Structural and mechanistic studies of herpes virus proteins

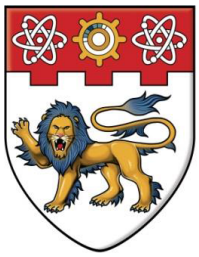
Patel, Vidhi Kanaiyalal

2015

Patel, V. K. (2015). Structural and mechanistic studies of herpes virus proteins. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/65385

https://doi.org/10.32657/10356/65385



NANYANG TECHNOLOGICAL UNIVERSITY

STRUCTURAL AND MECHANISTIC STUDIES OF HERPES VIRUS PROTEINS

VIDHI KANAIYALAL PATEL

SCHOOL OF BIOLOGICAL SCIENCES

2015

STRUCTURAL AND MECHANISTIC STUDIES OF HERPES VIRUS PROTEINS VIDHI K. PATEL 2015

STRUCTURAL AND MECHANISTIC STUDIES OF HERPES VIRUS PROTEINS

VIDHI KANAIYALAL PATEL

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2015

STRUCTURAL AND MECHANISTIC STUDIES OF HERPES VIRUS PROTEINS VIDHI K. PATEL 2015

ACKNOWLEDGEMENTS

The journey of every post-graduate student is like that of an ocean tide, with the highs and lows, and the twists and turns. But we need to remain afloat and accomplish our goals. This journey cannot be completed without the support of people surrounding us and I would like this opportunity to thank those who have helped me achieve my goals.

I am, first and foremost, grateful to my doctoral advisor, Professor Pär Nordlund, for providing me the opportunity to work with him. His expertise on the subject has been a learning experience for me. His views on protein structure and how each structure tells a story have helped me view structural biology in a different perspective.

I am wholeheartedly grateful to Dr. Venkatachalam Rajakannan (Raja) and Dr. Sue-Li Dahlroth for their guidance and time. I especially appreciate Raja's patience in guiding me through crystallography and structure determination of ICP27. His expertise has helped me immensely in achieving the targets of my project. I also appreciate Sue-Li's knowledge in viral protein purification and molecular biology, and her help in laying the foundations of my projects. I would like to thank all the other members of the PN group, especially Dr. Anna Jansson, Dr. Andreas Larsson for their expert inputs and the tips and tricks they shared that have always been helpful. I'm grateful to my fellow graduate students, Kelly, Dan, Rachel and Lekshmy, for all their support during the times of need. The synchrotron trips would have been dull without their presence. I especially thank Kelly for the times she has helped me keep up my morale. I am exceptionally thankful to Dr. Fariborz Nasertorabi for sharing his experiences on working with herpes proteins, for all the informal discussions on science, biology, and crystallography. His enthusiasm in doing science has encouraged me to continue ahead.

I would like to immensely thank the Protein Production Platform for all their help that was needed in achieving the goals of the multi-construct approach project. I would like to extend my thanks to Agnes, Dina and Ignacio. When I faced any problem related to cloning or protein purification or use of the various instruments, they were always ready to help, unreservedly.

I am grateful to my thesis advisory committee members, Dr. Curt Davey and Dr. Koh Cheng Gee, for their invaluable inputs on the projects and guidance. I am also extremely

indebted to all my previous mentors: my mentors from the Microbiology department (Ramnarain Ruia College), especially Dr. Ravi Phadke, Urmi Palan ma'am and Leena Phadke ma'am, and from the Life Sciences department (University of Mumbai), Dr. Sivakami and Dr. Anju George, all of whom have influenced me immensely to develop a scientific perspective and always encouraged me in my journey. I am immensely thankful to Dr. Sonal Srikanth from the David Geffen School of Medicine (University of California, LA), who has helped me in all possible ways to pursue science.

My term as post-graduate student would have been acutely difficult had it not been for my dearest friends. Being with them has helped me keep my sanity and keep me moving forward. I am extremely indebted to Kripa, my roommate, partner in biology discussions and my best friend; Aditya, my Zen-guru, partner in discussions on life and an extremely patient better-half; Heidi, my partner in arts and in biology discussions; Payal, who has been available, unconditionally, every single time I have needed her; Ram, who is responsible for all my travels across South-east Asia and is always just a phone-call away; and Vivek, my partner in all activities and in discussing any topic under the sun. Each of them has played a significant role in boosting my morale, especially during my trying times. I would also take this opportunity to thank all my friends from #ApnaGang, who, in spite of being away, have always been a source of encouragement and discussions on various topics.

I am immensely filled with gratitude towards my family, especially my parents, who have supported every step in my life as a post-graduate student. They have been loving, caring, understanding and completely unconditional for the entire term of my study in Singapore. I am also grateful to my brother, Nirav, who has always supported me in all my decisions and always been ready to help when needed.

