

**Structural and biochemical studies of proteins
implicated in Kaposi's Sarcoma-associated
Herpesvirus pathobiology**

Hyunah Lee

Thesis submitted to the University College London
for the degree of Doctor of Philosophy
2019

Division of Structural and Molecular Biology
UCL

Declaration

I, Hyunah Lee, confirm the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The Kaposi's sarcoma-associated Herpes virus (KSHV) infects hundreds of millions of people world-wide contributing to the development of Kaposi's sarcoma (the most common HIV-related cancer), Multicentric Castleman's disease and Primary effusion lymphoma. The project focused on two key proteins: KSHV SOX and KSHV-vFLIP that operate during the lytic and latent phases of the KSHV life-cycle respectively. KSHV SOX is a virally encoded highly conserved alkaline exonuclease which plays a key role in the global and rapid degradation of host mRNA in a process termed host shutoff as well as the packaging of viral genomes into capsids following replication. Although published (Bagneries et al., 2011) and unpublished (Lee, 2015, Master's thesis, UCL) crystal structures of SOX bound to DNA has given key insights into the mechanism of DNA recognition and cleavage, it was unclear how RNA could be recognised owing to the lack of canonical RNA binding motifs and therefore uncertain whether the mechanism of cleavage proposed for the exonucleolytic cleavage of DNA substrates would apply. A combination of structural and biochemical approaches were therefore used to address these key outstanding questions which are presented in this thesis. They include the first crystal structure of SOX-E244S mutant bound to the KSHV pre-miRNA K2-31 which was determined at 3.3 Å by molecular replacement. Analysis of this structure has revealed a distinct binding mode for RNA binding relative to DNA in which the "bridge" motif, spanning the C- and N-terminal lobes of the SOX molecule, has an essential role in substrate

recognition. Despite these differences, endonucleolytic processing of RNA transcripts could still be achieved by the S_N2 (bimolecular nucleophilic substitution) mechanism originally proposed for the exonucleolytic cleavage of DNA. The results of biochemical and biophysical assays performed on wild-type and mutant proteins defective in host shut off further revealed that the degradation of RNA is largely sequence non-specific, but requires structured elements such as stem loop or bulge motifs. They also revealed that impaired host shut off activity could be explained by impaired RNase activity against structured substrates for a subset of the mutants. The ability to compare SOX-DNA and SOX-RNA structures afforded by these studies lead to the identification of phytic acid as a potential inhibitor of both its DNase and RNase activities confirmed by the crystal structure of SOX-phytic acid complex refined to 2.3 Å. Here it is shown that phytic acid not only has the potential to physically block the association of DNA/RNA substrates but also inhibit nucleolytic cleavage by directly co-ordinating to a catalytically important magnesium ion.

During latency, KSHV produces a limited number of proteins essential for its survival. One of these is the viral FLICE inhibitory proteins (vFLIPs) that directly activates the canonical and alternative NF- κ B pathways resulting in increased cell proliferation, transformation, cytokine secretion, and protection against growth factor withdrawal-induced apoptosis. Previous studies have shown that KSHV-vFLIP forms a ternary complex with the transcription factor p100 and the kinase IKK α for persistent activation of the alternative pathway shown to have an important role in cellular transformation. Although the mechanism underlying this process is unclear, a physical interaction between KSHV-vFLIP and p100 has been implicated. Having

successfully produced all three proteins in *E. Coli* and baculovirus expression systems, the nature of this complex was investigated using pull down assays and site-directed mutagenesis. From these results, it has been possible to deduce that KSHV-vFLIP interacts with residues 860-900 located at the C-terminal end of p100 and does not interact with the death domain of p100.

Impact Statement

The findings of this thesis provide the first novel atomic structure of a viral RNase bound to RNA and has been recently published in *Nucleic acids* journal (Lee. H., Patschull. A., Bagneris. C., Ryan. H., Sanderson. C., Ebrahimi. B., Nobeli. I., and Barrett.T. KSHV SOX mediated host shutoff: the molecular mechanism underlying mRNA transcript processing (2017). *Nucleic Acids Research*). These studies have provided a platform for the identification of potential inhibitors which is further analysed in Results chapter 5 and a manuscript will be prepared for publication. Moreover, this work will be expanded by other members of the Barrett group for the identification of more potent inhibitors or compounds for KSHV anti-cancer therapy. This thesis also describes the involvement of KSHV-vFLIP in the alternative NF- κ B pathway which sheds light on another defence mechanism utilised by KSHV to counteract human immune surveillance systems. Together these findings could aid in the development of novel drugs or anti-viral therapies that specifically target KSHV.

Acknowledgements

First and foremost, I am deeply thankful to Dr Tracey Barrett for supervising this PhD. Throughout my PhD, she provided me tremendous support, advice, invaluable discussion about ideas and results which I found extremely helpful.

I would like to thank Dr Claire Bagneris for the support and help in the Franklin lab and her contribution to the work presented in the thesis; Dr Ambrose Cole for advice and discussion on collecting X-ray data. I would like to also thank my PhD committee members: Prof John Christodoulou and Prof Chris Kay for advice and invaluable discussions. I would like to thank Prof Snezana Djordjevic and Miss Manu Davies for advice and support at UCL. I am especially grateful to all our Franklin lab members for their help, technical support, sharing equipment and reagents and being wonderful friends. I am grateful to all our collaborators for the opportunity to be involved in interesting projects and their support.

And finally I would like to thank my family; especially my mum, dad and my brother who have always supported me and pushed me to achieve my goals. I have to give special thanks to my closest friends for supporting me throughout my PhD and being always there for me.

Table of Contents

Abstract	3
Impact Statement	6
Acknowledgements	7
List of Figures	14
List of Tables	17
Abbreviations	18
Chapter 1: Introduction	26
1.1 Human Herpesviruses.....	26
1.2 Kaposi's Sarcoma-associated Herpesvirus (KSHV)	29
1.2.1 Infection and lifecycle of KSHV.....	31
1.2.2 KSHV miRNAs.....	34
1.3 Mechanisms of Viral host shutoff	36
1.3.1 HSO in γ -herpesviruses	37
1.3.2 The mechanics of DNA and RNA turnover in BGLF5 and KSHV SOX	39
1.4 The DNase activity of KSHV-SOX and its involvement in viral DNA packaging	42
1.5 The structural basis of DNA processing by SOX.....	43
1.6 The RNase Activity of KSHV-SOX.....	45
1.7 Evidence of structured RNA recognition by SOX.....	47
1.8 KSHV-associated antiviral therapy	51

1.9 Structure-based design of KSHV-SOX inhibitors	53
1.9.1 Cancer preventive and therapeutic properties of Phytic acid	53
1.10 Viral latency and the NF- κ B pathway.....	57
1.10.1 NF- κ B	57
1.10.2 NF- κ B related proteins.....	58
1.10.3 NF- κ B family.....	59
1.10.4 I κ B Kinase (IKK) complex.....	62
1.10.5 The catalytic subunits: IKK α and IKK β	63
1.10.6 The regulatory protein NEMO/IKK γ	65
1.10.7 Other IKK-associated components.....	66
1.11 Activation of the Canonical NF- κ B pathway	67
1.12 The Non-canonical (alternative) NF- κ B pathway.....	69
1.13 P100 processing and the p100 death domain (DD).....	71
1.14 KSHV- encoded viral flce inhibitory protein (vFLIP)	72
1.15 Aim.....	76
Chapter 2: Biophysical background theory	78
2.1 X-ray Crystallography	78
2.1.1 Theory of Crystallisation and practical approaches.....	78
2.2 Crystal geometry.....	81
2.3 X-ray scattering and Bragg's law	83
2.4 Data collection.....	86
2.5 Structure factors and the Fourier Transform.....	88
2.6 The Phase problem.....	90

2.7 Solving the phase problem by molecular replacement (MR)	91
2.7.1 Patterson-based MR.....	92
2.7.1.1 Rotation function.....	95
2.7.1.2 Translation function.....	95
2.7.2 Maximum likelihood molecular replacement.....	97
2.7.2.1 Likelihood based rotation function.....	98
2.7.2.2 Likelihood based translation function	100
2.8 Electron density maps.....	101
2.9 Refinement and validation of macromolecular models	102
2.10 Biophysical methods to probe protein-ligand interactions.....	105
2.10.1 Fluorescence Anisotropy (FA) background theory	105
2.10.2 ThermoFluor Assay.....	109
Chapter 3: Materials and Methods	111
3.1 WTSOX Protein Expression and Purification.....	111
3.2 Preparation of SOX mutants	111
3.2.1 Mutagenesis	111
3.2.2 Expression and Purification of SOX mutants	114
3.3 Protein analysis and quantification	114
3.3.1 Protein concentration	114
3.3.2 Sodium dodecyl sulfate (SDS)-Polyacrylamide gel electrophoresis (PAGE)	115
3.3.3 Western Blots.....	115
3.4 Preparation and Annealing of RNA substrates.....	116
3.5 RNase assays using TBE-Urea gel electrophoresis	118

3.6 Fluorescence polarization anisotropy (FPA)	119
3.7 Crystallization of SOX-inhibitor complexes	120
3.7.1 Formation of the SOX-inhibitor complexes	120
3.7.2 Initial Crystallization trials of SOX:Phytic acid and SOX:Pyranine complexes	120
3.8 ThermoFluor Assay	121
3.9 Cloning, expression and purification of p100 and IKK α	121
3.10 p100 and IKK α Expression	126
3.10.1 Transposition of pFASTBAC-p100 and pFASTBAC-IKK α into bacmids	126
3.10.2 Transfection of Insect cells and protein overexpression	127
3.10.3 Expression analysis using SDS-PAGE	128
3.11 Purification of p100 and IKK α	128
3.12 Expression and Purification of p100 short c-terminal constructs	131
3.12.1 Expression of pETM6T1-p100 746-860 and pETM6T1-P100 746-848	131
3.12.2 Purification of pETM6T1-P100 746-860 and pETM6T1-P100 746-848	132
3.12.3 Expression and Purification of p100 746-900	134
3.13 Expression and Purification of KSHV-GB1-vFLIP	134
3.14 Expression and Purification of untagged KSHV-vFLIP	136
3.15 Pull Down Assays	136

Chapter 4: Structural and Functional Analysis of the interaction between KSHV-SOX and the pre-micro RNA fragment K2-31	137
4.1 SOX-RNA processing targets stem loop or bulge motifs.....	137
4.2 Solving the SOXE244S:K2-31 crystal structure by molecular replacement and Refinement.....	138
4.3 Interactions between SOX and K2-31	146
4.4 Comparison of the SOX:DNA and RNA binding modes.....	148
4.5 Potential mechanism of SOX-mediated RNA degradation	152
4.6 SOX-mediated RNA endonucleolytic cleavage appears to be sequence non specific.....	155
4.7 Probing the importance of stem loop size.....	158
4.8 Binding affinity of SOX for different RNA substrates.....	159
4.9 The HSO mutants have distinct functions in SOX-mediated host mRNA degradation	160
4.10 Discussion	162
4.10.1 Structure of SOX-K2-31 and its comparison to the SOX:DNA complex 3POV	164
4.10.2 SOX-mediated turnover appears not to require a targeting sequence	167
4.10.3 The HSO mutants have distinct roles in SOX-mediated RNA degradation	168
4.10.4 Comparison between RNase activity of EBV BGLF5 and KSHV SOX.....	169
 Chapter 5. Structural Analysis of SOX bound to Phytic acid.....	171
5.1 Phytic acid and Pyranine as two potential inhibitors of SOX.....	171
5.2 Crystallisation of SOX with Phytic acid.....	172
5.3 Structure determination of the SOX:Phytic acid complex.....	174

5.4 Phytic acid binds to key residues implicated in catalysis	176
5.5 ThermoFluor Studies of SOX:Phytic acid complex.....	181
5.6 Discussion	182
Chapter 6: Towards understanding the process of KSHV-vFLIP mediated subversion of the non-canonical NF-κB pathway	185
6.1 Expression and purification of IKK α	185
6.2 Expression and purification of p100	186
6.3 Pull-down assays.....	187
6.4 Interaction between p100 and KSHV-vFLIP	191
6.5 Pull-down assay to probe the interaction between KSHV-vFLIP and the death domains of p100	192
6.6 Discussion	196
Future work	200
Summary and Conclusions.....	203
Bibliography	205

List of Figures

Figure 1.1: Classification of the <i>Herpesviridae</i> Family.....	28
Figure 1.2: Schematic diagram of KSHV gene map.....	30
Figure 1.3: Schematic diagram of <i>de novo</i> infection of KSHV	32
Figure 1.4: A schematic diagram showing the seven conserved motifs of SOX	40
Figure 1.5: Crystal structure of SOX bound to DNA.....	44
Figure 1.6: Schematic diagram representing the locations of the residues involved in non-catalytic host shutoff.....	47
Figure 1.7: TBE urea gel demonstrating that HSO mutants	48
Figure 1.8: The predicted lowest energy structures of GFP51, DsRed61 and HBB58.....	51
Figure 1.9: Structure of phytic acid	55
Figure 1.10: Overview of NF- κ B related proteins.....	61
Figure 1.11: IKK α , IKK β and NEMO form the core of the IKK complex	63
Figure 1.12: Crystal structure of dimeric IKK α	64
Figure 1.13: Crystal structure of IKK β	65
Figure 1.14: Overall dimeric structure of CC2-LZ domains of IKK γ	66
Figure 1.15: Overview of the canonical and non-canonical NF- κ B pathways	68
Figure 1.16: Schematic diagram showing the trimeric IKK α	70
Figure 1.17: Schematic diagram of p100.....	72
Figure 1.18: Crystal structure of KSHV-vFLIP and vFLIP-IKK γ complex.....	74
Figure 1.19: Western blot revealing that KSHV-vFLIP interacts with the C- terminus of p100	75
Figure 2.1: Crystallisation Phase Diagram.....	79
Figure 2.2: Crystal systems	81
Figure 2.3: Overview of the general geometry of a unit cell	82
Figure 2.4: Constructive interference and Destructive interference.....	84
Figure 2.5: X-rays Crystallography experimental set up.....	87
Figure 2.6: Schematic diagram of MR method	93
Figure 2.7: Simple 2D structure in a P2 unit cell.....	94

Figure 2.8: A random walk is used to generate a probability function for F_{hkl}	99
Figure 2.9: Fluorophores excited by vertically polarized light.....	106
Figure 2.10: Fluorescence Anisotropy Experimental setup.....	106
Figure 2.11: Overview of ThermoFluor analysis.....	110
Figure 4.1: Structure of K2-31 and the RNase assay of K2-31 treated with WT SOX.....	138
Figure 4.2: Initial $2F_o-F_c$ electron density map of K2-31.....	139
Figure 4.3: $2F_o-F_c$ electron density map of K2-31 after anisotropy correction.....	140
Figure 4.4: Improved K2-31 density after several rounds of refinement.....	141
Figure 4.5: Schematic overview of K2-31 (nucleotides U3-G28).....	142
Figure 4.6: Crystal packing diagram illustrating that key to formation of the lattice are RNA-RNA interactions	143
Figure 4.7: Several key residues in SOX form hydrogen bonds with K2-31	146
Figure 4.8: A20 forms a pi-stacking interaction with F179	147
Figure 4.9: Schematic diagram illustrating the SOX-K2-31 interactions and SOX-DNA interactions	148
Figure 4.10: View of the SOXA:DNA complex superimposed on the SOXE244S:K2-31 complex.....	149
Figure 4.11: Important interactions between the A20 base of K2-31 and F179 in the bridge region	150
Figure 4.12: SOX RNase assays using F179A and C247S mutants	151
Figure 4.13: Schematic diagram illustrating the potential cleavage geometry of K2-31 by SOX	153
Figure 4.14: The MC-Fold MC-Sym lowest energy secondary structure predictions for GFP51 substitutions.....	156
Figure 4.15: SOX RNase assays using GFP51 substitutions	156
Figure 4.16: Time course cleavage assays involving SOX and the substrates GFP51-UCUCU and GFP51	157
Figure 4.17: RNase assays using K2-31, K2-31A4 and K2-31A9	158
Figure 4.18: FPA binding curves obtained for WTSOX in the presence of the substrates GFP51, K2-31, HBB58 and dsUN51.....	159
Figure 4.19: Schematic diagram highlighting the location of the residues involved in HSO.....	160
Figure 4.20: SOX RNase assays using HSO mutants.....	161

Figure 5.1: Molecular Structures of pyranine and phytic acid.....	171
Figure 5.2: Crystals of WTSOX:Phytic acid.....	172
Figure 5.3: Stereoview of the initial Fo-Fc and 2Fo-Fc maps	175
Figure 5.4: Initial Fo-Fc and 2Fo-Fc maps of N-terminus of symmetry related monomer	175
Figure. 5.5: Stereoview of phytic acid together with the associated 2Fo-Fc omit map density	176
Figure 5.6: Interactions involved in SOX:Phytic acid complex.....	179
Figure 5.7: Phytic acid is co-ordinated to the magnesium ion Mg _A	180
Figure 5.8: The SOX:Phytic acid interactions involving the bridge region.....	181
Figure 5.9: Thermal denaturation curves for SOX and SOX in the presence of phytic acid	182
Figure 6.1: SDS-PAGE gel of IKK α	186
Figure 6.2: SDS-PAGE gels of p100	186
Figure 6.3: SDS-PAGE analysis of both p100 746-860 and p100 746-846	187
Figure 6.4: SDS-PAGE analysis of a pull-down assay involving p100, IKK α and KSHV-vFLIP	188
Figure 6.5: SDS-PAGE analysis of a pull-down assay involving p100, IKK α , RelB and KSHV-vFLIP.....	189
Figure 6.6: SDS-PAGE analysis of control pull-down assays involving RelB, p100 and KSHV-vFLIP.....	190
Figure 6.7: Putative interactions between p100 and KSHV-vFLIP investigated using gel filtration.....	191
Figure 6.8: Pull-down assay of p100 746-900 and KSHV-vFLIP	193
Figure 6.9: SDS-PAGE analysis of the pull-down assay to establish the location of the KSHV-vFLIP binding region of p100	194
Figure 6.10: Gel filtration trace of p100 746-846 with or without KSHV-vFLIP	195
Figure 6.11: Amino acid sequences of p100 701-900	197
Figure 6.12: Schematic diagram of possible mechanisms of KSHV-vFLIP	199

List of Tables

Table 1.1: Antitumour effects of IP6 <i>in vitro</i>	54
Table 2.1: The seven Bravais lattices and associated unit cell constraints.....	83
Table 3.1: Primers used for SOX mutants	111
Table 3.2: Overview of the PCR reaction	113
Table 3.3: RNA and DNA sequences	117
Table 3.4: Composition of in house 15 % Urea gels.....	119
Table 3.5: Oligonucleotide DNA primers used to generate the p100 and IKK α	122
Table 3.6: PCR Reaction composition	122
Table 3.7: The program used for PCR for p100 and IKK α	123
Table 3.8: Composition of Restriction Digest	124
Table 3.9: Composition of Gibson assembly reaction mixture	124
Table 3.10: Buffer compositions used for p100 purification	129
Table 3.11: Buffers used for IKK α purification	130
Table 4.1: Data collection and refinement statistics for the final SOXE244S:K2-31 model.....	144
Table 5.1: Crystallization conditions of KSHV-SOX	173
Table 5.2: Summary of data collection and refinement statistics for final model of the SOX:Phytic acid complex	177

Abbreviations

#

2D	Two-Dimensional
3D	Three-Dimensional
3'UTR	Three Prime Untranslated Region
5'UTR	Five Prime Untranslated Region
6-FAM	6-Carboxyfluorescein

A

Å	Ångström
APS	Ammonium Persulfate
ARD	Ankyrin repeat domain

B

β-TrCP	Beta-transducing repeats-containing proteins
Blast	Basic Local Alignment Search Tool
BCLAF	Bcl-2-associated transcription factor
Bp	Base Pair
BSA	Bovine Serum Albumin

C

°C	Degree Celsius
CC	Coiled coil
cFLIP	Cellular FLICE-like inhibitory protein
CCP4	Collaborative Computational Project, Number 4
CHUK	Conserved helix-loop-helix ubiquitous kinase

ciAP	Cellular inhibitor of apoptosis
CMV	Cytomegalovirus
COPD	Chronic obstructive pulmonary disease
Coot	Crystallographic Object-Oriented Toolkit
CV	Column Volume
CXCR2	C-X-C Motif Chemokine Receptor 2

D

DC	Dendritic cells
DC-SIGN	Dendritic cell-specific ICAM-3-grabbing nonintegrin
DD	Death domain
DED	Death effector domain
DISC	Death-inducing signalling complex
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
DR	Death receptor
Ds	Double stranded
dsDNA	Double stranded Deoxyribonucleic Acid
DsRed2	Red Fluorescent Protein
dsRNA	Double stranded Ribonucleic Acid
DTT	Dithiothreitol

E

EBV	Epstein-Barr Virus (HHV-4)
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electron Mobility Shift Assay
EPHA2	Ephrin type-A receptor 2

F

FA	Fluorescence Anisotropy also Fluorescence Polarization Anisotropy
FLICE	Fas-associated death domain-like interleukin-1 β -converting enzyme
FW	Forward primer

G

GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GFP	Green Fluorescent Protein
GRK2	G protein-coupled receptor kinase 2

H

HAART	Highly Active Anti-Retroviral Therapy
HBB	Human Beta-Globin
HCl	Hydrochloric Acid
HCMV	Human Cytomegalovirus
HEK293	Human Embryonic Kidney-derived 293
HHV	Human Herpesvirus
HIV	Human Immunodeficiency Virus
HLX	Helical domain
HLH	Helix-loop-helix
HPLC	High-Performance Liquid Chromatography
HSO	Host Shutoff
HSP	Heatshock protein
HSV	Herpes Simplex Virus
HTLV	Human T-lymphotropic virus

I

IBD	Inflammatory bowel disease
ICP	Infected cell protein
IEX	Ion Exchange
I κ B	Inhibitors of κ B
IKK	I κ B Kinase
IL	Interleukin
IPTG	Isopropyl β -D-1-Thiogalactopyranoside
IP ₆	Inositol hexakisphosphate

K

K	Kelvin
Kb	Kilo Base
K _d	Dissociation Constant
KD	Kinase domain
kDa	Kilo Dalton
KICS	KSHV inflammatory cytokine syndrome
KS	Kaposi's Sarcoma
KSHV	Kaposi's Sarcoma-associated Herpesvirus

L

LANA	Latency-associated nuclear antigen
LB	Luria Broth
LZ	leucine zipper

M

MCD	Multicentric Castleman's Disease
-----	----------------------------------

MAD	Multi-Wavelength Anomalous Dispersion
MES	2-(N-Morpholino) Ethanesulfonic Acid
MHV68	Murine Herpesvirus 68
MR	Molecular Replacement
MIR	Multiple Isomorphous Replacement
mRNA	Messenger Ribonucleic Acid
miRNA	Micro Ribonucleic Acid
MST	Microscale Thermophoresis
MT4	HTLV-1-transformed T-cell line 4
mTOR	Mammalian Target Of Rapamycin
MWCO	Molecular Weight Cut Off
N	
NCS	Non-Crystallography Symmetry
NEMO	NF- κ B essential modulator
NF- κ B	Nuclear factor κ light-chain enhancers of activated B cells
NIK	NF- κ B-inducing kinase
NLS	Nuclear Localisation Signal
NMD	Nonsense-mediated mRNA Decay
NUB	NEMO ubiquitin binding
NusA	N-Utilization Substance Protein A
O	
OD ₆₀₀	Optical Density 600 nm
ORF	Open Reading Frame
P	
PAGE	Polyacrylamide Gel Electrophoresis

PAN	Polyadenylated nuclear RNA
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEL	Primary Effusion Lymphoma
PEST	proline (P), glutamate (E), serine (S), and threonine (T)
PHENIX	Python-based Hierarchical Environment for Integrated Xtallography
PMSF	Phenylmethanesulfonyl fluoride
Pol I	Polymerase I
Pol II	Polymerase II
Pol III	Polymerase III
Poly(A)tail	Polyadenylation Tail

R

RA	rheumatoid arthritis
RHD	Rel-Homology Domain
RNA	Ribonucleic Acid
RNase	Ribonuclease
RS	reverse primer
Rta	Replication and transcriptional activator (ORF50)

S

SAD	Single-Wavelength Anomalous Diffraction
SCF/SCRF	Skp1-Culin-Roc1/Rbx1/Hrt-1-F-box
SDS	Sodium Dodecyl Sulfate
SDD	Scaffold/dimerisation domain
siRNA	Small Interfering Ribonucleic Acid
SIRAS	Single Isomorphous Replacement with Anomalous Scattering

SMARD	Single-Molecule Analysis of Replicating DNA
snRNA	Small Nuclear Ribonucleic Acid
SOX	Shutoff and Exonuclease
ssRNA	Single Stranded Ribonucleic Acid
T	
TAD	Transactivation Domain
TBE	Tris/Borate/EDTA
TBE-Urea	Tris/Borate/EDTA-Urea
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus Protein
TLR	Toll-like receptors
TNF α	Tumour necrosis factor alpha
TPM1	Tropomyosin 1
TWEAKR	Tumour Necrosis Factor-like Weak Inducer of apoptosis Receptor protein
U	
UL	Unique Long
ULD	Ubiquitin-like domain
V	
vFLIP	viral FLICE-inhibitory proteins
Vhs	Viral Host Shutoff
VZV	Varicella-zoster virus
W	
WTSOX	Wild-Type host Shutoff Exonuclease

X

Xrn1	Exoribonuclease I
xCT	Cystine-glutamate transporter
VEGFA	Vascular Endothelial Growth Factor A

Z

ZF	Zinc finger-like domain
----	-------------------------

Chapter 1: Introduction

1.1 Human Herpesviruses

Human herpesviruses (HHV) belong to the *Herpesviridae* family, whose members characteristically have relatively large, double-stranded, linear DNA genomes within an icosahedral capsid surrounded by a proteinaceous tegument layer. They are capable of encoding 100-200 genes that express viral proteins important for viral cell survival and infection of new host cells, causing various diseases in humans (Mettenleiter, Klupp, & Granzow, 2009). The *Herpesviridae* family can be classified into α -, β -, and γ -herpesviruses depending on their genome sequences, tissue tropism and the duration of the replication cycle (Figure 1.1). Human α -herpesviruses are known to primarily target mucoepithelial cells. Their reproductive cycle is short, which leads to the development of latency in sensory ganglia. These include herpes simplex virus 1 and 2 (HSV-1 and HSV-2, also known as human herpesvirus-1, HHV-1, and HHV-2, respectively) which predominantly cause oral blisters and genital blisters, respectively. Varicella-zoster virus (VZV, HHV-3), classified as an α -herpesvirus, causes shingles and chickenpox (Wang, Baldi, & Gaut, 2007; Liu et al., 2007). Human β -herpesviruses are characterised by long replication cycles and include human cytomegalovirus (HCMV, HHV-5), which mainly targets monocytes and epithelial cells and causes retinitis together with approximately 10 % of cases of infectious mononucleosis-like syndrome (Bravender, 2010; Brunson, Khoretonenko, & Stokes, 2016). Roseolovirus (HHV-6A), human B-

lymphotropic virus (HHV-6B) and HHV-7 are also β -herpesviruses that mainly target T cells and the cause of rashes that are extremely common (Brunson et al., 2016; Cruz-Muñoz & Fuentes-Pananá, 2018). Human γ -herpesviruses include Kaposi's sarcoma-associated herpesvirus (KSHV, HHV-8) and Epstein-Barr virus (EBV, HHV-4), whose main targets are lymphocytes and B cells, respectively, and are known to cause various lymphoproliferative disorders. For instance, EBV is the main causative agent of approximately 90% of mononucleosis and can cause Hodgkin lymphoma, nasopharyngeal carcinoma, and Burkitt lymphoma (Bravender, 2010; Brunson et al., 2016). KSHV is the most recently identified HHV and is the main causative agent of Kaposi's sarcoma (KS), which is the most common HIV-related cancer. It is also responsible for primary effusion lymphoma (PEL, a rare HIV-related cancer arising in body cavities) and Multicentric Castleman's disease (MCD) (Ganem, 2010; Kaplan, 2013). PEL is a lymphomatous effusion tumour developed from clonally expanded malignant B cells and is mainly found in body cavities manifested as pleural, pericardial, and peritoneal lymphomatous effusions (Wen & Damania, 2010). It is an incredibly aggressive lymphoma for which there are currently no effective treatments. On the other hand, MCD is a rare polyclonal B cell angio-lymphoproliferative disorder mainly found in the mantle zone of the lymph node follicle (Katano et al., 2000). KS is the most common malignancy in patients after organ transplantation, as well as the most common neoplasm in patients with untreated AIDS, especially in men in Sub-Saharan Africa (Raeisi et al., 2013; Cesarman et al., 2019). KS leads to the development of endothelial tumours and is often characterised by skin lesions, and other damaged areas including the lungs,

the liver, lymph nodes and the gastrointestinal tract.

The latest KSHV associated disease identified is inflammatory cytokine syndrome, or KICS. Patients with KICS show a subsequent increase in the level of both interleukin-6 and high viral loads (Uldrick et al., 2010).

The *Herpesviridae* Family

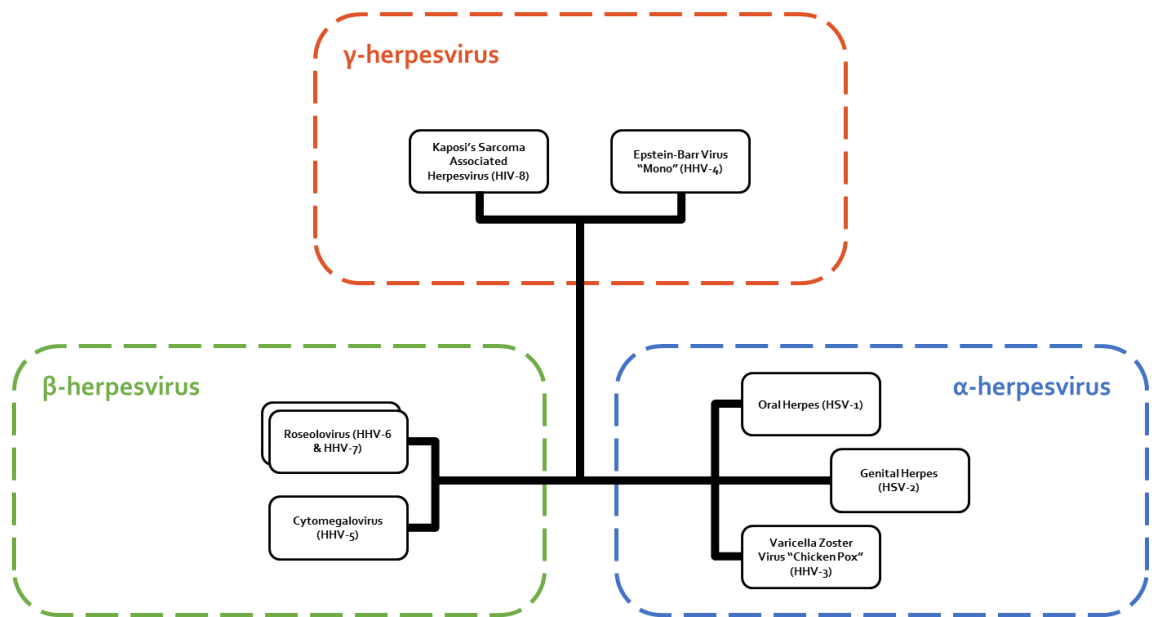


Figure 1.1: Classification of the *Herpesviridae* Family. α-herpesviruses include Herpes simplex virus 1 and 2 (HSV-1 and HSV-2), and Varicella Zoster Virus (HHV-3). β-herpesviruses include Roseolovirus (HHV6, HHV7), and Cytomegalovirus (HHV-5). γ-herpesviruses include Epstein-Barr Virus (HHV-4) and Kaposi's sarcoma-associated herpesvirus (HHV-8).

1.2 Kaposi's Sarcoma-associated Herpesvirus (KSHV)

The KSHV genome consists of a ORFs flanked at each end by variable number of a direct repeat giving a genome size of approximately 170 kb (Figure 1.2) (Rezaee et al., 2006; Russo et al., 1996). These ORFs (ORF4 to ORF75) are responsible for expressing 90 viral proteins whose functions for the majority are poorly understood. Those that have been characterised, however, reveal roles in the modulation of host-virus interactions, the biogenesis of structural virion components and viral replication (Yan et al., 2019). In addition, KSHV also encodes a unique set of genes (K1-K15) that are associated with the subversion of host cellular mechanisms, several long non-coding RNAs including polyadenylated nuclear RNA (PAN, 1.1kb long) that are highly abundant in the lytic phase, and a set of viral microRNAs (miRNAs) (Neipel et al., 1997; Russo et al., 1996; Staskus et al., 1997).

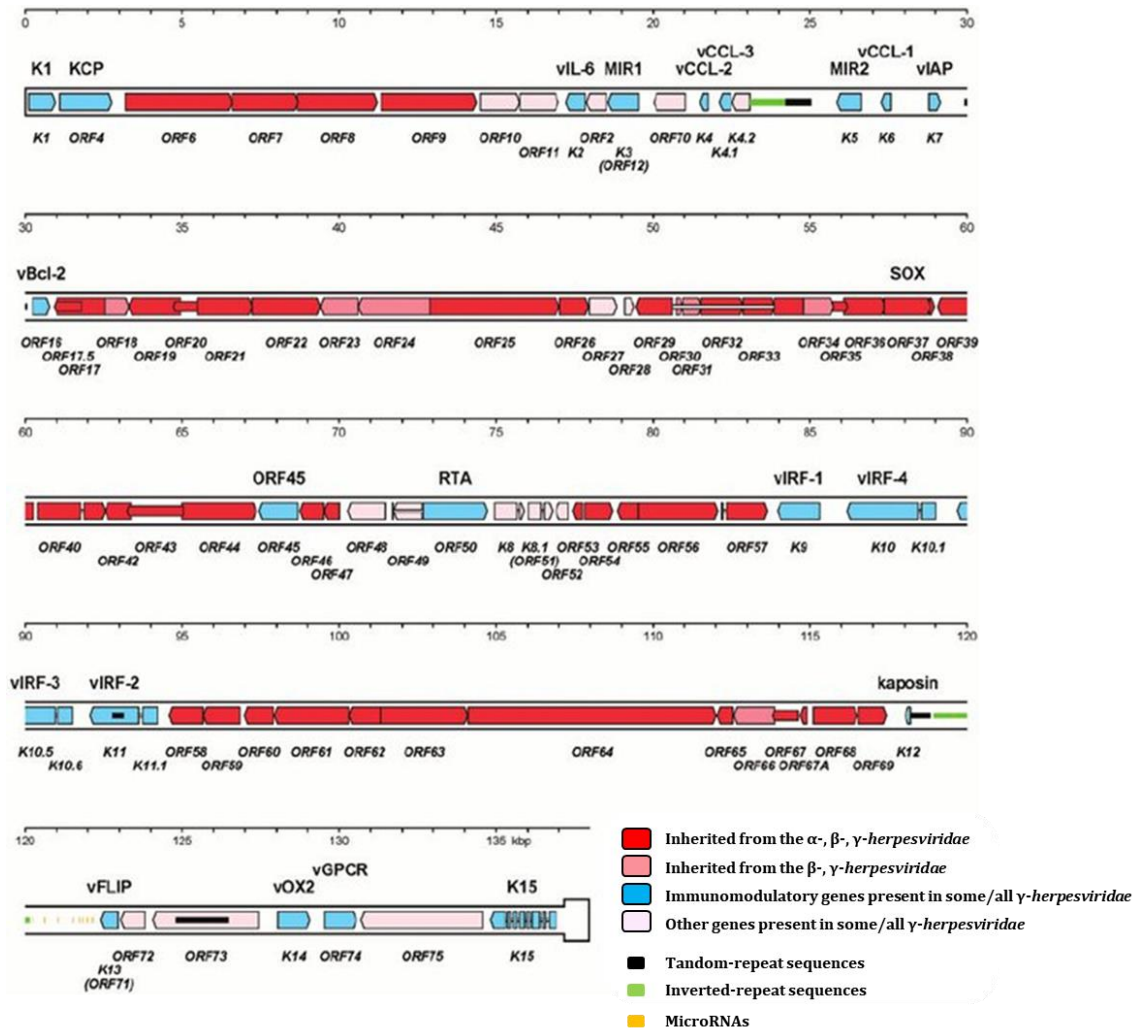


Figure 1.2: Schematic diagram of KSHV gene map. KSHV genome consists of a unique region flanked at each end by variable number of a direct repeat giving a genome size of approximately 170 kb (Figure adapted and modified from Rezaee et al., 2016).

1.2.1 Infection and lifecycle of KSHV

During infection, KSHV virions attach themselves via membrane fusion, mediated by the attachment of their viral glycoproteins such as gB, gH and gL, to the host cell membrane (Figure 1.3) (Pertel, 2002; Krishnan et al., 2005; Avitabile et al., 2009; Veettil et al., 2014; Yan et al., 2019). Host proteins that serve as KSHV binding receptors include heparan sulfate, integrins ($\alpha3\beta1$, $\alpha V\beta5$, and $\alpha V\beta3$), dendritic cell-specific ICAM-3-grabbing nonintegrin (DC-SIGN), tyrosine protein kinase receptor ephrin type-A receptor 2 (EPHA2), and the cystine-glutamate transporter xCT (Hensler et al., 2014; Hahn et al., 2012; Rappocciolo et al., 2006). Depending on the target cells, the modes of attachment as well as the types of receptors that KSHV binds to varies (Jarousse et al., 2008). For example, KSHV can bind to integrins and EPHA2 for attachment to induce macropinocytosis (fluid endocytosis resulting in a suspension of large vesicles inside the cell) in certain cell types (Dutta et al., 2013). In contrast, clathrin-mediated endocytosis is used to enter into human foreskin fibroblast cells. Once KSHV enters the target cell, it deposits its viral capsid and tegument proteins into the cytoplasm. Here, with the aid of microtubules, the capsid is transported into the nuclear pore where it injects its double-stranded linear DNA into the nucleus. Once inside the nucleus, the genome becomes rapidly circularised in association with cellular histones. This ensures the protection of viral DNA ends from nucleases that operate in host DNA damage response mechanisms, the initiation of viral DNA replication and regulation of viral gene expression (Uppal et al., 2015) (Figure 1.3).

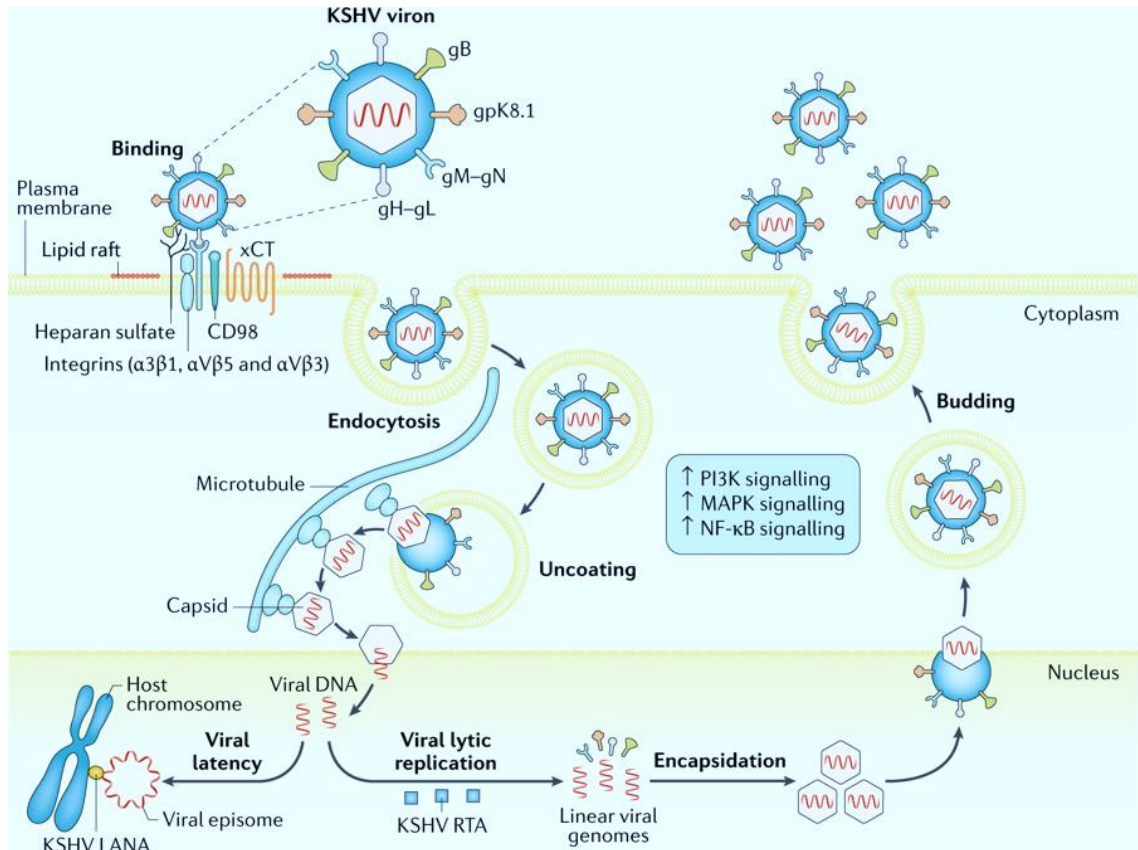


Figure 1.3: Schematic diagram of *de novo* infection of KSHV. Once KSHV is attached to the host cell, the viral capsid containing double-stranded linear DNA is deposited into the cytoplasm. Subsequently, in association with microtubules, the viral DNA is ejected into the nucleus where it undergoes rapid circularisation and initiates replication and regulation of viral gene expression. (Figure adapted and modified from Cesarman et al., 2019).

The life cycle of KSHV is not completely understood; however, it is known to be biphasic, consisting of a latent phase (infection phase) and a lytic phase (reactivation phase) (Cai et al., 2010; Glaunsinger & Ganem, 2004). During this latent phase, the viral genome remains circularised (as an episome) and associated with chromatin which prevents viral replication. Replication of the viral episome, however, is

achieved through expression of the virally encoded protein latency-associated nuclear antigen (LANA) (Lieberman, 2013). LANA physically tethers the viral genome to host chromosomes enabling replication of the viral episomes through co-option of host DNA polymerases and other replication factors (Lieberman, 2013). Detailed analysis of the proteins produced during latency revealed that many share significant homology with those within the host and operate through de-regulating normally tightly controlled cellular mechanisms (Ballestas et al., 1999; Glaunsinger & Ganem, 2004) (Figure 1.3). It is unclear what triggers the transition from the latent to lytic phase. Physiological stimuli and external/environmental factors such as hypoxia and oxidative stress have been shown to disrupt the latent state (Davis et al., 2001; Ye et al., 2011) which may lead to reactivation of viral DNA replication leading to induction of the lytic phase (Chen et al., 2001; Haque et al., 2003).

During the lytic phase, the majority of viral genes (~ 130 kb) encoding what are referred to as lytic transcripts are expressed to produce viral proteins (Figure 1.3). This sequence of events is tightly controlled temporally where the viral genome is expressed in “waves” where expression follows a sequential order of immediate early (IE), early (E) and late (L) genes (Jenner et al., 2001). Typically, IE genes include transcription factors and regulators such as ORF50 and ORF45. They are responsible for encoding regulatory proteins for viral replication, while L genes encode most of the structural components of the virions. Many E genes function in assembly of the highly conserved machinery for viral DNA replication and therefore have enzymatic functions (ORF59). They are also involved in modulation of the immune system (modulator of immune recognition 1/2 (MIR1/2)), regulation of gene expression,

and mediation of RNA turnover of the host mRNA (ORF37). Upon expression of lytic genes, virus assembly occurs in the nucleus (Figure 1.3) (Gradoville et al., 2000).

1.2.2 KSHV miRNAs

MicroRNAs (miRNAs) are a class of short (19-23 nucleotides), single-stranded non-coding RNAs which function primarily in gene regulation at the post-transcriptional level by binding to target mRNAs (Ameres & Zamore 2013). In KSHV, there are 25 mature miRNAs that originate from 12 pre-miRNAs (K1-12), which are important players in cell immune responses, regulation of the KSHV life cycle and virus-induced angiogenesis (Pfeffer et al., 2005; Grundhoff et al., 2006). Mature miRNAs are expressed during latency in KS or MCD patients (Sullivan, 2007). Studies have shown that varying levels of some specific miRNAs are expressed at different phases of the KSHV life cycle (Qin et al., 2017). For example, viral latency is maintained by K12-7, K12-9, and K9-5p through inhibition of replication and downregulated expression of the transcription activator Rta (ORF50) (Lin et al., 2011). In addition, the cellular G protein-coupled receptor kinase 2 (GRK2) is directly targeted by K3, which subsequently initiates C-X-C Motif Chemokine Receptor 2/protein kinase B (CXCR2/AKT) signalling induction of cell migration and invasion resulting in the maintenance of viral latency (Hu et al., 2015; Li et al., 2016b). K3 and K11 also target cellular factors including BCL2 associated transcription factor 1 (BCLAF1), nuclear factor (I/B), and I κ B α to promote latency (Lu et al., 2010; Ziegelbauer et al., 2009). Some KSHV miRNAs function by mimicking cellular miRNAs to target the same

transcripts. For example, K12-11 mimics host cell-miRNA 155 (Skalsky et al., 2007).

To promote long-term latency, host immune-surveillance mechanisms are also affected by KSHV miRNAs, which include K12-11 that targets IKK ϵ and causes suppression of antiviral immunity in the IFN signalling pathway (Liang et al., 2011). Additionally, K5 and K9 target the myeloid differentiation primary response protein 88 (MYD88) and the 3'UTR of interleukin-1 receptor (IL-1R)-associated kinase (IRAK1), respectively, which leads to the overexpression of cellular miRNAs and subsequently decreased levels of IL-6 and IL8 (Abend et al., 2012). This inhibits the innate immune response and hence benefits the virus. K10a functions by binding to tumour necrosis factor-like weak inducer of apoptosis receptor protein (TWEAKR), which results in the suppression of inflammatory cytokine (IL8) and inhibition of TWEAK-induced apoptosis (Abend et al., 2010).

Several KSHV encoded miRNAs have been implicated in angiogenesis. These KSHV miRNAs include K5 and K2, which bind to tumour suppressor protein tropomyosin 1 (TPM1) resulting in the expression of Vascular Endothelial Growth Factor A (VEGFA) and tube formation that facilitates pathogenesis and angiogenesis of KSHV (Kieffer-Kwon et al., 2015). Interestingly, miRNA K12-2, an RNA polymerase II transcript, suppresses the expression of high molecular weight tropomyosin 1 splice variants (HMW-TPM1) thereby regulating cytoskeletal organisation (Kieffer-Kwon et al., 2015). The exact manner in which K12-2 does this is still unclear. However, studies have shown that this suppression leads to downregulated anoikis (a type of anchorage dependent programmed cell death) which has been linked to metastasis

in several non-viral cancers. In addition, the over expression of K12-2 may have an important role in viral trafficking and it is interesting to speculate that SOX may play as an important contributor in pre-miRNA maturation especially given that RNA polymerase II transcripts that so far appear to be the sole targets of SOX (Feng et al., 2012).

1.3 Mechanisms of Viral host shutoff

Several viruses inhibit or down-regulate cellular gene expression in the host cell, resulting in a subsequent reduction of cellular transcripts and proteins in a phenomenon known as host shutoff (HSO). In order to promote and enhance viral translation, upon infection, HSO functions by degrading cellular mRNAs allowing “hijacking” of the cellular translational machinery to promote viral gene regulation (Thompson & Sarnow, 2003). Consistently, *in vivo* studies have suggested that HSO plays a key role in evading immune surveillance mechanisms in host cells, which in turn enhances viral replication (Smiley, 2004; Smith et al., 2002; Strelow & Leib, 1995).

In 1965, HSO was first identified in the α -herpesviruses HSV-1 and HSV-2 (Roizman et al., 1965) and since then much work has been carried out, especially on HSV-1. These studies have shown that in HSV-1, HSO requires the viral proteins 27 (ICP27, encoded from unique long segment 54, UL54) and the virion host shutoff protein vhs (Smiley, 2004). ICP27 contributes to the inhibition of splicing of cellular mRNA by redistributing the major components of the spliceosome, the snRNP (Sciabica, Dai,

& Sandri-Goldin, 2003). In addition, ICP27 also promotes transcriptional activation of viral genes and mRNA export upon viral infection (Sandri-Goldin, 2011). vhs is encoded by the HSV-1 gene UL41 and acts as an mRNA-specific RNase (Everly et al., 2002). HSV-1 vhs operates by degrading host mRNAs thus promoting host shutoff (Kwong & Frenkel, 1987). However, vhs does not support viral replication, as seen in deletion mutation studies where most wild type viral replication activity is seen in tissue culture even in the absence of vhs (Kwong & Fenkel, 1987). On the other hand, vhs mutants in a virus-infected mouse model have shown that it is essential for pathogenesis and establishment of latent infection (Strelow & Leib, 1995). This is due to the failure of these mutants to suppress the interferon response, which subsequently disables the virus' ability to evade the host immune responses (Duerst & Morrison, 2004). vhs is also found in other members of the human α -herpesvirus family (but not in β - or γ -herpesviruses) including varicella-zoster virus; however their activity is not as extensive as that seen in HSV-1 (Smiley, 2004). Interestingly, HSO activity had not been found in β -HHV (Smiley, 2004).

