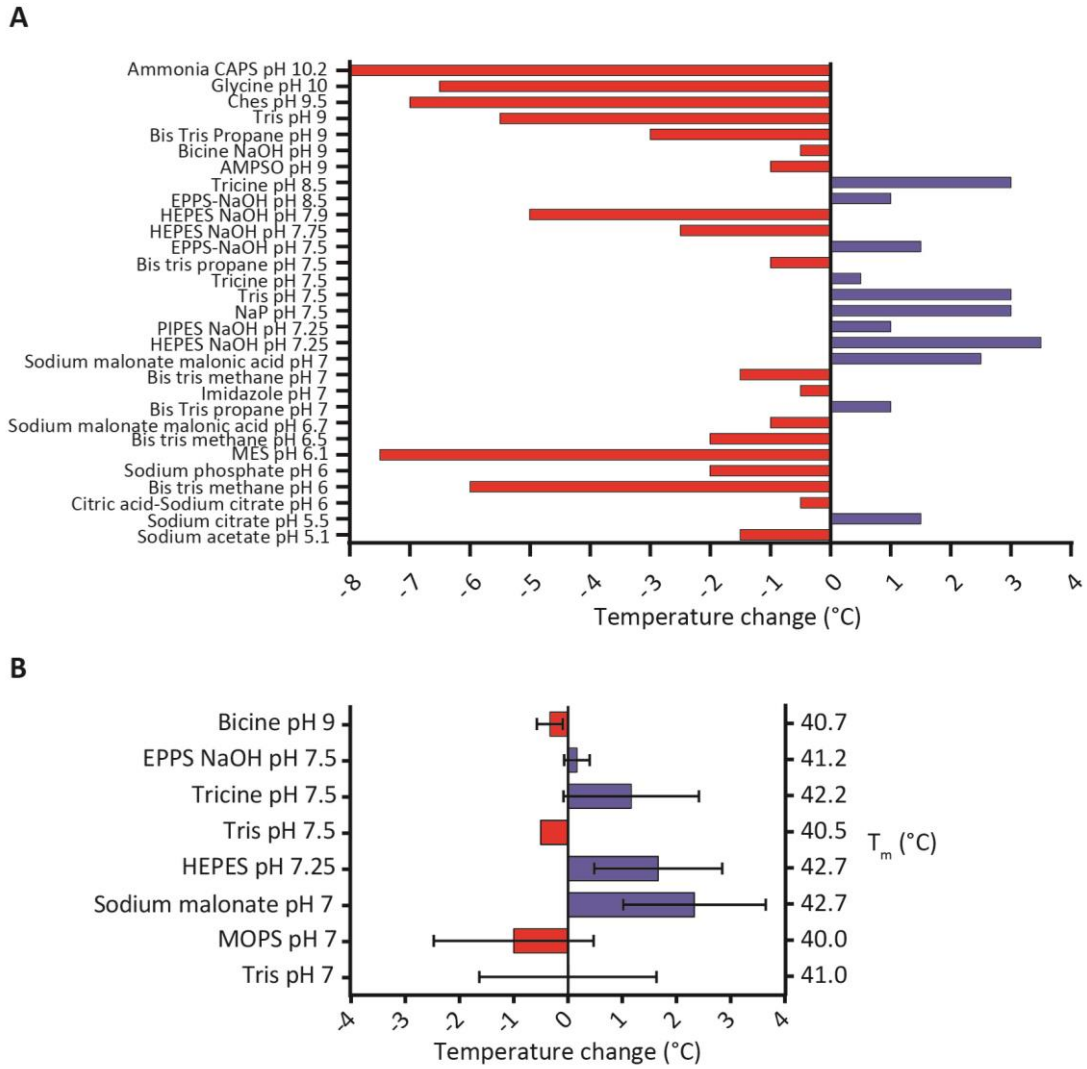

Figure 4.2| LSH-6His T_m change in buffer ranging from pH 5-10
A| Bar graph of LSH-6His T_m in buffers ranging from pH 5.1 to 10.2 subtracted from LSH-6His T_m in 50 mM Tris pH 7 buffer. A buffer condition is which is LSH-6His is more stable has a positive temperature change (blue) and less stable has a negative temperature change (red) **B|** Bar graph of LSH-6His Tm retested in 7 of the best thermally stabilising buffers from **A|** subtracted from the average of LSH-6His in 50 mM Tris pH 7 buffer. A buffer condition is which is LSH-6His is more stable has a positive temperature change (blue) and less stable has a negative temperature change (red). Each buffer tested was performed in triplicate and standard error is represented for each average.

The thermal stability of LSH-6His in 50 mM HEPES pH 7.25 was measured with differing concentrations of the additional buffer components, NaCl, glycerol or $MgCl_2$ (Figure 4.3). The T_m of LSH-6His in NaCl concentrations as low as 10 mM or as high as 500 mM was 38-38.5°C, therefore LSH-6His is equally stable in a range of 10-500 mM NaCl concentrations (Figure 4.3A and Figure 4.3B).

Glycerol can stabilise proteins by forming a amphiphilic layer between the hydrophobic patches on the protein surface and the polar solvent (Vagenende, Yap and Trout, 2009). The addition of 10% glycerol in buffers is used during LSH-6His purification to reduce aggregation and during cryo-freezing to reduce water crystal formation, which can damage the protein. Increasing the concentration of glycerol in the buffer increased LSH-6His thermal stability, with 10% glycerol raising the T_m of LSH-6His by 3.2°C to 42.2°C, and 20% glycerol by 5.7°C to 44.7°C when compared to no glycerol (Figure 4.3C and Figure 4.3D). However, the presence of 20% glycerol is too viscous for many downstream applications including crystallisation, therefore 10% glycerol was the best compromise.

LSH has ATP and Mg^{2+} binding site at helicase motifs I and II as conserved for all chromatin remodelers and helicases (Richmond and Peterson, 1996b; Dürr *et al.*, 2006). I tested if the cofactor Mg^{2+} might have an effect on LSH stability. There was no significant change in the T_m of LSH-6His up to a concentration of 20 mM $MgCl_2$ (Figure 4.3E and Figure 4.3F).

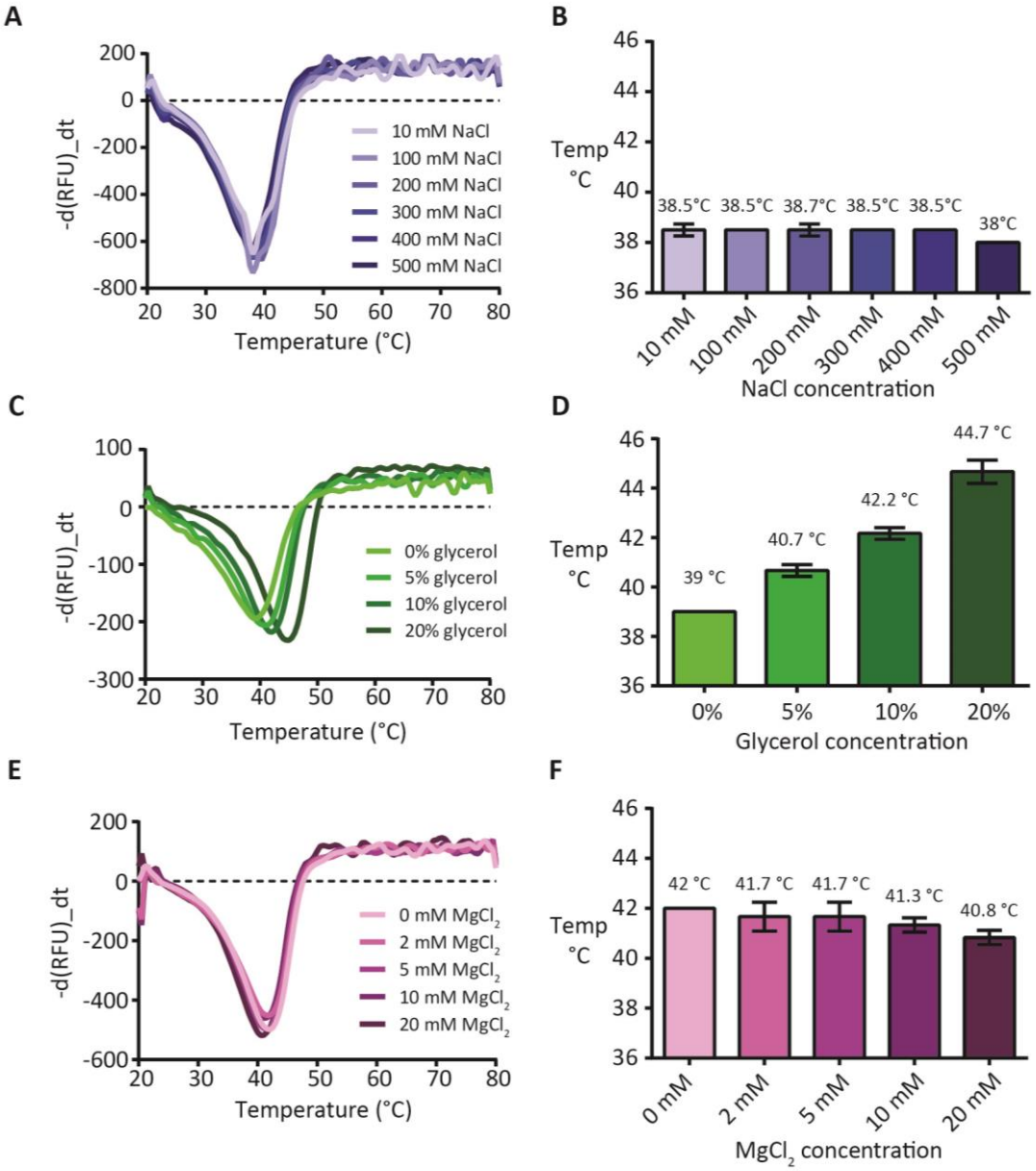

Figure 4.3| Thermal denaturation of LSH-6His in different NaCl, Glycerol and $MgCl_2$ concentrations

A| Derivitised fluorescence intensity of LSH-6His thermal denaturation in NaCl concentrations ranging from 10-500 mM in 50 mM HEPES and 1mM DTT.

B| Bar graph of LSH T_m from **A|**. Averages and standard error were calculated from 3 replicates.

C| Derivitised fluorescence intensity of LSH-6His thermal denaturation in Glycerol concentrations ranging from 0-20% in 50 mM HEPES, 200 mM NaCl and 1mM DTT.

D| Bar graph of LSH T_m from **C|**. Averages and standard error were calculated from 3 replicates.

E| Derivitised fluorescence intensity of LSH-6His thermal denaturation in $MgCl_2$ concentrations ranging from 0-20 mM in 50 mM HEPES, 200 mM NaCl, 10% Glycerol and 1mM DTT.

F| Bar graph of LSH T_m from **E|**. Averages and standard error were calculated from 3 replicates.

As stated above LSH has ATP binding motifs, and has been shown to have ATPase activity (Burrage *et al.*, 2012). This confirms LSH binds ATP which could be used as a specific stabilising factor for LSH. Therefore I measured the thermal stability of LSH-6His with ATP, the non-hydrolysable ATP analogue AMP-PNP, and the products of ATP hydrolysis ADP and AMP (Figure 4.4A and Figure 4.4B). ATP and AMP-PNP increased the T_m of LSH-6His by 2.6°C and 2.3°C respectively (Figure 4.4A and Figure 4.4B), whereas ADP and AMP increased the T_m of LSH by 5.6°C and 2°C respectively (Figure 4.4A and Figure 4.4B). When ADP - a product of ATP hydrolysis is present, the T_m of LSH increases by 3°C more than when ATP – the substrate is present. The greater increase in stability from ADP indicates the product of ATP hydrolysis may compact the nucleotide binding region more than ATP.

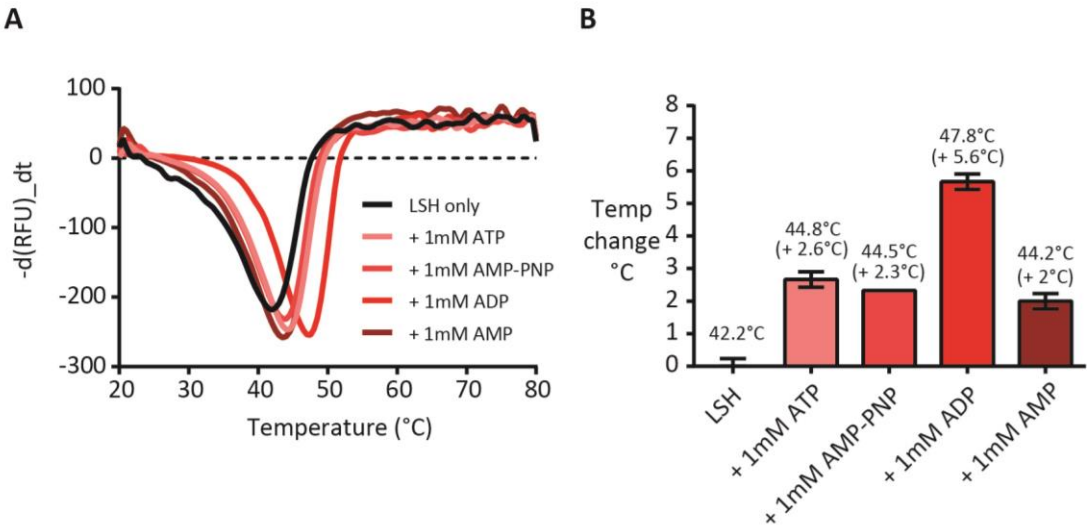

Figure 4.4| Thermal denaturation of LSH-6His in the presence of ATP, AMP-PNP, ADP or AMP

A| Derivitised fluorescence intensity of LSH-6His thermal denaturation only, or with ATP, AMP-PNP, ADP or AMP.

B|Bar graph of LSH-6His Tm change with ATP, AMP-PNP, ADP or AMP from **A|** subtracted from the LSH-6His only T_m.

TDA experiments have shown LSH is folded with a thermostability increase upon ATP binding indicating LSH binds ATP indicating it is likely a functional protein. Based on the TDA results, LSH-6His is most stable in a buffer comprising 50 mM HEPES pH 7.25, 100-500 mM NaCl, 10% Glycerol, 1 mM DTT and 1 mM ADP.

4.2.2| LSH-6His is homogeneous in low and high ionic strength buffer

In any biochemical or structural studies, protein polydispersity and aggregation is a common cause of reduced activity, uninterpretable data or failed crystallisation trials. Therefore, DLS was chosen to assess protein homogeneity and aggregation.

Dynamic light scattering applies monochromatic light to a sample of particles. The particles scatter the light in all directions (Rayleigh scattering) which constantly changes due to constructive and destructive interference as a result of time and Brownian motion. The environmental factors of temperature and solution viscosity influence the level of particle scattering. Particle size also influences scattering with severe scattering detected with larger particles. If the particle is homogenous, the hydrodynamic radius can be used to calculate the mass of the particle using the model of an ideal sphere. DLS is a fast, non-destructive method which requires a small sample volume at low concentrations (60µl at 0.5 mg/ml).

I used DLS to test thawed LSH-6His at 0.6 mg/ml (6.25 µM) for aggregation after the SEC purification and cryo-freezing. LSH-6His is a DNA binding protein and a 500 mM NaCl condition would break protein:DNA contacts which is unsuitable for studying LSH:DNA interactions, therefore a buffer containing 100 mM NaCl which should not disrupt DNA binding of LSH-6His and 500 mM NaCl were tested (Figure 4.5). A peak with ~ 1% intensity and a 0.5 nm diameter was detected for LSH-6His in both 100 mM NaCl and 500 mM NaCl buffers (Figure 4.5A and Figure 4.5B), probably due to the scattering of glycerol, which has a diameter of 5-6 Å. This is the most abundant small molecule in the solution at 1.37 M (Figure 4.5A and Figure 4.5B). LSH-6His was homogenous with a symmetrical peak with a diameter of 10.56 nm in 100 mM NaCl buffer and 10.77 nm in 500 mM NaCl buffer (Figure 4.5A and Figure 4.5B). Particles averaging a diameter of 300 nm were also present in the 500 mM NaCl LSH-6His sample (Figure 4.5B). This was likely a small percentage of aggregates as it is only 4% the intensity of the LSH-6His, making it insignificant.

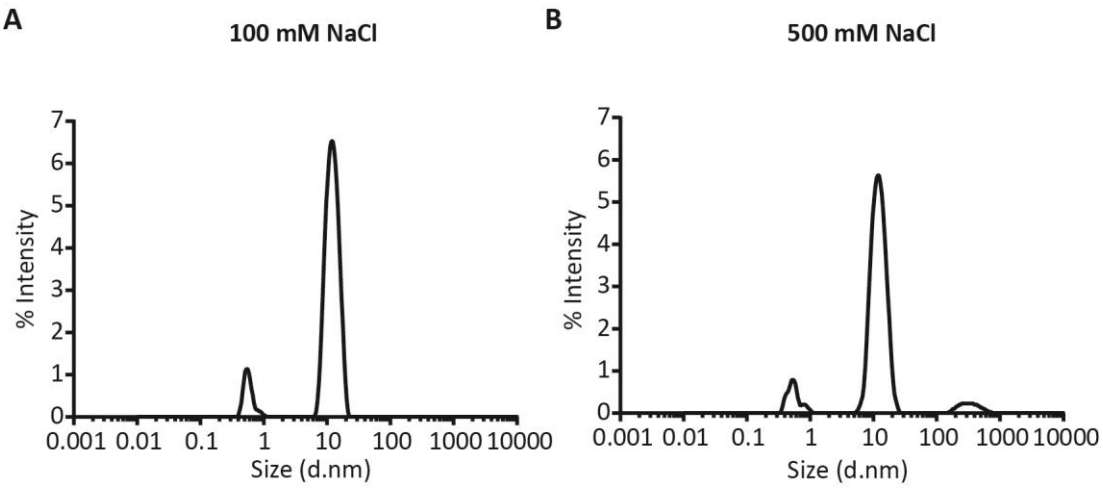

Figure 4.5| Size distribution by intensity of LSH-6His measured by DLS
A| The size distrubution by intensity of LSH-6His in 100 mM NaCl buffer.
B| The size distribution by intensity of LSH-6His in 500 mM NaCl buffer.

However, the LSH-6His has an estimated Mw of 157 kDa (100 mM NaCl buffer) or 177 kDa (500 mM NaCl) calculated from the hydrodynamic radius of the second peaks which is in-between the mass of a LSH monomer (96 kDa) or a LSH dimer (192 kDa). Two potential hypotheses exist for the oligomeric nature of LSH-6His drawn from the DLS data. The first hypothesis is the protein is a mixture of dimer and monomer. This is unlikely as the SEC of LSH-6His is a symmetrical peak at 13 ml suggests one Mw species. The second hypothesis is LSH-6His is a monomer due to the estimated masses being less than that predicted for a dimer. However, as the molecular mass is estimated based on the protein forming a sphere, the mass can be overestimated if the actual shape is a prolate spheroid or an oblate spheroid.

Therefore, size exclusion chromatography coupled with multi angle light scattering (SEC-MALS) was used to accurately measure the oligomeric nature of LSH-6His.

4.2.3| LSH-6His is a monomer in low and high ionic strength buffer

SEC separates molecules based on their hydrodynamic size. Calculating the mass of a protein using SEC with protein standards is inaccurate, because the calculated mass is dependent on the protein having a similar shape to the globular protein standards. For example, if a protein is partially globular and elongated it will elute earlier than expected from the SEC column and the mass is overestimated.

SEC-MALS uses SEC to separate a protein based on its hydrodynamic size and MALS to determine its mass. When light from an incident beam is absorbed by a protein it is scattered in all directions. In general larger proteins scatter light more strongly than smaller proteins do. The ratio of scattered light intensity to incident light intensity at the measurement angle defines the Rayleigh ratio. The Rayleigh ratio is measured at 20 different angles to the incident beam. An accurate concentration of the protein is measured using UV and refractive index (RI) during a SEC-MALS measurement. With all these parameters known an accurate molecular mass can be calculated using the Rayleigh equation:

$$\frac{\mathrm{KC}}{\mathrm{R_\theta}} = \left(\frac{1}{Mw} + 2A_2C\right)\frac{1}{\mathrm{P_\theta}} \qquad \text{Equation (3.1)}$$

C = sample concentration
θ = measurement angle
R_θ = Rayleigh ratio (ratio of scattered light intensity to incident light intensity) at the measurement angle (θ)
Mw = molecular weight
A_2 = second virial coefficient
P_θ = A term defining angular dependence
K = A constant, which is system, solvent and sample dependent

K is defined in the equation below:

$$\mathrm{K} = \frac{4\pi^2}{\lambda_0{}^4 N_A}\left(n_0\frac{dn}{dc}\right)2 \qquad \text{Equation (3.2)}$$

λ = laser wavelength in a vacuum
N_A = Avogadro's number
n_0 = refractive index of the solvent
dn/dc = difference in refractive index between the sample and the solvent

I performed SEC-MALS of LSH-6His in the 100 mM NaCl and 500 mM NaCl buffers tested using DLS (Figure 4.6A and Figure 4.6B respectively). The average molecular mass from triplicate SEC-MALS runs of LSH-6His in the 100 mM NaCl buffer was 96.9 kDa and 90.9 kDa in the 500 mM NaCl buffer (table 4.1). The predicted mass of LSH is 95.8 kDa therefore LSH-6His is a monomer in 100 mM and 500 mM NaCl buffers. Interestingly, in the 100mM buffer LSH-6His might be taking a different conformation than LSH in 500 mM NaCl due to the difference in average mass. However, as the elution volume is the same this is negligible, and when factoring in the standard deviation it is within error of the instrument which is ± 5%. The mass changes across the peak are due to peak broadening, a result of the 10% glycerol present in the buffer which causes drift on the baseline.

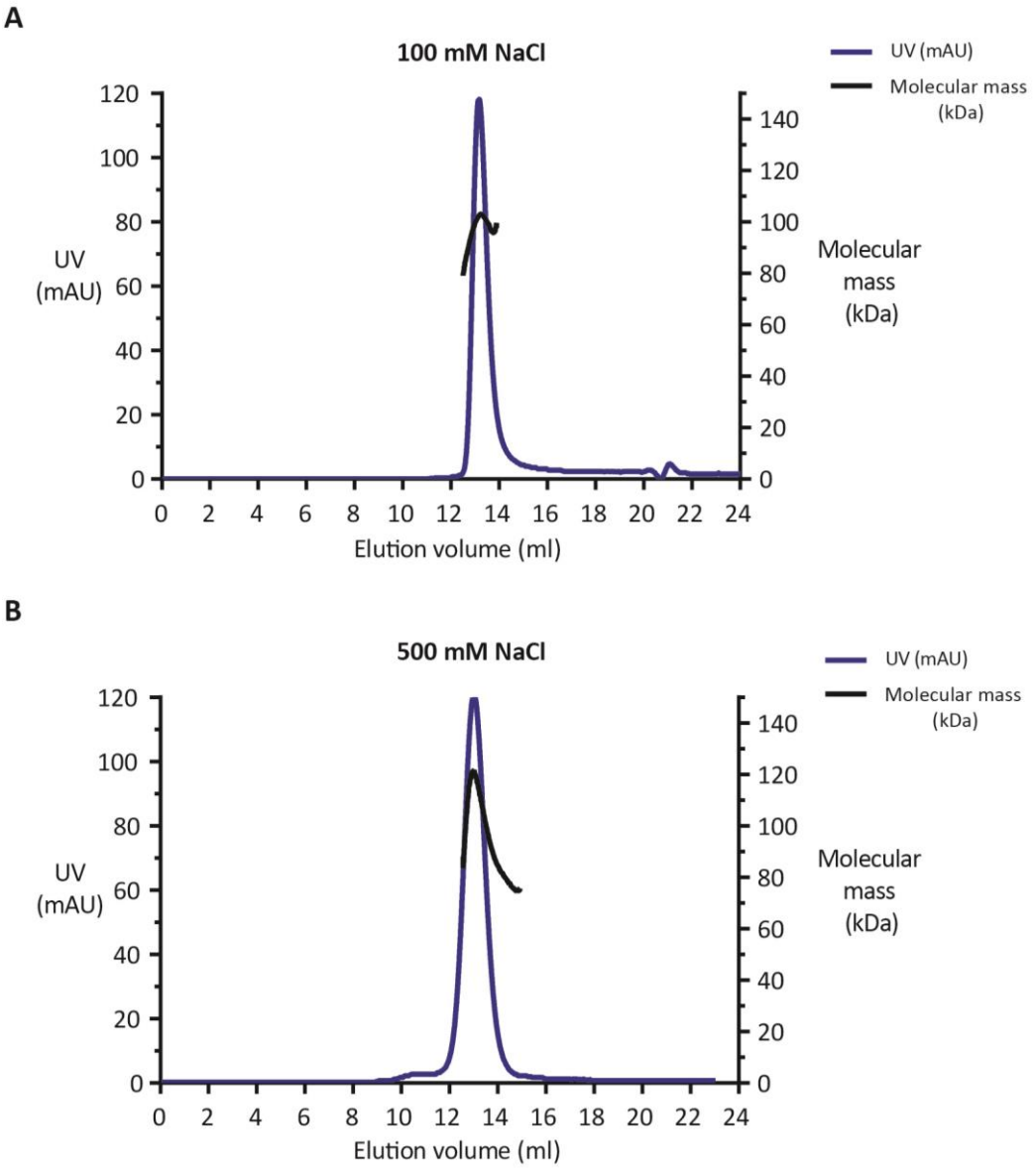

Figure 4.6| LSH-6His is a monomer in 100 mM NaCl and 500 mM NaCl buffers
SEC-MALS chromatogram of LSH-6His in 100 mM NaCl or 500 mM NaCl buffer with UV detection (280 nm) and calculated molecular mass (kDa).
A| LSH-6His elution volume in 100 mM NaCl buffer is 13.1 ml and average molecular mass across elution peak of 96.3 kDa.
B| LSH-6His elution volume in 500 mM NaCl buffer is 13.1 ml and average molecular mass across elution peak of 93.5 kDa.

Table 4.1| The peak elution volume and calculated molecular mass of LSH-6His at 2 mg/ml in 100 mM NaCl or 500 mM NaCl buffer using SEC-MALS

Buffer ionic strength	Elution volume (ml)	Average Molecular Mass (kDa)	Std deviation of Molecular mass (kDa)
100 mM NaCl	13.1	96.9	2.7
500 mM NaCl	13.1	90.8	3.1

As LSH-6His was expressed in insect cells, PTMs such as the addition of N-linked glycans can be added during translation which may alter protein function (Palomares, Estrada-Mondaca and Ramírez, 2004). Therefore, an intact mass of LSH-6His was measured using ESI-Mass spectrometry. The Intact mass of LSH-6His measured by ESI-MS was 95.857 kDa, and the predicted mass of LSH-6His is 95.817 kDa (data not shown). The actual mass is 40 Da more than the predicted, which indicates no glycosylation as this would add at least 200 Da or more to the LSH mass, indicating there could only be small PTMs such as methylation (14 Da), which may be added during expression in eukaryotic cells. An investigation into LSH PTMs was not taken forward, as it was not an aim of this project.

4.2.4| LSH-6His and LSHK237Q is monomeric at 3mg/ml or 22 mg/ml

Protein concentration can affect the oligomeric state of protein and can influence crystallisation strategies. Therefore, I used SEC to assess if protein concentration affected the elution volume (and hence the oligomeric state/shape) of LSH-6His at two different concentrations: 3 mg/ml (3.1 µM) and a 22 mg/ml (22.9 µM) (Figure 4.7A). The elution volume for both samples was 13.06 ml, indicating LSH-6His is monomeric at both concentrations (Figure 4.7A).

I also compared the elution volume of the ATPase inactive mutant LSH^{K237Q}-6His, with LSH-6His (Figure 4.7B).The LSH^{K237Q} point mutant is an inactive ATPase which perturbs LSH function with phenotypes similar to LSH knockdown cells (Burrage *et al.*, 2012; Lungu *et al.*, 2015). LSH^{K237Q}-6His and LSH-6His both eluted at 13.08 ml indicating LSH^{K237Q}-6His is also a monomer, indicating that the mutation does not influence the oligomeric state of the protein. (Figure 4.7B).

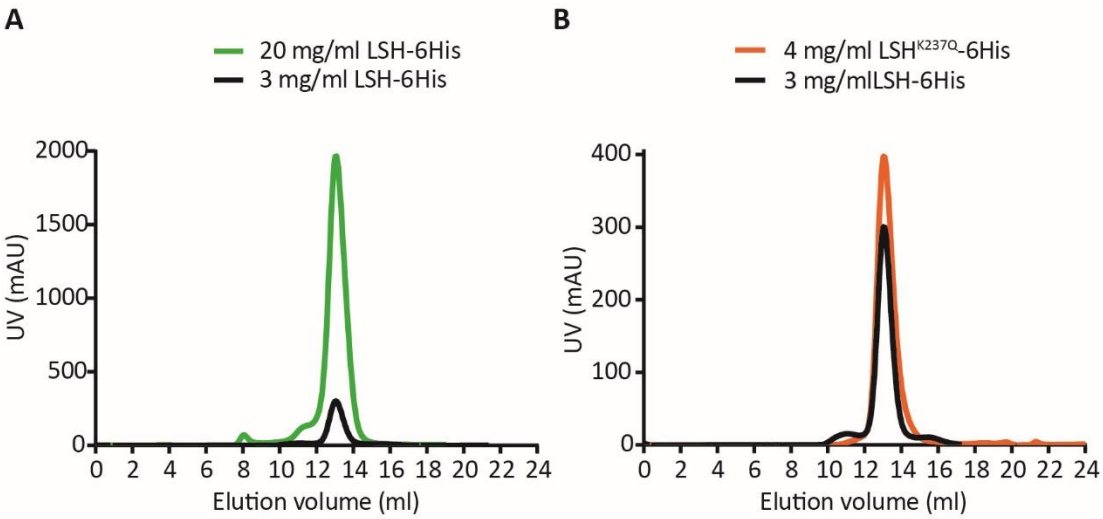

Figure 4.7| SEC elution volume of LSH-6His at low and high concentrations is unchanged and equivalent to LSH^{K237Q}-6His.
A| Chromatogram of LSH-6His SEC with UV 280 nm detection. LSH-6His injected at a concentration of 3 mg/ml or 20 mg/ml elute at 13.06 ml.
B| Chromatogram of LSH-6His and LSH^{K237Q}-6His SEC with UV 280 nm detection. LSH-6His injected at a concentration of 3 mg/ml and LSH^{K237Q}-6His injected at a concentration of 4 mg/ml elute at 13.08 ml.

4.2.5| Crystallisation trials

I have shown that LSH-6His is homogenous and monomeric in low and high ionic strength buffer and at concentrations of 10-100 µM. LSH-6His is stabilised upon ATP or ADP binding. The main aim of this project was to solve a 3D structure of LSH to an atomic resolution of 3Å or less in order to further understand its function. X-ray crystallography was chosen for LSH-6His because it is too large for NMR (should be less than 50 kDa) and too small for cryo-EM (should be greater than 150 kDa).

LSH-6His concentrations of 20-120 µM were trialled with over 1500 different crystallisation conditions in commercially available screens (See section 2.9.5). However, I could not obtain crystals of LSH-6His, either on its own or in the presence of the stabilising ligands ATP or ADP. As LSH is a multidomain protein, there might be disordered domains and flexibility between domains which could cause the formation of multiple conformations of LSH-6His which may prevent adequate protein stacking for crystal formation. The only chromatin remodelers with crystal structures of the conserved ATPase portions are CHD1, RAD54 (Thomä *et al.*, 2005; Hauk *et al.*, 2010) and SNF2 (Xi et al., 2016). Given the ATPase portions are globular I decided to reinvestigate the cloning and expression of the ATPase portion of LSH.

I previously cloned a core LSH ATPase defined as $LSH^{200-821}$ into the Multibac system (section 3.2). However, weak expression and absence of the 6His tag for purification prevented the use of this construct for structural studies (Section 3.2).

CHD1 and LSH are in the SNF2 subfamily, making both proteins more phylogenetically similar than LSH to RAD54 (Flaus *et al.*, 2006). Therefore, I aligned LSH with CHD1 and two other characterised SNF2 subfamily chromatin remodelers ISWI and SNF2 to identify the conserved regions of the chromatin remodelers (Figure 4.8). The N-terminal portions of LSH (residues 1-210) share no similarity to the other remodelers (Figure 4.8). The ATPase lobes 1 and 2 and the C-terminal bridge defined in the crystal structure of CHD1 are conserved in LSH, ISWI and SNF2 chromatin remodelers (Figure 4.8). LSH has a unique stretch of residues from 476-530 which might add greater length and/or flexibility to protrusion 2 between the linker and recA domain lobe as seen in Figure 1.6B. The alignment shows the region of $LSH^{210-821}$ is conserved between the remodelers and is mostly the ATPase region which is structured and therefore potentially a good construct for crystallisation.

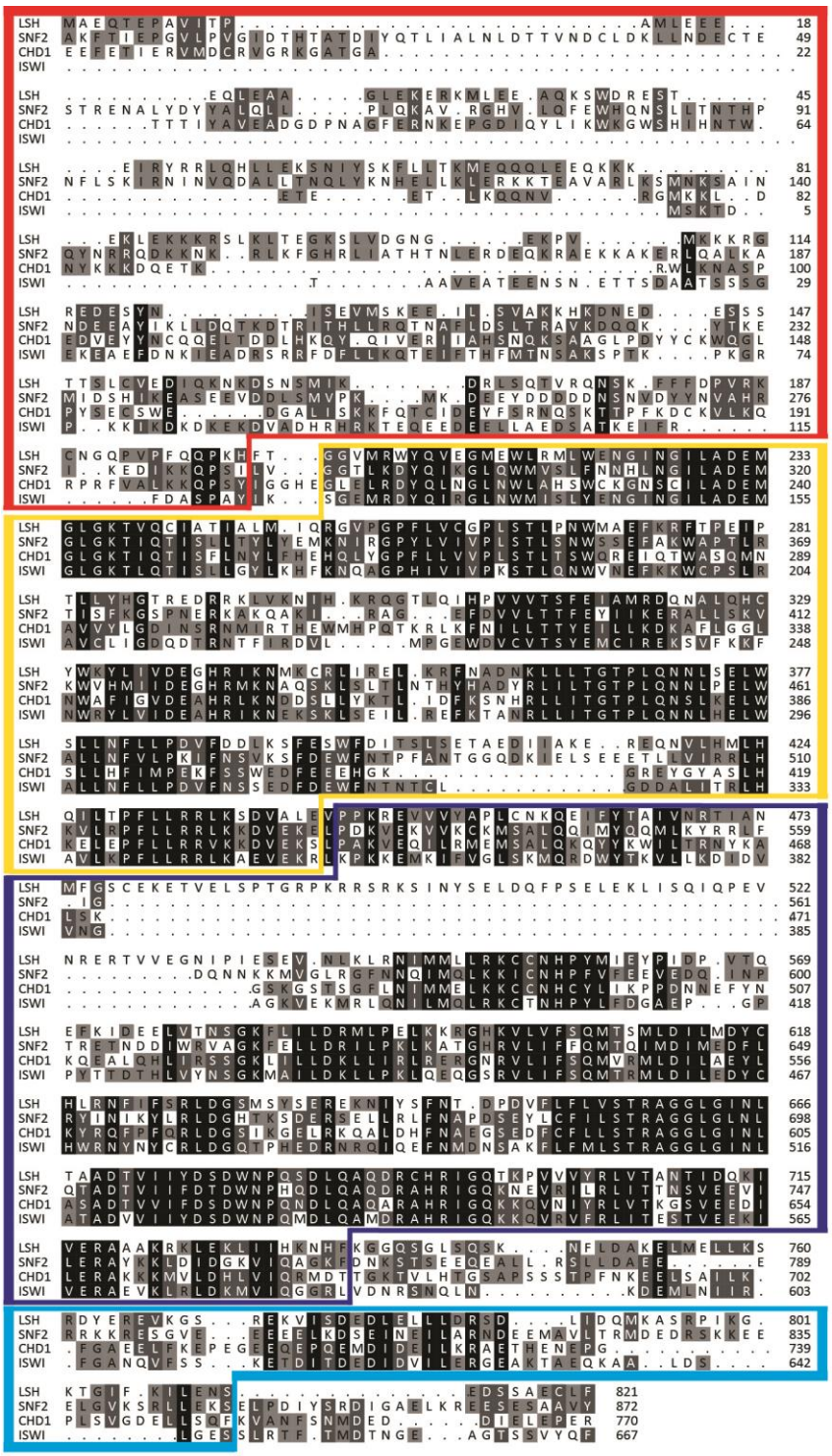

Figure 4.8| Protein sequence alignment of LSH from *M. musculus*, SNF2 from *S. cerevisiae*, CHD1 from *M. musculus* and ISWI from *D. melanogaster*

The main structural domains are highlighted.

The sequence alignments are coloured in light grey to black (low to high similarity) for the groupings of amino acids (Hydrophilic, Hydrophobic, Basic, Acidic and Aromatic).

4.2.6| The N-terminal domain of LSH is predicted to be disordered.

I used the disorder prediction software IUPred to identify disordered regions in LSH (Dosztanyi et al. 2005). During folding, entropy loss is overcome by the stabilising energy provided from inter-residue contacts in globular proteins (Dosztanyi et al. 2005). Intrinsically unstructured proteins (IUPs) or domains do not have inter-residue interactions due to more hydrophilic residues and a higher net charge (Uversky et al. 2000; Dosztanyi et al. 2005). IUPred estimates the potential of polypeptides to form the stabilizing contacts using known structures of globular and disordered proteins (Dosztanyi et al. 2005; Dosztanyi et al. 2005).

IUPred predicted LSH$^{1-195}$ to be disordered (Figure 4.9). This was expected as there is a CC domain in this region. LSH$^{485-500}$ was also predicted to be disordered and falls within LSH$^{481-530}$ which is the region between conserved boxes B and C in which LSH has an extra 50 residues compared to CHD1, ISWI or SNF2. This indicates the linker or protrusion 2 of LSH have added length part of which might be unstructured.

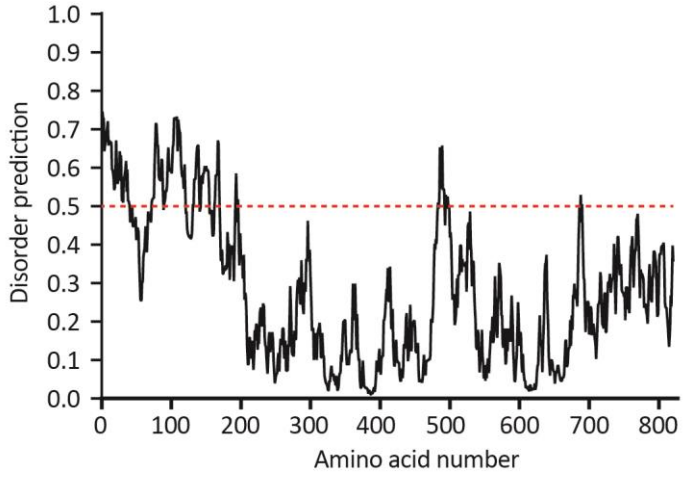

Figure 4.9| LSH$^{1-196}$ and LSH$^{485-500}$ are predicted as being disordered by IUPred
The disorder pediction cut off is 0.5. Any measurement higher than 0.5 is 95% likely to be disordered (Dosztanyi et al. 2005).

The secondary structure elements of LSH were predicted using PSIpred software (Jones, 1999), which identifies evolutionary related proteins through a BLAST search of the input amino acid sequence and uses this information to predict secondary structure (Jones, 1999). LSH$^{1-200}$ is more than 80% helical as predicted by PSIpred (see Appendix - Figure A1).

The half-life of a protein is in part dictated by its N-terminal amino acid, as according to the N-end rule (Bachmair, Finley and Varshavsky, 1986). Studies with a eukaryotic system identified three groups of primary destabilising N-terminal amino acids: basic (R,K,H) bulky hydrophobic (F,L,W,Y) and small uncharged (A,S,T) (Gonda *et al.*, 1989). Therefore, these amino acids were avoided when selecting the N-terminal amino acid for expression of an LSH core ATPase domain.

I chose LSH$^{193-821}$, LSH$^{196-821}$, LSH$^{209-821}$ and LSH$^{219-821}$ sequences for cloning based on protein disorder, secondary structure prediction and the N-end rule. I also chose LSH$^{226-821}$ as this was previously cloned and expressed in mammalian cells in a luciferase reporter assay to study LSH as a transcriptional repressor (Myant and Stancheva, 2008).

4.2.7| Expression of the core ATPase domain of LSH

The five ΔN LSH sequences were each cloned with a C-terminal 6Histag and expressed in insect cells as described previously (section 3.2.1). Western blotting was used to compare expression of the five ΔN LSH proteins to LSH-6His (Figure 4.10). Expression of all ΔN LSH-6His proteins was ~10-fold weaker than LSH-6His (Figure 4.10). The best expressing ΔN protein, LSH$^{226-821}$-6His had 9.5 times lower expression than LSH-6His (Figure 4.10B). The weak expression of the five ΔN LSH constructs was similar to LSH$^{200-821}$-6His (section 3.2.1), indicating the LSH$^{193-226}$ region is important for the stable expression of LSH. Therefore, these constructs were not taken further in this study.