I would like to extend my thanks to Nanyang Technological University for providing me the opportunity to pursue my studies. I also acknowledge the financial support provided by the NTU Research Scholarship. I am also thankful to all the SBS office employees, especially May, Eugene and Raymond, for their prompt help whenever needed.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION TO THE HUMAN HERPESVIRUSES

2. MULTIFUNCTIONAL REGULATORY PROTEINS OF HUMAN HERPESVIRUSES

4. RESULTS

LIST OF TABLES

1. INTRODUCTION TO THE HUMAN HERPESVIRUSES

2. MULTIFUNCTIONAL REGULATORY PROTEINS OF HUMAN HERPESVIRUSES

3. MATERIALS AND METHODS

4. RESULTS

ABBREVIATIONS

CD	Circular Dichroism
CTD	C-terminal domain
DNA	DeoxyriboNucleic acid
EBV	Epstein-Barr Virus
eIF	eukaryotic Initiation Factor
GC-rich	Guaninosine / Cytosine-rich
HCMV	Human CytoMegalo Virus
HHV	Human Herpes Virus
His6-Tag	hexaHistidine-tag
HSV	Herpes Simplex Virus
ICP27 / 8	Infectious Cell Protein 27 / 8
IMAC	Ion-Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KH domain	K-Homology domain
KSHV	Kaposi's Sarcoma-associated Herpes Virus
LEX system	Large scale EXpression System
LIC	Ligation Independent Cloning
MAD	Multiple Anomalous Dispersion
MWCO	Molecular Weight Cut-Off
NES	Nuclear Export Signal
Ni-NTA	Nickel-Nitrilotriacetic Acid
NLS	Nuclear Localization Signal
NMR	Nuclear Magnetic Resonance
PAPB	Poly(A)-Binding Proteins
PEG	PolyEthylene Gycol
Poly-(A) tail	Poly-Adenylated tail
REF	RNA Export Factor
RI	Refractive Index
RNA	RiboNucleic Acid
RNAP II	RNA polymerase II
RRM	RNA recognition motif
SAD	Single Anomalous Dispersion
SDS-PAGE	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SeMet	SelenoMethionine
SR proteins	Ser/Arg-rich proteins
SRPK1	SR-Protein Kinase 1
TCEP	(Tris(2-CarboxyEthyl)Phosphine)
TEV-tag	Tobacco Etch Virus- tag
TREX complex	TRanscription Export Complex
UV	UltraViolet
VP16	Viral Protein 16
VZV	Varicella Zooster Virus
Zn	Zinc

ABSTRACT

Herpes viruses are nuclear-replicating viruses, known to cause a variety of diseases in humans ranging from painful skin and genital lesions to life-threatening cancers. They have successfully evolved to evade the immune system of humans, establishing lifelong infections. Herpes Simplex Viruses (HSV) 1 and 2 are the most prevalent amongst the eight known human herpes viruses (HHV). All herpes viruses have a characteristic life cycle, divided into latent and lytic phases. They cause recurrent infections long after the primary infection, leading to more stress to the patients. Though evolutionarily each of the HHV has diverged from each other, key survival processes are still similar.

Infectious Cell Protein 27 (ICP27) from HSV is a prototype of a multifunctional regulatory protein that is highly conserved in all HHVs. It is a key regulatory protein of the HSV and has the potential to interact with an array of cellular proteins and viral intronless mRNAs. ICP27 plays an essential role in viral transcription, nuclear export of intronless mRNAs, translation of viral transcripts and virion host shut-off function. The only other well-characterized functional homolog of ICP27 is ORF57 from Kaposi's Sarcoma-associated herpes Virus (KSHV). Other homologues, especially EB2 from Epstein-Barr Virus (EBV) and ORF69 from Human Cytomegalovirus (HCMV), are being studied currently but not yet characterized in detail.

In spite of its central role in viral replication and infection, very little is known about the structure and mechanistic properties of ICP27 and its homologues. We present the first crystal structure of ICP27 at a resolution of 2.0 Å. The structure reveals the C-terminal half of ICP27 to be a bundle of α-helices along with a unique zinc-fingerlike motif. The overall fold of protein suggests it to be novel compared to protein structures available in the Protein Data Bank. We present data that confirm ICP27 exists as dimer in solution, using biophysical assays. Regulation of ICP27 by phosphorylation of the C-terminal domain has also been probed. In vitro phosphorylation assays reveal the residues that may be involved in regulation of ICP27. We also present preliminary data on EB2 protein from EBV. From sequence alignments and secondary structural analyses, we demonstrate that EB2 is not only functionally similar, but probably structurally too. Further probing will be needed to analyse if all the homologues are structurally similar in spite of divergence in

sequence similarity. Structural knowledge of ICP27 and its homologues may aid in a better understanding of the mechanism by which they function. It can give a better insight to how the virus survives and replicates in host cells. Development of drugs against the herpes viral infections in humans may be guided more effectively.

Herpes viral proteins have posed several challenges during trials for their expression and purification on a large scale. In this thesis, we try to outline the multi-construct approach that has been successful for human protein expression. This strategy is applied to herpes proteins and we observe that the hurdles in herpes protein expression can be circumvented to a certain extent. We present data on 94 different herpes proteins and the success we obtained in expressing otherwise difficult proteins. Few of these proteins have been successfully expressed, purified, crystallized and their structures determined to a good resolution, including ICP27. The multi-construct approach has the potential to drive viral structural biology forward and boost our knowledge in this otherwise fledgling field.

1. INTRODUCTION TO THE HUMAN HERPESVIRUSES

The word '***Herpes***' is derived from the Greek word '***herpein***', meaning 'to crawl', referring to the spreading nature of the skin lesions. Mankind has known of herpes infections since ancient times. However, it was only in the early 20th century that research on herpes accelerated. From an evolutionary perspective, herpes viruses seem to have successfully expanded their host range in numerous vertebrates including humans, and also in oysters, an invertebrate (Davison, 2002).

1.1. THE HERPESVIRIDAE FAMILY / HUMAN HERPESVIRUSES

Herpesviridae is a large family of double stranded (ds) DNA viruses, having a linear genome of 125-240 kbp, enclosed in an icosahedral protein capsid of about 125-130 nm in diameter (differs for different viral species). A lipid bilayer envelope derived from the host membrane with embedded viral glycoproteins, in turn, covers the capsid. In between the capsid and the envelope, lie the amorphous tegument regions consisting of over thirty proteins (Figure 1.1) (Davison, 2002; Davison 2007).