1.3.1 HSO in γ -herpesviruses

There are three host shutoff proteins found in γ -herpesvirus: MHV-68 ORF37, KSHV-SOX and EBV BGLF5 (Glaunsinger & Ganem 2004; Rowe, 2007; Covarrubias et al., 2009). The genes encoding these host shutoff proteins are conserved across the γ -herpesvirus family.

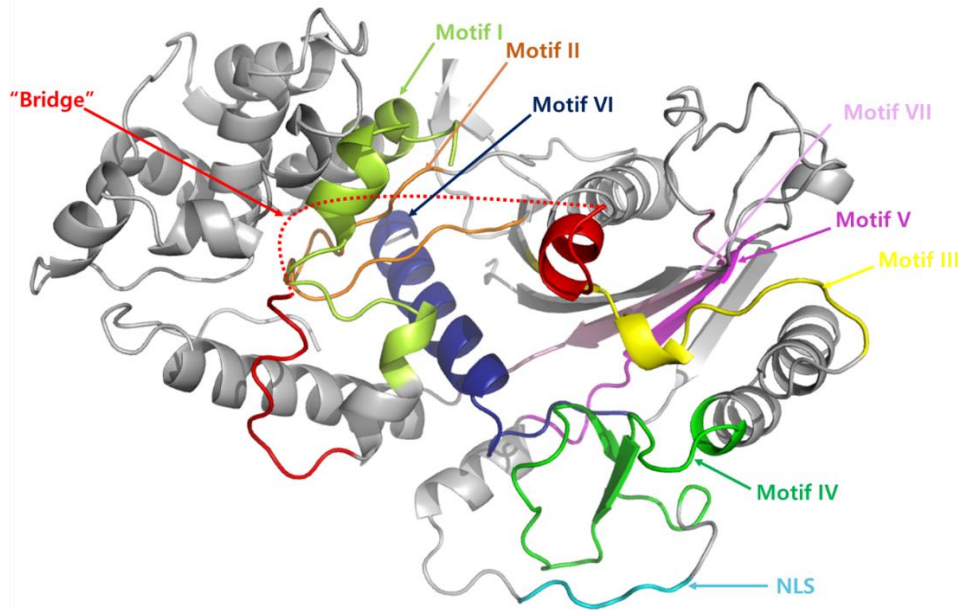
HSO in γ -herpesviruses was first identified in MHV68 using microarray technology

(Ebrahimi et al., 2003). This observation was then followed by the identification of HSO during the lytic phase of KSHV-infected cells (Glaunsinger & Ganem 2004). Previous studies identified ORF37 as a factor involved in the resection (DNase activity) of viral genomes (to be further discussed in section 1.4) following replication. On this basis it was classified as an alkaline exonuclease that is widely conserved amongst herpesviruses. However, ORF37 was subsequently shown to also mediate HSO resulting in the suppression of gene expression in host cells (Glaunsinger et al., 2005). This additional RNase activity of ORF37 could be genetically separated from its DNase activity and appears to be distinct to γ -herpesviruses. The fact that ORF37 possesses both host shutoff and exonuclease activities prompted the name change to SOX (ShutOff and Exonuclease) (Glaunsinger & Ganem 2004). The related EBV gene BGLF5 also has host shutoff activity (Rowe, 2007).

1.3.2 The mechanisms of DNA and RNA turnover in BGLF5 and KSHV SOX

A combination of structural biology, cell biology and biochemistry gave the first insights into host shutoff in the γ -herpesviruses and differences in the processing of DNA and RNA substrates (Dahlroth et al., 2009; Bagneris et al., 2011; Buisson et al., 2009). The crystal structures of BGLF5 and KSHV SOX bound to DNA confirmed a type II-restriction enzyme like fold (Buisson et al., 2009; Bagneris et al., 2011). The crystal structure of Apo SOX (PDB=3FHD) was solved to a resolution of 1.85 Å and can be subdivided into two large domains: five β -sheets, each flanked by an equal number of α -helices to form a C-terminal domain (amino acid residues from 7 to 164 and 446 to 481), while 10 α -helices form the N-terminal domain (amino acid residues from 179 to 445) (Dahlroth et al., 2009). Several structural studies have shown that SOX contains highly conserved motifs (Figure 1.4): Motifs I-VII, which are homologous to the members of the type II restriction enzyme like families (Dahlroth et al., 2009; Bagneris et al., 2011).

A



B

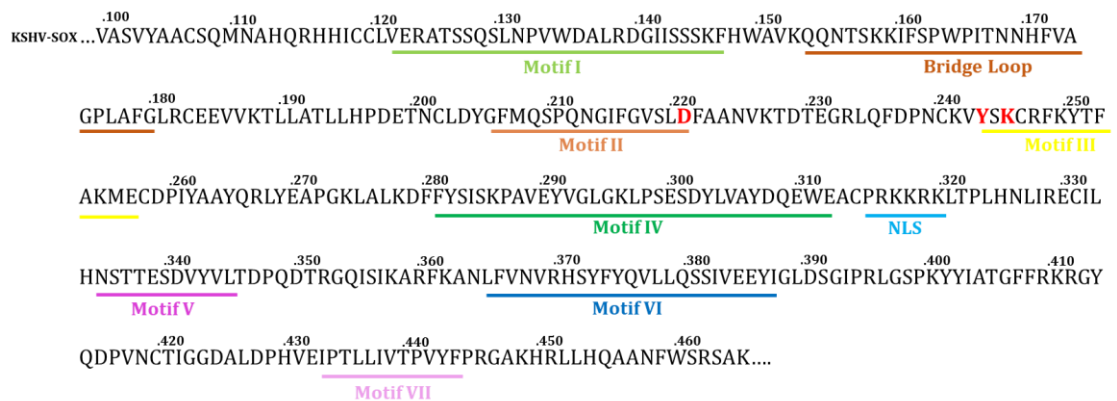


Figure 1.4: (A) A schematic diagram showing the seven conserved motifs of SOX, nuclear localisation signal (NLS) and “Bridge” in SOX. Motif I (lime), Motif II (orange), Motif III (yellow), Motif IV (green), Motif V (magenta), Motif VI (blue), Motif VII (pink), “bridge” (dark orange) and NLS (cyan). (PDB=3FHD, figure adapted and modified from Dahlroth et al., 2008). (B) The corresponding amino acid sequences in these motifs. The PD-(D/E)XK (D221 in motif II, E244 and K246 in motif III) sequences are highlighted in red.

Previous experiments have shown that these seven motifs are crucial for their proper nuclease activity (Bagneris et al., 2011; Goldstein & Weller, 1998). In addition, this PD-(D/E)XK (D221 in motif II, E244 and K246 in motif III) sequence was found in many other virally encoded proteins such as ribonucleases (RNase) that endonucleolytically process RNA substrates (Bujnicki & Rychlewski., 2001; Morin et al., 2010; Yuan et al., 2009). Motif III and Motif VI are involved in substrate binding and structural studies involving SOX bound to a 20mer DNA duplex have confirmed that Motif III and residues K246, F249, K250 and Y373, found in Motif VI, are involved in substrate recognition (Bagneris et al., 2011). The active site is found at the centre of the SOX protein in a region known as the “canyon” that subdivides the SOX structure into N- and C-terminal lobes. These are spanned by an unstructured polypeptide known as the “bridge” (Bagneris et al., 2011). This bridge structure is comprised of residues P164 to G180 and is positioned above the active site. Therefore, the bridge was thought to be involved in the catalytic activity of SOX. Other unrelated endonucleases are also found to have bridge motifs; for example, the bacteriophage T5 flap endonuclease bridge structure shown to have an important role in substrate binding (Nishino et al., 2006; Tomlinson et al., 2010; AlMalki et al., 2016). A nuclear localisation signal (NLS), which is highly conserved throughout the members of the γ -herpes subfamily of HHV, has also been identified between Motifs IV and V, just below the active site (Bagneris et al., 2011; Dahlroth et al., 2009; Buisson et al., 2009; Glaunsinger et al., 2005).

The 5'-3' nuclease activity of BGLF5 against DNA and single stranded RNA substrates was found to be enhanced in the presence of Mn^{2+} compared to Mg^{2+} which

confirmed the requirement for di-valent metal ions (Buisson et al., 2009). Interestingly, the processing of single-stranded RNA substrates required 30 times more protein for complete degradation relative to DNA, suggesting much weaker RNase activity (Buisson et al., 2009). In keeping with this, the fluorescence anisotropy performed using KSHV SOX demonstrated poor affinity for single-stranded substrates, with a K_d of $\sim 87 \mu\text{M}$. These results therefore presented the possibility that SOX/BGLF5 might not be the main factor in RNA processing but that other cellular factors or viral proteins might also be required (Bagneries et al., 2011). Curiously, several non-catalytic mutants that impair the DNase activity of SOX in cells showed mostly wild-type RNase activity *in vitro* indicative of different configurations associated with the binding of DNA and RNA substrates (Glaunsinger et al., 2005; Dahlroth et al., 2009).

1.4 The DNase activity of KSHV-SOX and its involvement in viral DNA packaging

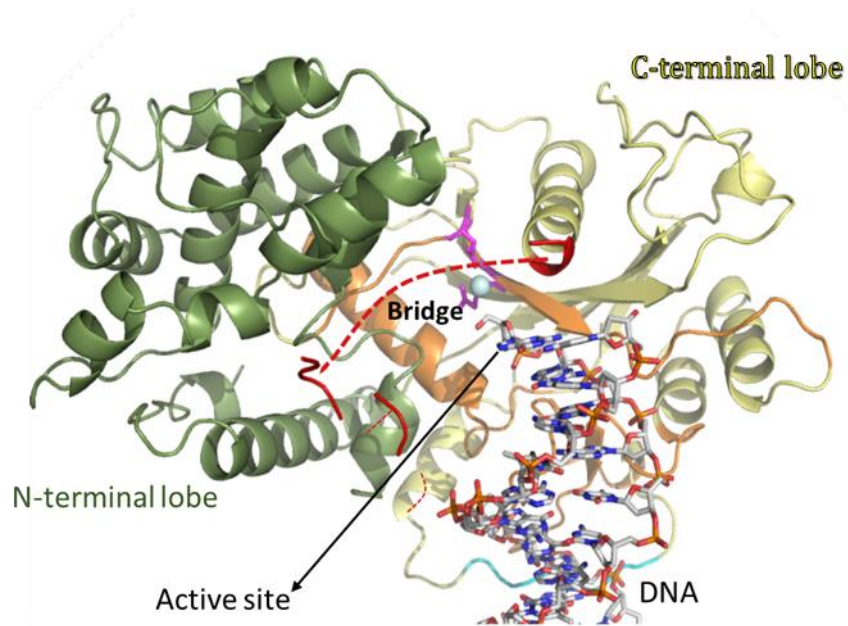
Replication of the KSHV genome results in the production of covalently linked concatemers consisting of multiple copies. These have to be resolved into single copy genomes prior to incorporation into capsids for the production of viable progeny viruses. This is achieved by means of the intrinsic 5'-3' DNase activity of SOX. The nature of the intermediates targeted is unclear, however, evidence obtained from HSV-1 suggests that alkaline exonuclease activity targets 3-way junctions (Martinez et al., 1996; Severini et al., 1994; Severini et al., 1996). The newly generated, single copy genomes are then encapsidated to form DNA-filled capsids (C capsids). Failure

to achieve this results in reduced viral egress and downstream infectivity (Buisson et al., 2009). After budding through the nuclear membrane, the C capsids mature in the cytoplasm, where they are tegumented and enveloped by the trans-Golgi network and subsequently released into the extracellular space to form a virion (Mettenleiter, 2002).

1.5 The structural basis of DNA processing by SOX

Initially, it was hypothesised that since most of the non-catalytic residues implicated in host shutoff are located at the N-terminal region distant from the catalytic centre, SOX might recognise its DNA substrates as structural motifs and is not sequence-specific (Buisson et al., 2009). In 2011, the crystal structure of KSHV SOX-DNA complex was solved confirming this and also the involvement of the PD-(D/E)XX motif and NLS for DNA substrate binding (Bagneries et al., 2011). In this structure, the bridge region was disordered and a role in endo or exo-nucleolytic cleavage could only be speculated (Figure 1.5A). Interestingly, two magnesium ions were identified associated with SOX, one in a canonical position in relation to E244 and D221 and the other, in a non-canonical position. Based on these and the position of the serine cluster known to function in sequestering a 5' phosphate, a putative model for cleavage was proposed consistent with a classical S_N2 -metal nuclease mechanism (Figure 1.5B) (Pingoud et al., 2005; Yang, 2011; Yang et al., 2006; Bagneries et al., 2011).

A



B

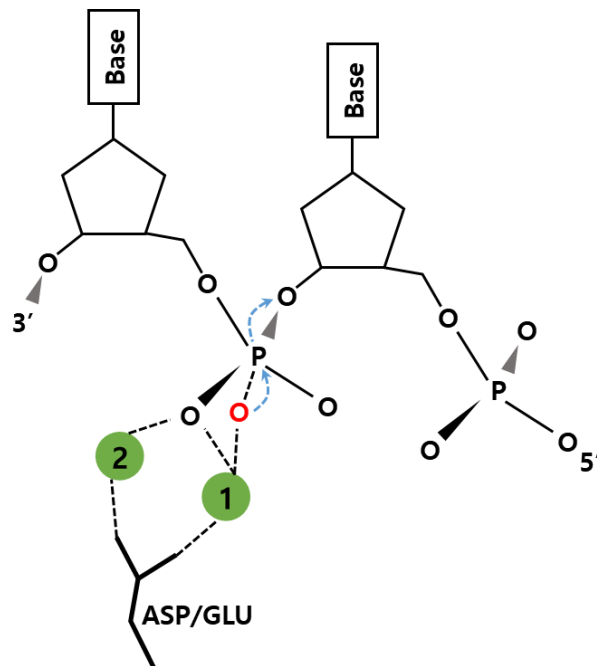


Figure 1.5: (A) Crystal structure of SOX bound to DNA (PDB=3POV, figure adapted and modified from Bagneries et al., 2011). (B) Classical S_N2 -metal nuclease mechanism that involves the carboxylate groups of E244 and D221 chelated to two divalent metal ions. (Figures adapted and modified from Bagneries et al., 2011).

Recently, Q129 was shown to have an essential role in DNA turnover (Uppal et al., 2019). A Q129H, mutant was reported to completely abolish the DNase activity of KSHV SOX. Consistent with biochemical observations, it has been shown to be crucial for the processing of KSHV's genome during DNA replication during the lytic phase. In addition, using single-molecule analysis of replicating DNA (SMARD) assays, electron microscopy, and pulsed-field and Gardella gel analysis, the Q129H mutant shows a more severe defect on the intranuclear processing of progeny virions, compared to the apo enzyme confirming that the DNase activity of SOX is important for viral particle maturation as well as processing and encapsidation of the viral genome (Uppal et al., 2019).

1.6 The RNase Activity of KSHV-SOX

Several studies have been performed to establish the exact mechanism underlying the RNase activity of KSHV SOX. In initial cell based experiments, human embryonic kidney-derived 293 (HEK293) cells expressing the mRNA-GFP (green fluorescent protein) reporter gene were transfected with active wild-type SOX (WTSOX). These were among the first to show that the nuclease activity of SOX significantly decreased the concentration of mRNA-GFP in the cytoplasm, but not in the nucleus. This indicated that SOX-mediated mRNA degradation is more likely to occur in the cytoplasm (Glaunsinger et al., 2005). To investigate this further, a GFP reporter plasmid, which allowed transcription by either RNA polymerase II (pol II), RNA polymerase I (pol I) or RNA polymerase III (pol III), was constructed and introduced

into separate cells. Once SOX was transfected, the levels of GFP-mRNA were quantified. According to the results, the expression of SOX caused a significant decline in pol-II-mRNA transcripts, while pol-I-and pol-III-mRNA transcripts were not significantly degraded (Covarrubias et al., 2011). Furthermore, the presence of a 5' cap on the host mRNA did not abolish SOX-mediated host mRNA degradation (Covarrubias et al., 2011). *In vivo* studies have demonstrated that SOX targets host mRNA transcripts that are being actively translated and that SOX over-expression causes a significant decline in polysome levels but increases the levels of eukaryotic ribosomes (80S) (Covarrubias et al., 2011). Moreover, it has been shown that the mRNAs present in the cytoplasm are associated with polysomes. These results suggest that targeting translationally active mRNAs enhances SOX-mediated turnover as part of an efficient mRNA degradation strategy (Hendrickson et al., 2009).

Although it has been shown that the same catalytic site is utilised for RNA and DNA degradation by SOX, residues were identified that were capable of inhibiting HSO while having no impact on its DNase activity based on the screening of random mutants in cell based assays (Glaunsinger & Ganem, 2004). These HSO mutants include T24I, A61T (within the N-terminal lobe), P176S (located in the bridge region), V369I (in close proximity to the active site region), D474N and Y477stop (both positioned at the far C-terminus) (Figure 1.6) (Buisson et al., 2009; Covarrubias et al., 2011; Bagneries et al., 2011). These mutants had differential effects on HSO, but the most profound defects were observed for P176S and V369I. However, the exact mechanism of how these residues contribute to the catalytic

activity of SOX was unclear due to their disparate nature and broad spatial distributions (Glaunsinger et al., 2005; Bagneries et al., 2011).

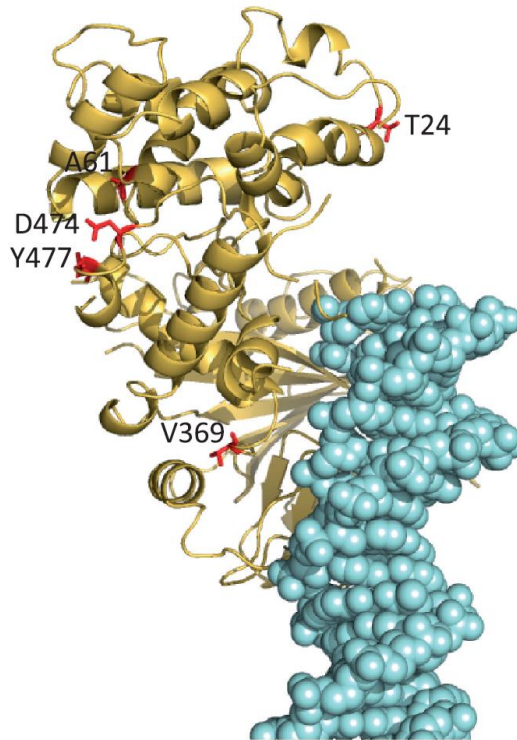


Figure 1.6: Schematic diagram representing the locations of the residues involved in non-catalytic host shutoff. (Figure adapted and modified from Bagneries et al., 2011).

1.7 Evidence of structured RNA recognition by SOX

The studies performed by the Barrett Group and Buisson et al. demonstrated that SOX is able to mediate degradation of single-stranded RNA (ssRNA), but curiously, the HSO mutants demonstrated near wild-type activity against ssRNA (Figure 1.7) (Bagneries et al., 2011). These experiments highlighted the possibility that SOX recognises either structured mRNA or a specific nucleotide sequence as well as

raising the prospect that it might have a more indirect role. These possibilities were investigated by the Glaunsinger group using three distinct, fluorescently labelled mRNA-reporter genes to establish any common features associated with cleavage. In addition to GFP, red fluorescent protein (DsRed2) and human beta-globin (HBB) reporter genes were analysed and the degradation products quantified (Covarrubias et al., 2011). The results indicated that SOX binds to similar lengths of mRNA (1.2-1.5kb) but cleaved each of the reporter-genes into variable lengths of mRNA intermediates (Covarrubias et al., 2011).

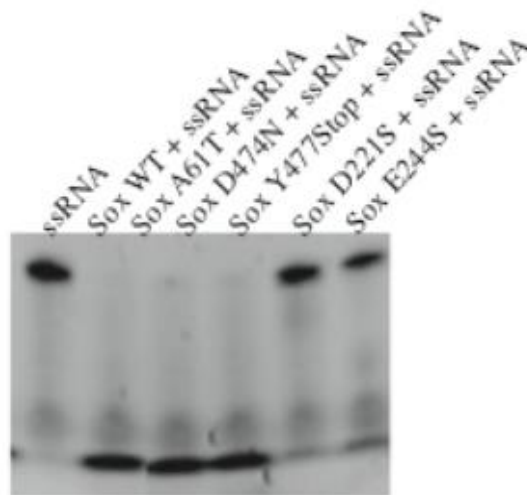


Figure 1.7: TBE urea gel demonstrating that HSO mutants (SOXA61T, D474N and Y477Stop) had near wild-type activity against the single-stranded RNA substrate. D221S and E244S were used as negative controls. (Figures adapted and modified from Bagneris et al., 2011).

Similarly, variable lengths of cleavage product were also identified with a GAPDG (glyceraldehyde 3- phosphate dehydrogenase) reporter gene. These effects were similar to those observed in nonsense mediated decay (NMD) which degrades

mRNAs that terminate translation more than 50-55 nucleotides upstream of a splicing-generated exon-exon junction preventing the production of potentially toxic truncated proteins that they encode (Maquat, 2004; Nagy & Maquat, 1998). These studies appeared to indicate that positional cues around the translation initiation site do not regulate the accumulation of mRNA intermediates but instead suggested that SOX-mediated mRNA degradation required binding to a specific sequence or location in the target host mRNA (Covarrubias et al., 2011). They further suggested that degradation occurred via endo but not exonucleolytic cleavage. Interestingly, it was also observed that the activities of SOX were insufficient for the complete degradation of mRNA transcripts. Xrn1, a host scavenger exoribonuclease that functions in mRNA decay, was subsequently identified as having a key role in the turnover of the SOX generated mRNA intermediates. These studies thus indicate that the endonucleolytic activity of SOX introduces a “nick” into the target transcript followed by degradation via the exonucleolytic activity of Xrn1 (Covarrubias et al., 2011).

To gain further insights into the requirements for mRNA cleavage, sequence alignments were performed by Glaunsinger and co-workers on the GFP, DsRed and HBB reporter gene degradation products which identified a UGAAG motif 2 or 3 nucleotides upstream of the cleavage sites in all cases (Covarrubias et al., 2011; Kronstad & Glaunsinger, 2012). When a 201 nucleotide (nt) fragment of GFP-mRNA possessing the UGAAG motif was transfected into cells, nucleolytic cleavage was observed. However, when a 25nt fragment possessing the UGAAG motif was introduced, degradation was inhibited suggesting that RNA tertiary structure might

also have an important role in cleavage (Covarrubias et al., 2011; Clyde & Glaunsinger, 2011). A more detailed, transcriptome-wide analysis of targeted transcripts and their cleavage patterns has shown that SOX mediated RNA degradation favours substrates with two features: first, it likely recognises six nucleotides containing two or three adenine nucleotides located 5' to the cleavage site and second, binding is favoured by the presence of particular structures such as a stem loop or bulge motifs in the vicinity of the cleavage sites (Gaglia et al., 2015). These studies lead to a more detailed computational analysis in the Barrett group by a former PhD student (Anathe Patschull), who sequentially folded shorter lengths of the 201nt GFP mRNA transcript cited in the Glaunsinger studies using the RNA folding software MC-Fold and MC-Sym to establish whether structural motifs centred around the UGAAG sequence could be identified. From these studies, it was possible to ascertain that despite the absence of highly conserved motifs, cleavage appeared to preferentially occur in the vicinity of loop regions (within stem loops) or areas corresponding to unpaired bases (Figure 1.8) (Lee et al., 2017).

GFP51 (117-168)

DsRed61 (489-550)

HBB58 (179-237)

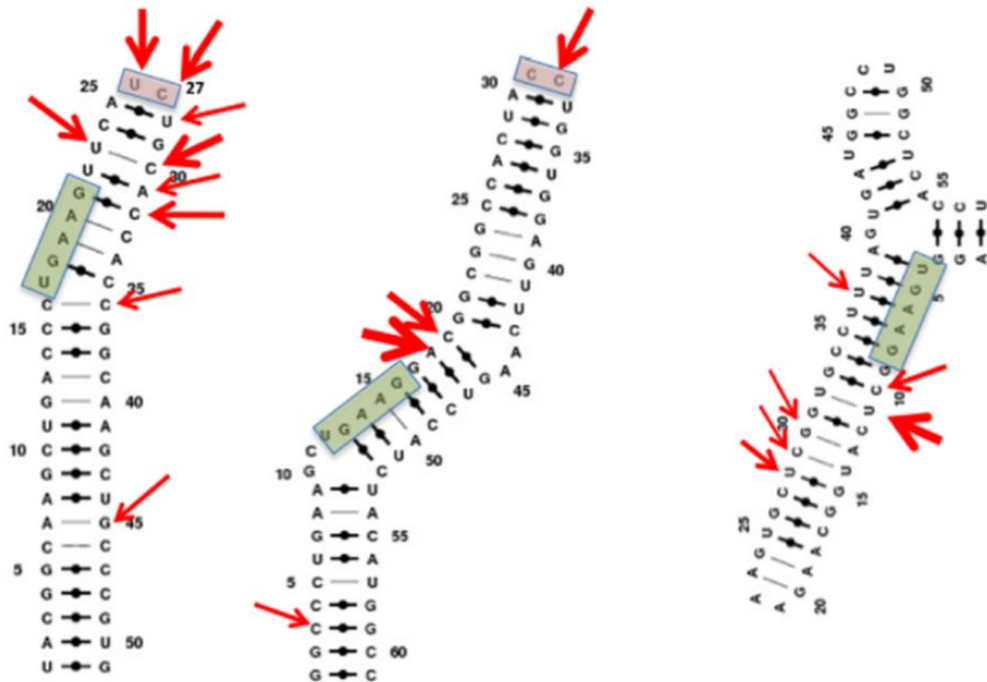


Figure 1.8: The predicted lowest energy structures of GFP51, DsRed61 and HBB58 generated by the pipeline MC-Fold | MC-Sym. Red arrows indicate the cleavage sites of SOX. Green boxes highlight the conserved sequence UGAAG motif. (Figure adapted and modified from Lee et al., 2017).

1.8 KSHV-associated antiviral therapy

The standard treatments for KS include chemotherapy, surgery, and radiation therapy with 22 to 80 % of HIV-positive patients initially responding positively. However, complete remission of KS is rarely achieved (Union for International Cancer Control EML Review, 2014). In 2002, the natural product rapamycin emerged as a novel therapeutic drug for treating patients with transplantation-associated KS (Guba et al., 2002). Rapamycin was shown to suppress KS lesions in KSHV-infected

patients (Stallone et al., 2005). It also showed anti-angiogenic effects in a murine tumour model (Guba et al., 2002) and directly affected mammalian target of rapamycin (mTOR), which is crucial for KSHV pathogenesis (Chang & Ganem, 2013; Roy et al., 2013). Nonetheless, Rapamycin also acts as an immunosuppressive agent; therefore, it was not selected as a standard therapeutic option for KS. The highly active antiretroviral therapy (HAART) was introduced in 1996 for HIV-associated KS treatment in Europe and the US (Mocroft et al., 2004), which significantly reduced the prevalence of AIDS-associated malignancies and deaths (Cohen et al., 2005). Although HAART was able to impair the activity of HIV reverse transcriptase to inhibit replication, it did not have any effect on the replication cycle of KSHV, thereby limiting the success of KS treatment. Use of other antiviral drugs such as Phosphonoformic acid (PFA), Ganciclovir, and Valganciclovir, which mimic both pyrophosphates and nucleosides, are able to target KSHV DNA polymerase activity. Unfortunately, these drugs, are no longer effective owing to drug resistance (Cohen et al., 2005).

Only a small proportion of KSHV-infected cells undergo lytic reactivation, while most remain latently infected (Dittmer, 2003; Hosseinipour et al., 2014). Recently, several antiviral treatment options were suggested that target lytic reactivation of KSHV (Coen et al., 2014). However, since both the latent and lytic phases of the KSHV life cycle are involved in tumourigenesis, simply impairing the lytic viral replication will not clear KSHV infection from the infected host.

Until now, licensed vaccines that target KSHV-induced malignancies are not available due to the lack of scalable cell culture systems, the strict specificity of KSHV for humans, limited development of *in vivo* models, and incomplete identification and screening of antiviral compounds (Dubich et al., 2019). These limitations highlight the need for novel antiviral therapy and antiviral drugs for treating KSHV.

1.9 Structure-based design of KSHV-SOX inhibitors

The DNase activity of SOX was shown to be essential for the processing and encapsidation of the viral genome (Martinez et al., 1996; Severini et al., 1994; Severini et al., 1996; Mettenleiter, 2002). Furthermore, the RNase activity of SOX was shown to be essential for degrading host mRNA (Lee et al., 2017). Together, these results indicate that SOX could be targeted for designing antiviral agents to block KSHV virion production and hence the production of progeny. Additionally, SOX inhibitors may also impact on downstream latency and viral trafficking given the link between the activities of SOX and these important processes.

1.9.1 Cancer preventive and therapeutic properties of Phytic acid (Inositol Hexaphosphate, IP₆)

Interestingly, a number of oncology studies have shown that phytic acid plays an important role in cancer prevention, and inhibition of tumour growth and progression (Vucenik & Stains, 2010; Vucenik & Shamsuddin, 2003). For example,

phytic acid was shown to suppress cell proliferation and promote apoptosis and differentiation of malignant cells via MAPK, PI3K, NFκB, AP-1 and PK (Vucenik & Stains, 2010). Moreover, several anticancer effect of phytic acid was demonstrated in different *in vitro* experimental models (Table 1.1).

Table 1.1: Antitumour effects of IP6 *in vitro*. (Table adapted and modified from Vucenik & Shamsuddin, 2003).

Organ or tissue	Species	Cell line
Blood	Human	Erythroleukemia K562 cell line
Colon	Human	Adenocarcinoma HT-29 cell line
Lung	Rat	Tracheal epithelium
Liver	Human	HepG2 cells
Mammary	Human	Adenocarcinoma
Uterine cervix	Human	HeLa cells
Prostate	Human	Adenocarcinoma PC-3 cell line
Skin	Mouse	JB6 cells
Soft tissue	Human	Rhabdomyosarcoma

Phytic acid (also known as inositol hexakisphosphate (IP₆)), is composed of six phosphate groups bound to each carbon to form a ringed structure (Figure 1.9A). Phytic acid is ubiquitous and is particularly abundant in plants, serving as a phosphate reserve in the seeds and being the foremost inositol phosphate discovered (Joy & Balaji, 2015). Phytic acid is also found in most mammalian cells and its lower phosphorylated forms (IP₁₋₅) play key roles in regulating vital cellular functions (Shamsuddin et al., 1989; Shamsuddin, 2002).

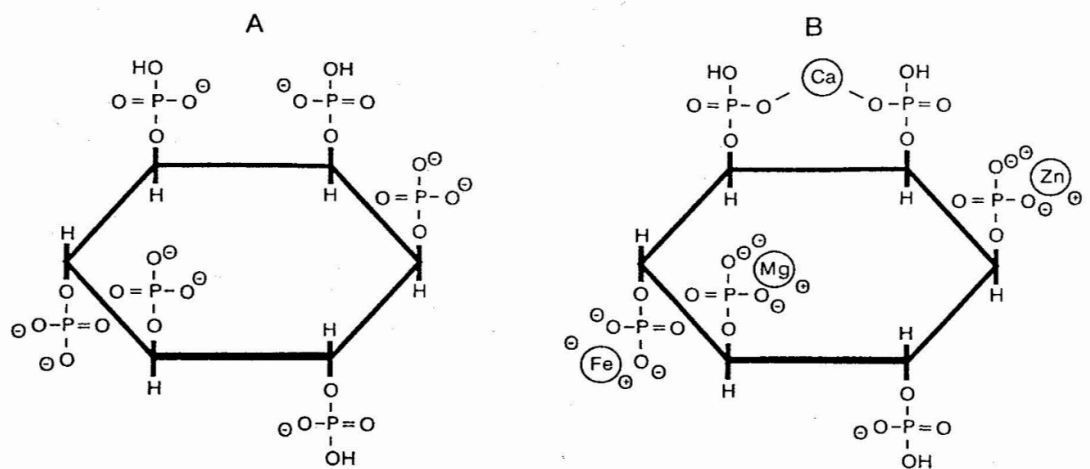


Figure 1.9: (A) Structure of phytic acid (IP₆); (B) Structure of phytic acid chelated to divalent cations.

Phytic acid can undergo dephosphorylation and phosphorylation, whose products play different roles in mammalian cells: Lower forms of inositol phosphates (IP₁₋₄) act as intracellular messengers; for example, inositol 1,4,5-trisphosphate (I_(1,4,5)P₃) and inositol 1,3,4,5-tetrakisphosphate (IP₄) are recognised as second intracellular messengers that induce calcium sequestration involved in mitosis. Alternatively, the higher inositol phosphates, inositol 1,3,4,5,6-pentakisphosphate (IP₅) and inositol hexakisphosphate (IP₆), are found in higher concentrations with more recent studies showing that IP₅ and IP₆ have physiological functions in the regulation of the affinity of avian haemoglobin for oxygen and neuronal excitation, respectively (Menniti et al., 1993).

It has also been shown that phytic acid undergoes dephosphorylation to IP₁₋₅ and IP₃ which is involved in both intracellular functions and cellular signal transduction (Shamsuddin, 1992). It is well known that divalent cations such as magnesium ions (Mg²⁺) function as important second messengers within the cell, and zinc (Zn²⁺) is essential for cell division as well as DNA synthesis (Shamsuddin, 2002). Interestingly, previous experiments have shown that phytic acid induces anticancer effects by removing these cations (Jariwalla et al., 1988; Thompson & Zhang, 1991). This is based on the ability of the six phosphate groups to chelate divalent cations (Figure 1.9B).

It has been demonstrated that phytic acid inhibited the cytopathic effect of the human immunodeficiency virus (HIV) and HIV-specific antigen expression in HTLV-I-transformed T-cell line 4 (MT-4) cells, indicating that phytic acid may play an important role in preventing and inhibiting HIV infection and associated HIV-related disorders (Otake et al., 1989). Recently, *in silico* experiments have demonstrated the interactions of phytic acid with various drug targets and uncovered possible analogues using bioinformatics to prove their drug-likeness (Joy & Balaji, 2015). It has been also shown that phytic acid inhibited the colony formation of KS cell lines such as KS Y-1 (AIDS-related KS), and KS SLK (Iatrogenic KS) in a dose-dependent manner (Tran et al., 2003).

Since phytic acid has almost no toxic effects to the normal human cells and is naturally present in almost all plant and mammalian cells, phytic acid has received much attention for its role in anticancer therapeutics (Vucenik & Stains, 2010).

Although there are many studies relating phytic acid as an anti-cancer therapy, due to the chelate formation and rapid elimination inside the body, it is not popular in chemotherapeutic treatment (Arya et al., 2019).

1.10 Viral latency and the NF- κ B pathway

The expression of hundreds of genes primarily involved in immune responses are regulated by the NF- κ B transcription factors. Signalling events leading to NF- κ B activation constitute a major antiviral immune pathway. Viruses target these mechanisms to replicate and persist within their hosts and as a second part of this project, the remainder of this chapter focuses on KSHV-vFLIP, a protein produced during latency that has evolved diverse strategies to evade and exploit cellular NF- κ B immune signalling cascades to benefit the virus.

1.10.1 NF- κ B

The nuclear factor kappa light-chain enhancers of activated B cells (NF- κ B) are transcription factors that exist in virtually all cell types and play key roles in regulating several biological processes such as apoptosis, cell development, responses to cellular stresses, inflammation and angiogenesis (Pahl, 1999; Tolani et al., 2014). The activation of NF- κ B is a rapid process, as the transcription factors do not require co-activators and are upregulated by diverse stimuli which include growth factors, cyclins, chemokines, cytokines and adhesion molecules (Pahl, 1999).

Activation of NF- κ B pathways promotes the expression of over 150 target genes that are important for processes associated with cell proliferation and survival, and of genes that both suppress and promote apoptosis and have a role in immune responses (Hayden & Ghosh, 2004). Studies have shown that factors that are associated with the maturation of multiple cell types and embryonic development also lead to the activation of NF- κ B pathways (Hayden & Ghosh, 2004; Oeckinghaus & Ghosh, 2009). Hence deregulation of NF- κ B can lead to the development of cancer, viral infection, septic shock, inflammation and autoimmune deficiencies.

The following sections will focus on the components involved in NF- κ B signalling pathways and two major NF- κ B signal transduction pathways: the canonical and non-canonical (or alternative) NF- κ B pathways.

1.10.2 NF- κ B related proteins

In general, NF- κ B signalling pathways involve three major protein families: NF- κ B transcription factors, the inhibitors of κ B (I κ B) family and the I κ B Kinase (IKK) complex (Zhang et al., 2017; Sun, 2017; Hayden & Ghosh, 2008). In resting cells, the I κ B proteins sequester the NF- κ B transcription factors in the cytoplasm. In order for the canonical pathway to be activated, these I κ B proteins need to be disrupted by the IKK kinase complex (a multi-subunit assembly) leading to post-translational modifications of the I κ Bs leading to their degradation. By contrast, activation of the non-canonical pathway is IKK kinase independent and instead is dependent on an upstream kinase NIK (NF- κ B inducible kinase).

1.10.3 NF- κ B family

There are a total of five NF- κ B transcription factors that constitute the mammalian NF- κ B family (Figure 1.10A). These can be divided into two classes: Class 1 includes NF- κ B1 (p105/p50) encoded by the NFKB1 gene and NF- κ B2 (p100/p52) encoded by the NFKB2 gene; Class 2 includes RelA (p65) encoded by the gene RELA, RelB encoded by the gene RELB, and c-Rel, encoded by the gene REL (Gilmore, 2006).

Different members of NF- κ B proteins can associate to form 15 homo- or heterodimers that are either transcriptionally active or thought to have a regulatory role. Each member contains a 300 amino acid N-terminal Rel-Homology Domain (RHD), which is responsible for dimerisation, binding to DNA and I κ B (Ghosh et al., 1998). The Class II family of NF- κ B possesses a non-homologous transactivation domain (TAD, activated by one or more receptors) at their C-termini, which is responsible for co-activator and repressor recruitment (Gilmore, 2006). By contrast, the Class I family of NF- κ B proteins are synthesised as large precursors, p105 and p100, in which the inhibitor regions of p105 and p100 at the C-termini are proteolytically cleaved to generate the mature NF- κ B subunits, p50 and p52, respectively. Unlike the Rel proteins, an ankyrin repeat domain (ARD) is found at the C-terminal halves in the p100 and p105 precursors (Hoffmann et al., 2006). Compared to p100, whose processing is inducible and is mediated by upregulating signals, the processing of p105 is constitutive (Karin & Ben-Neriah, 2000; Liou et al., 1994).

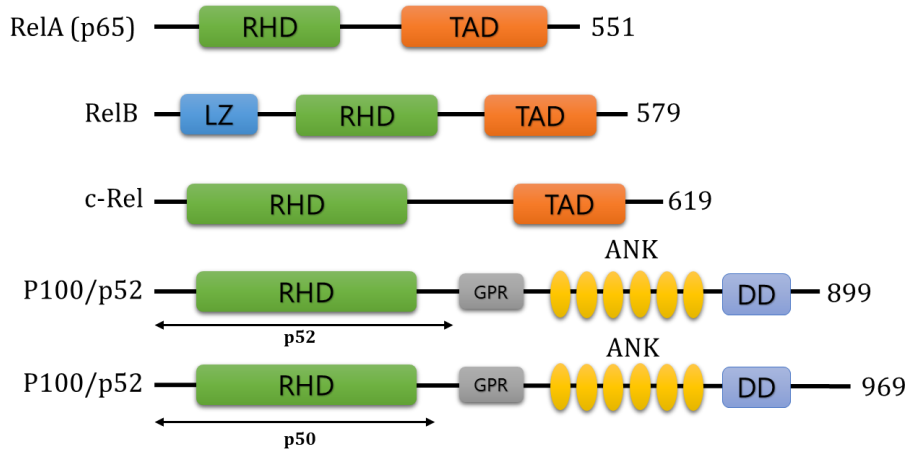
Although the TADs are absent in P50 and p52 proteins, they are thought to act as transcriptional repressors as they are still able to bind to NF- κ B consensus sites in

DNA (Solan et al., 2002). Although transcriptionally inactive, they can both form hetero-dimers with cofactor-recruiting proteins such as BCL3 or one of the TAD containing NF- κ B subunits (Zhong et al., 2002; Hayden & Ghosh, 2008). RelB is unique in that it contains a leucine zipper-like (LZ) motif at N-terminus, which also plays a key role in the activation of NF- κ B.

Although almost all combinations of homo- and heterodimers of NF- κ B have been identified *in vivo*, the RelA/p50 heterodimer is most commonly found in the NF- κ B complex, in contrast to RelB, which has been shown to only form a heterodimer with p50 and p52 (Gosh & Karin, 2002). 12 out of the 15 possible dimers have been identified in various tissues, and the composition of NF- κ B dimers depends on the type of cell, duration and signalling stimulus (Sen & Smale, 2009).

In most cells, the NF- κ B inhibitor proteins (I κ Bs) remain bound to the NF- κ B dimers, sequestering them in the cytoplasm. I κ Bs are polypeptides which bear Ankyrin repeat domains (ARDs) that are required for binding to NF- κ B dimers. I κ B α and I κ B β harbours sequences rich in PEST (proline (P), glutamate (E), serine (S), and threonine (T) amino acids) domains at their C-termini which are required for protein turnover (Figure 1.10B) (Zheng et al., 2011).

A. NF- κ B family



B. I κ B family

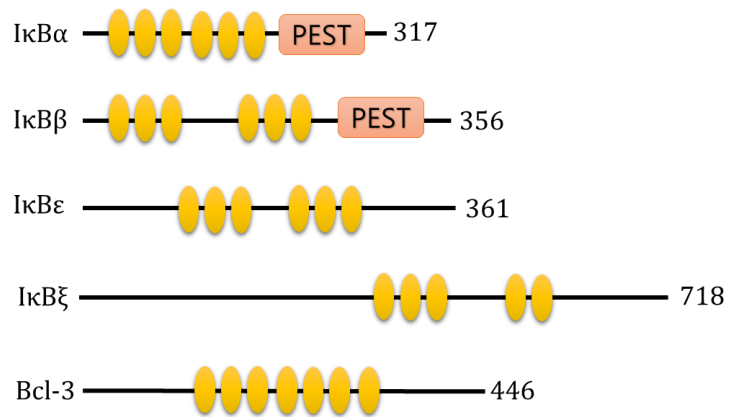


Figure 1.10: Overview of NF- κ B related proteins. The total number of amino acids in each protein is indicated on the right. (A) NF- κ B family proteins. (B) The I κ B family proteins.

1.10.4 I κ B Kinase (IKK) complex

For activation of the canonical pathway, the NF- κ B-I κ B complexes have to be disrupted and this process is initiated by the IKK kinase complex.

The I κ B Kinase (IKK) is a highly regulated protein complex responsible for the phosphorylation of serine residues at the N-termini of the I κ Bs. This then leads to their rapid polyubiquitination and proteasomal degradation (Figure 1.11). This phosphorylation step plays a key role in activating the NF- κ B pathways (Oeckinghaus & Ghosh 2009; Hayden & Ghosh, 2012).

The IKK complex is a large assembly (~700-900 kDa) composed of two enzymatic subunits: IKK α (IKK1) and/or IKK β (IKK2) and a regulatory subunit known as IKK γ or NF- κ B essential modulator (NEMO) (Scheidereit, 2006) (Figure 1.11). Due to the large discrepancy between the molecular weights of the complex observed during gel filtration of the IKK complex, the exact number of subunit components and the stoichiometry have been a matter of debate. However, it is generally accepted that the IKK core complex is constituted of a dimer of dimers, IKK α ₁IKK β ₁IKK γ ₂, in which 1 IKK α associates with a dimer of NEMO, hetero dimerised with 1 IKK β . It was also suggested that other combinations of IKK core complex can exist which include NEMO with or without homodimers of IKK α and IKK β (Zhang et al., 2017; Liu et al., 2012).

IKK proteins

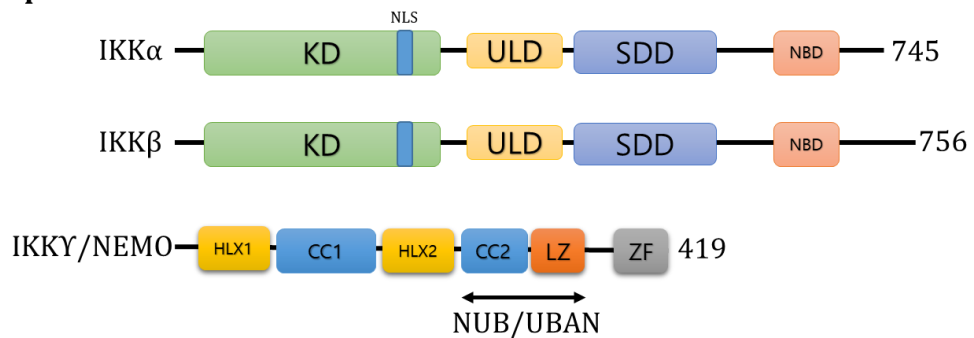


Figure 1.11: IKK α , IKK β and NEMO form the core of the IKK complex. Serine residues in the IKK α / β kinase domain (KD) that are phosphorylated are highlighted. NBD: NEMO-binding domain, SDD: scaffold/dimerisation domain; HLX: helical domain; ULD: ubiquitin-like domain; CC: coiled coil; LZ: leucine zipper; HLH: helix-loop-helix; NUB/UBAN: NEMO ubiquitin binding domain, also called UBAN.

1.10.5 The catalytic subunits: IKK α and IKK β

IKK α (Figure 1.12A), also known as CHUK (conserved helix-loop-helix ubiquitous kinase), was the first component of the complex to be identified with a molecular weight of 85 kDa and is a serine-threonine kinase (Polley et al., 2016). Similarly, IKK β is an 87 kDa serine-threonine kinase which is 51 % identical to IKK α . Both IKK α and IKK β contain N-terminal kinase domains (KD) capable of phosphorylating multiple members of the I κ B family (DiDonato et al., 1997; Polley et al., 2016) (Figure 1.12B). In order to perform their kinase activities, the two serine residues within the activation loop of the KD need to be phosphorylated (S176 and S180 for IKK α , S177 and S181 for IKK β) (Liu et al., 2012).

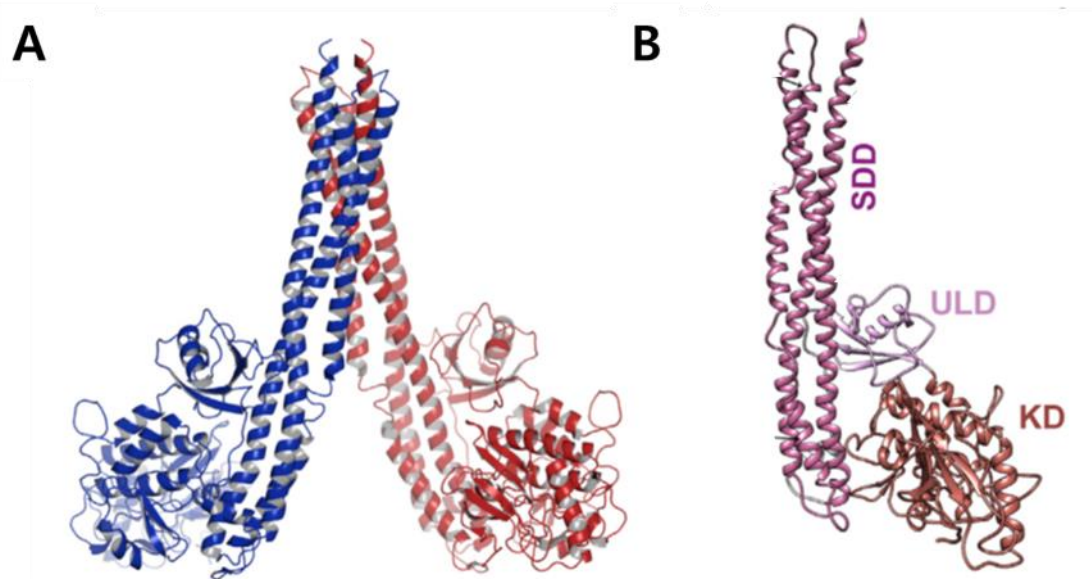


Figure 1.12: (A) Crystal structure of dimeric IKK α . (B) The IKK α monomer with three domains: SDD; ULD; and KD. (Figure adapted and modified from Polley et al., 2016)

IKK α contains a putative nuclear localisation signal (NLS) associated with NF- κ B-independent nuclear activities (Sil et al., 2004). The recently solved crystal structure of IKK β revealed that IKK β is comprised of an N-terminal KD, an ubiquitin-like domain (ULD) at the centre and a scaffold/dimerization domain (SDD) at the C-terminal end. In addition, it also contains a leucine zipper (LZ) domain and a helix-loop-helix (HLH) motif as a part of the SDD (Xu et al., 2011) (Figure 1.13). The SDD plays an important role in kinase activity by mediating dimerization of IKK β , but when the activation loop is phosphorylated, the SDD is no longer important for maintaining the kinase activity of IKK β . The ULD, together with the SDD, determines the substrate specificity for I κ B α , thereby enhancing the catalytic activity of IKK β

(Xu et al., 2011). IKK γ binds to the IKK α IKK β complex via a C-terminal hexapeptide sequence (Leu-Asp-Trp-Ser-Trp-Leu) present in the nemo binding domain (NBD) (May et al., 2002).