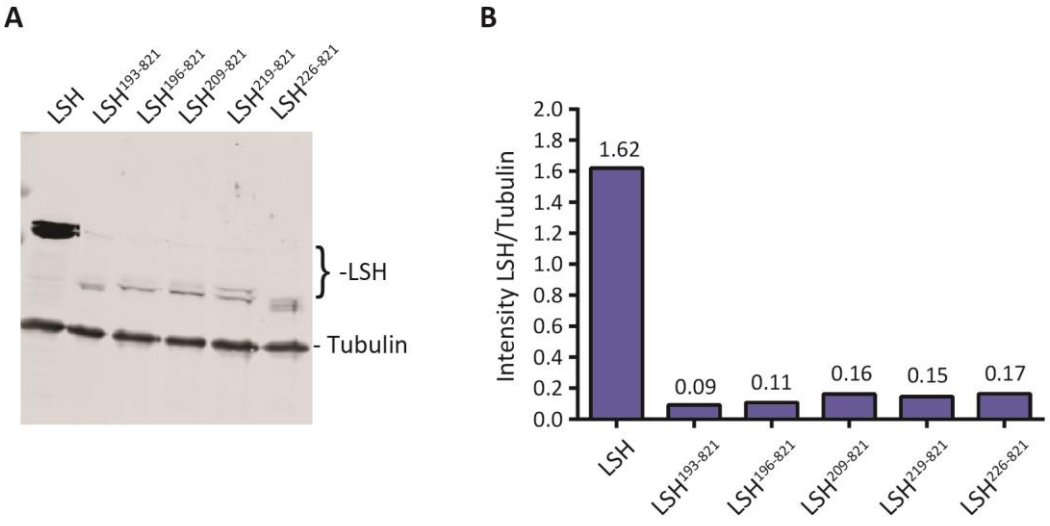

Figure 4.10| Comparing the expression of five ΔN LSH-6His constructs to LSH-6His
A| Anti-LSH and anti-Tubulin Western blot of LSH-6His and the five ΔN LSH-6His constructs.
B| Graph showing the relative normalised intensities of LSH/Tubulin.

4.2.8| A core LSH ATPase domain is defined by limited proteolysis

Limited proteolysis followed by mass-spectrometry was used to determine stable domains of LSH that could be better targets for improved expression of a stable truncated LSH and which could be more amenable for crystallisation trials (Gao *et al.*, 2005). During limited proteolysis low concentrations of protease are used to digest flexible loops, creating protein fragments (Fontana *et al.*, 2004). The protein fragments are separated by size using SDS-PAGE and the protein region that the fragment of digested protein originated from can be identified by MALDI-ToF (Fontana *et al.*, 2004). I decided to digest LSH-6His using the broadly specific endopeptidases subtilisin and elastase (Figure 4.11).

Elastase and subtilsin digested LSH-6His into a product with a mass of ~70 kDa as estimated by SDS-PAGE (Figure 4.11A). MALDI-ToF mass spectrometry confirmed this fragment was $LSH^{178-821}$-6His with coverage of 75% (data not shown), which has a predicted mass of 76 kDa which is in agreement with the SDS-PAGE size (Figure 4.11B). The MALDI-ToF MS result indicates the region preceding LSH^{178} is a disordered region linked to the globular SNF2 ATPase domain. This could explain why all previous ΔN LSH constructs were weakly expressed, as they were likely positioned in the folded region of a globular domain. Therefore, I decided to clone the core ATPase domain of LSH starting around the region LSH^{178}.

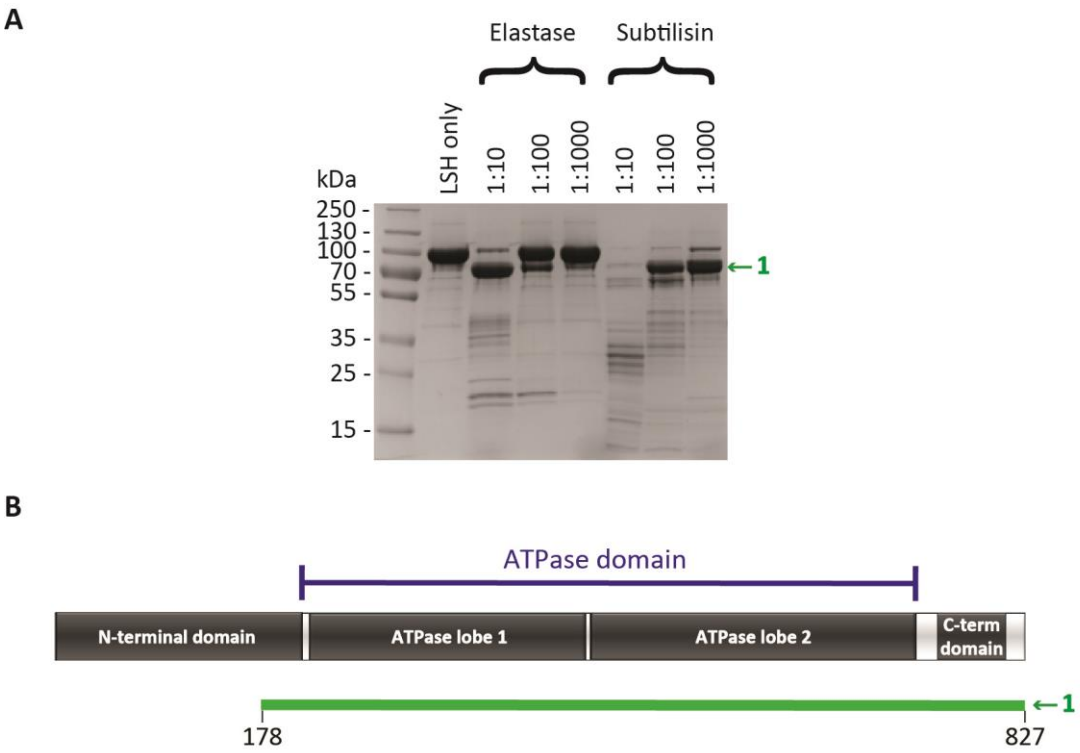

Figure 4.11| Limited proteolysis of LSH-6His with elastase and subtilisin
A| Coomassie stained SDS-PAGE gel of LSH-6His proteolysis using elastase or subtilisin at protease:LSH-6His (w/w) ratios of 1:10, 1:100 and 1:1000 .A LSH-6His digestion product defined as 1 is seen in both the elastase and subtilsin treatments.
B| Domain architecture of LSH-6His with the digestion product defined as 1 as being $LSH^{178-827}$ from MALDI-ToF mass spectrometry measurements.

ΔN LSH constructs were designed to start close to the limited proteolysis product $LSH^{178-821}$, whilst also factoring in the N-end rule, secondary structure prediction and hydrophobic cluster analysis (HCA). HCA is based on the statistical studies of 3D structures which identified that hydrophobic amino acids gathered into clusters are likely secondary structures in a globular region (Woodcock, Mornon and Henrissat, 1992; Callebaut *et al.*, 1997). Five ΔN LSH constructs were designed: $LSH^{146-821}$, $LSH^{162-821}$, $LSH^{176-821}$, $LSH^{178-821}$ and $LSH^{184-821}$.

4.2.9| The N-terminal domain of LSH is required for stability

The five ΔN constructs were all cloned and expressed as in section 4.2.7. Western blotting was used to compare expression of the five ΔN LSH proteins to LSH-6His (Figure 4.12). Expression of all ΔN LSH-6His constructs was weaker than LSH-6His (Figure 4.12). The best expressing ΔN construct, $LSH^{176-821}$-6His had a 1.5 times lower expression than LSH-6His (Figure 4.12B). However LSH-6His expression was 5 times weaker than LSH-6His expressed previously (Figure 4.10).

This result made it difficult to confirm whether expression of the new ΔN LSH-6His constructs (Figure 4.12) was as weak as the old ΔN LSH-6His constructs (Figure 4.10) or expression of all the new ΔN LSH-6His constructs and LSH-6His was weak due to poor transfection.

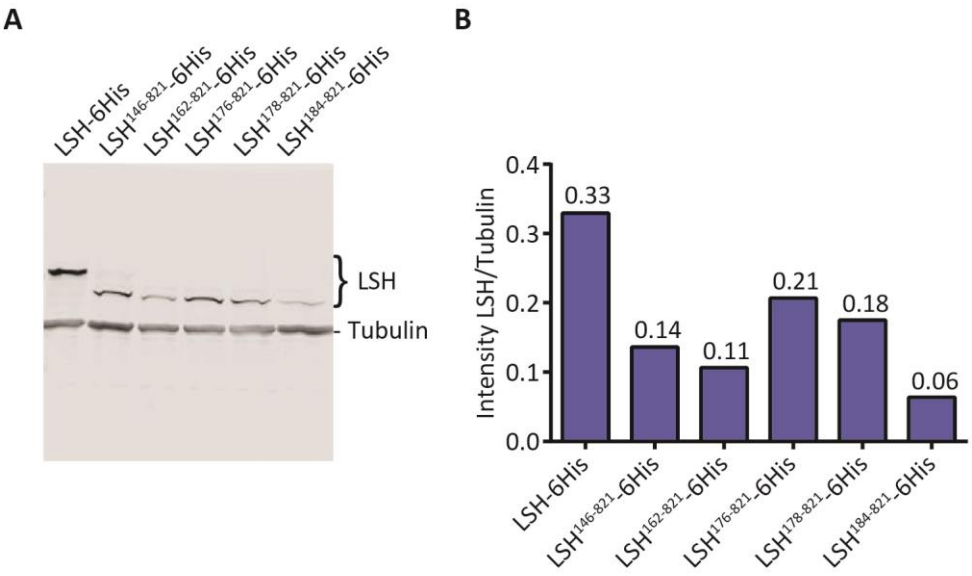

Figure 4.12| Comparing the expression of five ΔN LSH-6His constructs to LSH-6His
A| Anti-LSH and anti-Tubulin Western blot of LSH-6His and the Five ΔN LSH-6His constructs.
B| Bar graph showing the relative normalised intensities of LSH/Tubulin.

LSH$^{176-821}$-6His was the best expressing ΔN LSH protein, therefore expression was scaled up to 1L of HighFive™ cells and the protein was purified by IMAC and phosphocellulose steps, as described for LSH-6His (Section 2.6.4). LSH$^{176-821}$-6His could be partially purified to ~60% purity after the IMAC and phosphocellulose purifications, however the protein degraded during and after purification, diminishing the protein yield to less than 100 µg before a SEC purification step (Data not shown). The insufficient yield and instability of LSH$^{176-821}$-6His made it an unsuitable construct to use. 11 different ΔN LSH proteins were unstable with weak expression making them unsuitable for structural studies. Therefore, an alternative method termed *in situ* limited proteolysis was chosen to be used for crystallisation trials.

4.2.10| *In situ* proteolysis of LSH-6His with subtilisin indicates proteolytic digestion of the LSH C-terminus first requires proteolytic cleavage of LSH$^{1-178}$

In situ limited proteolysis is the addition of trace amounts of protease to the protein before crystallisation (Dong *et al.*, 2007). 270 proteins that could not be crystallised or which formed crystals with poor diffraction were used in a large trial of *in situ* limited proteolysis (Wernimont and Edwards, 2009). 34 of the 270 protein formed crystals which diffracted to 2.8 Å (Wernimont and Edwards, 2009). Therefore, I decided to use this method and trialled subtilisin, as 10 times less of it was required for digestion of the LSH-6His N-terminal domain compared to elastase (Figure 4.11B). Before setting up crystallisation trials, I tested proteolysis with three subtilisin:LSH-6His ratios (w/w) of 1:50000, 1:10000 and 1:5000 at room temperature following the published method (Wernimont & Edwards 2009). However, cloudy precipitate occurred within 1 hr and LSH-6His was completely digested to fragments smaller than 15 kDa at all ratios (data not shown). Therefore, I tested the same ratios at 6°C to find the condition which digested LSH-6His into the LSH$^{178-821}$-6His fragment after 1 day or longer to provide time for crystallisation without over digestion (Figure 4.13A). A subtilisin:LSH-6His ratio of 1:50000 digested ~50% of LSH$^{178-821}$-6His after 5 days. As LSH-6His was not fully digested into LSH$^{178-821}$-6His it was not optimal to use in crystallisation trials (Figure 4.13A). The 1:10000 and 1:5000 ratios of subtilisin:LSH-6His digested more than 95% of LSH-6His after 1 day into LSH$^{178-821}$-6His and a smaller ~60 kDa digestion product as estimated by SDS-PAGE (Figure 4.13A). MALDI-ToF mass spectrometry of the ~60 kDa polypeptide had a coverage of 60% in the LSH$^{178-704}$ region, which would have a predicted mass of 61 kDa which is in agreement with the size estimated by SDS-PAGE (Figure 4.13A and Figure 4.13B and S4.4). LSH$^{178-704}$ is the 'core' ATPase portion of LSH-6His and suggests the N and C-terminal regions are connected to the ATPase region with disordered linkers and possibly flexible linkers. A ~20 kDa digestion product produced in all ratios of subtilisin:LSH-6His was defined as the N-terminal domain. However, establishing the exact sequence was problematic due to the large numbers of lysine residues in this region but based on size, it was probably a LSH$^{1-178}$ digestion product (Figure 4.13A and Figure 4.13B).

LSH *In situ* proteolysis was performed with a subtilisin:LSH ratio of 1:10000 with commercial screens, however only amorphous precipitate was visualised.

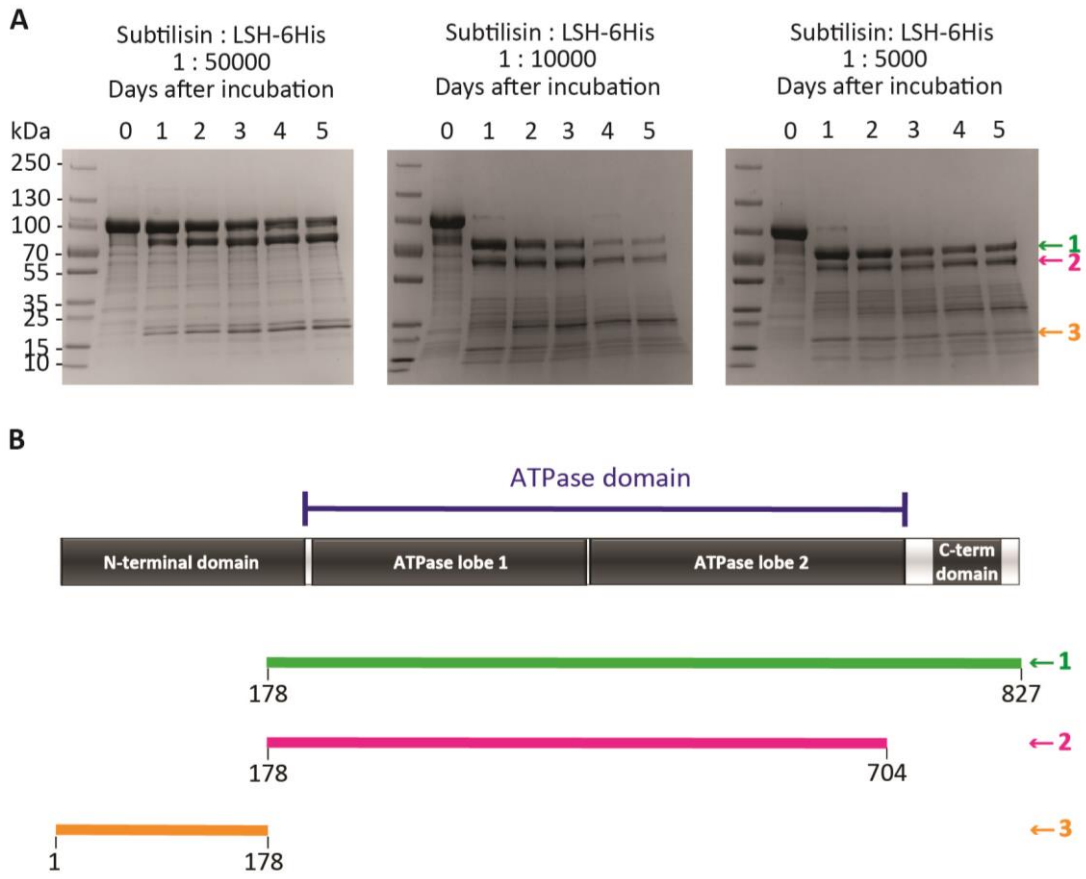

Figure 4.13| *In situ* limited proteolysis of LSH-6His with subtilisin
A| Coomassie stained SDS-PAGE gels of LSH-6His proteolysis at subtilisin:LSH-6His (w/w) ratios of 1:50000, 1:10000 and 1:5000 (left to right) over five days. The LSH-6His digestion product defined as 1 is detected in all ratios. The LSH-6His digestion product defined as 2 is seen in 1:10000 and 1:5000 ratios. The LSH-6His digestion product defined as 3 is seen in all ratios.
B| Domain archietcure of LSH-6His with the bands defined as 1 being $LSH^{178-827}$, 2 being $LSH^{178-704}$ and 3 being LSH^{1-178} from MALDI-ToF mass spectrometry measurements and western blotting experiments.

$LSH^{178-704}$ which is the 'core' ATPase portion and does not contain the flexible N and C-termini might provide better expression and would be a better crystallisation target than the $LSH^{176-821}$-6His which was unstable. However, due to time constraints I could not do this.

LSH^{1-178} was a probable product of LSH-6His limited proteolysis with subtilisin and is unique to LSH with no sequence conservation in the closely related chromatin remodelling subfamilies; CHD1, SNF2 and ISWI (Figure 4.13A and Figure 4.8). However, the N-terminal domain of LSH is conserved across higher eukaryotes indicating its functional importance (Figure 4.14). Functionally, the LSH N-terminal domain is required for the nuclear localisation of LSH to the nucleus, and binding to DNMT3B *in vitro* (Myant & Stancheva 2008; Lee et al. 2000; Yan et al. 2003). Studying the structure and biochemistry of this domain would provide insights into its role in LSH function.

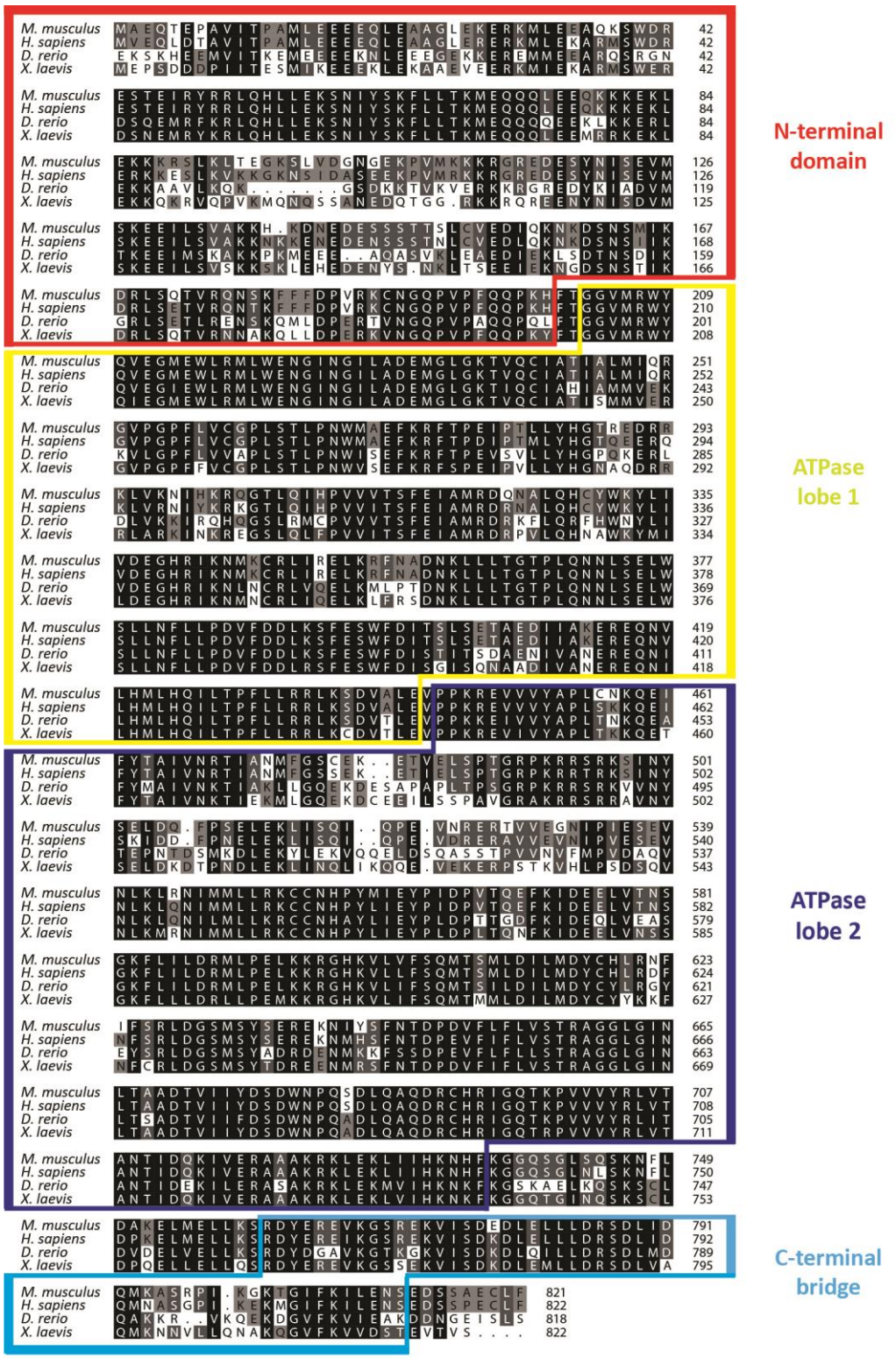

Figure 4.14| Protein sequence alignment of LSH from the species, *M. musculus*, *H. sapiens*, *D. rerio* and *X. laevis*

The main structural domains are highlighted.

The sequence alignments are coloured in light grey to black (low to high similarity) for the groupings of amino acids (Hydrophilic, Hydrophobic, Basic, Acidic and Aromatic). Structural domains are based on the crystal structure of Chd1 (PDB: 3MWY) (Hauk et al. 2010).

4.2.12| Expression and purification of LSH$^{1-176}$

LSH$^{1-176}$ was chosen for cloning and expression as its C-terminal amino acid was in an unstructured region based on PSIpred secondary structure prediction and HCA prediction (Appendix S4.2). Mari Eltermann, a MChem student, successfully cloned LSH$^{1-176}$ with an N-terminal TEV cleavable 6His tag and expressed soluble 6His-TEV-LSH$^{1-176}$ in *E. coli* (Mari Eltermann, 2015). An optimised purification protocol was developed with an initial IMAC step followed by 6His tag cleavage using TEV protease (Figure 4.14A-C). An anion exchange column did not bind LSH$^{1-176}$ but DNA contaminants and protein contaminants between 55-250 kDa bound and eluted during a 1 M NaCl gradient (Figure 4.14D and Figure 4.14E). The LSH$^{1-176}$ which flowed-through this resin was greater than 95% pure and underwent SEC to remove aggregates. LSH$^{1-176}$ eluted as a single asymmetric peak at 12.85 ml during SEC (Figure 4.14F and Figure 4.14G). This is 0.2 ml earlier than LSH-6His (Figure 4.8) indicating that LSH$^{1-176}$ has a similar hydrodynamic radius as full-length LSH-6His. This could be due to it adopting an elongated conformation or forming dimers or higher order oligomers. Therefore SEC-MALS was used to assess the oligomeric state of LSH$^{1-176}$ at four different concentrations (10, 5, 2.5 and 1.25 mg/ml).

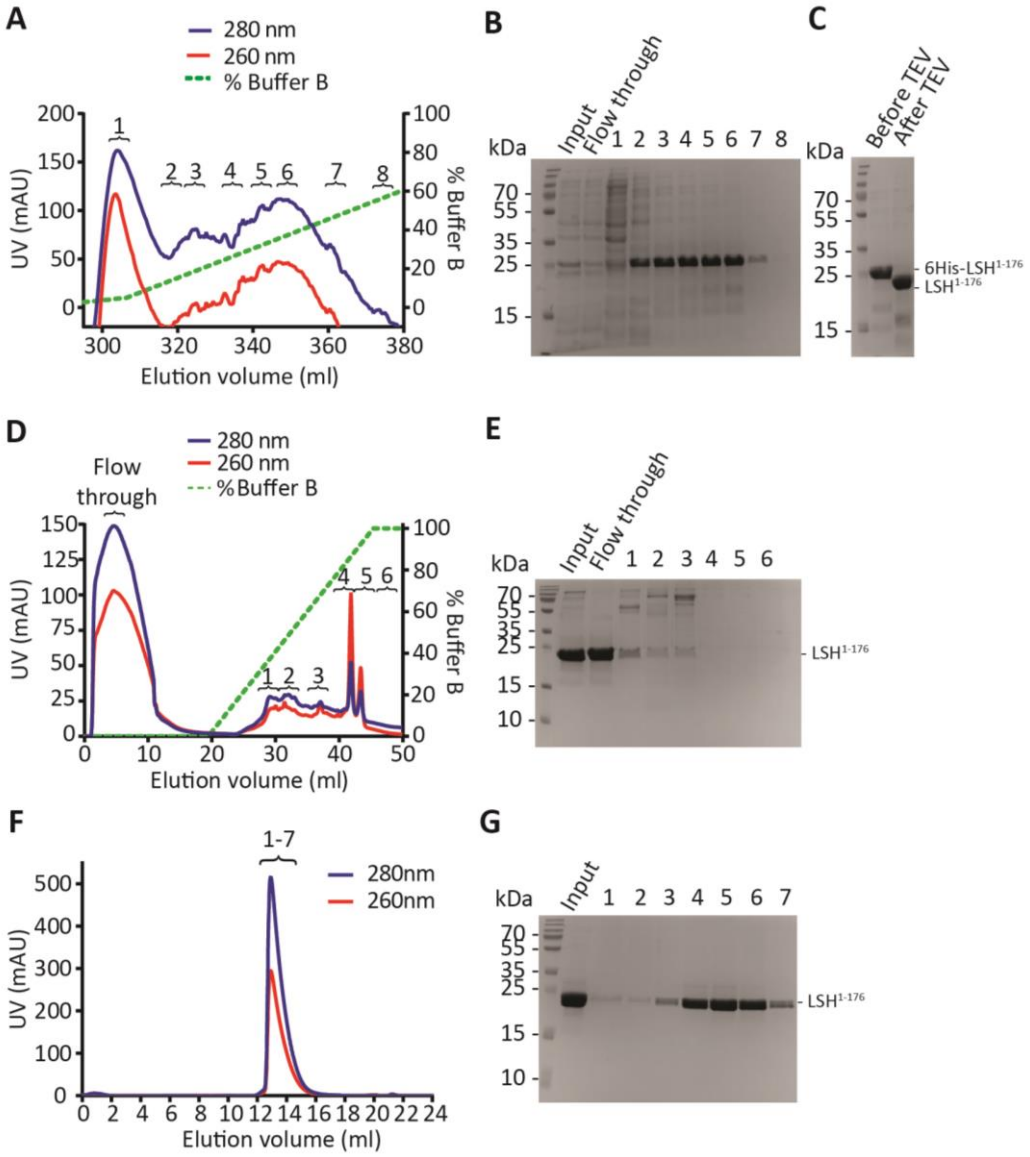

Figure 4.15| His-TEV-LSH$^{1-176}$ purification using IMAC, anion exchange and SEC

A| Chromatogram of the 6His-TEV-LSH$^{1-176}$ IMAC purification with UV detection (260 nm and 280 nm).

B| Coomassie stained SDS-PAGE of the 6His-TEV-LSH$^{1-176}$ IMAC purification.

C| TEV cleavage of 6His-TEV-LSH$^{1-176.}$.

D| Chromatogram of the LSH1-176 anion exchange chromagraphy purification with UV detection (260 nm and 280 nm)

E| Coomassie stained SDS-PAGE of the LSH$^{1-176}$ anion exchange chromatography purification.

F| Chromatogram of the LSH1-176 SEC purification with UV detection (260 nm and 280 nm).

G| Coomassie stained SDS-PAGE of the LSH$^{1-176}$ SEC purification.

4.2.13| LSH$^{1-176}$ is a monomer and takes different concentration-dependent conformations.

LSH$^{1-176}$ injected at a concentration of 1.25 mg/ml eluted at a peak volume of 14.6 ml, 0.8 ml later than LSH$^{1-176}$ injected at a concentration of 10 mg/ml which eluted at a peak volume of 13.8 ml (Table 4.2).

The intact mass (using ESI-MS) and predicted mass of LSH$^{1-176}$ was 20.649 kDa and 20.650 kDa respectively, indicating no PTMs on the protein (Appendix S4.1). The estimated mass of LSH$^{1-176}$ injected at a concentration of 10 mg/ml was 23.4 kDa, close to the intact and predicted mass, therefore it can be concluded that LSH$^{1-176}$ is a monomer. The estimated mass of LSH$^{1-176}$ injected at a concentration 1.25 mg/ml was 1.3 kDa lighter than LSH$^{1-176}$ injected at a concentration 10 mg/ml.

The later elution time and heavier mass of LSH$^{1-176}$ at higher concentrations indicates LSH$^{1-176}$ molecules have flexible movement which could be taking the form of a more compact molecule at higher concentrations.

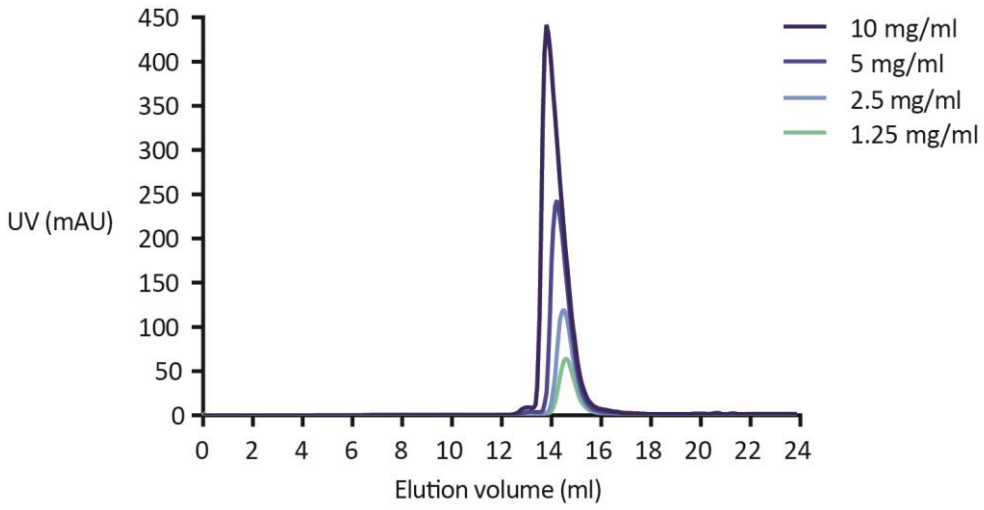

Figure 4.16| LSH$^{1-176}$ SEC elution profile is concentration dependent
Chromatogram of LSH$^{1-176}$ SEC performed at different concentrations (10, 5, 2.5 and 1.25 mg/ml) with UV detection (280 nm).

Table 4.2| The peak elution volume and estimated molecular mass of LSH$^{1-176}$ at 10, 5, 2.5 and 1.25 mg/ml using SEC-MALS

Concentration (mg/ml)	Elution volume (ml)	Molecular mass (kDa)
10	13.8	23.4
5	14.2	22.7
2.5	14.5	22.5
1.25	14.6	22.1

NMR and CD experiments performed by Mari Eltermann indicated that LSH$^{1-176}$ has helical secondary structure and no tertiary structure indicating the protein was disordered (Mari Eltermann, 2015). Crystallisation trials of LSH$^{1-176}$ with 400 commercial conditions were performed, however no crystals formed, likely due to unstructured nature of the protein (Mari Eltermann, 2015). As LSH$^{1-176}$ is helical and unstructured, it would likely have dynamic flexibility with open and closed conformations based upon concentration and intermolecular interactions.

4.3| Discussion

The major aim of this work was to gain further insights into the LSH structure and how its flanking domain regulates enzymatic function. In this chapter I have biophysically characterised LSH and identified unstructured and structured domains of LSH.

LSH-6His was determined to be a monomer using SEC-MALS and did not show oligomerisation when LSH was injected into a SEC column at 230 µM (Figure 4.6 and Figure 4.7). This indicates the CC region of the LSH N-terminal domain does not cause self-dimerisation or self-oligomerisation as witnessed in other coiled-coil domain containing proteins (Lupas and Gruber, 2005). Native LSH from nuclear extracts of human cells which underwent SEC was calculated as a monomer from elution volumes which is in agreement with my data (Myant and Stancheva, 2008). This indicates LSH is likely biologically functional as a monomer.

Although LSH-6His is folded, binds ATP and is homogenous, crystals of LSH-6His could not be obtained using more than 1500 crystal screens. Lack of crystal growth was predicted to be because of unstructured regions located at the N-terminal domain and in the extended linker-protrusion 2 region (Figure 4.9). The first 177 residues of the N-terminal domain of LSH are cleaved in limited proteolysis using broadly specific proteases showing that the region around LSH^{178} is unstructured (Figure 4.11 and Figure 4.13). A longer proteolysis time led to cleavage of $LSH^{178\text{-}821}$ into $LSH^{178\text{-}704}$ (Figure 4.13) indicating the first 177 residues of the N-terminal domain could be linked to the ATPase region via a flexible linker and may contact and cover the region at LSH^{704} which encompasses the ATPase region without the C-terminal brace (Figure 1.6).

The majority of the LSH ATPase region ($LSH^{178\text{-}704}$) is globular and compact as it was not digested with subtilisin or elastase proteases at low concentrations (Figure 4.11 and 4.13). $LSH^{1\text{-}176}$ is unstructured, elongated and has intrinsic flexibility as determined from SEC-MALS (Figure 4.15) and previous studies using CD and NMR methods (Mari Eltermann,. 2015). This is in agreement with the biochemical data suggesting $LSH^{1\text{-}176}$ may form contacts along the length of ATPase lobe 1 and 2.

Given the LSH N-terminal domain (LSH^{1-215}) sequence is unique to LSH (Figure 4.8) it is difficult to draw any comparisons with other chromatin remodelers whose structure has been solved. In order to dissect the molecular function of the unique N-terminal domain of LSH to the ATPase region and C-terminal domain expression of a ΔN-LSH protein was tested (Figure 4.10 and Figure 4.12). 11 different ΔN-LSH constructs were expressed but all expressed weakly and showed instability (Figure 4.10 and Figure 4.12). Purifying the greatest expressing ΔN LSH construct ($LSH^{176-821}$-6His) was problematic due to extensive degradation and less than a 100 µg yield of less than 60% pure protein from 1 L of insect cells (data not shown). This indicated the N-terminal domain is critical for stabilising the ATPase domain of LSH. The chromatin remodeler RAD54 also has an unstructured N-terminal domain as defined through limited proteolysis experiments (Raschle *et al.*, 2004), but shares no sequence similarity with LSH. The N-terminal region of RAD54 is required for RAD51 binding, but acts independently of the ATPase region (Raschle *et al.*, 2004). The N-terminal domain of LSH is required for E2F3 binding with its CC region and potentially DNMT3B (von Eyss et al. 2012, Myant & Stancheva 2008). As with LSH, a ΔN-RAD54 protein degraded significantly, however several ion exchange steps enabled the purification of pure non-degraded protein (1mg yield from 24 L of *E. coli* cells) (Raschle *et al.*, 2004).

If the N-terminal domain interacts with a binding partner such as E2F3 or DNMT3B it may regulate ATPase activity by allosteric movement or may function independently of the ATPase region as seen with RAD54 (Raschle *et al.*, 2004). In order to study this, purification of the core ATPase domain from the N-terminal domain would be necessary. Since cloning and expression of 11 different LSH N-terminal truncation constructs failed due to protein instability I tried purifying $LSH^{178-821}$-6His after subtilisin digestion. I used an IMAC step to bind $LSH^{178-821}$-6His and remove LSH^{1-178} with imidazole washing; however, both fragments co-eluted indicating the N-terminal domain has hydrogen contacts or electrostatic interactions with $LSH^{178-821}$-6His (data not shown).

Therefore, a purification protocol would need to be optimised as for RAD54 and large scale expression (greater than 10 L insect cells used) of $LSH^{176-821}$ is required to study the ATPase region only. Time constraints prevented this.

The structure and domain interactions of LSH could be studied in more detail using limited proteolysis with a specific protease such as trypsin, which cleaves the carboxyl side of lysine or arginine except when either is bound to a C-terminal proline. This has been used to define domain boundaries of an intrinsically disordered protein (Adams *et al.*, 2007). Limited proteolysis could be monitored with LSH bound to ligands (ATP or ADP) and DNA to monitor what conformational changes are made to the domains in the protein, as performed with ISWI (Mueller-Planitz *et al.*, 2013). If proteolysis in the region of LSH^{178} is slower, faster or abolished upon DNA binding then the N-terminal domain would be regulating ATP hydrolysis through DNA binding. A lack of time meant these experiments could not be performed.

ADP-BeF has been used as a transition state mimic to trap ATPases and more specifically chromatin remodelling ATPases such as ACF (Racki *et al.*, 2009) or Chd1 as two examples (Sundaramoorthy *et al.*, 2017). Therefore ADP-BeF could be used for trapping LSH in a transition state which could provide a better condition for LSH crystallisation. A more accurate mass determination by SEC-MALS could be possible following the removal of the 10% glycerol in the sample through dialysis. This would enable a flatter and more stable baseline lessening mass fluctuations when the mass is calculated. As seen in figure 4.11 elastase digested LSH into a major truncated species spanning LSH176-821. This could prove an alternative and more promising method for generating LSH with its core ATPase and if time allowed this would have been pursued.

To investigate how the N-terminal domain and C-terminal domain of LSH interacts with its ATPase region complementary structural biology methods were used and are addressed in the next chapter.

Chapter 5| A low resolution structure of LSH

5.1| Introduction

One of the main aims of this study was to investigate the molecular function of LSH by studying its structure. Hybrid structural methods were implemented to study a LSH structure. In this chapter, I have determined LSH has an elongated tri-lobal structure using the combined methods of negative staining electron microscopy and Small Angle X-ray scattering (SAXS). I have also used crosslinking mass spectrometry to show the N-terminal domain of LSH is proximal to the C-terminal domain.

5.2| Results

5.2.1| Intramolecular crosslinking and mass spectrometry of LSH-6His

In the last decade crosslinking mass spectrometry (XL-MS) of proteins or multi-protein complexes has become a complementary tool for high resolution structural methods (NMR, crystallography, cryo-EM) and low resolution methods (negative stain EM, SAXS). Intramolecular contacts between the domains of an individual protein or intermolecular crosslinks between proteins (or protein subunits) in a complex can be crosslinked natively in solution and detected using MS (Rappsilber, 2011; Leitner *et al.*, 2016).

XL-MS involves incubating a purified protein or protein complex in its native state with a chemical crosslinking reagent, which forms covalent bonds to reactive exposed amino acid side chains (typically amino groups or thiols) (Rappsilber, 2011; Leitner *et al.*, 2016). Trypsin digestion of the crosslinked protein/protein complex results in crosslinked peptides, which are enriched and then analysed by liquid chromatography tandem MS (LC-MS/MS), a non-crosslinked trypsin digested protein sample is used as a control. Raw mass spectra are processed into peak lists using MaxQuant software (Cox and Mann, 2008b). The peak lists are searched against the protein or protein complex sequence/s for the identification of cross-linked peptides and non-crosslinked linear peptides using computational software (Xi-software - ERI Edinburgh) (Rappsilber, 2011; Leitner *et al.*, 2016). This information can be used to determine domain-domain interactions, the orientation of the subunits and the maximum distance constraint (based on the spacer arm of the crosslinking reagent) between the peptides for molecular modelling purposes (Rappsilber, 2011; Leitner *et al.*, 2016).

One of the most commonly used crosslinking reagents is bis(sulfosuccinimidyl) suberate (BS^3), an amine-to-amine homobifunctional crosslinker, which couples to the ε-amine- group on lysine (Leitner *et al.*, 2016). The BS^3 crosslinker has a spacer arm length of 11.4 Å plus the 5 Å distance between the reactive amine groups, making a total distance of 22.4 Å. LSH has 75 lysine residues, therefore, BS^3 was a good crosslinking reagent to study LSH-6His structure by measuring intramolecular crosslinks.

Initially crosslinking of LSH-6His with BS^3 crosslinker at eight LSH-6His:BS^3 w/w ratios was performed on a small scale (10 ug of LSH) to optimise conditions (Figure 5.1A). Crosslinked LSH-6His was resolved by SDS-PAGE into monomeric crosslinked species and higher-order crosslinked species (Figure 5.1A). Monomeric crosslinked LSH-6His species at a LSH-6His:BS^3 w/w ratio of 1:3 was considered optimal for producing the greatest number of crosslinks.

For mass spectrometry analysis 80 µg of LSH-6His was crosslinked at a LSH-6His:BS^3 w/w ratio of 1:3, SDS-PAGE performed, the monomeric crosslinked LSH-6His species band cut from the gel, trypsin digested and strong cation exchange performed for peptide enrichment (Figure 5.1B). The mass of the enriched peptides and analysis of data for crosslinks was performed by Dr Juan Zou. 46 auto-validated spectra matches with score cut-off of 7.0 to support 33 link pairs (Appendix Table A1).

Intramolecular LSH-6His lysine-lysine crosslinks are shown as arcs along its domain architecture in Figure 5.1C. A cluster of 12 crosslinks (red arcs) between lysines within the N-terminal domain, and 5 crosslinks (blue arcs) within the ATPase lobe 2 or C-terminal domain were measured (Figure 5.1C). Due to the crosslinks covering short distances, some of the crosslinks might be intrapeptide links along the same domain also known as loop-links, indicating these regions could be helical/unstructured and exposed to nucleophilic attack by the BS^3 crosslinker (Maiolica *et al.*, 2007). This data is in agreement with the limited proteolysis experiments where the non-specific protease subtilisin which cleaves in unstructured regions digested LSH-6His firstly at LSH^{178} and then LSH^{704}, which are in the same region where the crosslinks were measured.