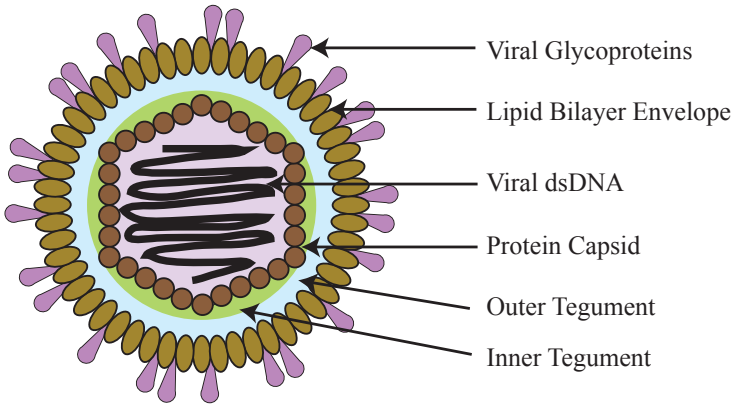


Figure 1.1 Schematic structure of the human herpesvirus.

The general structure of herpes virus consists of dsDNA genome packaged within an icosahedral protein capsid. The capsid is coated with several viral proteins that form the inner and outer tegument regions. The assembly is then covered with the lipid bilayer that is derived from the host cell membrane. Embedded in this bilayer are some of viral glycoproteins that help in attaching to and fusing with the target host cell membranes.

The *Herpesviridae* family is classified into 3 sub-families, depending on the cell-type of the host they reside in during the latency period: *Alphaherpesvirinae* (neurons), *Betaherpesvirinae* (monocyte lineage) and *Gammaherpesvirinae* (lymphocytes) (Whitley, 1996; Davison, 2007). The schematic in Figure 1.2 shows the classification of the eight known Human Herpesviruses (HHV) into these three sub-families.

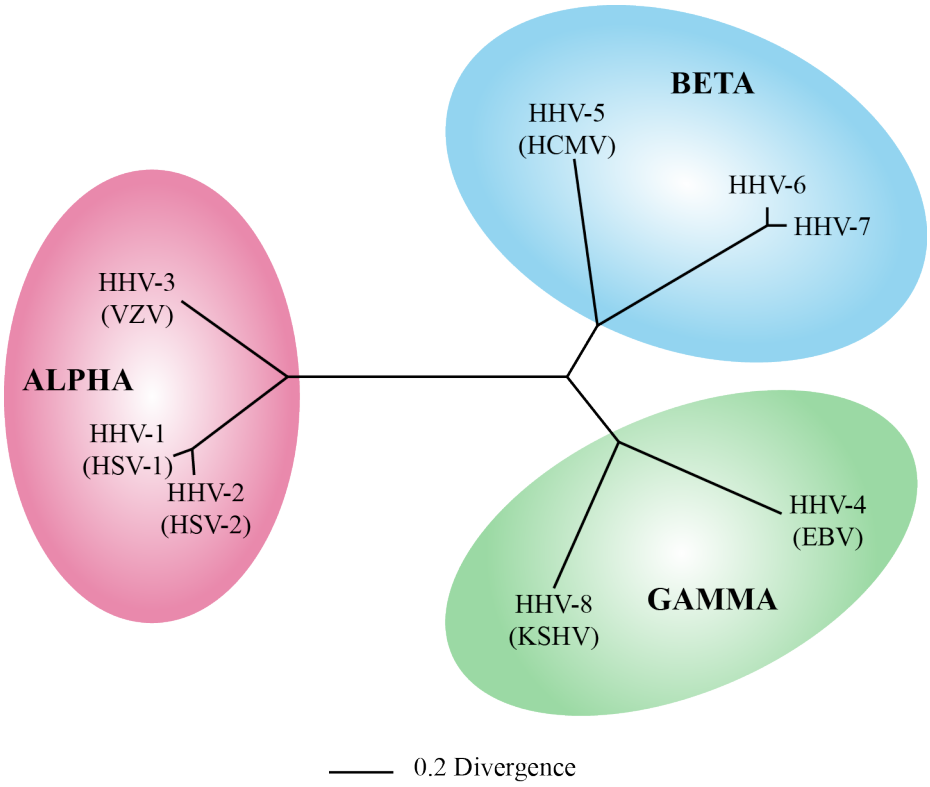


Figure 1.2 Human *Herpesviridae* subfamily tree.

Three Subfamilies of Herpes viruses showing the eight known Human Herpesviruses (HHV)

α: *Alphaherpesvirinae* β: *Betaherpesvirinae* γ: *Gammaherpesvirinae*

HSV- Herpes Simplex Virus; VZV- Varicella Zoster Virus; HCMV- Human Cytomegalo Virus; EBV- Epstein-Barr Virus; KSHV- Kaposi Sarcoma-associated Herpes Virus

(Adapted from McGeoch et al., 1995)

Herpes viral infections are prevalent worldwide, in both developing and developed countries. These infections spread amongst individuals by close contact between the person shedding the virus and a susceptible host (Whitley, 1996). The virus introduces itself into a host at the point of entry causing primary infection. It then undergoes lytic cycle and spreads to other cells, where it can establish latency and remain within the host system for prolonged periods of time without being noticed by

host immune system. The virus genome is maintained in an extra-chromosomal state. On trigger, the virus enters lytic state again, resulting in recurring infection. This is very well documented in case of the Human Herpesvirus 3 (or more commonly known as Varicella Zoster Virus) infection where the virus enters the susceptible host, generally children, causes Chicken Pox and remains within the nervous system for years before it re-emerges in the form of Shingles (Whitley, 1996; Davison, 2007). (Table 1.1) summarizes the major diseases caused by the eight known Human Herpesviruses. However, many additional diseases are also attributed to the HHV, which are not mentioned here.