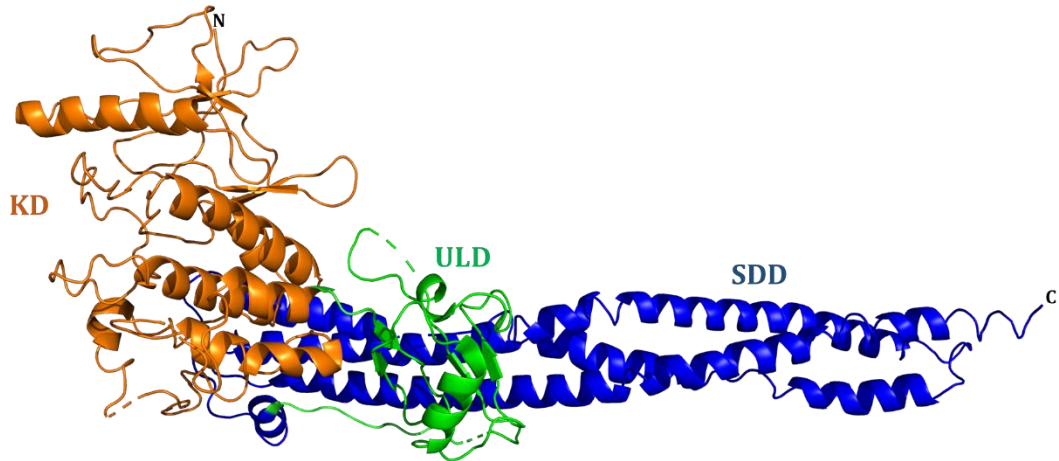


Figure 1.13: Crystal structure of IKK β comprised of an N-terminal KD, ULD at the centre and SDD at the C-terminal end. (Figure adapted and modified from Xu et al., 2011).

1.10.6 The regulatory protein NEMO/IKK γ

IKK γ , (or NEMO), is a non-catalytic scaffold/regulatory protein of 49 kDa containing two helical domains (HLX) at the N-terminus flanked by coiled-coil (CC) domains, and zinc finger-like (ZF) domains at the C-terminus together with a leucine zipper motif (Zheng et al., 2011). The C-termini of IKK α and IKK β interact with the N-terminal region of NEMO (May, 2000). The LZ with adjacent CC2 domain, forms the ubiquitin-binding domain of NEMO known as the NEMO ubiquitin binding (NUB) (Figure 1.14) (Bloor et al., 2008). Previous studies have suggested that the ZF motif

of NEMO is also required for regulating the substrate-specificity of IKK β towards I κ B α (Schrofelbauer et al., 2012).

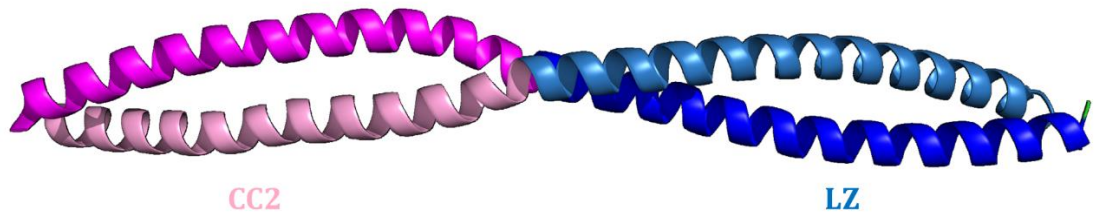


Figure 1.14: Overall dimeric structure of CC2-LZ domains of IKK γ , (or NEMO). (Figure adapted and modified from Lo et al., 2009).

1.10.7 Other IKK-associated components

It was suggested that apart from IKK α , IKK β and IKK γ , other factors are involved in formation of the active IKK complex. They include the chaperone heat shock protein 90 (HSP90), HSP70 and CDC37. Although their exact role is still unclear, it was suggested that HSP90 in association with CDC37 stabilises the complex through enhancing folding of IKK α /IKK β , where HSP70 appears to have a regulatory role through preventing the interaction between NEMO and IKK α /IKK β (Ran et al., 2004).

1.11 Activation of the Canonical NF- κ B pathway

Robust signalling in the canonical NF- κ B pathway is mainly achieved through the activation of the IKK β subunit of the I κ B kinase (IKK) complex, which phosphorylates two serine residues, Ser32 and Ser36 of I κ B α , at the N-terminal end of the I κ B proteins in a NEMO-dependent manner (Hayden & Ghosh 2008, 2012). NEMO is a non-catalytic subunit that binds IKK α and IKK β into a regulatory holocomplex and is required association with ubiquitin chains for activation of the kinase (Hayden & Ghosh, 2008). The phosphorylated I κ B proteins are recognised and subjected to ubiquitination by the ubiquitin ligase family members involving Skp1-Culin-Roc1/Rbx1/Hrt-1-F-box (SCF/SCRF) (Ben-Neriah, 2002). In this process, the phosphorylated E3 recognition sequence (DS*GXXS*) found on I κ B proteins are recognised and targeted by the β -transducing repeats-containing proteins (β -TrCP, a subunit of the SKP1-cullin-F box (SCF) ubiquitin ligase complex) of the SCF family ubiquitin ligase complex (Wu & Ghosh, 1999; Hayden & Ghosh 2008, 2012). In addition, two conserved lysine acceptor sites on I κ B (Lys21 and Ly22) at the N-terminus are in turn subjected to poly-ubiquitination by an E3 ubiquitin ligase enzyme. Following ubiquitination, the 26S proteasome fully degrades the I κ Bs. Once I κ B is degraded, NF- κ B is released and translocated to the nucleus. This is then recruited to its target promoter where it binds to the enhancer regions containing the κ B consensus binding motif and initiates gene transcription (Figure 1.15).

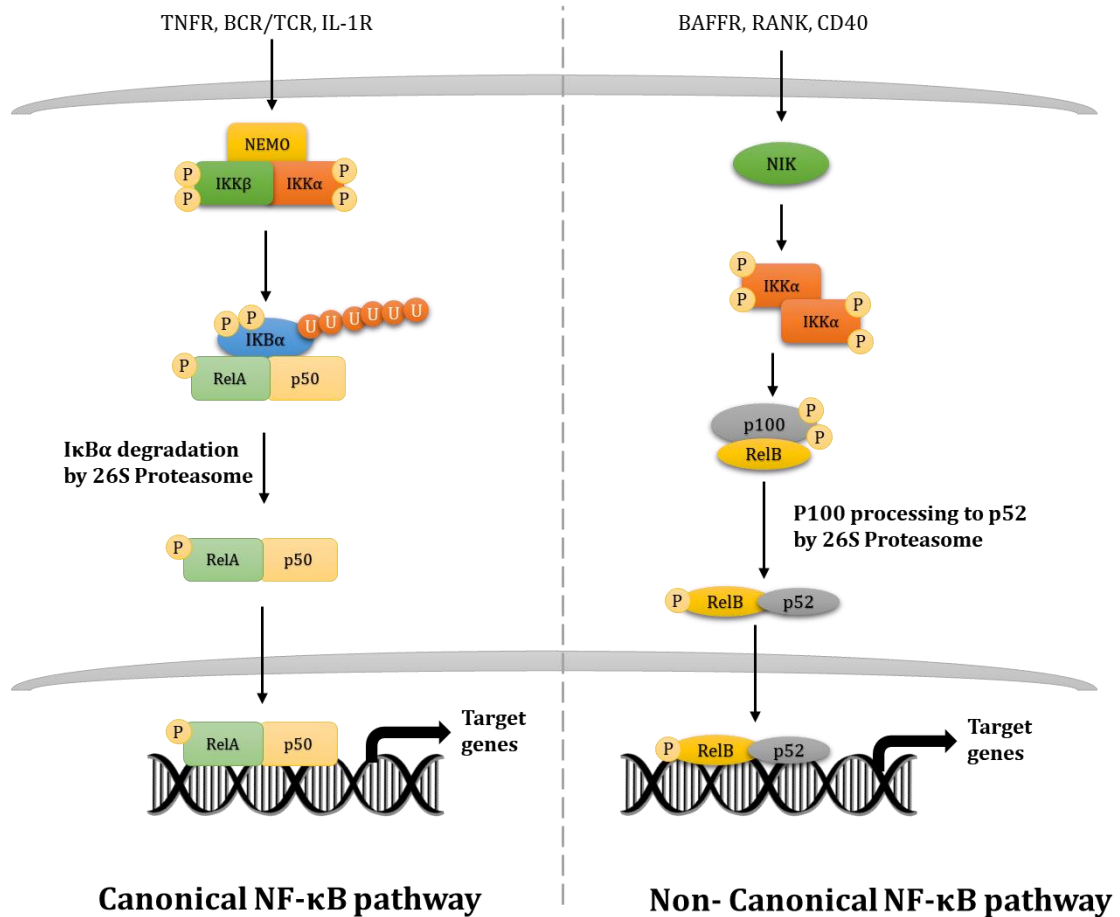


Figure 1.15: Overview of the canonical and non-canonical NF-κB pathways. The canonical NF-κB pathway involves the activation of the IKKβ which phosphorylates the IκB thereby recruited to its target promoter and initiates gene transcription. The non-canonical NF-κB pathway involves activation of IKKα which phosphorylates the p100 that processed to p52 which then initiates the transcriptional activation.

1.12 The Non-canonical (alternative) NF- κ B pathway

The non-canonical or alternative NF- κ B pathway is activated by a limited number of receptors (CD27, CD40, Fn14, B-cell activating factor receptor, lymphotoxin β receptor) that belong to the Tumour necrosis factor (TNF) receptor superfamily (Figure 1.15) (Razani et al., 2011). Although the signals that activate the non-canonical NF- κ B pathway are able to activate the canonical pathway, not all signals activate non-canonical pathway (Bonizzi & Karin, 2004).

Unlike the canonical NF- κ B pathway, the non-canonical pathway appears to be independent of NEMO and IKK β and relies solely upon the activity of IKK α (Figure 1.15) (Senftleben et al., 2001). The key activation protein of this pathway is NF- κ B-inducing kinase (NIK) which is a member of the mitogen-associated protein 3 kinase (MAP3K) family (Xiao et al., 2001, 2004). It is thought that NIK activates IKK α by phosphorylating the two serine residues Ser176 and Ser180 at the activation loop. It has been shown that the trimeric surface of IKK α is essential for interaction with NIK (Figure 1.16). These surfaces form an IKK α hexamer that, is required for the processing of p100 to p52 (Polley et al., 2016) although not detectable in cells. It has been shown that IKK α is recruited to p100 by NIK in a poorly understood process where it initiates phosphorylation of the C-terminal serines 866 and 870 although there is uncertainty in the exact sites owing to discrepancies between experiments conducted *in vitro* and *in vivo* (Polley et al., 2016). The phosphorylation of p100 initiates subsequent K-48 linked phosphorylation by the ω -TrCp ubiquitin ligase leading to targeted degradation by the 26S proteasome in which only the C-terminus

is processed. Although it remains unclear as to the exact manner in which processing of p100 and formation of the p52:RelB heterodimer takes place, it has been suggested that RelB controls both p100 processing and the formation of p52:RelB heterodimer that in addition to p100, sequester other NF- κ B transcription factors to inhibit gene expression (Fusco et al., 2016). Also unclear is how the balance between the transcriptional and these inhibitory activities of p100 are maintained. It is assumed that newly synthesised p100 molecules are required for p100 processing (Yilmaz et al., 2014).

The alternative NF- κ B pathway controls genes that are mainly involved in regulating adaptive immunity, B-cell maturation, dendritic cells (DC) maturation, osteoclastogenesis and T-cell differentiation (Zhang et al., 2017; Sun, 2017).

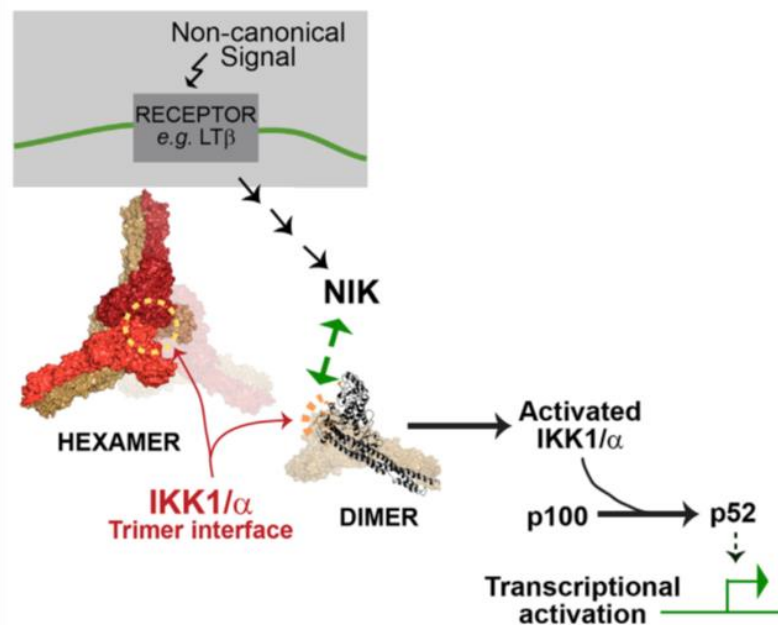


Figure 1.16: Schematic diagram showing the trimeric IKK α surfaces required for interaction with NIK. (Figure adapted and modified from Polley et al., 2016).

1.13 P100 processing and the p100 death domain (DD)

The processing of p100 into p52 represents another important regulatory mechanism in the activation of NF- κ B pathways; however, the exact mechanism in which p100 processing occurs is still unclear. The presence of the C-terminal ARD in p100 indicates that it functions similarly to I κ B-like inhibitors but other factors have also been shown to be required (Tao et al., 2014). The p100 DD plays a key role in the phosphorylation-induced recruitment of a β -TrCP. It was shown that β -TrCP interacts with a phosphorylated sequence in p100 termed the PEST region (sequence enriched in proline (P), glutamate (E), serine (S), and threonine (T)) (Senftleben et al., 2001). Moreover, p100 ubiquitination is regulated by the β -TrCP binding site created by NIK/IKK α -dependent phosphorylation of p100 at serines 866, 870 and possibly 872 (Liang et al., 2006). The N-terminal Lys855 in p100 also serves as a β -TrCP binding site (Amir et al., 2004). Notably, p100 mutants were observed in B and T cell lymphomas, which are caused by truncations at DD resulting from chromosomal translocations (Migliazza et al., 1994). These mutations led to the reduction of p100 and resulted in the constitutive production of p52. These studies also showed that constitutive p52 production led to an increase in the level of p52:RelB translocation into the nucleus, resulting in the suppression of apoptosis and aberrant growth regulation in lymphocytes, thereby promoting lymphomagenesis (Figure 1.17) (Karin et al., 2002).

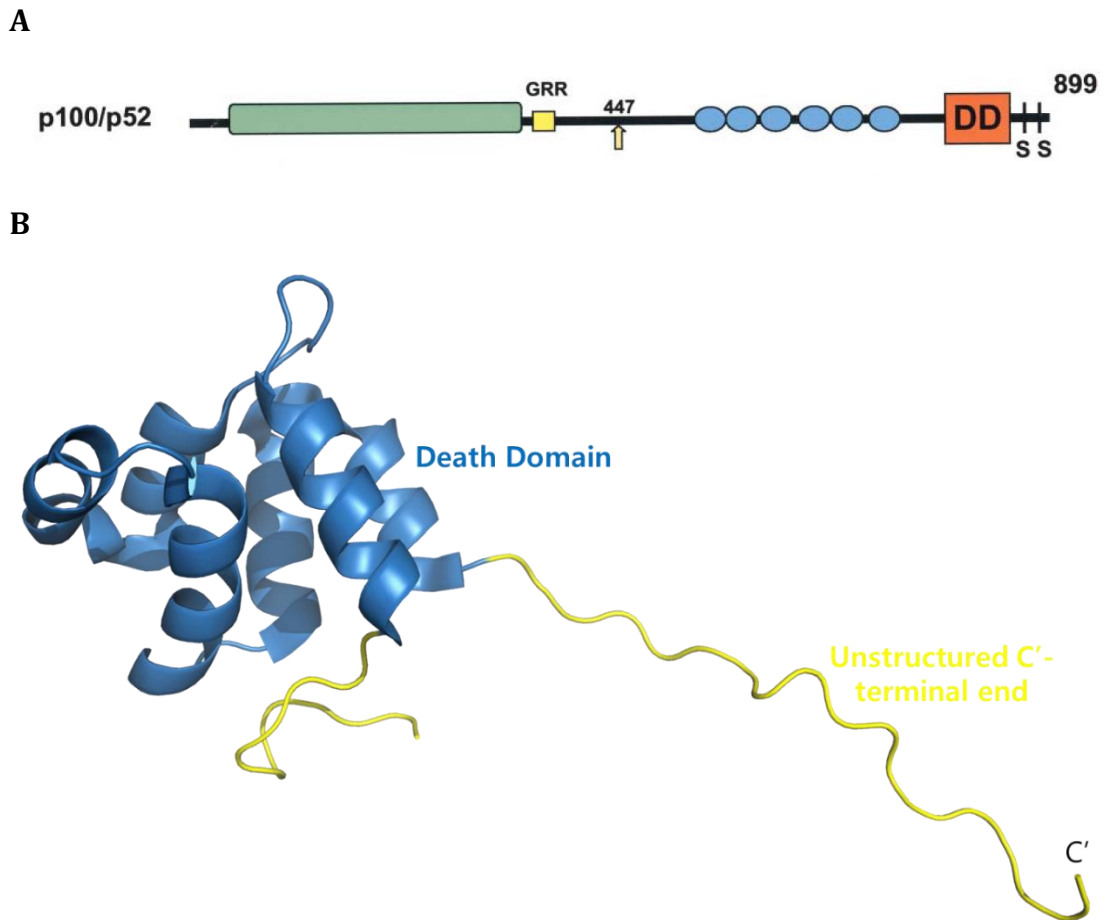


Figure 1.17: Schematic diagram of p100. (A) The precursor p100 contains an N-terminal RHD that is important for DNA binding and dimerization. It also contains a C-terminal region that contains Ankyrin repeats which mask its nuclear localisation motif together with the C-terminal DD. (B) Schematic diagram of the p100 DD.

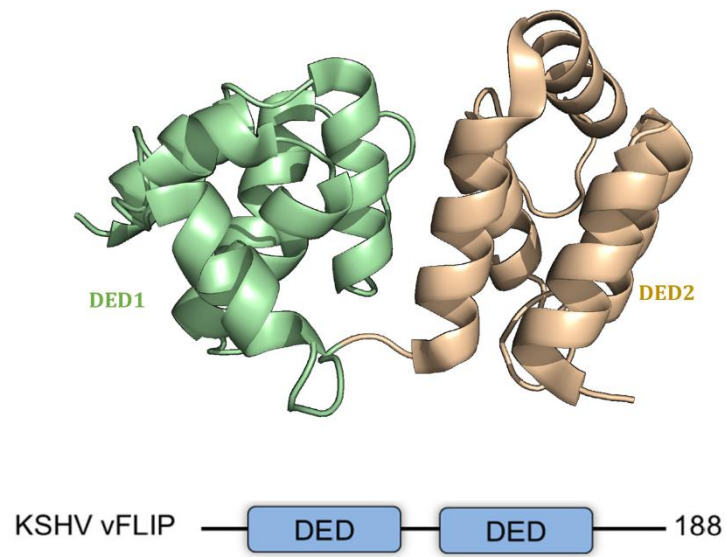
1.14 KSHV- encoded viral fllice inhibitory protein (vFLIP)

In order to evade the host immune response, several viruses induce expression of inhibitory proteins that suppress apoptosis and promote replication and survival in host cells (Thome & Tschopp, 2001). Amongst important apoptosis inhibitory proteins is the KSHV viral FLICE-inhibitory protein (vFLIPs). KSHV-vFLIP is the gene

product of open reading frame K13 (ORF71) and resembles the caspase 8/FLICE in comprising two tandem death effector domains (DEDs) (Figure 1.18A). It is a viral analogue of the cellular FLIP homologues with which it shares ~33 % sequence identity (Field et al., 2003).

Unlike the host FLIPs, studies have shown that KSHV-vFLIP possesses the unique ability to activate the canonical NF- κ B pathway by interacting with IKK γ to modulate the expression of a number of genes involved in cell proliferation, transformation, cytokine secretion, and protection against growth factor withdrawal-induced apoptosis (Field et al., 2003). Initially, it was suggested that KSHV-vFLIP facilitates IKK activation by interacting with TRAF2 via a putative TRAF-binding motif. However, later reports proved that independent of TRAF2 or TRAF3, KSHV-vFLIP is capable of binding to NEMO leading to constitutive IKK complex activation (Matta et al., 2007; Field et al., 2003). The crystal structure of KSHV-vFLIP bound to IKK γ has been determined which confirmed that KSHV-vFLIP binds to the HLX2 motif in IKK γ in some way inducing or stabilising an active conformation. In this configuration IKK α /IKK β are persistently phosphorylated resulting in constitutive activation of the canonical NF- κ B pathway (Figure 1.18B) (Bagneris et al., 2008).

A



B

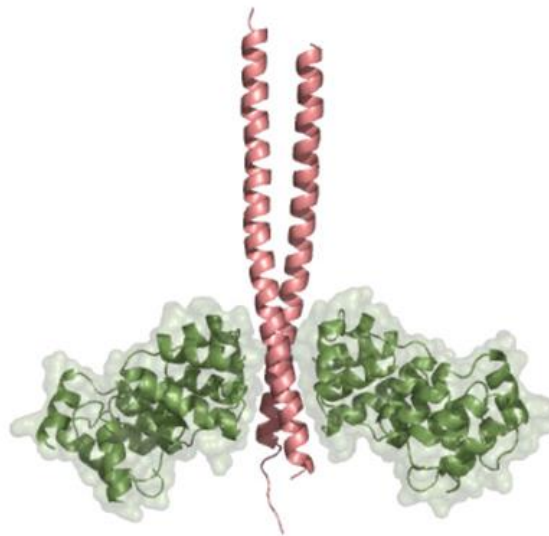


Figure 1.18: (A) Crystal structure of KSHV vFLIP containing two tandem death effector domains (DEDs). (B) Crystal structure of the vFLIP-IKK γ complex (KSHV-vFLIP highlighted in green and IKK γ highlighted in pink) (Figure adapted and modified from Bagn ris et al., 2008).

Previous experiments, however, have also suggested that KSHV-vFLIP can activate the non-canonical NF- κ B pathway by forming a ternary complex with the transcription factor p100 and IKK α , especially the C-terminal end of p100 (Figure 1.19).

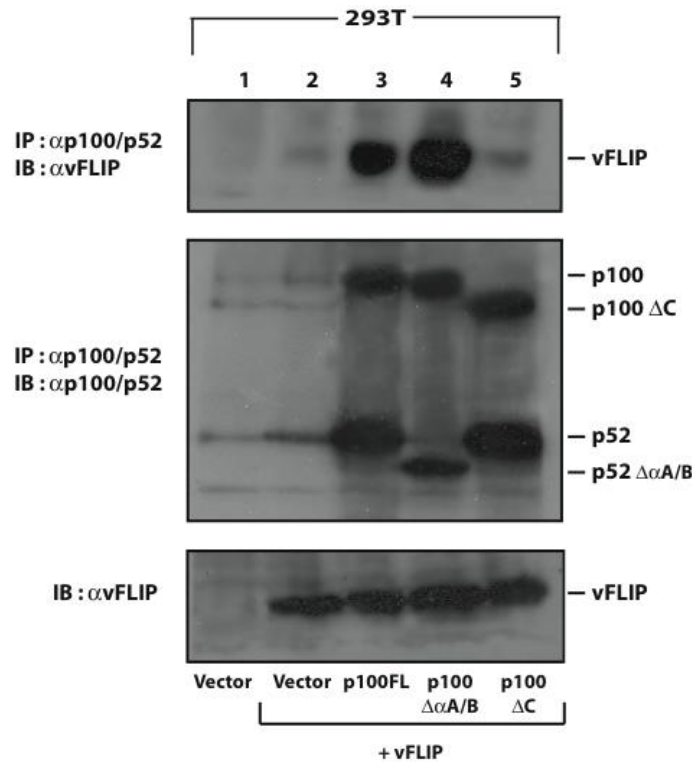


Figure 1.19: Western blot revealing that KSHV-vFLIP interacts with the C-terminus of p100. 293T cells were transfected with KSHV-vFLIP together with the full length p100 or mutants lacking α A/ α B (N-terminal helices implicated in p100 binding by the Human T cell leukemia virus Tax) or the C-terminal region (spanning amino acids 746-900). The p100 proteins were isolated from cell lysates by immunoprecipitation followed by detection of KSHV-vFLIP by immunoblotting (upper panel). The lysates were also subjected to immunoblotting to monitor p100 processing and expression using a p100/p52 antibody (middle panel). The lower panel shows KSHV-vFLIP expression (unpublished data reproduced from the PhD thesis of Sofia Eflkidou).

1.15 Aim

Endonucleolytic mRNA cleavage mediated by SOX during lytic infection with KSHV has been linked to the degradation of cellular mRNA which plays an important role in modulating host immune responses to viral invasion and disease progression. Thus, further understanding of the molecular mechanisms by which SOX induces host shutoff will lead to insights into how KSHV evades immune responses and establishes latency, both of which are directly related to viral pathogenesis. An additional aim was to gain insights into the mechanism of cellular transformation mediated by KSHV-vFLIP, a virally encoded protein that constitutively activates the alternative NF- κ B pathway through its association with the transcription factor p100 during latency. The objectives of the research reported in this thesis were therefore:

- To investigate the structural basis underlying how SOX recognizes mRNA targets and to deduce the most likely cleavage geometry using X-ray crystallography and biochemical assays coupled with site-directed mutagenesis given published reports of distinct binding modes for DNA and RNA.
- To establish whether the defects in host shut off exhibited by a set of reported mutants could be attributed to their inability to process structured DNA elements using fluorescence anisotropy and RNase assays.
- To identify potential inhibitors of SOX based on the structural studies and the available SOX-DNA complexes since impaired HSO results in attenuated

viral trafficking and substantially reduced latent infection.

- To elucidate the manner in which KSHV-vFLIP dysregulates the non-canonical NF- κ B pathway. To achieve this, IKK α and p100 constructs were designed and attempts made to study their interactions with KSHV-vFLIP and RelB using pull-down assays.

Chapter 2: Biophysical background theory

The aim of this chapter is to give a brief theoretical review of how X-ray crystallography is used to determine the molecular structures of proteins, with a particular emphasis placed on Molecular Replacement (MR), as this was the method applied to phase the SOX complex structures. This chapter also includes theories underlying the biophysical methods used.

2.1 X-ray Crystallography

2.1.1 Theory of Crystallisation and practical approaches.

Protein crystallisation requires the slow and controlled precipitation of proteins from an aqueous solution where the formation of crystals depends on the packing of irregularly ordered protein molecules into an ordered periodic lattice. Key to this is protein solubility and its behaviour in the presence of precipitants. Crystallisation occurs when a protein is brought up to a state of supersaturation following the addition of certain precipitants. The crystallisation process can be described by a phase diagram in which protein concentration is plotted as a function of precipitant concentration (Rhodes, 2006; Rupp, 2010) (Figure 2.1). This diagram consists of four zones. In the soluble region the protein remains in solution, whilst in the precipitation zone, it precipitates amorously. At the solubility line that separates these two phases, the protein solution becomes saturated and beyond that,

supersaturated. Supersaturation is a metastable state which can lead to the formation of small nuclei (ordered aggregates) in the nucleation zone. These pre-crystals can then form crystals that during their growth, depletes the available protein in the surrounding solution resulting in a reduction in protein concentration and a transition back to the solubility line (Figure 2.1).

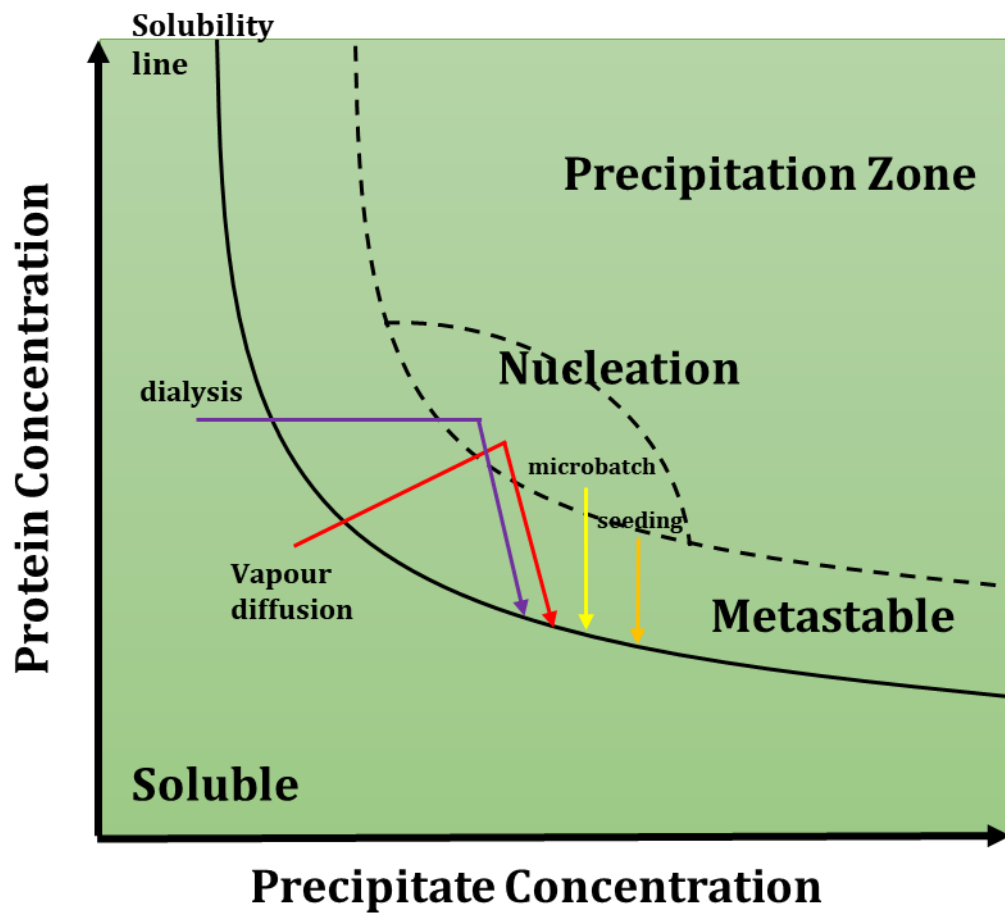


Figure 2.1: Crystallisation Phase Diagram. This shows the four regions (or phases) associated with protein concentration when varied as a function of precipitate concentration (Figure adapted and modified from Huub Driessen, px17.cryst.bbk.ac.uk).

There are several methods of producing gradual changes in protein concentration required for the growth of successful crystals (Rhodes, 2006; Rupp, 2010). In vapour diffusion, water evaporates gradually concentrating both the protein and precipitant in the mix which then slowly brings them to a supersaturated state resulting in the formation of crystals. In the microbatch method, crystals form when protein solution and precipitant are mixed to bring about a supersaturated state instantaneously. Seeding can also be used to increase the size of small crystals by introducing them into a supersaturated solution where they can act as platforms for the deposition of additional, ordered layers of protein. The dialysis method relies on the protein being separated from the supernatant by a semi-permeable membrane that only allows the flow of small molecules. Over time, there will be the flow of precipitants into the protein across the membrane from higher to lower concentration (or water in the reverse direction) until equilibrium is reached and hopefully supersaturation leading to the production of nuclei and crystal growth (Rhodes, 2006; Rupp, 2010).

Although the phase diagram shows crystallisation as a 2-parameter problem, in reality the crystallisation process is multi-factorial being also dependent on factors such as temperature, pH, types of buffer, types of additive and crystallisation technique. In order to supersaturate the solution, a trial-and-error variation of a multitude of biochemical and biophysical parameters can be used to decrease the solubility of the protein. This has been substantially aided by commercially available crystallisation screening kits that allow researchers to trial multiple conditions for subsequent optimisation (Rhodes, 2006; Rupp, 2010).

2.2 Crystal geometry

Crystals are *lattices* built from assemblies of repeated single units (represented by three vector operators) (Figure 2.2). As a consequence, a basic repeat unit can be defined that encompasses a motif/basis (generally DNA, RNA or protein) known as a *unit cell* which will generate the entire lattice when repeated in three dimensions (Rhodes, 2006; Rupp, 2010) (Figure 2.2).

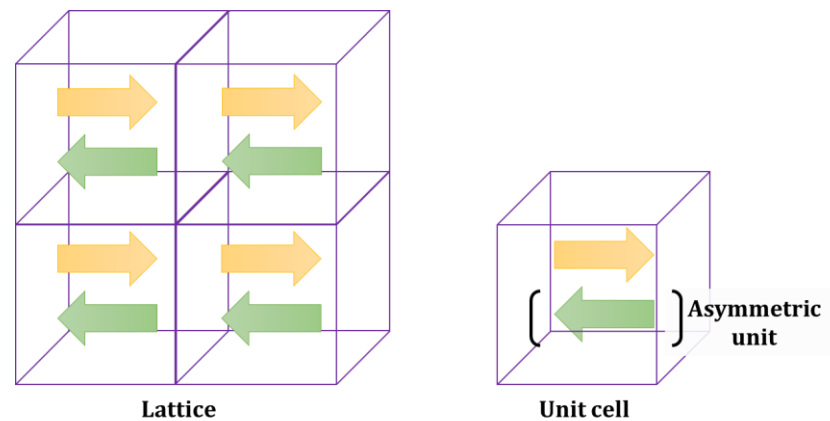


Figure 2.2: Crystal systems. A crystal is a symmetrical set of repeating units. Each basic repeating unit is called the unit cell. Identical numbers of molecules are arranged inside each unit cell and these molecules are related to each other by internal symmetry, generating a unit cell with multiple, symmetry related copies of the same molecule.

Each unit cell can be characterised by six parameters: three grid length (a , b and c) and their related three angles (α , β , and γ) (Figure 2.3).

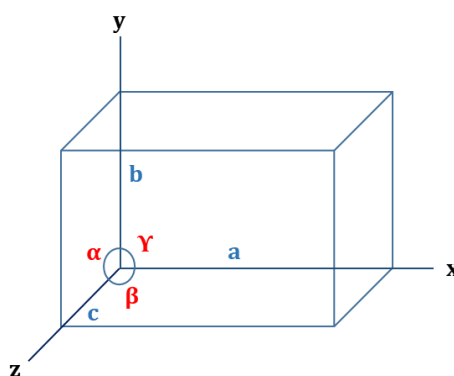


Figure 2.3: Overview of the general geometry of a unit cell.

Depending on the combination of orders of rotational symmetry within the lattice, a crystal can be classified into 7 groups, collectively called Crystal Systems. These are referred to as triclinic, monoclinic, orthorhombic, trigonal, hexagonal, tetragonal and cubic systems where the rotational axes place restrictions on the cell axes and angles of the associated unit cells. The combination of the 7 crystal systems with the seven lattice centering options leads to a maximum of 14 unique lattice types, which are referred to as the Bravais lattices. All crystals have at least a primitive Bravais lattice (Rhodes, 2006; Rupp, 2010) (Table 2.1). Through different combinations of rotational symmetry and translational symmetry, there are a total of 65 different ways that crystals can assemble from asymmetric chiral protein molecules into three-dimensional periodic crystals. Based on these 65 possible assemblies, the crystal can belong to 65 chiral space groups. Space groups provide full information on the rotational and translational elements that need to be applied to a protomer in order to generate the crystal. Knowledge of the space group is essential for determining the protein structure (Rhodes, 2006; Rupp, 2010).

Table 2.1: The seven Bravais lattices and associated unit cell constraints.

Crystal System	Variations	Unit cell Parameters	
		Faces	Angles
Triclinic	Primitive	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$
Monoclinic	Primitive, Base-Centred	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ; \beta \neq 90^\circ$
Orthorhombic	Primitive, Body-, Base-, and Face- Centred	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Tetragonal	Primitive, Body- Centred	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Trigonal (Rhombohedral)	Primitive	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$
Trigonal/Hexagonal	Primitive	$a = b \neq c$	$\alpha = \beta = 90^\circ;$ $\gamma = 120^\circ$
Cubic	Primitive, Body-, and Face-Centred	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$

2.3 X-ray scattering and Bragg's law

X-rays undergo scattering by electrons when applied to matter. When the objects take the form of a crystal whose constituents are of a comparable size to the incident X-rays, the various scattered waves will undergo a process of either constructive or destructive interference giving rise to what is known as a diffraction pattern (Rhodes, 2006; Drenth, 2007; Rupp, 2010) (Figure 2.4A).

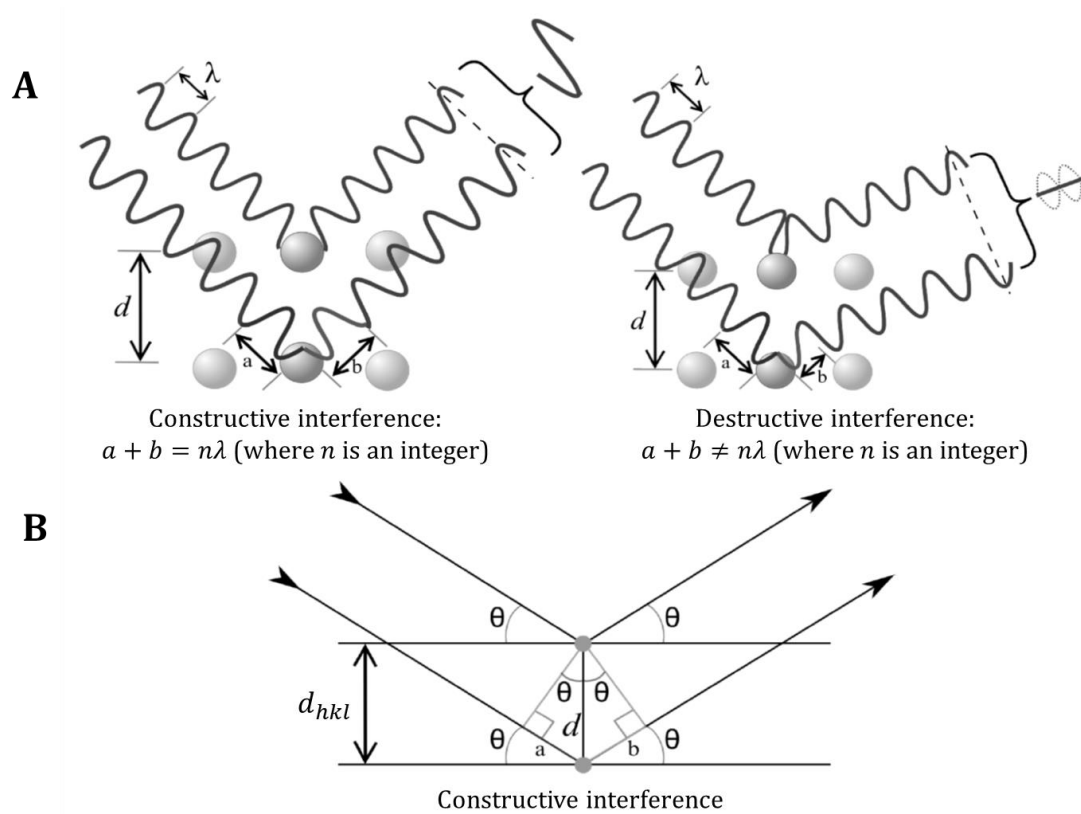


Figure 2.4: (A) Constructive interference (left) in which two waves are in phase with one another. Destructive interference (right) in which two wave are out of phase. (B) Schematic diagram of two waves interfering constructively. In this case, the path difference is $2d \sin(\theta)$, which give rise to the Bragg equation. (Figures adapted and modified from <http://px17.cryst.bbk.ac.uk>).

The regions of darkness and brightness associated with the diffraction pattern will depend on the angle of observation and the structure of the object. The conditions that lead to constructive interference were considered by Bragg where for the purposes of simplicity, a crystal can be thought of as being comprised of an infinite series of parallel planes defined by where they intersect the axes of the unit cell (defined by the lengths of the cell edges a, b and c). Each set of parallel planes is assigned the indices hkl depending on where they intersect. Bragg showed that the

conditions under which diffraction can occur can be considered as if arising from X-rays reflected from these sets of parallel planes of atoms (Rhodes, 2006, Drenth, 2007, Rupp, 2010). This is illustrated in Figure 2.4B. Furthermore, Bragg showed that the diffraction process in the crystal can be treated as if the X-rays were reflected by sets of parallel planes of atoms in the crystal, thereby referring to the diffracted X-ray intensities as 'reflections'. When considered, Bragg's law (illustrated in equation 2.1) demonstrates that diffraction maxima will only be observed at particular values of $\sin\theta$. This gives rise to a lattice in diffraction space when taking into account all possible sets of parallel planes (Figure 2.4B). The real and diffraction space lattices, however are reciprocally related leading to the diffraction space lattice being referred to as the reciprocal lattice. The major simplification of Bragg's law is that all atoms are assembled in planes which is not the reality so all atoms contribute to each diffraction maxima or reflection.

Equation 2.1: Bragg's Law

$$2d_{hkl} \sin \theta = n\lambda$$

Where, θ = incident angle

d_{hkl} = space between parallel planes on which atoms are located

n = integer

λ = wavelength of the incident wave.

2.4 Data collection

In order to determine the crystal structure, the diffraction pattern first has to be collected. This is achieved using a set up similar to that shown in Figure 2.5 where a crystal is irradiated with X-rays and the diffraction pattern intercepted by an X-ray sensitive detector (Rhodes, 2006; Drenth, 2007; Rupp, 2010). A typical data collection geometry is one in which the crystal is oscillated through 0.1 to 1.0° (depending on the type of detector and factors such as the unit cell dimensions,) over an angular range determined by crystal symmetry. The first step is to determine the unit cell and likely crystal system from test images typically 45-90° apart to enable sufficient sampling of reciprocal space using Software suites such as the XDS pipeline (Kabsch, 2009, iMOSFLM (Battye et al., 2011) and DIALS (Diffraction Integration for Advanced Light Sources, Brewster et al., 2018) in a process known as auto-indexing. These results will determine the best parameters for data collection influencing factors such as the amount of data that needs to be collected and optimal oscillation angles. Once collected, the data will then undergo processing where each reflection (diffraction spot) is assigned an index hkl and its corresponding intensity determined either through a process of simple integration (where intensity contributions in each pixel are summed) or by profile fitting where the intensity is obtained by scaling an average profile obtained from a number of a diffraction spots in a particular region of the detector. At this stage, the crystal symmetry and space group can be ascertained by comparing diffraction spots that should be symmetry related to establish whether they do have similar intensities and then combining them to produce a unique set in the event that this is the case. The final stage of the

data collection process is truncation where the diffraction intensities are reduced to amplitudes (known as structure factor amplitudes) given that $(I_{hkl})^2 = (F_{hkl})^2$, where I_{hkl} is the diffraction spot intensity and F_{hkl} the amplitude (referred to as the structure factor amplitude) (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

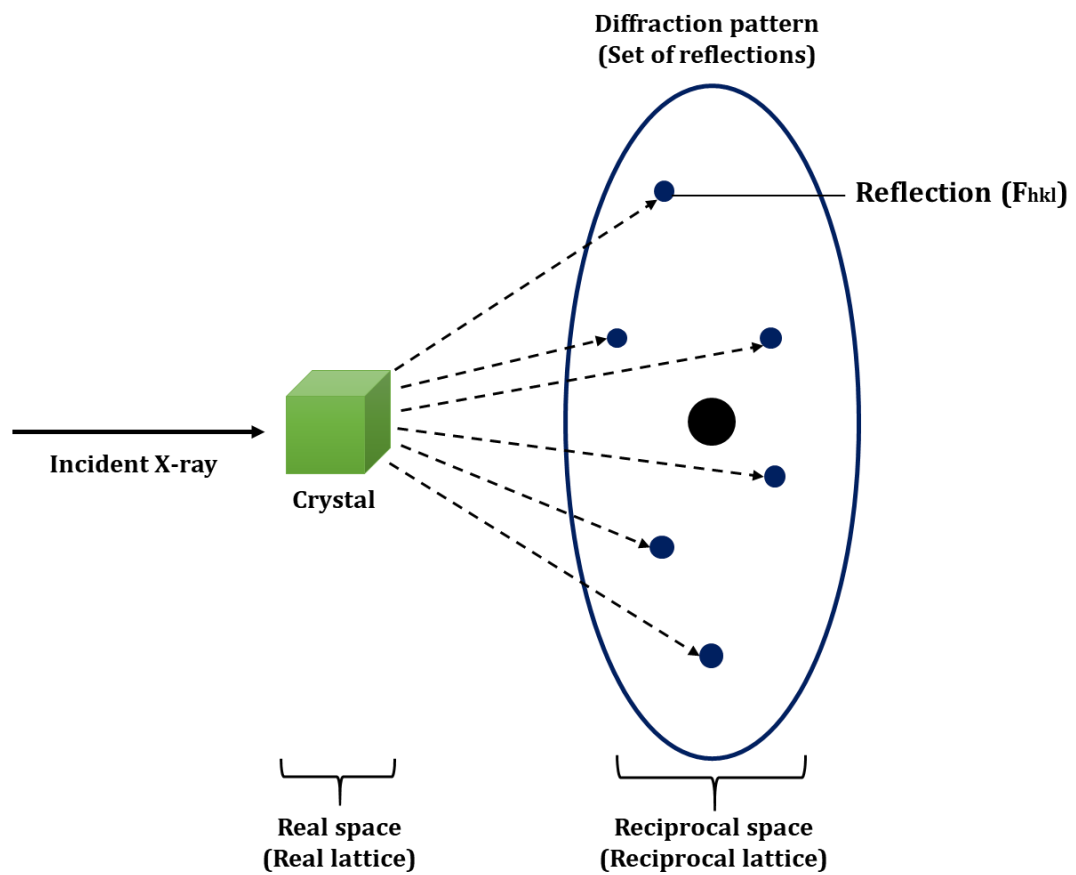


Figure 2.5: X-rays Crystallography experimental set up. Upon incident X-ray beam, the crystal produces a diffraction pattern of spots. These spots are a product of the scattering power of the individual atoms within the unit cell that make up a crystal, with each reflection F_{hkl} corresponding to all atoms that effect the scattering.

2.5 Structure factors and the Fourier Transform

As previously described, the diffraction pattern can be seen as the sum of the scattered waves of all atoms that contribute to sets of parallel planes in the crystal lattice. These “reflections” are defined by where they intersect the unit cell (coordinates h , k , and l , called reflection indices), and, since reflections are produced by waves, they can be characterised by three different properties: frequency (or wavelength), amplitude and phase (Rhodes, 2006; Drenth, 2007; Rupp, 2010). The frequency is defined by the length of a single repeat/wavelength, which in turn defines how frequently the terms $F(x)$ are repeated. The amplitude is defined by the minimum and maximum of the wave over this frequency/wavelength. The phase is specified by the position of the wave from its origin, which in turn defines the angular position of the wave from the origin, indicating the degree that the wave has moved from the origin (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

Each diffraction spot contains scattering contributions from each atom that can be added vectorially and replaced by a Fourier series to produce the structure factor that represents the overall scatter in a direction defined by hkl as represented in equation 2.2.

Equation 2.2: Structure Factor as a Fourier Sum.

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

Where, (F_{hkl}) is the structure factor for a particular reflection hkl that is the sum of the scattering contributions from all of the j atoms (f_j) ranging from (1 to n) where x_j , y_j and z_j correspond to their fractional co-ordinates.