There are 9 cross-links (black arcs) between lysines in the N-terminal domain and in the ATPase lobe 2/C-terminal domain region (Figure 5.1C). This indicates the N-terminal domain of LSH is in the proximity of the C-terminal in the ATPase region ($LSH^{634-794}$). 7 of the 9 links are between lysines in LSH^{97-179} and $LSH^{698-736}$, suggesting these residues are surface accessible and are separated by less than 22.4 Å. This indicates the LSH N-terminal domain is within the proximity of the C-terminal region of the ATPase region, in particular, the C-terminal domain.

There are 7 (grey) crosslinks, 5 of which link the C-terminus of ATPase lobe 2 or the C-terminal domain to the ATPase lobes and the region between them (Figure 5.1C). This indicates the C-terminal domain is within 22.4 Å of the linker between the ATPase lobes 1 to ATPase lobe 2 and regions of each ATPase lobe.

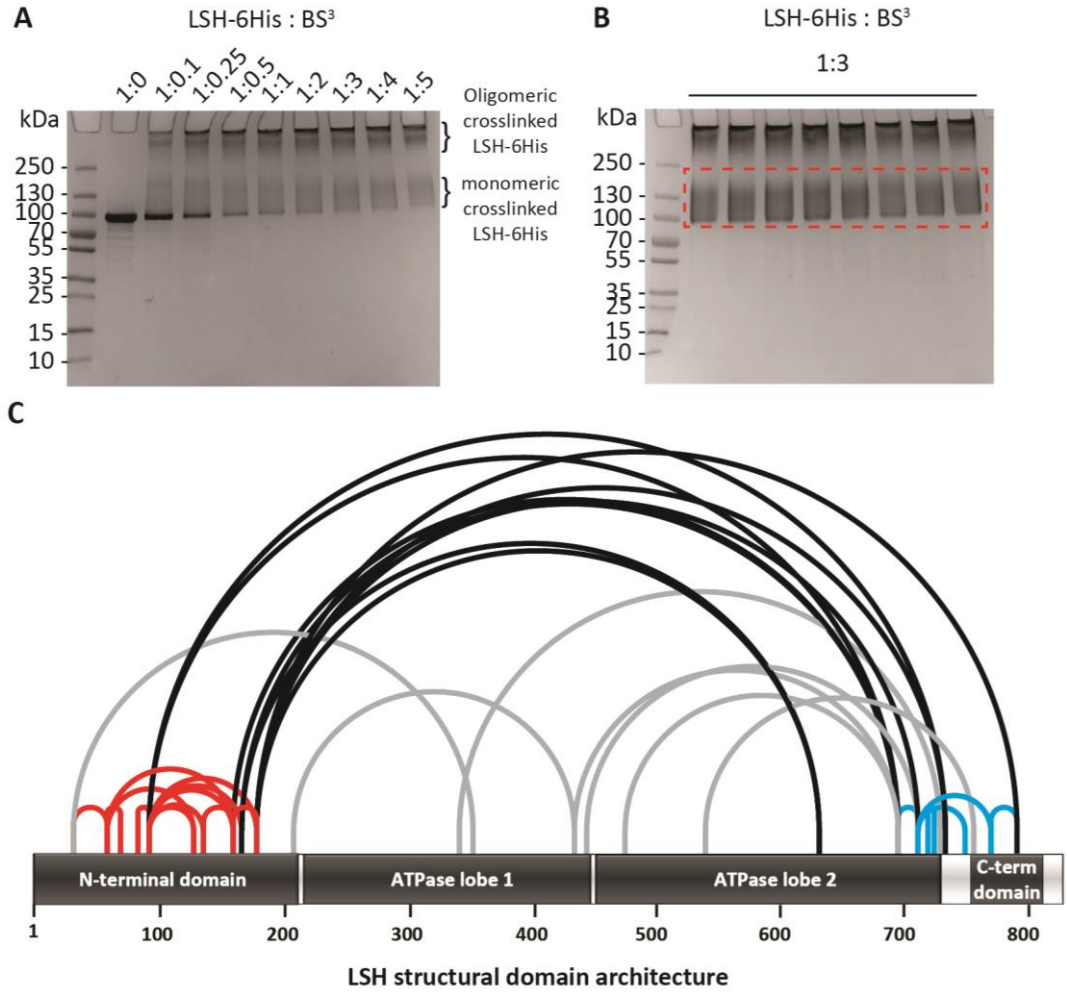

Figure 5.1| BS³ intramolecular cross-linking of LSH-6His

A| Coomassie stained SDS-PAGE gel of a titration series of 2µg LSH-6His incubated with increasing w/w ratios of BS³.

B| Coomassie stained SDS-PAGE gel of 10 µg LSH-6His incubated with 30 µg BS³ in each lane. Monomeric cross-linked LSH species were cut out and processed for mass spectrometry.

C| Intramolecular Lysine-Lysine cross-links of monomeric cross-linked LSH-6His represented on the LSH structural domain architecture. 12 red links indicate loop links or crosslinks in the N-terminal domain. 10 black links indicate crosslinks between the N-terminal domain and ATPase lobe 2 - C-term bridge. 5 cyan links indicates loop links or cross-links in ATPase lobe 2 and the C-term bridge. 7 grey links indicate crosslinks along LSH-6His.

5.2.2| Negative staining EM of LSH-6His

To analyse the general shape of LSH-6His, negative-staining electron microscopy (negative staining EM) was used in collaboration with Dr Laura Spagnolo and Dr Giuseppe Cannone.

Negative staining EM involves placing the protein solution (LSH-6His) onto a charged carbon copper to fix proteins molecules to the surface. Fixed LSH-6His particles are then embedded in a layer of heavy metal staining solution (uranyl acetate) which is dried by gentle blotting and evaporation. This was performed by Dr Giuseppe Cannone.

Negative staining generates high contrast during imaging with electron microscopy and is useful for studying proteins and protein complexes (Ohi *et al.*, 2004; Booth, Avila-Sakar and Cheng, 2011; Rames, Yu and Ren, 2014). The resolution of negative staining EM is limited to ~20 Å due to the microcrystals of heavy metal stain formed during embedding of the sample. Negative staining EM is typically performed with proteins larger than 100 kDa, however, proteins smaller than 50 kDa have been visualised, with a clear contrast to distinguish protein domains (Ohi *et al.*, 2004; Rames, Yu and Ren, 2014).

Individual particles of LSH-6His with minimal aggregation were visualised clearly with uranyl acetate staining (Figure 5.2A). Due to signal to noise ratio being low on certain micrographs, an automated particle picking was not performed due to a risk of selecting false positive LSH-6His particles. Therefore, 10 049 particles were selected manually by myself and 2D class averaging of the particles was performed by Dr Giuseppe Cannone by selecting for 10 2D classes based on the view and shape of the particle using statistical analysis with EMAN 2.1 software (Tang *et al.*, 2007) (Figure 5.2B). The 2D class averages show LSH-6His has a tri-lobal architecture with an elongated arc shape, with a length of ~140 Å and a maximum width of ~90 Å (Figure 5.2B).

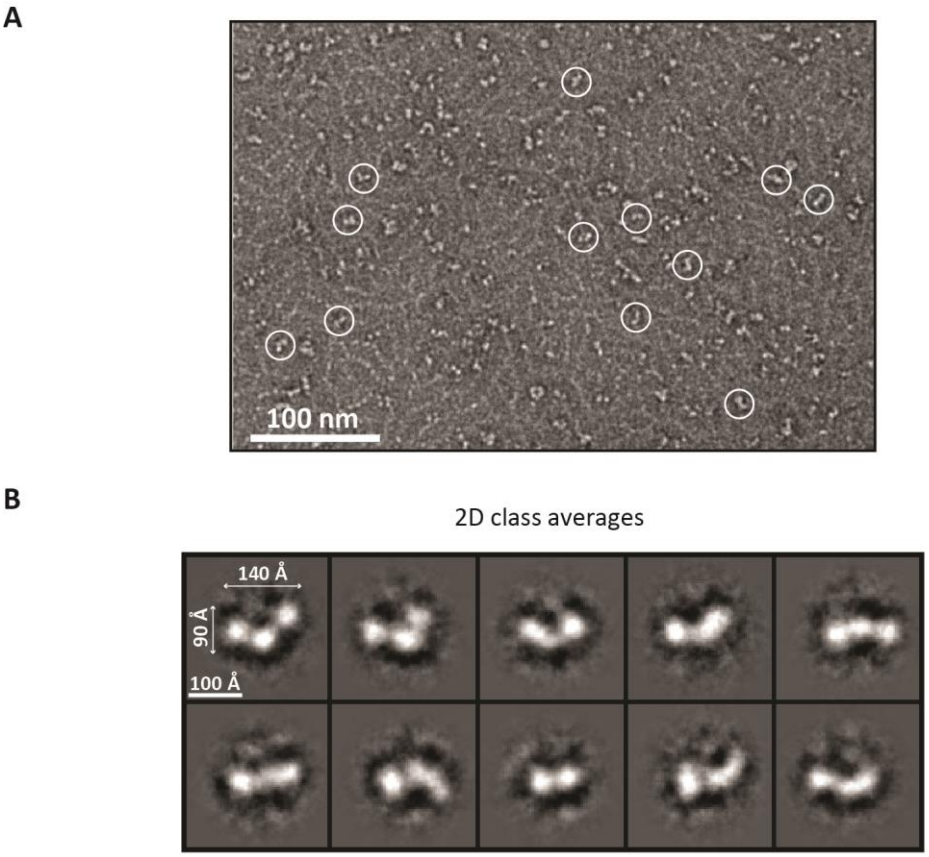

Figure 5.2| Negative staining EM micrograph shows a trilobal and elongated structure of LSH-6His

A| A micrograph of LSH-6His negative staining EM with a 100 nm scale bar. Particles circled in white are individual monomeric LSH particles.

B| 2D class averages of 10049 LSH-6His particles into 10 classes with a 100 Å scale bar.

Generating a 3D map from the 10049 particles was possible however it was not tried firstly because 3D maps calculated with this method alone often shown deformations and distortions in the EM map due to particle flattening and incomplete stain embedding (Ohi *et al.*, 2004). Therefore, a greater number of particles would be required and a random tilt method is used to image particles at a greater number of angles and therefore views for greater statistical significance (Ohi *et al.*, 2004). Given there is no known high-resolution structure available for LSH it is impossible to assign symmetry to its 3D reconstruction or even which domain to fit into the space. If a shape for LSH^{1-176} could be measured or calculated it could be fit to the full-length protein, however, as the protein is 20.5 kDa as a monomer it is too small to be used in negative staining EM to gain any noticeable structural information.

Therefore the complementary structural biology method; small angle X-ray scattering (SAXS) was used to obtain independent confirmation of the LSH-6His structure from the negative staining EM 2D class averages. SAXS would also be used to model a 3D structure of LSH-6His and LSH^{1-176} in solution.

5.2.3| Small Angle X-ray Scattering of LSH-6His

SAXS is a complementary structural method to negative staining EM, X-ray crystallography, NMR and XL-MS. SAXS gives the molecular dimensions, the radius of gyration (R_g) and is a method to accurately establish the oligomerisation state of a protein. SAXS can provide a low resolution (typically 10-50 Å) solution structure of biomolecules of sizes in the range of several kDa to several MDa, which cannot be achieved using NMR or EM (Jacques & Trewhella 2010; Svergun & Koch 2002; Skou et al. 2014). SAXS measures the solution structure of a protein, which is physiological and provides a more accurate representation of the protein structure compared to negative staining EM where proteins are orientated on a grid or X-ray crystallography where proteins are in a tightly packed crystal lattice. SAXS can also be used to measure conformational changes with complex formation and movement of flexible domains (Rambo and Tainer, 2011; Ando *et al.*, 2012; Skou, Gillilan and Ando, 2014).

A general setup of a SAXS experiment is illustrated and explained in Figure 5.3.

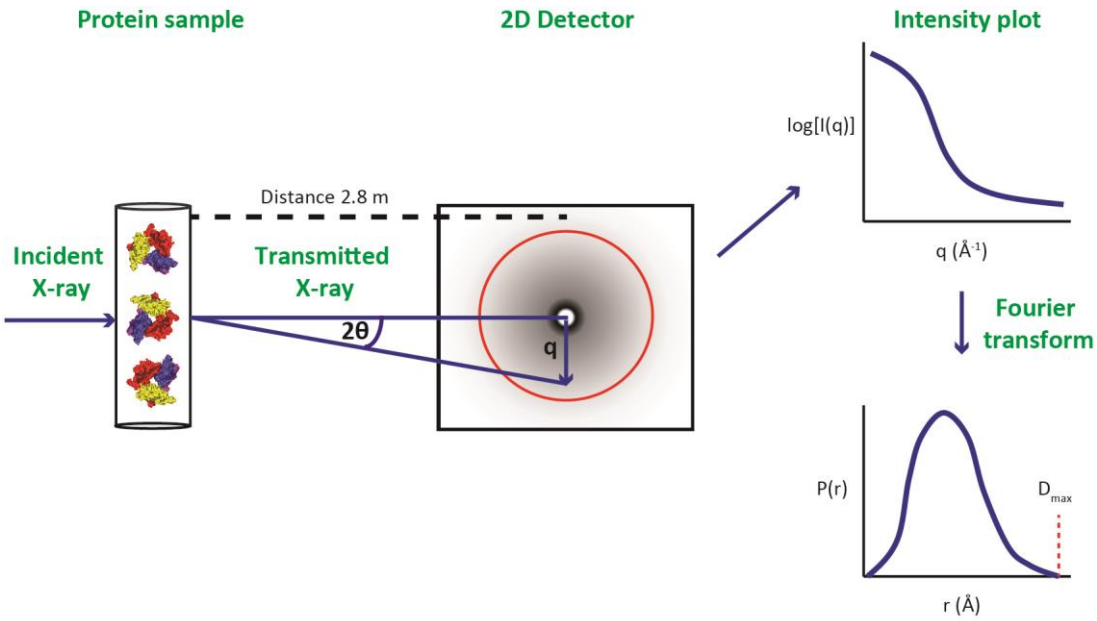

Figure 5.3| Schematic of SAXS method
High energy monochromatic X-rays (wavelength - λ = 0.1 nm) are directed at a protein sample in buffer flowing through a capillary at concentrations ranging from 0.5-100 mg/ml (protein dependent). X-ray scattering at a low angle (2θ - angle between the incident and scatter radiation) is detected on a 2D detector. An isotropic intensity distribution is a result of random orientations and positions of the proteins flowing through the capillary tube. For monodisperse samples this is equal to the scattering from a single particle averaged in all orientations.
The scattered intensity I(q) is recorded as a function of momentum transfer q ($q = 4\pi\sin\theta/\lambda$). A scattering intensity is shown as a 1D function of q by integration of the images. I(q) versus q is essentially the intensity as a function of scattering angle because θ is small (generally < 3˚). A Fourier transform of the I(q) versus q plot is used to interpret the scattering profile in terms of protein structure and provides the interatomic distance distribution function, P(r). P(r) can be used to determine the probable frequency of interatomic vector lengths (r) within a protein. The longest interatomic vector length D_{max} can be measured from P(r) distribution, which provides information about how long the protein is. Adapted from (Jacques & Trewhella 2010, Svergun & Koch 2002).

The intensity plot, I(q) versus q and the Fourier transform of the intensity plot (an interatomic distance distribution (P(r)) can be used to calculate several pieces of information about the sample including: molecular weight (Mw), disorder and flexibility, radius of gyration (R_g) and maximum interatomic distance (D_{max}) (Jacques & Trewhella 2010, Mertens & Svergun 2010).

From high-quality scattering data, the zero angle scattered intensity I(0) and R_g can be calculated and provide the molecular mass and shape information of the protein respectively (Jacques & Trewhella 2010, Mertens & Svergun 2010). The beam stop prevents the zero angle (2θ = 0°) from being measured, therefore I(0) is an extrapolation of the radiation intensity scattered through the zero angle. I(0) is related to the particle volume (V) and contrast (Δρ) which is the mean difference of the scattering density between the particle and the solvent (Jacques & Trewhella 2010, Mertens & Svergun 2010).

R_g is the average of square centre-of-mass distances in the molecule weighted by the scattering length density and measures the overall size of the molecule (Jacques & Trewhella 2010, Mertens & Svergun 2010).

I(0) and R_g can be estimated quickly using Guinier analysis at small values of q:

$$I(q) = I(0)e^{\frac{-q^2R_g^2}{3}} \qquad \text{Equation (5.1)}$$

Therefore I(0) and R_g can be calculated from the y-intercept and slope respectively from a linear fit of ln[(I)q] versus q^2 (Guinier 1938). This relationship is true only when $qR_g < 1.3$ Å; this limitation ensures the estimated parameters are within 10% of the true value (Jacques and Trewhella, 2010). Deviations from a linear plot indicate interparticle interference (downturn at low q) or aggregation (upturn at low q) (Jacques & Trewhella 2010, Mertens & Svergun 2010).

A more accurate I(0) and R_g can be calculated by using the entire scattering profile with the P(r) distribution function, where R_g is the second moment of P(r):

$$R_g^2 = \frac{\int P(r)r^2dr}{2\int P(r)dr} \qquad \text{Equation (5.2)}$$

I(0) is the zeroth moment of P(r):

$$I(0) = 4\pi \int_0^{D_{max}} P(r)dr \qquad \text{Equation (5.3)}$$

P(r) distribution can also provide the D_{max} as explained in Figure 5.3.

The relationship between I(0) and the particles volume through extrapolation can provide a Mw for the molecule (Jacques and Trewhella, 2010).

$$M_w = \frac{I(0).N_A}{C(\Delta p.V)^2}$$ Equation (5.4)

Using a scattering standard such as BSA to calibrate I(0) enables the molecular mass and volume of the scattering particle to be calculated when an accurate concentration is known.

If the inside of a particle has uniform electron density, a hydrated particle volume (Vp) can be estimated on a relative scale, which avoids an accurate concentration measurement being required, and therefore errors in concentration will not influence the estimated Mw. Vp is estimated using Porod's equation (Porod, 1982):

$$V_p = (2\pi\text{^}2\ I(0))/Q$$ Equation (5.5)

Where Q is the Porod invariant: $Q = (\int_0^\infty q^2\ I(q).dq$

The electron density is not uniform for macromolecules; however, a reasonable estimate can be calculated for macromolecules larger than 30 kDa by subtracting a constant from scattering data generating an approximation to the scattering of the corresponding homogeneous body (Mertens and Svergun, 2010). For a globular protein, the Vp (in nm^3) can be divided by 2 or 1.5 to give an estimated range the molecular mass (in kDa) of the particle (Mertens and Svergun, 2010).

SAXS of LSH-6His and LSH^{K237Q}-6His were performed to identify if there were any structural differences between the inactive ATPase mutant and wild-type LSH and to corroborate the EM results of a tri-lobal architecture.

The intensity plot of LSH-6His at 5.6mg/ml has no sharp down scattering close to 0 indicating there was no aggregation of LSH-6His at this concentration (Figure 5.4A), consistent with DLS data (Chapter 4, section 4.2.2). A linear Guinier plot with negligible upturn or downturn confirms the sample is free of aggregation (Figure 5.4B). The P(r) distribution was calculated using scattering data from 0-0.37 $Å^{-1}$ as scattering data from q= 0.37-0.5 $Å^{-1}$ had a too high signal-to-noise ratio to fit a D_{max} accurately (Figure 5.4C). The P(r) distribution shows multiple shoulders after 80 Å until 138 Å when the D_{max} is reached (Figure 5.4C) indicative of a multidomained protein (Putnam *et al.*, 2007). Back calculating a scattering curve from the modelled P(r) distribution fits well to the actual scattering data (Figure 5.4D).

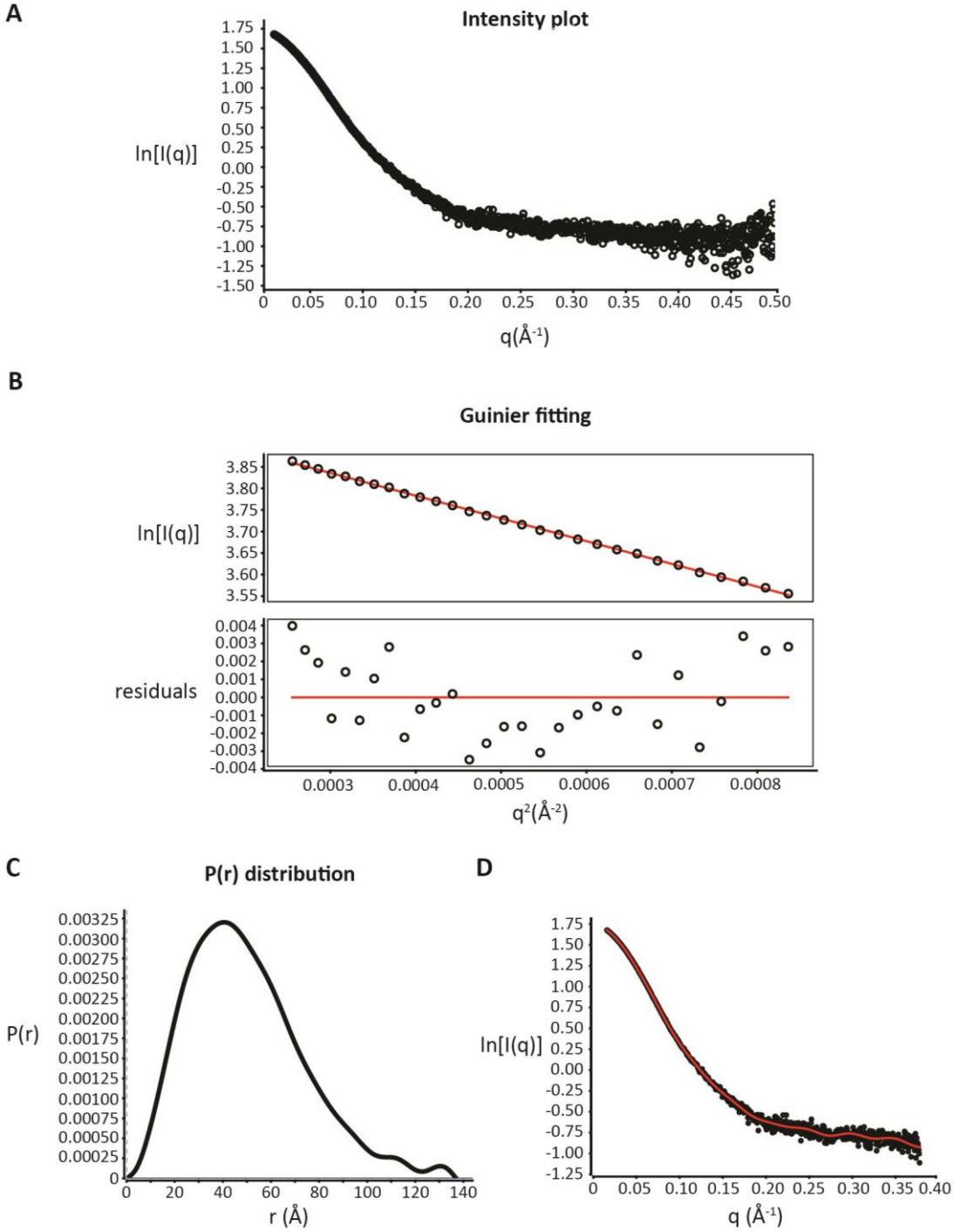

Figure 5.4| SAXS scattering of LSH-6His

A| Intensity plot of LSH-6His at 5.6 mg/ml.

B| Guinier fitting of the intensity plot from **A|**. The q x R_g limits range from 0.6286 to 1.1456 (residuals 28-56). Data fits along a line of best fit without significant upturn. A Rg of 39.5 Å was calculated.

C| P(r) distribution of LSH-6His of data up to a q(Å^{-1}) of 0.37. The Rg is 39.8 Å and the D_{max} is 138 Å.

D| Back calculated P(r) distribution to model reciprocal space (red line) overlaid to the reciprocal space (black spheres) up to a q(Å^{-1}) of 0.37.

SAXS plots of LSH-6His at a series of concentrations were measured. If the oligomerisation state does not change across a range of protein concentrations, the I(0) and R_g should remain constant; an increase in both parameters as protein concentration increases indicates aggregation or oligomerisation. Conversely a decrease in both parameters as concentration increases indicates interparticle interference.

The SAXS parameters Mw, R_g, D_{max} and Porod volume of LSH-6His are shown in table 5.1. The I(0) does not change significantly with concentration (5.9 µM – 58.3 µM) with a difference of 1.78 kDa between the smallest (53.56 kDa – 5.9 µM) and largest values (55.34 kDa – 28.3 µM), indicating there is no concentration dependent aggregation or oligomerisation (Table 5.1). However, the mass calculated is less than that for a LSH-6His monomer (95.8 kDa), indicating the concentration measured by UV spectroscopy was an overestimation. The measurement may have been influenced by oxidation of DTT in the buffer (Jacques and Trewhella, 2010).

The problem encountered in determining the protein concentration for Mw determination by SAXS could be resolved by dialysis to remove the oxidised DTT, or measuring the protein concentration through peptide bond absorbance at 216 nm (this would require buffer exchange to an alternative buffer such as NaP pH 7.5 or Tris pH 7.5 as HEPES absorbs strongly at 216 nm).

The molecular mass calculated from the Porod volume for all concentrations tested (5.9 µM – 58.3 µM) was between 83 – 120 kDa which is in the range of LSH-6His being a monomer in solution (Table 5.1). This result is in agreement with measurements from SEC-MALS (Chapter 4, section 4.2.3). The R_g values calculated either by Guinier fitting or from the P(r) distribution are in agreement (Table 5.1). The R_g decreases slightly as LSH-6His concentration decreases, however, the error between the lowest (37.7 Å) and highest (39.8 Å) measurements is ± 0.6 Å. This indicates there is negligible interparticle interference of LSH-6His particles at different concentrations. The D_{max} ranges from 130.1 – 138.6 Å and is close to the maximum length of LSH-6His of ~140 Å measured from the 2D class averages by negative staining EM (Figure 5.2).

Table 5.1| SAXS parameters of LSH-6His along a concentration series

LSH-6His concentration	I(0) Mw (kDa)	MM (kDa) estimated from Porod volume	Guinier fitting R_g (Å)	P(r) R_g (Å)	D_{max} (Å)
5.6 mg/ml (58.3 µM)	54.12 ± 0.07	90-120	39.6 ± 1.9	39.8	138.6
2.72 mg/ml (28.3 µM)	54.91 ± 0.11	88-117	38.8 ± 1.8	38.6	130.1
1.24 mg/ml (12.9 µM)	55.34 ± 0.11	86-114	38.6 ± 1.6	38.9	133.7
0.57 mg/ml (5.9 µM)	53.56 ± 0.11	83-111	37.8 ± 1.3	37.7	132.1
Average	54.47	-	38.7	38.8	133.6
Standard deviation	0.69	-	0.6	0.8	3.1

SAXS measurements can also be used to establish the flexibility of the protein. A Kratky plot of $q^2I(q)$ against q is used to qualitatively distinguish disordered and flexible proteins from globular compact proteins. The scattering intensity of a globular protein decays at high angles as $I(q) \sim Iq^4$ producing a bell shaped curve (in a Kratky plot) with a well-defined maximum, which falls close to a $q^2I(q)$ of 0 (Receveur-Brechot and Durand, 2012; Kikhney and Svergun, 2015). However, a Gaussian coil (unstructured protein) plateaus at large q values because $1/q^2$ is asymptotic of I(q). To compare the folded state and flexibility of different proteins the data is normalised so that I(0)=1 and q is multiplied by R_g creating a dimensionless Kratky plot (Receveur-Brechot and Durand, 2012; Kikhney and Svergun, 2015).

The dimensionless Kratky plot of globular proteins shows a maximum of 1.104 for a qR_g of $\sqrt{3}$ = 1.73, independent of protein concentration, composition or size. For an unstructured protein, the curve rises with increasing angle and plateaus at a maximum between 1.5-2 with another increase observed over the qR_g range 2-3 $Å^{-1}$ if the protein is disordered and extended (Receveur-Brechot and Durand, 2012). Partly disordered proteins have intermediate values between a globular and unfolded protein (Receveur-Brechot and Durand, 2012).

A dimensionless Kratky plot of LSH-6His indicates the protein is asymmetric because the peak maxima is 1.25 at a qR_g of 2.05 instead of 1.1 at a qR_g of 1.73 seen for a globular protein (Figure 5.5).

The Kratky plot of LSH-6His follows the shape of a bell curve often seen with globular proteins (Figure 5.5) (Receveur-Brechot and Durand, 2012; Kikhney and Svergun, 2015). Instead of going back to 0 the peak slopes off, reaching a minima of 0.28 at a qR_g of 8, with a continual raising slope after this (Figure 5.5). This indicates LSH has an asymmetric shape with flexible regions, which is common with a multidomain protein (Receveur-Brechot and Durand, 2012; Kikhney and Svergun, 2015).

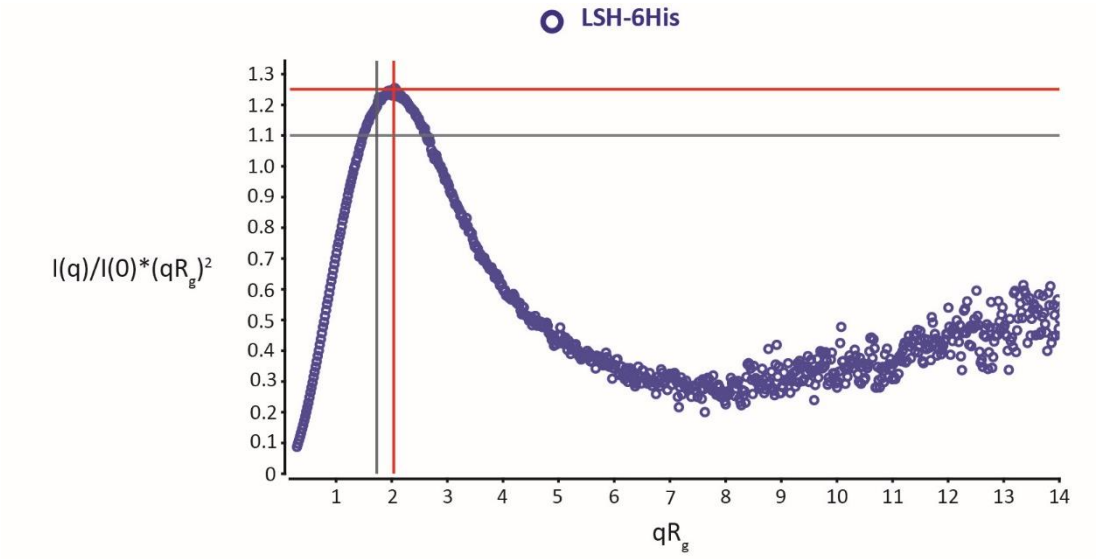

Figure 5.5|Dimensionless Kratky plot of LSH-6His SAXS scattering at 5.6 mg/ml.
The dimensionless Kratky plot is based from the R_g generated from Guinier analysis. The grey line indicates the maxima at which a globular protein should lie (1.1 at a qR_g of 1.73). LSH-6His has a maxima at (1.25 at a qR_g of 2.05).

LSH-6His flexibility was not influenced by the protein concentration, with negligible difference in normalised Kratky plots between different LSH concentrations (data not shown). LSH^{K237Q}-6His SAXS measurements were similar to LSH-6His (data not shown).

The addition of ADP to LSH-6His increased the thermal stability (Chapter 4, section 4.2.1). To investigate if this was due to structural changes, SAXS with ADP was performed. ADP was added to LSH-6His (2.5 mg/ml or 26 µM) to final concentrations of 1, 2, 5 or 10 mM. The addition of ADP at concentrations ranging from 1-10 mM did not change I(0), molecular mass estimated from Porod volume, R_g or D_{max} from values measured for LSH without ADP with

any significance (Table 5.2). This indicates the addition of ADP did not cause aggregation, interparticle interference or oligomerisation of LSH-6His. Similar results were obtained for LSH^{K237Q}-6His (data not shown).

Table 5.2| The effect of varying concentrations of ADP on the SAXS parameters of LSH-6His at 2.5 mg/ml.

ADP concentration	I(0) Mw (kDa)	MM (kDa) estimated from Porod volume	Guinier fitting R_g (Å)	P(r) R_g (Å)	D_{max} (Å)
0 mM	57.86 ± 0.08	91 – 121	41.5 ± 1.1	41.3	145.2
1 mM	58.46 ± 0.08	90 – 120	40.8 ± 0.9	41.0	143.0
2 mM	59.67 ± 0.08	87 – 116	41.3 ± 1.3	41.0	144.6
5 mM	58.43 ± 0.07	89 – 118	39.9 ± 1.0	40.4	139.6
10 mM	57.41 ± 0.08	89 – 118	40.4 ± 0.8	40.7	141.4
Average	58.37	-	40.8	40.9	142.8
Standard deviation	0.85	-	0.7	0.3	2.3

Although there were no changes in parameters, ADP binding might have affected the compactness of LSH by reducing the flexibility of the protein. To measure this effect a dimensionless Kratky plot of LSH-6His with or without 10 mM ADP was made (Figure 5.6). The Kratky plots overlap almost completely with negligible differences around the peak maxima (Figure 5.6). This indicates ADP binding did not cause any significant structural changes by influencing protein compactness or flexibility. All concentrations of ADP tested showed overlapping Kratky plots to LSH-6His (data not shown). As the ATPase lobes require DNA for allosteric movement into a 'closed state' for ATPase activity, the structural changes might be local to the ATP binding pocket and therefore not cause major structural changes, which SAXS does not have the resolution to measure for.

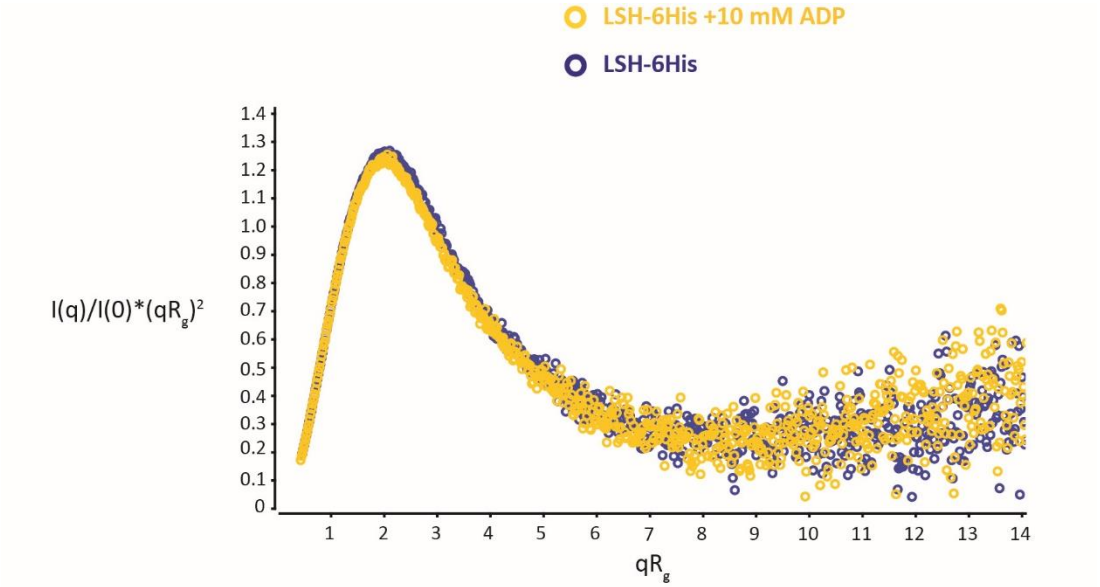

Figure 5.6|Dimensionless Kratky plot of LSH-6His at 2.5 mg/ml with or without 10 mM ADP
The dimensionless Kratky plot is based from the Rg generated from Guinier analysis. LSH-6His and LSH-6His + 10 mM ADP Kratky plots overlap indicating addition of ligand does not alter the compactness or flexibility of LSH-6His.

5.2.4| SAXS of LSH$^{1-176}$

SAXS of LSH$^{1-176}$ was performed with the aim of determining its structural characteristics in solution and using the data for modelling into a SAXS envelope of LSH-6His, defining where the N-terminal domain fits in relation to full-length LSH. The intensity plot of LSH$^{1-176}$ at 10.01mg/ml has no sharp down scattering close to 0 indicating there was no aggregation of LSH-6His (Figure 5.7A). A linear Guinier plot with negligible upturn or downturn confirms the sample was free of aggregation (Figure 5.7B). The P(r) distribution was fit using scattering data from 0-0.37 q(Å^{-1}) as scattering data from 0.37-0.5 had a signal-to-noise too high to fit a D_{max} accurately (Figure 5.4C). The P(r) distribution shows a broad shoulder from 60 Å until 130 Å, followed by a shallow slope to the D_{max} of 168 Å (Figure 5.7C). This is observed with unstructured proteins which typically have long sloping shoulders until the D_{max} (Putnam *et al.*, 2007). Back calculating a scattering curve from the modelled P(r) distribution fits well to the actual scattering data (Figure 5.7D).

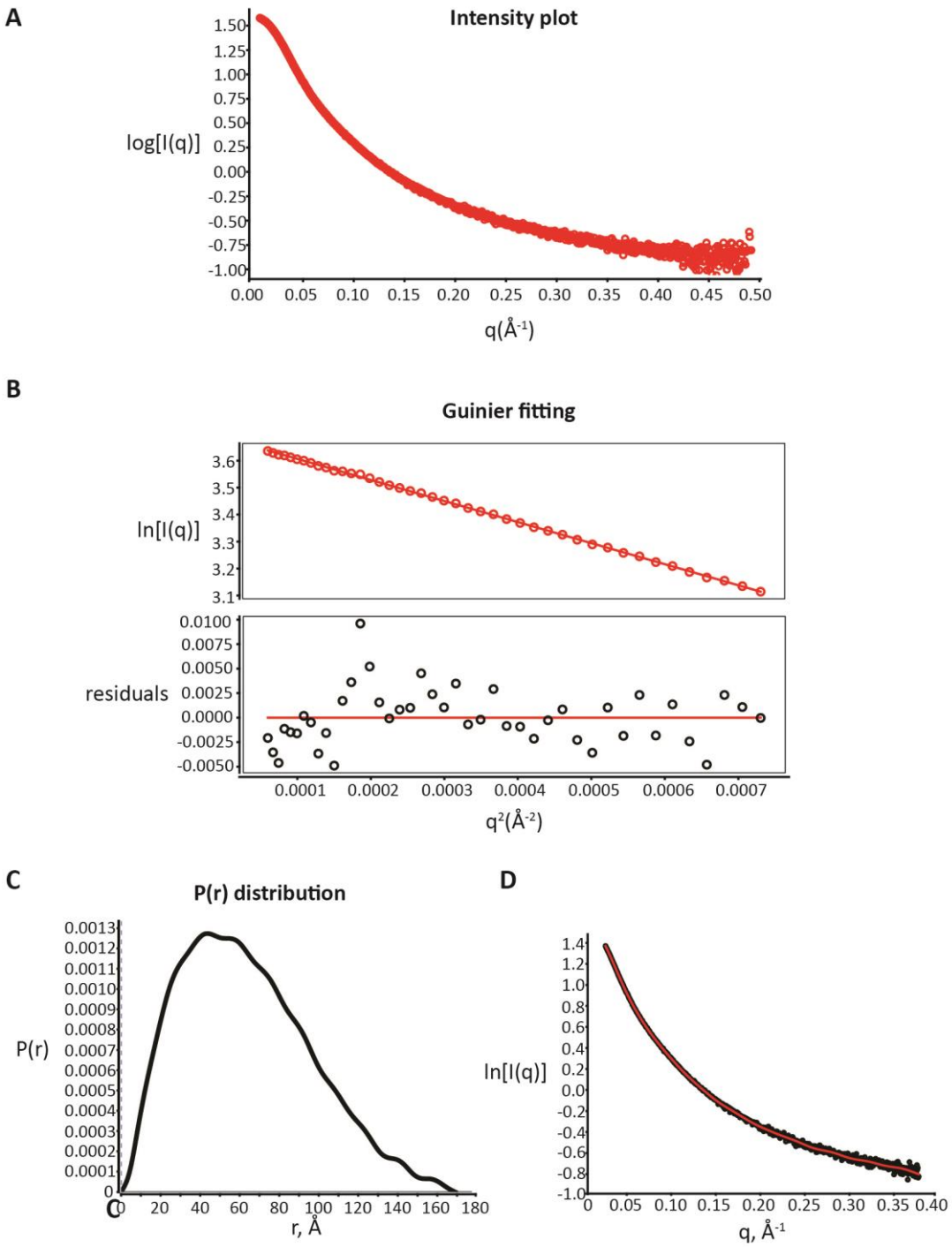

Figure 5.7| SAXS scattering of LSH$^{1-176}$.

A| Intensity plot of LSH$^{1-176}$ at 10.01 mg/ml

B| Guinier fitting of the intensity plot from A|. The q x R_g limits range from 0.3567 to 1.3032 (residuals 10-52). Data fits along a line of best fit without significant upturn or downturn. A Rg of 48.1 Å was calculated.

C| P(r) distribution of LSH$^{1-176}$ of data up to a q(Å^{-1}) of 0.37. The Rg is 49.8 Å and the D_{max} is 168 Å.

D| Back calculated P(r) distribution to model reciprocal space (red line) overlaid to the reciprocal space (black spheres) up to a q(Å^{-1}) of 0.37.

As with LSH-6His, SAXS of LSH^{1-176} were measured over a range of concentrations (from 0.63 mg/ml to 10.01 mg/ml) (Table 5.3).