Table 1.1: Major Clinical Diseases caused by 8 Human Herpesviruses

SUBFAMILY	HUMAN HERPESVIRUS*	MAJOR CLINICAL DISEASES
Alphaherpesvirinae	HHV-1 (Herpes Simplex Virus – 1)	Mucocutaneous infections, Genital herpes
	HHV-2 (Herpes Simplex Virus – 2)	Mucocutaneous infections, Genital herpes
	HHV-3 (Varicella Zoster Virus)	Chicken pox, Shingles
Betaherpesvirinae	HHV-5 (Human Cytomegalo Virus)	Mononucleosis syndrome, Congenital Cytomegalic inclusion disease, CMV pneumonia esp. in immune-compromised patients
	HHV-6	Mild early childhood Roseola
	HHV-7	Mild early childhood Roseola
Gammaherpesvirinae	HHV-4 (Epstein-Barr Virus)	Mononucleosis syndrome, Hepatitis, Cancers like Burkitt's lymphoma and Hodgkin's lymphoma
	HHV-8 (Kaposi's Sarcoma-associated Herpes Virus)	Cancers like Kaposi's Sarcoma and Primary effusion lymphoma

* Common names written in brackets

1.2. HERPES SIMPLEX VIRUS

Of the eight known human herpesviruses, herpes simplex virus (HSV) 1 and 2 are the most prevalent epidemiologically. HSV-1, belonging to the *Alphaherpesvirinae* sub-family, is generally considered a prototype of the *Herpesviridae* family and extensive studies are performed on HSV-1 to understand the basic functioning of the other

HHVs. HSV infections range from mild subcutaneous conditions such as cold sores to more serious infections like genital herpes. HSV is a leading cause of genital ulcers and can be transferred vertically from mother to child either during the act of giving birth or postpartum via close contact with infected mother (Anzivino et al, 2009; Pinninti et al., 2013). HSV establishes its primary infection in epithelial cells and spreads to neurons, wherein it establishes latent state, before reactivation.

HSV-1 and HSV-2 are similar in genetic structure and have 40-50 % homology (Kieff et al., 1972). HSV infects a variety of cells via different mediators (Karasneh & Shukla, 2011). Upon infection, the dsDNA genome is transported to the host nucleus, where it remains dormant till triggered for lytic cycle. The 80 viral genes are divided into 3 temporal groups, depending on sequence of their expression: Immediate Early (IE), Early (E) and Late (L) genes. The IE genes have been extensively studied since they are responsible for the expression of the early and late stage genes (Deluca, Mccarthy, & Schaffer, 1985; Fontaine-Rodriguez & Knipe, 2008; Jean, LeVan, Song, Levine, & Knipe, 2001; Rice & Knipe, 1990). A virion trans-activator protein VP16 in turn controls the IE gene expression. The genes encoding ICP0, ICP4, ICP22, ICP27 and ICP47 constitute the IE genes in HSV.

1.3. EPSTEIN-BARR VIRUS

Epstein-Barr virus (EBV) is a widely studied HHV from the *Gammaherpesvirinae* sub-family along with Kaposi's sarcoma-associated herpes virus (KSHV). It was the first known virus to cause tumours in humans and is associated with lympho-proliferative diseases, Burkitt's lymphoma, Hodgkin's disease and many more lymphomas. There are 2 EBV types: EBV-1 and EBV-2; both share a high level of homology that is greater than the one between HSV-1 and HSV-2. The virus infects children and can persist in the human body in a latent, asymptomatic state for years. At the onset of infection, the B-cells undergo proliferation leading to symptomatic infectious mononucleosis. EBV infects B-cells and epithelial cells. It can remain latent in these cells and can be reactivated into lytic state. The triggers for this switch are unknown (Sergeant, Gruffat, & Manet, 2008).

1.4. THERAPEUTICS AGAINST HERPES INFECTIONS

Vaccines have been a highly effective means to prevent various diseases, with those against smallpox and poliovirus being among the prominent ones. Herpes viruses have proven to be extremely difficult when it comes to vaccine development (Belshe et al., 2012). With approximately 90 % of the world's population infected with at least one herpes virus, effective vaccines and drugs against herpes infections are a must, especially in cases of neonates and immuno-compromised patients (Xu et al., 2002). A recent study also hints at co-prevalence of HSV-1, EBV and HCMV in children and women belonging to certain ethnicities (Delaney et al., 2014). Currently, vaccines are available only against VZV infections and even these can be administered only to healthy children and not immune-compromised individuals. New efforts are being made to assess the efficacy of the vaccine in the elderly (Schmader et al., 2012). Drugs like Acyclovir, Valaciclovir, Famiclovir, Ganciclovir and Foscarnet are available against some herpes infections; however HSV-1 and -2 seem to be developing resistance to the drugs. Prospective vaccines have been identified against EBV, HCMV and KSHV, and efforts are being channelled towards their development (Coen et al., 2014; Cohen et al., 2013; Cui et al., 2013; Griffiths et al., 2013; Wu et al., 2012). However, we still face numerous challenges in vaccine and drug development against human herpesviruses. Understanding the mode of infections of herpes viruses, and the structure and function of proteins involved in these processes can pave a path for us to develop improved drugs and protective vaccines.