The structure factor (F_{hkl}) will have an associated amplitude and phase and can be re-written as:

$$F(hkl) = |F(hkl)|e^{i\phi(hkl)}$$

It can be directly related to electron density by the electron density equation where the structure factor and electron density are seen to be related by a Fourier transform. This can be visualised as electron density resulting from the summation of waves represented by the F_{hkl} 's that have different amplitudes/phases and are therefore analogous to the sine and cosine components of Fourier series used to represent periodic functions (Rhodes, 2006; Drenth, 2007; Rupp, 2010). The electron density equation can therefore be written as:

Equation 2.3

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

Where, $\rho(x, y, z)$ is the electron density at point (x,y,z),

V=volume of unit cell

F_{hkl} = the structure factor amplitude

x,y and z =coordinates of unit cell

h,k,l= Miller indices

F_{hkl} is the complex structure factor for reflection indices h, k, l that can be written $|F_{hkl}|e^{i\varphi}$, where $|F_{hkl}|$ is the amplitude and φ the phase. Whilst all of the parameters are known, the only exception is the structure factor phase φ , which is essential for determining the electron density. This is referred to as the phase problem in protein crystallography (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

2.6 The Phase problem

Experimentally, all phase information is lost in a typical diffraction experiment owing to the absence of an X-ray lens that would maintain the phase relationships between the various diffracted rays. Since calculation of electron density requires phase information, this is referred to as the 'phase problem'. Further computational or experimental methods need to be carried out in order to recover the missing phases. There are three main methods to estimate the missing phases: molecular replacement (MR), direct methods, and experimental methods. Since MR was the only technique used in this study, direct and experimental methods will not be discussed in detail.

In brief, experimental methods rely on heavy atom substitution of the protein. These include *isomorphous replacement* or exploitation of the phenomenon of *anomalous scattering for single- or multiple-wavelength anomalous dispersion (SAD/MAD)*. The incorporation of heavy atoms subsequently results in greater electron density and scattering power, leading to significant differences between the equivalent native and heavy atom substituted structure factors. These differences can be utilised to

obtain estimates of the missing protein phases. To generate heavy atom derivatives of proteins, classically, existing crystals are soaked or co-crystallised in solutions containing heavy atoms. However, binding can disrupt crystal packing leading to changes in monomer conformation and in extreme cases space group transitions which can all lead to the native and derivative datasets no longer being isomorphous (Rhodes, 2006; Rupp, 2010). A second method that circumvents this problem makes use of the phenomenon known as anomalous scattering. Atoms exhibit anomalous scattering when the incident X-rays have a wavelength near to the absorption edge of the atoms. Part of the radiation is absorbed by the atoms and reemitted with a change in phase. The scattering factor for the atom can be written in terms of real and imaginary parts:

Equation 2.4:

$$f = f_0 + f' + if''$$

Where, f_0 is the normal scattering factor away from the adsorption edge and $(f' + if'')$ is the anomalous scattering correction.

2.7 Solving the phase problem by molecular replacement (MR)

In molecular replacement (MR), the phases relating to the target of interest are estimated from an existing set of homologous co-ordinates (Rhodes, 2006; Drenth, 2007; Rupp, 2010). In order to be successful, the similarity between the molecular replacement model and the target protein structure should be as high as possible (at

least more than 30 % sequence identity). Typically, the root mean squared deviation between the alpha carbons of the target protein and the molecular replacement model need to be less than 1.5 Å (Rhodes, 2006; Drenth, 2007; Rupp, 2010). It cannot be assumed that structures of high similarity will always be successful in molecular replacement since they may undergo conformational changes that will have a negative impact. There are two different approaches to MR: Patterson-based and maximum likelihood-based MR. Software such as Phaser uses maximum likelihood for MR while other software such as Molrep uses a Patterson-based MR (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

2.7.1 Patterson-based MR

The Patterson function, $P(u,v,w)$ is derived through multiplying electron density by itself in a process known as autocorrelation (equation 2.5). Whilst the calculation of electron density requires phases, simply squaring the electron density equation produces one that is phase independent (equation 2.5).

Equation 2.5: Patterson function

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{(hkl)}|^2 \cos 2\pi(hu + kv + lw)$$

Where, $P(u, v, w)$ is the Patterson function at point (u,v,w) on the Patterson map; V is the unit cell volume; $|F_{(hkl)}|$ is the structure factor amplitude; u, v, w are the Patterson map coordinates and h, k, l are the three indices of the corresponding reflection.

A Patterson map has coordinates of u,v,w rather than x,y,z as it has peaks corresponding to vectors between atoms. The goal of MR is to first rotate the model into the same orientation as the target followed by positioning via translation to fully superpose the model and target (Rhodes, 2006; Drenth, 2007; Rupp, 2010). The MR process can therefore be simplified from a 6 dimensional problem to two three dimensional problems involving rotation and translation as illustrated in Figure 2.6.

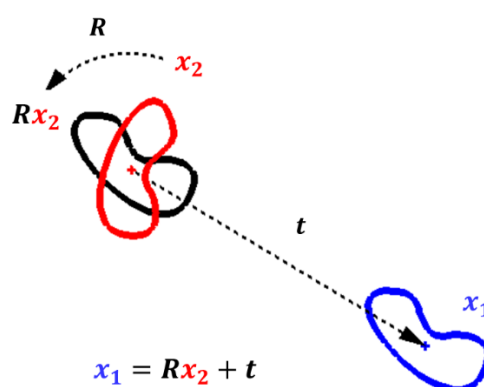


Figure 2.6: Schematic diagram of MR method. x_1 =target co-ordinates, x_2 = model co-ordinates, R = rotation matrix that when applied to x_2 , places it in the same orientation as x_1 and t the translation vector that superposes the two sets of co-ordinates. (Figures adapted and modified from <http://px17.cryst.bbk.ac.uk>)

Traditionally, this was made possible by the fact that a protein crystal can be seen as being comprised of atoms that are either related by intramolecular vectors (within protomers) or intermolecular vectors (i.e., between protomer atoms that are related by crystallographic symmetry). This means that a rotation function can be constructed from Pattersons generated using intramolecular vectors and the translation function from those involving intermolecular vectors (Figure 2.7).

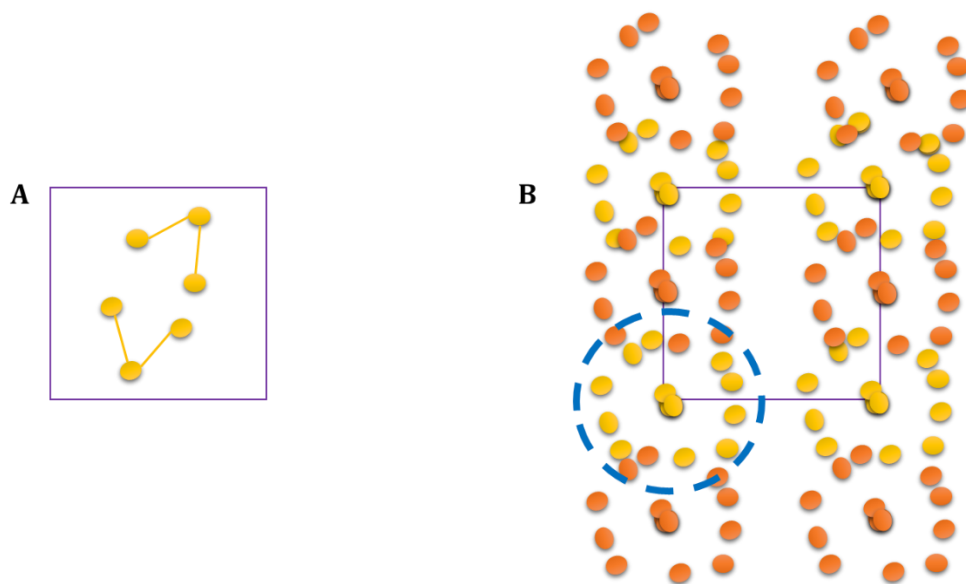


Figure 2.7: (A) Simple 2D structure in a P2 unit cell. (B) Corresponding Patterson map showing intramolecular vectors coloured in yellow and intermolecular vectors coloured in orange. An example of a radius of integration set to separate *intra* from *inter* molecular vectors highlighted in blue dotted circle. (Figures adapted and modified from <http://px17.cryst.bbk.ac.uk>)

Although it's computationally possible to force six dimensional searches, the task of computing structure factors for an entire macromolecule in all possible orientations on a suitably large search grid is extremely demanding. A more computationally efficient approach is to split the problem into two three dimensional searches, a rotation function search, followed by a translation function search (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

2.7.1.1 Rotation function

In the rotation function, the search Patterson map is rotated incrementally around the target Patterson map with a correlation score calculated for each angular increment where the search Patterson map is reduced to the strongest few thousand peaks (Rhodes, 2006; Drenth, 2007; Rupp, 2010). Therefore the overlap function between Patterson functions (\mathbf{R}) is the integrated (over all points \mathbf{u}) product of the observed crystal Patterson $P_{observed}(\mathbf{u})$ and search model Patterson $P_{model}(\mathbf{R}, \mathbf{u})$, with \mathbf{R} as the rotation matrix within a sphere of radius r_{max} centred on the origin and excluding the origin peak out to a radius r_{min} (Equation 2.6).

Equation 2.6: Rotation function

$$RF(\mathbf{R}) = \int_{r_{min}}^{r_{max}} P_{observed}(\mathbf{u}) \times P_{model}(\mathbf{R}, \mathbf{u}) d\mathbf{u}$$

The results from the rotation function are scored by ranking the strongest peaks in order of units of σ above the average solution background.

2.7.1.2 Translation function

In order to correctly position a molecule in the unit cell, a set of high scoring orientations from the rotation function are then trialled in the translation function (Rhodes, 2006; Drenth, 2007; Rupp, 2010). Similar to the rotation function search, a translation function is accomplished by calculating the correlation between the data and a correctly orientated model at various points in the asymmetric unit cell

(equation 2.7). For the translation function search the longer intermolecular vectors are relevant.

The number of definable parameters required for the translation function varies according to the symmetry of the target. For example, since there are no symmetry elements in a space group P1, the position of the molecule inside the cell is arbitrary. By considering possible multiple origins for differing space groups, different 'Cheshire cells' can be defined that delineate the minimum cell volumes that must be searched, the size of which is proportional to space group symmetry.

Equation 2.7: Translation function

$$T(\omega, \tau) = \sum_{\text{inter}} P_o(\mathbf{u}_i) \cdot P_c(\mathbf{u}_i, \omega, \tau)$$

Where, P_o = observed (target) Patterson, a function of intermolecular vector

P_c = calculated Patterson, a function of intermolecular vector

ω = Rotation angle (orientation matrix)

τ = Translation vector

\mathbf{u}_i = Intermolecular vector

Translation functions can be scored by correlation coefficients or by R_{factor} (discussed in session 2.9). It is known that several MR programs also incorporate more advanced packing function analysis and also include a penalty for clashing molecules.

2.7.2 Maximum likelihood molecular replacement

Maximum likelihood (ML) methods are based on the assumption that the most likely model is the one with the highest probability of generating the observed data. This allows more realistic target functions that account for incompleteness and errors in models (Rhodes, 2006; Drenth, 2007; Rupp, 2010). Key to the maximum likelihood is conditional probability known as Bayes' theorem (Equation 2.8) which relates the posterior probability ($P(\text{model}:\text{data})$), to the prior ($P(\text{model})$) and likelihood ($P(\text{data}:\text{model})$) probabilities.

Equation 2.8

$$P(\text{model}:\text{data}) = P(\text{model}) \times P(\text{data}:\text{model})$$

In MR, the maximum likelihood treatment begins with X-ray data (intensities or amplitudes) and the prior probability ($P(\text{model})$), therefore the task becomes one of selecting the model with the maximum likelihood (or highest probability). In MR, the likelihood or probability is computed where the experimental structure factor amplitude (F_o) serves as the *data*, and the model structure factor amplitude (F_c), the *model*.

2.7.2.1 Likelihood based rotation function

Given the lack of defined position of the search model at each orientation in the unit cell, it is more difficult to apply maximum likelihood methods to the rotation function search. The structure factor amplitudes for the whole unit cell (F_c) cannot be calculated as undefined positions in real space correspond to undefined phases of the structure factors and therefore the likelihood cannot be estimated. Moreover, since the model structure factor amplitudes remain without phases, they cannot be combined directly with the symmetry related molecules (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

To overcome this, contributions to the structure factor amplitudes can be determined from each of the symmetry related molecules given that all that needs to be known are the various atoms involved and the atomic scattering factors f_j which can be calculated at different angles. Since the relative phases of the various contributions are unknown and can be anywhere in the complex plane, this is equivalent to a random walk with steps of known length, but unknown direction and can be estimated by a 2D Gaussian function where the variance is equivalent to the sum of the squares of the individual contributions. Since a probability distribution function can be defined, a rotation likelihood function can now be constructed of the form: $p(\text{data} | \text{model}) = p(F_{\text{obs}} | (F_m)_{\text{sym}})$, where $(F_m)_{\text{sym}}$ are the contributions from each symmetry related monomer. These probability functions, however, are normally based on the full phased structure factor F_o , so to counteract the fact that the phase is unknown, a relative phase between F_o and one of the F_m components is

normally introduced. Usually, the $(F_m)_{\text{sym}}$ that makes the largest contribution to the total amplitude (F_{big}) is chosen. The remaining symmetry related contributions will therefore be centred at its end and can be represented by a random walk but with a much sharper distribution (Figure 2.8). The variance may be inflated to account for any addition model errors by the introduction of a factor D. Final analysis involves the phase variable between (F_o) and (F_{big}) being integrated out to give the probability as a function of amplitude (Rice distribution). Since the reflections can be assumed to be independent, the likelihood function for the rotation search can simply be taken as the product of all of the individual reflection likelihoods which will be computed for all possible orientations. Since the probability values are usually very small, -log-likelihood is normally used as the target function (also in translation searches) (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

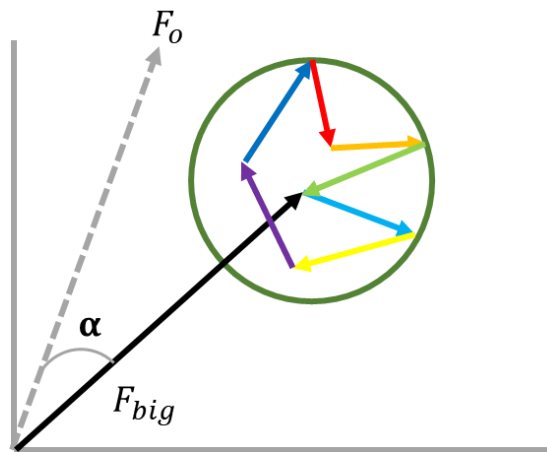


Figure 2.8: A random walk is used to generate a probability function for F_{hkl} as a two dimensional Gaussian centred on F_{big} to overcome the lack of relative phase information in the rotation problem. The relative phase α between F_o and F_{big} can be integrated out by evaluating the probability at all values of phase between 0 and 360°.

Since this search is computationally expensive, fast Fourier transform methodologies can be applied using a Taylor series expansion to expedite the computation of full likelihood targets, followed by a re-scoring of the top peaks using full likelihood (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

2.7.2.2 Likelihood based translation function

Maximum likelihood methods are straightforward to apply to the translation function search given that the molecule has at this stage been correctly orientated and can be positioned at some search point. Therefore, the locations of all of the atoms are known and the structure factor amplitude (F_c) can be easily calculated. A sum of individual atomic structure factors (F_{atom}) from the model are used to produce the structure factor for the model in the asymmetric unit. This is then combined with any symmetry relatives in the unit cell to derive the observed structure factor amplitude for the unit cell (F_c). The likelihood function can then be determined on a grid of search positions within the unit cell to deduce the translational operator (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

Different potential solutions for a molecular replacement are compared by their log-likelihood-gain (LLG) and Z-score values. While the LLG measures how much better the positioned model is for a potential solution when compared to a randomly placed set of atoms, the Z-score measures the standard deviations above the mean of LLG value. Generally, when a Z-score is more than 5, the potential molecular replacement solution is usually correct. If the LLG gives a negative value, this indicates that a poor

search model is used as it describes the target data more poorly than a random collection of atoms (a low signal-to-noise ratio). However, the LLG can improve with every rotation or translation step when used in the right direction decreasing the randomness as the matching between the data and the model increases. Also, the LLG score can be improved by maintaining the highly conserved residues between the unknown structure and the existing model (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

Software, namely *Phaser*, can be employed to increase the probability of determining an MR solution by using maximum-likelihood (McCoy et al., 2007) as well as enabling the use of low homology and ensemble models which can be appropriately weighted. This can lead to success in difficult molecular replacement cases where traditional Patterson based methods fail.

2.8 Electron density maps

Once a solution is found, the estimated phases from MR are then used to calculate an electron density map in which an atomic model of the target protein can be built. These maps are calculated using different combinations of coefficients into the Fourier synthesis. An (F_o-F_c) difference map is generated by subtracting the calculated structure factor amplitudes (F_c) from the observed structure factor amplitudes (F_o). Negative electron density map indicates regions where the model has density that should be absent and is indicative of differences between the model and target owing to conformational changes or differences in amino acid sequence

for example. Positive electron density indicates regions where the observed data suggests the presence of atoms which are not included in the model. (F_o-F_c) difference maps are commonly used in conjunction with a $(2F_o-F_c)$ map where there is increased weight placed on the experimental data that “up-weights” density that is not present in the current model (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

The next step in the process is to optimize the fit between the model and the data. This is achieved through minimising the difference between the set of observed structure factors that were experimentally derived and those calculated from the model.

2.9 Refinement and validation of macromolecular models

The purpose of refinement is to improve the fit between the experimental observations and the model parameters in order to produce the model most consistent with the data. Historically, this was performed by minimising an energy equation of the form given below using least squares methods:

$$E \text{ (Total)} = E(\text{X-ray}) + E(\text{stereochemical})$$

Where $E(\text{x-ray})$ corresponds to the residual $W(F_o-F_c)/F_o$, where W is the relative weight, and $E(\text{stereochemical})$, terms that include bond lengths, bond angles and chirality restraints, along with non-bonded interactions such as Van der Waals and electrostatic interactions. This least squares based method of refinement has been largely superseded by the application of maximum likelihood that enables errors in

the model to be taken into account such as overall completeness (Rhodes, 2006; Drenth, 2007; Rupp, 2010).

In the refinement process, the two values, R_{factor} and the R_{free} are used as a quality indicator to assess whether the refinement strategy was successful or not. The R_{factor} is a measurement of the agreement between the calculated structure factors from the model and the observed structure factors from the diffraction experiment (Equation 2.9).

Equation 2.9:

$$R = \frac{\sum_{hkl} (|F_{\text{obs}}| - |F_{\text{calc}}|)}{\sum_{hkl} |F_{\text{obs}}|}$$

Where, F_{obs} are the observed structure factors, and F_{calc} are the calculated structure factors.

The conventional R_{factor} is open to manipulation by for example sigma and resolution cut-off, inappropriate weighting and overfitting by the introduction of more parameters than the data allow. In an ideal case, the value of R_{factor} (or R_{work}) would be 50 % for a molecular replacement solution and should be kept below 30 % after several refinement. R_{free} is free from model bias as it is calculated using only a small subset of the total (typically 5%) of the experimental reflection that are systematically excluded from the refinement process. This percentage is chosen to ensure that a sufficient number of reflections remain for the production of accurate electron density.

In the atomic models derived from molecular replacement or experimental phasing, all atoms have a position (x_j, y_j, z_j) and a natural oscillation that is determined by the local environment of that atom. This oscillation is denoted by the temperature factor, B (or B-factor), and it is assumed that side chain atoms will exhibit more flexibility than in the main chain.

There are two main aspects of refinement: maintaining the geometry of the atoms whilst optimising the agreement between the model and the electron density. A molecular modelling program such as *Coot* is generally used to manually build the residues inside the electron density and regularise the stereochemical properties of peptides including bond lengths, bond angles and side chain rotamers (Emsley et al., 2010) where large changes are required (such as the re-building of loops and substitution of amino acid residues/inclusion of solvent molecules depending on the resolution). These models are then submitted to programs such as Buster or Refmac 5 (Murshudov et al., 2011; Vagin et al., 2004) where they are refined to better fit the diffraction data by altering the x,y,z (and B-factor) parameters. The phase estimates of F_c are therefore improved by these processes that iterate between model rebuilding and program based refinement, resulting in sequentially improved electron density maps. This is continued until convergence is reached where no further improvements are observed.

Throughout the processes, the geometry of the model is continuously analysed through calculation of a Ramachandran plot, which highlights residues of poor geometry by plotting the relative Φ and Ψ angles of each amino acid (Ramachandran

and Sasiskharan, 1968) in programs such as MOLPROBITY (Chen et al., 2010) and PROCHECK (Laskowski et al., 1993) in addition to COOT.

2.10 Biophysical methods to probe protein-ligand interactions.

2.10.1 Fluorescence Anisotropy (FA) background theory

Fluorescence Anisotropy (FA) is a method used to determine the affinity of proteins for their ligands (Singh et al., 2000; Shi and Herschlag, 2009). It is based on the principle of photoselective excitation of fluorophores by polarized light. When excited by polarized light, fluorophore molecules whose absorption transition dipoles are parallel to the electric vector of the excitation become selectively elevated to an excited-state. In other words, if the fluorophores are excited by the vertically polarized light, light is emitted in the same polarized plane, provided that the molecule remains stationary throughout the fluorescence lifetime (time between excitation and emission). However, in a homogeneous solution, the fluorophores will be oriented randomly and upon excitation, small fluorescent molecules tumble rapidly resulting in largely depolarized emitted light (Figure 2.9). On the other hand, a fluorescently tagged larger molecules will undergo slower motions during the fluorescence lifetime and therefore the emitted light will remain highly polarized. Large molecules rotate slowly/little during the excited state resulting in high values of polarization while small molecules rotate rapidly resulting in low levels of polarization. A spectrometer can be used to measure the intensity of the emission (Figure 2.10).

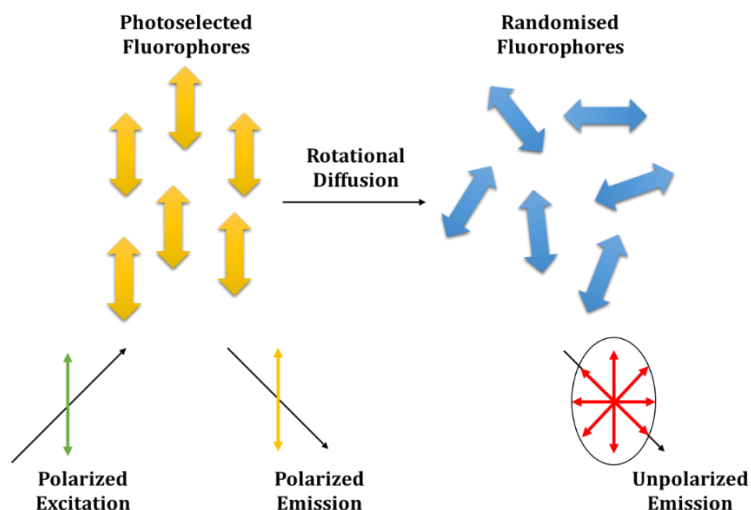


Figure 2.9: Fluorophores excited by vertically polarized light. The emitted light is in the same polarized plane, provided that the molecule remains stationary. However if the fluorophores are randomised, they tumble rapidly upon excitation and the emitted light is largely unpolarised.

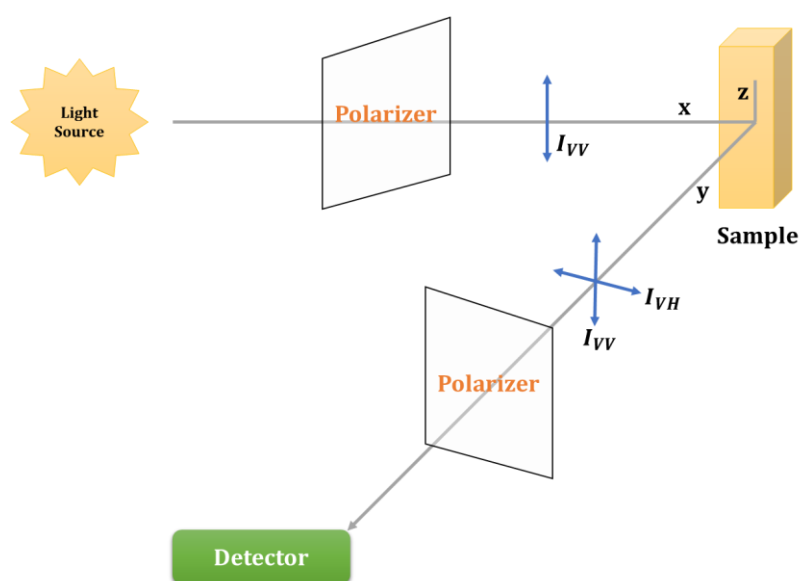


Figure 2.10: Fluorescence Anisotropy Experimental setup. The sample is excited with vertically polarized light which is oriented parallel to the vertical (z-axis) and a polarizer is used to measure the intensity of the emitted light.

The two intensity values, I_{VV} and I_{VH} , can be obtained to calculate the anisotropy r which is shown in equation 2.10.

Equation 2.10

$$r = \frac{I_{VV} - I_{VH}}{I_{VV} + 2I_{VH}}$$

Where, r is the anisotropy and VV denotes vertical excitation, vertical emission and VH denotes vertical excitation, horizontal emission.

The anisotropy r can be used to measure the dissociation constant (K_d) between two molecules independently of the protein concentration, given that it is a function of the overall rotational correlation time which varies with the rotation rate of the molecule. This has been shown to be dependent on the viscosity of the solvent, the molecular volume and temperature and thus changes when the protein associates with its ligand. The rotational correlation time can be directly linked to the intensity changes of polarized light emitted by fluorophores that are covalently attached to either the protein or ligand, following irradiation (Jameson and Sawyer, 1995). The extent of this reduction or difference in anisotropy (r) is dependent on the rotational correlation time of the fluorophore, during the lifetime of fluorescence, and in the absence of other processes that would further reduce the anisotropy, is given by the Perrin equation illustrated in equation 2.11 (Weber, 1953).

Equation 2.11: Perrin Equation

$$r = \frac{r_0}{1 + \tau/\phi}$$

where, r is the observed anisotropy

r_0 is the intrinsic anisotropy of the molecule

τ is the fluorescence lifetime

ϕ is the rotational correlation time for the diffusion process

The dissociation constant (K_d) is determined by measuring the change in anisotropy for samples in a titration series where the concentration of the ligand is kept constant, and the concentration of the protein increased (Pollard, 2010) and can be calculated from:

Equation 2.12: Dissociation constant K_d

$$K_d = \frac{[P][L]}{[C]}$$

Where, K_d is the dissociation constant (M); $[P]$ is the molar concentration of the Protein (M); $[L]$ is the molar concentration of the Ligand (M); $[C]$ is the molar concentration of the protein-ligand complex (M).

2.10.2 ThermoFluor Assay

The ThermoFluor assay is a temperature-based assay that utilizes SYPRO Orange and other chromophores that serve as a solvatochromic fluorescent dyes which have the ability to bind to hydrophobic areas within proteins that stimulate fluorescence emission (Figure 2.11). Initially, the protein is in its folded (native) state and when the temperature increases, slowly undergoes thermal denaturation (protein melting). When the protein starts to unfold, it exposes its hydrophobic core to SYPRO orange that binds emitting higher levels of fluorescence. As the temperature increases, a fluorescence maximum is reached and the protein begins to aggregate dissociating the dye. The dye becomes unquenched and fluorescence emission decreases. The melting temperature (T_m) of the protein is the inflection (mid) point of the initial slope (Huynh & Partch, 2015). Addition of a ligand will shift the melting temperature of the protein. Although ligand binding often produces greater stability and positive shifts in melting temperature, it may also result in protein distortion which can then lead to the complex being less thermally stable than the protein alone resulting in a negative temperature shift.

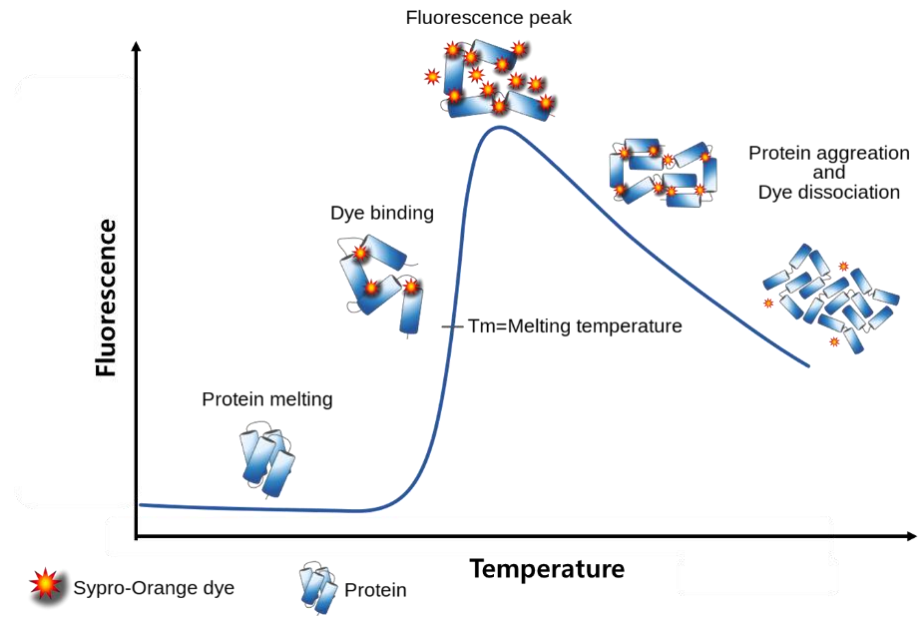


Figure 2.11: Overview of ThermoFluor analysis of protein stability and ligand interactions. (Figure adapted and modified from Huynh & Partch, 2015).

Chapter 3: Materials and Methods

3.1 WTSOX Protein Expression and Purification

The details of cloning, transformation and expression of WTSOX were performed using the same protocol from *Bagneris et al., 2011*.

3.2 Preparation of SOX mutants: SOXE244S, SOXC247S, SOXF179A, SOXV369I, SOXP176S, SOXA61T, SOXD221S, SOXD474N and SOXY473Stop

3.2.1 Mutagenesis

To investigate the role of the HSO mutants in SOX-mediated host mRNA degradation, several amino acids were mutated using the designed primers shown in Table 3.1. All other mutants were kind gift from Dr Claire Bagneris.

Table 3.1: Primers used for SOX mutants

Mutant name	Primer sequences
F179A	FW 5' CGGGCATGAGGTTCTGGTCAAGGTGAACATAAGATCATG-3' RV 5'CATGATCTTATGTTACCTTGACCAGGAACCTCATGCCCG-3'
C247S	FW 5' GGTATACGAGATCAAATCCCGCTTTAAGTATACG-3' RV 5'CGTATACTTAAAGCGGGATTTGATCTCGTATACC-3'

All mutants were generated using site directed ligase independent mutagenesis (SLIM) to introduce point mutations in WTSOX using the KAPA HiFi PCR Kit (KAPA Biosystems). Similar to the inverse PCR mutagenesis, this technique introduces desired point mutations into the oligonucleotide primers with complementary tails at the opposite termini and uses amplification of the whole vector. The primers were designed to be fully complementary to the protein sequence with the exception of the mutation sites. The PCR reactions consisted of 2XKAPA HiFi Buffer (KAPA Biosystems), 0.3 μ M Forward Primer, 0.3 μ M Reverse Primer, 2 ng Template DNA and PCR-grade water up to 50 μ L. For the mutagenesis reaction (PCR), the mixture was heated to \sim 95 $^{\circ}$ C to fully denature the double stranded DNA template. The duration ranged from 30 s to more than 5 mins, based on the GC base-pair content of the template. This was next followed by, 20-30 cycles of three successive steps: denaturation, annealing, and extension. In the denaturation step, the temperature was again elevated to above 90 $^{\circ}$ C for strand separation. Subsequently, the temperature was lowered to approximately 50 $^{\circ}$ C to allow primer annealing onto the target sequences. Finally, the temperature was increased to 80 $^{\circ}$ C to activate the DNA polymerase for primer extension (Table 3.2).

PCR products were treated with the restriction enzyme DpnI at 37 $^{\circ}$ C for 1 hour to digest all methylated and hemi-methylated parental DNA.

Table 3.2: Overview of the PCR reaction

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	5 mins	1
Denaturation	98	20s	30
Annealing	63	15s	
Extension	72	4.5 mins	
Final extension	72	10 mins	1
Cooling	4	Hold	1

Two different types of competent *E. coli* cells were used to either propagate plasmid DNA or to overexpress the protein of interest. In order to obtain high quality plasmid DNA for sequencing purposes, SOX mutant plasmids were amplified in the NEB 5- α *E. coli* strain. After transformation, they were plated onto solidified agar media containing 25 $\mu\text{g}/\text{mL}$ of kanamycin for selection and were incubated at 37 °C overnight to produce colonies similar to WT SOX. Individual colonies were picked and inoculated into 5 mL of LB media containing 25 $\mu\text{g}/\text{mL}$ of kanamycin. After overnight incubation at 37 °C, the cultures were pelleted and the plasmids purified using the Wizard® Plus SV Miniprep Start-Up Kit (PROMEGA). 30 μL of 100 $\text{ng}/\mu\text{L}$ of the purified plasmid samples were sent to GATC (Germany) for sequencing to verify whether mutagenesis had been successful. The Expasy Translate tool (Gasteiger et al., 2003) was used to translate the obtained sequences and the mutants were verified using BLAST2seq (NCBI).

3.2.2 Expression and Purification of SOX mutants

The SOX mutant proteins were expressed and purified using a similar protocol as for WTSOX with the exception that the concentrated SOX mutants were applied to a gel filtration column (Superdex 200 HR 16/60), which had been previously equilibrated with gel filtration buffer (32 mM Tris-Base, pH 8.3, 189 mM NaCl, 1.6 µg bovine serum albumin (BSA), 10 mM DTT). The eluted fractions were analyzed using 10 % SDS-PAGE and the purest SOX mutant proteins were then concentrated to 8 mg/mL, separated into 50 µL aliquots and stored at – 80 °C.

3.3 Protein analysis and quantification

3.3.1 Protein concentration

All purified protein concentrations were derived from the average of 3 absorbance readings at 280 nm (A_{280}) using a Nano Drop spectrophotometer (Thermo Scientific). They were determined using the rearranged formulae from Beer and Lambert's law (Equation 3.1).

Equation. 3.1:

$$c = \frac{A_{280} \times MW}{\epsilon \times l}$$

c = Concentration of the protein (mg/mL)

A_{280} = Absorbance at 280 nm

ϵ = molar extinction coefficient ($M^{-1}cm^{-1}$)

MW = molecular weight (g/mol)

l = Path length (cm)

The molecular weight and molar extinction coefficient for protein concentration calculation were estimated using web resources such as ProtParam (web.expasy.org/protparam/).

3.3.2 Sodium dodecyl sulfate (SDS)-Polyacrylamide gel electrophoresis (PAGE)

Protein samples (10 μ L) were first mixed with 3 μ L of 4 X LDS Sample Buffer (NuPAGE, Invitrogen). They were then resolved on 4-12 % Bis-Tris 12-well Mini-Gels (NuPAGE, Invitrogen). 3 μ L of 1X Blue Prestained Protein marker molecular weight markers (Invitrogen) were used for analysis. SDS-PAGE was performed in 1 X MOPS buffer (NuPAGE, Novex) for 40 minutes, at 200 V, and the gels were stained with InstantBlue stain (Expedeon).

3.3.3 Western Blots

SDS-PAGE gels were prepared as previously described with the omission of staining with instant blue. Following electrophoresis, the gels were washed with Milli-Q water (Milli-Q[®]) and the proteins transferred onto a nitrocellulose membrane using the Invitrogen iBlotR 7-Minute Blotting System. The membrane was blocked using 1 x TBST buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1 % Tween20) blocking solution containing 5 % non-fat milk powder for 1 hour at room temperature, followed by three washes with 1 x TBST buffer (5 min each). Next, 1 x TBST

containing 5 % non-fat milk powder and 10 μ L of Anti-6His antibody was added to the membrane to give the specified 1:1000 dilution ratio and left shaking for 1 hour at room temperature or overnight at 4 °C. The membrane was subsequently washed 5 times with 1x TBST buffer (5 minutes for each wash) at room temperature. It was then incubated with 1 x TBST buffer containing 5 % non-fat milk powder, and Anti Mouse IgG (Fab)-Peroxidase Alkaline Phosphatase (SIGMA-ALDRICH) and incubated for 1 hour at room temperature. The membrane was next washed with 1x TBST buffer 5 times (5 minutes for each wash) at room temperature. 5 mL of distilled water containing 3,3'-Diaminobenzidine (DAB) tablet (SIGMAFAST) and a precipitating substrate Urea Hydrogen Peroxidase (Sigma-Aldrich) were then added to the membrane for the detection of 6His-tagged proteins.

3.4 Preparation and Annealing of RNA substrates

The 31 nucleotide truncated form of KSHV miRNA K2-31 (RNA 31nt, Table3.3) was purchased from Eurogentec and annealed to give a final concentration of RNA in excess of SOX by a factor of 1.2 for crystallization. To achieve this, the RNA was annealed in 50 μ L aliquots containing 10 mM Tris-HCL pH 7.6, 50 mM NaCl, 1 mM EDTA and RNase and DNase free PCR water, using a PCR machine (Peqlab Primus 96 Gradient PCR). For annealing, the sample was incubated at 90 °C for 1 min and the temperature was reduced by 1 °C every minute until a temperature of 4 °C was reached.

Similarly, GFP51, HBB58, GFP51-UCUCU, GFP51-UGCAC, K2-31A4, K2-31A9,

dsDNA5'P were annealed using the same PCR conditions as for K2-31 RNA. The various sequences are provided Table 3.3.

To avoid any potential exonucleolytic cleavage by WTSOX in assays probing endonucleolytic processing, the RNA constructs were designed not to be conjugated to any fluorescent tags or monophosphate groups at the 5' end. They were alternatively labelled with 3' 6-carboxyfluorescein (6-FAM).

Table. 3.3: RNA and DNA sequences.

Oligonucleotide	Sequences
GFP51	5'-UACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUG-3'FAM
HBB58	5'AGGUGAAGGCUCAUGGCAAGAAAGUGCUCGGUGCCUUUAGUGAUGGCCUGGCUCACC U-3'FAM
GFP51-UCUCU	5'-UACGGCAAGCUGACCCUCUCUUUCAUCUGCACCACCGGCAAGCUGCCCGUG-3'FAM
GFP51-UGCAC	5'-UACGGCAAGCUGACCCUGCACUUCAUCUGGUGCACCACCGGCAAGCUGCCCGUG-3'FAM
K2-31	5'-GAUCUGAGCCAUUGAAGCAAGCUUCCAGAUC-3'FAM
K2-31A4	5'-GAUCUGAGCCAUUGAAGCAAAAAGCUUCCAGAUC-3'FAM
K2-31A9	5'-GAUCUGAGCCAUUGAAGCAAAAAAAAAAAGCUUCCAGAUC-3'FAM
UN51	5'-GGCCAUCCUGUUUUUUUCCCUUUUUUUUUUUCUUUUUUUUUUUUUUUUUUU- 3'FAM
dsUN51	5'-GGCCAUCCUGUUUUUUUCCCUUUUUUUUUUUCUUUUUUUUUUUUUUUUUUU- 3'FAM 3'-CCGGUAGGUCAAAAAAAGGGAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAA-5'
dsDNA5'P	5'-pGGGGATCCTCCCAGTCGACC-3' 3'FAM-CCCCTAGGAGGATCAGCTGG-5'

3.5 RNase assays using TBE-Urea gel electrophoresis

The ability of WTSOX and SOX mutants to bind and degrade RNA was assessed via RNase assays using TBE-Urea gel Electrophoresis. This assay was adapted from that reported in *Bagneris et al., 2012* given that the binding conditions originally used (1.35 μ M SOX, 0.2 μ M RNA, 25 mM Tris-HCL pH 9.0, 200 mM NaCl, 10 mM MgCl₂ and 5 mM β -mercaptoethanol) resulted in poor endonucleolytic cleavage. Therefore, optimal processing conditions had to be determined using a reaction mix containing 0.2 μ M RNA combined with 1.35 μ M SOX in a buffer comprised of 10 mM MgCl₂, 5 mM BME, 25 mM Tris-HCl at a range of pHs between 6.0 to 9.0 and NaCl (50 to 200 mM) concentration. Maximal endonucleolytic processing was obtained at a pH of 9.0 and NaCl concentration of 50 mM. All RNA endonuclease assays were therefore performed in a buffer comprising 10 mM MgCl₂, 5 mM BME, 25 mM Tris-HCl pH 9.0 and 50mM NaCl. The reaction mixes comprising optimal buffer, 0.2 μ M RNA and 1.35 μ M SOX were then incubated for 1 hour at 37 °C. Time course assays were also conducted in order to compare the rates of RNA processing more quantitatively where samples were taken at 20 minute intervals for one hour. The endonuclease reactions were halted by the addition of 7.5 μ L Novex TBE (Tris-borate, EDTA)-urea sample buffer (Invitrogen) to 7.5 μ L of reaction mixture. The samples were then heated for 3 mins at 70 °C and loaded onto 15 % TBE-urea gels (either pre-cast (Invitrogen) or generated in-house (Table 3.4)). The samples were electrophoretically separated using a PowerPac™ power supply (BioRad) using a 1 X TBE running buffer for 100 minutes at 100 V. The gels were visualized using a FLA3000 transilluminator (FujiFilm) Imager, at an excitation wavelength of 473 nm

and emission wavelength of 520 nm consistent with 6-FAM.

Table 3.4: Composition of in house 15 % Urea gels:

Name	Volume
40 X Acrylamide	5.625 mL
5 x TBE	3.0 mL
Urea	6.3 g
40 % Ammonium Persulfate (APS)	30 μ L
Tetramethylethylenediamine (TEMED)	12 μ L
Distilled water (dH ₂ O)	Up to 15 mL

3.6 Fluorescence polarization anisotropy (FPA)

Fluorescence polarisation anisotropy assays were conducted using 1 μ M concentrations of oligonucleotides that were tagged with 3'FAM in serial dilutions of SOX (75 μ M, 50 μ M, 33.3 μ M, 22.2 μ M, 14.8 μ M, 9.9 μ M, 6.6 μ M, 4.4 μ M, 2.9 μ M, 1.9 μ M, 1.3 μ M, 0.87 μ M). Experiments were performed in a buffer comprising 25 mM Tris-HCl, pH 8.5, 300 mM NaCl and 10 % glycerol to give a final volume of 60 μ L. Each sample was incubated for 30 mins and its anisotropy measured using a fluoromax-3 spectrofluorimeter (Jobin Yvon Horiba) where the assay was conducted at 25 °C with a slit width of 5 nm. An excitation wavelength of 492 nm was used that gave rise to an emission wavelength of 515 nm. The measurements were repeated 10 times at each concentration of SOX and the affinity constants determined following data processing in Graphpad PRISM6 where all curves were fitted to a single site binding equation.

3.7 Crystallization of SOX-inhibitor complexes

3.7.1 Formation of the SOX-inhibitor complexes

Two SOX inhibitors, Pyranine and Phytic acid, were investigated for their capacity to form complexes with SOX. Since both the inhibitors are highly acidic (pH= ~5.0), they were diluted to maintain a pH of ~7.5. SOX (8 mg/mL) and was combined with each inhibitor at a concentration 4 mM. The samples were subsequently incubated on ice for 3 hours prior to crystallisation.

3.7.2 Initial Crystallization trials of SOX:Phytic acid and SOX:Pyranine complexes.

Both SOX:Phytic acid and SOX:Pyranine complexes were initially screened using the commercially available Proplex HT – 96 (Molecular Dimensions), JCSG – plus HT – 96 (Molecular Dimensions), and Structure Screen 1 & 2 (30 Molecular Dimensions) kits. 80 μ L of each condition was pipetted by hand into the “mother liquor” well of 96 well MRC 2 sitting drop vapour diffusion crystallization 96 well plates (Swissci) using a multi-channel pipette. A Mosquito Crystal Liquid handling robot (TTP Labtech) was used to combine the sample (SOX complexes) and mother liquor (precipitant) in 1 (250 nL) : 1 (250 nL) and 1 (250 nL) : 2 (500 nL) (volume/volume) drop ratios and dispense sitting drops. The screens were then incubated at 20 °C.

3.8 ThermoFluor Assay

In this assay, a 25 μ L reaction mix was assembled in a 96 well plate format comprising WTSOX (1 mg/mL), 1 mM Phytic acid and 1 μ L of SYPRO Orange (1:1000 dilution). Each assay was repeated 3 times and performed using the MyiQ Real Time PCR System (Bio-Rad-PCR Machine). The starting temperature for each run was 10 °C increasing to 95 °C in 0.5 steps. Trials were also performed in the absence of phytic acid for comparison along with buffer controls lacking WTSOX.

3.9 Cloning, expression and purification of p100 and IKK α

In order to investigate the interaction between p100 and IKK α in the non-canonical NF- κ B pathway, full length p100 and IKK α constructs were expressed using a baculovirus system.

For cloning, the polymerase chain reaction (PCR) was used to amplify the full length open reading frames of both p100 and IKK α from existing plasmids. The appropriate forward and reverse primers for Gibson assembly were designed using the NEBuilder Assembly Tool (BioLabs, www.nebuilder.neb.com) and were purchased from SIGMA ALDRICH. They are listed in Table 3.5.

Table 3.5. Oligonucleotide DNA primers used to generate the p100 and IKK α full length constructs by Gibson Assembly.

Name	Primer sequence (5' → 3')
p100 Fwd	TCAGGGCGCCATGGATCCGGAATTCATGGAGAGTTGCTACAACCC
p100 Rev	TCCTCTAGTACTTCTCGACAAGCTTCAGTGCACCTGAGGCTG
Ikkα Fwd	TCAGGGCGCCATGGATCCGGAATTCATGGAGCGGCCCGGGG
Ikkα Rev	TCCTCTAGTACTTCTCGACAAGCTTCATTCTGTAAACCAACTCCAATCAAGATTC

Each PCR reaction was prepared according to the manufacturer using KAPA HiFi Hotstart ready mix (KAPA Biosystems) and were made up to 25 μ L using PCR grade water (Milli-Q[®]) (Table 3.6).

Table 3.6 PCR Reaction composition.

Reagent	Final Concentration
Template DNA	1 ng/ μ L
Forward primer	0.4 μ M
Reverse primer	0.4 μ M
2 X KAPA HiFi Hotstart ready mix (KAPA Biosystems)	1 X
PCR grade water (Millipore)	Up to 25 μ L

PCR reactions were carried out using the Peqlab Primus 96 Gradient PCR-machine with the program provided in Table 3.7.

Table 3.7. The program used for PCR.

Step	Temperature (°C)	Time	Number of cycles
Lid heating and Initial denaturation	96	5 min	1
Denaturation	96	1 min	25
Annealing	Within 5 °C of the primer melting temperature	15 s	
Extension	72	30-60 seconds per kilobase of DNA being amplified	
Final extension	72	10 min	1
Cooling	4	Hold	1

Successful amplification was confirmed using agarose gel electrophoresis. 50 mL gels were prepared using 1 % agarose and TAE buffer (40 mM Tris-base, 20mM acetic acid, and 1 mM EDTA, pH 8.0) containing 1 µL of SYBR Safe (Invitrogen) DNA stain. 10 µL of Quick-Load DNA Ladder (NewEngland BioLabs) was used for quantification and analysis. 25 µL of DNA sample were mixed with 2.5 µL of Gel Loading Dye Purple (NewEngland BioLabs) prior to gel loading. Gels were run in 1 X TAE buffer for 1 hour at 80 V. Successful amplification was confirmed using a UV transilluminator (Fisher Scientific). In order to clone p100 and IKK α into the pFASTBAC vector, the HindIII and EcoRI restriction sites were used. To achieve this, restriction digests were initially performed to produce linearized vector. The composition of the digests is provided in Table 3.8. The digestion mixture was

incubated at 37 °C in a water bath for 2 hours.

Table 3.8 Composition of Restriction Digest

P100 and IKKα	
Name	Volume (Final Concentration)
pFASTBAC	15 μ L (3 μ g)
Hind III	4 μ L
EcoRI	4 μ L
Buffer (10X)	5 μ L
ddH₂O	Up to 50 μ L

Agarose gel electrophoresis was carried out to verify digestion of the vector and the products were purified using a DNA extraction kit (QIAGEN).