Based on sequence the monomeric mass of LSH^{1-176} is 20.5 kDa. Guinier analysis of SAXS data indicate from I(0)) that LSH^{1-176} is a monomer at 0.63 mg/ml, and a dimer at 5.18 mg/ml. The average molecular mass calculated from I(0) for samples at 1.24 mg/ml and 2.37 mg/ml is 27.32 and 34.17 kDa respectively, intermediate between a monomer and dimer. This indicates LSH^{1-176} is in equilibrium between a monomer and dimer from 0.63 mg/ml until 5.18 mg/ml, forming predominantly monomers at the lowest concentration measured and dimers at the highest concentration. However, if the concentrations measured are inaccurate and overestimated as was done for LSH-6His (Table 5.1 and Table 5.2) then this indicates LSH^{1-176} is going from a dimer to a hexamer as measured by the porod volume rather than the monomer to dimer measured by I(0). At 0.63 mg/ml LSH^{1-176} is a monomer (23.31 kDa) according to I(0) and a dimer (38-50 kDa) according to the MM from the Porod volume. However at 10.01 mg/ml LSH^{1-176} is a dimer (39.68 kDa) according to I(0) and a hexamer (113-151 kDa) according to the MM calculated from the porod volume. Given SEC-MALS indicates LSH^{1-176} is in a monomer-dimer equilibrium, at concentrations ranging from 1.25-10 mg/ml (Chapter 4, section 4.2.13), it provides confidence the monomer-dimer equilibrium calculated from I(0) is correct.

The R_g values calculated by Guinier analysis range from 46.1 Å to 51.5 Å, but these differences are within the error of the measurement. There is an increase in R_g at 2.37 mg/ml and 5.18 mg/ml, however, the increases are within a 5 % error, indicating negligible differences in R_g data. The R_g values calculated from the P(r) distribution is higher than the Guinier fitting data for all concentrations, indicating the Guinier fitting is underestimating the R_g, or the P(r) is overestimating the Rg or the calibration is off. However, it is within the 5% error for all concentrations of LSH^{1-176} indicating the differences are negligible. This indicates LSH^{1-176} dimer formation is not compacting the protein complex any more than a LSH^{1-176} monomer. The D_{max} ranges from 161.3 Å to 180.1 Å, which is within error (± 4%).

Table 5.3|SAXS parameters of LSH$^{1-176}$ along a concentration series

LSH$^{1-176}$ concentration	**I(0) Mw kDa**	**MM (kDa) estimated from Porod volume**	**Guinier fitting R_g (Å)**	**P(r) R_g (Å)**	**D_{max} (Å)**
10.01 mg/ml (487.9 µM)	39.68 ± 0.03	113 – 151	48.1 ± 0.5	49.8	168.4
5.18 mg/ml (252.7 µM)	38.37 ± 0.05	116 – 154	50.5 ± 0.5	52.3	176.8
2.37 mg/ml (115.6 µM)	34.17 ± 0.06	63 – 84	51.5 ± 1.7	52.4	180.1
1.24 mg/ml (60.5 µM)	27.32 ± 0.18	50 – 66	47.7 ± 0.7	49.4	165.5
0.63 mg/ml (30.7 µM)	23.31 ± 0.28	38 - 50	46.1 ± 1.0	47.2	161.3
Average	32.57	-	48.8	50.2	170.4
Standard deviation	6.33	-	2.0	2.0	7.0 (± 4%)

A dimensionless Kratky plot of LSH$^{1-176}$ at different concentrations was calculated (Figure 5.8). LSH$^{1-176}$ at 10.01 mg/ml shows a peak maxima of 1.24 at a qR_g of 2.4, followed by a continuous plateau (Figure 5.8). This indicates the protein is disordered but not completely unstructured as the peak maxima would be between 1.5-2 instead of 1.24. However, as protein concentration decreases there is no peak maxima with a rising plateau increasing to a maximum between 1.5-2 indicating LSH$^{1-176}$ is unstructured as a monomer and becomes slightly more structured when it dimerizes (Figure 5.8).

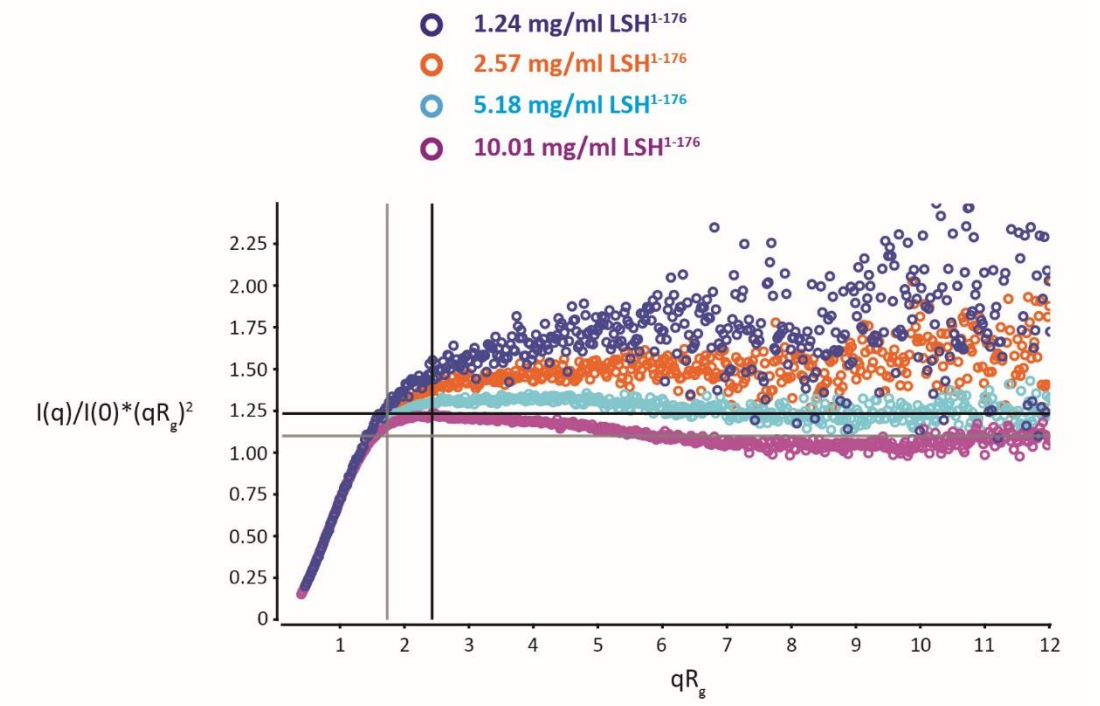

Figure 5.8|Dimensionless Kratky plot of LSH$^{1-176}$ shows the protein becomes more structured and less flexible as protein concentration increases.
Kratky plot is based from the R_g generated from Guinier analysis. LSH$^{1-176}$ at a concentration of 10 mg/ml has a maxima at 1.25 at a qRg of 2.4 (black lines). The grey intersecting lines are where the maxima for a globular protein would peak. LSH$^{1-176}$ concentrations at 2.57 mg/ml or less have no clear maxima.

The R_g and D_{max} do not change significantly between LSH$^{1-176}$ as a monomer or dimer (Table 5.3) and shows some structural compaction upon dimerization as observed from the dimensionless Kratky analysis (Figure 4.8). This indicates LSH$^{1-176}$ forms a dimer through contact along a surface area which does not drastically compact the protein nor reduce flexibility significantly.

The aim was to use SAXS data of LSH$^{1-176}$ to fit this domain into the SAXS envelope of LSH-6His by modelling. As LSH$^{1-176}$ is a unique portion of LSH with no homologous proteins with a known structure, structure prediction software such as Phyre and I-TASSER cannot use homology modelling to predict a structure. Instead, *ab-initio* modelling based on the predicted secondary structure of the N-terminal domain is used. Using Phyre or I-TASSER to predict a structure of LSH works well for the ATPase region as the LSH sequence is conserved with an 80% homology to the known structures of the chromatin remodelers; RAD54, CHD1 and SNF2.

However, the predicted N-terminal domain structure varies dependent on which prediction software is used as the structure is predicted solely from a predicted secondary structure because of no sequence homology. Phyre represents LSH^{1-176} as an elongated helix with structured regions at each end, whereas I-TASSER predicts LSH^{1-176} as a helical bundle (Figure 5.9). The Phyre model is in closer agreement with the LSH^{1-176} SAXS data as it is not compacted and has a D_{max} of 120 Å closer to the experimental 170 Å (Figure 5.9). Whereas the I-TASSER model is more compact with a D_{max} of 68 Å which is significantly shorter than the SAXS experimental D_{max} value (Figure 5.9).

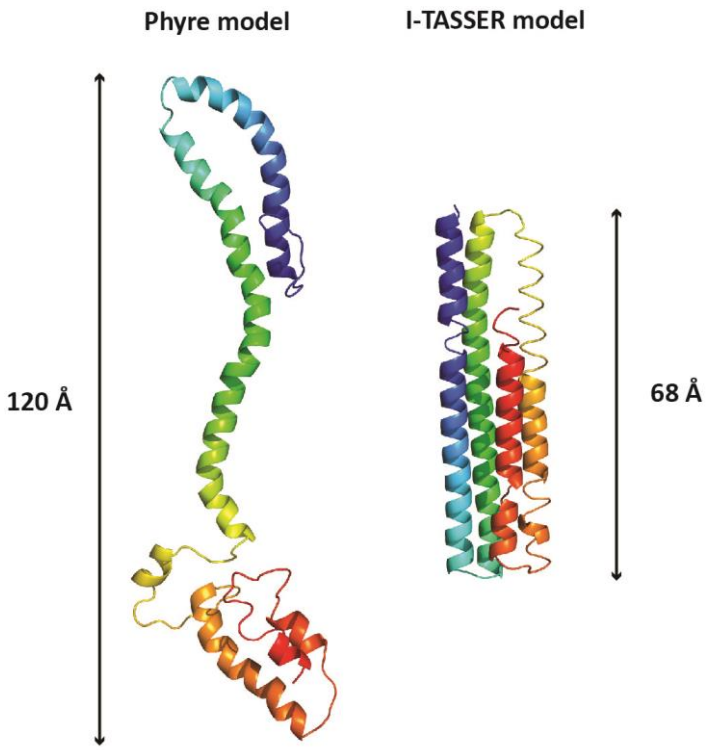

Figure 5.9|*Ab-initio* structure prediction of LSH^{1-176} using Phyre or I-TASSER
Blue (N-terminus) to red (C-terminus).

LSH^{1-176} is in a monomer-dimer equilibrium where the dimer dominates at a concentration at or greater than 5 mg/ml, is disordered and has a D_{max} 30 Å larger than LSH-6His which is a monomer across all concentrations tested and mostly structured. This indicates LSH^{1-176} folds into a more compact and less flexible domain as part of the full-length LSH-6His, which is supported by XL-MS data which shows the N-terminal domain is close to ATPase lobe 2 and the C-terminal domain LSH (Figure 5.1) This makes it impossible to use SAXS data of LSH^{1-176} for fitting to the LSH^{1-176} density of LSH-6His.

Therefore I decided to model a SAXS envelope of LSH-6His using GASBOR and predict a structure of LSH using the SAXSTER program, which uses SAXS data as a major constraint for protein structure, as structure prediction from LSH sequence alone is unreliable in regions with unique sequences such as the N-terminal domain.

5.2.5| A low-resolution structure of LSH-6His

When modelling a structured domain or protein from SAXS the Mw, R_g and D_{max} provide structural information which is essential for *ab initio* modelling a 3D molecular structure. A common *ab initio* bead modelling program is Dummy Atom Model Minimisation (DAMMIN) (Svergun, 1999; Mertens and Svergun, 2010). This modelling method uses densely packed beads (5 Å in diameter) in a sphere which is larger in diameter than the D_{max} of the particles. Shape reconstruction is performed in the sphere of beads by simulated annealing, in which one bead is changed, creating a new model until a final model of compact and connected beads which fit the experimental data is reached (Svergun, 1999; Mertens and Svergun, 2010). Another *ab initio* bead modelling program known as GASBOR which is similar to DAMMIN can be used and has several advantages over DAMMIN. Firstly GASBOR defines each amino acid of the protein as a bead which has its centre 3.8 Å away from the adjacent bead and must anneal as a chain to simulate a polypeptide chain (Putnam et al. 2007, Svergun et al. 2001). Therefore GASBOR can create more realistic models of multidomain proteins as it can factor in linker or hinge regions (Putnam et al. 2007, Svergun et al. 2001). GASBOR also uses more accurate penalties which constrain shape and do not prohibit the generation of anisotropic shapes such as cavities, when the number of residues is known for the protein sample (Putnam et al. 2007, Svergun et al. 2001).

Therefore I used GASBOR to generate an *ab initio* model of LSH-6His from the SAXS data. Given I do not have a crystal structure of LSH and the N-terminal domain of LSH is modelled only from the secondary structure, I decided to use the SAXS assisted protein fold recognition (SAXSTER) software to model a high-resolution protein structure (dos Reis, Aparicio and Zhang, 2011). Firstly the best 10 template structures matching the input protein sequence are selected from the PDB library (dos Reis, Aparicio and Zhang, 2011). The template which fits the SAXS data the best is used as the template for modelling a high-resolution protein structure with SAXS structural parameters as constraints using MODELLER (dos Reis, Aparicio and Zhang, 2011).

A representative GASBOR model of LSH-6His fit the SAXS data with a χ^2 value of 1.22 showing high confidence in the model produced. The GASBOR model itself is elongated in shape, with three regions encompassing an elongated flanking region, a central spherical lobe structure and a smaller lobe structure next to it (Figure 5.9A). This model has a similar shape and tri-lobal features observed from 2D class averages of negative staining EM of LSH-6His (Figure 5.2).

Two high-resolution LSH structures were predicted using I-TASSER or SAXSTER. I-TASSER and SAXSTER predict different secondary structures for the N-terminal domain: the I-TASSER model of this domain is predominantly helical and the SAXSTER predicts coils (Figure 5.9B). Each structure was overlaid manually onto the LSH-6His GASBOR model to help provide an idea of where the domains of LSH might lie (Figure 5.10B). ATPase lobes 1, 2 and the C-terminal domain of the SAXSTER and I-TASSER models appear to fit in the space of the two spherical lobe regions in the GASBOR model (Figure 5.10B). The N-terminal domain of the SAXSTER model is elongated and may fit the SAXS envelope if angled in a different orientation. By contrast, the N-terminal domain of the I-TASSER model is too compact to fill the SAXS envelope (Figure 5.9B). Density in the left area of the spherical central part of the SAXS envelope is unoccupied for both the SAXSTER and I-TASSER models (Figure 5.9B).

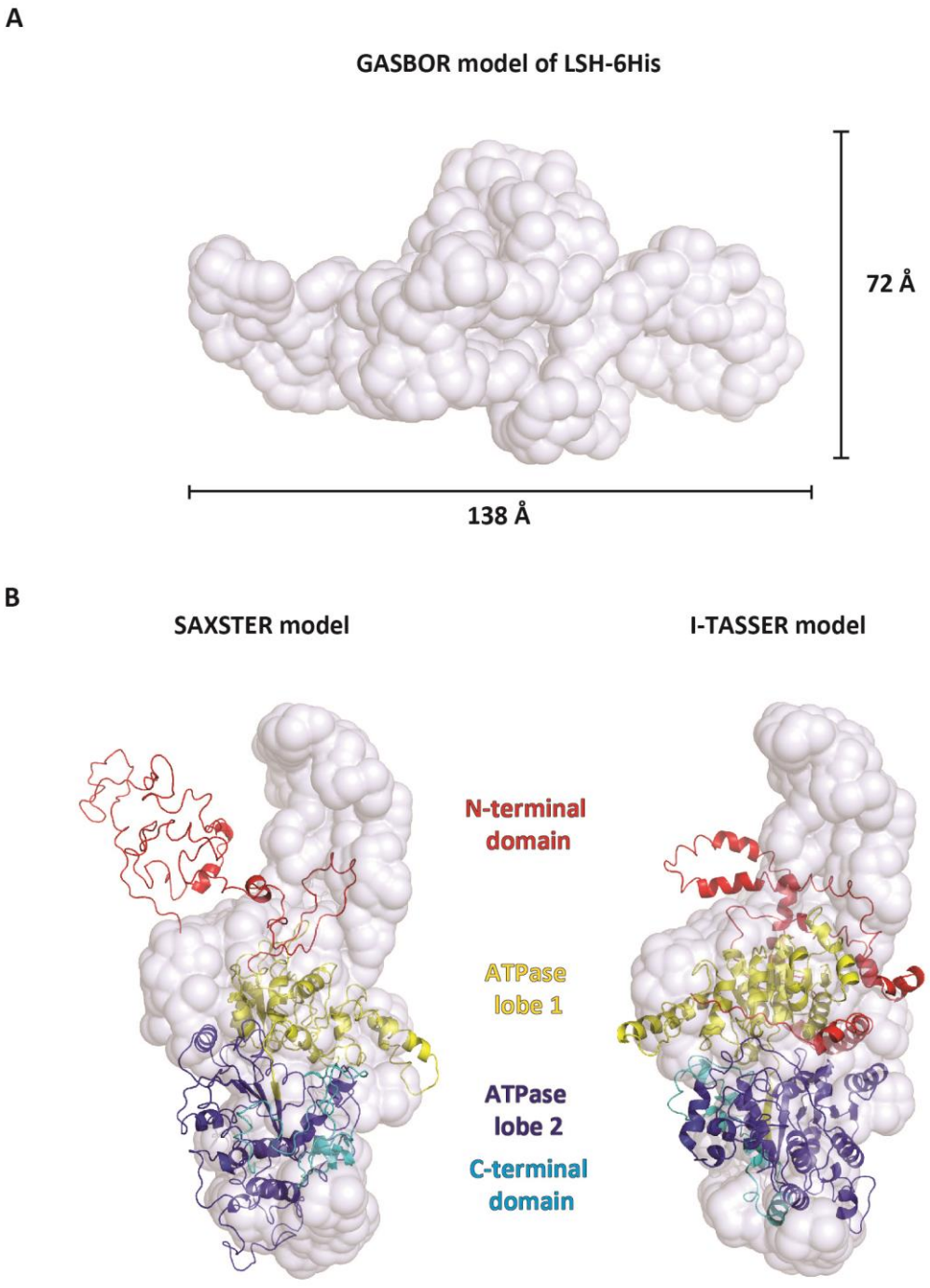

Figure 5.10|SAXS envelope of LSH-6His overlaid with SAXSTER and I-TASSER models of LSH
A| GASBOR model of LSH-6His calculated from SAXS data. The model (left to right) has an elongated region followed by two spherical regions, with the central lobe having the greatest density.
B| One the left the highest scored SAXSTER model of LSH overlaid in the GASBOR model and on the right the highest scored I-TASSER model of LSH overlaid in the GASBOR model. The high resolution models are coloured in functional regions corresponding to the N-terminal domain (red), ATPase lobe 1 (yellow), ATPase lobe 2 (blue) and C-terminal domain (cyan).

Due to the lack of a high-resolution 3D structure of LSH, it is difficult to model accurately where the N-terminal domain of LSH is in relation to the ATPase lobes and how it fits along the SAXS and EM maps. However, given there is extra density from both EM and SAXS maps after fitting both lobes of the LSH ATPase region, it is likely the N-terminal domain occupies this space. However, if this is the case then the C-terminal domain must be within ~30Å of the N-terminal as detected from XL-MS, which neither model has factored in. Therefore, modelling the orientation of domain from these tools alone is currently limited due to software available for modelling and high resolution structural information missing. This will be discussed in more detail later. Further work would need to be performed to determine how the N-terminal domain of LSH interacts with the ATPase region and how it may function to regulate ATPase and chromatin remodelling activity.

5.3| Discussion

The ATPase region of ATP-dependent chromatin remodelers is conserved and is necessary for ATP hydrolysis, DNA binding and translocation (Clapier and Cairns, 2009, Dürr et al., 2006, Flaue and Owen-Hughes, 2011, Yodh, 2013). The ATPase lobes 1 and 2 of the ATP-dependent chromatin remodeler CHD1 are connected by the N-terminal chromodomain and the C-terminal domain, holding the enzyme in an inactive conformation (Figure 1.7) (Hauk *et al.*, 2010). However, the ATPase lobes of the ATP-dependent chromatin remodeler SNF2 are braced together through the C-terminal helix and the disordered region of ATPase lobe 2 (Figure 1.7) (Xia *et al.*, 2016). The N-terminal HSA domain is connected only to ATPase lobe 1 (Figure 1.7) (Xia *et al.*, 2016). Due to the ATPase region flanking domains being different between different families of chromatin remodelers it is difficult to accurately define how ATPase region flanking domains of LSH are structurally and functionally acting on its ATPase region. The C-terminal domain of LSH, which has sequence similarity with the C-terminal domains of CHD1, ISWI and SNF2 (Figure 4.8), may bridge or brace the ATPase lobes as seen in the CHD1 crystal structure (Hauk *et al.*, 2010).

The N-terminal domain of LSH is unique in sequence with functions including nuclear localisation and binding with DNMT3B (Lee et al. 2000, Myant & Stancheva 2008). The N-terminus of LSH (1-200 aa) has a maximum of 5% sequence similarity to other proteins from a BLAST search. Structural characterisation of the N-terminal domain of LSH using NMR and CD has confirmed the prediction that the domain is disordered (Mari Eltermann., 2015). SAXS of LSH^{1-176} also shows it is disordered, elongated and flexible (Figure 5.7 and Figure 5.8). LSH^{1-176} also dimerises at concentrations from 60 µM to 250 µM (Table 5.3). As LSH-6His is monomeric, it is unlikely the dimerisation of LSH^{1-176} observation is physiological in the context of a full-length protein, and interactions of the N-terminal domain with the other domains of LSH block its dimerisation. The ability of the N-terminal domain to dimerise indicates that the domain could bind to other proteins such as DNMT3B, potentially through the use of the predicted coiled-coil (LSH^{14-96}) (Myant and Stancheva, 2008). Given the D_{max} of LSH-6His (~140 Å) is 20-30 Å shorter than that of LSH^{1-176} (D_{max} of 165-180 Å), LSH^{1-176} likely forms contacts with the ATPase region which compact and reduce the flexibility of this domain (Table 5.1 and Figure 5.4). Ideally, a SAXS model of $LSH^{177-821}$ would be needed to determine how LSH^{1-176} may interact with the ATPase region of LSH. However, due to $LSH^{177-821}$ being unstable this could not be performed.

From SAXS and EM data alone I cannot determine whether the N-terminal domain of LSH is connected to ATPase lobe 1 only or ATPase lobes 1 and 2 (Figure 5.2 and Figure 5.9). However, XL-MS data has shown the N-terminal domain LSH^{97-179} is within 22.4 Å contact of $LSH^{698-736}$ indicating the N-terminal domain likely forms contacts with ATPase lobe 1 and is within 22.4 Å from the C-terminal end of ATPase lobe 2 and the C-terminal domain (Figure 5.1). Limited proteolysis with subtilisin cleaves LSH into LSH^{1-176} and $LSH^{178-821}$-6His and then into $LSH^{178-704}$ (Chapter 4, section 4.2.10). Thus, the N-terminal domain may occlude the region around LSH^{704} (C-terminal end of ATPase lobe 2) protecting it from proteolysis, or cleavage of the N-terminal domain causes allosteric changes in the ATPase region making the LSH^{704} region accessible for protease cleavage. Furthermore, the N-terminal domain of LSH is required for stability of the ATPase region ($LSH^{200-821}$) which is unstable and weakly expressed without the N-terminal domain (Chapter 4, section 4.2.9). Therefore, it is likely the N-terminal domain forms contacts along parts of the ATPase region.

An alternative approach to modelling protein structures to fit within the SAXS volume would be to use BUNCH (Petoukhov and Svergun, 2005). *Ab initio* and rigid modelling are used to determine the 3D domain structure of a multi-domained protein based on multiple scattering data sets from deletion mutants when the high resolution structures of the individual domains are known (Petoukhov and Svergun, 2005). Dummy residue chains in place of the domains which do not have a high resolution structure to fit the experimental scattering data (Petoukhov and Svergun, 2005). However, the issue here is there is only a crystal structure of the CHD1 SNF2 ATPase region which is most closely related to LSH which can be used. As there in no high-resolution structure of the N-terminal domain of LSH it may be difficult to gain further structural information of this region with *ab initio* predicted structures only.

Further experiments are required to define how the N-terminal region might influence the action of the ATPase region structurally. Firstly determining spatially where the N and C-terminal domains of LSH are in relation to the EM averages and SAXS envelope can be performed by tagging or coupling the N or C-terminus with GFP or MBP to define the location of the respective domain (Ciferri *et al.*, 2012; Dambacher and Lander, 2015).

2D class averages of ~10000 LSH-6His particles showed a tri-lobal and elongated shape (Figure 5.2). SAXS data could be used as a template for 3D reconstruction of the negative staining EM particles with a, however, a greater number of particles for each 2D class would allow a better 3D model to be generated (greater than 20000 particles). Greater contrast and therefore resolution (down to 15 Å) for smaller proteins has been achieved by using uranyl formate instead of uranyl acetate to stain grids as it produces smaller sized grains (Ohi *et al.*, 2004; Rames, Yu and Ren, 2014). This method could be used for greater resolution of LSH-6His negative staining EM particles.

Revisiting XL-MS for more distance information between domains could be used for adding constraints for modelling the N-terminal domain and the ATPase region. The zero length (0 Å spacer arm) chemical crosslinker 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) activates carboxylic acids (aspartate and glutamate) to crosslink with amines (lysine). EDC is coupled with sulfo-N-hydroxysuccinimide (sulfo-NHS) to improve efficiency and create a stable amine-reactive intermediate. EDC-sulfo NHS can provide a greater number of crosslinks which are 11.4 Å shorter than BS^3. Therefore, comparing the EDC-sulfo NHS crosslinking with BS^3 crosslinking of LSH could be used for distance modelling constraints.

A major limitation of using crosslinking mass spectrometry is it is undemocratic as a method in that specific regions of the protein may provide greater crosslinks than others and this may alter the conformation of the protein potentially inhibiting the movement of flexible regions and fixing them in one conformation biasing the protein structure to one state from a pool of physiological states (Maiolica *et al.*, 2007; Rappsilber, 2011; Fischer, Chen and Rappsilber, 2013).

Combining the complementary structural methods of XL-MS, EM and SAXS could enable to obtain a detailed structure of LSH in different conditions and potentially with substrates such as DNA, nucleosomes or protein binding partners such as DNMT3B. Combining the XL-MS, negative staining EM and SAXS data could provide certain information about the protein structure and domain interactions, however, the tools are not yet developed enough to model a structure using the parameters from the data for all methods combined.

Significant advancements have been made in Cryo-EM technology in the last 5 years enabling the near atomic resolution structures of large protein complexes to be solved (Cheng, 2015; Doerr, 2015). Recently a 3.8 Å structure of a 93 kDa protein was achieved, making even proteins less than the expected cut off 150 kDa achievable using cryo-EM (Merk *et al.*, 2016). As the LSH-6His sample is 96 kDa, greater than 95 % pure and monodisperse, a cryo-EM structure is achievable and would be the best technique to use to obtain a high resolution structure of LSH.

5.4| Conclusion

To summarise I have determined through XL-MS studies of LSH-6His with BS^3 that the N-terminal region (LSH^{97-179}) is within a 22.4 Å distance of the C-terminal region ($LSH^{698-736}$) indicating an interaction of the LSH N-terminal domain along portions of the ATPase region. SAXS of the LSH N-terminal domain (LSH^{1-176}) showed this domain is disordered and in a monomer-dimer equilibrium, indicating this domain is capable of protein binding. The maximum dimensions of LSH^{1-176} are 30 Å more than of LSH-6His, indicating it must be in contact with the ATPase region in the context of full length LSH. Dimensionless Kratky analysis has shown LSH-6His is mostly globular with some flexibility indicating the N-terminal domain is compacted in the full length LSH. Negative staining EM 2D class averaged particles and the SAXS envelope of LSH-6His are in agreement that LSH is a monomer with an elongated trilobal shape with a D_{max} of ~140 Å.

A GASBOR model of LSH-6His fits the ATPase region of a predicted LSH-6His structure in an inactive state, with the N-terminal domain likely occupying the rest of the density. To complement the structural studies of LSH, the next step would be to study how LSH interacts with DNA and nucleosomes and its binding partner DNMT3B using biochemical assays.

Chapter 6| LSH interactions with DNA and nucleosomes

6.1| Introduction

The function of the conserved ATPase region of chromatin remodelers is to bind and translocate dsDNA using the energy from ATP hydrolysis. This translocating action results in the breakage of DNA-nucleosome contacts causing the sliding or unwinding of nucleosomes making previously inaccessible dsDNA regions available for transcription factors and DNA modifying enzymes (Clapier and Cairns, 2009; Ho and Crabtree, 2010).

In a previous study recombinant LSH was shown to bind dsDNA and hydrolyse ATP *in vitro* (Burrage *et al.*, 2012). The aim of this chapter was to further characterise the interaction of recombinant LSH with DNA and nucleosomes *in vitro* using biochemical assays in order to gain further insights into the molecular function of LSH.

6.2| Results

6.2.1| Detecting a LSH-6His:dsDNA interaction

The electrophoretic mobility shift assay (EMSA), also known as a band shift assay, is used to detect protein:nucleic acid interactions. A polyacrylamide or agarose gel is used to separate free DNA and a protein:DNA complex using electrophoresis under non-denaturing conditions (Fried and Crothers, 1981; Garner and Revzin, 1981). A decrease in the mobility of a protein:nucleic acid complex compared to nucleic acid alone is detected by staining for nucleic acid or using radiolabelled or fluorophore-labelled nucleic acid (Fried and Crothers, 1981; Garner and Revzin, 1981; Rye *et al.*, 1993; Hellman and Fried, 2009). The EMSA is a rapid assay (1-3 hrs) and sensitive enough for quantitative analysis of the interaction (Rye *et al.*, 1993; Hellman and Fried, 2009).

Previous experiments, using the systematic evolution of dsDNA ligands by exponential enrichment (SELEX) with a library of random dsDNA sequences, showed that LSH had no preference for a specific DNA sequence (Myant – unpublished data). However, there is a preference for GC-rich sequences (Myant – unpublished data).

EMSA experiments were first carried out on 5% TBE polyacrylamide gels, in an attempt to measure binding of increasing concentrations of LSH-6His to either 12 bp or 56 bp GC-rich dsDNA -labelled at the 5' end with the fluorophore IR700. The 56 nt DNA was a random library of sequences. The 12 nt DNA sequence used was GCTGCGTGCGTT (Chapter 2, section 2.11, table 2.16). Whilst LSH-6His shifted the dsDNA, smeary undefined bands with complexes stuck in the loading wells were consistently visualised (data not shown). Changing the polyacrylamide concentration, between 4-10%, the temperature (6°C or 20°C) and voltage (50V, 75V, or 100V) of electrophoresis, showed no improvement (data not shown).

Therefore an agarose matrix, which has larger pores was used instead (Hellman and Fried, 2009; Stellwagen, 2009). Shifting of a 12 bp dsDNA by LSH-6His was observed on a 2% (w/v) horizontal agarose gel EMSA in TBE (pH 7.5) buffer and there was reduced band smearing compared to the acrylamide EMSA. However, LSH-6His:dsDNA complexes were still present in the wells (data not shown).

The binding buffer was kept physiological (pH 7.25) and would keep the protein positively charged and the DNA negatively charged. A range of buffers with different pH capacity (pH 7.5 to 10.3) was tested in both polyacrylamide and agarose EMSAs (McLellan, 1982) to test an optimal buffer to stabilise a LSH:DNA complex whilst enabling migration into the gel. Tris-CAPS buffer (pH 9.3) in an agarose gel was found to be optimal, as LSH-6His shifted a 12 bp DNA duplex which migrated into the gel (data not shown).

The formation of 12 bp 5'^{IR700} DNA duplex was confirmed as it migrated more slowly than the non-annealed 12 bp 5'^{IR700} ssDNA (Figure 6.1A) on a polyacrylamide gel due to its larger size. LSH-6His shifted a 12 bp DNA duplex, with complexes entering the wells (Figure 6.1B). To confirm that the LSH:dsDNA complex was not an artefact resulting from the C-terminal 6His tag a 6His tagged cytosolic protein, Cyclophilin A (CypA) was used in an EMSA alongside LSH-6His (Figure 6.1B). The lack of any noticeable shift of the dsDNA with CypA, whilst a clear shift with LSH at equivalent molar concentrations was observed, indicated a true LSH-6His:dsDNA binding event independent on the 6His tag (Figure 6.1B).

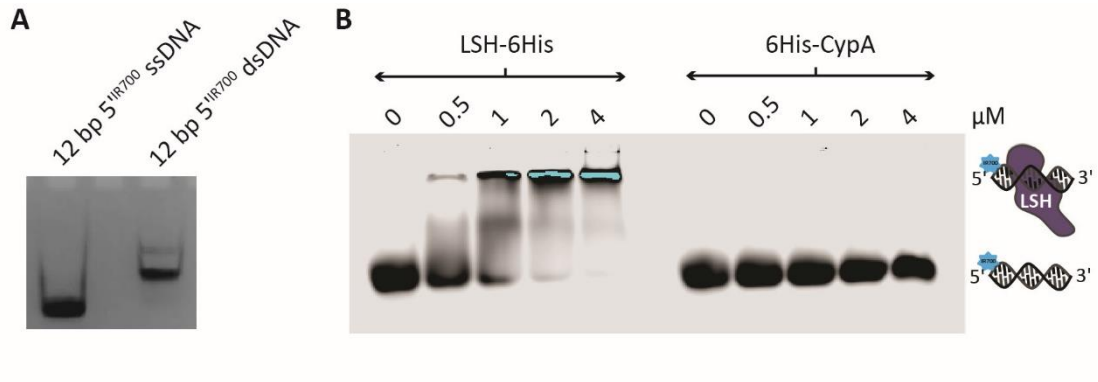

Figure 6.1| EMSA of LSH-6His binding to dsDNA.
A| Methylene blue stained acrylamide gel of the 12 bp 5'^{IR700} fluorescent ssDNA and annealed 12 bp 5'^{IR700} fluorescent dsDNA. DNA sequence 5'-GCTGCGTGCGTT-3'.
B| EMSA agarose gel of increasing molar concentrations (0.5-4 µM) of LSH-6His or 6His-CypA and 1 µM 12 bp 5'^{IR700} fluorescent dsDNA.

6.2.2| LSH-6His has a K_D of 1.2 µM for dsDNA based on EMSA

The intensity of the bound DNA band in the EMSA was measured to estimate a K_D of the 12 bp dsDNA for LSH-6His over a fine range of LSH-6His concentrations between 0.2 µM and 4 µM (Figure. 6.2A and Figure 6.2B). The K_D is defined as the concentration of the receptor at which 50% occupancy is achieved. From three separate EMSAs the estimated K_D was 0.87 µM ± 0.02 µM (Figure 6.2B).

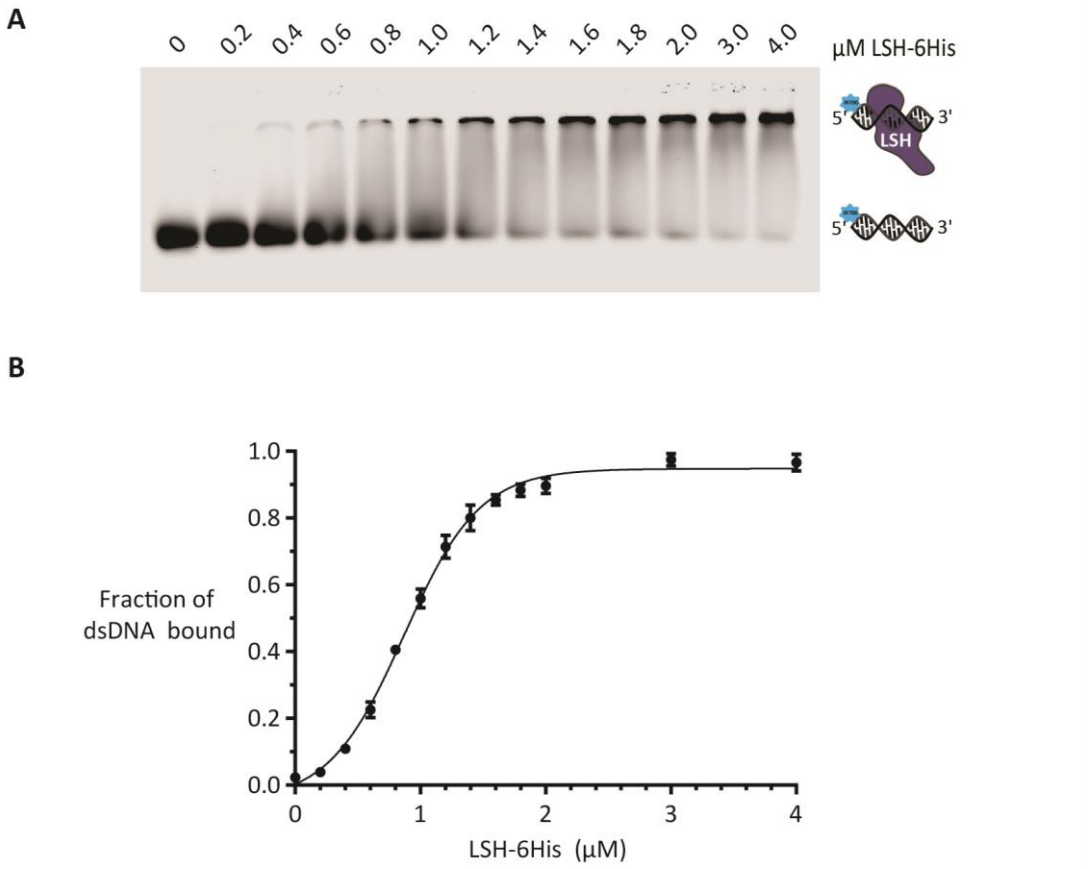

Figure 6.2| Quantitative EMSA of LSH-6His binding to dsDNA.
A| EMSA agarose gel of increasing molar concentrations (0.2-4 µM) of LSH-6His and 1 µM 12 bp 5$'^{IR700}$ fluorescent dsDNA.
B| The plot shows the fraction of dsDNA bound as LSH-6His was titrated, with a line of best fit using non-linear regression (sigmoidal - variable dose response). The value for the K_D was obtained from the line of best fit and is 0.87 µM (±0.02 µM). Results are averaged from three separate EMSAs.

Because the EMSA is a gel-based assay, complex formation might be artificially stabilised compared to what it would be in solution, also the agarose gel and running buffer were at a non-physiological pH of 9.3. Therefore, a fluorescence polarisation assay was used to quantify the K_D of a LSH-6His:12 bp dsDNA complex in solution, as a complementary approach to validate this finding.

6.2.3| Fluorescence Polarisation assay measures a K_D of 0.38 µM for a LSH-6His:dsDNA complex

Fluorescence polarisation (FP) is a non-disruptive method to measure the association of a fluorescent ligand, commonly DNA, a peptide or a drug, with a larger receptor such as a protein (Rossi and Taylor, 2011b). FP measures the light emitted from the fluorescent ligand that has been excited with vertically polarised light. Anisotropy is calculated using the intensities of the light detected in the vertical (parallel) and horizontal planes (perpendicular) with respect to the excitation light using the following formula:

$$A = \frac{I - I\perp}{I + 2I\perp}$$ Equation (4.1)

A = Fluorescence anisotropy
I = Intensity from parallel polariser
$\mathbf{I}_\perp$ = Intensity from perpendicular polariser

A typical FP assay is shown schematically in Figure 6.3.

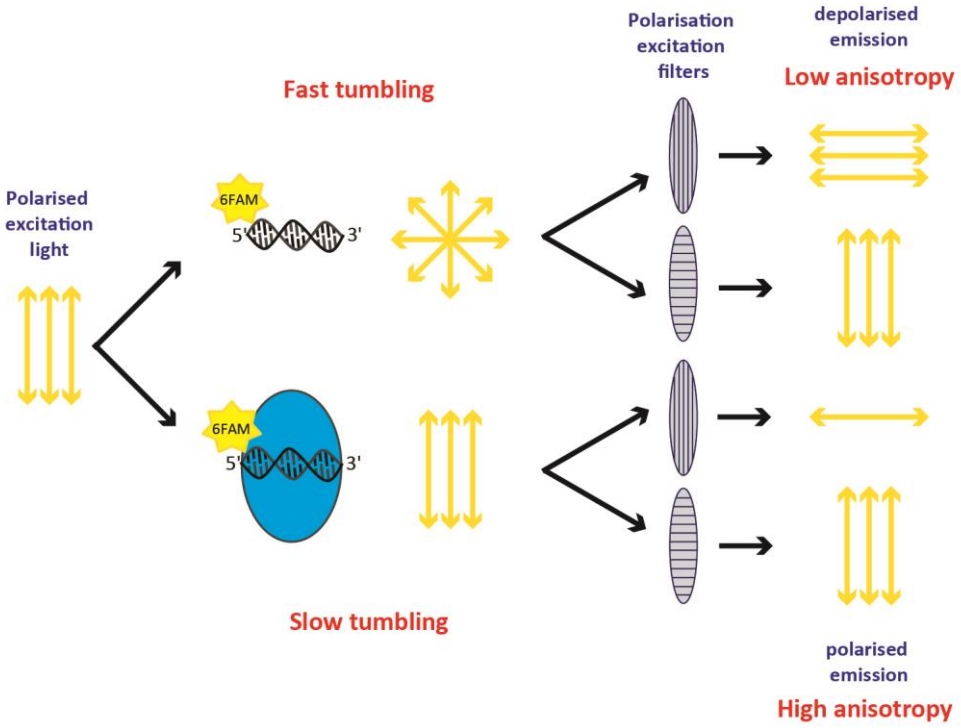

Figure 6.3| Fluorescence anisotropy model.
A fluorescein fluorophore (6FAM) is able to tumble in solution in the excited state within its fluorescence lifetime (4 ns). Polarisers that are parallel and perpendicular to the plane of excitation measure the emitted light from the fluorophore. The small fluorescently labelled sample, $5'^{6FAM}$ dsDNA is excited with vertically polarised light. In this scenario the fluorophore will tumble quickly and the orientation of the fluorophore will be randomised and the emitted light is less polarised with a low anisotropy. However a large complex of the small $5'^{6FAM}$ dsDNA with a large protein (shown in cyan) will slow down the rate of fluorophore tumbling. In this scenario the orientation of the fluorophore is less randomised and the emitted light is more polarised with a high anisotropy. Adapted from Rossi & Taylor 2011.