1.5. MULTI-CONSTRUCT APPROACH FOR EXPRESSION OF RECOMBINANT HERPES VIRAL PROTEINS

Herpes viruses and the infections they cause have been studied widely. Some anti-viral therapies have been developed over time, though not against all diseases. In spite of the extensive research, we lack an insight on the molecular mechanisms of many of the key viral proteins. Systematic tapping of this knowledge can provide new avenues for effective drug development. Structural biology combined with biophysical techniques is widely used to establish the molecular mechanisms of proteins. These techniques require pure, homogenous and well-folded proteins at high

concentrations. Bacterial system, specifically *Escherichia coli*, has been widely used for large-scale expression of proteins and has been successful for less challenging proteins. However, viral and eukaryotic proteins often fail in standard expression trials in *E. coli* and more sophisticated techniques, such as multi-construct approach or eukaryotic expression systems, are required.

Expression of recombinant viral proteins in bacterial systems, including those from herpes viruses, have always posed many difficulties (Bowman et al., 2006; Fogg et al., 2006; Henson et al., 2011; Le Sage et al., 2013; Sam et al., 2009; Tait & Straus, 2011; Tarbouriech et al., 2006). Viral proteins have been proposed to be biophysically different from eukaryotic and bacterial proteins, and having intrinsically disordered regions (Tokuriki et al., 2009; Xue, et al., 2012). Tokuriki et al. have linked this property to the adaptation of archebacterial proteins having extremely compact structures, but only opposite in nature with viral proteins having loose, flexible structures (Tokuriki et al., 2009). Additionally, Bahir et al. have suggested that viral proteins may even have extreme codon preferences, restricting to the host they replicate in (Bahir et al., 2009). Such highly flexible proteins are difficult express and purify due to their strict codon usage and high susceptibility to protease activity. Gräslund *et al.* have described a systematic approach to express human proteins, wherein they make multiple constructs covering the domain(s) of interest of the protein (Gräslund et al., 2008). This ensured that the well-folded domains of the protein could be expressed and purified much more successfully than the attempts to purify whole proteins. This approach increased the chances of obtaining pure, homogenous protein samples from bacterial expression systems. The proteins thus obtained are often successfully used for biophysical techniques as well as for structural studies. Similar multi-construct approach has been used for expression of the challenging herpes viral proteins. This concept has been briefly touched upon in this thesis and aims to provide an insight on the strategies to improve viral protein production.

2. MULTIFUNCTIONAL REGULATORY PROTEINS OF HUMAN HERPES VIRUSES

The focus of this thesis has been to shed light on the structure and function of a key immediate-early (IE) protein of the herpes viruses. ICP27 from Herpes Simplex Virus-1 is one of the few non-structural proteins that is conserved across the *Herpesviridae* family and has known homologues in all eight human herpes viruses. The homologues include IE4 in VZV, UL69 in HCMV, EB2 protein in EBV and ORF57 in KSHV. ICP27 and ORF57 are most widely studied amongst the ICP27 family. Though they all share similar functions, the overall sequence homology is less than 25 %. This chapter describes the various functions of ICP27 and EB2.

2.1. INTRODUCTION TO ICP27 AND EB2

The UL54 gene of HSV-1 encodes Infectious Cell Protein 27 (ICP27). This protein of 512 amino acids is absolutely necessary for the virus to establish infection in the host. It shuttles between nucleus and cytoplasm, across the nuclear membrane carrying along the intronless viral mRNAs. It is modified post-translationally by phosphorylation and Arg-methylation events that regulate its function through the course of infection (Mears & Rice, 1996; Souki et al., 2009; Souki & Sandri-Goldin, 2009; Zhi & Sandri-Goldin, 1999). ICP27 has numerous interacting partners both within the nucleus as well as the cytoplasm, which will be briefly described in the following sections. *In vivo*, *in vitro* and bioinformatics studies have determined the various domains of ICP27 depicted in Figure 2.1 A.

The BMLF1 gene of EBV encodes the EB2 protein. It consists of 479 amino acids (Chevallier-Greco et al., 1986). Similar to ICP27, EB2 accumulates the intronless viral mRNAs in cytoplasm (Semmes et al., 1998). It is regulated by phosphorylation and displays interactions with various proteins (Medina-Palazon et al., 2007). This protein is not as widely studied as ICP27 but some of its interacting partners have been identified. The domain structure of EB2 is described in Figure 2.1 B.

Open Reading Frame 57 (ORF57) protein from Kaposi's Sarcoma-associated Herpes Virus (KSHV) is also studied on a wide scale. However this thesis is primarily focused on ICP27 and EB2. Since many functions of EB2 overlap with those of ICP27, both the proteins will be discussed in detail together.