The purified p100 and IKK α PCR products were then inserted into purified, linearized pFASTBAC vector using Gibson assembly ligation. This was performed following the manufacturer's instructions where the composition of each reaction is listed in Table 3.9.

Table 3.9 Composition of Gibson assembly reaction mixture.

Name	Volume (Final concentration)
Digested pFASTBAC vector	1 μ L of 50 ng
Fragments	2 μ L of 25 ng
Gibson Assembly Master Mix (2X)	10 μ L
dH₂O	Up to 20 μ L

The mixture was subsequently incubated in a thermocycler (Peqlab Primus 96 Gradient PCR-machine) at 50 °C for 15 minutes. 2 µL of the Gibson assembly mixture were then transferred into 50 µL of NEB competent cells (New England Biolabs) and incubated on ice for 30 minutes. The cells were then heat shocked at 42 °C for 45 seconds followed by incubation on ice for 2 minutes. 250 µL of SOC media (kept at room temperature) was added to the mixture which was incubated at 37 °C for 1 hour in a shaker incubator (shaken at 250 rpm) to aid recovery of the cells. 100 µL of this mixture was spread onto agar plates containing 100 µg/mL ampicillin and incubated at 37 °C overnight.

Plasmids were amplified from three colonies for both p100 and IKK α . Following selection, each was inoculated into 5 mL of LB media containing 100 µg/mL ampicillin and grown overnight at 37 °C. The plasmids were then purified using a Wizard® Plus SV Miniprep DNA purification kit (Promega). 30 µL of 70 - 100 ng/µL of the purified plasmids were sent to GATC (Germany) for sequencing. The ExPASy Translate tool (Gasteiger et al., 2003) was used to translate the obtained sequences that were verified as correct using BLAST2seq (NCBI).

3.10 p100 and IKK α Expression.

3.10.1 Transposition of pFASTBAC-p100 and pFASTBAC-IKK α into bacmids.

1 μ L of pFASTBAC-p100 or pFASTBAC-IKK α , was transformed into 50 μ L of electro competent DH10 EMBacY cells (Thermo Scientific). The cell/DNA mix was transferred into a chilled electroporation cuvette (1 mm) and was electroporated using the Bio-Rad Gene Pulser Xcell™. 950 μ L of SOC media was then added directly to the cuvette and the suspension transferred to a 17mm x 100 mm round-bottom culture tube which was incubated at 37 °C for 6 hours with shaking at 250 rpm. 50 μ L of this suspension was plated on LB agar plates containing Kanamycin (for DH10Bac selection), Gentamycin (for DH10Bac selection), Tetracycline (from helper plasmid selection antibiotic), IPTG and X-Gal for blue/white colony selection. The plate was incubated for 2 days at 37 °C. White colonies indicated that the transfer vector was successfully transposed into the bacmid. Single white colonies were then inoculated into 3 mL of LB media containing Kanamycin, Gentamycin and Tetracycline and grown overnight at 37 °C in 50 mL tubes with shaking at 200 rpm. The bacmids were then purified using a DNA purification kit (Qiagen). The DNA concentration was determined using a NanoDrop spectrophotometer (ND 1000, Thermo Scientific).

3.10.2 Transfection of Insect cells and protein overexpression

Initially, *Spodoptera frugiperda* (sf9) cells (Gibco) were counted using an EVOS light cube (Thermo Fisher Scientific) in a Countess II FL Automated Cell Counter (Invitrogen). The sf9 cells and the Insect-XPRESS Medium + L-glutamine (Lonza) were dispensed into a 6-well plate to give 1 million cells per well, in a total volume of 3 mL. The plate was incubated at 27 °C for 30 minutes to allow the cells to adhere. 95 µL of insect cell medium, 2 µg of bacmid DNA and 3 µL of FuGene HD transfection reagent were gently mixed in a 1.5 mL tube and incubated at 27 °C for 15 minutes to allow the bacmid DNA to form a complex with the FuGene. 100 µL of the transfection complex was added drop-wise into each well of a 6-well plate. The plate was then incubated at 27 °C for 3 days in a plastic bag containing damp tissue to maintain humidity. After 3 days, the supernatant (containing the baculovirus) was transferred into a tube containing 50 mL of sf9 cell (at 1 million cells per mL). This viral suspension (termed V0) was incubated at 27 °C for 3 days with shaking at 100 rpm. 1 million cells were then aliquoted into a 1.5 mL tube for protein expression analysis using SDS PAGE and the remaining cell suspension transferred into a 50 mL tube that was centrifuged at 2000 rpm for 5 minutes. The supernatant containing the baculovirus (now called V1) was transferred into a dark-walled 50 mL tube and stored at 4 °C. The remaining pellet was then used for protein expression analysis using SDS PAGE. For large scale expression, 2 L of insect cell media were infected with 1 % V1 (v/v). This was then allowed to incubate for 3 days at 27 °C. The cells were then centrifuged at 2000 rpm for 20 minutes and the pellets stored at -80 °C.

3.10.3 Expression analysis using SDS-PAGE

In order to assess protein expression, 1 million cells per mL of sf9 cells were aliquoted into a 1.5 mL tube and analysed by SDS-PAGE using 1 million uninfected sf9 cells as a control. Prior to SDS-PAGE, these cells were centrifuged at 13,000 rpm for 1 minute and the resultant pellets resuspended in 500 μ L of 1 X PBS buffer. 5 μ L of 4X SDS loading buffer were added to 15 μ L of cell suspension and incubated at 95 °C for 5 minutes. 2.5 μ L, 5 μ L and 10 μ L aliquots of this mix were loaded onto SDS-PAGE gels.

3.11 Purification of p100 and IKK α

Unless otherwise stated, all purification protocols involving columns were conducted using an Akta Explorer purifier (GE Healthcare).

p100

Cell pellets were resuspended in 20 mL of lysis buffer composed of Buffer A (Table 3.10) with an added EDTA free protease inhibitor cocktail tablet (Roche), using a handheld homogeniser. Lysed cells were pelleted by centrifugation (20,000 rpm for 50 minutes at 4 °C). The supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with Buffer A using a peristaltic pump (~1 mL/min, 4 °C). Once loaded, the HisTrap column was washed with 10 column volumes (CV) of Buffer A (Table 3.10) supplemented with 5 mM imidazole to remove non-specifically bound proteins. p100 was eluted with buffer B (Table 3.10) using a

gradient of 5 – 500 mM imidazole over 20 CV collected in 1 mL fractions. Protein purity was determined using SDS-PAGE and the fractions containing P100 at a purity of greater than 95 % were combined. The sample was then digested with TEV protease overnight in dialysis Buffer (25mM Tris-Base, 300 mM NaCl, and 0.5 mM DTT, pH 8.5) at 4 °C to remove the 6His-tag before concentration using a Vivaspin 10 kDa MWCO centrifugal concentrator (GE Healthcare) to a volume of approximately 2 mL. The concentrated sample was then applied to a Superdex™ S200 16/60 gel filtration column which had been pre-equilibrated with Buffer C (Table 3.10), and 0.5 mL fractions collected. All peak fractions were analysed by SDS-PAGE and those containing p100 were pooled and concentrated to 10 mg/mL in a Vivaspin 10 kDa MWCO centrifugal concentrator (GE Healthcare).

Table 3.10. Buffer compositions used for p100 purification.

Name	Composition
Buffer A	20 mM Tris, pH 7.5, 250 mM NaCl, 10 % Glycerol, 7 mM 2-mercaptoethanol (BME)
Buffer B	20 mM Tris, pH 7.5, 250 mM NaCl, 500 mM imidazole, 10 % Glycerol, 7 mM BME
Buffer C	20 mM Tris, pH 7.5, 250 mM NaCl, 10 % Glycerol

IKK α

Cell pellets were resuspended in 20 mL of lysis buffer comprising Buffer A (Table 3.11), with an added EDTA free protease inhibitor cocktail tablet (Roche), and 0.2 % NP-40 (non-ionic polyoxyethylene surfactant). Lysed cells were pelleted by centrifugation (14,000 rpm for 45 minutes at 4 °C). All purification steps for IKK α were carried out at 4 °C owing to the lability of the protein. First, the supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with Buffer A using a peristaltic pump (1 mL/min, 4 °C). Once loaded, the HisTrap column was washed for 10 CV with Buffer A supplemented with 30 mM imidazole to remove non-specifically bound proteins. IKK α eluted using buffer B (Table 3.11) in a gradient of 30 – 250 mM imidazole over 20 CV, where samples were collected in 1 mL fractions. Protein purity was determined using SDS-PAGE and the fractions containing the purest IKK α combined. The N-terminal 6His-tag was left intact for pull down assays. In order to remove imidazole, the collected samples were dialysed in Buffer C (Table 3.11) overnight at 4 °C.

Table 3.11. Buffers used for IKK α purification.

Name	Composition
Buffer A	25 mM Tris-Base, pH 8.0, 200 mM NaCl, 10 % Glycerol, 5 mM BME.
Buffer B	25 mM Tris-Base, pH 8.0, 200 mM NaCl, 250 mM imidazole 10 % Glycerol, 5 mM BME.
Buffer C	25 mM Tris, pH 8.0, 200 mM NaCl, 10 % Glycerol, 5 mM BME.

3.12 Expression and Purification of p100 short c-terminal constructs

3.12.1 Expression of pETM6T1-p100 746-860 and pETM6T1-P100 746-848

Two short constructs of p100: pETM6T1-p100 746-860 and pETM6T1-p100 746-848 were provided by Dr. Tracey Barrett. Both of these plasmids were transformed into Rosetta™ 2(DE3) cells (Novagen) as described in section 3.2.1 using Chloramphenicol and Kanamycin for antibiotic selection. These plasmids contained an N-terminal His-tag, followed by a TEV cleavage site and an N-utilization substance protein A (NusA) tag for enhanced solubilization.

Following growth on agar, colonies containing the pETM6T1-p100 746-860 and pETM6T1-p100 746-848 plasmids were selected and cultured in 100 mL of Lysogeny broth (LB, 10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride at pH 7.5) containing 34 µg/mL Chloramphenicol and 25 µg/mL Kanamycin and were grown overnight at 37 °C with shaking (225 rpm, Certomat BS1 incubator) to produce seed cultures. 10 mL of these cultures were transferred into each of 5 x 2 L flasks of LB media containing 34 µg/mL Chloramphenicol and 25 µg/mL Kanamycin to give a 1: 100 ratio. These expression cultures were then incubated at 37 °C and agitated at 225 rpm to an optical density (OD₆₀₀) of 0.8 – 0.9. At this stage, 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma) was added to each flask to induce the expression of pETM6T1-P100 746-860 and pETM6T1-P100 746-848. Once added, the cultures were incubated at 18 °C overnight and the cells harvested by centrifugation for 20 mins at 5000 x g (4 °C Beckman Avanti J – 20 I Centrifuge).

The resulting pellets were transferred into 50 mL centrifuge tubes and stored at -80 °C.

3.12.2 Purification of pETM6T1-P100 746-860 and pETM6T1-P100 746-848

Both pETM6T1-p100 746-860 and pETM6T1-p100 746-848 were purified using the same protocol. Cell pellets were first re-suspended in 50 mL of His Trap buffer A (25 mM Tris-HCl pH 8.5, 300 mM NaCl), supplemented with an EDTA- free protease cocktail inhibitor tablet (Roche) and 10 µg/mL DNase I (NEB). The cells were lysed using a sonicator (Vibra Cell) at an amplitude of 40 Watts for 1 min interspersed with 30 s rest intervals for a total of 5 mins on ice. The lysate was then centrifuged at 20,000 rpm at 4 °C for 45 mins (Beckman Avanti J -20 XP rotor) to separate the soluble fraction from the cell debris and then passed through a 0.45 µm syringe filter (Sartorius Biotech) to remove large particulates. The supernatant (soluble protein) was then applied to a 5 mL HisTrap FF column (GE Health Care), which had been previously equilibrated with HisTrap column buffer A (25 mM Tris-Base, pH 8.5, 300 mM NaCl). The column was then washed with 20 CV of His Trap column buffer A. Protein samples were then eluted with 60 CVs of HisTrap column buffer B (25 mM Tris-Base, pH 8.5, 500 mM imidazole, 300 mM NaCl) in a gradient of 30 – 500 mM imidazole, where samples were collected in 1 mL fractions. The eluted fractions that showed the highest concentrations of protein were confirmed as p100 short constructs using SDS-PAGE. The fractions containing p100 short constructs were

collected and the NusA-tag removed by TEV cleavage. 2 mg of TEV protease was added to the p100 samples in a Pierce Snakeskin membrane (ThermoFisher) and were dialyzed overnight at 4 °C into 4 L of dialysis buffer (25 mM Tris-Base, pH 8.5, 200 mM NaCl, 1 mM dithiothreitol (DTT)) to remove the 6His-NUSA tag. Following dialysis, the samples were diluted to reduce the NaCl concentration to 50 mM using dilution buffer (25 mM Tris-Base, pH 8.5). The samples were then applied to a 10 mL Hi Trap Q HP column (GE Health Care), which had been pre-equilibrated with Q column buffer A (25 mM Tris-Base, pH 8.5, 50 mM NaCl). The bound p100 constructs were then eluted with Q column buffer B (25 mM Tris-Base, pH 8.5, 1M NaCl) using a gradient of 5 % to 100 % over 30 CV. The protein fractions were analyzed by SDS-PAGE and those containing the purest proteins collected and concentrated to 1 mL using a Vivaspin (3 kDa MWCO) centrifugal concentrator (Sartorius Biotech). The concentrated protein samples were next applied to a gel filtration column (Superdex S75 HR 16/60), which had been previously equilibrated with gel filtration buffer (25 mM Tris-HCl, pH 8.5, 250 mM NaCl). The eluted fractions were analyzed using SDS-PAGE and those containing proteins with a purity of greater than 98 % concentrated to 4 mg/mL using a Vivaspin (3 kDa MWCO) centrifugal concentrator (Sartorius Biotech). Concentrated samples were then separated into 50 μ L aliquots, flash frozen in liquid nitrogen and stored at – 80 °C.

3.12.3 Expression and Purification of p100 746-900

The p100 746-900 protein was expressed and purified using a similar protocol as for the p100 short constructs. The eluted fractions were analyzed in 10 % SDS-PAGE and the purest p100 746-900 proteins were stored at – 80 °C.

3.13 Expression and Purification of KSHV-GB1-vFLIP

The pET22b-GB1-vFLIP-1-188 construct was supplied by Dr Tracey Barrett. pET22b-GB1-vFLIP-1-188 was transformed into BL21(DE3) cells (Novagen) using 100 mg/mL ampicillin (Melford) as the selection antibiotic. Colonies from successful transformations were used to inoculate 100 mL seed culture comprised of LB media containing 100 mg/mL ampicillin (Melford). The seed cultures were grown overnight at 37 °C with shaking at 220 rpm. 10 mL of these cultures were then used to inoculate 2 X 2 L flasks of LB broth containing 100 mg/mL ampicillin and when an OD₆₀₀ of 1.2 was reached, the cells were induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1mM. The cells were then grown for 18 hrs at 25 °C. This was followed by harvesting via centrifugation at 5000 rpm for 15 mins at 4 °C and storage at – 80 °C.

For protein production, cell pellets were resuspended in 50 mL of lysis buffer which constituted Buffer A (25 mM Tris-Base, pH 8.5, 200 mM NaCl) with an added EDTA free protease inhibitor cocktail tablet (Roche) and DNase I (10 µg/mL final concentration; Roche). Resuspended cells were disrupted using a sonicator (Virba Cell) at an amplitude of 40 Watts, interspersed with rest intervals of 30 s for a total

of 2 mins and kept on ice throughout. The soluble lysate was separated by centrifugation at 20000 rpm for 30 mins at 4 °C and the supernatant filtered through a 0.45 µm syringe filter (Sartorius Biotech).

For GB1-vFLIP, a two-step purification strategy was employed that involved affinity and size exclusion chromatography. Two 5 mL HisTrap™ HP columns (GE Healthcare) were connected in series and loaded with the clarified lysates using a peristaltic pump, following pre-equilibration with Buffer A (25 mM Tris-Base, pH 8.5, 200 mM NaCl). This was then washed with 15 CV of buffer A supplemented with 25 mM imidazole to remove any residually bound *E.coli* proteins and GB1-vFLIP eluted with a 10 CV step gradient to 100 % Buffer B (25 mM Tris-Base, pH 8.5, 200 mM NaCl, 1 M imidazole). The eluted fractions were analysed using SDS-PAGE. Those containing GB1-vFLIP were collected and in order to prevent aggregation, buffer exchanged into Buffer C (25 mM Tris-HCl, pH 8.5, 200 mM NaCl, 250 mM imidazole and 5 mM DTT) via gel filtration (Superdex™ S75 16/60 pre-equilibrated Buffer C). Fractions containing GB1-vFLIP were collected and incubated with prescission protease at room temperature for 3 hrs for removal of the GB1 tag. The digested sample was next loaded onto a Superdex™ S75 26/60 gel filtration column which was pre-equilibrated in gel filtration buffer (25 mM Tris-HCl, pH 8.5, 200 mM NaCl). All protein containing fractions were analysed by SDS-PAGE and those containing vFLIP pooled and stored at 4 °C.

3.14 Expression and Purification of untagged KSHV-vFLIP

Untagged KSHV- vFLIP was produced as described in *Bagneris et al, 2008*.

3.15 Pull Down Assays

For pull down analysis, 200 μ L of Ni-NTA resin (ThermoFisher Scientific) was first washed with 3 X 1 mL of distilled water followed by 3 X 1 mL Buffer A (25 mM Tris-Base, pH 8.5, 250 mM NaCl). The washed Ni-NTA resin was combined with the protein sample and incubated for 1 hour at room temperature or 4 °C to allow binding to the resin. The mixture was then applied to a gravity flow column (Bio-Rad) and washed with 3 CVs of Buffer A, 3 CVs of Buffer B (25 mM Tris-Base, pH 8.5, 250 mM NaCl, 30 mM Imidazole) and the bound proteins eluted with 3 CVs of Elution buffer (25 mM Tris-Base, pH 8.5, 250 mM NaCl, 500 mM Imidazole). Each fraction was collected in separate 1.5 mL tube. The samples from the flow-through, wash, beads and 3 X elution fractions were also collected for analysis by SDS-PAGE.

Chapter 4: Structural and Functional Analysis of the interaction between KSHV-SOX and the pre-micro RNA fragment K2-31

4.1 SOX-RNA processing targets stem loop or bulge motifs

Following *in silico* experiments as described in section 1.7, and the identification of potential SOX targets involving fragments of GFP51, DsRed61 and HBB58 transcripts, none could be successfully co-crystallised with SOX. This led to alternative sequences from the KSHV transcriptome (owing to its smaller size and susceptibility to SOX cleavage published in the literature) were computationally analysed, in particular, those containing UGAAG motifs in the vicinity of a bulge/stem loop region similar to those in the DsRed2, GFP and HBB. From these studies, the SOX pre-microRNA K2-31 was identified by Anathe Patschull as a possible candidate and was subsequently shown to be susceptible to SOX mediated cleavage when reduced to a 31mer fragment (Figure 4.1). As part of a subsequent Masters project (Lee, 2015), this was successfully co-crystallised with SOX and data collected to 3.3 Å at the Diamond synchrotron source (station i24). This was the starting point for the structural studies described in this chapter.

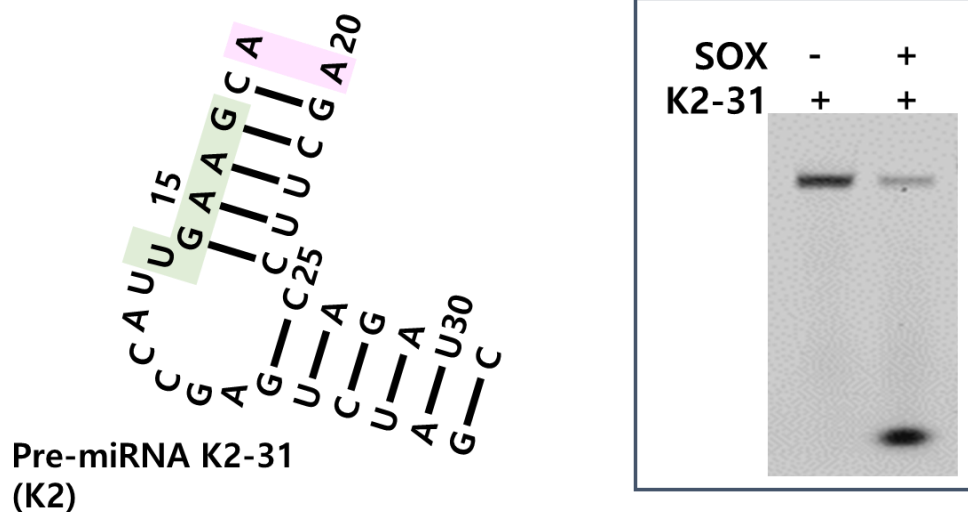


Figure 4.1: Structure of K2-31 and the RNase assay of K2-31 treated with WT SOX. When added, the K2-31 oligonucleotide degraded (Figure adapted and modified from Lee et al., 2017).

4.2 Solving the SOXE-244S:K2-31 crystal structure by molecular replacement and Refinement

Following collection, the data were scaled, merged and truncated to produce structure factor amplitudes in the CCP4 program AIMLESS. The structure of the SOXE-244S:K2-31 complex was solved by molecular replacement using PHASER (McCoy et al., 2007) and the native KSHV-SOX co-ordinates (PDB code 3FHD) as a search model. Initial $2F_o - F_c$ and $F_o - F_c$ electron density maps confirmed the presence of RNA nucleotides within the catalytic region (Figure 4.2A). After initial molecular replacement, the programs Nautilus and REFMAC 5 were used to obtain an improved electron density map (Figure 4.2B). However, although this showed evidence of additional density it was weak and poorly ordered.

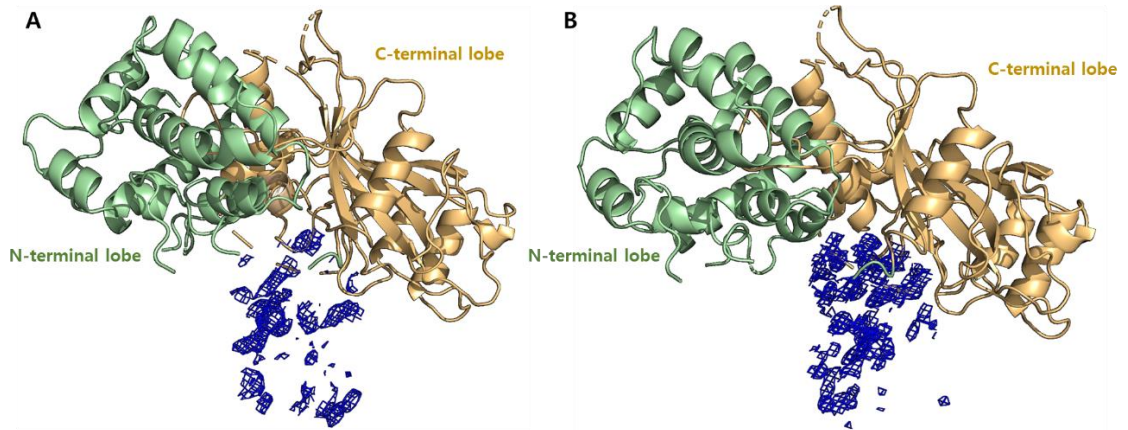


Figure 4.2: (A) Initial $2F_o-F_c$ (contoured at 1σ) electron density map of K2-31. (B) Improved $2F_o-F_c$ (contoured at 1σ) electron density map of K2-31 after several cycles of refinement using REFMAC 5 and Nautilus (CCP4).

To improve this, AUTOBUSTER was run in $-L$ mode for ligand location which is facilitated by placing water molecules in the F_o-F_c electron density and removing them in the final stages in the event that the density is large. This resulted in limited improvements to the extent that some nucleotides in the K2-31 loop could be fitted, but the remaining density was still poor. Several refinement cycles were subsequently performed using PHENIX but no further improvements were obtained. Re-evaluation of the original data, however, showed that it was severely anisotropic, i.e., diffraction was limited to only 4 \AA in one direction compared to $\sim 3.2 \text{ \AA}$ in the remaining two. It was therefore reprocessed to the highest resolution limit (3.2 \AA) and submitted to the anisotropy correction server STARANISO (<http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>) for structure factor amplitude correction. Using these modified structure factor amplitudes, molecular

replacement was repeated in PHASER using the SOX co-ordinates alone. The resulting density maps were significantly improved allowing placement of most of the nucleotides in the loop region (Figure 4.3).

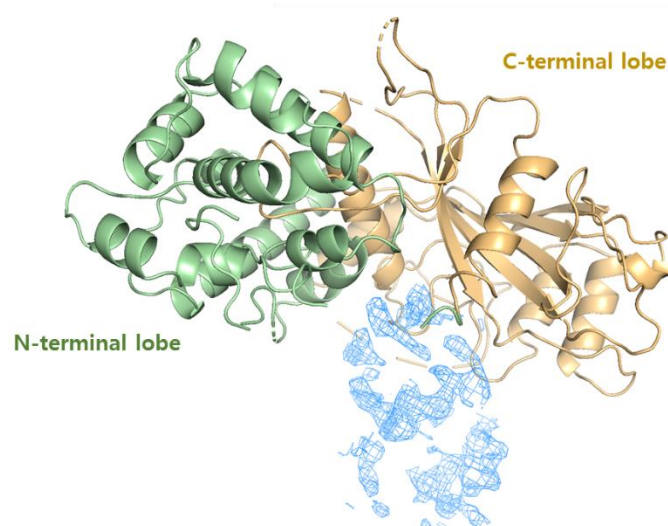


Figure 4.3: $2F_o - F_c$ (contoured at 1σ) electron density map of K2-31 after anisotropy correction and molecular replacement using PHASER. This significantly improved allowing placement of most of the backbone of nucleotides in the loop region.

Refinement using AUTOBUSTER enabled the placement of a substantial number of the missing nucleotides (20 out of the 31 bases could be unambiguously placed). Initial analysis indicated that K2-31 had been partially processed in the loop region of the stem loop given that nucleotide A19 appeared to be missing (Figure 4.4).

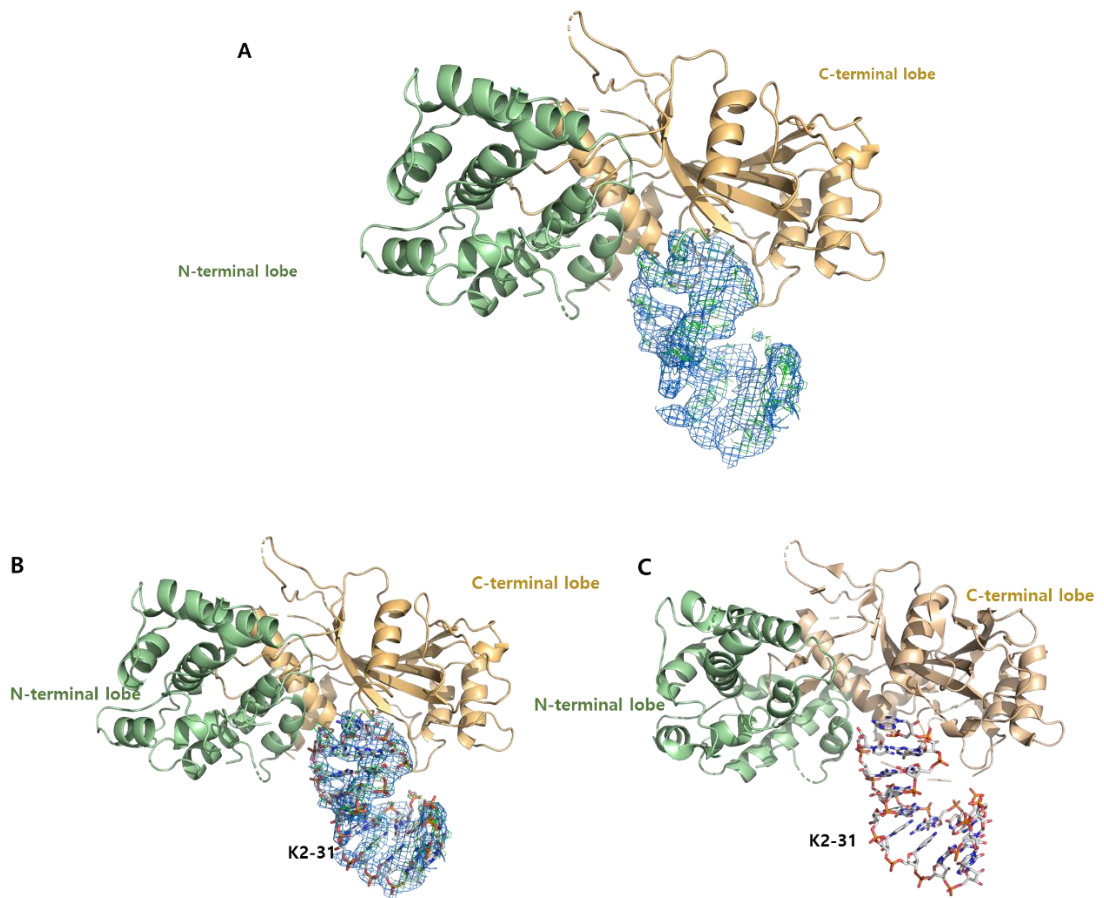


Figure 4.4: (A) Improved K2-31 density after several rounds of refinement. $2F_o-F_c$ map density (contoured at 1σ) and F_o-F_c map (contoured at 3σ). (B) The improved electron density superposed with the co-ordinates of the final SOX-K2-31 model. This significantly improved allowing placement of K2-31 nucleotides U3-G28. (C) Co-ordinates of the final SOX-K2-31 model.

Notably, the location of the RNA in the catalytic site between the N- and C-terminal lobes of SOX indicated that the complex was biologically significant. Nucleotides constituting the bulge directly 5' to the UGAAG site could not be observed most likely due to either disorder or enzymatic processing. The overall structure of K2-31 is

concomitant with the lowest energy structures predicted by MC-Fold and MC-Sym in which A19 and A20 present at the apex of the loop are unpaired with the majority of the remaining nucleotides stabilised by Watson Crick base pairing. The only deviation from the predicted structure is nucleotide C26 that fails to form a Watson Crick base pair with G6 (Figure 4.5).

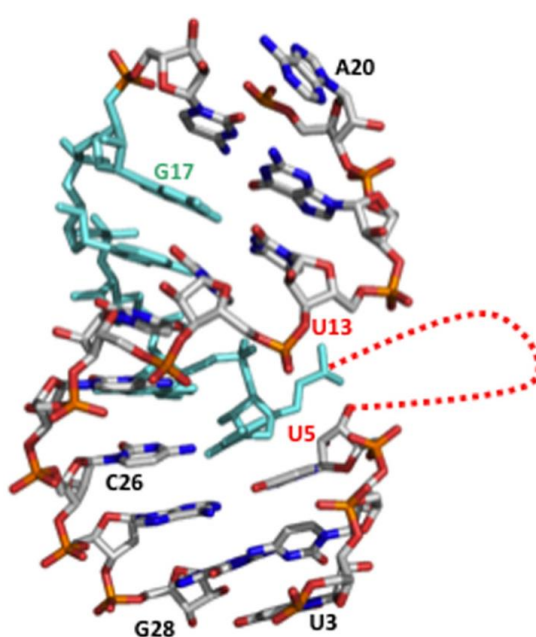


Figure 4.5: Schematic overview of K2-31 (nucleotides U3-G28). Due to the lack of density, the internal loop (3' U5 to 5' U13) consisting of 7 nucleotides could not be modelled (red dotted line). The UGAAG sequence is highlighted in cyan. (Adapted and modified from Lee et al., 2017).

Similar to the SOX-DNA complex, analysis of the SOX-E244S:K2-31 complex crystal contacts revealed that crystal lattice was also stabilised by RNA-RNA interactions

mediated by symmetry-related molecules involving end nucleotides. The nucleotides forming the crystal contacts are disordered (Figure 4.6).

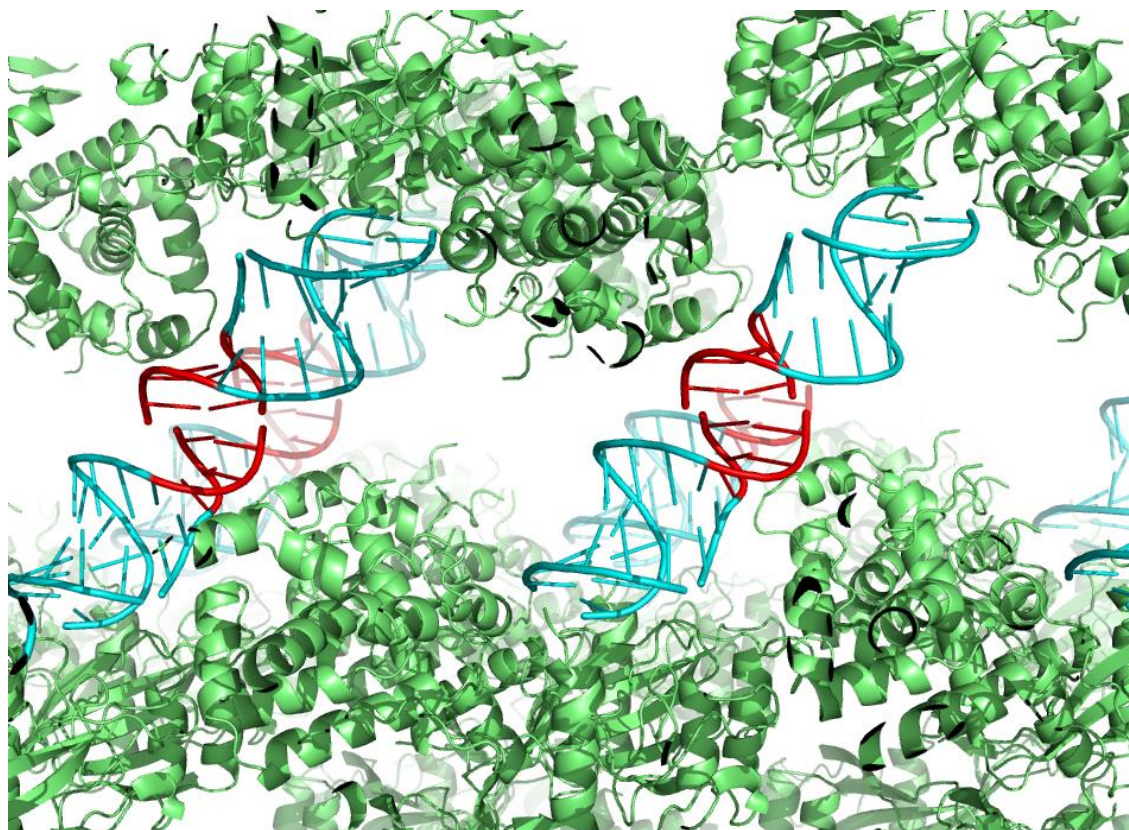


Figure 4.6: Crystal packing diagram illustrating that key to formation of the lattice are RNA-RNA interactions (cyan). Nucleotides highlighted in red forming the interface between symmetry related monomers are poorly ordered. Protomers are coloured green. (Adapted and modified from Lee et al., 2017).

The final model was refined to a R_{work} and R_{free} of 20.8 % and 26.3 % respectively with Table 4.1 summarising the data collection and refinement statistics.

Table 4.1: Data collection and refinement statistics for the final SOX-E244S:K2-31 model.

Crystal	SOX-E244S:RNA complex
Unit Cell	
a, b, c (Å)	198.5, 45.9, 67.6
α, β, γ (°)	90, 90.1, 90
Space Group	C2
Resolution (Å)	44-3.3 (3.3-3.7)
Number of Measured Reflections	30812
Number of Unique Reflections	9408
R _{merge} (Outer Shell)	0.19 (0.967)
Multiplicity	3.5 (4.1)
Completeness (%)	99.80 (99.8)
Mean I/ (Outer Shell)	6.4 (1.8)
Wavelength (Å)	0.9686
R _{pim}	0.19 (0.961)
CC _{1/2}	0.978 (0.673)
Refinement	
Number of atoms Protein, Nucleic acid, Solvent	3232, 380, 9
Nucleic acid atoms	380
Solvent atoms	9
R _{work} ^b /R _{free} ^c (%)	20.8/26.3
Estimated co-ordinate error based on R _{free} (Å)	0.527
Mean B-factor (Å ²) Protein, Nucleic acid, solvent	100.75, 151.23, 53.77
Deviations from ideal stereochemistry	
RMSD bonds (Å)	0.008
RMSD angles (°)	0.95
Wilson B-factor (Å ²)	83.0
Ramachandran plot analysis	
Most favoured (%)	91.79
Additionally allowed (%)	6.76
Disallowed (%)	1.45

R_{pim} is a measure of the quality of the data after averaging the multiple measurements and $R_{pim} = \frac{\sum_{hkl} [n/(n-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where n is the multiplicity, other variables as defined for R_{merge}.

$\sigma(I)$ = standard deviations of the measured intensity (I); $R_{\text{merge}} = \sum(I - \langle I \rangle) / \sum \langle I \rangle$, where $\langle I \rangle$ is the mean intensity for all observations

$R_{\text{work}} = \sum(|F_{\text{obs}} - F_{\text{calc}}|) / \sum F_{\text{obs}}$, F_{obs} are the observed structure factor amplitudes, and F_{calc} those calculated from the model

R_{free} is equivalent to R_{work} but where 5% of the measured reflections have been excluded from refinement and set aside for cross-validation purposes.

Ramachandran plot analysis was performed using molprobit (Chen et al., 2010).

4.3 Interactions between SOX and K2-31

K2-31 interacts with residues R248, F249, and Y373 in conserved motifs III that in other type II restriction endonucleases have important roles in substrate binding (Figure 4.7 and session 1.3.2). In particular, R248, F249 and Y373 at the catalytic region mediate hydrogen bonds with A20 and G21 located at the apex (Figure 4.7) and have all be shown to be involved in DNA binding in the SOX-DNA complex (Bagneries et al., 2011). Specifically Y373 acts as a hydrogen bond donor mediating contacts to both O1P and O4' of A20 (which acts as a hydrogen bond acceptor) while Nε1 of R248 acts as a hydrogen bond donor to O4' of A20 which donates a hydrogen bond to the carbonyl oxygen of C247. In addition, a hydrogen bond is observed between O1P of G21 and the peptide NH group of F249 and the 2'OH group of A20 forms a hydrogen bond with the carbonyl oxygen of C247 (Figure 4.7).

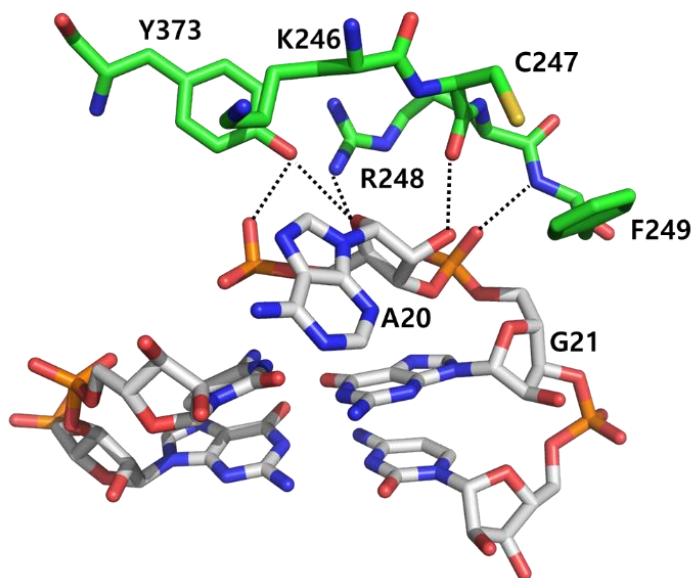


Figure. 4.7: Several key residues in SOX form hydrogen bonds with K2-31. Here, R248, F249 and Y373 form hydrogen bonds (black dotted line) with A20 and G21 located at the apex.

F179, located in the bridge motif, further stabilizes the SOX-K2-31 complex by forming a stacking interaction with the adenine base of A20. This interaction appears to be favoured by a disulfide bond formed between the two cysteine residues, C183 and C247 (Figure 4.8). Interestingly, there are no specific interactions observed between nucleotides 5' to A19 and SOX. Crucially, no interactions are observed between SOX and the UGAAG motif.

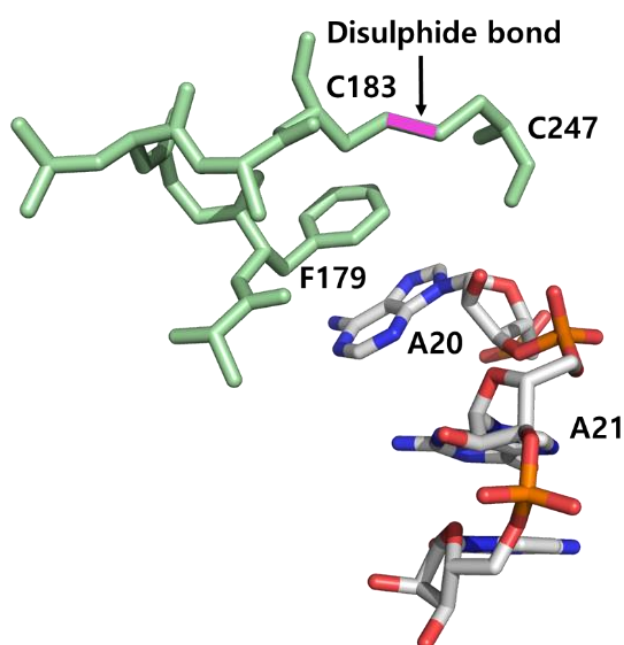


Figure 4.8: A20 forms a pi-stacking interaction with F179 which appears to be further stabilized by a disulfide bond (coloured in magenta) formed between C183 and C247.

4.4 Comparison of the SOX:DNA and RNA binding modes.

Although the RNA and DNA substrates share the same catalytic site, there are significant differences in the way in which both are accommodated as shown in Figure 4.9.

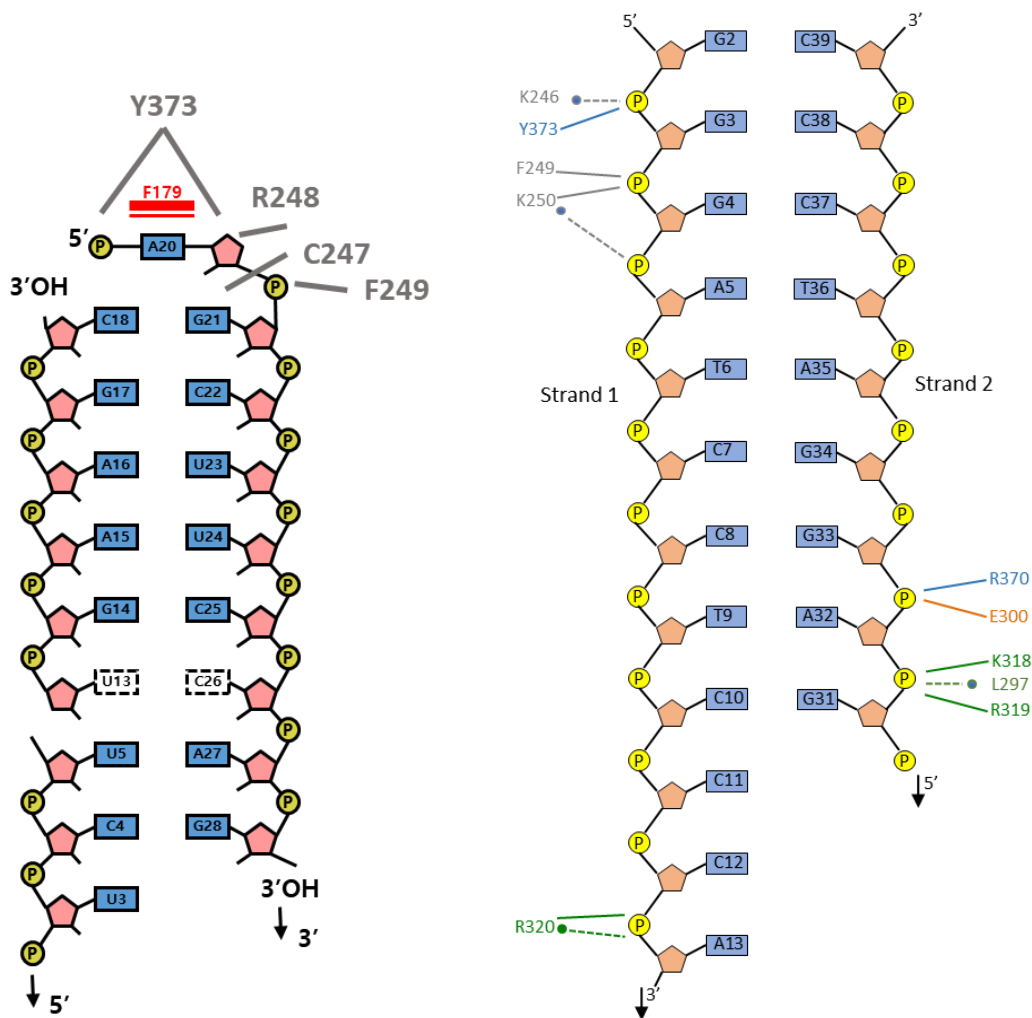


Figure 4.9: Schematic diagram illustrating the SOX-K2-31 interactions (left) and SOX-DNA interactions (right). Hydrogen bonds are highlighted in grey, π -stacking interactions are shown in red, dashed lines and asterisks are mediated by water molecules and unbroken lines are mediated by main or side chain groups. (Figure modified and adapted from Lee et al., 2017; Bagnieries et al., 2011).

The differences in the DNA and RNA binding modes are highlighted when the SOX-DNA and SOX-RNA structures are superimposed. It is clear that a 90 ° rotation would be required to map the RNA co-ordinates onto the DNA (Figure 4.10). Interestingly, the DNA duplex in the SOX-DNA complex interacts with residues in the second nuclear localization sequence (NLS). These interactions are absent in the SOX-RNA complex where the RNA alternatively extends into the solvent (Figure 4.10). The differences are also reflected in the accessible surface areas buried upon DNA and RNA binding that constitute 480 Å² and 240 Å² respectively.

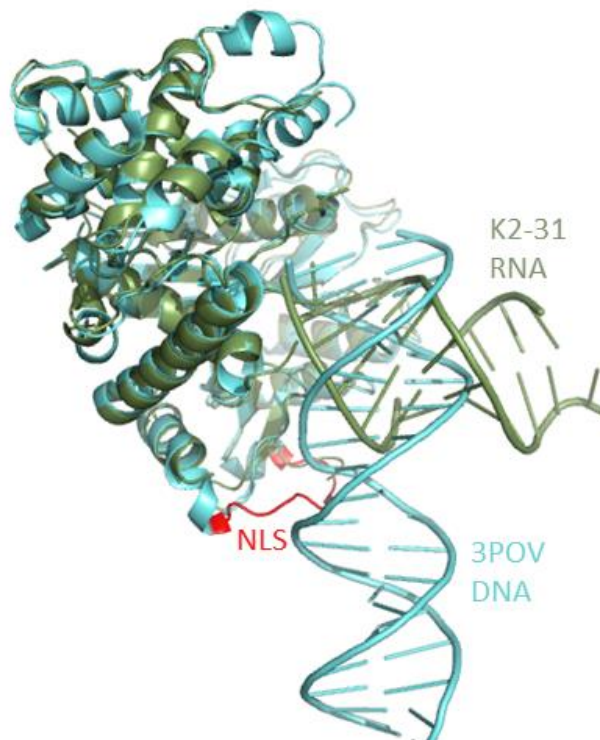


Figure 4.10: View of the SOXA:DNA complex superimposed on the SOXE244S:K2-31 complex. Both interact with the catalytic region but are highly divergent beyond. The SOX:DNA structure is coloured in cyan and SOX:K2-31 in green. The DNA in the SOX-DNA complex makes additional interactions with the nuclear localisation sequence (NLS highlighted in red). (Adapted and modified from Lee et al., 2017).

As described previously, K2-31 is further stabilised by stacking interactions between the aromatic side chain of F179, situated in the bridge region, and the adenine base of A20 (Figure 4.11). However, this configuration is absent in the SOX-DNA structure where F179 is positioned $\sim 5 \text{ \AA}$ away from A20 relative to its position in the SOX-RNA complex along with the apo structure. As previously mentioned, this rearrangement is further stabilised by a disulphide bridge formed between the sulphhydryl groups of C247 and C183 at the N-terminal end of the bridge motif that is also absent in the apo and DNA bound structures (Figure 4.11).