Fluorescein is typically used for labelling the ligand because the fluorescence lifetime (4ns) is long enough to measure an anisotropy change when the protein is bound. A range of protein concentrations below and above the ligand concentration are used and anisotropy is plotted against protein concentration to determine a K_D.

The formation of 12 bp 5'^{6FAM} DNA duplex was confirmed as it migrated more slowly than the non-annealed 12 bp 5'^{6FAM} ssDNA (Figure 6.4A) on a polyacrylamide gel due to its larger size.

The FP assay is suitable for quantifying LSH-6His binding to the 12 bp 5'^{6FAM} dsDNA because the ligand (12 bp 5'^{6FAM} dsDNA) is 10 times smaller than the protein (LSH-6His) which enables a greater change in anisotropy to be detected (Rossi & Taylor 2011).

The highest fluorescence signal without saturating the detector (data not shown) was achieved with 50 nM 12 bp 5'^{6FAM} dsDNA. Initially, buffer conditions of 100 mM, 50 mM and 10 mM NaCl were tested in the FP assay to measure the effect of ionic strength on LSH-6His dsDNA binding (Figure 6.4B). In the buffer containing 100 mM NaCl there was no change in anisotropy, indicating that no DNA binding occurs (Figure 6.4B). The 50 mM and 10 mM NaCl buffer conditions resulted in an increase in anisotropy with increasing LSH-6His concentration with a peak in anisotropy at 2 µM (Figure 6.4B). The anisotropy change for the 10 mM buffer condition was 2.6 times greater than the 50 mM NaCl buffer condition indicating that dsDNA binds LSH-6His with a slower off rate in the 10 mM NaCl buffer (Figure 6.4B). An accurate K_D could not be accurately determined with the 50 mM NaCl buffer condition due to small anisotropy changes, however, the curve was similar to the 10 mM NaCl buffer condition indicating the affinity of LSH-6His for 12 bp dsDNA in 50 mM NaCl and 10 mM NaCl are similar. The FP assay with 10 mM NaCl buffer condition was performed in triplicate for calculating a K_D with greater confidence and accurate error (Figure 6.4C). The calculated K_D from the line of best fit was 0.45 µM (± 0.1 µM).

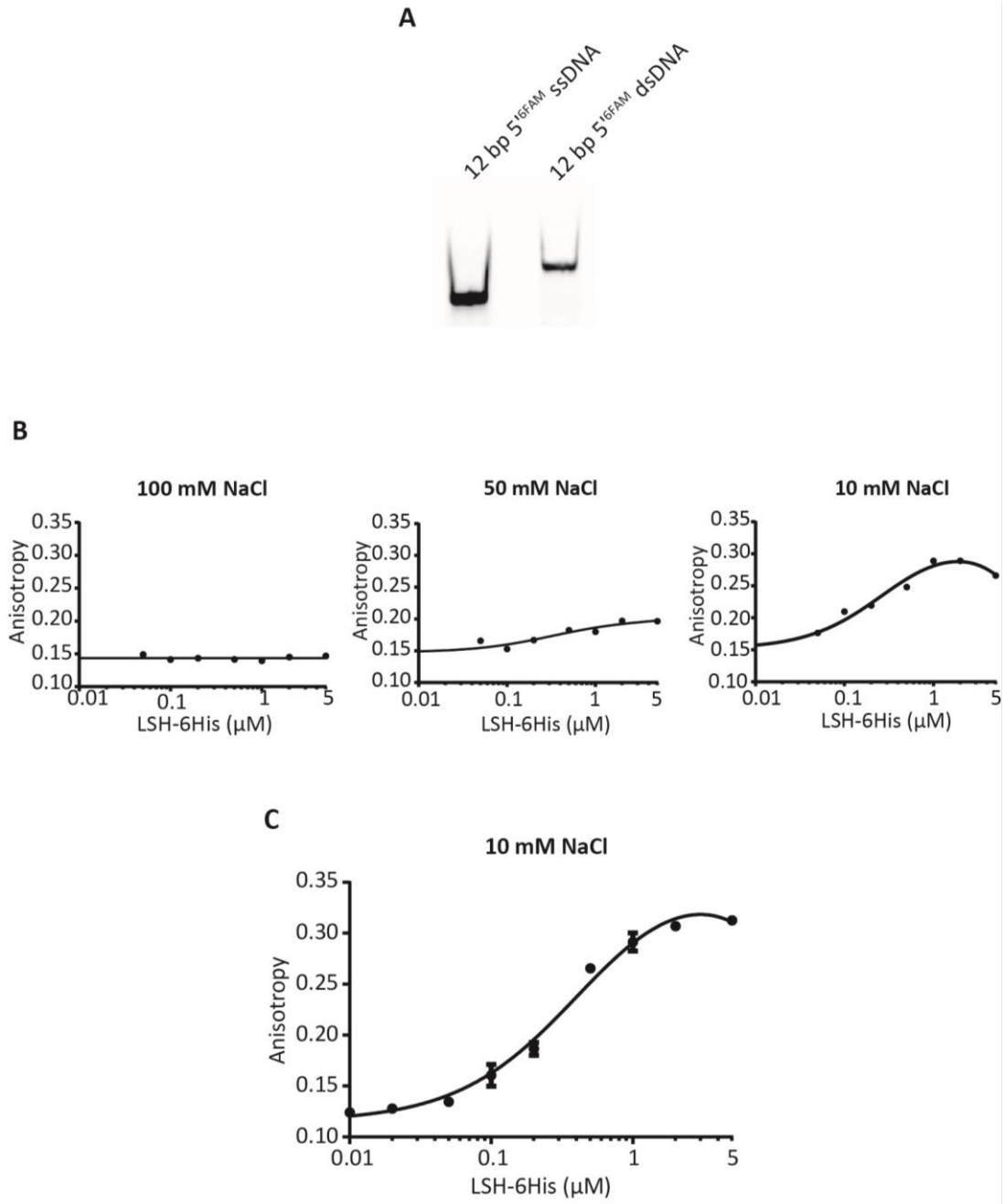

Figure 6.4| FP assay of LSH-6His with dsDNA

A| Methylene blue stained acrylamide gel of the 12 nt 5'^{6FAM} ssDNA and annealed 12 bp 5'^{6FAM} dsDNA.

B| Excitation was 493 nm and emission was 521 nm (515 nm cut-off). The plot shows the anisotropy of 12 bp 5'^{6FAM} dsDNA as LSH-6His was titrated in 100 mM, 50 mM or 10 mM NaCl buffer, with a line of best fit using non-linear regression (sigmoidal - variable dose response).

C| Same as in **B|** with averages and standard deviation from triplicates for the 10 mM NaCl buffer. The value for the K_D was obtained from the line of best fit is 0.45 (± 0.1 μM)

There are several differences between the EMSA and FP assay results. Firstly, a clear binding event between LSH-6His and the 12 bp dsDNA occurs in EMSA in buffer containing 100 mM NaCl, whereas in the FP assay this does not occur. The FP assay was performed in physiological pH buffer (50 mM HEPES pH 7.25, 10% Glycerol, 1mM DTT and different NaCl concentrations - 10, 50 or 100 mM NaCl). However, as the EMSA agarose gel and running buffer do not contain NaCl, when the mixture in undergoing electrophoresis, NaCl is being diluted lowering local ionic strength allowing the formation of a LSH-6His:12 bp dsDNA complex.

The K_D of 0.45 µM calculated from the FP assay is 0.32 µM less than the K_D calculated from the EMSA of 0.87 µM. However, the K_D values cannot be directly compared because the concentrations of the dsDNA were 50 nM for the FP assay and 1 µM for the EMSA assay, therefore, if I wanted to directly compare the K_D values, I would need to use 50 nM 12 bp dsDNA in the EMSA. However, due to time constraints, this experiment could not be performed.

Next, I wanted to investigate if LSH-6His is able to bind ssDNA as this has not been tested *in vitro*. The ATPase core of *Sulfolobus solfataricus* RAD54 binds to ssDNA with a 100 fold lower affinity than dsDNA and can only hydrolyse ATP with dsDNA (Dürr et al. 2005, Mazin et al. 2010; Bugreev et al. 2006). SsoRad54cd strongly prefers dsDNA (Kd = 0.10 ± 0.02 µM) over ssDNA (Kd =11 ± 5 µM), consistent with the dsDNA- but not ssDNA-stimulated ATPase activity. The conserved ATPase region of chromatin remodelers cannot bind to ssDNA as efficiently as the SF1 ATP-dependent helicase because it does not have a single-stranded DNA binding domain to firmly grab onto the nucleotides of ssDNA (Dürr *et al.*, 2005). Therefore I hypothesise LSH-6His has low or no affinity for ssDNA.

6.2.4| LSH-6His has negligible binding to ssDNA

A variant of the 12bp dsDNA used in EMSA and FP assays, because the sequence used (Figure 6.1) could form hairpin structures with a T_m of 27°C which LSH-6His could bind to (data not shown). Therefore the sequence was designed preventing the formation of hairpin structures at temperatures greater than 6°C (G**C**TGCGTGC**G**TT into G**T**TGCGTGC**T**TT).

An EMSA of LSH-6His with 12 bp 5'^{IR700} dsDNA used for Figure 6.1 and 6.2 was run alongside the EMSA of LSH-6His with the 12 nt 5'^{IR700} ssDNA (Figure 6.5A and Figure 6.5B). The EMSA with the 12 bp 5'^{IR700} dsDNA shifted with LSH-6His (Figure 6.5A), but there was negligible binding of LSH-6His with 12 nt 5'^{IR700} ssDNA at LSH-6His concentration of 4 µM, indicating that LSH-6His does not possess an affinity for ssDNA (Figure 6.5B), in agreement with the hypothesis.

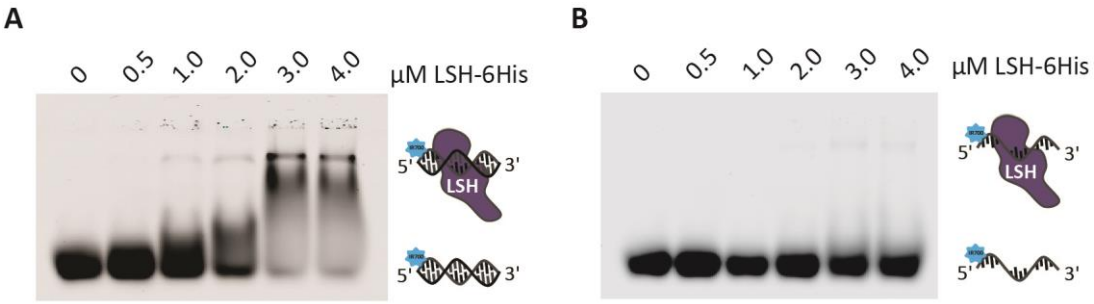

Figure 6.5| EMSA of LSH-6His with dsDNA or ssDNA.
A| EMSA agarose gel of increasing molar concentrations (0.5-4 µM) of LSH-6His and 1 µM 12 bp 5'^{IR700} fluorescent dsDNA.
B| EMSA agarose gel of increasing molar concentrations (0.5-4 µM) of LSH-6His and 1 µM 12 nt 5'^{IR700} fluorescent ssDNA.

6.2.4| LSH-6His has similar affinity for unmethylated and cytosine methylated duplex DNA

Next, I investigated the ability of LSH-6His to bind cytosine methylated dsDNA. LSH$^{-/-}$ mouse embryonic fibroblasts (MEFs) show a 50-70% reduction in cytosine DNA methylation at repeat elements at distinct genomic sites, demonstrating that LSH is necessary for *de novo* DNA methylation (Dennis et al., 2001, Tao et al., 2011, Myant et al., 2011, Sun et al., 2004, Huang et al., 2004).

This finding is supported by the interaction of LSH with the *de novo* DNA methyltransferase DNMT3B from co-immunoprecipitation studies, indicating LSH is a primary regulator of *de novo* DNA methylation through recruitment of DNMT3B to genomic sites as a result of LSH chromatin remodelling activity (Myant and Stancheva, 2008).

Therefore, unmethylated GC-rich dsDNA might direct where on chromatin LSH localises to in order to regulate DNA methylation. It was hypothesised that LSH has a greater affinity for unmethylated DNA than methylated DNA due to the significant loss of methylation in these parts in LSH$^{-/-}$ cells.

To this end, I compared LSH-6His binding to methylated dsDNA and unmethylated dsDNA *in vitro* (Figure 6.6). CpG methylation at two sites in the 12 bp 5'^{IR700} dsDNA and on both strands of the duplex was used with LSH-6His in an EMSA alongside unmodified 12 bp 5'^{IR700} dsDNA as a positive control (Figure 6.6). The formation of 12 bp 5'^{IR700} methylated dsDNA post-annealing was confirmed as it migrated slower due to a larger size when compared to the non-annealed 12 nt 5'^{IR700} methylated ssDNA (Figure 6.6A). LSH-6His appeared to bind the methylated and non-methylated dsDNA with similar affinity as the band shifts for methylated dsDNA and dsDNA were too similar to see any significant difference (Figure 6.6B and Figure 6.6C).

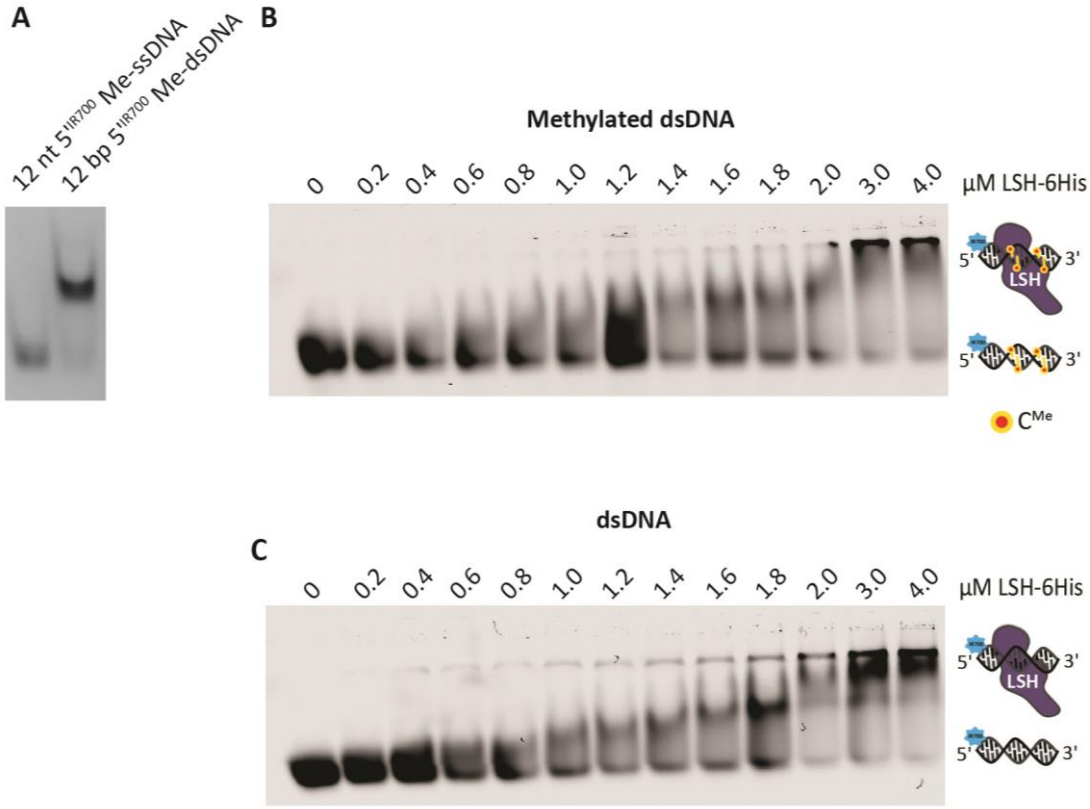

Figure 6.6|EMSA of LSH-6His with non-methylated or methylated DNA.
A| Methylene blue stained acrylamide gel of the 12 nt 5'^{IR700} fluorescent Me-ssDNA and the annealed 12 bp 5'^{IR700} fluorescent Me-dsDNA. Me-DNA sequence 5'-GCTGC^{Me}GTGC^{Me}GTT-3'.
B| EMSA agarose gel of increasing molar concentrations (0.2-4 µM) of LSH-6His and 1 µM 12 bp 5'^{IR700} fluorescent Me-dsDNA.
C| EMSA agarose gel of increasing molar concentrations (0.2-4 µM) of LSH-6His and 1 µM 12 bp 5'^{IR700} fluorescent dsDNA.

Although no clear difference for LSH-6His binding to methylated or non-methylated dsDNA was detected by EMSA, a more sensitive assay such as the FP assay would be a useful as a next step to confirm if LSH-6His affinity for dsDNA is not affected by cytosine methylation.

As LSH-6His is a putative chromatin remodeler, we sought to test LSH-6His binding to nucleosomes, which is a more native chromatin component than dsDNA.

6.2.5| The reconstitution of mononucleosomes

LSH has been shown to localise to the condensed form of chromatin – heterochromatin and more specifically pericentromeric heterochromatin (Lungu et al. 2015, Yan et al. 2003). It is not known whether LSH binds to DNA linking nucleosomes, to histone tails of nucleosomes or the DNA wrapped around the nucleosome, therefore, I reconstituted mononucleosomes both with and without flanking dsDNA (Figure 6.7).

The recombinant core histones (H2A, H2B, H3 and H4) of *X. laevis* origin were expressed in *E. coli* and purified from inclusion bodies in unfolding buffer. Equimolar ratios of the core histones were dialysed into 2M NaCl buffer for histone octamer formation. SEC was performed on dialysed core histones to purify a homogeneous population of histone octamers, observed as peak 1 (Figure 6.7A and Figure 6.7B). The Widom 601 DNA sequence that binds to histone octamers for nucleosome formation was PCR amplified from the pUC-601 plasmid. A forward primer with a $5'^{IR700}$ fluorescent probe was used to amplify a nucleosome binding sequence ($5'^{IR700}$ 147 bp dsDNA) or a nucleosome binding sequence with or without the 26 bp symmetrical overhanging DNA (Figure 6.7C). For mononucleosome reconstitution, the DNA was dialysed with histone octamers with increasing w/w ratios. A w/w ratio of DNA:histone octamer of 1:4 and 1:2.5 was optimal for complete reconstitution of a core mononucleosome (147 bp) and a mononucleosome with linkers (200 bp) respectively (Figure 6.7D and Figure 6.7E).

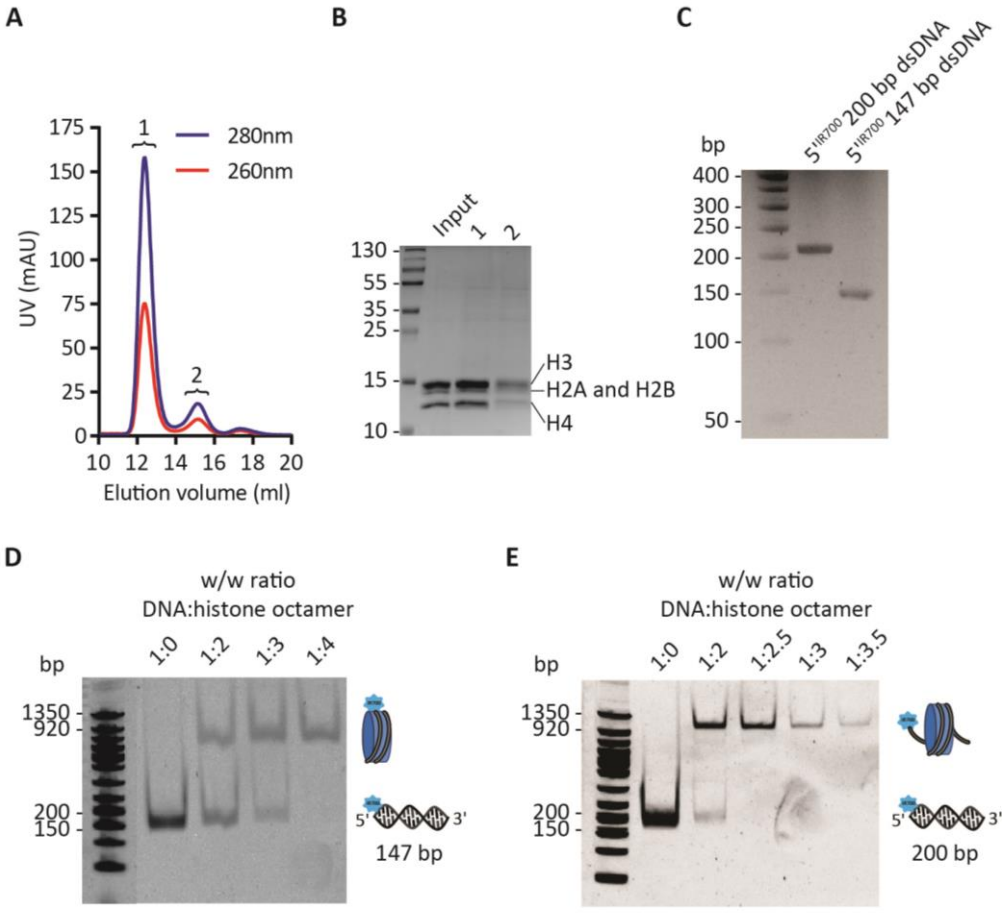

Figure 6.7|Mononucleosome reconstitution.
A| Chromatogram of histone octamer SEC with UV detection (280 nm and 260 nm).
B| Coomassie stained SDS-PAGE gel of Histone octamer SEC peak elution's 1 and 2.
C| SYBR safe stained agarose gel of purified fluorescent widom 601 binding sequences: 5'^{IR700} 200 bp dsDNA and 5'^{IR700} 147 bp dsDNA.
D| SYBR safe stained polyacrylamide gel of w/w ratios of 5'^{IR700} 147 bp dsDNA:Histone octamers.
E| SYBR safe stained polyacrylamide gel of w/w ratios of 5'^{IR700} 200 bp dsDNA:Histone octamers.

6.2.6| LSH-6His binds only to flanking DNA on mononucleosomes

An EMSA was performed to detect LSH-6His binding to the mononucleosomes with and without linkers (200 and 147bp). Because of the larger size of nucleosomes compared to 12 bp dsDNA, 0.5% agarose was used instead of 2% to increase the separating resolution (Figure 6.8). LSH-6His does not bind to the core mononucleosome, however, it does completely bind the free dsDNA which is present as a result of incomplete nucleosome reconstitution (Figure 6.8A). LSH-6His binds to a mononucleosome with 26 bp overhangs with a K_D of 28.4 nM (±6 nM) (Figure 6.8B and Figure 6.8C). This result indicates LSH-6His is only able to bind to the DNA that links nucleosomes and cannot bind histone tails or DNA wrapped around the histone octamer.

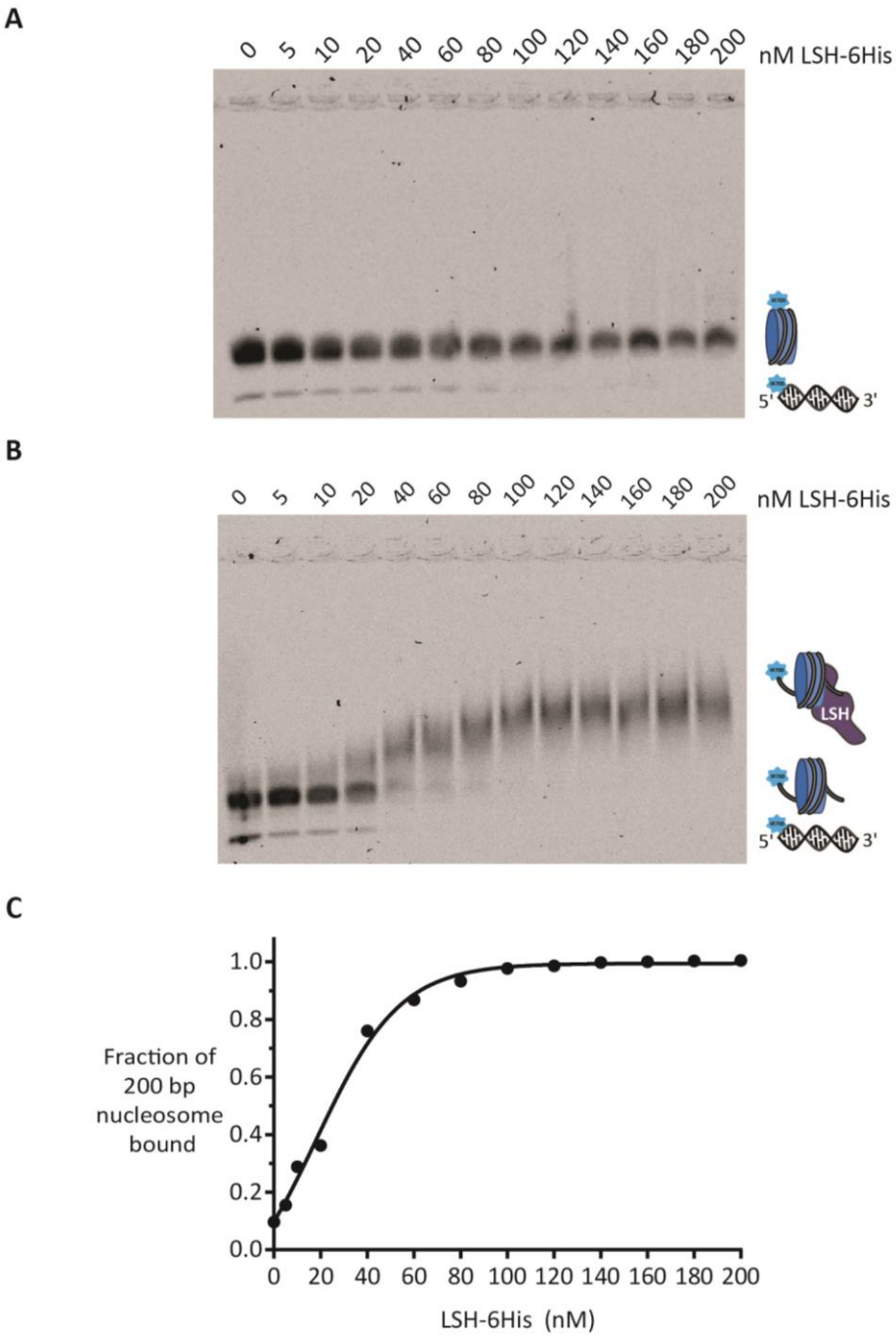

Figure 6.8| EMSA of LSH-6His with nucleosome with or without linker dsDNA
A| EMSA agarose gel of increasing molar concentrations (5-200 nM) of LSH-6His and 5 nM 147 bp (no linker) fluorescent nucleosome.
B| EMSA agarose gel of increasing molar concentrations (5-200 nM) of LSH-6His and 5 nM 200 bp (symmetric linker) fluorescent nucleosome.
C| The plot shows the fraction of 200 bp nucleosome bound as LSH-6His was titrated, with a line of best fit using non-linear regression (sigmoidal - variable dose response). The value for the K_D was obtained from the line of best fit and is 28.4 nM (±6 nM).

The K_D of LSH-6His for the mononucleosome with 26 bp overhangs is 31 times stronger than the K_D calculated for 12 bp dsDNA (28 nM vs 870 nM). However, the two K_D values cannot be directly compared as the concentration of substrate was 1 µM for the dsDNA EMSA and 5 nM for the nucleosome EMSA. However, the K_D calculated for a LSH-6His: 12 bp dsDNA complex from the FP assay was 450 nM which is ~16 times weaker than 28 nM K_D for the complex of LSH-6His:mononucleosome with 26 bp overhangs. This indicates the EMSA is likely overestimating the K_D, which could be due to a number of factors including the agarose gel matrix stabilising the complexes through a 'caging effect', the low ionic strength of the running buffer and gel matrix as well as electrophoresis causing an increase in local concentration of the protein and DNA. Therefore, this experiment should be repeated again and using an in-solution method to measure the K_D of an LSH-6His: mononucleosome in a more native environment. An FP assay would not be suitable here because a mononucleosome with 26 bp flanking dsDNA is ~240 kDa, and for anisotropy change to be detected the receptor should be 10 times larger, which LSH is not. An alternative method that could be used is MicroScale Thermophoresis, which would quantitatively detect complex formation of LSH to mononucleosomes.

To summarise I have optimised an EMSA which shows LSH-6His binds dsDNA and not ssDNA in vitro. LSH-6His binds dsDNA in solution with 50 mM NaCl or less, with an affinity of 0.45 µM in a 10 mM NaCl buffer. LSH-6His binds to cytosine methylated dsDNA with a similar affinity to dsDNA and LSH-6His is able to bind to mononucleosomes with linker DNA but not core mononucleosomes lacking linker DNA.

6.3| Discussion

The main aim of this chapter was to characterise the biochemical interactions of recombinant LSH with DNA and nucleosomes. I optimised an EMSA to analyse/quantify the binding of LSH-6His with 12 bp dsDNA by using an agarose gel and running buffer with Tris-CAPS pH 9.3 buffer. These conditions enabled the migration and subsequent visualisation of a LSH:DNA complex within the gel (Figure 6.1). LSH-6His has a K_D of 0.87 µM for 12 bp dsDNA. To analyse the interaction in physiological conditions an in-solution FP assay was developed: at pH 7.2 LSH-6His has a K_D of 0.45 µM for 12 bp dsDNA; binding occurs in low ionic buffer conditions, between 10-50 mM NaCl, but not in 100 mM NaCl (Figure 6.3).

The physiological monovalent salt concentration in a mammalian cell is ~150 mM (12 mM Na^+ 139 mM K^+ ions) (Lodish, Berk and Zipursky, 2000). Therefore, LSH binds *in vitro* to DNA at an NaCl concentration at least 3 times lower than physiological concentrations.

Early EM-based structural studies found chromatin fragments form different structures dependent on NaCl concentration (Huang and Cole, 1984): aggregates formed in 150 mM NaCl, solenoidal chromatin formed at 100 mM NaCl, compacted chromatin forms at 50 mM NaCl and at 10 mM NaCl chromatin is extended (Huang and Cole, 1984). My finding that LSH interacts *in vitro* with dsDNA between 10-50 mM NaCl, suggests that LSH may bind to compacted and extended chromatin, *in vivo* which is related well to heterochromatin.

The affinity of LSH to dsDNA *in vitro* is difficult to compare directly with other chromatin remodelers for several reasons, which will be discussed. Other chromatin remodelers have accessory dsDNA-binding domains, such as the HSS domains, whereas LSH only possesses the DNA binding motifs in the SNF2 ATPase region.

In contrast to LSH, two well-studied chromatin remodelers, CHD and ISWI, are both unable to bind dsDNA less than 15 bp (ISWI) or 30 bp (CHD2). Optimal DNA binding requires more than 40 bp for both ISWI and CHD1 (Whitehouse *et al.*, 2003b; Liu, Ferreira and Yusufzai, 2015). The ATPase activity of RSC, RAD54, ISWI and CHD2 increases with increasing DNA duplex longer than 25 bp (Liu et al. 2015, Mazina & Mazin 2004; Saha et al. 2002, Whitehouse et al. 2003). However, the RAD54 ATPase domain structure indicates 12 bp dsDNA is the minimal required for binding (Dürr *et al.*, 2006), which matches my observations that LSH is also able to bind the minimal length of 12 bp dsDNA required for binding.

The affinity of certain chromatin remodelers for dsDNA is length dependent. For example, ISWI has a reported K_D of 18 nm for 20 bp dsDNA (± 2 nm) (Al-Ani *et al.*, 2014), whilst CHD2 has a reported K_D of 160 nm for 40 bp dsDNA (Liu, Ferreira and Yusufzai, 2015). Therefore quantifying the affinity of LSH for different lengths of dsDNA and how its ATPase activity changes would enable a direct comparison to be made with other ATP-dependent chromatin remodelers.

LSH did not show any binding to 12 nt ssDNA (Figure 6.5), which was expected given *S. solfactaricus* RAD54 has a 100 fold lower affinity for ssDNA than dsDNA (Dürr *et al.*, 2006). In comparison, the chromatin remodelers and remodelling complexes RAD54, RSC, SWI/SNF and ISWI can bind ssDNA, and possess ATPase activity in the presence of ssDNA. In these experiments the ssDNA used was 40 nt or circular ssDNA over 1000 nt in length (Cairns *et al.*, 1994, 1996; Petukhova, Stratton and Sung, 1998; Whitehouse *et al.*, 2003b). Therefore hairpin secondary structures may have formed, mimicking dsDNA. Therefore testing if LSH can bind to longer ssDNA, which may form secondary structures would identify if LSH has a similar affinity to other chromatin remodelers shown to bind longer ssDNA.

The chromatin remodeler RAD54, which is involved in double-strand break repair (DSBR) binds to holiday junction-like branched DNA structures with a 200-fold higher affinity than ssDNA or dsDNA alone, consistent with its function in late stage homologous recombination (Bugreev, Mazina and Mazin, 2006). The ATPase activity of LSH is necessary for efficient H2AX phosphorylation and repair of DNA damage in irradiated cells (Burrage *et al.*, 2012). One hypothesis is that LSH could act locally at double-strand breaks (DSBs) that occur in the vicinity of chromatin-bound LSH (Burrage et al. 2012). Therefore it would interesting to study if LSH has a greater affinity for branched DNA structures over dsDNA alone, which would aid in determining the function of LSH in DSBR.

LSH bound to both methylated or unmethylated 12 bp dsDNA with similar affinity (Figure 6.6). LSH is necessary for *de novo* methylation at repeat sequences, therefore it was hypothesised that LSH has a greater affinity for unmethylated DNA due to the significant loss of methylation in these regions in $LSH^{-/-}$ cells. However, a clear difference in affinity would need to be determined first using a lower concentration of DNA (5 nM instead of 1 µM) used in this study, as there may be nanomolar differences in affinity of LSH for methylated dsDNA compared to dsDNA.

The discrete complex formed with DNA is a suitable candidate for crystallisation. Crystallisation of LSH with dsDNA (12, 16 or 20 bp in length) was trialled with and without ADP with 400 crystallisation conditions (for each DNA length), however no hits were detected. If more time allowed further crystallisation trials of LSH with DNA and different NTP's to assimilate a transitions state (e.g. ADP-beryllium fluoride) would have been tested.

Recombinant LSH-6His was tested for ATPase activity through use of a colorimetric ATPase assay (PiColorLock – Innova Biosciences) or a luminescent ATPase assay (ADP-Glo – Promega) (Data not shown). The assay was performed with LSH-6His in the presence of ATP or ATP with dsDNA at differing LSH-6His, dsDNA and NaCl concentrations (Data not shown). However, no ATPase activity was detected from two separate batches of purified LSH-6His (Data not shown). Recombinant LSH purified with an N-terminal 6His-tag by Kevin Myant was shown to show ATPase activity, with increased activity with dsDNA or mononucleosomes, the LSH^{K237Q} ATPase mutant was used as a negative control (Burrage et al., 2012). A radioactive ^{32}P-γATP TLC ATPase assay was used, therefore this should be tested for LSH-6His. The major difference between the two studies is the location of the 6His-tag which is on the C-terminus for LSH in my study and on the N-terminus for Kevin's study. This could influence activity of LSH, and the comparing the activity of 6His-LSH and LSH-6His is required.

I observed that LSH-6His binds only to mononucleosomes that have flanking dsDNA extending from the core 147 bp DNA, which make contacts with the histone octamer (Figure 6.8). This indicates LSH-6His does not bind to the dsDNA in contact with the histone octamer either through the DNA binding ATPase or the flanking domains. It also suggests the flanking domains of LSH do not bind to the histone tails of nucleosomes.

The ATPase region of ISWI, CHD1 and SNF2 have all been shown to bind SHL2 on a nucleosome, 20 bp away from the dyad (Zofall *et al.*, 2006; Jeffrey N. McKnight *et al.*, 2011). LSH does not bind to this region as indicated by it not binding to a 147 bp mononucleosome (Figure 6.8). Potentially the N-terminal domain, C-terminal domain or extended linker-protrusion 2 of LSH might be preventing the binding of the ATPase cleft to nucleosome-bound DNA but not free DNA. To visualise this, a high-resolution structure of LSH in complex with a 200 bp nucleosome using cryo-EM would be required. This would definitively explain what portions of LSH might be preventing the ATPase cleft to bind DNA at SHL2 seen with the other chromatin remodeler.

It has been found that the plant LSH homolog, DDM1 is required at heterochromatic sequences for DNA methylation and this is dependent on the linker histone H1 (Zemach *et al.*, 2013). Therefore it can be speculated that LSH might require both dsDNA and linker histone H1 for localisation to heterochromatin and thus provide DNA methyltransferases access to H1-containing heterochromatin. Therefore future experiments would be to quantify binding of LSH to nucleosomes with the linker histone H1. Further experiments could investigate the affinity and activity of LSH on nucleosomes with or without H1 using ATPase assays and nucleosome remodelling assays. In addition, structural investigations of LSH with chromatin in different structural states mediated by NaCl concentration between 10 mM and 100 mM NaCl would aid in understanding a relationship between chromatin structure and LSH enzymatic function. If time allowed it would be interesting to determine if LSH in an active (ATP) or inactive transition state (ADP-BeF) had a different affinity for nucleosomes.

Chapter 7| Investigating LSH-DNMT3B complex formation *in vitro*

7.1| Introduction

LSH is a putative chromatin remodelling protein but is also an important regulator of DNA methylation. $LSH^{-/-}$ mouse embryonic fibroblasts (MEFs) show loss of cytosine methylation at distinct genomic sites demonstrating that LSH is necessary for *de novo* DNA methylation (Myant *et al.*, 2011; Tao, Xi, Shan, Maunakea, Che, Briones, Eunice Y Lee, *et al.*, 2011). This finding is supported by immunoprecipitation of LSH with the *de novo* DNA methyltransferases DNMT3A and DNMT3B (Zhu *et al.*, 2006). Co-immunoprecipitation and pull-down assays demonstrated a direct interaction of the LSH N-terminus (1-503) with DNMT3B (Myant and Stancheva, 2008). To further our understanding of LSH interactions with DNMT3B, we aimed to biochemically study if LSH and DNMT3B interact directly *in vitro*.

7.2| Results

7.2.1| Cloning of DNMT3B for simultaneous expression of LSH and DNMT3B using the MultiBac system.

I used the MultiBac system, which enables cloning of multiple genes of interest for simultaneous expression of all genes from one transfective EMBacY bacmid (Fitzgerald *et al.*, 2006). Previously, DNMT3B with an N-terminal 6His tag has been expressed in an insect cell system, purified and shown to be active (Suetake *et al.*, 2003). As LSH was expressed with a C-terminal 6His tag, I needed to use a differential tagging method for DNMT3B for use in pulldowns with LSH-6His. I cloned a DNMT3B gene encoding a StrepII tag linked to the N-terminus via a TEV-cleavable sequence (named StrepII-DNMT3B) under the control of the p10 promoter into a pFL plasmid containing LSH with a C-terminal 6His tag sequence under control of the polH promoter (Figure 7.1).

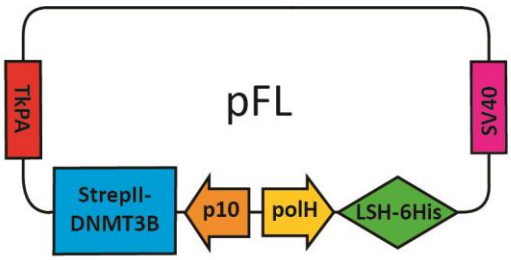

Figure 7.1|MultiBac pFl plasmid cloned with StrepII-DNMT3B and LSH-6His.
Plasmid cartoon showing the pFL plasmid containing StrepII-DNMT3B (under control of the p10 promoter and the TkPA terminator) and LSH-6His (under control of the polH promoter and SV40 Temrinator). The pFL plasmid would be transformed into the EMBacY bacmid and purified for the expression of StrepII-DNMT3B and LSH-6His from one transfective bacmid. Plasmid name - pFL-p10-StrepII-DNMT3B-polH-LSH-6His.

7.2.2| Small-scale expression and co-purification of StrepII-DNMT3B and LSH-6His

The EMBacY bacmid containing StrepII-DNMT3B and LSH-6His ORFs (pFL-p10-StrepII-DNMT3B-polH-LSH-6His) was transfected into *Sf9* cells. BIICs were generated from V1 baculovirus and a 200 ml culture of *Sf9* cells at 2x10 ^{6}cells/ml was infected with a volume ratio 1:20 (V1 baculovirus:*Sf9* cells) and pelleted 4 days post infection.