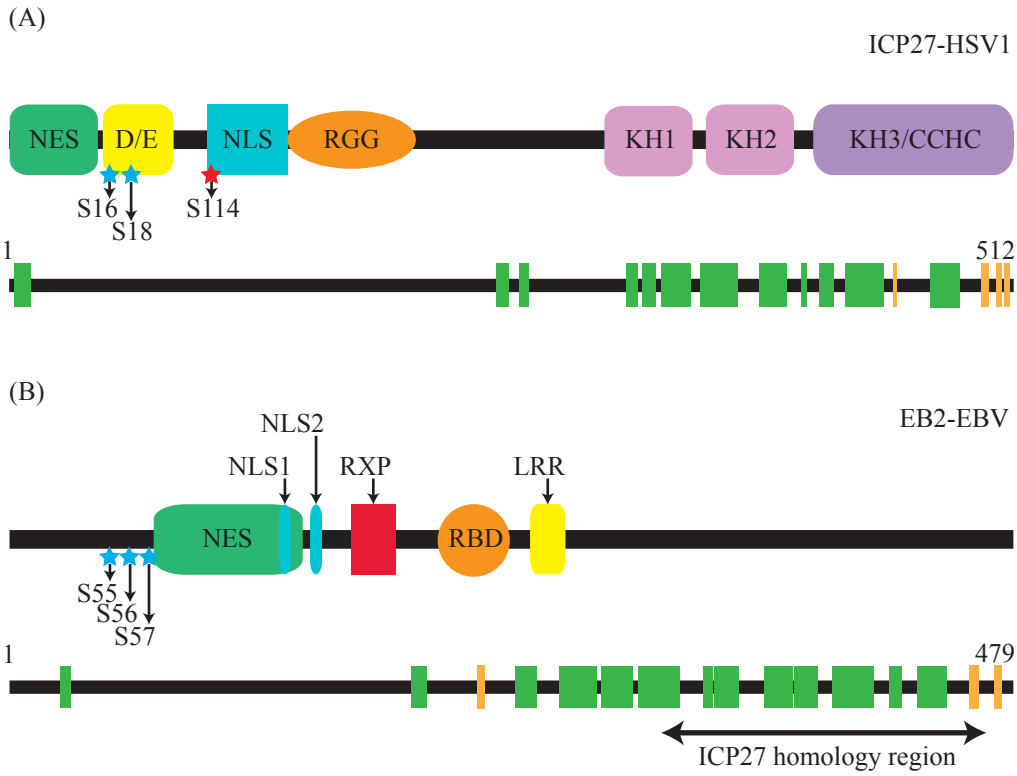


Figure 2.1 Domain and secondary structure predictions of ICP27 and EB2.

(A) and (B) depict the predicted domains of ICP27 and EB2, respectively. (A) The NES of ICP27 consists of the leucine-rich region that helps in its export out of the nucleus, D/E represents the acidic region, NLS is involved in localizing the protein into the nucleus, RGG box that binds to target mRNAs, the three KH domains that bind to the mRNA and the CCHC type of zinc-finger-like domain. S16 and S18 (blue stars ★) are CK II phosphorylation sites and S114 (red star ★) is PKA phosphorylation site. (B) The EB2 protein comprises of NES, two NLSs, the RNA-binding domain (RBD) that is the main RNA-binding element, the Arginine-X-Proline (RXP) repeats that also bind to RNA and the leucine rich region (LRR). S55, S56 and S57 (blue stars ★) are CK II phosphorylation sites. The secondary structure predictions for ICP27 and EB2 were performed using PSIPRED method (Jones, 1999; Buchan et al., 2013). The helices are represented as green box, strands as yellow box and coils as black line. (Domain structure of ICP27 is modified from Sandri-Goldin, 2011)

2.2. FUNCTIONS OF ICP27 AND EB2

2.2.1. VIRAL TRANSCRIPTION

ICP27 interacts with the cellular RNA polymerase II (RNAP II) and helps to activate viral transcription of the early and late genes (Dai-Ju et al., 2006; Jean et al., 2001; Rice & Knipe, 1990). Both N- and C-termini of ICP27 interact simultaneously with the carboxy-terminal domain (CTD) of RNAP II, recruiting it to the viral replication

centre and helping in formation of the transcription complex (Dai-Ju et al., 2006; Zhou & Knipe, 2002). It is speculated that ICP27-mRNA interaction begins as soon as the viral mRNA is transcribed. A recent study shows that ICP27 is a constituent of the viral transcription complex along with ICP4, ICP22 and cellular transcriptional factors (Kalamvoki & Roizman, 2011). RNAP II is found to interact with ICP27 in co-immunoprecipitation assays (Dai-Ju et al., 2006). However the nature of this interaction is not fully understood. Zhou and Knipe found that ICP27 co-precipitates when pulled down with anti-RNAP II (Zhou & Knipe, 2002). They reported that the binding is mediated via protein-protein interactions. Residues 407 to 504 of ICP27 are important for this interaction whereas N-terminal half may not be sufficient to bind. ICP4 binds to viral DNA and can interact with ICP27. ICP27 is speculated to bring RNAP II and ICP4 in close proximity to each other, thus indirectly assisting in viral gene transcription (Zhou & Knipe, 2002).

EB2 has not been reported to be part of the viral transcription complex. Neither has it been shown that it recruits RNAP II or other proteins to the transcription centre. EB2 is mostly involved in post-translational modifications of RNAs, which is discussed below.

2.2.2. ALTERING EXPRESSION OF CELLULAR PROTEINS

Eukaryotic pre-mRNAs undergo a series of modifications before they can be eligible for translation, such as 5'-capping, 3'-poly(A) tail addition, intron splicing and transport out of the nucleus. Though not all eukaryotic genes contain introns (e.g. Histones), most of them do. Multiple pathways exist for export of eukaryotic mRNAs, including: export of unspliced mRNA transcripts, splicing-dependent transcripts and Crm-1 dependent export (Carmody & Wente, 2009). The model for splicing-dependent export is provided in Figure 2.2 (Köhler & Hurt, 2007). Splicing and export of mRNA are tightly linked and hence a breakdown of the splicing machinery results in the accumulation of immature RNAs within the nucleus (Luo et al., 2001; Luo & Reed, 1999).