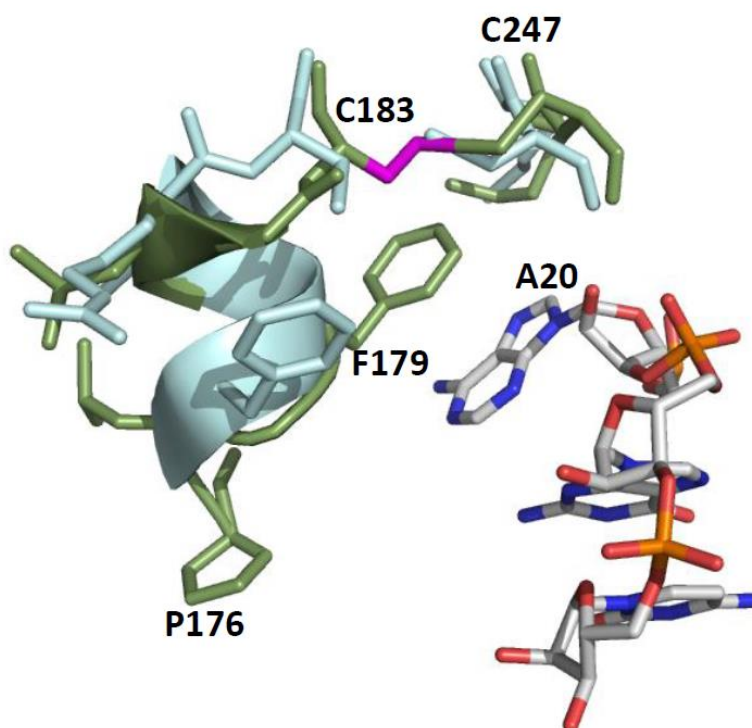


Figure 4.11: Important interactions between the A20 base of K2-31 and F179 in the bridge region (dark green). The protein co-ordinates have been superposed with those of the SOX-DNA structure (cyan, 3POV) in which significant conformational rearrangements were identified. The disulphide bond formed between C247 and C183 in the SOX-RNA structure is highlighted in magenta. (Figure adapted and modified from Lee et al., 2017).

These rearrangements suggested that the bridge plays a key role in SOX-mediated endonucleolytic cleavage of RNA substrates. Therefore to investigate the roles of these residues, both F179 and C247 were mutated to alanine and serine respectively. The F179A and C247S mutants were tested with GFP51 as a substrate and their activities assessed using RNase assays. GFP51 was used instead of K2-31 owing to its previous use as a model substrate. These experiments revealed that although C247S was only mildly defective compared to wild-type SOX, the F179A mutant was highly impaired in endonucleolytic activity. By contrast, both mutants were able to incise duplex DNA substituted with a 5' phosphate group (dsDNA5'P), suggesting that F179A functions specifically in the endonucleolytic cleavage of RNA substrates (Figure 4.12).

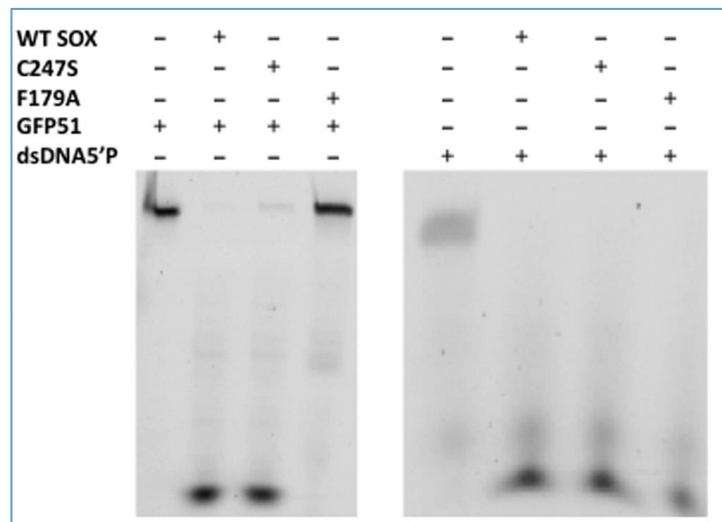


Figure 4.12: SOX RNase assays using F179A and C247S mutants illustrating that F179A is significantly impaired while C247S is largely unimpaired. Importantly, both the F179A and C247S mutants were able to incise duplex DNA substituted with a 5' phosphate group (dsDNA5'P), suggesting that both remained natively folded. (Figures adapted and modified from Lee et al., 2017).

4.5 Potential mechanism of SOX-mediated RNA degradation

Due to the usage of high concentrations of the SOXE244S mutant and duration of the crystallisation process, the RNA bound to the SOXE244S mutant seemed to undergo degradation as the density for the A19 was absent. Moreover, the lack of density for the phosphate group in close proximity to the catalytic residue D221 indicated that K2-31 was being endonucleolytically processed during the crystallisation experiment to produce a product complex. Therefore, to probe the potential mechanism of endonucleolytic cleavage, an 'intact loop' model was generated for K2-31 by combining the co-ordinates for loop residues 18-22 from the lowest energy structure prediction of K12-2 (MC-Fold MC-Sym pipeline) with the remaining stem nucleotides from the crystal structure (Figure 4.13A). Here, the amino acid sequences 221-249 containing the PD(D/E)-XK motif (D221, E244, K246 highlighted in green in Figure 4.13A) was used to define the active site with the two magnesium ions (Mg_A and Mg_B) derived from the SOX-DNA complex (PDB: 3POV) and the bacteriophage lambda DNA substrate complex (PDB: 3SM4) respectively given their reported roles in catalysis. The full K2-31 RNA sequences containing UGAAG was taken and analyzed using the MC-Fold | MC-Sym pipeline to produce secondary and tertiary structure predictions (Parisien and Major, 2008). Using this server, the lowest energy model generated for K2-31 was taken from the MC-FOLD | MC-Sym output and its intact stem loop (nucleotides from 18 to 22) was used to replace the cleaved loop in the K2-31 structure. This model was manually docked onto the SOX-E244S-RNA crystal structure (Figure 4.13A).

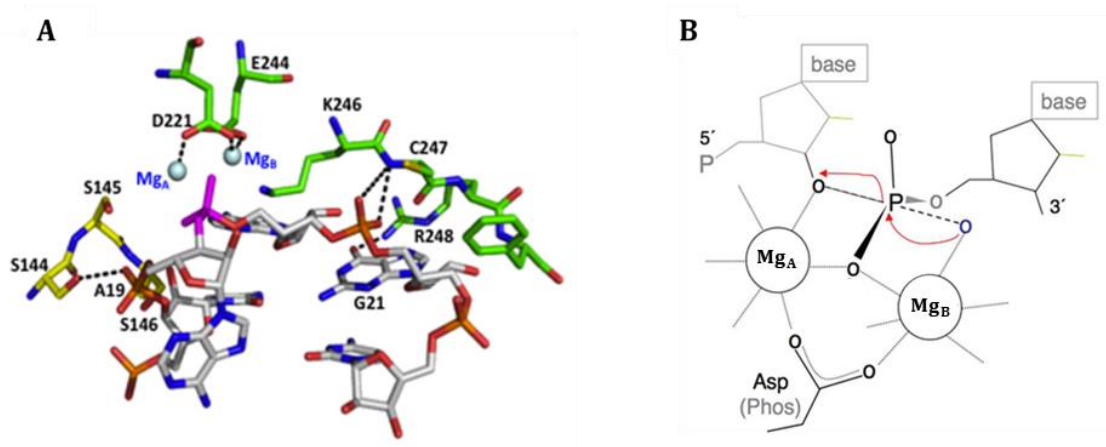


Figure 4.13: (A) Schematic diagram illustrating the potential cleavage geometry of K2-31 by SOX. The scissile phosphate (highlighted in magenta) is located between magnesium ions A (Mg_A derived from 3POV) and B (Mg_B derived from 3SM4). In this position, K2-31 is not only able to contact with the residues 221 to 249 (highlighted in green) but also the serine cluster (highlighted in yellow) which is reported to be essential for cleavage that is conserved in a number of type II restriction-like enzymes. (Figure adapted and modified from Lee et al., 2017). (B) Overview of the S_N2 cleavage geometry. In this mechanism, Mg_B deprotonates a water molecule resulting in the generation of a hydroxyanion attacking nucleophile and the Mg_A then act as a stabilizer allowing generation of a pentavalent transition state intermediate.

Analysis of this intact loop model revealed the most likely cleavage geometry of SOX-mediated RNA processing to be consistent with an SN2 bi-metal nuclease mechanism (Figure 4.13B). In this mechanism, one di-valent metal ion deprotonates a water molecule resulting in the generation of a hydroxyanion attacking nucleophile. The second metal ion could then act as a stabilizer allowing generation of a pentavalent transition state intermediate. This is similar to the mechanism put forward for SOX-mediated exonucleolytic DNA processing (Bagneris et al., 2011). Interestingly, a highly conserved residue S144 which has been shown to be essential for cleavage *in vivo*, failed to make contacts with K2-31 in the crystal structure, but in the intact loop model is favourably placed to interact with A19 that is directly 5' to the scissile phosphate. This serine residue is one of three highly conserved serine residues (S144, S145 and S146), reported to form a serine cluster that plays an important role in the stabilisation of the 5' phosphate in other type II restriction like enzymes such as bacteriophage lambda, highlighted in the bacteriophage lambda-DNA substrate complex (3SM4) (Bagneris et al., 2011).

Unlike the SOX:DNA complex, however, in which one of the putative catalytic metal ions appeared in a non-canonical position, the relative juxtapositions of the catalytic residues K246, E244 and D221 suggest that the configuration for in-line attack involves Mg_A and Mg_B in their canonical positions similar to the alignment of the scissile phosphate observed in 3SM4. In this configuration, the requirement for contacts with the binding residues of motif II, catalytic centre and serine cluster would be fulfilled.

4.6 SOX-mediated RNA endonucleolytic cleavage appears to be sequence non specific

As previously observed, there appear to be no interactions between either the bases or the phosphodiester backbone of the UGAAG motif and SOX. Therefore, to investigate whether this result might be due to a product complex binding configuration or a specific feature of K2-31, the UGAAG motif in GFP51 was substituted for UCUCU and UGCAC.

The UCUCU substitution was chosen given that all purines within the motif would be replaced by pyrimidines while in the case of UGCAC, the AAA/G sequence reported to potentially affect substrate binding would be substituted with CAC to maintain Watson Crick base pairing (Figure 4.14). To ensure maintenance of the overall stem loop configuration of both GFP51 mutants, MC-Fold was used to predict their secondary structures. The lowest energy structures suggested that their folds remained largely unchanged apart from a small bulge directly 5' to the apex of the stem loop (Figure 4.14).

Using these GFP51 mutants as substrates, RNase assays showed that they were still susceptible to degradation by SOX, suggesting that SOX-mediated endonucleolytic cleavage *in vitro* is not strictly sequence-specific (Figure 4.15).

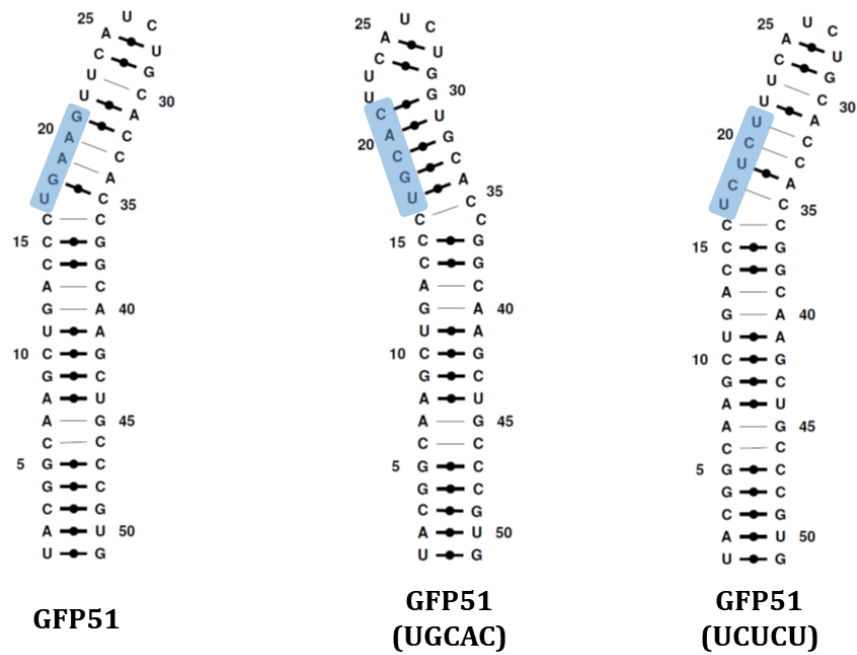


Figure 4.14: The MC-Fold | MC-Sym lowest energy secondary structure predictions for GFP51 (left) where the UGAAG motif has been substituted for UGCAC (centre) and UCUCU (right). The UGAAG motif substitutions are highlighted in blue.

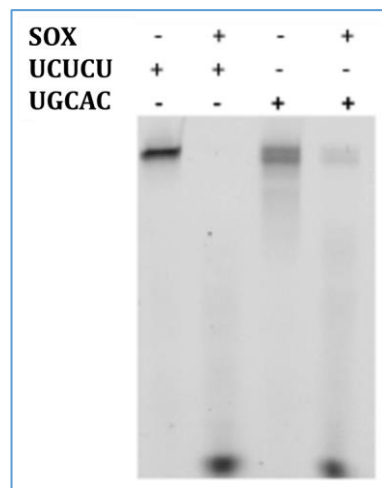


Figure 4.15: SOX RNase assays using GFP51 substitutions. Both the GFP51-UCUCU and GFP51-UGCAC mutants did not affect the activity of WTSOX. (Figure adapted and modified from Lee et al., 2017).

Also, to rule out the possibility that this failure to impair SOX-mediated RNA cleavage could be due to the concentration of SOX used and the duration of incubation, a time course RNase assay was carried out using GFP51-UCUCU and GFP51 as a control (Figure 4.16). This revealed that WT SOX could still degrade GFP51-UCUCU as efficiently as GFP51.

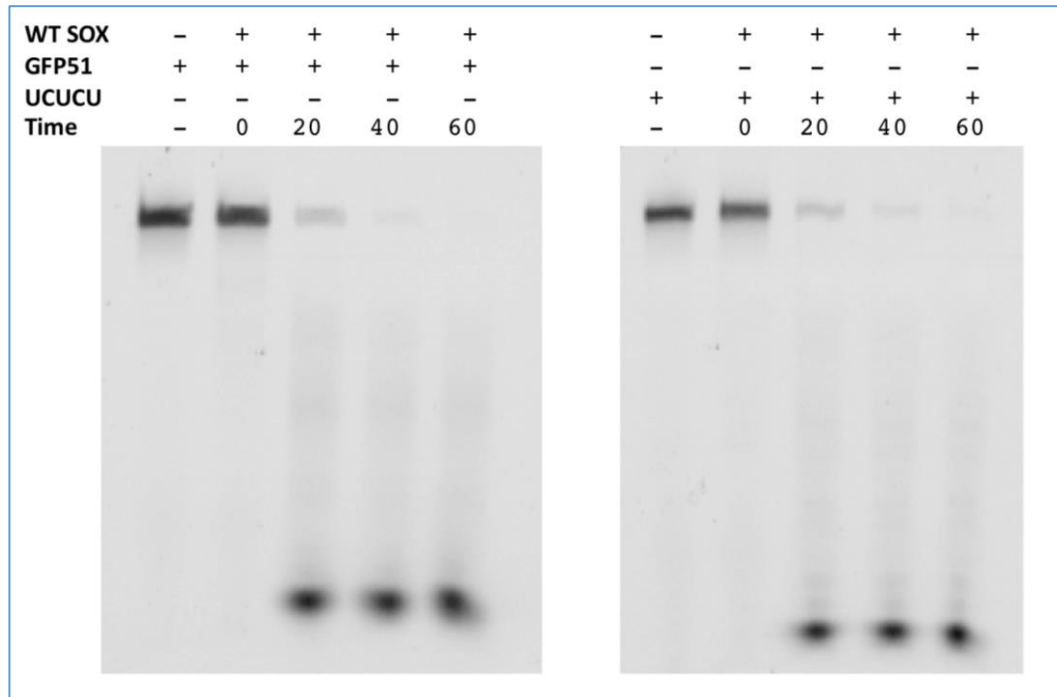


Figure 4.16: Time course cleavage assays involving SOX and the substrates GFP51-UCUCU and GFP51 illustrating that WT SOX can still degrade GFP51-UCUCU as efficiently as GFP51. Time is measured in 20 minute intervals. (Figures adapted and modified from Lee et al., 2017).

4.7 Probing the importance of stem loop size

To investigate the impact of stem loop size and configuration on SOX-mediated RNA turnover, four and nine additional adenines were introduced into the K2-31 stem loop regions (termed A4 and A9 respectively) and these oligonucleotides tested in RNase assays along with dsUn51, an unstructured polyadenine 51mer annealed to its complementary polyuracil sequence to generate double stranded RNA in the absence of a stem loop (Figure 4.17). The results revealed that SOX was able to readily degrade both confirming that although SOX requires stem/bulge structures for optimal endonucleolytic activity there appear to be no major limitations on the sizes of the loops incised. Significantly, however, the rate of turnover of these variants is greater than observed for K2-31 suggesting that they form better substrates.

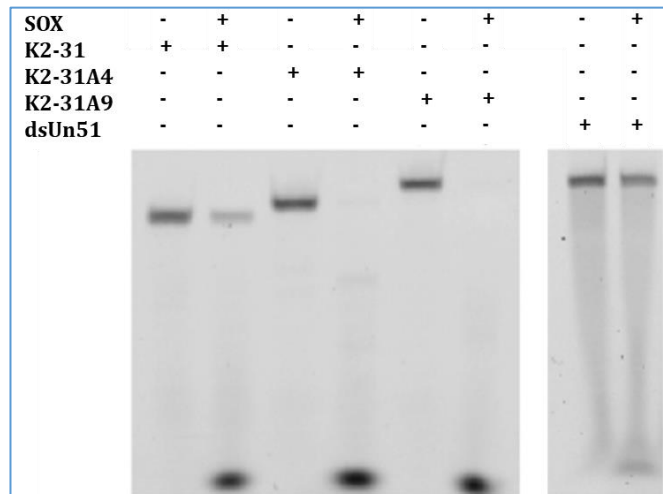


Figure 4.17: RNase assays using K2-31, K2-31A4 and K2-31A9 illustrating that all are readily degraded by SOX suggesting that endonucleolytic cleavage is non-sequence specific. The rate of turnover for K2-31A9 is greater than that for K2-31A4 suggesting that SOX favours larger loop. (Figure adapted and modified from Lee et al., 2017).

4.8 Binding affinity of SOX for different RNA substrates

Having proved that SOX-mediated RNA degradation requires stem loop/bulge elements and is not strictly dependent on consensus sequence, the binding affinity of SOX towards different substrates was investigated in order to establish whether any trends in cleavage preference could be identified. To enable this, RNA oligonucleotides corresponding to K2-31, GFP51, HBB58 and dsUn51 were tested using a Fluorescence Polarisation Assay (FPA). These studies revealed that WTSOX had the highest binding affinity for GFP51 ($K_d \sim 5 \mu\text{M}$), followed by HBB58 ($K_d \sim 42 \mu\text{M}$), K2-31 ($K_d \sim 49 \mu\text{M}$) and the lowest affinity for dsUn51 ($K_d \sim 73 \mu\text{M}$) (Figure 4.18). These results demonstrated that RNA substrates with simple stem loop structures appear to bind preferentially although a larger pool of substrates would be required before firm conclusions can be drawn.

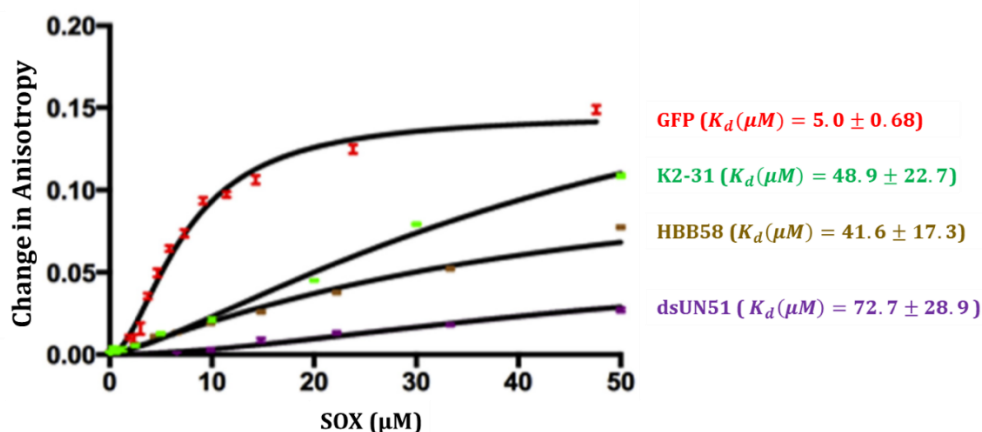


Figure 4.18: FPA binding curves obtained for WTSOX in the presence of the substrates GFP51, K2-31, HBB58 and dsUN51 illustrating that WTSOX had the highest binding affinity for GFP51, followed by HBB58, K2-31 and the lowest affinity for dsUn51. (Figure adapted and modified from Lee et al., 2017).

4.9 The HSO mutants have distinct functions in SOX-mediated host mRNA degradation

Since structures for both SOX:RNA and SOX:DNA complexes were available, the HSO mutants were re-evaluated to establish whether their defects could be explained by a failure to specifically cleave structured RNA (Bagneris et al., 2011; Gaglia et al., 2015) given that several appeared to cluster around the RNA interaction sites (Figure 4.19). Apart from SOXT24I, which was insoluble as previously reported (Bagneris et al., 2011), aliquots of the HSO mutants that constitute V369I, P176S, A61T, D474N and Y477Stop were available from previous studies in the lab.

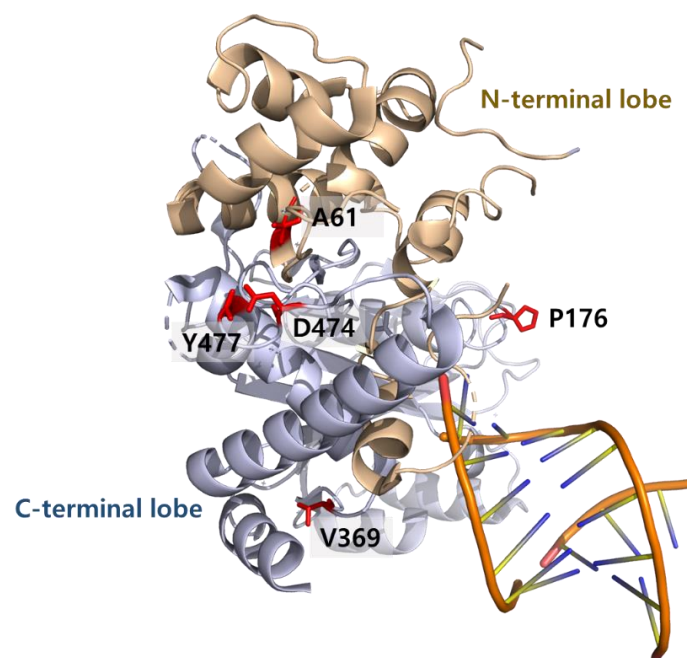


Figure 4.19: Schematic diagram highlighting the location of the residues involved in HSO. P176 is found at the centre of the bridge region, V369 is found at the centre of the C-terminal lobe, Y477 & D474 are also located at the C-terminal lobe ~ 30 Å away from the catalytic site. A61 is located at the N-terminal lobe and ~ 29 Å away from the catalytic site.

RNase assays using GFP51 as a substrate revealed that V369I (located at the centre of the C-terminal lobe) and P176S (located at the centre of the bridge region) were strongly impaired in endonucleolytic activity (Figure 4.20). This is consistent with the previous *in vivo* studies where both of these mutants abolished HSO (Glaunsinger et al., 2005).

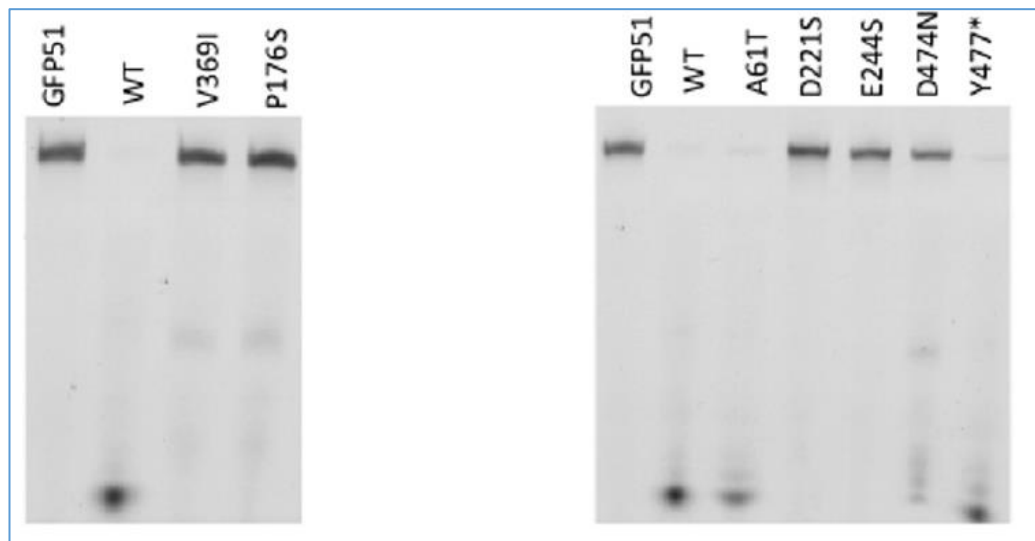


Figure 4.20: SOX RNase assays using HSO mutants (A61T, P176S, V369I, D474N and Y477stop) and GFP51 as a substrate illustrating that SOX mutants: V369I, P176S, and D474N impaired RNase activity of SOX. The catalytically inactive E244S and D221S were included as negative controls. (Figure adapted and modified from Lee et al., 2017).

Similarly, P176S, located at the centre of the bridge region close to F179, was also highly defective. These results suggest that the P176S mutation might disrupt the conformational rearrangement of F179 required for endonucleolytic processing. V369 is located within a small hydrophobic region that is in close proximity to the

interaction site involving A20 and the catalytic region. When mutated to isoleucine, the larger side chain would remodel the region disrupting substrate binding and catalysis.

By contrast, the RNase activity of both Y477Stop and D474N had significantly lower impacts. This may be due to the fact that since these mutants are located at the C-terminus and more than 30 Å away from the catalytic region, they may not have a direct impact on the endonucleolytic cleavage of RNA substrates. Consistent with this argument are the results obtained for the A61T mutant which showed almost WTSOX RNase activity. This residue is also present at the N-terminus and around 29 Å from the catalytic site.

4.10 Discussion

In order to survive and proliferate, viruses have evolved different strategies to escape the host's immune surveillance mechanisms. SOX, a highly conserved alkaline exonuclease in the KSHV family, is expressed at the lytic phase of infection to induce global and rapid degradation of host mRNA, thereby allowing the virus to 'hijack' the host's translational machinery (Glaunsinger & Ganem, 2004; Bagneries et al., 2011). Although the method by which SOX binds to DNA had been demonstrated, little could be ascertained about its mode of binding to or selection of RNA targets. Despite studies performed both *in vitro* and *in vivo* which suggested that SOX possesses an intrinsic RNase, the exact nature of its participation was called into question given that knock down of the SOX gene using single interference RNA (siRNA) still resulted

in HSO (albeit at reduced levels) (Hutin et al., 2013; Clyde & Glaunsinger, 2011). Although this could be attributed to poor affinity associated with ssRNA, it could also suggest that HSO may also function in the co-option of other factors for the efficient degradation of mRNA transcripts.

Studies were therefore performed to identify the nature of the RNA elements targeted by SOX and the variations in their sequences. Initial investigations found that the cleavage sites appeared to be restricted to regions within, or flanked by, unpaired nucleotides within RNA transcripts (Gaglia et al., 2015). Using this information, studies performed in the Barret group were able to identify cleavage elements in three transcripts known to be SOX targets corresponding to: GFP51, HBB58 and DsRed61 which were all endonucleolytically degraded in *in vitro* RNase assays. Although distinct in terms of topology, all appeared to have stem loops or bulge motifs adjacent to or within a UGAAG motif which was reported to be a targeting motif for cleavage. Based on this information, a unifying feature appeared to be the requirement for a structured RNA target containing stem loops or bulge motifs. In support of this, the two RNA targets, HBB58 or GFP51 transcripts were readily degraded by SOX while dsUn51 was resistant to processing. This led to the identification of KSHV pre-miRNA K12-2 (31mer) fragment (K2-31) as a potential substrate. Previous biochemical studies in which the intrinsic RNase activity of SOX was identified (Bagneris et al., 2011; Glaunsinger et al., 2005) revealed the E244S mutant to be catalytically inactive in both DNase and RNase assays. E244 forms part of the highly conserved PD(D/E)XK shown to have an important catalytic role (Bagneris et al., 2011, Glaunsinger et al., 2005). Therefore, the E244S mutant was

used in crystallisation trials. Following the identification and confirmation of K2-31 as a potential target, crystallisation of a SOX E244S-K2-31 complex was successful. The structure was subsequently determined using molecular replacement and refined to 3.3 Å.

4.10.1 Structure of SOX-K2-31 and its comparison to the SOX:DNA complex 3POV.

Closer inspection of SOX:K2-31 revealed that similar to the SOX-DNA structure, the RNA also binds to SOX in a configuration enabling engagement with the catalytic site where it makes several interactions with residues (namely R248, F249 and Y373) that are important in both DNA and RNA processing.

As one nucleotide A19 was missing in the SOX:K2-31 structure, it appeared that a biologically significant product complex had been trapped where it was possible to predict the potential mode of RNA-processing using the 3SM4 co-ordinates along with an intact loop model. These predictions point to an S_N2 type mechanism consistent with that put forward for SOX mediated cleavage of DNA. Based on the position of residues important for both nucleotide binding (residues from 247 to 249 and S144) and the two conserved magnesium ions associated with the catalytic residues D221 and E244, it is more likely that the scissile phosphate makes an interaction with the Mg_B ion in its canonical position. This is similar to other type II restriction endonuclease-like enzymes for example, bacteriophage lambda in the reported DNA substrate complex (3SM4) (Bagneris et al., 2011).

The major difference between DNA and RNA is that RNA has a 2' hydroxyl group, and this additional 2' hydroxyl group possesses both hydrogen donors and acceptors therefore allowing RNA to form highly stable, branched structures from the created loops and junctions (Baidya & Uhlenbeck, 1995). Moreover, this reactive 2' hydroxyl group plays a critical role in protein-RNA interactions in many enzymatic reactions, such as ribonuclease A, as these proteins rely on recognition and specific binding of the 2' hydroxyl group in RNA as observed by Baidya & Uhlenbeck, (1995) and others. In the SOX:K2-31 complex, the 2' hydroxyl groups of 20A H-bonds with the hydroxyl group of C247. Recognition of this 2'OH group renders SOX capable of recognizing RNA.

In addition, SOX-K2-31 is further stabilized by F179 which is part of the bridge structure where mutagenesis studies have confirmed that the bridge has a function in RNA but not in DNA processing. As previously hypothesised, this result also confirms that during endonucleolytic processing, the bridge plays a key role in substrate recognition or product stabilisation by associating with the nucleotide directly adjacent to that targeted for cleavage (Horst et al., 2012). Interestingly, differences are observed in the amino acid composition of the bridge region which is not strictly conserved amongst the γ -herpes viruses. Particularly notable is the repositioning of P158 (analogous to P176) as a result of a truncation involving two upstream amino acid residues. Interestingly, differences in the cleavage patterns of the SOX homologues have been observed (Covarrubias et al., 2011), which may originate from structural differences in the bridge conformations as a direct consequence of variations in their sequences (Gaglia et al., 2012).

Some disulfide bonds constitute a redox-active centre and participate directly in redox reactions, usually via reversible oxidation of a cysteine residue leading to a cysteine-sulfenic acid (Baker et al., 2008). For example, several disulfide bonds are found in oxidoreductases play an important role in regulating to prevent the accumulation of reactive oxygen species in the endoplasmic reticulum (Baker et al., 2008). Although it was speculated that the disulphide bond formed between C183-C247 might have an essential role in SOX, this is not in the case of SOX as the RNase assay result have shown that C247S mutant was only slightly able to impair the activity of SOX.

Previous studies have shown that three serine residues, S144, S145 and S146 in Motif I form a serine cluster with serine residue, S219, found in Motif II (Buisson et al., 2009; Dahlroth et al., 2009). *In vivo* studies have shown that when one of these serine residues, S146, was mutated, the catalytic activity of HCMV SOX on DNA substrates was completely abolished (Martinez et al., 1996; Goldstein and Weller, 1998) indicating that this serine cluster is also important for cleavage. More recently, the importance of S144 in the specific context of γ -herpesviruses was confirmed in RNA turnover catalysed by muSOX (the murine SOX homologue) (Mendez et al., 2018). The fact that this residue is well placed to form hydrogen bonds with the phosphate group directly 5' to the scissile phosphate in the SOX-K2-31 substrate model is indicative of an important role in substrate stabilisation (Goldstein & Weller, 1998; Mikhailov et al., 2004).

4.10.2 SOX-mediated turnover appears not to require a targeting sequence

Surprisingly, the UGAAG motif, originally proposed as being an important targeting element, does not participate in SOX-RNA interactions where K2-31 substrate fails to mediate additional contacts beyond the catalytic site. This lack of sequence specificity was supported by RNase assays where the conserved UGAAG motif was substituted by UGCAC or UCUCU in GFP51 which failed to impair cleavage. This is also supported by the experiment conducted with DsRed61 in which the loop structure was incised but lacked the UGAAG motif (Covarrubias et al., 2011). RNase assays additionally showed that SOX seems to prefer RNA targets that consist of simple stem loop structures rather than more complex loop configurations or bulge motifs, in keeping with the higher rates of turnover observed with UCUCU and GFP51. There are other known RNA endonucleases that are non-sequence-specific. For example, Ribonuclease V1 (RNase V1) is a ribonuclease enzyme which mediate cleavage of a double-stranded RNA in a structure specific manner, usually requiring a substrate of at least six stacked nucleotides. Similar to SOX, the RNase V1 requires the presence of magnesium ions for cleavage (Duval et al., 2013). Based on RNA assays involving K2-31 loop variants, larger loop sizes do not seem to have a negative impact on RNA turnover. By contrast, the rates of turnover of these variants appears greater than observed for K2-31 suggesting that they form better substrates.

The issue of whether SOX mediated cleavage is more structure than sequence orientated remains to be definitively addressed. Recent studies by *Mendez et al, 2018*

indicate a 100-fold difference compared to K2-31 in binding affinity compared to experiments conducted with a fragment of the LIMD mRNA transcript that contains the AAA/G sequence originally put forward as a targeting motif. Their studies further show that this fragment is more efficiently incised. Mutation of the AAA/G sequence results in a substantial reduction in both binding and turnover. This leads the authors to conclude that the sequence is essential for cleavage given that computational studies indicated that the mutations left the overall stem loop configuration unchanged. Conversely, a substantial increase in turnover for assays conducted with the K2-31 A4 and A9 variants detailed above favour a mechanism independent of the UGAAG motif given its increased distance away from the loop region in both fragments. These seemingly contradictory results can only be resolved by a detailed analysis of a greater number of transcripts with systematic variation of the sequences in and around their major cleavage sites once identified and/or further structural studies.

4.10.3 The HSO mutants have distinct roles in SOX-mediated RNA degradation

Based on the substrate preference of SOX-mediated RNA degradation, the HSO mutants were re-evaluated to ascertain whether the failure to process structured substrates would account for the loss of SOX activity at least for those in close proximity to RNA binding regions. Interestingly, the RNase assays gave a range of results that largely mirrored those reported *in vivo* (Glaunsinger et al., 2005). The

two mutants, P176S and V369I, showed the greatest defects in activity, which was most likely due to their important roles in stabilising the substrate/product. On the contrary, A61T and Y477* had near wild-type activity while D474N was moderately defective in RNase activity. This could be explained by their locations that are distant from the catalytic region and any motifs involved in RNA binding/stabilisation. It is possible therefore, that the HSO defects associated with these mutants may originate from their participation in processes such as the recruitment of Xrn1, given that it's been shown that the two proteins interact (Lee et al., 2017), but also the ribosome in order to target mRNA transcripts that are actively undergoing translation. This would be consistent with its co-localisation with polysomes and translation initiation factors (Glaunsinger et al., 2005).

4.10.4 Comparison between RNase activity of EBV BGLF5 and KSHV SOX

Previous biochemical studies have shown that EBV BGLF5 has both DNase and RNase activity and shares 42 % sequence identity with KSHV SOX (Buisson et al., 2012). Similar to KSHV SOX, BGLF5 contains catalytic residues D203 and E225 as a part of the highly conserved PD(D/E)XK (Buisson et al., 2009). Moreover, the D203S point mutation results in severely impaired DNase and RNase activities, similar to the E244S mutant in KSHV SOX. Although a complex of BGLF5 bound to RNA has yet to be reported, both wild-type and E203S mutant structures has been published. In the mutant, the bridge region of BGLF5 is ordered and comprises a 12-residue long (145-156) linking the N- and C-terminal lobes (Buisson et al., 2009). Whilst the

studies in this thesis indicate that the involvement of the bridge is an important factor in RNA cleavage in KSHV SOX, the positioning and potential for conformational flexibility suggest a similar role in BGLF5. It has been shown, however, that SOX and BGLF5 have different substrate preferences and patterns of cleavage: BGLF5 prefers unstructured and involves Manganese ions whilst SOX prefers structured substrate and involves Magnesium ions (Buisson et al., 2009; Lee et al., 2017). Notably, there is poor amino acid conservation in this region where the bridge in BGLF5 is 2 amino acids shorter than in SOX owing to a deletion. It is interesting to speculate that these differences maybe factors in conferring specificity, but needs to be further investigated. It would seem however, that similar to SOX, BGLF5 alone is insufficient to support host shut off and required the activities of a downstream exoribonuclease and possibly other factors.

Chapter 5. Structural Analysis of SOX bound to Phytic acid

5.1 Phytic acid and Pyranine as two potential inhibitors of SOX

From the SOX:DNA and SOX:RNA structures, SOX is observed to target exclusively the phosphodiester backbone of substrates and as a consequence of this non-specific interacting mode, lacks specialised targeting modules such as a specificity pocket. Given the nature of these interactions, potential inhibitors would therefore have to interfere with catalysis or RNA/DNA binding more generally in the absence of allosteric sites. Based on this premise, two small compounds, phytic acid and pyranine were selected as potential inhibitors (Figure 5.1) given that both have been shown to chelate metal ions and associate with nucleases with in the case of phytic acid (Akond et al., 2011). The open nature of the KSHV catalytic site further suggested that they could both be favourably accommodated prompting preliminary crystallisation trials.

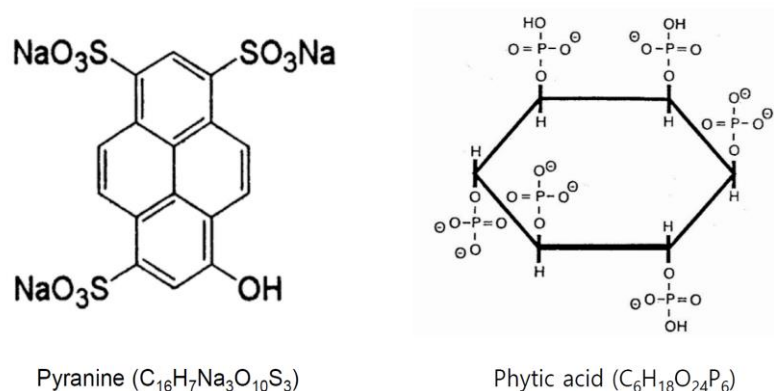


Figure 5.1: Molecular Structures of pyranine and phytic acid.

5.2 Crystallisation of SOX with Phytic acid

WTSOX was expressed and purified from *E.coli* as previously described (Bagneries et al., 2011) and used in co-crystallisation trials. Whilst trials with pyranine failed to produce crystals, initial screening of the WTSOX:Phytic acid complex using commercial screening kits was successful and produced multiple hits with similar crystal morphologies after approximately 7 days of incubation at 20 °C (Table 5.1). Of all the conditions screened, crystals grew best from the conditions: 0.1M MES, pH 6.5; 15% w/v PEG 6000 at 20 °C (Figure 5.2).

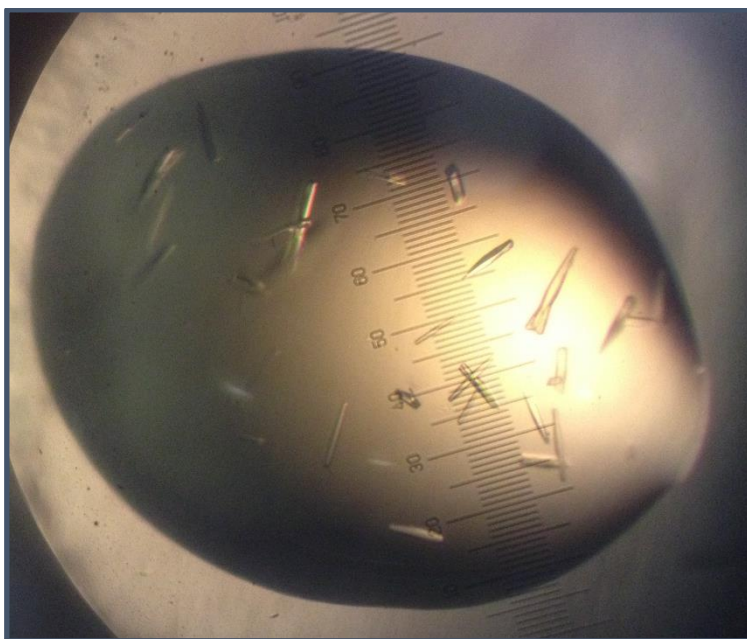

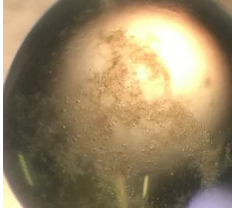
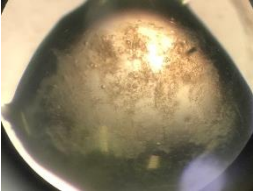
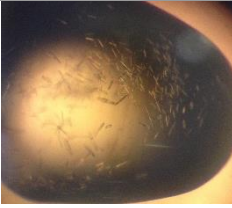
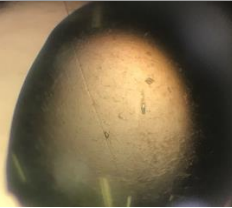


Figure 5.2: Crystals of WTSOX:Phytic acid that appeared in the optimized screen in conditions: 0.1M MES, pH 6.5; 7% w/v PEG 6000 at 20 °C.

Table 5.1: Crystallization conditions of KSHV-SOX (8 mg/mL)/Phytic acid (4 mM) complex screen.

Crystal conditions	Temperature (°C)	Crystal morphology
0.1M MES, pH 6.5 15% w/v PEG550 MME	20	
0.1M NaHEPES pH 7.0 15% w/v PEG 4000	20	
0.1M Magnesium Chloride 0.1M NaHEPES pH7.5 10% w/v PEG 4000	20	
0.1M MES, pH 6.5 15% w/v PEG 6000	20	
0.1 M Potassium Chloride 0.1M NaHEPES pH 7.0 15% w/v PEG5000 MME	20	

These crystals were then cryo protected in mother liquor solutions supplemented with 25% ethylene glycol and stored in liquid nitrogen prior to synchrotron data collection.

5.3 Structure determination of the SOX:Phytic acid complex

Data collection was performed on the cryo-cooled crystals at the Diamond Light Source (Didcot, UK) on the microfocus beamline I24 owing to the small size of the crystals that were typically 5 μm x 10 μm x 10 μm . Based on the recommended strategy from preliminary data processing on three 0.5 ° oscillation images taken at 0, 45 and 90 °. 3280 images were collected from a single crystal with an oscillation of 0.1 ° per image and an exposure time of 0.01 s. The resulting diffraction images were subsequently indexed and integrated using the XDS pipeline (Kabsch, 2009). The data were then scaled using the CCP4 program AIMLESS (Evans, 2006) and structure factor amplitudes generated using TRUNCATE (French & Wilson, 1978) to 2.33 Å. The structure of the potential WTSOX:Phytic acid complex was solved by molecular replacement using PHASER (McCoy et al., 2007) and the native KSHV-SOX co-ordinates (PDB accession code 3FHD) as a search model. Density consistent with phytic acid could be identified in preliminary F_o-F_c and $2F_o-F_c$ maps (Figure 5.3) within the catalytic site region of the canyon. Closer inspection revealed that this density could have been attributed to the N-terminus of a symmetry related monomer. However when N-terminal amino acids were built into the electron density and subjected to several rounds of refinement (Figure 5.4A), steric clashes were observed between the peptide and neighbouring residues where only one hydrogen bond could be observed between the hydroxyl group of Thr4 and hydroxyl group of Tyr373 (Figure 5.4B). Taken together, these observations were inconsistent with the density originating from a symmetry related monomer.

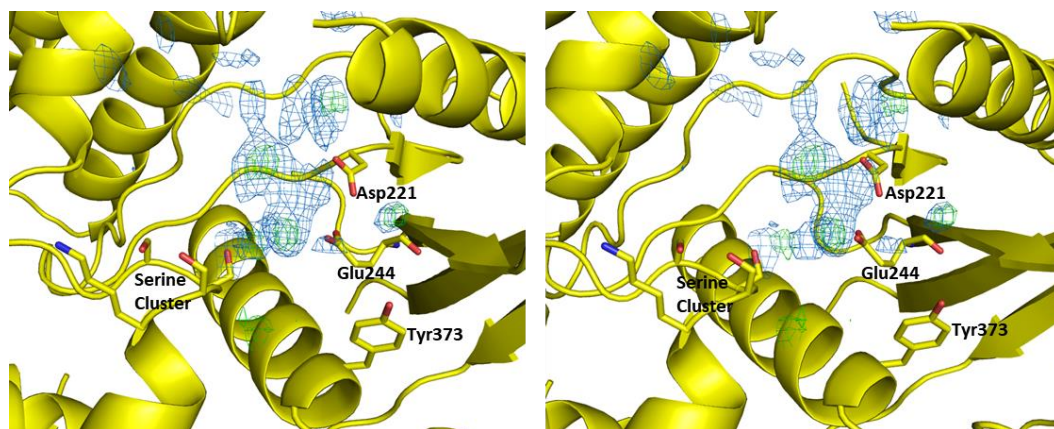


Figure 5.3: Stereoview of the initial F_o-F_c (green) and $2F_o-F_c$ maps (contoured at 3σ and 1σ respectively) consistent with phytic acid in the vicinity of the active site region.

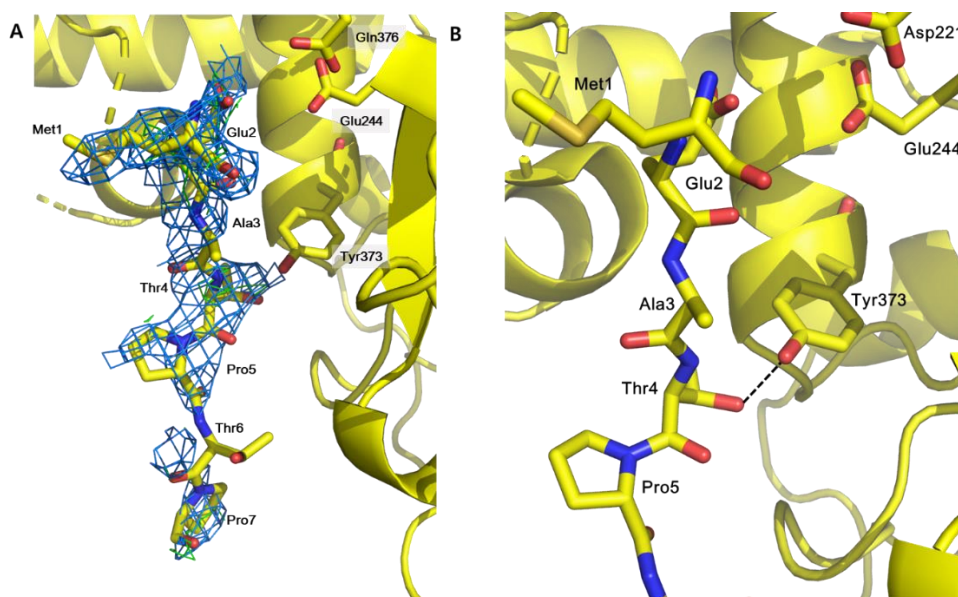


Figure 5.4: (A) Initial F_o-F_c (green) and $2F_o-F_c$ maps (contoured at 3σ and 1σ respectively) of N-terminus of symmetry related monomer. (B) Amino acid residue from 2 to 7 of N-terminus of symmetry related monomer (highlighted in yellow) built into the initial electron density after several cycles of refinement in AUTOBUSTER. A single hydrogen bond (dotted line) is observed between the hydroxyl (OH) group of residue Thr4 and hydroxyl (OH) group of Tyr373.