Cells were lysed in 100 mM NaCl and treated with benzonase to digest DNA and RNA. To test if StrepII-DNMT3B co-purified with LSH-6His I first immobilised the LSH-6His on IMAC resin. IMAC purification was performed with 100 mM NaCl or 500 mM NaCl to test if low or high ionic buffers disrupted a LSH-DNMT3B complex (Figure 7.2A and Figure 7.2C). LSH-6His, but not StrepII-DNMT3B was visible on a coomassie stained SDS-PAGE (Figure 7.2A and Figure 7.2C). Western blotting was performed with anti-LSH and anti-StrepII antibodies to detect LSH-6His and StrepII-DNMT3B (Figure 7.2B and Figure 7.2D). Both LSH-6His and StrepII-DNMT3B are present in fractions eluted in 100 mM NaCl. The anti-LSH signal intensity is low in the flowthrough fraction, indicating successful immobilisation of LSH-6His, whereas anti-StrepII signal does not change from the lysate to the flowthrough suggesting some StrepII-DNMT3B remains unbound (Figure 7.2B). LSH and StrepII-DNMT3B also appear in the fractions eluted in 500 mM NaCl buffer (Figure 7.2D) albeit with reduced anti-StrepII signal compared to the 100 mM NaCl purification (Figure 7.2B). Moreover, StrepII-DNMT3B is washed out in W2, when the 500 mM NaCl buffer is used (Figure 7.2D), indicating that high salt concentration disrupts DNMT3B-LSH contacts.

First, purification of StrepII-DNMT3B on Streptactin sepharose (StrepTrap column) was attempted using desthiobiotin to elute the immobilised StrepII tagged protein. StrepII-DNMT3B was barely visible in elution fractions on a coomassie stained SDS-PAGE (Figure 7.2E). An anti-StrepII Western blot confirmed StrepII-TEV-DNMT3B was in both the elution and the unbound fractions (Figure 7.2F), suggestive of inefficient binding to Streptactin sepharose, therefore, an alternative procedure was sought.

Overall less StrepII-TEV-DNMT3B eluted than LSH-6His during the IMAC step in both buffer conditions and could not be successfully purified using Streptactin purification. Therefore, reforming a DNMT3B-LSH complex after purifying the proteins separately was decided as an improved strategy.

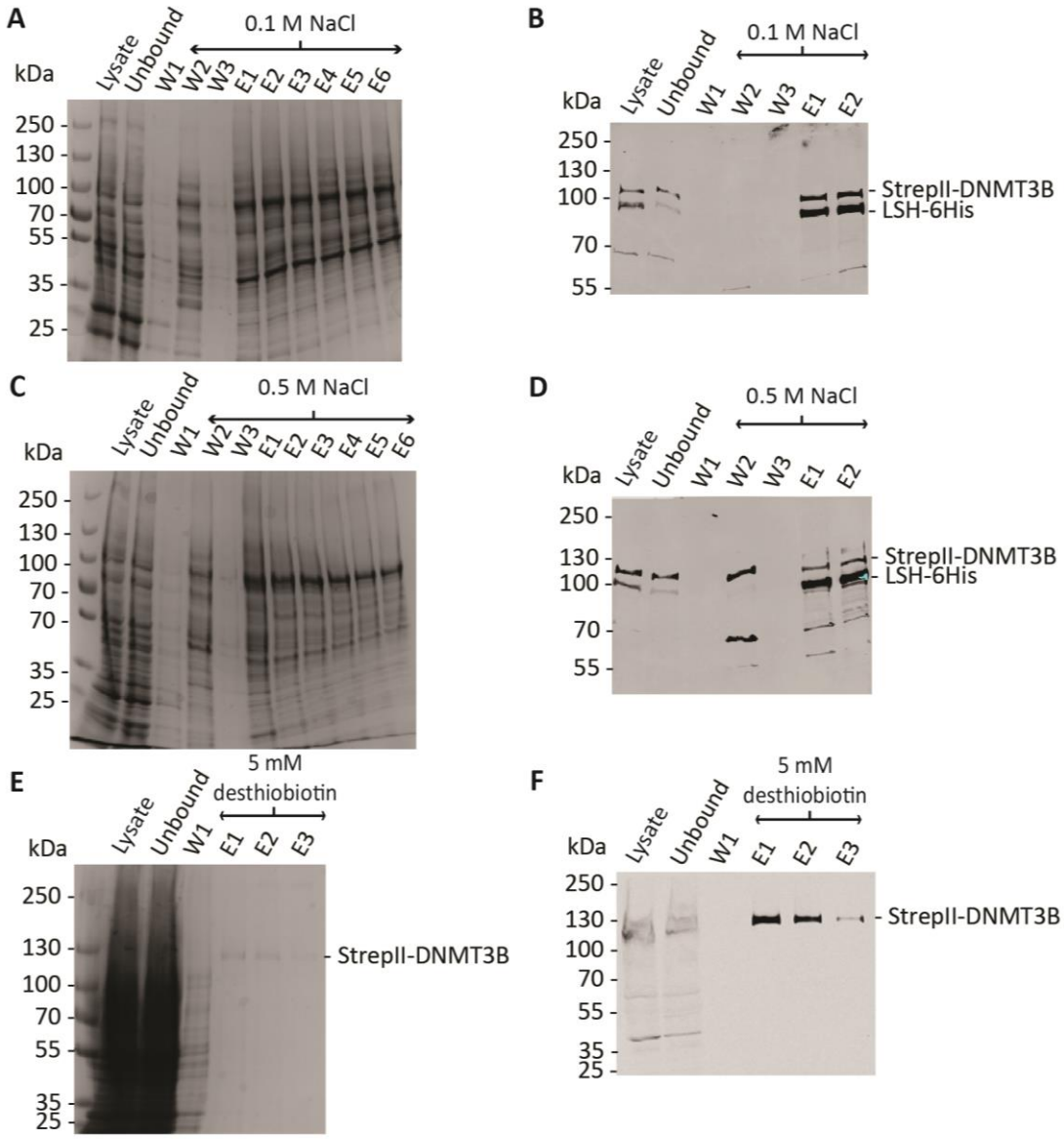

Figure 7.2| Small scale co-expression and purification of StrepII-DNMT3B and LSH-6His
A| Coomassie stained SDS-PAGE gel of LSH-6His and StrepII-DNMT3B IMAC purification which used 100 mM NaCl buffer (0.01% of the total volume loaded per well for the Lysate and Unbound fractions). 10 ml washes were performed with W2 and W3 containing 40 mM Imidazole buffer (0.15% of the total volume loaded per well for the wash fractions). 1 ml elutions used 500 mM Imidazole buffer (1% of the total volume loaded per well for each elution).
B| Anti-StrepII and anti-LSH Western blot of fractions from **A|.**
C| Coomassie stained SDS-PAGE gel of LSH-6His and StrepII-DNMT3B IMAC purification which used 500 mM NaCl buffer (0.01% of the total volume loaded per well for the Lysate and Unbound fractions). 10 ml washes were performed with W2 and W3 containing 40 mM Imidazole buffer (0.15% of the total volume loaded per well for the wash fractions). 1 ml elutions used 500 mM Imidazole buffer (1% of the total volume loaded per well for each elution).
D| Anti-StrepII and anti-LSH Western blot of fractions from **C|**.
E| Coomassie stained SDS-PAGE gel of StrepII-DNMT3B using a 1ml StrepTrap column. 20 ml of lysis buffer (50 mM HEPES pH 7.2, 500 mM NaCl, 10% Glycerol and 1 mM DTT) was used for the wash step (W1, 0.075% total volume loaded on the gel). 1 ml elutions used 5 mM desthiobiotin in lysis buffer (1% of the total volume loaded per well for each elution).
F| Anti-StrepII Western blot of the StrepII-DNMT3B StrepTrap purification from **E|**

7.2.3| Improving DNMT3B expression and purification

Since IMAC media has a higher binding capacity than Streptactin resin, a 6His tag was added to the StrepII-DNMT3B construct to allow primary purification by IMAC followed by capture on Streptactin resin.

In previous studies DNMT3B with an N-terminal 6His tag was expressed and purified with greater than 85% purity, but low yield (7%) after an IMAC purification (Suetake *et al.*, 2003). This was due to the occlusion of the tag since the N-terminal region is important for strong binding to nucleosomes (Jeong *et al.*, 2009) thereby likely making the N-terminal tags of nucleosome-bound DNMT3B inaccessible. Therefore, I decided to add a flexible linker between DNMT3B and both affinity tags to enhance binding to the resin.

I cloned DNMT3B with a multiple N-terminal tag comprising a 6His tag and GAGA linker between the StrepII tag and TEV cleavage site making the plasmid pFL-StrepII-6His-GAGA-TEV-DNMT3B (named StrepII-6His-DNMT3B for the rest of the study). Constructs encoding only DNMT3B were used to remove any decrease in expression due to co-expression of LSH. Expression of StrepII-DNMT3B under the control of the p10 promoter (Figure 7.3A) and of StrepII-6His-GAGA-TEV-DNMT3B under the control of the polH promoter (Figure 7.3B) were compared and quantified with Western blotting with anti-StrepII and anti-Tubulin antibodies (Figure 7.2C). StrepII-6His-DNMT3B expression was 6-fold higher under the control of the polH promoter than StrepII-DNMT3B under the control of the p10 promoter. The large difference was unexpected as both promoters are late stage and expected to have similar expression levels (Fitzgerald *et al.*, 2006). This may be due to the separation between the start codon and the promoter affecting expression. Therefore, StrepII-6His-DNMT3B under the control of the polH promoter was used for the rest of the study.

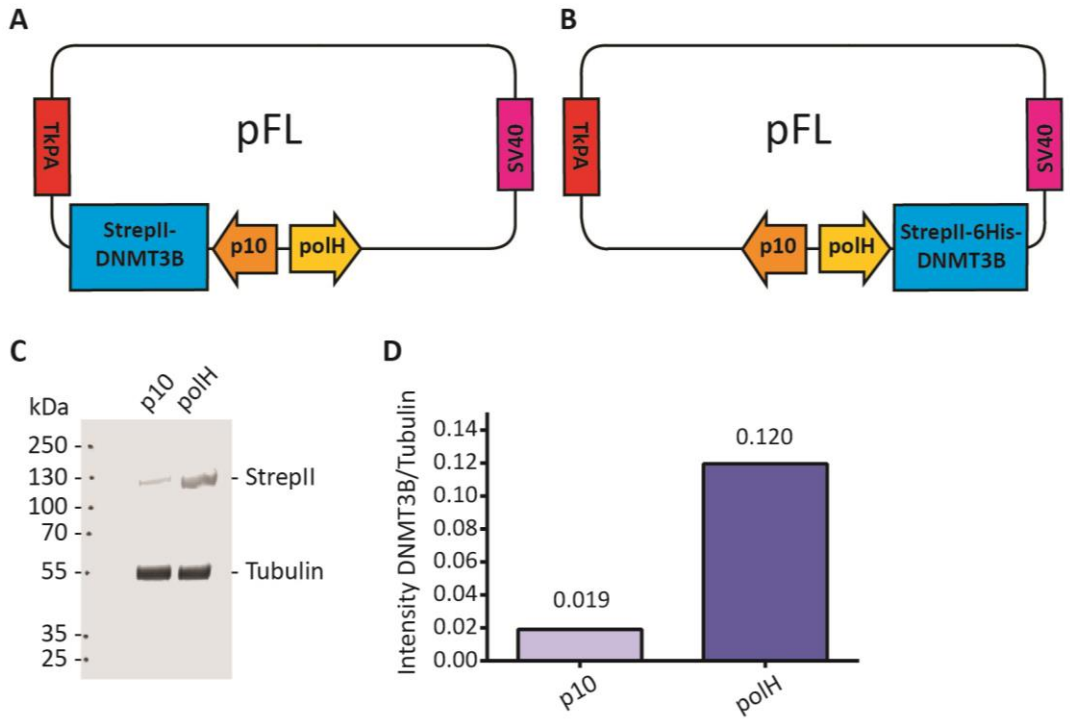

Figure 7.3| DNMT3B expression under control of a p10 or a polH promoter
A| Plasmid map of StrepII-DNMT3B under control of the p10 promoter and the TkPA terminator. Plasmid name - pFL-p10-StrepII-TEV-DNMT3B.
B| Plasmid map of StrepII-6His-DNMT3B under control of the polH promoter and the SV40 terminator. Plasmid name - pFL-polH-StrepII-6His-GAGA-TEV-DNMT3B.
C| Anti-StrepII and anti-Tubulin Western blot of lysates 4 days post infection of DNMT3B under the expression of the p10 promoter or the polH promoter.
D| Graph showing the relative normalised intensities of DNMT3B/Tubulin from the Western blot.

StrepII-6His-DNMT3B was expressed using optimised insect cell expression methods (as detailed in Chapter 3 section 3.2.5), but in *Sf9* cells at high cell density ($4x10^6$ cells/ml) were used instead of HighFive cells as this yielded the best protein expression (data not shown). StrepII-6His-DNMT3B expression was confirmed by the presence of a band at ~120 kDa in lysed infected cells that was not present in uninfected cells (Figure 7.4A). A small-scale purification from 100 ml of cells was performed using IMAC followed by StrepII affinity chromatography (Figure 7.4). StrepII-6His-DNMT3B was detected in the unbound and eluted fractions along with contaminants (Figure 7.4A). To check if unbound DNMT3B retained the tags I performed anti-6His Western blotting of the IMAC purification fractions (Figure 7.4B).

The 6Histag was detected in the unbound and elution fractions ruling out tag cleavage by proteases. Therefore, the 6His tag on some StrepII-6His-DNMT3B molecules may be inaccessible, thus precluding binding to affinity resin, as noticed previously (Figure 7.2F). Therefore, the GAGA flexible linker did not improve affinity tag accessibility.

The eluted StrepII-6His-DNMT3B (Figure 7.4A and Figure 7.4B) was further purified by Streptactin affinity chromatography and StrepII-6His-DNMT3B eluted with greater than 80% purity (Figure 7.4C). However, as with the IMAC purification, more than 90% of StrepII-6His-DNMT3B was in the unbound fraction. Cloning DNMT3B with a C-terminal StrepII and 6His tag was not performed due to time constraints and it potentially hindering the methylase activity of the C-terminal domain (Figure 1.2 Domain organisation of DNMT3B). Therefore, non-affinity based purification methods were investigated to improve recovery and purity of StrepII-6His-DNMT3B.

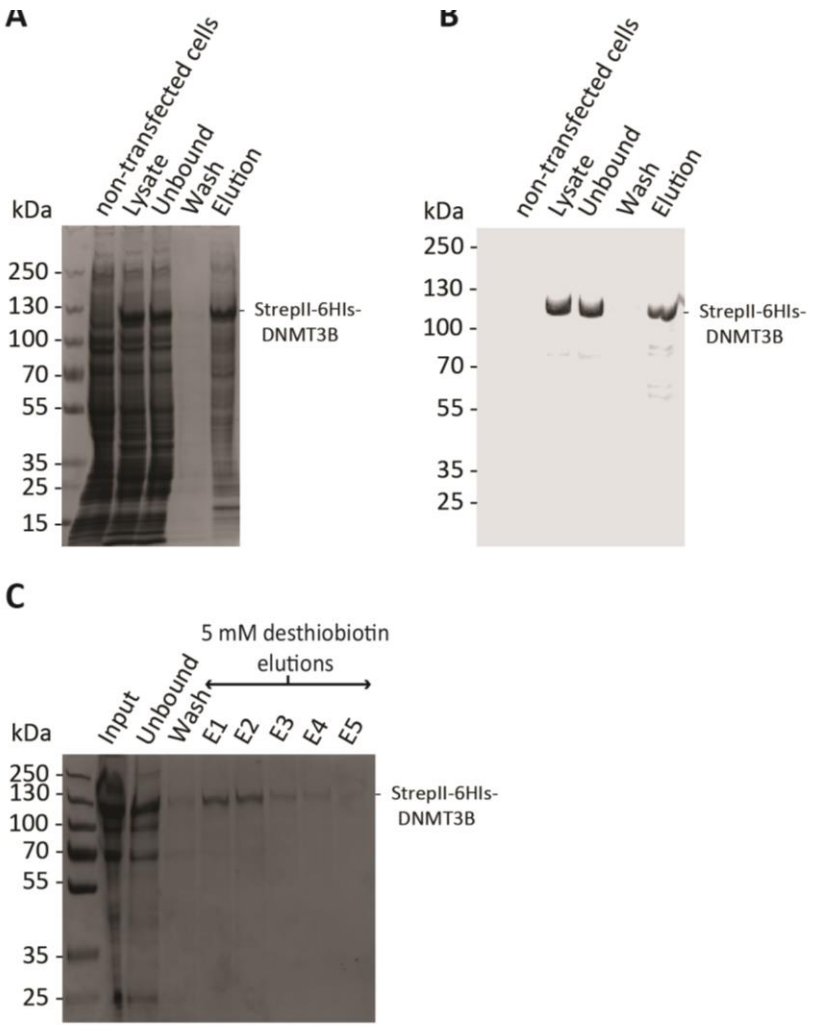

Figure 7.4|Small scale purification of StrepII-6His-DNMT3B using IMAC and Streptactin Sepharose (StrepTrap)

A| Coomassie stained SDS-PAGE gel of StrepII-6His-DNMT3B IMAC purification (0.01% of the total volume loaded per well for the Lysate and Unbound fractions). The wash step contained 40 mM Imidazole, (0.075% of the total volume loaded). The elution step contained 500 mM Imidazole (1% of the total volume loaded).

B| Anti-6His Western blot of the StrepII-6His-DNMT3B IMAC purification from **A|**

C| Coomassie stained SDS-PAGE gel of StrepII-6His-DNMT3B using a 1ml StrepTrap column (0.1% of the total volume loaded per well for the Input and Unbound fractions). 20 ml of lysis buffer was used for the wash step (0.075% of the total loaded). 1 ml elutions with 5 mM desthiobiotin (2% of the total volume loaded).

In previous studies DNMT3B and other DNMT proteins with N-terminal 6His tags were extracted from whole cell extracts with non-ionic detergent and the nuclear fraction (DNMT3B bound to chromatin) separated from the cytoplasmic fractions (Yokochi and Robertson, 2002; Suetake *et al.*, 2003). Non-ionic detergents above the critical micelle concentration (CMC) disrupt cell membranes thereby removing lipid-protein contacts, aiding in cell lysis. Therefore, I wanted to test if isolating the nuclear fraction and purifying StrepII-6His-DNMT3B with 1% (v/v) NP40 detergent improved recovery and purity after IMAC purification.

I lysed StrepII-6His-DNMT3B containing cells with or without 1% NP40 in 100 mM NaCl buffer and pelleted the nuclear fractions. The cytoplasmic fractions with or without 1% NP40 lysis did not contain StrepII-6His-DNMT3B (Figure 7.5A). The nuclear fractions were lysed in 500 mM NaCl buffer (with or without 1% NP40) and centrifuged. The supernatant of both samples contained StrepII-6His-DNMT3B (Figure 7.5A) and so a small scale IMAC purification was performed. StrepII-6His-DNMT3B was present in the unbound fraction and eluted with similar intensity with or without 1% NP40 treatment (Figure 7.5A), therefore 1% NP40 was an unnecessary additive.

Separating the nuclear and cytoplasmic fractions was a useful primary step in reducing lysate contaminants. Therefore, I performed a step-wise salt extraction cell lysis in the low ionic strength buffer, followed by pelleting the nuclear fraction, repeated consecutively with a buffer of increasing ionic strength. The supernatant from each lysis step was analysed by SDS-PAGE (Figure 7.5B). StrepII-6His-DNMT3B begins to dissociate from chromatin in buffer composed of more than 200 mM NaCl (Figure 7.5B). StrepII-6His-DNMT3B was present in 1 M NaCl buffer extractions (Figure 7.5B), indicating that StrepII-6His-DNMT3B is bound tightly to chromatin and 1 M NaCl recovers the majority of StrepII-6His-DNMT3B. Therefore 150 mM NaCl was chosen as the optimal buffer for solubilising proteins whilst leaving StrepII-6HIs-DNMT3B bound to chromatin.

Alongside the step-wise salt extraction I decided to test if ammonium sulphate precipitation could also be used as a native protein purification method to remove contaminants before IMAC. For this reason, an ammonium sulphate (w/v) gradient was performed on StrepII-6His-DNMT3B lysate (Figure 7.5C). Coomassie staining of an SDS-PAGE of the ammonium sulphate supernatant gradient fractions was not sensitive enough to visualise StrepII-6His-DNMT3B (Figure 7.5C), therefore anti-6His Western blotting was performed (Figure 7.5D).

StrepII-6His-DNMT3B was completely precipitated in 30% (w/v) ammonium sulphate with minimal precipitation at 20% (w/v) ammonium sulphate (Figure 7.5D). At the same time the majority of contaminants were retained in the 30% (w/v) ammonium sulphate supernatant. Therefore, I tried a stepwise salt extraction followed by ammonium sulphate precipitation to purify DNMT3B.

A stepwise salt extraction with two rounds of lysis with 150 mM NaCl to remove the majority cytoplasmic proteins and a 1 M NaCl treatment of the nuclear pellet to dissociate StrepII-6His-DNMT3B from chromatin was performed (Figure 7.5E). The 1 M NaCl StrepII-6His-DNMT3B lysate was precipitated with 30% (w/v) ammonium sulphate and both the pellet and supernatant after centrifugation were analysed by SDS-PAGE (Figure 7.5E). The ammonium sulphate precipitation had a resulted in the loss of a large 30 kDa contaminant (Figure 7.5E). However, only 80% of the precipitated StrepII-6His-DNMT3B could be resuspended after ammonium sulphate precipitation and was considered a disadvantageous step.

Ion exchange was contemplated however the NaCl concentration would need to be less than 150 mM which caused more than 50% of StrepII-6His-DNMT3B to precipitate (data not shown) making the method impractical. I concluded that a stepwise salt extraction followed by IMAC purification of StrepII-6His-DNMT3B was the optimal purification procedure.

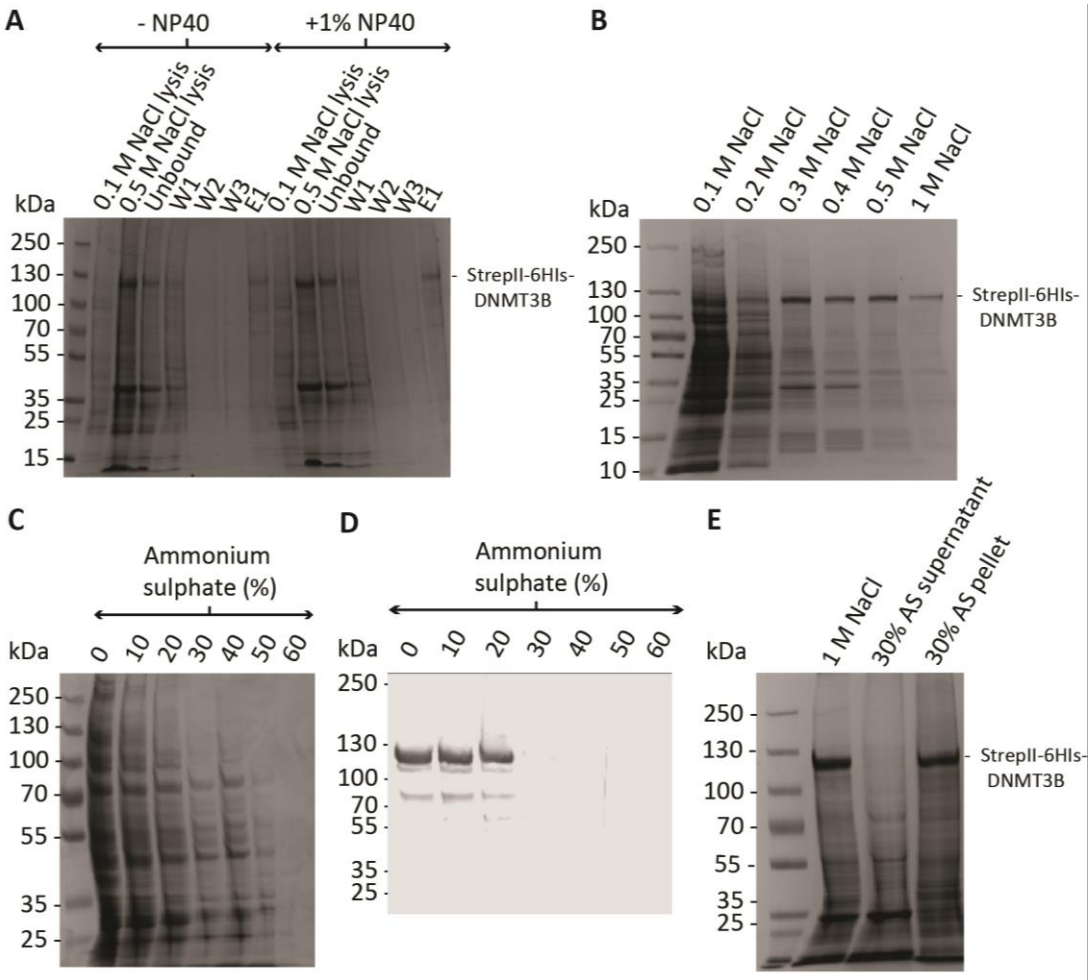

Figure 7.5| Preperative purification of DNMT3B with salt extraction, detergent and ammonium sulphate precipitation

A|Coomassie stained SDS-PAGE of StrepII-6His-DNMT3B IMAC with stepwise salt extraction with or without 1% NP40. (0.1% of the total volume loaded for the 0.1M NaCl, 0.5M NaCl and Unbound fractions. 1% of the total volume of W1-W3 and 5% of E1 fractions loaded).

B| Coomassie stained SDS-PAGE of the step-wise salt extraction of StrepII-6His-DNMT3B. Lysis of cell pellet in 50 mM HEPES pH 7, 10% Glycerol and 1 mM DTT in 0.1M NaCl. Lysate is spun down seperating nuclear pellet from supernatant. Nuclear pellet is lysed in 0.2M NaCl buffer, continuing the procedure until 1M NaCl. 0.4% total volume of loaded per well. (1% of the total volume of each fraction is loaded).

C| Coomassie stained SDS-PAGE of supernatant from ammonium sulphate (0-60%) precipitated StrepII-6His-DNMT3B lysate. (0.4% of the total volume loaded per well for each fraction).

D| Anti-6His Western blot of **C|**.

E| Coomassie stained SDS-PAGE of StrepII-6His-LSH lysate in IM NaCl after a stepwise salt extraction followed by 30% ammonium sulphate precipitation.

7.2.4| Large-scale expression and purification of DNMT3B

Expression of StrepII-6His-DNMT3B was scaled up to 2 litres of Sf9 cells. Cells were lysed by stepwise salt extraction (as described in 7.2.3) and StrepII-6His-DNMT3B, the dominant protein in the lysate (Figure 7.6A) was captured on an IMAC column. Three peaks eluted along the Imidazole gradient (Figure 7.6B). SDS-PAGE showed fractions 1 to 7 and fractions 8 to 11 were contaminant proteins (Figure 7.6B and Figure 7.6C), whereas the third peak eluting at 200 mM Imidazole was StrepII-6His-DNMT3B. The purity was estimated to be greater than 80% (Figure 7.6B and Figure 7.6C). The UV ratio (260/280 nm) was ~0.8 indicative of the presence of DNA. In total 2 mg of StrepII-6His-DNMT3B were purified after IMAC as measured by UV spectroscopy. As before (section 7.22), the unbound IMAC fraction contained StrepII-6His-DNMT3B, decreasing the overall obtainable protein yield. After buffer exchange, the pooled IMAC peak 3 fractions were further purified by SEC (Figure 7.6D). A symmetrical peak (1) eluting in the void volume (47 ml) contained StrepII-6His-DNMT3B and low molecular weight contaminants (Figure 7.6D and Figure 7.6E). The UV ratio (260/280 nm) of 1.2 indicated DNA contamination. However, a broader and smaller peak eluting later at 61 ml also containing StrepII-6His-DNMT3B with greater than 90% purity (Figure 7.6D and Figure 7.6E) had negligible DNA contamination (UV ratio 260/280 nm = 0.65). The total yield of pure, DNA-free StrepII-6His-DNMT3B was 200 µg, with the other 1.8 mg eluting with DNA and the low MW contaminants as measured by UV spectroscopy.

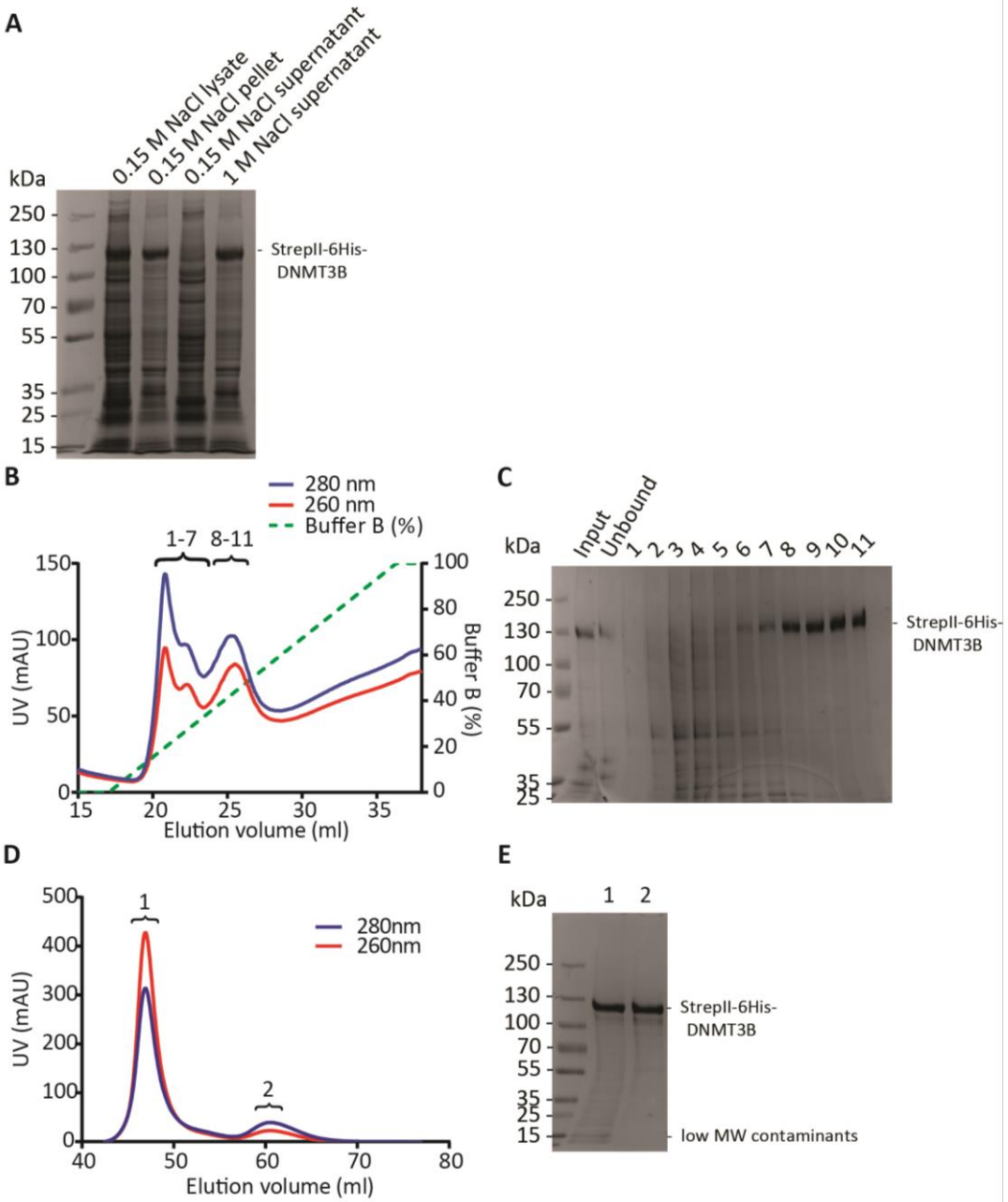

Figure 7.6| StrepII-6His-DNMT3B purification with step-wise salt extraction, IMAC and SEC
A| Coomassie stained SDS-PAGE of a step-wise salt extraction of StrepII-6His-DNMT3B. Lysis of cell pellet in 0.15 M NaCl lysis buffer keeps nuclear fraction in pellet and cytoplasmic fractions in the supernatant. Lysis of nuclear fraction pellet in 1 M NaCl buffer resuspends most proteins into solution. 1% of the total fraction volume loaded.
B| Chromatogram of StrepII-6His-DNMT3B IMAC with UV detection (280 nm and 260 nm) and elution buffer B (%).
C| Coomassie stained SDS-PAGE of IMAC fractions (1-11, 1% of the total fraction volume loaded).
D| Chromatogram of StrepII-6His-DNMT3B SEC (Superdex 200 16/60 pg) with UV detection (280 nm and 260 nm).
E| Coomassie stained SDS-PAGE of SEC fractions 1 (47 ml) and 2 (61 ml). 1% of the total fraction volume loaded.

It was hypothesised that the low MW contaminants in peak 2 (Figure 7.7A) might be histones due to their migration on SDS-PAGE (Figure 7.7B). Anti-H3 Western blotting confirmed the presence of Histone H3 in peak 1 as well as StrepII-6His-DNMT3B (Figure 7.7C). As peak 1 also contains DNA, it was hypothesised StrepII-6His-DNMT3B was bound to chromatin and was eluting as a large complex. To test the presence and size of DNA, peak 1 fractions were treated with proteinase K to digest proteins, thus dissociating DNA (Figure 7.7D) and analysed by AGE. DNA in the range of 350-1350 bp was detected after proteinase K treatment (Figure 7.7E). This could suggest a large oligomeric complex or aggregation of StrepII-6His-DNMT3B with chromatin. Binding of DNMT3B to H3 tails (Zhang *et al.*, 2010) could maintain histone-DNA contacts in high ionic strength buffers (1 M NaCl). The DNA-free sample of StrepII-6His-DNMT3B (peak 2) was used to test DNA and nucleosome binding activity and in attempts to reconstitute an LSH-DNMT3B complex *in vitro*.

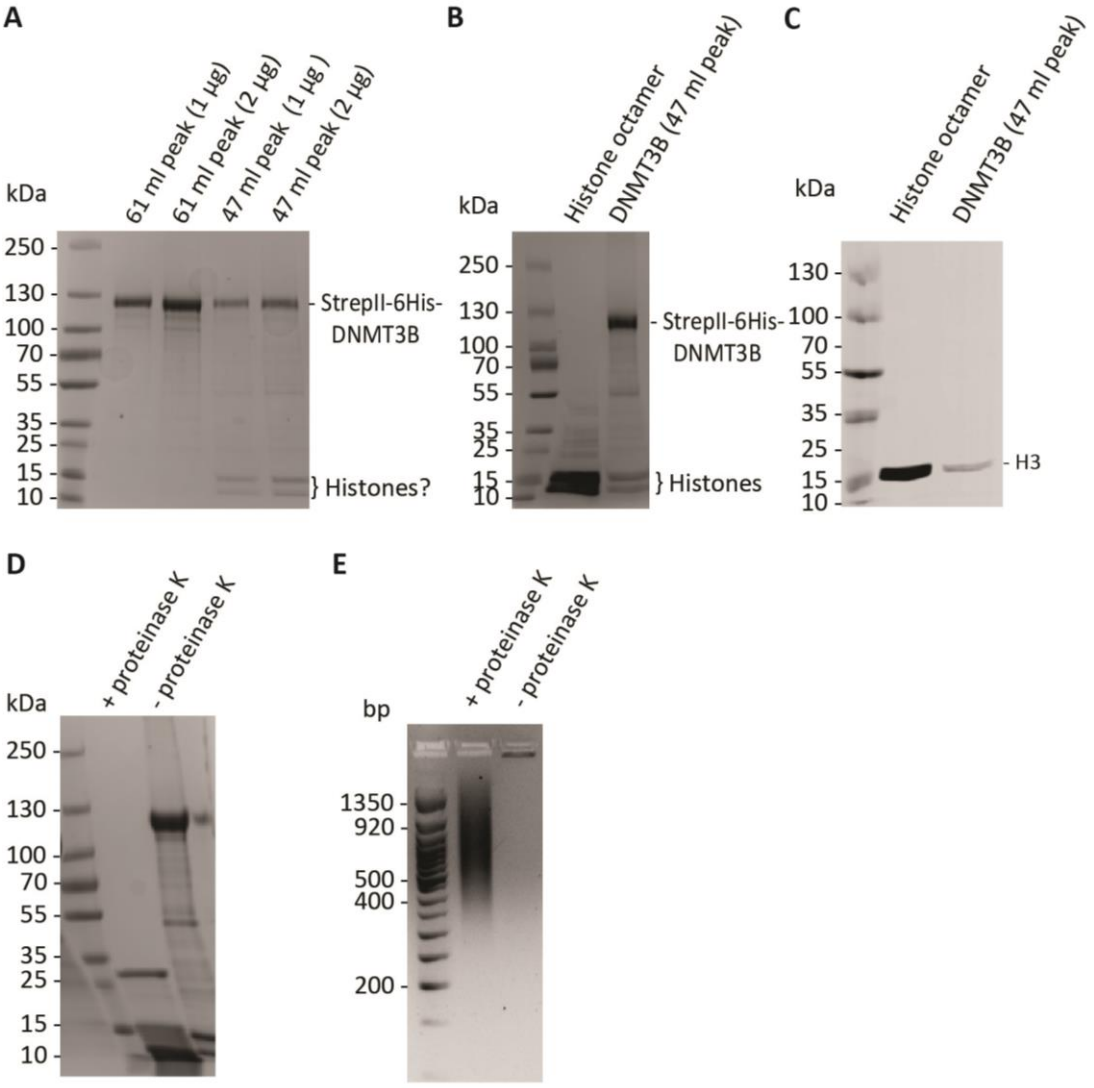

Figure 7.7| Presence of histones and DNA in the 47 ml peak of StrepII-6His-DNMT3B SEC
A| Coomassie stained SDS-PAGE gel of DNMT3B SEC-peaks of DNMT3B eluting at 47 ml and 61 ml. The 47 ml elution has contaminants speculated to be histones based on the molecular weight.
B| Coomassie stained SDS-PAGE gel of histone octamers and the 47 ml SEC fraction of StrepII-6His-DNMT3B.
C| Anti-H3 Western blot of histone octamers and the 47 ml SEC fraction of StrepII-6His-DNMT3B.
D| Coomassie stained SDS-PAGE gel of StrepII-6His-DNMT3B with or without proteinase K treatment.
E| 2% AGE of StrepII-6His-DNMT3B with or without proteinase K treatment.

7.2.5| StrepII-6His-DNMT3B nucleosome binding

To establish if StrepII-6His-DNMT3B is functional I tested if it could bind nucleosomes using an EMSA performed with symmetric mononucleosomes with or without 25 bp overhangs (Figure 7.8). StrepII-6His-DNMT3B shifted both types of mononucleosomes (Figure 7.8), suggesting that overhangs are not required for binding. Free DNA binding is also evident, as the residual free DNA migration below the nucleosome band in samples without DNMT3B disappears in the presence of 20 nM StrepII-6His-DNMT3B (Figure 7.8A and Figure 7.8B). Therefore StrepII-6His-DNMT3B may bind to histone octamer bound DNA and/or the H3 tail of the core mononucleosome via the ADD and PWWP domains (H3 binding) and the MT domain (DNA and SAM binding) (Cheng 2014; Cheng & Blumenthal 2008, Gowher et al. 2005). The K_D of DNMT3B interaction with 147 bp mononucleosomes was ~80 nM and ~20 nM for 200 bp mononucleosomes, however, a direct comparison was difficult given nucleosome concentrations were not accurately measured. A previous study found DNMT3B binds to DNA with a K_D of 183 nM (Van Emburgh and Robertson, 2011). This is a weaker K_D than my results, however, they did not purify a homogenous sample using SEC, therefore activity cannot be accurately compared.

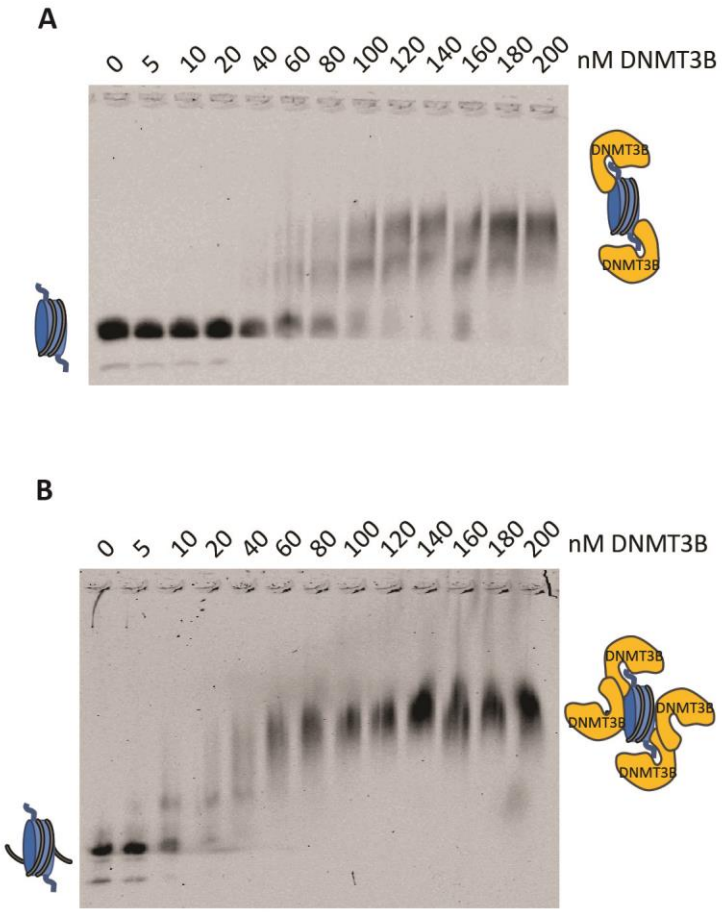

Figure 7.8| EMSA of StrepII-6His-DNMT3B and mononucleosomes
A| EMSA of 5 nM 147bp mononucleosome with increasing concentrations of StrepII-6His-DNMT3B (5-200 nM)
B| EMSA of 5 nM 200bp mononucleosome with increasing concentrations of StrepII-6His-DNMT3B (5-200 nM)

7.2.6| Reconstituting DNMT3B-LSH complex *in vitro* in a Streptactin pull-down assay

In the previous section, StrepII-6His-DNMT3B was shown to bind DNA and nucleosomes. Next, I wanted to test if DNMT3B-LSH complex could be formed *in vitro*. I used a pull-down assay with StrepII-6His-DNMT3B as the bait protein immobilised on Streptactin agarose beads. Low ionic strength buffer containing 50 mM Hepes pH 7.0, 100 mM NaCl, 10% (v/v) glycerol, 1 mM DTT was used as this was successful for the complex stability of co-expressed LSH and DNMT3B (Figure 7.2).