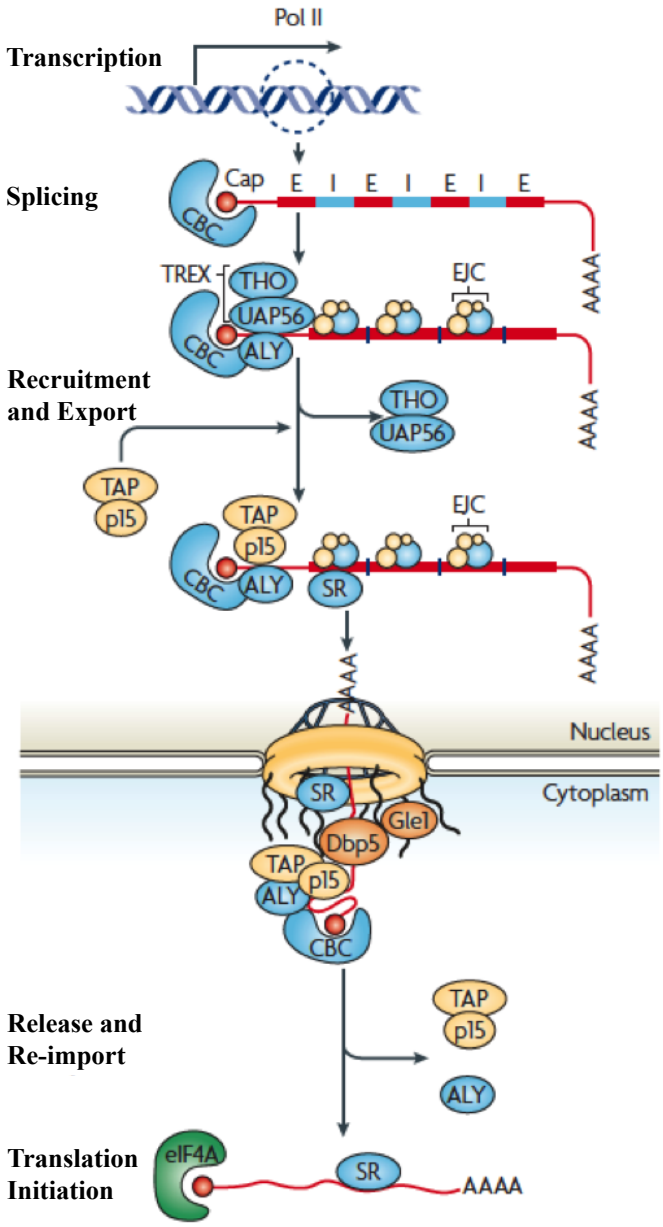


Figure 2.2 Overview of mRNA export in Metazoa.

This diagram depicts the maturation and export of mRNA out of the nucleus that are dependent on splicing and Cap-binding to the RNA. The transcribed RNA undergoes splicing, poly-adenylation and binding of the Cap-protein at its 5' end. This RNA-protein complex is called mRNP. The human TREX complex and the exon junction complex (EJC) proteins get recruited to the mRNP. Adaptor proteins like TAP and p15 are the recruited to the mRNP by Aly protein from TREX complex. This mRNP then associates with the nuclear pore complex (NPC) and is transported to the other side of the nucleus, into the cytoplasm. Most export factors then dissociate from the mRNP, allowing translation initiation factors to now bind to the mRNA and initiate translation. Shuttling proteins are generally recycled back into the nucleus. Some may need modifications, like phosphorylation, for this process. The EJC remains associated with the mRNP till it reaches the cytoplasm; however for simplicity, it is not depicted on the exported mRNA.

Pol II: RNA polymerase II; E:exon; I: intron; CBC: cap-binding complex; TREX: transcription export complex; EJC: exon junction complex; SR: SR-proteins; eIF4A: eukaryotic translation initiation factor 4A; Tap-p15: mRNA export receptor. (Image Source: Köhler & Hurt, 2007)

Herpes viruses have evolved to efficiently disrupt host the cell machinery during the lytic cycle. In HSV, the gene UL41 encodes the virion host shut-off (vhs) protein, which is localized in the tegument region of the virus. Homologues of this protein are present in all α-herpes viruses but absent in β- and γ-herpes viruses. Vhs protein is involved in rapid shut-off of the host nuclear processes. During this, it causes global increase in the degradation rate of cellular mRNA. This is supported in parallel by suppression of cellular mRNA synthesis, which is a result of repression of primary transcription and alteration in pre-RNA splicing of intron-containing cellular genes (Smiley et al., 2001). As shown in figure 2.2, splicing is an important event for RNA maturation and also for association to export proteins. The viral genes being mostly intronless remain unaffected. ICP27 has been found to be the key player in this process of stalling splicing-dependent RNA maturation. The resulting immature cellular RNAs cannot exit the nucleus and are subsequently degraded by nuclear exosomes (Smiley, 2004).

RNA polymerase II mediates transcription of many cellular proteins. It undergoes multiple phosphorylations that enable it to interact with transcription factors and activators. However, in a herpes-infected cell, RNAP II is phosphorylated differently and is re-localized to viral replication compartments within the nucleus. This results in unavailability of RNAP II for cellular transcription. ICP27 interacts with RNAP II and helps in its re-localization (Zhou & Knipe, 2002).