The co-ordinates for phytic acid (IHP) were obtained directly from protein databank entry 2XAM and fitted into the initial difference electron density obtained after molecular replacement. This was then followed by several rounds of refinement using AUTOBUSTER and re-building in COOT (Figure 5.5). The refinement statistics for the final model are shown in Table 5.2.

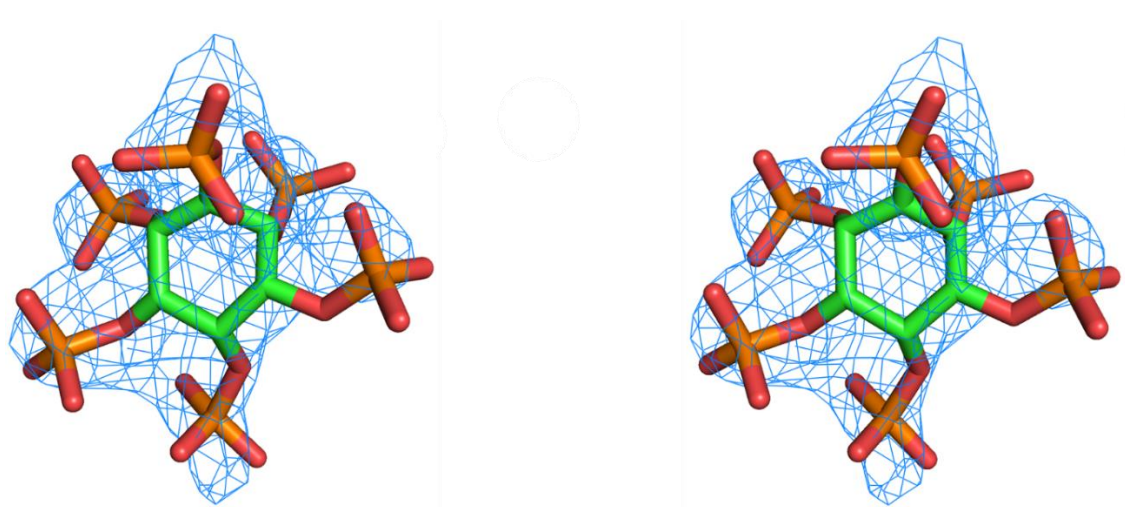


Figure. 5.5: Stereoview of phytic acid together with the associated $2F_o-F_c$ map density (contoured at 1).

5.4 Phytic acid binds to key residues implicated in catalysis

Inspection of the complex revealed that phytic acid makes contacts with residues at the active site of SOX. These include Leu181, Glu184, Lys246, Tyr373, Gln376 and Ser146 which are highlighted in Figure 5.6.

Table. 5.2: Summary of data collection and refinement statistics for final model of the SOX:Phytic acid complex.

Crystal	SOX:Phytic acid complex
Protein solution	8 mg/mL WT SOX
Phytic acid	4 mM
Precipitant solution	0.1M MES, pH 6.5; 15% w/v PEG 6000
Unit Cell	
a, b, c (Å)	66.6, 49.0, 72.9
α , β , γ (°)	90, 98.2, 90
Space Group	P2 ₁
Resolution (Å)	24-2.33 (2.33-2.46)
Number of Measured Reflections	19830
Number of Unique Reflections	2867
R _{merge} (Outer Shell)	0.21 (0.863)
Multiplicity	3.4 (4.3)
Completeness (%)	98.35 (98.7)
Mean I/ (Outer Shell)	6.2 (1.3)
Wavelength (Å)	0.9686
R _{pim}	0.036 (0.479)
CC _{1/2}	0.982 (0.746)
Refinement	
Number of atoms Protein, others, Solvent	3293, 54, 108
R _{work} ^b /R _{free} ^c (%)	22.6/28.8
Estimated co-ordinate error based on R _{free} (Å)	0.509
Mean B-factor (Å ²) Protein, others, solvent	34.8, 31, 20.2
Deviations from ideal stereochemistry	
RMSD bonds (Å)	0.015
RMSD angles (°)	1.439
Wilson B-factor (Å ²)	33.0
Ramachandran plot analysis	
Most favoured (%)	96.29
Additionally allowed (%)	2.11
Disallowed (%)	0.47

R_{pim} is a measure of the quality of the data after averaging the multiple measurements and

$R_{pim} = \sum_{hkl} [n/(n-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where n is the multiplicity, other variables as defined for R_{merge} .

$\sigma_{(I)}$ = standard deviations of the measured intensity (I); $R_{merge} = \sum(I - \langle I \rangle) / \sum \langle I \rangle$, where $\langle I \rangle$ is the mean intensity for all observations

$R_{work} = (|F_{obs} - F_{calc}|) / (F_{obs})$, F_{obs} are the observed structure factor amplitudes, and F_{calc} those calculated from the model

R_{free} is equivalent to R_{work} but where 5% of the measured reflections have been excluded from refinement and set aside for cross-validation purposes.

Ramachandran plot analysis was performed using molprobity (Chen et al., 2010).

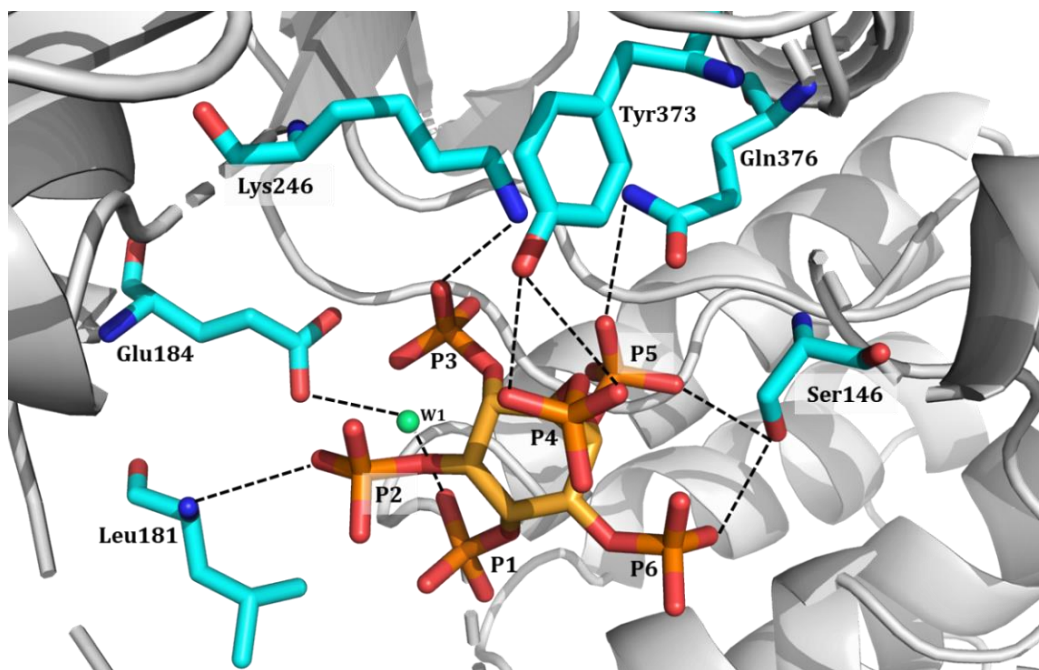


Figure 5.6: Several residues make hydrogen bonds with a phytic acid. Ser146 forms hydrogen bonds with phosphate groups of P5 and P6 of Phytic acid. The OH group of Tyr373 makes a hydrogen bonds with a phosphate group of P4. The N ζ group of Lys246 and the amide group of Gln376 make hydrogen bonds with phosphate groups P3 and P5 respectively. A hydrogen bond is also formed between the P2 phosphate group and the peptide NH group of Leu181. The carbonyl oxygen of Glu184 makes a hydrogen bond with P1 phosphate group via a water molecule.

Notably, Ser146 that forms part of the serine cluster present in Motif 1 donates a hydrogen bond to phosphate groups P5 and P6 of Phytic acid. Phytic acid is also stabilized by hydrogen bonds formed between the phosphate group of P4 and OH group of Tyr373. The phosphate groups P3 and P5 also forms hydrogen bonds with the N ζ group of Lys246 and the amide group of Gln376 (Figure 5.6). The phytic acid is further stabilized by a hydrogen bond between its P2 phosphate group and the

peptide NH group of Leu181. A hydrogen is also formed between the carbonyl oxygen of Glu184 and the P1 phosphate group via a water molecule (Figure 5.6). Phytic acid is also incorporated into the coordination sphere of Mg_A linking three residues Glu184, Asp221 and Glu244 (Figure 5.7).

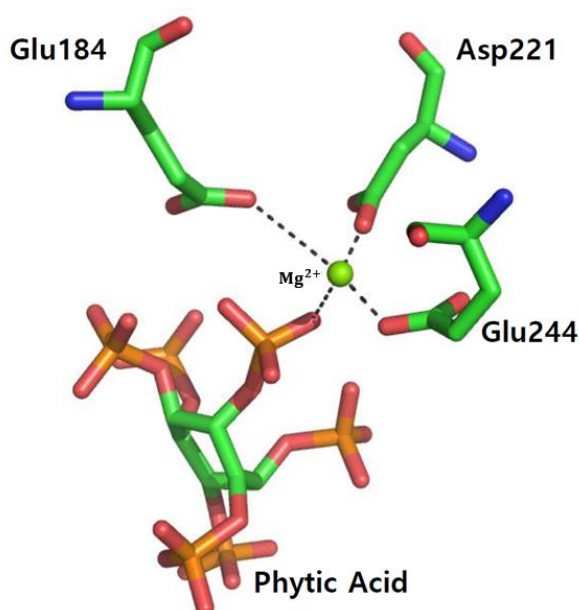


Figure 5.7: Phytic acid is co-ordinated to the magnesium ion Mg_A that co-ordinates to the carboxylate groups of Glu184, Asp221 and Glu244.

Interestingly, when compared to the SOX:DNA structure (highlighted in yellow in Figure 5.8), the bridge region (residues P164 to V180) appears re-oriented in response to the Leu181 interaction that moves $\sim 2 \text{ \AA}$ towards Phytic acid (Figure 5.8). Unexpectedly, this modest movement coupled with the interaction between phosphate P2 and the side chain of Glu184 is also associated with major remodelling of the bridge region that had to be substantially rebuilt and allowed the

identification of residues from 173-175 that were previously disordered. Overall, the location of phytic acid would suggest the capacity to impact on the association of both DNA and RNA substrates whilst also preventing their subsequent processing.

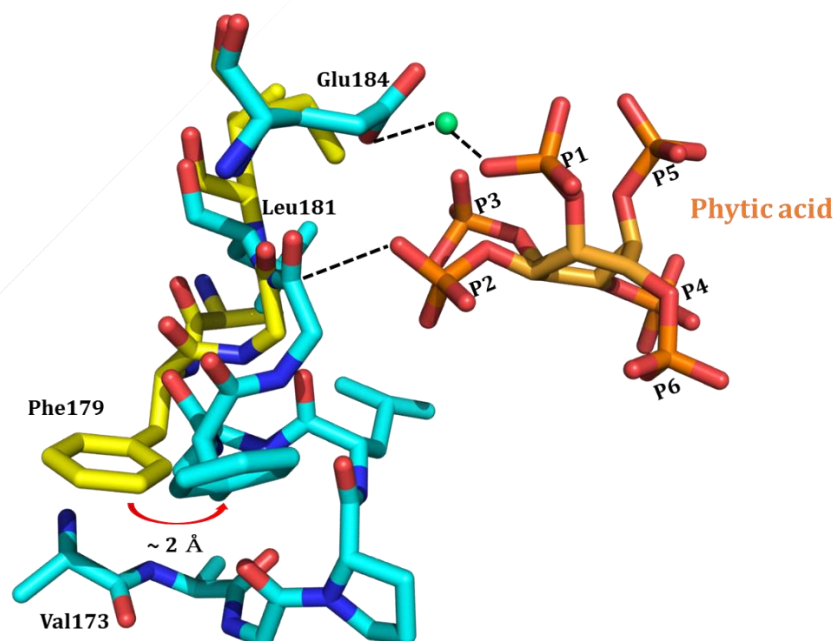


Figure 5.8: The SOX:Phytic acid interactions involving the bridge region (P164 to V180) and the conformational re-arrangements observed in transitioning from the SOX:DNA (3POV, highlighted in yellow) to SOX:Phytic acid complex (highlighted in cyan) by ~ 2 Å towards Phytic acid.

5.5 ThermoFluor Studies of SOX:Phytic acid complex

Thermal shift assays were used to establish the effects of phytic acid binding to SOX. The addition of phytic acid demonstrated an average negative thermal shift around 1.5 °C (Figure 5.9).

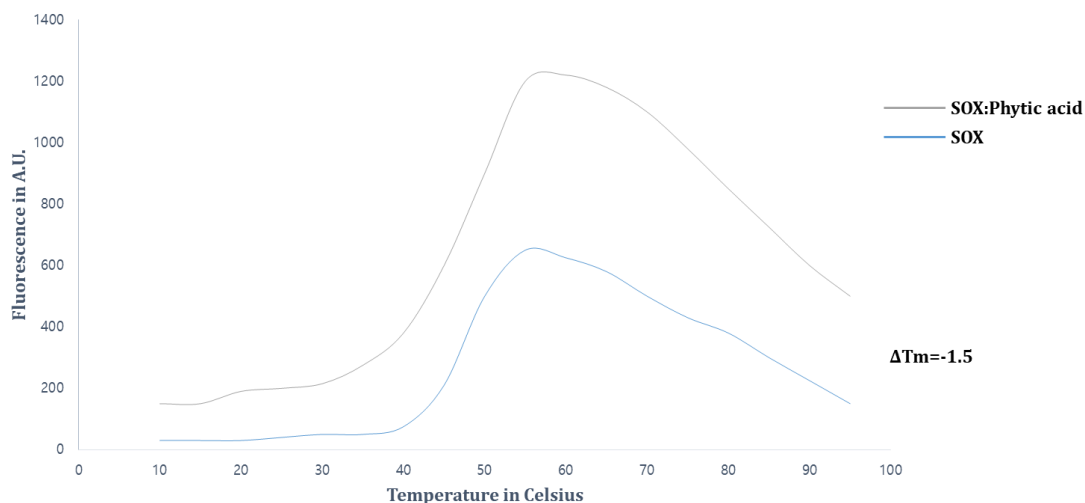


Figure 5.9: Thermal denaturation curves for SOX (blue curve) and SOX in the presence of phytic acid (grey curve) in which the addition of phytic acid resulted in a negative thermal shift around 1.5 °C.

5.6 Discussion

As discussed, there are no specific structural elements associated with the recognition of RNA and DNA oligonucleotides by SOX. Inhibition of SOX would therefore require either disruption of nucleic acid binding or catalysis. Since both the structures of SOX bound to DNA and SOX bound to RNA are now available, two potential inhibitors were selected Phytic acid and Pyranine which contained potentially mimetic phosphate. It has also been shown that phytic acid chelates very strongly with divalent metal ions and has been suggested to bind to nucleases (Shamsuddin, 2002).

Despite numerous attempts to crystallise SOX with pyranine, only crystals of SOX in the presence of phytic acid were obtained where the crystal structure of a SOX:Phytic acid complex was solved to a resolution of 2.3 Å. Similar to the SOX:RNA complex, phytic acid makes interaction with catalytically important residues including Ser146. Serine 146 is part of the serine cluster (S144, S145 and S146 in Motif 1 and S219 in Motif II) that are highly conserved in the type II restriction endonuclease like superfamily (Buisson et al., 2009; Dahlroth et al., 2009). Whilst S146 is dispensable for host shutoff, S144 is essential along with Q376 that is also co-ordinates to phytic acid. Moreover, phytic acid is further stabilised by interacting with Lys246 which is part of the highly conserved PD-(D/E)XK (D221, E244, K246) motif in the type II restriction like endonuclease superfamily (Pingoud et al., 2005) and is also essential for catalysis. Interestingly, the bridge region in the SOX-phytic acid structure is reoriented when compared to the SOX:DNA complex where it is observed to move towards phytic acid on account of the hydrogen bond mediated by Leu181. This configuration of the bridge is intermediate between those observed in the SOX-DNA and SOX-RNA structures and would be unlikely to support RNA cleavage based on the position of F179.

As discussed in chapter 4, magnesium ion A (Mg_A) is implicated in catalysis which is most likely to proceed via a canonical S_N2 bi-metal nuclease mechanism for both endonucleolytic and exonucleolytic cleavage (Bagneries et al., 2007; Lee et al., 2017). In terms of this cleavage mode, the interaction of Mg_A with phytic acid would disfavour the necessary interactions with the scissile phosphate whilst blocking access to key residues required for binding potentially including the serine cluster.

This type of inhibition is seen in other type II restriction endonucleases. For instance, EcoRV endonuclease is inhibited by the deoxyribo-3'-S-phosphorothiolates which ligates to one of two catalytic Mg²⁺ ions thereby inhibiting cleavage (Horton et al., 2000). Other example includes inhibition of the restriction enzyme EcoRI by phenethylephenylphthalimide derivatives which contain dibydroxyphenyl group that makes hydrogen bonds with catalytically important residues thereby inhibiting cleavage (Motoshima et al., 2011).

The ability of SOX to associate phytic acid was further investigated using a ThermoFluor assay. The results suggested that phytic acid does indeed interact with SOX, but not strongly. Curiously, this association appears to be de-stabilising given the decrease in melting temperature relative to SOX alone. Whilst the kinetics of the SOX-phytic interaction have yet to be determined, the co-crystal structure of phytic acid bound to WT-SOX has provided insights into features that are important for the future design of more potent inhibitors.

Chapter 6: Towards understanding the process of KSHV-vFLIP mediated subversion of the non-canonical NF- κ B pathway

As discussed in the introduction, the latently expressed viral oncoprotein KSHV-vFLIP has been directly implicated in KSHV cellular transformation and in particular, is essential for the survival of both PEL and KS cells. Whilst its role in constitutively activating the normally tightly controlled canonical NF- κ B pathway has been extensively characterised, its capacity to activate the non-canonical pathway is unclear, in particular, the nature of complexes involving KSHV-vFLIP, p100 and IKK α that are key players. In order to establish the nature of these complexes, pull down assays were performed using recombinantly expressed proteins.

6.1 Expression and purification of IKK α

Full length IKK α was successfully cloned into the baculovirus pFASTBAC vector and expressed in sf9 insect cells. Although affinity chromatography resulted in highly purified His-tagged IKK α (Figure 6.1), subsequent attempts to purify IKK α using gel filtration resulted in the protein being retained on the column following TEV cleavage to remove the 6His-tag. This could not be resolved despite trialling different buffer compositions and concentrations of IKK α . It was therefore decided that it was best to leave the tag *in situ* given the high levels of purity observed.

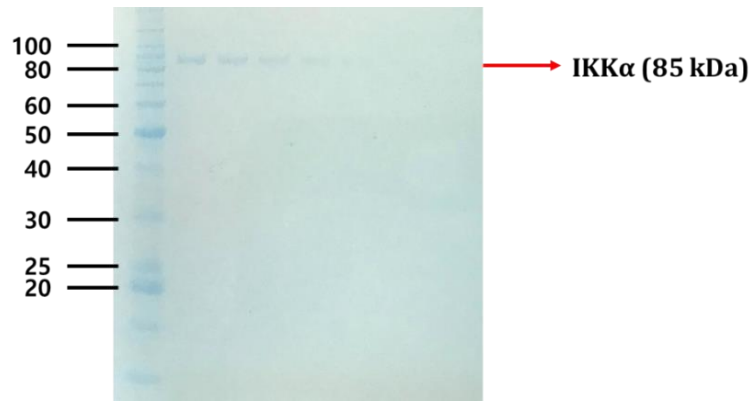


Figure 6.1: SDS-PAGE gel of IKK α (~85 kDa) after affinity chromatography using a 5ml HisTrap column.

6.2 Expression and purification of p100

p100 was successfully cloned into the baculovirus pFASTBAC vector and expressed in sf9 insect cells. It was purified using affinity chromatography (5 mL HisTrap) and size exclusion chromatography, yielding ~500 μ L of protein at a concentration of 2.3 mg/mL from 250 mL of sf9 insect cells (Figure 6.2).



Figure 6.2: SDS-PAGE gels of p100 (~110 kDa as a monomer and ~220 kDa as a dimer) after size exclusion chromatography.

Short constructs of p100 (746-846 and 746-860) to test the role of the death domain in the recruitment of KHSV-vFLIP were also designed and successfully cloned into the pETM6T1 vector where they were subsequently expressed in the BL21 DE3(Star) *E.coli* strain. Both constructs yielded approximately 500 μ L of 3.4 mg/mL per 5 L of *E.coli* cells (Figure 6.3).

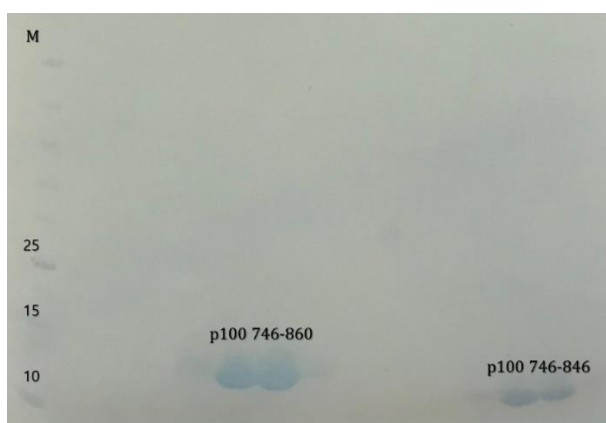


Figure 6.3: SDS-PAGE analysis of both p100 746-860 and p100 746-846 fractions after gel filtration.

6.3 Pull-down assays

In order to investigate the interaction between p100, KSHV-vFLIP and IKK α , a batch pull-down assay was carried out using 100 μ L of Ni-NTA resin (Thermo Scientific). Due to the lack of information on the stoichiometry of the putative ternary complex, p100, IKK α and KSHV-vFLIP were combined in a \sim 1:1:1 ratio. The pull-down assay results indicate that although there was evidence of some co-elution between p100, IKK α , and KSHV-vFLIP, band intensities were weak and non-stoichiometric making the assessment of whether binding had taken place difficult (Figure 6.4).

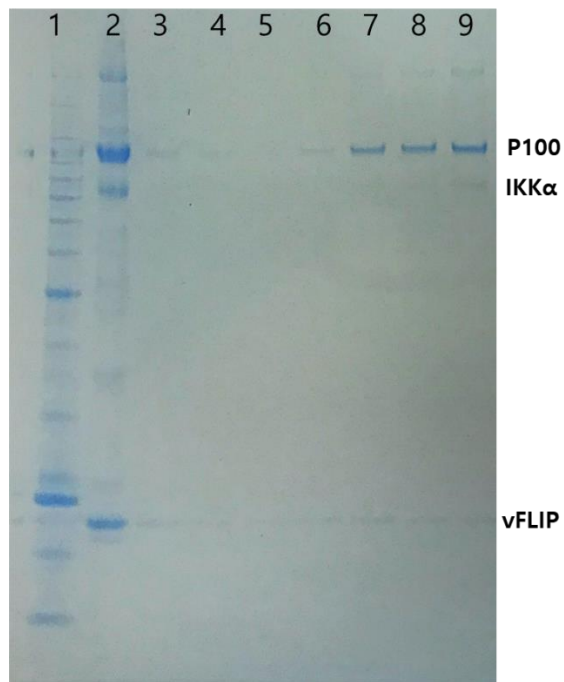


Figure 6.4: SDS-PAGE analysis of a pull-down assay involving p100, IKK α and KSHV-vFLIP. Lane1: Molecular Marker; Lane2: Sample load (P100 = 100 kDa, IKK α =85 kDa, and KSHV-vFLIP= 20 kDa); Lane3= Flow through fraction; Lane4-5: Wash fractions without imidazole; Lane6-7= Wash fractions containing 30 mM Imidazole; Lane8-9: Elution fractions containing 500 mM imidazole.

This was also true of a similar experiment performed in the presence of RelB (supplied by Dr Tracey Barrett) that associates with p100 to form a transcriptionally active heterodimeric complex following p100 processing to p52. In this assay, 100 μ L of resin was incubated with IKK α , p100 (untagged), KSHV-vFLIP (derived from precision cleavage of GB1-vFLIP), and/or RelB (Figure 6.5).

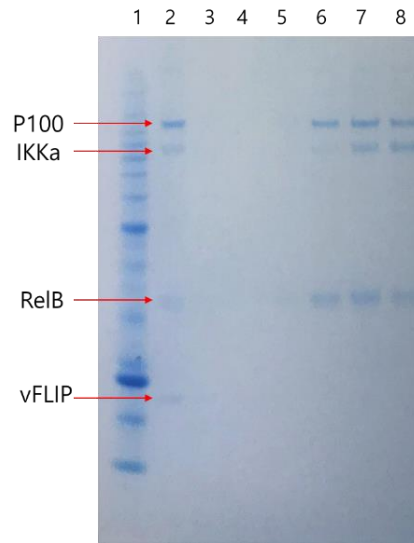
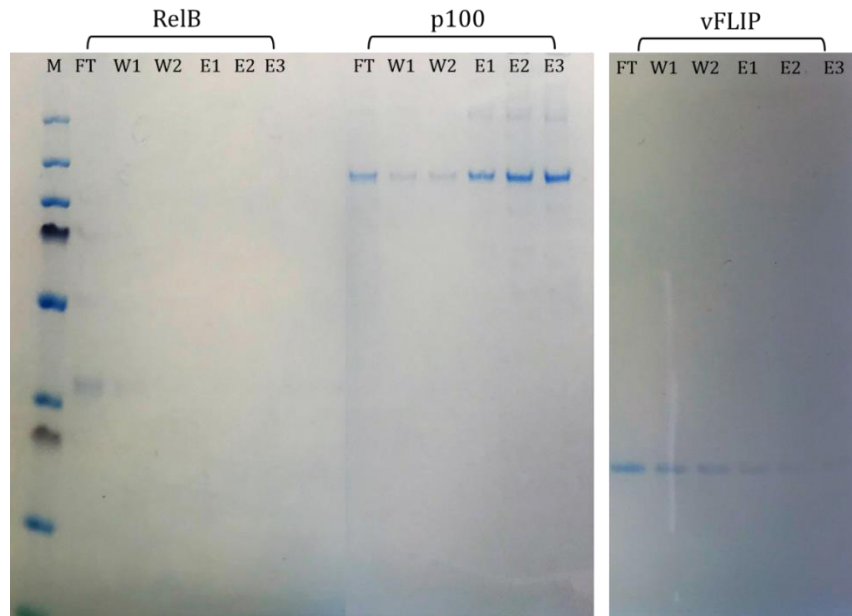


Figure 6.5: SDS-PAGE analysis of a pull-down assay involving p100, IKK α , RelB and vFLIP. Lane1= Molecular weight marker; Lane2= Mixture of samples prior to loading; Lane3= Flowthrough; Lane4=Wash fraction without imidazole; Lane5= Wash fractions containing 30 mM Imidazole; Lanes 6,7 and 8= Elution fractions containing 500 mM imidazole.

Interesting whilst p100, and IKK α co-eluted with RelB, there was no evidence of KSHV-vFLIP that was mainly in the flow through fractions. Given the lack of stoichiometric complexes in both experiments, all “untagged” proteins were therefore analysed for their capacity to non-specifically associate with Ni-NTA resin (Figure 6.6A). These results revealed that whilst untagged KSHV-vFLIP has residual affinity for Ni-NTA, p100 was proficient in binding on account of incomplete removal its 6-His tag after cleavage as observed by Western blotting (Figure 6.6B). Given the various complexities and time constraints, it was decided to focus on the interaction between KSHV-vFLIP and p100 that had previously been verified in cell based assays by Chaudhuri and co-workers along with our collaborators.

A



B

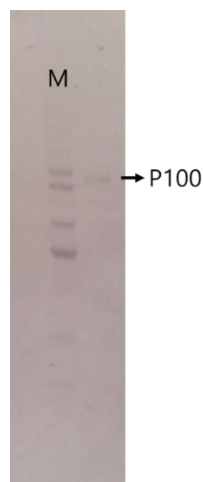


Figure 6.6: (A) SDS-PAGE analysis of control pull-down assays involving RelB, p100 and KSHV-vFLIP. M= Molecular weight marker, FT= Flowthrough; W1=Wash fraction without imidazole; W2=Wash fraction containing 30mM imidazole; E1-3=Elution fractions containing 500 mM imidazole. (B) Western blot analysis of p100 showing the 6-His tag was still retained after cleavage.

6.4 Interaction between p100 and KSHV-vFLIP

The interaction between p100 and KSHV-vFLIP alone was first investigated using an S200 10/300 gel filtration column. Figure 6.7 shows the peak fractions analysed using SDS-PAGE. Although most of the KSHV-vFLIP eluted in fractions 12 13 and 14, fractions 3 to 9 showed evidence of co-elution between p100 and KSHV-vFLIP although the KSHV-vFLIP bands were weak. These results were possibly indicative of a low affinity interaction.

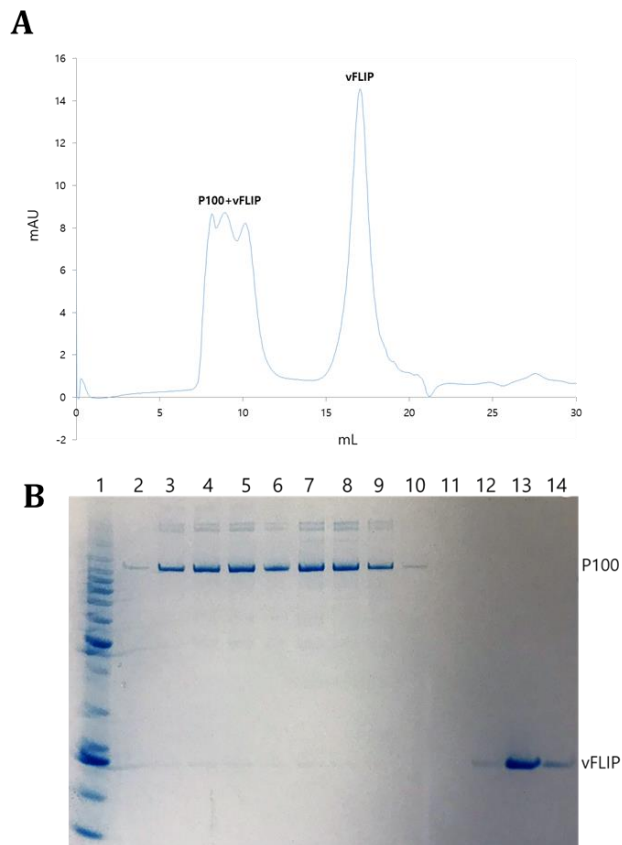


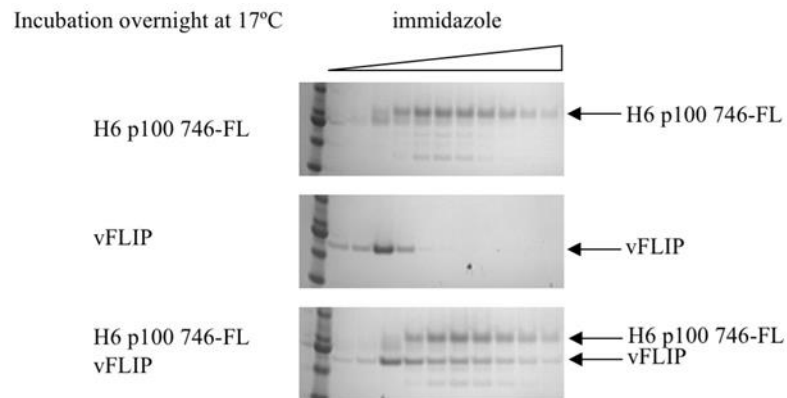
Figure 6.7: Putative interactions between P100 and vFLIP investigated using gel filtration. (A) Gel filtration trace of p100 and vFLIP. (B) Bands corresponding to p100 (~110 kDa) and KSHV-vFLIP (~ 20 kDa) were visible in fractions 2 to 10 indicating a possible weak complex given that KSHV-vFLIP mainly elutes in fractions 12 to 14.

6.5 Pull-down assay to probe the interaction between KSHV-vFLIP and the death domains of p100

Previous cell-based experiments performed by our collaborators implicated the C-terminus of p100 as being involved in the interaction with KSHV-vFLIP. Their interaction studies involved a p100 mutant truncated from the death domain that appeared defective in binding to KSHV-vFLIP (Figure 1.19 in session 1.15). This was confirmed by preliminary pull-down and gel filtration experiments performed in the Barrett lab involving the p100 construct 746-900 incorporating the death domain and far C-terminus (Figure 6.8). This assay result showed that while KSHV-vFLIP as the negative control did not show any specific binding to the 1 mL HisTrap column, the overnight incubation of KSHV-vFLIP and p100 746-900 (746-FL) co-eluted at high concentrations of imidazole.

To ascertain whether the p100 death domain was directly involved in KSHV-vFLIP binding, or the region between the death domain and far C-terminus, p100 mutant constructs corresponding to the death domain alone (residues 746-846) and a longer construct lacking the unstructured C-terminal tail, but retaining the first putative phosphorylation site (746-860) were tested for pulldown experiments with KSHV-vFLIP-6His (Figure 6.9).

A



B

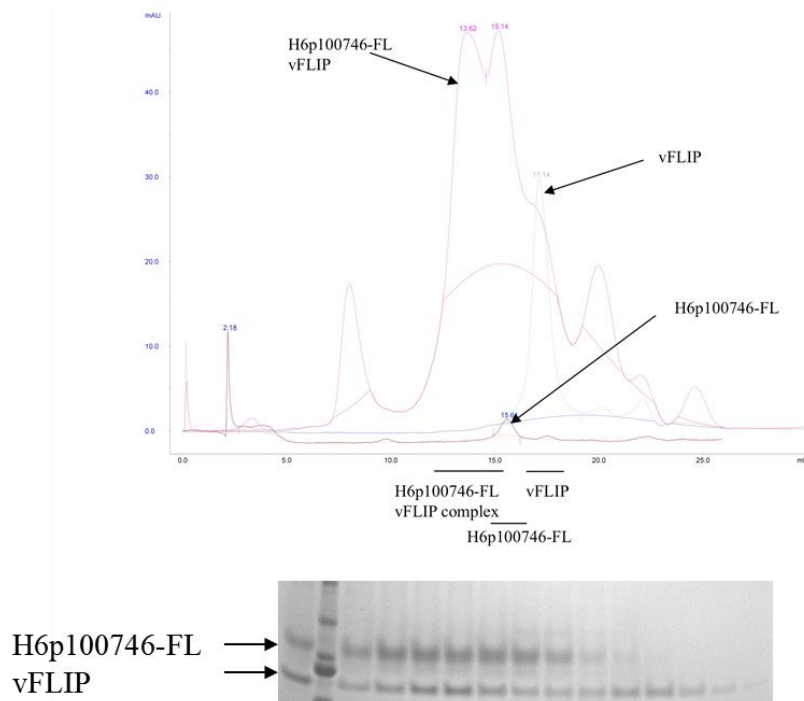


Figure 6.8: (A) Pull-down assay of p100 746-900 and KSHV-vFLIP using a 1 mL Histrap Column. (B) Gel filtration analysis of p100 746-900 following incubation of vFLIP. Both vFLIP and H6-p100 746-900 not only co-eluted, but shifted relative to their positions when analysed in isolation (Courtesy of Dr Claire Bagneris).

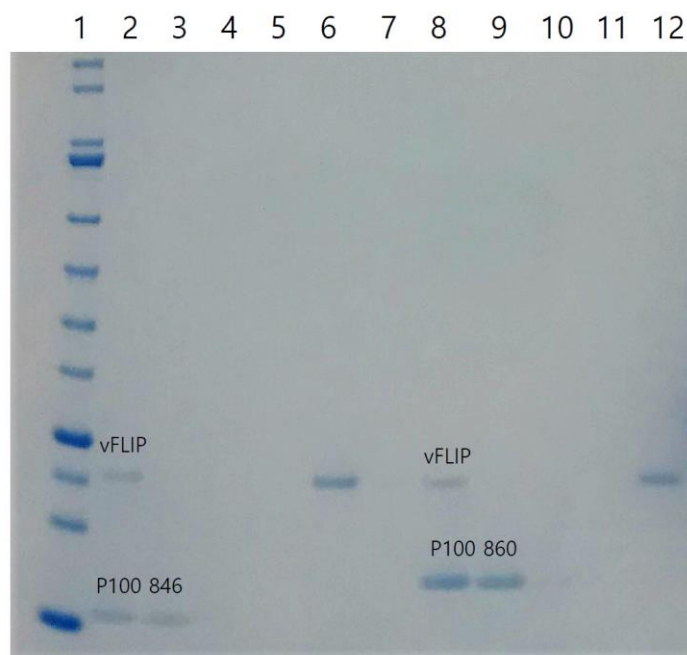


Figure 6.9: SDS-PAGE analysis of the pull-down assay to establish the location of the KSHV-vFLIP binding region of p100 using the constructs 746-846 and 746-860. Lane1: Molecular weight marker; Lane2: KSHV-vFLIP and p100 746-846 load sample; Lane3: Flow through; Lane4: Wash fraction; Lane5: Wash fraction containing 30 mM imidazole; Lane6: Elution fraction containing 500 mM imidazole; Lane7: Blank; Lane8: KSHV-vFLIP and P100 846-860 load sample; Lane9: Flow through; Lane10: Wash fraction; Lane11: Wash fraction containing 30 mM imidazole; Lane12: Elution fractions containing 500 mM imidazole.

As shown in Figure 6.9, both death domain constructs failed to bind to KSHV-vFLIP (His-tagged). Attempts to verify these results using pull down assays involving the p100 746-900 construct and GB1-vFLIP incorporating a C-terminal 6His-tag were unclear given similarities in molecular weight of the two proteins (results not shown). However, the lack of substantial quantities of KSHV-vFLIP eluting in the wash fractions coupled with the initial gel filtration studies shown in Figure 6.8

involving the p100 746-900 construct and KSHV-vFLIP indicate its capacity to form a complex unlike the shorter constructs. This is further confirmed using gel filtration in which p100 746-846 elutes in the same position as KSHV-vFLIP (Figure 6.10). The formation of a complex could be eliminated given that p100 764-846 elutes in the same position whether KSHV-vFLIP is present or absent.

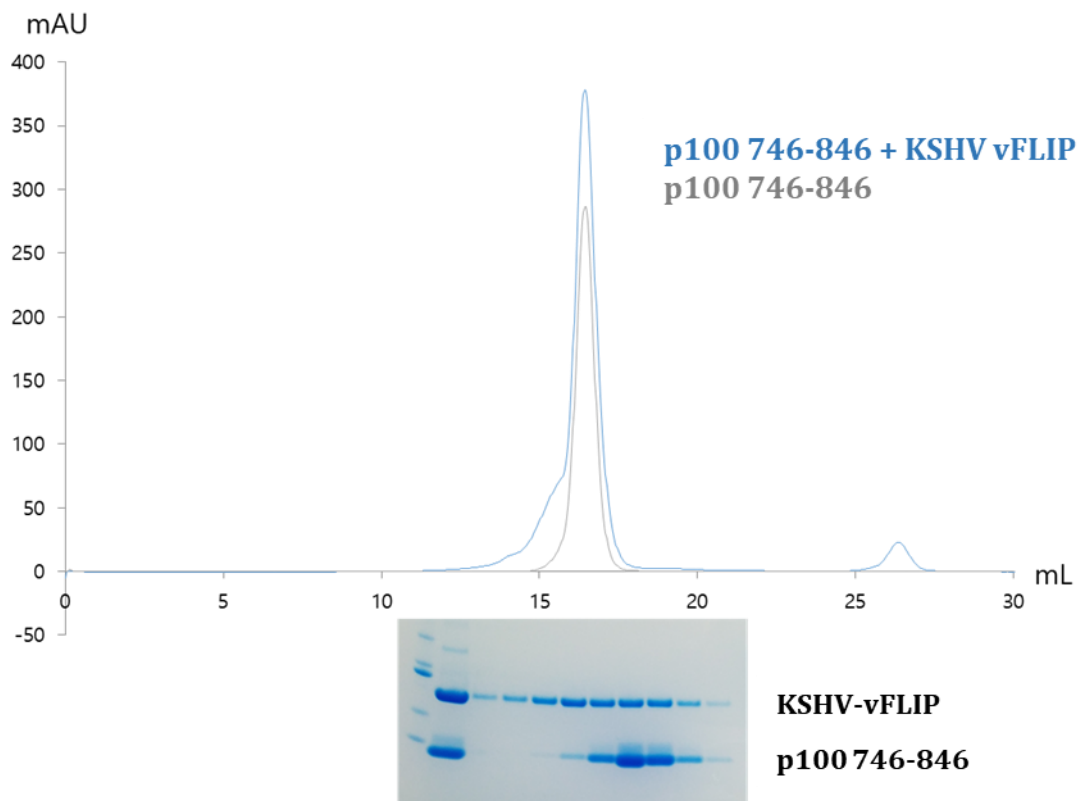


Figure 6.10: Gel filtration trace of p100 746-846 with (highlighted in blue) or without KSHV-vFLIP (highlighted in grey) and the SDS-PAGE analysis gel corresponds to p100 746-846 and KSHV-vFLIP gel filtration result.

6.6 Discussion

Whilst KSHV-vFLIP has been shown to be involved in upregulation of the canonical NF- κ B pathway, how it participates in non-canonical NF- κ B signalling is unclear. It was previously suggested that KSHV-vFLIP formed a ternary complex with the transcription factor p100 and IKK α (Matta and Chaudhary, 2004). Additionally, sequence analysis indicated a potential site of interaction towards the C-terminus of IKK α that bears weak homology to IKK γ , suggestive of a binding site for KSHV-vFLIP. Therefore, in this study, the interactions between p100, RelB, IKK α and KSHV-vFLIP were investigated using pull-down assays. Given that the capacity of these proteins to form a ternary complex was inconclusive owing to non-specific binding, trials involving RelB were perhaps more definitive in terms of the participation of KSHV-vFLIP. No KSHV-vFLIP was observed co-eluting with p100 in these trials suggesting possible occlusion of the KSHV-vFLIP binding site or alternatively, that binding of RelB favours dissociation of the KSHV-vFLIP from p100.

Although the exact structure of full-length p100 has yet to be elucidated, the C-terminal region incorporating the death domain has been solved where studies performed by our collaborators have indicated an important role in KSHV-vFLIP binding (Figure 1.19 in session 1.15). Therefore, in this study, three short constructs of p100 were expressed: p100 746-846 containing the death domain back lacking the far C-terminus, p100 746-860 (860 being one of the serines implicated in IKK α phosphorylation) and p100 746-900, incorporating the death domain and far C-terminus. Whilst it had been previously shown that KSHV-vFLIP binds to the p100

746-900 construct using pull-down assays and gel filtration, both p100 746-846 and 746-860 failed to associate with KSHV-vFLIP. Although there was the suggestion of possible co-elution of KSHV-vFLIP and p100 746-846, this could be discounted based on the gel filtration profile of 746-846 alone where it appears to naturally elute at this position. Furthermore, when compared to experiments performed with co-incubated KSHV-vFLIP and H6-p100 746-900, there is no shift towards higher molecular weight either of p100 746-846 or KSHV-vFLIP. Whilst p100 746-900 and KSHV-vFLIP bind, the stoichiometry appears not be 1:1 and may alternatively indicate a 2:1 ratio (p100:KSHV-vFLIP).

Although there are no known defined tertiary structures between residues 860 to 900 in p100, the region is predicted to mainly unstructured containing a short helical stretch (Figure 6.11). Studies have shown that phosphorylation induced by NF- κ B stimuli occurs at residue Ser866 and Ser870 which ultimately leads to p100 cleavage into p52 (Busino et al., 2012).

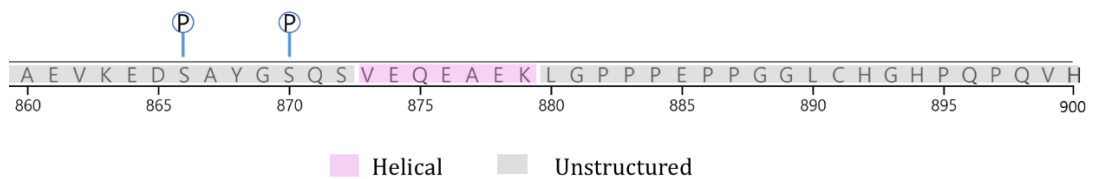


Figure 6.11: Amino acid sequence of p100 residues 860-900 and secondary structure predictions (Analysed and modified using JPRED, <http://www.compbio.dundee.ac.uk/jpred/>).

It has been shown that the death domain of p100 and the far C-terminus form an inhibitory motif along with the ankyrin repeats. This is thought to prevent the processing of p100 into p52 where the overall morphology of the full length protein has been proposed to resemble that of a horse-shoe with the N and C-termini in close proximity. This configuration is thought to disfavour phosphorylation of the C-terminal serines by IKK α . It can thus be envisaged that in the NIK dependent pathway, NIK and/or IKK α either alone or in concert act as a “wedge” to displace the C-terminus enabling not only phosphorylation but subsequent K48-linked ubiquitination and proteosomal degradation (Figure 6.12). It is possible that KSHV-vFLIP may be a functional analogue of NIK, but also that a combination of KSHV-vFLIP and IKK α are required for displacement of the C-terminal region. This could explain why KSHV-vFLIP associates very weakly with full length p100 when compared to the 746-900 construct. It may also explain why no detectable complex formation is observed between KSHV-vFLIP and p100 in the presence of RelB that may further occlude the KSHV-vFLIP binding site given that it associates with the N-terminal Rel homology domain that is in close proximity to the death domain and p100 C-terminus. Alternatively, the interaction between p100 and KSHV-vFLIP could only be transient, but nonetheless sufficient to enable the recruitment of IKK for the phosphorylation and downstream processing of p100. It also remains unclear whether KSHV-vFLIP may have a role in stimulating the phosphorylation and activation of IKK most likely through a process of autophosphorylation in the absence of co-activators.

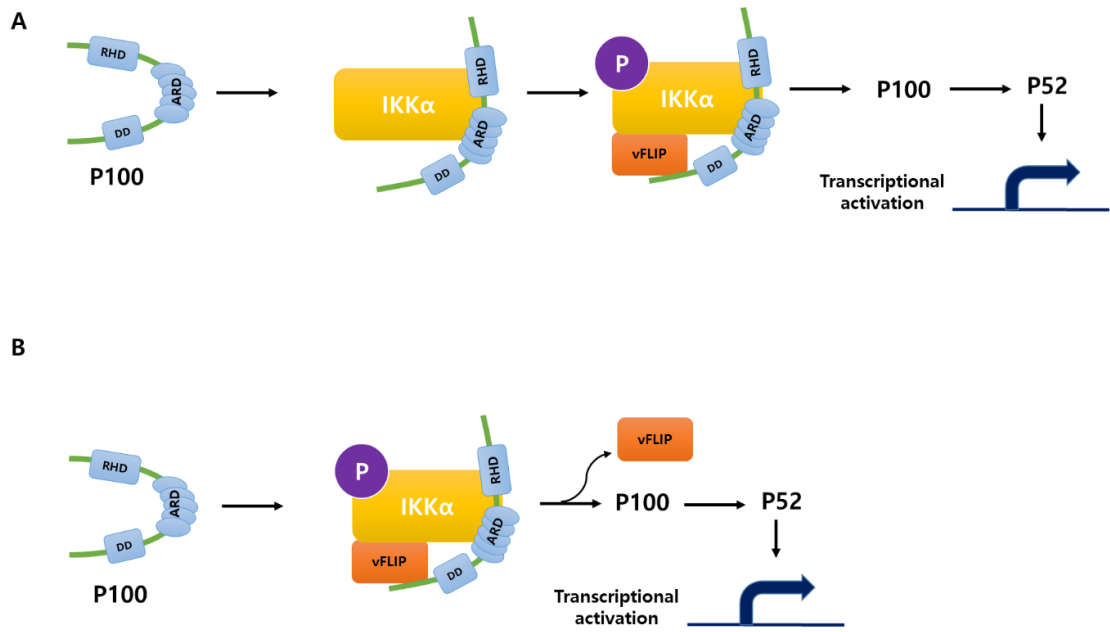


Figure 6.12: Schematic diagram of possible mechanisms of KSHV-vFLIP induced p100 processing. (A) The recruitment of IKK α and/or KSHV-vFLIP wedges open the horseshoe conformation of p100 and linearises it. This may lead to stimulation of IKK α 's kinase activity and subsequent phosphorylation. (B) KSHV-vFLIP only transiently interacts with p100 enabling IKK α to phosphorylate p100.