StrepII-6His-DNMT3B (2 µg in 50 µl volume) did not elute in 10 mM desthiobiotin bound to Streptactin agarose beads (Figure 7.9), which might be due DNMT3B aggregation in the low ionic strength buffer or StrepII-6His-DNMT3B binding to the agarose beads with high affinity. LSH (2 µg in 50 µl) did not bind to the beads as the majority of protein was in the flowthrough and wash fractions, with minor amounts visible on the Streptactin agarose beads (Figure 7.9).

Alternative elution methods, including increasing dethiobiotin to 50 mM or using Streptactin recharging buffer containing Glycine (pH 3) were tested but did not improve the elution of DNMT3B (data not shown). Therefore, I performed SDS-PAGE of the Streptactin agarose beads to analyse potential complexes.

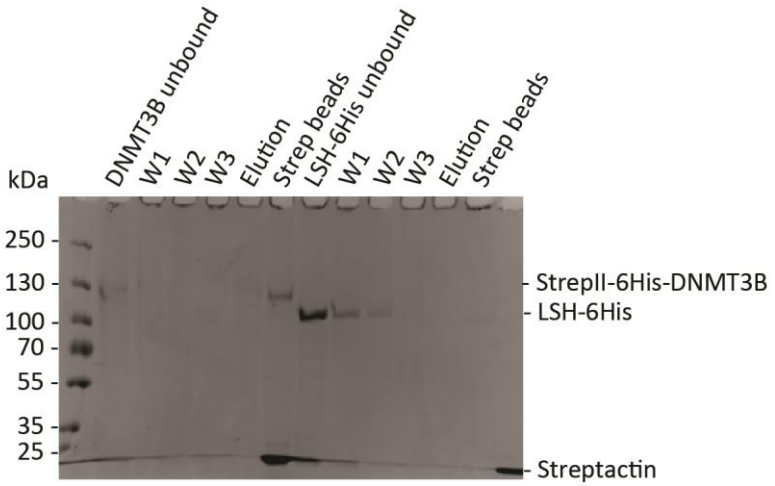

Figure 7.9| StrepII-6His-DNMT3B and LSH-6His binding and elution to Streptactin agarose. Coomassie stained SDS-PAGE gel of StrepII-6His-DNMT3B and LSH-6His binding and elution to Streptactin agarose (100% of DNMT3B loaded on gel). Binding buffer was composed of 50 mM HEPES pH 7, 100 mM NaCl, 10% Glycerol and 1 mM DTT. W1-W3 were 1 ml washes of the Streptactin agarose beads using binding buffer (1.5% of the total volume of W1-W3 was loaded on the gel). The Elution steps contained binding buffer with 10 mM desthiobitoin (50% of elution loaded on the gel). The Strep beads was the Streptactin agarose beads (100% of the strep beads loaded on the gel).

A streptactin agarose pulldown assay with 2 µg of StrepII-6His-DNMT3B as the immobilised bait protein was performed with 2 µg of LSH-6His with a 30-minute mixing incubation at 6°C. LSH^{1-176} was also tested as this region was previously shown to be necessary for DNMT3B binding to LSH (Myant and Stancheva, 2008). As a negative control I used StrepII-tagged MicroTubule Organiser 1 (StrepII-Mto1, a gift from the Sawin lab) expressed in insect cells. This yeast cytoplasmic protein was unlikely to bind LSH-6His or LSH^{1-176} as these are nuclear proteins.

After incubation and washing steps (5 times with 500 µl buffer) LSH was present on Streptactin agarose beads baited with StrepII-6His-DNMT3B or StrepII-Mto1 or not baited (Figure 7.10A). However, because of the non-specific binding of LSH-6His to Streptactin agarose, the results are inconclusive. The result was the same for the LSH^{1-176} pulldown and in 3 repeats (Figure 7.10B). To reduce or abolish LSH-6His and LSH^{1-176} non-specific binding to Streptactin agarose beads, BSA or Triton X-100 were added to the buffer, however, this did not block binding (data not shown).

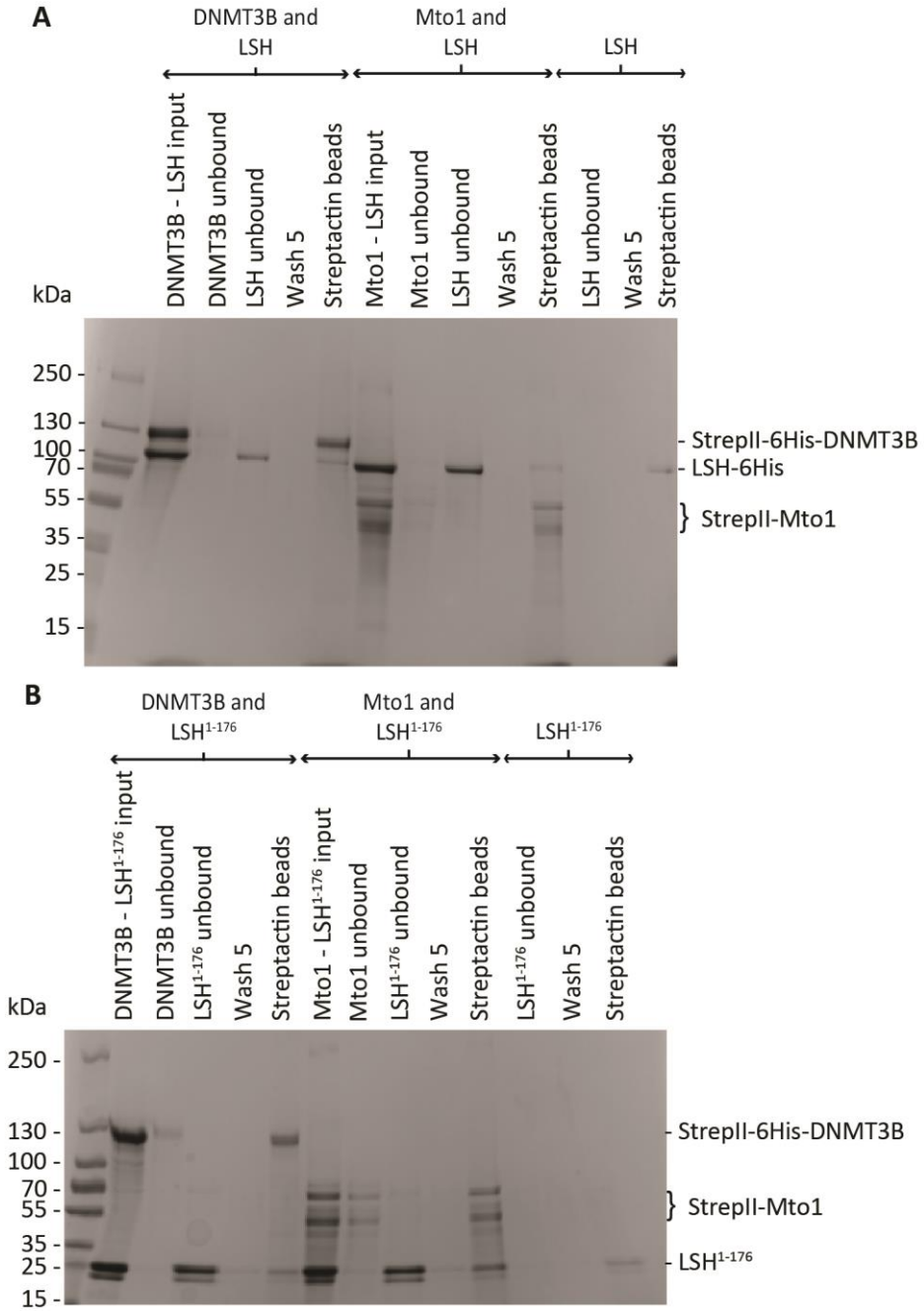

Figure 7.10| StrepII-6His-DNMT3B Streptactin agarose pulldown with LSH-6His or LSH$^{1-176}$
A|Coomassie stained SDS-PAGE gel of StrepII-6His-DNMT3B Streptactin agarose pulldown of LSH-6His Binding and wash buffer (Wash 5) was composed of 50 mM HEPES pH 7, 100 mM NaCl, 10% Glycerol and 1 mM DTT. StrepII-Mto1 was used as a negative control for LSH-6His binding. Non-specific binding to Streptactin agarose beads was tested with LSH-6His only.
B|Coomassie stained SDS-PAGE gel of StrepII-6His-DNMT3B Streptactin agarose pulldown of LSH$^{1-176}$. Binding and wash buffer (Wash 5) was composed of 50 mM HEPES pH 7, 100 mM NaCl, 10% Glycerol and 1 mM DTT. StrepII-Mto1 was used as a negative control for LSH$^{1-176}$ binding. Non-specific binding to Streptactin agarose beads was tested with LSH$^{1-176}$ only.
2 µg of each protein was loaded as the input. The unbound fractions are 100% of the sample volume after the incubation with streptactin resin. 3% of the total sample volume for Wash 5 was loaded on the gel.

As an alternative strategy I wanted to perform a Ni-NTA pull-down with DNMT3B using LSH-6His as the immobilised bait protein. In order to do so, the 6His tag of StrepII-6His-DNMT3B had to be cleaved, but it could not be removed with TEV cleavage, likely due to TEV inaccessibility (data not shown). Alternatively, the Ni-NTA pull-down could be performed using StrepII-6His-DNMT3B as the immobilised bait protein. However, LSH-6His did not have a TEV cleavage site, so LSH-6His could not be used. On the other hand, LSH^{1-176} did have the 6His tag successfully removed by TEV cleavage, therefore Ni-NTA pulldown with StrepII-6His-DNMT3B as the immobilised bait protein and LSH^{1-176} could be attempted.

7.2.7| Reconstituting a DNMT3B-LSH complex in a Ni-NTA pulldown assay

A Ni-NTA pulldown assay with StrepII-6His-DNMT3B immobilised via the 6His tag was performed with LSH^{1-176} (6His cleaved) under low ionic strength buffer conditions used in the streptactin agarose pulldown (Section 7.2.6). The LSH-6His was used as a negative control as it was unlikely to form a heterodimer with LSH^{1-176}. StrepII-6His-DNMT3B could be eluted from the NTA beads with 500 mM Imidazole giving an advantage over streptactin agarose beads.

LSH^{1-176} was present in elutions from Ni-NTA agarose beads baited with StrepII-6His-DNMT3B or LSH-6His or not baited (Figure 7.11A). An anti-6His and anti-NLSH Western blot was performed to measure the intensity of proteins in the elution fractions (Figure 7.11B). The strongest signal for LSH^{1-176} was seen in the Ni-NTA agarose-beads-only elution (Figure 7.11B). Therefore, there was no definitive evidence of complex formation between StrepII-6His-DNMT3B and LSH^{1-176} under the conditions tested.

However, the assays have several limitations. Firstly a pulldown assay tethers the bait protein. In this study immobilising the N-terminus of StrepII-6His-DNMT3B might block access to the N-terminal domain which could be necessary for the interaction with LSH or LSH^{1-176}. Secondly, in the pulldown assays, 2 µg of each protein were used and each wash volume was 500 µl, this makes a concentration of 0.01 mg/ml. This is 100 nM LSH-6His or 500 nM LSH^{1-176} and if the K_D of the complex is weaker than 100 nM, a complex cannot form and would not be detected with the pulldown assay. Furthermore, sequentially diluting a complex 5 times in 500 µl wash buffer would allow LSH-6His to dissociate each time reducing the concentration of the complex each time. Therefore, an alternative assay which does not tether a protein and allows protein concentrations in the µM range to be used was investigated.

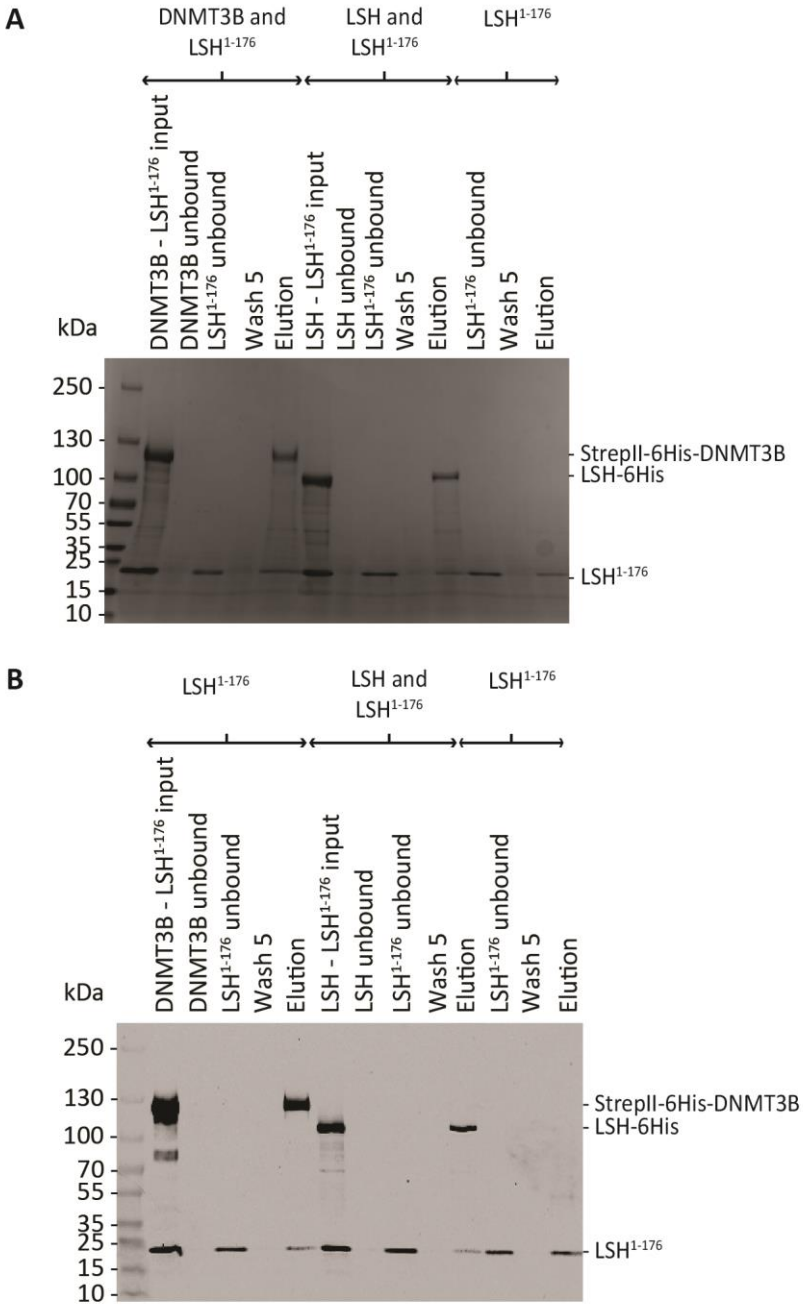

Figure 7.11|Ni-NTA pulldown of StrepII-6His-DNMT3B with LSH$^{1-176}$

A|Coomassie stained SDS-PAGE gel of StrepII-6His-DNMT3B Ni-NTA pulldown of LSH$^{1-176}$. LSH-6His was used as a negative control for LSH$^{1-176}$ binding. LSH$^{1-176}$ used on its own to test non-specific binding to Ni-NTA beads. W5 was wash 5 of the Ni-NTA beads using 50 mM HEPES pH 7, 100 mM NaCl, 10% Glycerol and 1 mM DTT. 500 mM Imidazole was used for all elutions.

2 µg of each protein was loaded as the input. The unbound fractions are 100% of the sample volume after the incubation with streptactin resin. 3% of the total sample volume for Wash 5 was loaded on the gel. The Elution is 100% of the elution buffer with Ni-NTA beads.

B|Anti-6His and anti-NLSH Western blot of StrepII-6His-DNMT3B Ni-NTA pulldown of LSH$^{1-176}$.

7.2.8| Native AGE of StrepII-6His-DNMT3B and LSH

A native AGE method used for multiprotein complexes was chosen because of the small sample volumes, quick assay time (Kim, 2011) and no complications from tethering proteins to beads. Moreover, the protein concentrations used in this assay can be in low µM range (rather than low nM used in the pulldown assays). For native AGE, sample wells are in the centre of a thin horizontal agarose gel (Kim, 2011). Proteins with a pI lower than the running buffer migrate to the anode due to a net negative charge, whilst proteins with a pI higher than the running buffer migrate to the cathode due to a net positive charge (Kim, 2011). As it is a native assay, protein mass also influences migration (speed of movement) through the gel (Kim, 2011). If an interaction between two proteins occurs, the overall charge and migration of the complex are different to that of individual proteins (Kim, 2011). Proteins are visualised by Coomassie staining. A 0.5% agarose gel was used for LSH-6His and StrepII-6His-DNMT3B which are both ~100 kDa to allow the greatest migration. The pH of the agarose gel and running buffers tested ranged from 6.1 to 9.4. Tris-CAPS buffer at pH 9.4 enabled the proteins to migrate far enough to visualise a difference between LSH-6His and StrepII-6His-DNMT3B (Figure 7.12). This buffer was chosen to test complex formation of LSH-6His or LSH^{1-176} with StrepII-6His-DNMT3B (Figure 7.12). A single band migrating inbetween the individual StrepII-His-DNMT3B and LSH-6His proteins was seen (Figure 7.12), suggesting that a LSH-6His and StrepII-6His-DNMT3B complex may have formed. However, the resolution of the native AGE made it difficult to confirm a clear shift. Increasing gel running time to enhance separation of each protein created undefined smears. Given the predicted pI's of DNMT3B and LSH are similar (8.2 and 8.35 respectively) this method would not enable a shift of the individual proteins great enough to define complex formation.

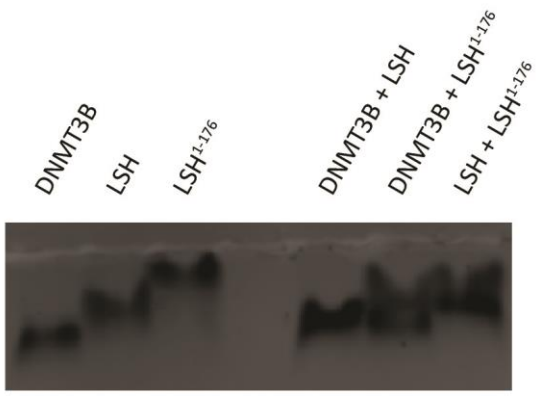

Figure 7.12| Native agarose gel electrophoresis to examine complex formation of StrepII-6His-DNMT3B with LSH-6His or LSH$^{1-176}$
0.5% agarose gel made with Tris-CAPS buffer (pH 9.4). 1 µg of each protein was used in a 10 µl reaction containing binding buffer (50 mM HEPES pH 7, 100 mM NaCl, 10% Glycerol and 1 mM DTT).

7.3| Discussion and future work

Recently the crystal structure of the DNMT3B PWWP domain in complex with H3K36me3 was been solved (Rondelet *et al.*, 2016). This work has enabled a proposed model in which DNMT3B may be directed to genomic sites of DNA for methylation. In the model, the PWWP domain binds H3K36me3, which in turn enables the ADD domain to bind H3K4 which prevents the inhibition of the catalytic C-terminal domain which is free to methylate DNA (Rondelet *et al.*, 2016).

CHIP based experiments have shown that in ES cells, LSH promotes DNMT3B association at Oct4 and Nanog promoters (Xi *et al.*, 2009). Co-immunoprecipitation and pulldown experiments have shown LSH binds to DNMT3B, suggesting LSH recruits DNMT3B directly to chromatin targets (Zhu et al. 2006, Myant & Stancheva 2008). More recent CHIP experiments have shown LSH with an active ATPase domain is essential for *de novo* methylation and recruiting DNMT3B at minor satellite sequences, retroviral repeat sequence (IAP) and LINES (Ren *et al.*, 2015). This suggests the chromatin remodelling activity of LSH is required for the DNMT3B association to repeat elements, in which LSH may also act as a scaffolding protein (Ren *et al.*, 2015).

It could be speculated that LSH may remodel chromatin for a favourable environment in which PWWP and ADD domains of DNMT3B bind to their respective substrates activating the catalytic methyltransferase domain or may interact directly with DNMT3B.

The pulldown and CHIP experiments describing the localisation of LSH and DNMT3B at the same genomic locations provide evidence to support LSH and DNMT3B could interact directly. We aimed to study if LSH and DNMT3B interact directly *in vitro* using purified recombinantly expressed LSH and DNMT3B from a baculoviral system. From the current work performed there is no evidence to support a direct interaction between recombinant LSH and DNMT3B, however not all experiments were tested due to time constraints.

Both proteins could be co-expressed but could not be co-purified with equivalent concentrations (section 7.22). The expression and purification of both proteins separately was successful, although yields of StrepII-6His-DNMT3B were low due to loss during the purification steps. Formation of a DNMT3B-LSH complex *in vitro* did not occur with LSH-6His or LSH^{1-176} using a pulldown assay or native AGE. If more time allowed I would perform the following experiments:

1| Co-expression and purification of LSH and DNMT3B.

A stable LSH-DNMT3B complex might only form in the nucleus of a cell with other interacting partners, therefore co-expression of both proteins could be revisited. Co-infection of LSH-6His and StrepII-6His-DNMT3B bacmids in insect cells and IMAC purification could be used to purify both proteins.

2| Studying a LSH-DNMT3B complex *in vitro*.

Analytical SEC of LSH and DNMT3B to test for the formation of a LSH-DNMT3B complex using small volumes resulting in low dilutions allowing protein concentrations in the µM range should be carried out. Once a complex is confirmed *In vitro* then Isothermal titration calorimetry can be used to determine a K_D for a LSH-DNMT3B complex.

DNMT3B recovery was low during affinity purification steps (section 7.23 and 7.24). For structural and biochemical studies of DNMT3B in complex with LSH, greater yields of more than 1 mg of pure DNMT3B would be required.

Cloning DNMT3B with a large cleavable tag such as Glutathione S-Transferase (GST) may improve recovery as it is larger (26 kDa) than a 6His tag (~1 kDa) and is unlikely to be occluded within the protein. Contaminating DNA was also an issue with DNMT3B purification, therefore lowering the NaCl concentration to a concentration which prevents DNMT3B precipitation and test the separation of DNA from DNMT3B via a heparin column purification which may compete away the DNA from DNMT3B due to the anionic sulphate groups which mimic the polyanionic nature of nucleic acid. Alternatively, given DNMT3B is only soluble in relatively high NaCl concentrations Hydrophobic interaction chromatography (HIC) could be used as an alternative purification strategy to purify the protein and remove DNA contamination. Given DNMT3B precipitates in ≥ 30% saturated ammonium sulphate keeping DNMT3B soluble in 20% saturated ammonium sulphate (0.82 M) and performing HIC along a reducing ammonium sulphate gradient could be a useful purification strategy.

3| Revisit pulldown experiments of LSH-DNMT3B

Previous evidence to support a direct LSH-DNMT3B interaction was a GST-LSH$^{1-503}$ pulldown with DNMT3B (Myant and Stancheva, 2008). However, LSH^{503} lies between ATPase lobes 1 and 2 (Figure 1.8) and therefore might not be folded or functional. Therefore it would be useful to test if LSH$^{1-503}$ is actually folded and can bind DNMT3B using the same pulldown assay as performed (Myant and Stancheva, 2008). A direct comparison with LSH-6His or LSH$^{1-176}$ would be necessary to determine if DNMT3B can bind full-length LSH and if the N-terminal domain is necessary for the interaction.

Other evidence to support an LSH-DNMT3B complex was LSH co-immunoprecipitation of DNMT3B (Zhu *et al.*, 2006; Myant and Stancheva, 2008) from nuclear lysate and therefore other interacting partners could have influenced binding. Affinity purification of a Flag-LSH or Strep-LSH expressed in embryonic stem cells (ESC) could be performed to pull down interacting proteins and Mass spectrometry could be used to identify them. This experiment would indicate what proteins may interact with LSH and some of them may be required for a LSH-DNMT3B interaction.

7.4| Conclusion

In order to study a LSH-DNMT3B complex *in vitro*, StrepII-DNMT3B was successfully co-expressed with LSH-6His using the MultiBac system. A DNMT3B-LSH complex failed to co-purify, therefore the reconstitution of a complex was attempted after purifying both proteins separately. StrepII-DNMT3B was cloned with an N-terminal 6His tag to aid purification. Expression of StrepII-6His-DNMT3B was increased 6-fold under the control of the polH promoter. A stepwise salt extraction removing cytoplasmic proteins improved the purification procedure of StrepII-6His-DNMT3B. Pure StrepII-6His-DNMT3B could bind DNA and core nucleosomes. A LSH-DNMT3B complex could not be detected using a pulldown assay or native AGE assay, suggesting complex formation occurs in µM concentrations or is mediated by other proteins.

Chapter 8| Conclusion and future outlook

There is little biochemical and no structural data on how LSH functions. The main aim of this work was to determine the structural characteristics and to elucidate the role N and C terminal domains have in regulating LSH enzymatic activity *in vitro* using biochemical, biophysical and structural methods.

Optimised expression and purification enabled a 25-fold increase in LSH-6His protein from 0.2 mg/L cells to 5 mg/L cells. Purified LSH-6His was homogeneous and a monomer which is in agreement with native LSH purified from HeLa cells eluting as a monomer from SEC (Myant & Stancheva 2008). The N-terminal domain (LSH^{1-178}) was found to be hinged on an unstructured region to $LSH^{178-821}$ and SAXS has shown LSH^{1-176} is disordered and elongated, yet takes a more compact shape in full length LSH. In order to dissect how the N-terminal domain may positively or negatively regulate the ATPase domain, the ATPase region was cloned. However the ATPase domain was poorly expressed and unstable preventing further study to be undertaken. LSH has a tri-lobal and elongated shape as determined by SAXS and negative staining EM. XL-MS suggest the N-terminal domain is in contact with ATPase lobe 1 and the region between ATPase lobe 2 and the C-terminal domain.

In $LSH^{-/-}$ cells the re-expression of wild-type LSH is required for restoring nucleosome and is dependent on functional ATP binding (Ren *et al.*, 2015). Chromatin remodelling by LSH is suggested to promote *de novo* DNA methylation at repeat elements, which has been indicated through the direct interaction of LSH with DNMT3B (Myant and Stancheva, 2008). Therefore the primary function of LSH is to remodel chromatin. LSH-6His can bind DNA and nucleosomes *in vitro*, with negligible binding to ssDNA, indicating the recombinant protein is active. Under the conditions tested LSH does not show any difference in affinity for methylated or un-methylated DNA suggesting the DNA methylation status of DNA is not important for recruiting LSH to particular regions on chromatin. LSH DNA binding occurs in solution only in physiological NaCl concentrations providing an important pre-requisite for the further study of LSH with nucleosomes.

However, it is unknown if LSH can remodel chromatin on its own, and how the N and C-terminal domains influence its activity. Therefore future work is required to determine if LSH alone can remodel chromatin *in vitro* using nucleosome sliding assays. ATPase activity assays and chromatin remodelling assays with and without its N and/or C terminal domain would allow us to determine if these domains are involved in enzymatic activity.

The second aim of this work was to test if LSH and DNMT3B interact directly *in vitro* and what effect such an interaction may have on LSH and on *de novo* DNA methylation by DNMT3B. Co-expression of LSH and DNMT3B in a baculoviral expression system was successful, however, DNMT3B expression was weak and purification methods could have been disrupting an interaction. Therefore DNMT3B expression and purification was optimised independently of LSH. No interaction between LSH or LSH^{1-176} with DNMT3B could be observed with pulldown assays, therefore other accessory proteins and/or DNA binding may be required for complex formation.

Further work is required to establish which protein complexes LSH is involved with. Defining which proteins or protein complexes can be pulled down with LSH and using mass spectrometry to identify them could provide insight into how LSH might be targeted to specific sites in chromatin. This could improve our understanding of LSH involvement in *de novo* DNA methylation and other regulatory roles of LSH.

Structural studies with LSH with nucleosomes and/or interacting proteins would provide insight into how the protein functions as a chromatin remodeler. I propose using high resolution cryo-EM and complementary methods (SAXS and XL-MS) to achieve this.

This work has provided important preliminary studies and a foundation into understanding how LSH functions at a molecular level through separate study of LSH structure, its interactions with DNA, nucleosomes and DNMT3B.

References|

Adams, V.H. et al., 2007. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. Journal of Biological Chemistry, 282(20), pp.15057–15064.

Al-Ani, G. et al., 2014. Quantitative determination of binding of ISWI to nucleosomes and DNA shows allosteric regulation of DNA binding by nucleotides. Biochemistry, 53(27), pp.4334–4345.

Alkhatib, S. G. and Landry, J. W. (2011) 'The nucleosome remodeling factor', FEBS Letters.

NIH Public Access, pp. 3197–3207.

Almo, S.C. et al., 2013. Protein production from the structural genomics perspective: Achievements and future needs. *Current Opinion in Structural Biology*, 23(3), pp.335–344.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes and Development, 16(1), pp.6–21.

Bird, A.P. & Wolffe, A.P., 1999. Methylation-Induced Repression— Belts, Braces, and Chromatin. Cell, 99(5), pp.451–454.

Bird, A.P., 1986. CpG-rich islands and the function of DNA methylation. Nature, 321(6067), pp.209–213.

Booth, D.S., Avila-Sakar, A. & Cheng, Y., 2011. Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. *Journal of visualized experiments : JoVE*, (58).

Boyer, L.A., Latek, R.R. & Peterson, C.L., 2004. The SANT domain: a unique histone-tail-binding module? Nature reviews. Molecular cell biology, 5(2), pp.158–163.

Briones, V. & Muegge, K., 2012. The ghosts in the machine: DNA methylation and the mystery of differentiation. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms, 1819(7), pp.757–762.

Brown, M. & Faulkner, P., 1977. A plaque assay for nuclear polyhedrosis virus using a solid overlay. *Journal of General Virology*, 36(May), pp.361–364.

Brownell, J.E. et al., 1996. Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell, 84(6), pp.843–851.

Bugreev, D. V., Mazina, O.M. & Mazin, A. V., 2006. Rad54 protein promotes branch migration of Holliday junctions. Nature, 442(7102), pp.590–593.

Burgess, R.J. & Zhang, Z., 2014. Histone chaperones in nucleosome assembly and human disease. Nat Struct Mol Biol., 20(1), pp.14–22.

Burrage, J. et al., 2012. The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. Journal of Cell Science, 125(22), pp.5524–5534.

Cai, Y. et al., 2007. YY1 functions with INO80 to activate transcription. Nature Structural & Molecular Biology, 14(9), pp.872–874.

Cairns, B.R. et al., 1994. A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. Proceedings of the National Academy of Sciences of the United States of America, 91(5), pp.1950–4.

Cairns, B.R. et al., 1996. RSC , an Essential , Abundant Chromatin-Remodeling Complex. Cell, 87, pp.1249–1260.

Callebaut, I. et al., 1997. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): Current status and perspectives. Cellular and Molecular Life Sciences, 53(8), pp.621–645.

Campos, E.I. et al., 2010. The program for processing newly synthesized histones H3.1 and H4. Nature structural & molecular biology, 17(11), pp.1343–51.

Chang, Y. et al., 2011. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. Nature communications, 2, p.533.

Chatterjee, N., Sinha, D., Lemma-Dechassa, M., Tan, S., Shogren-Knaak, M. A. and Bartholomew, B. (2011) 'Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms', Nucleic Acids Research. Oxford University Press, 39(19), pp. 8378–8391.

Chen, J. & Archer, T.K., 2005. Regulating SWI / SNF Subunit Levels via Protein-Protein Interactions and Proteasomal Degradation : BAF155 and BAF170 Limit Expression of BAF57 Regulating SWI / SNF Subunit Levels via Protein-Protein Interactions and Proteasomal Degradation : BAF155 and BA. Gene Expression, 25(20), pp.9016–9027.

Chen, T., Tsujimoto, N. & Li, E., 2004. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol, 24(20), pp.9048–9058.

Chen, Z.A. et al., 2010. Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. The EMBO Journal, 29(4), pp.717–726.

Cheng, X. & Blumenthal, R.M., 2008. Mammalian DNA methyltransferases: a structural perspective. *Structure (London, England : 1993)*, 16(3), pp.341–50.

Cheng, X., 2014. Structural and functional coordination of DNA and histone methylation. *Cold Spring Harbor perspectives in biology*, 6(8).

Cheng, Y., 2015. Single-particle Cryo-EM at crystallographic resolution. *Cell*, 161(3), pp.450–457.

Chua, E.Y.D. et al., 2012. The mechanics behind DNA sequence-dependent properties of the nucleosome. Nucleic Acids Research, 40(13), pp.6338–6352.

Ciferri, C. et al., 2012. Molecular architecture of human polycomb repressive complex 2. *eLife*, 2012(1), pp.250–261.

Clapier, C.R. & Cairns, B.R., 2009. The biology of chromatin remodeling complexes. Annual

review of biochemistry, 78, pp.273–304.

Clapier, C.R. & Cairns, B.R., 2012. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. Nature, 492(7428), pp.280–4.

Collins, N. et al., 2002. An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. Nature genetics, 32(4), pp.627–32.

Cook, A.J.L. et al., 2011. A Specific Function for the Histone Chaperone NASP to Fine-Tune a Reservoir of Soluble H3-H4 in the Histone Supply Chain. Molecular Cell, 44(6), pp.918–927.

Corona, D.F.V. & Tamkun, J.W., 2004. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 1677(1), pp.113–119.

Cosgrove, M.S. & Wolberger, C., 2005. How does the histone code work? Biochemistry and cell biology = Biochimie et biologie cellulaire, 83, pp.468–476.

Cosgrove, M.S., Boeke, J.D. & Wolberger, C., 2004. Regulated nucleosome mobility and the histone code. Nature structural & molecular biology, 11(11), pp.1037–43.

Côté, J. et al., 1994. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science (New York, N.Y.), 265(5168), pp.53–60.

Courilleau, C., Chailleux, C., Jauneau, A., Grima, F., Briois, S., Boutet-Robinet, E., Boudsocq, F., Trouche, D. and Canitrot, Y. (2012) 'The chromatin remodeler p400 atpase facilitates RAD51-mediated repair of DNA double-strand breaks', Journal of Cell Biology, 199(7), pp. 1067–1081.

Cox, J. & Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26(12), pp.1367–72.

Dambacher, C.M. & Lander, G.C., 2015. Site-specific labeling of proteins for electron microscopy. *Journal of Structural Biology*, 192(2), pp.151–158.

Dang, W., Kagalwala, M. N. and Bartholomew, B. (2006) 'Regulation of ISW2 by Concerted Action of Histone H4 Tail and Extranucleosomal DNA', Molecular and Cellular Biology. American Society for Microbiology, 26(20), pp. 7388–7396.

De Greef, J.C. et al., 2011. Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. American Journal of Human Genetics, 88(6), pp.796–804.

Dechassa, M.L. et al., 2012. Disparity in the DNA translocase domains of SWI/SNF and ISW2. Nucleic Acids Research, 40(10), pp.4412–4421.

Dee, K.U., Shuler, M.L. & Wood, H.A., 1997. Inducing single-cell suspension of BTI-TN5B1-4 insect cells: I. The use of sulfated polyanions to prevent cell aggregation and enhance recombinant protein production. *Biotechnology and Bioengineering*.

Dekker, J. (2008) 'Mapping in vivo chromatin interactions in yeast suggests an extended

chromatin fiber with regional variation in compaction', Journal of Biological Chemistry. American Society for Biochemistry and Molecular Biology, 283(50), pp. 34532–34540.

Dhalluin, C. et al., 1999. Structure and ligand of a histone acetyltransferase bromodomain. Nature, 399(6735), pp.491–496.

Dhayalan, A. et al., 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. Journal of Biological Chemistry, 285(34), pp.26114–26120.

Doerr, A., 2015. Structural biology: Cryo-EM goes high-resolution. *Nature Methods*, 12(7), pp.598–599.

Dong, A. et al., 2007. In situ proteolysis for protein crystallization and structure determination. Nature methods, 4(12), pp.1019–21.

Dong, J. et al., 2013. SLIDE, The protein interacting domain of imitation switch remodelers, binds DDT-domain proteins of different subfamilies in chromatin remodeling complexes. Journal of Integrative Plant Biology, 55(10), pp.928–937.

dos Reis, M.A., Aparicio, R. & Zhang, Y., 2011. Improving Protein Template Recognition by Using Small-Angle X-Ray Scattering Profiles. *Biophysical Journal*, 101(11), pp.2770–2781.

Dosztannyi, Z., Csizmok, V., et al., 2005. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics, 21(16), pp.3433–3434.

Dürr, H. et al., 2005. X-Ray structures of the sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. Cell, 121(3), pp.363–373.

Dürr, H. et al., 2006. Snf2 family ATPases and DExx box helicases: Differences and unifying concepts from high-resolution crystal structures. Nucleic Acids Research, 34(15), pp.4160–4167.

Eberharter, A. & Becker, P.B., 2002. Histone acetylation: A switch between repressive and permissive chromatin. Second in review on chromatin dynamics. EMBO Reports, 3(3), pp.224–229.

Ehrlich, M. et al., 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Research, 10(8), pp.2709–2721.

Eltermann, M 2015. Structural Studies of the N-terminal Domain of Chromatin Remodelling Protein Lymphoid Specific Helicase. MChem Thesis, University of Edinburgh.

Eltsov, M., Maclellan, K. M., Maeshima, K., Frangakis, A. S. and Dubochet, J. (2008) 'Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ.', Proceedings of the National Academy of Sciences of the United States of America, 105(50), pp. 19732–7.

Epsztejn-Litman, S. et al., 2008. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nature structural & molecular biology, 15(11), pp.1176–83.

Erdel, F. & Rippe, K., 2011. Binding kinetics of human ISWI chromatin-remodelers to DNA repair sites elucidate their target location mechanism. Nucleus (Austin, Tex.), 2(2), pp.105–12.

Fatica, A. & Bozzoni, I., 2014. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet, 15(1), pp.7–21.

Felle, M. et al., 2011. Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Research, 39(16), pp.6956–6969.

Finch, J. T. and Klug, A. (1976) 'Solenoidal model for superstructure in chromatin.', Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 73(6), pp. 1897–901.

Fitzgerald, D.J. et al., 2006. Protein complex expression by using multigene baculoviral vectors. *Nature methods*, 3(12), pp.1021–1032.

Flanagan, J.F. et al., 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature, 438(7071), pp.1181–5.

Flaus, A. & Owen-Hughes, T., 2001. Mechanisms for ATP-dependend chromatin remodelling. Current Opinion in Genetics & Development, 11(2), pp.148–154.

Flaus, A. & Owen-Hughes, T., 2011. Mechanisms for ATP-dependent chromatin remodelling: The means to the end. FEBS Journal, 278(19), pp.3579–3595.

Flaus, A. et al., 2006. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Research, 34(10), pp.2887–2905.

Fontana, A. et al., 2004. Probing protein structure by limited proteolysis. In Acta Biochimica Polonica. pp. 299–321.

Franke, D. & Svergun, D.I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography, 42(2), pp.342–346.

Fried, M. & Crothers, D.M., 1981. Equilibria and kinetics of the lac repressor - operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Research, 9, pp.6505–6525.

Fyodorov, D. V and Kadonaga, J. T. (2002) 'Binding of Acf1 to DNA involves a WAC motif and is important for ACF-mediated chromatin assembly.', Molecular and cellular biology. American Society for Microbiology, 22(18), pp. 6344–53.

Gao, X. et al., 2005. High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. Journal of Structural and Functional Genomics, 6(2-3), pp.129–134.

Garner, M.M. & Revzin, A., 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: applications to components of E.coli lactose operon regulatory proteins. Nucleic Acids Research, 9, pp.3047–3060.

Gasteiger, E. et al., 2005. Protein Identification and Analysis Tools on the ExPASy Server.

The Proteomics Protocols Handbook, pp.571–607.

Gévry, N., Ho, M. C., Laflamme, L., Livingston, D. M. and Gaudreau, L. (2007) 'p21 transcription is regulated by differential localization of histone H2A.Z', Genes and Development, 21(15), pp. 1869–1881.

Giese, S. H., Fischer, L. and Rappsilber, J. (2015) 'A study into the CID behavior of cross-linked peptides.', Molecular & cellular proteomics : MCP. American Society for Biochemistry and Molecular Biology, 15(3), pp. 1094–1104.

Gkikopoulos, T. et al., 2011. A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science (New York, N.Y.), 333(6050), pp.1758–60.

Golub, E.I. et al., 1997. Interaction of human recombination proteins Rad51 and Rad54. Nucleic Acids Research, 25(20), pp.4106–4110.

Gonda, D.K. et al., 1989. Universality and structure of the N-end rule. The Journal of biological chemistry, 264(28), pp.16700–16712.

Gowher, H. et al., 2005. De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. *Biochemistry*, 44(29), pp.9899–9904.

Graslund, 2008. Protein production and purification. *Nature methods*, 5(2), pp.135–146.

Groth, A. et al., 2007. Regulation of replication fork progression through histone supply and demand. Science (New York, N.Y.), 318(5858), pp.1928–31.

Guo, X. et al., 2015. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature, 517(7536), pp.640–4.

Guttman, M. & Rinn, J.L., 2012. Modular regulatory principles of large non-coding RNAs. Nature, 482(7385), pp.339–46.