Future work

The crystal structure of SOX bound to RNA revealed residues that are important for catalysis and recognition of its substrates. As discussed previously, disruption of the SOX gene using siRNA still resulted in HSO which suggests that HSO may also require the co-option of other external factors for the efficient degradation of mRNA transcripts or perhaps indicate alternative, compensatory mechanisms. Several experiments can be carried out to investigate the binding modes of known cellular factors such as Xrn1 and KSHV SOX as well as their binding mode with substrates. Firstly, structural studies could be conducted aimed at trapping a substrate complex using non-cleavable RNA, i.e., DNA where the scissile phosphate has been replaced with a phosphothioate group for example, could be used with a catalytically inactive mutant. This could perhaps have greater insights into the cleavage mechanism. SOX-E244S mutant, although useful, was limited owing to its incapacity to associate with metal ions unlike glutamic acid. This would address the investigating potential complex formation between human Xrn1 and SOX along with possible consequences for cleavage in terms of specificity at the structure and sequence levels. Secondly, investigate how SOX is recruited to its RNA targets and the possible role of the host shut off mutants that were not defective in RNase activity for their capacity to mediate potential protein-protein interactions with other proteins implicated in HSO such as translation initiation and elongation factors as well as Xrn1 and exosome components which have been implicated based on interactome studies.

Lastly, investigate the role of the bridge in specificity and driving the observed patterns of cleavage using site-directed mutagenesis. This could involve mutating residues to make it shorter/longer or changing those that exist to more closely mirror BGLF5 or MHV68 for example.

Although phytic acid has been shown to associate with SOX both from the crystal structure and ThermoFluor assay, more quantitative results need to be obtained. Future work will therefore involve the development of a competition assay to obtain either a K_d or obtain IC50 (half maximal inhibitory concentration) of phytic acid towards SOX. Also, it will be important to establish if the effects of phytic acid on RNA and DNA binding are similar. Whilst both the structural and ThermoFluor studies indicate that the SOX-phytic acid interaction is weak given the concentrations required and the modest shifts in melting temperature observed, more potent compounds need to be identified. Since the interaction between SOX and phytic acid revealed the importance of residues that are not directly involved in catalysis, this provides a further avenue for future testing and possible exploitation. This is particularly important given the need to minimise potential cross-reactivity with cellular nucleases. In the first instance, it will be necessary to investigate the importance of interactions involving the bridge region. This can be addressed by the combination of mutagenesis and biophysical techniques such as isothermal titration calorimetry (ITC). The screening of small molecule compound libraries containing molecules similar to phytic acid (typically tropolones) may also prove to be useful future directions.

Future work for KSHV-vFLIP and P100 complex would involve resolving the various difficulties and pursuing techniques such as High-Throughput Mass Spectrometry to assess the phosphorylation state of P100 and complex stoichiometries. Phosphorylation of p100 could be directly measured given the mass shift of +79.99Da can be readily detected using this technique. Given that subversion of the non-canonical pathway is IKK complex independent, it would also be interesting to establish whether the IKK γ and p100 binding regions on KSHV-vFLIP are overlapping since this is currently unknown. Additionally, solving the structure of KSHV-vFLIP and P100 complex would help clarify the role of KSHV-vFLIP in subversion of the no- canonical NF- κ B pathway.

Summary and Conclusions

The studies detailed in this thesis have focussed on two proteins involved in the lytic and latent phases of KSHV infection, SOX and vFLIP. SOX, a dual function nuclease, is the initiator of host shut off that involves the rapid degradation of host mRNA transcripts, whilst KSHV-vFLIP operates to constitutively activate NF- κ B pathways. To gain further insights into how these proteins function, structural studies were performed on SOX bound to a pre-miRNA implicated in anoikis. These culminated in the first crystal structure of a viral nuclease bound to a pre-miRNA fragment that has shed light on the mechanism utilised for endonucleolytic processing. They have highlighted the importance of motifs conserved amongst the γ -herpesviruses and throughout the type II restriction endonuclease like family more widely, of which SOX is a member. Characterisation of the SOX-RNA complex also led to the identification of phytic acid as a possible weak inhibitor of SOX that can be used as a platform for the design of more potent compounds that possibly need to incorporate groups that make them selective for SOX. Major unanswered questions remain, however, such as the mechanism underlying the recruitment of SOX to target transcripts, sequence versus structure specificity and the participation of other host/viral factors that are key to a comprehensive understanding of the host shut off phenomenon. Similarly, whilst KSHV-vFLIP has been shown to be a key factor in not only viral pathogenesis but also survival of KSHV, little is known about how it explicitly subverts the non-canonical NF- κ B pathway. Key to this is its involvement

in the recruitment of IKK α and p100 that are essential factors in this signalling pathway. Whilst the nature of these various complexes could not be verified, the studies documented in this thesis have at least enabled the identification of the KSHV-vFLIP-p100 association site that is located towards the far C-terminus of p100 consistent with current theories on p100 topology. Overall, the structural and biochemical approaches described have provided a basis for future studies involving structural biology, cell biology and biochemical/biophysical characterisation with the potential to guide novel therapeutic intervention strategies given the link between SOX, vFLIP and KSHV pathogenesis.

Bibliography

- Abend**, J. R., Uldrick, T., Ziegelbauer, J. M. (2010). Regulation of tumor necrosis factor-like weak inducer of apoptosis receptor protein (TWEAKR) expression by Kaposi's sarcoma-associated herpesvirus microRNA prevents TWEAK-induced apoptosis and inflammatory cytokine expression. *J Virol* 84:12139–12151.
- Abend**, J. R., Ramalingam, D., Kieffer-Kwon, P., Uldrick, T. S., Yarchoan, R., Ziegelbauer, J. M. (2012). Kaposi's sarcoma-associated herpesvirus microRNAs target IRAK1 and MYD88, two components of the toll-like receptor/interleukin-1R signaling cascade, to reduce inflammatory-cytokine expression. *J Virol* 86:11663–11674.
- Akond**, G. M., Crawford, H., Berthold, J., Talukder, Z. I., Hossain, K. (2011). Minerals (Zn, Fe, Ca and Mg) and Antinutrient (Phytic Acid) Constituents in Common Bean. *Am J Food Technol.* 2011; 6(3): 235–243.
- AlMalki**, F. A., Flemming, C. S., Zhang, J., Feng, M., Sedelnikova, S. E., Ceska, T., Rafferty, J. B., Sayers, J. R. and Artymiuk, P. J. (2016). Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.* 23, 640–646.
- Ameres**, S. L., Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 14:475–488.
- Amir**, R. E., Haecker, H., Karin, M., Ciechanover, A. (2004). Mechanism of Processing of the NF-Kappa B2 P100 Precursor: Identification of the Specific Polyubiquitin Chain-Anchoring Lysine Residue and Analysis of the Role of Nedd8-Modification on the Scf (Beta-Trcp) Ubiquitin Ligase. *Oncogene*, 23, 2540–2547.
- Aneja** KK, Yuan Y. 2017. Reactivation and lytic replication of Kaposi's sarcoma-associated herpesvirus: an update. *Front Microbiol* 8:613.
- Arya**, M., Singh, P., Tripathi, C. B., Parashar, P., Singh, M., Kanoujia, J., Guleria, A., Kaithwas, G., Gupta, K. P., Saraf, S. A. (2019). Pectin-encrusted gold nanocomposites containing phytic acid and jacalin: 1,2-dimethylhydrazine-induced colon carcinogenesis in Wistar rats, PI3K/Akt, COX-2, and serum metabolomics as potential targets. *Drug Deliv Transl Res.* 9(1):53-65.
- AuCoin** DP, Colletti KS, Xu Y, Cei SA, Pari GS. (2002). Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) contains two functional lytic origins of DNA replication. *J Virol* 76:7890 –7896.
- Avitabile**, E, Forghieri, C., Campadelli-Fiume, G. (2009). Cross talk among the glycoproteins involved in herpes simplex virus entry and fusion: the interaction between gB and gH/gL does not necessarily require gD. *J Virol* 83:10752–10760.
- Bagneris**, C., Ageichik, A. V, Cronin, N., Wallace, B., Collins, M., Boshoff, C., Waksman, G., and Barrett, T. (2008). Crystal structure of a vFlip-IKKgamma complex: insights into viral activation of the IKK signalosome. *Mol. Cell* 30, 620–631.

- Bagneris**, C., Briggs, L. C., Savva, R., Ebrahimi, B. and Barrett, T. E. (2011). Crystal structure of a KSHV-SOX-DNA complex: insights into the molecular mechanisms underlying DNase activity and host shutoff. *Nucleic acids research* 39:5744-5756.
- Baidya**, N., Uhlenbeck, O. C. (1995). The role of 2'-Hydroxyl groups in an RNA-protein interaction. *Biochemistry*, 34:12363-12368.
- Ballestas**, M. E., Chatis, P. A., Kaye, K. M. (1999). Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. *Science* 284:641-644.
- Battye**, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallography*, D67, 271-281.
- Ben-Neriah**, Y. (2002). Regulatory functions of ubiquitination in the immune system. *Nat. Immunol.* 3, 20-26.
- Bloor**, S., Ryzhakov, G., Wagner, S., Butler, P.J.G., Smith, D.L., Krumbach, R., Dikic, I., and Randow, F. (2008). Signal processing by its coil zipper domain activates IKK gamma. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1279-1284
- Bonizzi**, G., and Karin, M. (2004). The two NF-kappaB activation pathways and their role in innate and adaptive immunity. *Trends Immunol.* 25, 280-288.
- Bravender**, T. (2010). Epstein-Barr virus, cytomegalovirus, and infectious mononucleosis. *Adolesc Med State Art Rev.* 21(2):251-64.
- Brewster**, A. S., Waterman, D. G., Parkhurst, J. M., Gildea, R. J., Young, I. D., O'Riordan, L. J., Yano, J., Winter, G., Evans, G., Sauter, N. K. (2018). Improving signal strength in serial crystallography with DIALS geometry refinement. *Acta Crystallogr D Struct Biol* 74, 877-894.
- Brunson**, J. L., Khoretonenko, M. V. & Stokes, K. Y. (2016). Chapter 10 - Herpesviruses. In *Vascular Responses to Pathogens*.
- Buisson**, M., Geoui, T., Flot, D., Tarbouriech, N., Rensing, M. E., Wiertz, E. J. and Burmeister, W. P. (2009). A bridge crosses the active-site canyon of the Epstein-Barr virus nuclease with DNase and RNase activities. *J. Mol. Biol.*, 391, 717-728.
- Bujnicki**, J. M., Rychlewski, L. (2001). The herpesvirus alkaline exonuclease belongs to the restriction endonuclease PD-(D/E)XK superfamily: insight from molecular modeling and phylogenetic analysis. *Virus Genes*, 22, 219-230.
- Cai**, Q., Verma, S. C., Lu, J., Robertson, E. S. (2010). Molecular biology of Kaposi's sarcoma-associated herpesvirus and related oncogenesis. *Adv Virus Res* 78: 87-142.
- Cesarman**, E., Damania, B., Krown, S., Martin, J., Bower, M., Whitby, D. (2019). Kaposi's sarcoma. *Nature Reviews*. 5:9.
- Chang**, H. H., Ganem, D. (2013). A unique herpesviral transcriptional program in KSHV-infected lymphatic endothelial cells leads to mTORC1 activation and rapamycin sensitivity. *Cell Host Microbe*.13:429-440.

- Chen, J.**, Ueda, K., Sakakibara, S., Okuno, T., Parravicini, C., Corbellino, M., Yamanishi, K. (2001). Activation of latent Kaposi's sarcoma-associated herpesvirus by demethylation of the promoter of the lytic transactivator. *Proc Natl Acad Sci USA* 98:4119–4124.
- Chen, V. B.**, Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66:12-21.
- Clyde, K.**, Glaunsinger, B. A. (2011). Deep sequencing reveals direct targets of gammaherpesvirus-induced mRNA decay and suggests that multiple mechanisms govern cellular transcript escape. *PLoS One* 6:e19655.
- Cohen, A.**, Wolf, D. G., Guttman-Yassky, E., Sarid, R. (2005). Kaposi's sarcoma-associated herpesvirus: clinical, diagnostic, and epidemiological aspects. *Crit Rev Clin Lab Sci* 42, 101–153.
- Covarrubias, S.**, Richner, J. M., Clyde, K., Lee, Y. J., Glaunsinger, B. A. (2009). Host shutoff is a conserved phenotype of gammaherpesvirus infection and is orchestrated exclusively from the cytoplasm. *J Virol* 83:9554 –9566.
- Covarrubias, S.**, Gaglia, M. M., Kumar, G. R., Wong, W., Jackson, A. O. and Glaunsinger, B. A. (2011). Coordinated destruction of cellular messages in translation complexes by the gammaherpesvirus host shutoff factor and the mammalian exonuclease Xrn1. *PLoS Pathog.*, 7, e1002339.
- Cruz-Muñoz, M. E.**, & Fuentes-Pananá, E. M. (2018). Beta and gamma human herpesviruses: Agonistic and antagonistic interactions with the host immune system. *Frontiers in Microbiology*, 8.
- Davis, D. A.**, Rinderknecht, A. S., Zoetewij, J. P., Aoki, Y., Read-Connole, E. L., Tosato, G., Blauvelt, A., Yarchoan, R. (2001). Hypoxia induces lytic replication of Kaposi sarcoma-associated herpesvirus. *Blood*, 97:3244–3250.
- Dahlroth, S. L.**, Gurmu, D., Schmitzberger, F., Engman, H., Haas, J., Erlandsen, H., Nordlund, P. (2009). Crystal structure of the shutoff and exonuclease protein from the oncogenic Kaposi's sarcoma-associated herpesvirus. *FEBS j* 276: 6636–6645.
- DiDonato, J.A.**, Hayakawa, M., Rothwarf, D.M., Zandi, E., and Karin, M. (1997). A cytokine-responsive IkappaB kinase that activates the transcription factor NF-kappaB. *Nature* 388, 548–554.
- Dittmer, D. P.** (2003). Transcription profile of Kaposi's sarcoma-associated herpesvirus in primary Kaposi's sarcoma lesions as determined by real-time PCR arrays. *Cancer Res* 63, 2010–2015.
- Drenth, J.** (2007). Principles of Protein X-Ray Crystallography, *New York, Springer*.

- Dubich**, T., Lieske, A., Santag, S., Beauclair, G., Ruchert, J., Herrmann, J., Gorges, J., Busche, G., Kazmaier, U., Hauser, H., Stadler, M., Schulz, T., Wirth, D. (2019). An endothelial cell line infected by Kaposi's sarcoma-associated herpesvirus (KSHV) allows the investigation of Kaposi's sarcoma and the validation of novel viral inhibitors in vitro and in vivo. *Journal of Molecular Medicine*. 97:311-324.
- Duerst**, R. J., and L. A. Morrison. (2004). Herpes simplex virus 2 virion host shutoff protein interferes with type I interferon production and responsiveness. *Virology* 322:158-167.
- Dutta**, D., Chakraborty, S., Bandyopadhyay, C., Valiya Veetil, M., Ansari, M. A., Singh, V. V., & Chandran, B. (2013). EphrinA2 Regulates Clathrin Mediated KSHV Endocytosis in Fibroblast Cells by Coordinating Integrin-Associated Signaling and c-Cbl Directed Polyubiquitination. *PLoS Pathogens*, 9(7).
- Ebrahimi**, B., Dutia, B. M., Roberts, K. L., Garcia-Ramirez, J. J., Dickinson, P., Stewart, J. P., Ghazal, P., Roy, D. J. & Nash, A. A. (2003). Transcriptome profile of murine gammaherpesvirus-68 lytic infection. *J Gen Virol* 84, 99–109.
- Emsley**, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66, 486-501.
- Everly**, D. N. Jr, Feng, P., Mian, I. S. Read, G. S. (2002). mRNA degradation by the virion host shutoff (Vhs) protein of herpes simplex virus: genetic and biochemical evidence that Vhs is a nuclease. *J Virol* 76, 8560–8571.
- Feng**, Y., Zhang, X., Graves, P. and Zeng, Y. (2012). A comprehensive analysis of precursor microRNA cleavage by human Dicer. *RNA*, 18, 2083–2092.
- Field**, N., Low, W., Daniels, M., Howell, S., Daviet, L., Boshoff, C., and Collins, M. (2003). KSHV vFLIP binds to IKK-gamma to activate IKK. *J. Cell Sci.* 116, 3721–3728.
- French**, S. & Wilson, K.S. (1978). "On the treatment of negative intensity observations." *Acta Cryst.* A34, 517-525.
- Ganem**, D. (2010). KSHV and the pathogenesis of Kaposi sarcoma: Listening to human biology and medicine. *Journal of Clinical Investigation*, 120(4), 939–949.
- Gaglia**, M. M., Covarrubias, S., Wong, W. and Glaunsinger, B. A. (2012). A common strategy for host RNA degradation by divergent viruses. *J. Virol.*, 86, 9527–9530.
- Gaglia**, M. M., Rycroft, C. H. and Glaunsinger, B. A. (2015). Transcriptome-wide cleavage site mapping on cellular mRNAs reveals features underlying sequence-specific cleavage by the viral ribonuclease SOX. *PLoS Pathog.*, 11, e1005305.
- Gasteiger**, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31(13):3784-8.
- Ghosh**, S., May, M.J., and Kopp, E.B. (1998). NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu. Rev. Immunol.* 16, 225–260.

- Ghosh, S., and Karin, M. (2002).** Missing pieces in the NF- κ B puzzle. *Cell*, Vol. 109:S81-S96.
- Gilmore, T.D. (2006).** Introduction to NF-kappaB: players, pathways, perspectives. *Oncogene* 25, 6680–6684.
- Glaunsinger, B., D, Ganem. (2004).** Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover. *Molecular cell* 13:713-723.
- Glaunsinger, B., Chavez, L. and D, Ganem. (2005).** The exonuclease and host shutoff functions of the SOX protein of Kaposi's sarcoma-associated herpesvirus are genetically separable. *Journal of virology* 79:7396-7401.
- Goldstein, J. N., and Weller, S. K. (1998).** The exonuclease activity of HSV-1 UL12 is required for *in vivo* function. *Virology* 244:442-457.
- Gradoville, L., Gerlach, J., Grogan, E., Shedd, D., Nikiforow, S., Metroka, C., Miller, G. (2000).** Kaposi's sarcoma-associated herpesvirus open reading frame 50/Rta protein activates the entire viral lytic cycle in the HH-B2 primary effusion lymphoma cell line. *J Virol* 74:6207–6212.
- Grundhoff, A., Sullivan, C.S., Ganem, D. (2006).** A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. *RNA*, 12:1-18.
- Guba, M., von Breitenbuch, P., Steinbauer, M., Koehl, G., Flegel, S., Hornung, M., Bruns, C. J., Zuelke, C., Farkas, S., Anthuber, M., Jauch, K. W., Geissler, E. K. (2002).** Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: involvement of vascular endothelial growth factor. *Nat Med* 8:128–135.
- Hahn, A. S., Kaufmann, J. K., Wies, E., Naschberger, E., Panteleev-Ivlev, J., Schmidt, K., Holzer, A., Schmidt, M., Chen, J., Konig, S., Ensser, A., Myoung, J., Brockmeyer, N. H., Sturzl, M., Fleckenstein, B., Neipel, F. (2012).** The ephrin receptor tyrosine kinase A2 is a cellular receptor for Kaposi's sarcoma-associated herpesvirus. *Nat Med* 18:961–966.
- Haque, M., Davis, D. A., Wang, V., Widmer, I., Yarchoan, R. (2003).** Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) contains hypoxia response elements: relevance to lytic induction by hypoxia. *J Virol* 77:6761–6768.
- Hayden, M.S., and Ghosh, S. (2004).** Signaling to NF-kappaB. *Genes Dev.* 18, 2195–2224.
- Hayden, M.S., and Ghosh, S. (2008).** Shared principles in NF-kappaB signaling. *Cell* 132, 344–362.
- Hayden, Matthew S, & Ghosh, S. (2012).** NF- κ B, the first quarter-century: remarkable progress and outstanding questions. *Genes & development*, 26(3), 203–34.
- Hensler, H. R., Tomaszewski, M. J., Rappocciolo, G., Rinaldo, C. R., Jenkins, F. J. (2014).** Human herpesvirus 8 glycoprotein B binds the entry receptor DC-SIGN. *Virus Res* 190:97–103.

- Hendrickson**, D. G., Hogan D. J., McCullough, H. L. Myers, J. W. Herschlag, D. Ferrell, J. E and Brown, P. O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS biology* 7:e1000238
- Hoffmann**, a, Natoli, G., and Ghosh, G. (2006). Transcriptional regulation via the NFkappaB signaling module. *Oncogene* 25, 6706–6716.
- Horst**, D., Burmeister,W.P., Boer,I.G., van Leeuwen,D., Buisson,M., Gorbalenya,A.E., Wiertz,E.J. and Ressing,M.E., (2012). The ‘Bridge’ in the Epstein-Barr virus alkaline exonuclease protein BGLF5 contributes to shutoff activity during productive infection. *J. Virol.*, 86, 9175–9187.
- Horton**, N. C., Connolly, B. A., Perona., J. J. (2000). Inhibition of EcoRV Endonuclease by Deoxyribo-3'-S-phosphorothiolates: A High-Resolution X-ray Crystallographic Study. *J. Am. Chem. Soc.* 122, 3314-3324.
- Hosseini**pour, M. C and other 16 authors. (2014). Viral profiling identifies multiple subtypes of Kaposi's sarcoma. *mBio* 5, e01633–01614.
- Hu**, M., Wang, C., Li, W., Lu, W., Bai, Z., Qin, D., Yan, Q., Zhu, J., Krueger, B. J., Renne, R., Gao, S. J., Lu, C. (2015). A KSHV microRNA directly targets G protein-coupled receptor kinase 2 to promote the migration and invasion of endothelial cells by inducing CXCR113 and activating AKT signaling. *PLoS Pathog* 11:e1005171.
- Hutin**, S., Lee, Y. and Glaunsinger, B .A. (2013). An RNA element in human interleukin 6 confers escape from degradation by the gammaherpesvirus SOX protein. *J. Virol.*, 87, 4672–4682.
- Huynh**, K., Partch, C. L. (2015). Analysis of protein stability and ligand interactions by thermal shift assay. *Current Protocols in Protein Science.* 79:28.9.1-28.9.14.
- Jameson**, D. M., Sawyer, W. H. (1995). Fluorescence anisotropy applied to biomolecular interactions. *Methods Enzymology*, 246, 283-300.
- Jariwalla**, R. J., Sabin, R., Lawson, S., Bloch, D. A., Prender, M., Andrews, V. & Herman, Z. S. (1988). Effects of dietary phytic acid (phytate) on the incidence and growth rate of tumors promoted in Fisher rats by a magnesium supplement. *Nutr. Res.* 8: 813–827.
- Jarousse**, N., Chandran, B., Coscoy, L, (2008). Lack of heparan sulfate expression in B-cell lines: implications for Kaposi's sarcoma-associated herpesvirus and murine gammaherpesvirus 68 infections. *J Virol* 82 (24), 12591–12597.
- Jenner**, R. G., Alba, M. M., Boshoff, C., Kellam, P. (2001). Kaposi's sarcoma-associated herpesvirus latent and lytic gene expression as revealed by DNA arrays. *J Virol* 75, 891–902.
- Joy**, A. Balaji, S. (2015). Drug-likeness of phytic acid and its analogues. *Open Microbiology Journal.* 9:141-149.
- Kabsch**, W. (2009). XDS. *Acta Crystallographica Section D: Biological Crystallography*, D66, 125-132.

- Karin, M.**, and Ben-Neriah, Y. (2000). Phosphorylation meets ubiquitination: the control of NF- κ B activity. *Annu. Rev. Immunol.* 18, 621–663.
- Karin, M.**, Cao, Y., Greten, F.R., and Li, Z.-W. (2002). NF- κ B in cancer: from innocent bystander to major culprit. *Nat. Rev. Cancer* 2, 301–310.
- Kaplan, L. D.** (2013). *Kaplan-2013-Human herpesvirus-8.pdf*. 103–108.
- Katano, H.**, Sato, Y., Kurata, T., Mori, S., and Sata, T. (2000). Expression and localization of human herpesvirus 8-encoded proteins in primary effusion lymphoma, Kaposi's sarcoma, and multicentric Castleman's disease. *Virology* 269, 335–344.
- Kieffer-Kwon, P.**, Happel, C., Uldrick, T. S., Ramalingam, D., Ziegelbauer, J. M. (2015). KSHV microRNAs repress tropomyosin 1 and increase anchorage-independent growth and endothelial tube formation. *PLoS ONE* 10:e0135560.
- Krishnan, H. H.**, Sharma-Walia, N., Zeng, L., Gao, S. J., Chandran, B. (2005) Envelope glycoprotein gB of Kaposi's sarcoma-associated herpesvirus is essential for egress from infected cells. *J Virol* 79:10952–10967.
- Kronstad, L. M.** Glaunsinger, B. A. (2012). Diverse virus–host interactions influence RNA-based regulation during g-herpesvirus infection. *Current Opinion in Microbiology*, 15, 506–511.
- Kwong, A.D.**, Frenkel, N. (1987). Herpes simplex virus-infected cells contain a function(s) that destabilizes both host and viral mRNAs, *Proc. Natl. Acad. Sci. U. S. A.* 84 1926–1930.
- Laskowski, R. A.**, Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26, 283–291.
- Lee. H.** (2015). Crystal structures of a Shutoff and Exonuclease in complex with DNA and RNA provide insights into the processes of host mRNA degradation and viral maturation by Kaposi's sarcoma-associated herpesvirus. Thesis presented for the degree of Master of Research in Bioscience. *Institute of Structural and Molecular Biology*. UCL.
- Lee. H.**, Patschull, A., Bagneris, C., Ryan, H., Sanderson, C., Ebrahimi, B., Nobeli, I., and Barrett, T. (2017). KSHV SOX mediated host shutoff: the molecular mechanism underlying mRNA transcript processing. *Nucleic Acids Research*. Doi:10.1093/nar/lkx1340
- Li, W.**, Jia, X. M., Shen, C., Zhang, M., Xu, J., Shang, Y., Zhu, K., Hu, M., Yan, Q., Qin, D., Lee, M., Zhu, J., Lu, H., Krueger, B. J., Renne, R., Gao, S. J., Lu, C. (2016). A KSHV microRNA enhances viral latency and induces angiogenesis by targeting GRK2 to activate the CXCR164/AKT pathway. *Oncotarget* 7:32286–32305.
- Liang, C.**, Zhang, M., Sun, S. C. (2006). β -TrCP binding and processing of NF- κ B2/p100 involve its phosphorylation at serines 866 and 870. *Cell Signal*.18:1309–1317.

- Liang**, D., Gao, Y., Lin, X., He, Z., Zhao, Q., Deng, Q., Lan, K. (2011). A human herpesvirus miRNA attenuates interferon signaling and contributes to maintenance of viral latency by targeting IKK epsilon. *Cell Res* 21:793–806.
- Lieberman**, P. M. (2013). Keeping it quite: chromatin control of gammaherpesvirus latency. *Nat Rev Microbiol.* 11(12):863-875.
- Lin**, X., Liang, D., He, Z., Deng, Q., Robertson, E., Lan, K. (2011). miR-K12- 7-5p encoded by Kaposi's sarcoma-associated herpesvirus stabilizes the latent state by targeting viral ORF50/RTA. *PLoS ONE* 6:e16224.
- Liou**, H.C., Sha, W.C., Scott, M.L., and Baltimore, D. (1994). Sequential induction of NFkappa B/Rel family proteins during B-cell terminal differentiation. *Mol. Cell. Biol.* 14, 5349–5359.
- Liu**, F., Xia, Y., Parker, A.S., and Verma, I.M. (2012). IKK biology. *Immunol. Rev.* 246, 239–253.
- Lo**, Y., Lin, S., Rospigliosi, C. C., Conze, D. B., Wu, C., Ashwell, J. D., Eliezer, D., Wu, H. (2009). Structural basis for recognition of diubiquitins by NEMO. *Mol Cell.* 33(5): 602–615.
- Lu**, C. C., Li, Z., Chu, C. Y., Feng, J., Feng, J., Sun, R., Rana, T. M. (2010). MicroRNAs encoded by Kaposi's sarcoma-associated herpesvirus regulate viral life cycle. *EMBO Rep* 11:784–790.
- Maquat**, L. E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* 5(2):89-99.
- Martinez**, R., Sarisky, R. T., Weber, P. C. and Weller, S. K. (1996). Herpes simplex virus type 1 alkaline nuclease is required for efficient processing of viral DNA replication intermediates. *J. Virol.*, 70, 2075–2085.
- Matta**, H., Mazzacurati, L., Schamus, S., Yang, T., Sun, Q., and Chaudhary, P.M. (2007). Kaposi's sarcoma-associated herpesvirus (KSHV) oncoprotein K13 bypasses TRAFs and directly interacts with the IkappaB kinase complex to selectively activate NF-kappaB without JNK activation. *J. Biol. Chem.* 282, 24858–24865.
- May**, M.J. (2000). Selective Inhibition of NF-kappa B Activation by a Peptide That Blocks the Interaction of NEMO with the Ikappa B Kinase Complex. *Science* (80) 289, 1550–1554.
- May**, M.J., Marienfeld, R.B., and Ghosh, S. (2002). Characterization of the Ikappa Bkinase NEMO binding domain. *J. Biol. Chem.* 277, 45992–46000.
- McCoy**, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, C., L. & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658–674.
- Menniti**, F. S., Oliver, K. G., Pytney, J. W., Jr. & Shears, S. B.(1993). Inositol phosphates and cell signalling: new view of InsP5 and InsP6. *Trends Biochem. Sci.* 18: 53–65.

- Mendez**, A. S., Vogt, C., Bohne, J., Glaunsinger, B. A. (2018). Site specific target binding controls RNA cleavage efficiency by the Kaposi's sarcoma-associated herpesvirus endonuclease SOX. *Nucleic Acids Research*. Doi:10.1093/nar/ky932.
- Mettenleiter**, T. C. (2002). Herpesvirus assembly and egress. *J Virol* 76: 1537–1547. <https://doi.org/10.1128/JVI.76.4.1537-1547.2002>.
- Mettenleiter**, T. C., Klupp, B. G., & Granzow, H. (2009). Herpesvirus assembly: An update. *Virus Research*. <https://doi.org/10.1016/j.virusres.2009.03.018>
- Migliazza**, A., Lombardi, L., Rocchi, M., Trecca, D., Chang, C. C., Antonacci, R., Fracchiolla, N. S., Ciana, P., Maiolo, A. T., Neri, A. (1994). Heterogeneous chromosomal aberrations generate 3'truncations of the NFKB2/lyt-10 gene in lymphoid malignancies. *Blood*. 84:3850-3860.
- Mikhailov**, V. S., Okano, K. and Rohrmann, G. F. (2004). Specificity of the endonuclease activity of the baculovirus alkaline nuclease for single-stranded DNA. *J. Biol. Chem.*, 279, 14734–14745.
- Mocroft**, A., Kirk, O., Clumeck, N., Gargalianos-Kakolyris, P., Trocha, H., Chentsova, N., Antunes, F., Stellbrink, H. J., Phillips, A. N. & Lundgren, J. D. (2004). The changing pattern of Kaposi sarcoma in patients with HIV, 1994-2003: *the EuroSIDA Study*. *Cancer* 100, 2644-2654.
- Morin**, B., Coutard B., M, Lelke, F. Ferron, R. Kerber, S. Jamal, A. Frangeul, C. Baronti, R. Charrel, X. de Lamballerie, C. Vonrhein, J. Lescar, G. Bricogne, S. Gunther, B. Canard. (2010). The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription, *PLoS Pathog.* 6 e1001038.
- Murshudov**, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F and Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):355{367, 2011.
- Nagy**, E., Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends In Biochemical Sciences*. 23:198–199.
- Neipel**, F., Albrecht, J. C., Fleckenstein, B. (1997). Cell-homologous genes in the Kaposi's sarcoma-associated Rhadinovirus Human Herpesvirus 8: determinants of its pathogenicity? *J Virol* 71:4187–4192.
- Nishino**, T., Ishino, Y. and Morikawa, K. (2006). Structure-specific DNA nucleases: structural basis for 3D-scissors. *Curr. Opin. Struct. Biol.* 16, 60–67.
- Oeckinghaus**, A., and Ghosh, S. (2009). The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harb. Perspect. Biol.* 1, a000034.

- Otake**, T., Shimonaka, H., Kanai, M., Miyano, K., Ueba, N., Kunita, N., Kurimura, T. (1989). Inhibitory effect of inositol hexasulfate and inositol hexaphosphoric acid (phytic acid) on the proliferation of the human immunodeficiency virus (HIV) in vitro. *J-Stage*. 63(7):676-683.
- Pahl**, H. L. (1999). Activators and target genes of Rel/NF-kappaB transcription factors. *Oncogene*. 18(49):6853-66.
- Pertel**, P. E. (2002) Human herpesvirus 8 glycoprotein B (gB), gH, and gL can mediate cell fusion. *J Virol* 76:4390-4400.
- Pingoud**, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, 62, 685-707.
- Pollard**, T. D. (2010). A guide to simple and informative binding assays. *Molecular Biology of the Cell*, 21, 4061-4067.
- Polley**, S., Passos, D. O., Huang, D., Mulero, M. C., Mazumder, A., Biswas, T., Verma, I. M., Lyumkis, D., Ghosh, G. (2016). Structural basis for the activation of IKK1/a. *Cell Reports* 17, 1907-1914.
- Pfeffer**, S., Sewer, A., Lagos-Quintana, M. & 16 other authors. (2005). Identification of microRNAs of the herpesvirus family. *Nat Methods* 2, 269-276.
- Qin**, J., Li, W., Gao, S. J., Lu, C. (2017). KSHV microRNAs: tricks of the devil. *Trends Microbiol* 25:648-661.
- Ramachandran**, G. N. & Sasiskharan, V. (1968). Conformation of polypeptides and proteins. *Advances in Protein Chemistry*, 23, 283-437.
- Ran**, R., Lu, A., Zhang, L., Tang, Y., Zhu, H., Xu, H., Feng, Y., Han, C., Zhou, G., Rigby, A.C., et al. (2004). Hsp70 promotes TNF-mediated apoptosis by binding IKK gamma and impairing NF-kappa B survival signaling. *Genes Dev*. 18, 1466-1481.
- Rappocciolo**, G., Jenkins, F. J., Hensler, H. R., Piazza, P., Jais, M., Borowski, L., Watkins, S. C., Rinaldo, C. R. (2006). DC-SIGN is a receptor for human herpesvirus 8 on dendritic cells and macrophages. *J Immunol* 176:1741-1749.
- Razani**, B., Reichardt, A.D., and Cheng, G. (2011). Non-canonical NF-κB signaling activation and regulation: principles and perspectives. *Immunol. Rev.* 244, 44-54.
- Raeisi**, D., Madani, S. H., Zare, M. E. (2013) Kaposi' s sarcoma after kidney transplantation: a 21-years experience. *Int J Hematol Oncol Stem Cell Res*. 7(4):29-33.
- Rezaee**. S. A. R., Cunnigham. C., Davison. A. J., Blackbourn. D. J. (2006). Kaposi's sarcoma-associated herpesvirus immune modulation: an overview. *Journal of general virology*. 87, 1781-1804.
- Rhodes**, G. (2006). Crystallography Made Crystal Clear .*Oxford, Academic Press*.

- Rowe, M., B. Glaunsinger, D. van Leeuwen, J. Zuo, D. Sweetman, D. Ganem, J. Middeldorp, E. J. Wiertz, and M. E. Rensing. (2007).** Host shutoff during productive Epstein-Barr virus infection is mediated by BGLF5 and may contribute to immune evasion. *Proceedings of the National Academy of Sciences of the United States of America* 104:3366-3371.
- Roizman, T. (1965).** An inquiry into the mechanisms of recurrent herpes infections of man. Perspectives in *Virology* 4, 7-83.
- Roy, D., Sin, S. H., Lucas, A., Venkataramanan, R., Wang, L., Eason, A., Chavakula, V., Hilton, I. B., Tamburro, K. M., Damania, B., Dittmer, D. P. (2013).** mTOR inhibitors block Kaposi sarcoma growth by inhibiting essential autocrine growth factors and tumor angiogenesis. *Cancer Res* 73:2235–2246.
- Rupp, B. (2010).** Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology, *New York, Garland Science*.
- Russo, J. J., Bohenzky, R. A., Chien, M. C., Chen, J., Yan, M., Maddalena, D., Parry, J. P., Peruzzi, D., Edelman, I. S., Chang, Y., Moore, P. (1996).** Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc Natl Acad Sci USA* .93:14862–14867.
- Sandri-Goldin, R. M. (2011).** The many roles of the highly interactive HSV protein ICP27, a key regulator of infection. *Future Microbiology* 6:11.
- Scheidereit, C. (2006).** I κ B kinase complexes: gateways to NF- κ B activation and transcription. *Oncogene* 25, 6685–6705.
- Schrofelbauer, B., Polley, S., Behar, M., Ghosh, G., Hoffmann, A. (2012).** NEMO ensures signaling specificity of the pleiotropic IKK β by directing its kinase activity towards I κ B α . *Mol Cell*. 47(1):111-121.
- Senftleben, U., Cao, Y., Xiao, G., Greten, F.R., Krahn, G., Bonizzi, G., Chen, Y., Hu, Y., Fong, a, Sun, S.C., et al. (2001).** Activation by IKK α of a second, evolutionary conserved, NF- κ B signaling pathway. *Science* 293, 1495–1499.
- Sciabica, K. S., Dai, Q. J., & Sandri-Goldin, R. M. (2003).** ICP27 interacts with SRPK1 to mediate HSV splicing inhibition by altering SR protein phosphorylation. *EMBO Journal*, 22(7), 1608–1619. <https://doi.org/10.1093/emboj/cdg166>.
- Sen, R., and Smale, S.T. (2009).** Selectivity of the NF- κ B Response. *Cold Spring Harb. Perspect. Biol.* 2, a000257–a000257.
- Severini, A., Morgan, A. R., Tovell, D. R. and Tyrrell, D. L. (1994).** Study of the structure of replicative intermediates of HSV-1 DNA by pulsed-field gel electrophoresis. *Virology*, 200, 428–435.
- Severini, A., Scraba, D. G. and Tyrrell, D. L. (1996).** Branched structures in the intracellular DNA of herpes simplex virus type 1. *J. Virol.*, 70, 3169–3175.
- Shamsuddin, A. M., Ulah, A. (1989).** Inositol hexaphosphate inhibits large intestinal cancer in F344 rats 5 months after induction by azoxymethane. *Carcinogenesis* 10: 625–626.

- Shamsuddin**, A. M., Baten, A. & Lalwani, N. D. (1992). Effect of inositol hexaphosphate on growth and differentiation in K562 erythroleukemia cell line. *Cancer Lett.* 64: 195–202.
- Shamsuddin**, A. M. (2002). Anti-cancer function of phytic acid. *Int. J. Food Sci. Technol.* 37: 769–782.
- Shi**, X. & Herschlag, D. (2009). Fluorescence Polarization Anisotropy to Measure RNA Dynamics. *Methods in Enzymology*, 469, 287-302.
- Sil**, A.K., Maeda, S., Sano, Y., Roop, D.R., and Karin, M. (2004). IkappaB kinase-alpha acts in the epidermis to control skeletal and craniofacial morphogenesis. *Nature* 428, 660–664.
- Singh**, K. K., Rucker, T., Hanne, A., Parwaresch, R. & Krupp, G. (2000). Fluorescence polarization for monitoring ribozyme reactions in real time. *Biotechniques* 29, 344–348, 350-351.
- Skalsky**, R. L., Hu, J., Renne, R. (2007). Analysis of viral cis elements conferring Kaposi's sarcoma-associated herpesvirus episome partitioning and maintenance. *J Virol* 81:9825–9837.
- Smiley**, J. R. (2004). Herpes simplex virus virion host shutoff protein: immune evasion mediated by a viral RNase? *Journal of virology* 78:1063-1068.
- Smith**, T. J., L. A. Morrison, and D. A. Leib. (2002). Pathogenesis of herpes simplex virus type 2 virion host shutoff (vhs) mutants. *Journal of virology* 76:2054-2061.
- Solan**, N. J., Miyoshi, H., Carmona, E. M., Bren, G. D., Paya, C. V. (2002). RelB cellular regulation and transcriptional activity are regulated by p100. *J Biol Chem.* 277:1405-1418.
- Stallone**, G., Schena, A., Infante, B., di Paolo, S., Loverre, A., Maggio, G., Ranieri, E., Gesualdo, L., Schena, F. P., Grandaliano, G. (2005). Sirolimus for Kaposi's sarcoma in renal-transplant recipients. *N Engl J Med* 352:1317–1323.
- Staskus**, K. A., Zhong, W., Gebhard, K., Herndier, B., Wang, H., Renne, R., Beneke, J., Pudney, J., Anderson, D. J., Ganem, D., Haase, A. T. (1997) Kaposi's sarcoma-associated herpesvirus gene expression in endothelial (spindle) tumor cells. *J Virol* 71:715–719.
- Strelow**, L. I., and D. A. Leib. (1995). Role of the virion host shutoff (vhs) of herpes simplex virus type 1 in latency and pathogenesis. *Journal of virology* 69:6779-6786.
- Sullivan**, C. S. (2007). High conservation of Kaposi sarcoma-associated herpesvirus microRNAs implies important function. *J Infect Dis* 195:618–620.
- Sun**, S. (2017). The non-canonical NF-kB pathway in immunity and inflammation. *Nat Rev Immunol.* 17(9):545-558. Doi:10.1038/nri.2017.52.

- Tao, Z., Fusco, A., Huang, D. B., Gupta, K., Young Kim, D., C. F. Ware, G. D. Van Duyne, G. Ghosh.** (2014). p100/IkBd sequesters and inhibits NF- κ B through kappaBsome formation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15946–15951.
- Thome, M., Tschopp, J.** (2001). Regulation of lymphocyte proliferation and death by FLIP. *Nat Rev Immunol.* 1(1):50-8.
- Thompson, L. U. & Zhang, L.** (1991). Phytic acid and minerals: effect of early markers of risk for mammary and colon carcinogenesis. *Carcinogenesis* 12: 2041–2045.
- Thompson, S. R., and P. Sarnow.** (2003). Enterovirus 71 contains a type I IRES element that functions when eukaryotic initiation factor eIF4G is cleaved. *Virology* 315:259-266.
- Tolani, B., Matta, H., Gopalakrishnan, R., Punj, V., and Chaudhary, P.M.** (2014). NEMO is essential for Kaposi's sarcoma-associated herpesvirus-encoded vFLIP K13-induced gene expression and protection against death receptor-induced cell death, and its N-terminal 251 residues are sufficient for this process. *J. Virol.* 88, 6345–6354.
- Tomlinson, C.G., Atack, J.M., Chapados, B., Tainer, J. A. and Grasby, J.A.** (2010). Substrate recognition and catalysis by flap endonucleases and related enzymes. *Biochem. Soc. Trans.*, 38, 433–437.
- Tran, H. D., Brooks, J., Gadwal, S., Bryant, J. L., Shamsuddin, A. M., Lunardi-Iskandar, Y.&Vucenik, I.** (2003). Effect of inositol hexaphosphate (IP6) on AIDS neoplastic Kaposi's sarcoma, iatrogenic Kaposi's sarcoma and lymphoma. *Proc.Am.Assoc Cancer Res.* 44:499.
- Uldrick, T. S., Wang, V., Mahony, D. O., Aleman, K., Kathleen, M., Marshall, V., ... Yarchoan, R.** (2010). An interleukin-6-related systemic inflammatory syndrome in patients coinfecting with Kaposi sarcoma-associated herpesvirus and HIV but without multicentric Castleman disease. *NIH Public Access*, 51(3), 350–358.
- Union for International Cancer Control.** (2014). Review of cancer medicines on the WHO List of Essential Medicines: Kaposi's sarcoma.
- Uppal, T., Jha, H. C., Verma, S. C., Robertson, E. S.** (2015). Chromatinization of the KSHV genome during the KSHV life cycle. *Cancers (Basel)*, 7 pp. 112-142.
- Uppal, T., Meyer, D., Agrawal, A., Verma, S. C.** (2019). The DNase activity of Kaposi's Sarcoma-associated Herpesvirus SOX protein serves an important role in viral genome processing during lytic replication. *Journal of virology.* 93(8):e01983-18
- Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., Murshudov, G. N.** (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D*, 60(12 Part 1):2184-2195.
- Veetil, M. V., Bandyopadhyay, C., Dutta, D., Chandran, B.** (2014). Interaction of KSHV with host cell surface receptors and cell entry. *Viruses* 6:4024–4046.

- Vucenik, I., Shamsuddin, A. M. (2003).** Cancer inhibition by inositol hexaphosphate (IP6) and inositol: from laboratory to clinic. *American Society for Nutritional Sciences*. 133:3778S-3784S.
- Vucenik, I., Stains, J. (2010).** Cancer preventive and therapeutic properties of IP6: Efficacy and mechanisms. *Period Biol.* 112:4.
- Wang, N., Baldi, P. F., & Gaut, B. S. (2007).** Phylogenetic analysis, genome evolution and the rate of gene gain in the Herpesviridae. *Molecular Phylogenetics and Evolution*, 43(3), 1066–1075.
- Weber, G. (1953).** Rotational Brownian motion and polarization of the fluorescence of solution. *Advances in Protein Chemistry*, 8, 415-459.
- Wen, K. W., Damania, B. (2010).** Kaposi sarcoma-associated herpesvirus (KSHV): molecular biology and oncogenesis. *Cancer Letters*. 289, 140-150.
- Wu, C. (1999).** beta -TrCP Mediates the Signal-induced Ubiquitination of Ikappa Bbeta. *Journal of Biological Chemistry*, 274(42), 29591–29594.
- Xiao, G., Harhaj, E.W., and Sun, S.C. (2001).** NF-kappaB-inducing kinase regulates the processing of NF-kappaB2 p100. *Mol. Cell* 7, 401–409.
- Xiao, G., Fong, A., and Sun, S.-C. (2004).** Induction of p100 processing by NF-kappaB-inducing kinase involves docking IkappaB kinase alpha (IKKalpha) to p100 and IKKalpha-mediated phosphorylation. *J. Biol. Chem.* 279, 30099–30105.
- Xu, G., Lo, Y.-C., Li, Q., Napolitano, G., Wu, X., Jiang, X., Dreano, M., Karin, M., and Wu, H. (2011).** Crystal structure of inhibitor of kappaB kinase beta. *Nature* 472, 325–330.
- Yan, L., Majerciak, V., Zheng, Z., Lan, K. (2019).** Towards better understanding of KSHV life cycle: from transcription and posttranscriptional regulations to pathogenesis. *Virologica Sinica*. 34:135-161.
- Yang, W. (2011).** Nucleases: diversity of structure, function and mechanism. *Q Rev. Biophys.*, 44(Pt 1), 1–93.
- Yang, W., Lee, J. Y. and Nowotny, M. (2006).** Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol. Cell*, 22, 5–13.
- Ye, F., Zhou, F., Bedolla, R. G., Jones, T., Lei, X., Kang, T., Guadalupe, M., Gao, S. J. (2011).** Reactive oxygen species hydrogen peroxide mediates Kaposi's sarcoma-associated herpesvirus reactivation from latency. *PLoS Pathog* 7:e1002054.
- Yilmaz, Z. B., Kofahl, B., Beaudette, P., Baum, I. Ipenberg, F. Weih, J. Wolf, G. Dittmar, C. Scheidereit. (2014).** Quantitative dissection and modeling of the NF-kB p100-p105 module reveals interdependent precursor proteolysis. *Cell Rep.* 9, 1756–1769.
- Yuan, P., Bartlam, M., Z. Lou, S. Chen, J. Zhou, X. He, Z. Lv, R. Ge, X. Li, T. Deng, E. Fodor, Z. Rao, Y. Liu. (2009).** Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site, *Nature* 458 909–913.

- Zhang, Q., Lenardo, M. J., Baltimore, D. (2017).** 30 years of NF- κ B: A blossoming of relevance to human pathobiology. *Cell*, 168.
- Zheng, C., Yin, Q., and Wu, H. (2011).** Structural studies of NF- κ B signaling. *Cell Res.* 21, 183–195.
- Zhong, H., Voll, R.E., and Ghosh, S. (1998).** Phosphorylation of NF-Bp65 by PKA stimulates transcriptional activity by promoting a novel bivalent interaction with the coactivator CBP/p300. *Mol. Cell*, 661–671.
- Ziegelbauer, J. M., Sullivan, C. S., Ganem, D. (2009).** Tandem array-based expression screens identify host mRNA targets of virus-encoded microRNAs. *Nat Genet* 41:130–134.