Hagleitner, M.M. et al., 2008. Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). Journal of medical genetics, 45(2), pp.93–99.

Han, J. et al., 2007. Acetylation of lysine 56 of histone H3 catalyzed by RTT109 and regulated by ASF1 is required for replisome integrity. Journal of Biological Chemistry, 282(39), pp.28587–28596.

Hansen, K.D. et al., 2011. Increased methylation variation in epigenetic domains across cancer types. Nature Genetics, 43(8), pp.768–775.

Harshman, S.W. et al., 2013. H1 histones: current perspectives and challenges. Nucleic acids research, 41(21), pp.9593–609.

Hauer, M. H., Seeber, A., Singh, V., Thierry, R., Sack, R., Amitai, A., Kryzhanovska, M., Eglinger, J., Holcman, D., Owen-Hughes, T. and Gasser, S. M. (2017) 'Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates', Nature Structural & Molecular Biology. Nature Research.

Hauk, G. & Bowman, G.D., 2011. Structural insights into regulation and action of SWI2/SNF2

ATPases. Current Opinion in Structural Biology, 21(6), pp.719–727.

Hauk, G. et al., 2010. The Chromodomains of the Chd1 Chromatin Remodeler Regulate DNA Access to the ATPase Motor. *Molecular Cell*, 39(5), pp.711–723.

Hellman, L.M. & Fried, M.G., 2009. NIH Public Access. N, 2(8), pp.1849–1861.

Hitchman, R.B. et al., 2007. Quantitative real-time PCR for rapid and accurate titration of recombinant baculovirus particles. *Biotechnology and Bioengineering*, 96(4), pp.810–814.

Hizume, K., Yoshimura, S.H. & Takeyasu, K., 2005. Linker Histone H1 per se Can Induce Three-Dimensional Folding of Chromatin Fiber †. Biochemistry, 44(39), pp.12978–12989.

Ho, L. & Crabtree, G.R., 2010. Chromatin remodelling during development. Nature, 463(7280), pp.474–484.

Ho, L. et al., 2009. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proceedings of the National Academy of Sciences of the United States of America, 106(13), pp.5181–6.

Hopkins, R.F. & Esposito, D., 2009. A rapid method for titrating baculovirus stocks using the Sf-9 Easy Titer cell line. *BioTechniques*, 47(3), pp.785–788.

Huang Cole,R.D., H., 1984. The Distribution of H1 IsNonuniform in Chromatin and Correlates with Different Degrees ofCondensation. J. Biol.Chem., 259(22), pp.14237–14242.

Hur, S.K. et al., 2010. Roles of human INO80 chromatin remodeling enzyme in DNA replication and chromosome segregation suppress genome instability. Cellular and Molecular Life Sciences, 67(13), pp.2283–2296.

Iacovoni, J.S. et al., 2010. High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. The EMBO journal, 29(8), pp.1446–57.

Ikonomou, L. et al., 2004. Effect of Partial Medium Replacement on Cell Growth and Protein Production for the High-Five(TM) insect cell line. *Cytotechnology*, 44(1-2), pp.67–76.

Ikonomou, L., Schneider, Y.J. & Agathos, S.N., 2003. Insect cell culture for industrial production of recombinant proteins. *Applied Microbiology and Biotechnology*, 62(1), pp.1–20.

Ito, T. et al., 1997. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell, 90(1), pp.145–155.

Jacques, D.A. & Trewhella, J., 2010. Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Science*, 19(4), pp.642–657.

Jaenisch, R., 1997. DNA methylation and imprinting: Why bother? Trends in Genetics, 13(8), pp.323–329.

Jeong, S. et al., 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Molecular and cellular biology*, 29(19), pp.5366–76.

Jiang, H. et al., 1996. Direct association between the yeast Rad51 and Rad54 recombination proteins. Journal of Biological Chemistry, 271(52), pp.33181–33186.

Jiang, Y.L. et al., 2005. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Human Mutation, 25(1), pp.56–63.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology, 292(2), pp.195–202.

Jones, P.L. et al., 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature genetics, 19(2), pp.187–191.

Jónsson, Z.O. et al., 2004. Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. Molecular Cell, 16(3), pp.465–477.

Jorio, H., Tran, R. & Kamen, A., 2006. Stability of Serum-Free and Purified Baculovirus Stocks under Various Storage Conditions. *Biotechnology Progress*, 22(1), pp.319–325.

Jurkowska, R.Z., Jurkowski, T.P. & Jeltsch, A., 2011. Structure and Function of Mammalian DNA Methyltransferases. ChemBioChem, 12(2), pp.206–222.

Kamen, A.A. et al., 1996. On-line monitoring of respiration in recombinant-baculovirus infected and uninfected insect cell bioreactor cultures. *Biotechnology and Bioengineering*, 50(1), pp.36–48.

Kelley, L.A. et al., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, 10(6), pp.845–858.

Kikhney, A.G. & Svergun, D.I., 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*, 589(19), pp.2570–2577.

Kim, R., 2011. Native agarose gel electrophoresis of multiprotein complexes. *Cold Spring Harbor Protocols*, 6(7), pp.884–887.

King, L., 2012. *The baculovirus expression system: a laboratory guide*,

Konev, A. Y., Tribus, M., Park, S. Y., Podhraski, V., Lim, C. Y., Emelyanov, A. V., Vershilova, E., Pirrotta, V., Kadonaga, J. T., Lusser, A. and Fyodorov, D. V. (2007) 'CHD1 Motor Protein Is Required for Deposition of Histone Variant H3.3 into Chromatin in Vivo', Science, 317(5841), pp. 1087–1090.

Kornberg, R.D., 1974. Chromatin structure: a repeating unit of histones and DNA. Science (New York, N.Y.), 184(139), pp.868–871.

Kwon, S. Y., Xiao, H., Wu, C. and Badenhorst, P. (2009) 'Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities', PLoS Genetics. Edited by A. Akhtar, 5(7), p. e1000574.

Law, M. J., Lower, K. M., Voon, H. P. J., Hughes, J. R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A., Abbott, A., Wilder, S. P., Taylor, S., Santos, G. M., Cross, J., Ayyub, H., Jones, S., Ragoussis, J., Rhodes, D., Dunham, I., Higgs, D. R. and Gibbons, R. J. (2010) 'ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner', Cell, 143(3), pp. 367–378.

Lee, D.W. et al., 2000. Proliferation-associated SNF2-like gene (PASG): A SNF2 family member altered in leukemia. *Cancer Research*, 60(13), pp.3612–3622.

Leitner, A. et al., 2016. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends in Biochemical Sciences*, 41(1), pp.20–32.

Li, E., Bestor, T.H. & Jaenisch, R., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell, 69(6), pp.915–926.

Li, M., Hada, A., Sen, P., Olufemi, L., Hall, M. A., Smith, B. Y., Forth, S., McKnight, J. N., Patel, A., Bowman, G. D., Bartholomew, B. and Wang, M. D. (2015) 'Dynamic regulation of transcription factors by nucleosome remodeling', eLife. eLife Sciences Publications Limited, 4(JUNE), pp. 1–16.

Lia, G. et al., 2006. Direct observation of DNA distortion by the RSC complex. Molecular Cell, 21(3), pp.417–425.

Lim, D.H.K. & Maher, E.R., 2010. DNA methylation: a form of epigenetic control of gene expression. The Obstetrician & Gynaecologist, 12(1), pp.37–42.

Liu, J.C., Ferreira, C.G. & Yusufzai, T., 2015. Human CHD2 is a chromatin assembly ATPase regulated by its chromo- And DNA-binding domains. Journal of Biological Chemistry, 290(1), pp.25–34.

Lodish, H., Berk, A. & Zipursky, S., 2000. Intracellular Ion Environment and Membrane Electric Potential. In Molecular Cell Biology, 4th Edition. W. H. Freeman, pp. 1–5.

Luger, K. & Collins, F., 2001. Nucleosomes: Structure and Function. Encyclopedia of Life Sciences, pp.1–8.

Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature, 389(6648), pp.251–260.

Luk, E., Ranjan, A., FitzGerald, P. C., Mizuguchi, G., Huang, Y., Wei, D. and Wu, C. (2010) 'Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome', Cell, 143(5), pp. 725–736.

Lungu, C. et al., 2015. An ATPase-deficient variant of the SNF2 family member HELLS shows altered dynamics at pericentromeric heterochromatin. Journal of Molecular Biology, 427(10), pp.1903–1915.

Lupas, A.N. & Gruber, M., 2005. The structure of alpha-helical coiled coils. Advances in protein chemistry, 70, pp.37–78.

Maeshima, K., Hihara, S. and Eltsov, M. (2010) 'Chromatin structure: Does the 30-nm fibre exist in vivo?', Current Opinion in Cell Biology, pp. 291–297.

Maeshima, K., Imai, R., Tamura, S. and Nozaki, T. (2014) 'Chromatin as dynamic 10-nm fibers', Chromosoma, pp. 225–237.

Maiolica, A. et al., 2007. Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Molecular & cellular proteomics : MCP*, 6(12),

pp.2200–2211.

Manelyte, L. and Längst, G. (2013) 'Chromatin Remodelers and Their Way of Action', Chromatin Remodelling, pp. 3–28. doi: 10.5772/55683.

Manning, B.J. & Peterson, C.L., 2013. Releasing the brakes on a chromatin-remodeling enzyme. Nature structural & molecular biology, 20(1), pp.5–7.

Mansfield, R. E., Musselman, C. A., Kwan, A. H., Oliver, S. S., Garske, A. L., Davrazou, F., Denu, J. M., Kutateladze, T. G. and Mackay, J. P. (2011) 'Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9', Journal of Biological Chemistry, 286(13), pp. 11779–11791.

Marfella, C.G.A. & Imbalzano, A.N., 2007. The Chd family of chromatin remodelers. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis, 618(1-2), pp.30–40.

Mazin, A. V et al., 2010. Rad54, the motor of homologous recombination. DNA Repair, 9(3), pp.286–302.

Mazina, O.M. & Mazin, A. V., 2004. Human Rad54 protein stimulates DNA strand exchange activity of hRad51 protein in the presence of Ca2+. Journal of Biological Chemistry, 279(50), pp.52042–52051.

McDowall, A. W., Smith, J. M. and Dubochet, J. (1986) 'Cryo-electron microscopy of vitrified chromosomes in situ.', The EMBO journal, 5(6), pp. 1395–1402.

McKnight, J.N. et al., 2011. Extranucleosomal DNA binding directs nucleosome sliding by Chd1. Molecular and cellular biology, 31(23), pp.4746–59.

McLellan, T., 1982. Electrophoresis buffers for polyacrylamide gels at various pH. Analytical Biochemistry, 126(1), pp.94–99.

Merk, A. et al., 2016. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell*, 165, pp.1–10.

Mertens, H.D.T. & Svergun, D.I., 2010. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*, 172(1), pp.128–141.

Mizuguchi, G. et al., 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science (New York, N.Y.), 303(5656), pp.343–8.

Mohrmann, L. & Verrijzer, C.P., 2005. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochimica et Biophysica Acta - Gene Structure and Expression, 1681(2-3), pp.59–73.

Mosammaparast, N. et al., 2002. A role for nucleosome assembly protein 1 in the nuclear transport of histones H2A and H2B. The EMBO Journal, 21(23), pp.6527–6538.

Moshkin, Y. M., Chalkley, G. E., Kan, T. W., Reddy, B. A., Ozgur, Z., van Ijcken, W. F. J., Dekkers, D. H. W., Demmers, J. A., Travers, A. A. and Verrijzer, C. P. (2012a) 'Remodelers Organize Cellular Chromatin by Counteracting Intrinsic Histone-DNA

Sequence Preferences in a Class-Specific Manner', Molecular and Cellular Biology. American Society for Microbiology (ASM), 32(3), pp. 675–688.

Mueller-Planitz, F. et al., 2013. The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. Nature structural & molecular biology, 20(1), pp.82–9.

Myant, K. & Stancheva, I., 2008. LSH cooperates with DNA methyltransferases to repress transcription. Molecular and cellular biology, 28(1), pp.215–226.

Myant, K. et al., 2011. LSH and G9a / GLP complex are required for developmentally programmed DNA methylation. *Genome Res*, 1(21), pp.83–94.

Niesen, F.H., Berglund, H. & Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.

Niikura, Y. et al., 2015. CENP-A K124Ubiquitylation Is Required for CENP-A Deposition at the Centromere. Developmental Cell, 32(5), pp.589–603.

Ohi, M. et al., 2004. Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy. *Biol. Proced. Online*, 6(1), pp.23–34.

Ohki, T. et al., 2012. Improvement of the yields of recombinant actin and myosin V-HMM in the insect cell/baculovirus system by the addition of nutrients to the high-density cell culture. *Journal of Muscle Research and Cell Motility*, 33(5), pp.351–358.

Okano, M. et al., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell, 99(3), pp.247–257.

Palomares, L.A., Estrada-Mondaca, S. & Ramírez, O.T., 2004. Production of Recombinant Proteins: Challenges and Solutions. In Recombinant Gene Expression. pp. 1–39.

Palomares, L.A., López, S. & Ramírez, O.T., 2004. Utilization of oxygen uptake rate to assess the role of glucose and glutamine in the metabolism of infected insect cell cultures. *Biochemical Engineering Journal*, 19(1), pp.87–93.

Papamichos-Chronakis, M. & Peterson, C.L., 2008. The Ino80 chromatin-remodeling enzyme regulates replisome function and stability. Nature structural & molecular biology, 15(4), pp.338–345.

Papamichos-Chronakis, M. et al., 2011. Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling Enzyme Is Essential for Genome Integrity. Cell, 144, pp.200–213.

Partensky, P. D. and Narlikar, G. J. (2009) 'Chromatin Remodelers Act Globally, Sequence Positions Nucleosomes Locally', Journal of Molecular Biology. NIH Public Access, 391(1), pp. 12–25.

Parthun, M.R., Widom, J. & Gottschling, D.E., 1996. The major cytoplasmic histone acetyltransferase in yeast: Links to chromatin replication and histone metabolism. Cell, 87(1), pp.85–94.

Paulson, J. R. and Langmore, J. P. (1983) 'Low angle X-ray diffraction studies of HeLa metaphase chromosomes: Effects of histone phosphorylation and chromosome

isolation procedure', Journal of Cell Biology. The Rockefeller University Press, 96(4), pp. 1132–1137.

Pernot, P. et al., 2013. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. Journal of Synchrotron Radiation, 20(4), pp.660–664.

Petoukhov, M. V et al., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography, 45(2), pp.342–350.

Petoukhov, M. V. et al., 2007. ATSAS 2.1 - Towards automated and web-supported small-angle scattering data analysis. In Journal of Applied Crystallography.

Pettersen, E.F. et al., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25(13), pp.1605–1612.

Petoukhov, M. V. and Svergun, D. I. (2005) 'Global rigid body modeling of macromolecular complexes against small-angle scattering data.', Biophysical journal, 89(2), pp. 1237–1250.

Petukhova, G., Stratton, S. & Sung, P., 1998. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. Nature, 393(6680), pp.91–94.

Pointner, J. et al., 2012. CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. Embo Journal, 31(23), pp.4388–4403.

Polo, S.E. & Jackson, S.P., 2011. Dynamics of DNA damage response proteins at DNA breaks: A focus on protein modifications. Genes and Development, 25(5), pp.409–433.

Porod, G., 1982. in Small angle X-ray Scattering. *Academic Press*, p.515.

Purdy, M.M., Holz-Schietinger, C. & Reich, N.O., 2010. Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion. Archives of Biochemistry and Biophysics, 498(1), pp.13–22.

Putnam, C.D. et al., 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*, 40(3), pp.191–285.

Racki, L. R., Yang, J. G., Naber, N., Partensky, P. D., Acevedo, A., Purcell, T. J., Cooke, R., Cheng, Y. and Narlikar, G. J. (2009) 'The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes', Nature. Nature Publishing Group, 462(7276), pp. 1016–1021.

Rambo, R.P. & Tainer, J.A., 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, 95(8), pp.559–71.

Rames, M., Yu, Y. & Ren, G., 2014. Optimized Negative Staining: a High-throughput Protocol for Examining Small and Asymmetric Protein Structure by Electron Microscopy. *Journal of Visualized Experiments*, (90), pp.e51087–e51087.

Ransom, M., Dennehey, B.K. & Tyler, J.K., 2010. Chaperoning Histones during DNA

Replication and Repair. Cell, 140(2), pp.183–195.

Rappsilber, J., 2011. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of Structural Biology*, 173(3), pp.530–540.

Rappsilber, J., Ishihama, Y. & Mann, M., 2003. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. Analytical Chemistry, 75(3), pp.663–670.

Raschle, M. et al., 2004. Multiple interactions with the rad51 recombinase govern the homologous recombination function of Rad54. Journal of Biological Chemistry, 279(50), pp.51973–51980.

Rausch, M., Pörtner, R. & Knäblein, J., 2013. Increase of protein yield in high five cells in a single-use perfusion bioreactor by medium replacement. *Chemie-Ingenieur-Technik*, 85(1-2), pp.111–117.

Razin, A. & Riggs, A.D., 1980. DNA methylation and gene function. Science, 210(4470), pp.604–610.

Receveur-Brechot, V. & Durand, D., 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. *Current protein & peptide science*, 13(1), pp.55–75.

Ren, J. et al., 2015. The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. Nucleic acids research, 43(3), pp.1444–55.

Reuveny, S. et al., 1993. Production of recombinant proteins in high-density insect cell cultures. *Biotechnology and Bioengineering*, 42(2), pp.235–239.

Reynolds, N. et al., 2012. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. The EMBO Journal, 31(3), pp.593–605.

Rhiel, M., Mitchell-Logean, C.M. & Murhammer, D.W., 1997. Comparison of Trichoplusia ni BTI-Tn-5b1-4 (high five(TM) and Spodoptera frugiperda Sf-9 insect cell line metabolism in suspension cultures. *Biotechnology and Bioengineering*, 55(6), pp.909–920.

Richmond, E. & Peterson, C.L., 1996. Functional analysis of the DNA-stimulated ATPase domain of yeast SW12/SNF2. Nucleic Acids Research, 24(19), pp.3685–3692.

Rinn, J.L. et al., 2007. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell, 129(7), pp.1311–1323.

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E. and Langst, G. (2007) 'DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes', Proceedings of the National Academy of Sciences, 104(40), pp. 15635–15640.

Ristic, D. et al., 2001. The architecture of the human Rad54-DNA complex provides evidence

for protein translocation along DNA. Proceedings of the National Academy of Sciences of the United States of America, 98(15), pp.8454–60.

Robertson, K.D. et al., 1999. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: Coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Research, 27(11), pp.2291–2298.

Robinson, P. J. J., Fairall, L., Huynh, V. A. T. and Rhodes, D. (2006) 'EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure.', Proceedings of the National Academy of Sciences of the United States of America, 103(17), pp. 6506–11.

Rondelet, G., Dal Maso, T., Willems, L. and Wouters, J. (2016) 'Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B', Journal of Structural Biology, 194(3), pp. 357–367.

Rossi, A.M. & Taylor, C.W., 2011. Analysis of protein-ligand interactions by fluorescence polarization. Nature protocols, 6(3), pp.365–387.

Rothbart, S.B. & Strahl, B.D., 2014. Interpreting the language of histone and DNA modifications. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms, 1839(8), pp.627–643.

Rush, M. et al., 2009. Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. Epigenetics, 4(6), pp.404–414.

Ryan, D.P. et al., 2011. The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. The EMBO journal, 30(13), pp.2596–2609.

Rye, H.S. et al., 1993. Stable fluorescent dye-DNA complexes in high sensitivity detection of protein-DNA interactions. Application to heat shock transcription factor. Journal of Biological Chemistry, 268(33), pp.25229–25238.

Ryu, B. et al., 2007. Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression J. Najbauer, ed. PLoS ONE, 2(7), p.e594.

Saha, A., Wittmeyer, J. & Cairns, B.R., 2002. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. Genes and Development, 16(16), pp.2120–2134.

Samuelsson, J., Dumbovic, G., Polo, C., Moreta, C., Alibés, A., Ruiz-Larroya, T., Giménez-Bonafé, P., Alonso, S., Forcales, S.-V. and Perucho, M. (2016) 'Helicase Lymphoid-Specific Enzyme Contributes to the Maintenance of Methylation of SST1 Pericentromeric Repeats That Are Frequently Demethylated in Colon Cancer and Associate with Genomic Damage', Epigenomes. Multidisciplinary Digital Publishing Institute, 1(1), p. 2.

Scheffer, M. P., Eltsov, M. and Frangakis, A. S. (2011) 'Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei.', Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 108(41), pp. 16992–7.

Schlaeger, E.-J., 1996. Medium design for insect cell culture. *Cytotechnology*, 20(1-3), pp.57–70.

Schneider, C. a, Rasband, W.S. & Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods, 9(7), pp.671–675.

Schubert, H.L. et al., 2013. Structure of an actin-related subcomplex of the SWI/SNF chromatin remodeler. Proceedings of the National Academy of Sciences of the United States of America, 110(9), pp.3345–50.

Sen, P. et al., 2011. A new, highly conserved domain in Swi2/Snf2 is required for SWI/SNF remodeling. Nucleic Acids Research, 39(21), pp.9155–9166.

Sen, P. et al., 2013. The SnAC domain of SWI/SNF is a histone anchor required for remodeling. Molecular and cellular biology, 33(2), pp.360–370.

Shechter, D., Chitta, R. K., Xiao, A., Shabanowitz, J., Hunt, D. F. and Allis, C. D. (2009) 'A distinct H2A.X isoform is enriched in Xenopus laevis eggs and early embryos and is phosphorylated in the absence of a checkpoint.', Proceedings of the National Academy of Sciences of the United States of America, 106(3), pp. 749–754.

Shen, X. et al., 2000. A chromatin remodelling complex involved in transcription and DNA processing. Nature, 406(6795), pp.541–544.

Singh, M., D'Silva, L. and Holak, T. A. (2006) 'DNA-binding properties of the recombinant high-mobility-group-like AT-hook-containing region from human BRG1 protein', Biological Chemistry, 387(10/11), pp. 1469–78.

Singleton, M.R., Dillingham, M.S. & Wigley, D.B., 2007. Structure and mechanism of helicases and nucleic acid translocases. Annual review of biochemistry, 76, pp.23–50.

Skou, S., Gillilan, R.E. & Ando, N., 2014. Synchrotron-based small-angle X-ray scattering (SAXS) of proteins in solution HHS Public Access. *Nat Protoc*, 9116(7), pp.1727–1739.

Stein, R. et al., 1982. Clonal inheritance of the pattern of DNA methylation in mouse cells. Proceedings of the National Academy of Sciences of the United States of America, 79(1), pp.61–65.

Stellwagen, N.C., 2009. Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution. Electrophoresis, 30(SUPPL. 1), pp.S188–95.

Stockdale, C., Flaus, A., Ferreira, H. and Owen-Hughes, T. (2006) 'Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes', Journal of Biological Chemistry, 281(24), pp. 16279–16288.

Stopka, T. & Skoultchi, A.I., 2003. The ISWI ATPase Snf2h is required for early mouse development. Proceedings of the National Academy of Sciences of the United States of America, 100(24), pp.14097–102.

Suetake, I. et al., 2003. Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b. *Journal of Biochemistry*, 133(6), pp.737–744.

Suetake, I. et al., 2011. Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. The Biochemical journal, 437(1), pp.141–148.

Sundaramoorthy, R., Hughes, A. L., Singh, V., Wiechens, N., Ryan, D. P., El-Mkami, H., Petoukhov, M., Svergun, D. I., Treutlein, B., Quack, S., Fischer, M., Michaelis, J., Böttcher, B., Norman, D. G. and Owen-Hughes, T. (2017) 'Structural reorganization of the chromatin remodeling enzyme Chd1 upon engagement with nucleosomes', eLife. eLife Sciences Publications Limited, 6, pp. 339–346.

Svergun, D.I. & Koch, M.H.J., 2002. Advances in structure analysis using small-angle scattering in solution. *Current Opinion in Structural Biology*, 12(5), pp.654–660.

Svergun, D.I., 1999. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal*, 76(6), pp.2879–2886.

Svergun, D.I., Petoukhov, M. V & Koch, M.H., 2001. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical journal*, 80(6), pp.2946–2953.

Szerlong, H. et al., 2008. The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. Nature structural & molecular biology, 15(5), pp.469–476.

Tagami, H. et al., 2004. Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. Cell, 116(1), pp.51–61.

Tajima, S. & Suetake, I., 1998. Regulation and function of DNA methylation in vertebrates. Journal of biochemistry, 123(6), pp.993–9.

Talbert, P.B. & Henikoff, S., 2010. Histone variants - ancient wrap aartists of the epigenome. Nature reviews. Molecular cell biology, 11(4), pp.264–275.

Tang, G. et al., 2007. EMAN2: An extensible image processing suite for electron microscopy. Journal of Structural Biology, 157(1), pp.38–46.

Tang, L., Nogales, E. & Ciferri, C., 2010. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. Progress in Biophysics and Molecular Biology, 102(2-3), pp.122–128.

Tao, Y. et al., 2011. Correction for Tao et al., Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. *Proceedings of the National Academy of Sciences*, 108(37), pp.15535–15535.

Taunton, J., Hassig, C.A. & Schreiber, S.L., 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science (New York, N.Y.), 272(5260), pp.408–411.

Termanis, A. et al., 2016. The SNF2 family ATPase LSH promotes cell-autonomous de novo DNA methylation in somatic cells. Nucleic Acids Research, p.gkw424.

Thijssen, P.E. et al., 2015. Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nature communications, 6, p.7870.

Thoma, F. et al., 1979. Involvement of histone H1 in the organization of the nucleosome and of thesalt-dependent superstructures ofchromatin. J.Cell.Biol, 83(November), pp.403–427.

Thomä, N.H. et al., 2005. Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. Nature structural & molecular biology, 12(4), pp.350–6.

Trowitzsch, S. et al., 2010. New baculovirus expression tools for recombinant protein complex production. Journal of Structural Biology, 172(1), pp.45–54.

Tsukiyama, T. & Wu, C., 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell, 83(6), pp.1011–1020.

Udugama, M., Sabri, A. & Bartholomew, B., 2011. The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. Molecular and cellular biology, 31(4), pp.662–673.

Uversky, V., Gillespie, J. & Fink, A., 2000. Why are 'natively unfolded' proteins unstructured under physiologic conditions? Proteins, 41(3), pp.415–427.

Vagenende, V., Yap, M.G.S. & Trout, B.L., 2009. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. Biochemistry, 48(46), pp.11084–11096.

Van Attikum, H. et al., 2004. Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. Cell, 119(6), pp.777–788.

Van Emburgh, B.O. & Robertson, K.D., 2011. Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. *Nucleic Acids Research*, 39(12), pp.4984–5002.

van Vugt, J. J. F. A., de Jager, M., Murawska, M., Brehm, A., van Noort, J. and Logie, C. (2009) 'Multiple aspects of ATP-dependent nucleosome translocation by RSC and Mi-2 are directed by the underlying DNA sequence', PLoS ONE. Edited by M. Freitag. Academic Press, 4(7), p. e6345.

Varga-Weisz, P.D. et al., 1997. Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. Nature, 388(6642), pp.598–602.

Velankar, S.S. et al., 1999. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell, 97(1), pp.75–84.

Vincent, J.A., Kwong, T.J. & Tsukiyama, T., 2008. ATP-dependent chromatin remodeling shapes the DNA replication landscape. Nature structural & molecular biology, 15(5), pp.477–84.

Viré, E. et al., 2005. The Polycomb group protein EZH2 directly controls DNA methylation. Nature, 439(7078), pp.871–874.

Voigt, P., Tee, W.-W. & Reinberg, D., 2013. A double take on bivalent promoters. Genes & Development, 27(12), pp.1318–1338.

von Eyss, B. et al., 2012. The SNF2-like helicase HELLS mediates E2F3-dependent

transcription and cellular transformation. The EMBO journal, 31(4), pp.972–85.

Walsh, C.P., Chaillet, R. & Bestor, T.H., 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature Genetics, 20(october), pp.116–117.

Wang, M.Y., Kwong, S. & Bentley, W.E., 1993. Effects of oxygen/glucose/glutamine feeding on insect cell baculovirus protein expression: a study on epoxide hydrolase production. *Biotechnology progress*, 9(4), pp.355–61.

Wang, R. et al., 2015. The ratio of FoxA1 to FoxA2 in lung adenocarcinoma is regulated by LncRNA HOTAIR and chromatin remodeling factor LSH. Scientific Reports, 5(August), p.17826.

Warburton, P.E. et al., 1997. Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Current biology : CB, 7(11), pp.901–904.

Waseem, A. et al., 2010. Downstream targets of FOXM1: CEP55 and HELLS are cancer progression markers of head and neck squamous cell carcinoma. Oral Oncology, 46(7), pp.536–542.

Wasilko, D.J. et al., 2009. The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. *Protein Expression and Purification*, 65(2), pp.122–132.

Weber, C.M. & Henikoff, S., 2014. Histone variants: dynamic punctuation in transcription. Genes & development, 28(7), pp.672–82.

Weider, M., Küspert, M., Bischof, M., Vogl, M. R., Hornig, J., Loy, K., Kosian, T., Müller, J., Hillgärtner, S., Tamm, E. R., Metzger, D. and Wegner, M. (2012) 'Chromatin-Remodeling Factor Brg1 Is Required for Schwann Cell Differentiation and Myelination', Developmental Cell. Elsevier, 23(1), pp. 193–201.

Wernimont, A. & Edwards, A., 2009. In Situ proteolysis to generate crystals for structure determination: An update. PLoS ONE, 4(4), p.e5094.

Whitehouse, I. et al., 2003. Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. Molecular and cellular biology, 23(6), pp.1935–1945.

Wilde, M. et al., 2014. Tnao38, high five and Sf9—evaluation of host–virus interactions in three different insect cell lines: baculovirus production and recombinant protein expression. *Biotechnology Letters*, 36(4), pp.743–749.

Wong, T.K.K. et al., 1994. Relationship between oxygen uptake rate and time of infection of Sf9 insect cells infected with a recombinant baculovirus. *Cytotechnology*, 15(1-3), pp.157–167.

Woodcock, C. L. (1994) 'Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length.', The Journal of Cell Biology, 125(1).

Woodcock, S., Mornon, J.-P. & Henrissat, B., 1992. Detection of secondary structure

elements in proteins by hydrophobic cluster analysis. 'Protein Engineering, Design and Selection', 5(7), pp.629–635.

Xi, S., Geiman, T. M., Briones, V., Tao, Y. G., Xu, H. and Muegge, K. (2009) 'Lsh participates in DNA methylation and silencing of stem cell genes', Stem Cells. NIH Public Access, 27(11), pp. 2691–2702.

Xia, X. et al., 2016. Structure of chromatin remodeler Swi2/Snf2 in the resting state. *Nature Structural & Molecular Biology*, (July), pp.1–10.

Yamada, K. et al., 2011. Structure and mechanism of the chromatin remodelling factor ISW1a. Nature, 472(7344), pp.448–453.

Yan, L., Wang, L., Tian, Y., Xia, X. and Chen, Z. (2016) 'Structure and regulation of the chromatin remodeller ISWI', Nature. Nature Research, 540(7633), pp. 466–469.

Yan, Q. et al., 2003. Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin. Molecular and cellular biology, 23(23), pp.8416–28.

Yang, J. et al., 2015. The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12(1), pp.7–8.

Yang, J. G., Madrid, T. S., Sevastopoulos, E. and Narlikar, G. J. (2006) 'The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing', Nature Structural & Molecular Biology, 13(12), pp. 1078–1083.

Yang, L., Rau, R. & Goodell, M.A., 2015. DNMT3A in haematological malignancies. Nature reviews. Cancer, 15(3), pp.152–165.

Yano, M. et al., 2004. Tumor-specific exon creation of the HELLS/SMARCA6 gene in non-small cell lung cancer. International Journal of Cancer, 112(1), pp.8–13.

Yoda, K. et al., 2000. Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proc. Natl. Acad. Sci. USA, 97(13), pp.7266–7271.

Yodh, J., 2013. ATP-dependent chromatin remodeling. Advances in Experimental Medicine and Biology, 973, pp.263–295.

Yokochi, T. & Robertson, K.D., 2002. Preferential methylation of unmethylated DNA by mammalian de novo DNA methyltransferase Dnmt3a. *Journal of Biological Chemistry*, 277(14), pp.11735–11745.

Zemach, A. et al., 2013. The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell, 153(1), pp.193–205.

Zeng, L. & Zhou, M.M., 2002. Bromodomain: An acetyl-lysine binding domain. FEBS Letters, 513(1), pp.124–128.

Zhang, Y. et al., 2010. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Research, 38(13), pp.4246–4253.

Zhang, Y., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Bird, A. and Reinberg, D. (1999) 'Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation', Genes and Development. Cold Spring Harbor Laboratory Press, 13(15), pp. 1924–1935.

Zhao, J. et al., 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science (New York, N.Y.), 322(5902), pp.750–6.

Zhu, H. et al., 2006. Lsh is involved in de novo methylation of DNA. The EMBO journal, 25(2), pp.335–345.

Zhu, P. & Li, G., 2016. Structural insights of nucleosome and the 30-nm chromatin fiber. Current Opinion in Structural Biology, 36, pp.106–115.

Zofall, M. et al., 2006. Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. Nature structural & molecular biology, 13(4), pp.339–346.

Appendix|

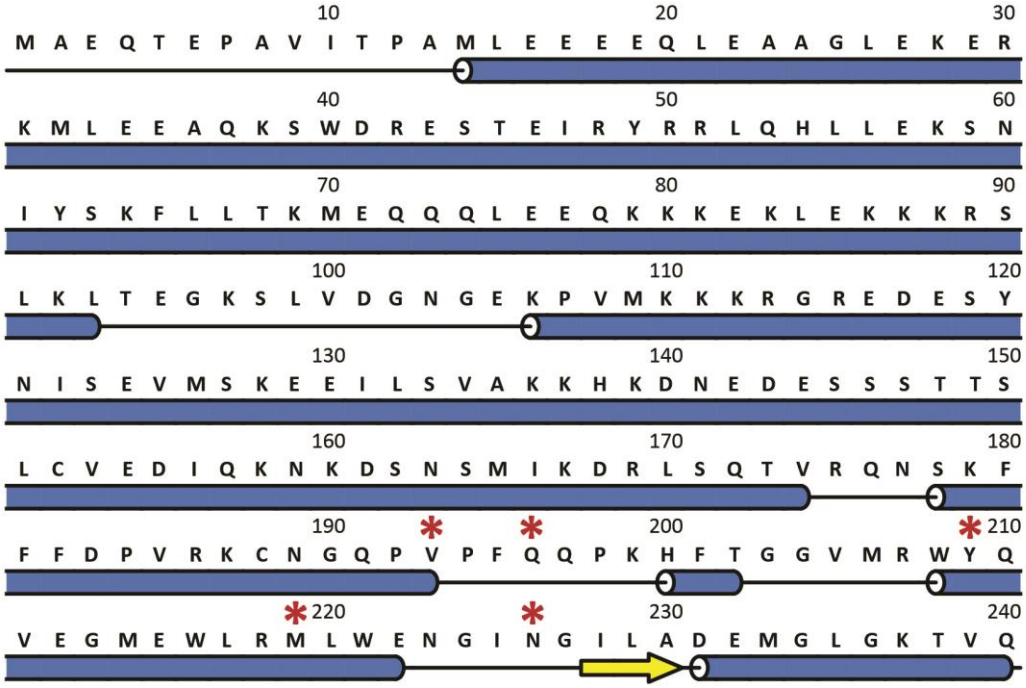

Figure A1| Secondary structure prediction of LSH using PSIpred
Amino acids with red asterisks above them are where truncated LSH constructs were cloned from, making five constructs: $LSH^{193-821}$, $LSH^{196-821}$, $LSH^{209-821}$, $LSH^{219-821}$ and $LSH^{226-821}$.

Table A1| Auto-validated (cut-off of 7) intramolecular BS^3 crosslinks of LSH

Protein1	PepPos1	PepSeq1	Protein2	PepPos2	PepSeq2	Score	AutoVal
LSH_mouse	65	FLLTKMEQQQLEEQK	LSH_mouse	52	LQHLLEKSNIYSK	11.93	TRUE
LSH_mouse	52	LQHLLEKSNIYSK	LSH_mouse	31	KMLEEAQK	11.637	TRUE
LSH_mouse	65	FLLTKMEQQQLEEQK	LSH_mouse	52	LQHLLEKSNIYSK	11.076	TRUE
LSH_mouse	694	IGQTKPVVVYR	LSH_mouse	161	DSNSMIKDR	10.929	TRUE
LSH_mouse	114	GREDESYNISEVMSKEEILSVAK	LSH_mouse	52	LQHLLEKSNIYSK	10.464	TRUE
LSH_mouse	176	QNSKFFFDPVRK	LSH_mouse	90	SLKLTEGK	10.287	TRUE
LSH_mouse	728	LIIHKNHFK	LSH_mouse	342	IKNMK	10.021	TRUE
LSH_mouse	435	LKSDVALEVPPK	LSH_mouse	694	IGQTKPVVVYR	9.87	TRUE
LSH_mouse	694	IGQTKPVVVYR	LSH_mouse	159	NKDSNSMIK	9.801	TRUE
LSH_mouse	31	KMLEEAQK	LSH_mouse	352	ELKR	9.791	TRUE
LSH_mouse	176	QNSKFFFDPVRK	LSH_mouse	90	SLKLTEGK	9.591	TRUE
LSH_mouse	705	LVTANTIDQKIVER	LSH_mouse	694	IGQTKPVVVYR	9.542	TRUE
LSH_mouse	114	GREDESYNISEVMSKEEILSVAK	LSH_mouse	90	SLKLTEGK	9.537	TRUE
LSH_mouse	733	NHFKGGQSGLSQSK	LSH_mouse	176	QNSKFFFDPVR	9.498	TRUE
LSH_mouse	747	NFLDAKELMELLK	LSH_mouse	705	LVTANTIDQKIVER	9.471	TRUE
LSH_mouse	694	IGQTKPVVVYR	LSH_mouse	161	DSNSMIKDR	9.334	TRUE
LSH_mouse	90	SLKLTEGK	LSH_mouse	161	DSNSMIKDR	9.314	TRUE
LSH_mouse	694	IGQTKPVVVYR	LSH_mouse	90	SLKLTEGK	9.259	TRUE
LSH_mouse	90	SLKLTEGK	LSH_mouse	82	EKLEK	9.173	TRUE
LSH_mouse	176	QNSKFFFDPVR	LSH_mouse	90	SLKLTEGK	9.149	TRUE
LSH_mouse	733	NHFKGGQSGLSQSK	LSH_mouse	90	SLKLTEGK	9.065	TRUE
LSH_mouse	52	LQHLLEKSNIYSK	LSH_mouse	159	NKDSNSMIKDR	9.057	TRUE
LSH_mouse	705	LVTANTIDQKIVER	LSH_mouse	694	IGQTKPVVVYR	9.034	TRUE
LSH_mouse	161	DSNSMIKDR	LSH_mouse	628	LDGSMSYSER	8.983	TRUE
LSH_mouse	772	EKVISDEDLELLLDR	LSH_mouse	787	SDLIDQMKASRPIK	8.946	TRUE
LSH_mouse	705	LVTANTIDQKIVER	LSH_mouse	694	IGQTKPVVVYR	8.853	TRUE
LSH_mouse	724	KLEKLIIHK	LSH_mouse	719	AAAKR	8.788	TRUE
LSH_mouse	52	LQHLLEKSNIYSK	LSH_mouse	159	NKDSNSMIK	8.714	TRUE
LSH_mouse	470	TIANMFGSCEK	LSH_mouse	694	IGQTKPVVVYR	8.658	TRUE
LSH_mouse	705	LVTANTIDQKIVER	LSH_mouse	435	LKSDVALEVPPKR	8.545	TRUE
LSH_mouse	176	QNSKFFFDPVRK	LSH_mouse	90	SLKLTEGK	8.504	TRUE
LSH_mouse	705	LVTANTIDQKIVER	LSH_mouse	176	QNSKFFFDPVR	8.141	TRUE
LSH_mouse	90	SLKLTEGK	LSH_mouse	159	NKDSNSMIK	8.076	TRUE
LSH_mouse	527	TVVEGNIPIESEVNLKLR	LSH_mouse	753	ELMELLKSR	7.892	TRUE
LSH_mouse	90	SLKLTEGK	LSH_mouse	161	DSNSMIKDR	7.891	TRUE
LSH_mouse	129	EEILSVAKK	LSH_mouse	90	SLKLTEGK	7.871	TRUE
LSH_mouse	129	EEILSVAKK	LSH_mouse	159	NKDSNSMIK	7.845	TRUE
LSH_mouse	161	DSNSMIKDR	LSH_mouse	90	SLKLTEGK	7.78	TRUE
LSH_mouse	90	SLKLTEGK	LSH_mouse	159	NKDSNSMIK	7.722	TRUE
LSH_mouse	176	QNSKFFFDPVR	LSH_mouse	628	LDGSMSYSER	7.689	TRUE
LSH_mouse	705	LVTANTIDQKIVER	LSH_mouse	694	IGQTKPVVVYR	7.633	TRUE
LSH_mouse	787	SDLIDQMKASRPIK	LSH_mouse	176	QNSKFFFDPVR	7.594	TRUE
LSH_mouse	176	QNSKFFFDPVR	LSH_mouse	159	NKDSNSMIKDR	7.561	TRUE
LSH_mouse	772	EKVISDEDLELLLDR	LSH_mouse	705	LVTANTIDQKIVER	7.475	TRUE
LSH_mouse	208	WYQVEGMEWLR	LSH_mouse	435	LKSDVALEVPPKR	7.228	TRUE
LSH_mouse	159	NKDSNSMIK	LSH_mouse	176	QNSKFFFDPVR	7.085	TRUE