The cytoplasmic SR Protein Kinase 1 (SRPK1) is a Ser/Arg kinase specific for the SR proteins, Ser/Arg-rich proteins that play a role in the spliceosome assembly within the nucleus (Makarovaet al., 2001; Misteli et al., 1998). SRPK1 regulates the distribution of these SR proteins through phosphorylation and hence controls splicing of the immature transcripts (Misteli et al., 1998). ICP27 is found to interact with SR proteins (Escudero-Paunetto et al., 2010). It also recruits SRPK1 into the nucleus and this interaction between the two proteins alters the way SRPK1 phosphorylates SR proteins, resulting in aberrant phosphorylation (Sciabica et al., 2003; Souki & Sandri-

Goldin, 2009). This, in turn, results in the stalling of the spliceosome assembly before it can start catalysis (Bryant et al., 2001; Lindberg & Kreivi, 2002; Sciabica et al., 2003).

Verma and Swaminathan (2008) report that EB2 protein when expressed in uninfected cells, leads to cytoplasmic accumulation of cellular RNAs. These transcripts include STAT1 and IFN-stimulated genes. STAT1 protein itself helps in activating gene transcription of IFN-response genes. EB2 acts as an alternative splicing factor of STAT1 mRNA and results in over-expression of STAT1β form whereas the STAT1α is repressed. This impairs the cellular response to viral infection (Verma & Swaminathan, 2008).

Juillard et al. (using co-immunoprecipitation assays and GST-pull down assays) show that EB2 interacts directly with SRp20, 9G8 and SF2/ASF in an RNA-independent manner (Juillard et al., 2012). They also show that these SR proteins may help in active transport of EB2 in spite of mutations in both the EB2 NLS regions. Also EB2 interaction with SRp20 has more dramatic effect on cytoplasmic accumulation of intronless mRNA than its interaction with ASF2 or 9G8.

2.2.3. TRANSPORT OF INTRONLESS VIRAL mRNAS

As discussed in section 2.1.2, splicing and nuclear-export are linked together. Nuclear export is brought about by transcription export (TREX) complex. TREX binds to the 5'-end of cellular mRNA, while it is still undergoing splicing thus marking it for export (Cheng et al., 2006; Masuda et al., 2005). The various components of TREX complex and individual functions are mentioned in Table 2.1.

Table 2.1 Components of the TREX complex.

COMPONENT	FUNCTION	REFERENCES
Aly/REF	Adaptor protein capable of binding to mRNA and export receptor TAP/NXF1 in nuclear membrane	(Luo et al., 2001)
UAP56	RNA helicase to recruit Aly/REF to TREX	(Luo et al., 2001; Taniguchi & Ohno, 2008)
THO	Multi-protein complex	(Rondón et al., 2010)
CIP29	Forms trimeric complex with Aly\REF and UAP56	(Dufu et al., 2010)

Overall, TREX associates with the cap-binding proteins at the 5'-end of mRNA and helps to transfer the mRNA to the export receptor protein TAP/NXF1, resulting in the export of mRNA out of the nucleus (Cheng et al., 2006; Rondón et al., 2010). Cellular intronless mRNAs are exported out of the nucleus with the help of Aly/REF and UAP56, but less efficiently since these proteins are associated with the spliceosome proteins (Luo et al., 2001; Taniguchi & Ohno, 2008; Zhou et al., 2000). Certain SR proteins (SRp20, 9G8 and ASF/SF2) involved in pre-RNA splicing, also play a role in export of intronless mRNAs by shuttling between the nucleus and cytoplasm (Huang & Steitz, 2001; Huang & Steitz, 2005). Such proteins are regulated via phosphorylation at Ser/Arg residues (Misteli et al., 1998).

HSV-1 mRNAs are predominantly intronless and hence not dependent on host splicing machinery for maturation. However, they need an adaptor to help them translocate out of the nucleus by interacting with the host export factors. ICP27 acts as a multi-adaptor protein since it not only binds to the viral mRNAs, but also some host export factors like Aly/REF, SRp20 and 9G8, and also export receptor TAP/NXF1 (Chen et al., 2005; Chen et al., 2002; Escudero-Paunetto et al., 2010; Johnson & Sandri-Goldin, 2009; Koffa et al., 2001). Hence ICP27 is involved in every step required for the transport of viral intronless mRNAs out of the nucleus. An NMR study by Tunnicliffe RB *et al.* found that a short stretch of ICP27 encompassing residues 103-110 is sufficient for binding to the RNA recognition motif (RRM) of REF (Tunnicliffe et al., 2011). Hernandez FP *et al.* have demonstrated that ICP27, in a head-to-tail dimeric form bound to viral mRNA can directly interact with TAP/NXF1 (Hernandez & Sandri-Goldin, 2010a, 2010b). Binding to host export factors may be contributing to make the export process efficient.

EB2 has been found to interact with the host export factors. EB2, Aly/REF and TAP co-immunoprecipitate from infected cell lysates (Hiriart et al., 2003). The C-terminal leucine-rich region (LRR) of EB2 is essential for direct interaction with Aly/REF (Hiriart et al., 2003). EB2-Aly-mRNA complex then interacts with the TAP/p15 protein that helps translocate the cargo across the nuclear pore complex. The human SPEN (split end) family proteins namely RNA-binding motif protein 15 (RBM15 or

OTT1), RBM15B (OTT3) and SHARP, consisting of multiple RNA recognition motifs (RRMs), have also been found to interact with EB2 (Ariyoshi & Schwabe, 2003; Hiriart et al., 2005). RMB15 (OTT1) functions as an mRNA export factor, though it is unclear if it is the same for RBM15B and SHARP proteins. RBM15 is also known to interact with TAP/p15 (Lindtner et al., 2006). Hence it is speculated that EB2 exports its target mRNAs with the help of export factors.

2.2.4. BINDING TO VIRAL mRNAS

ICP27 binds many intronless viral mRNAs mainly belonging to the immediate early and late stages of the viral lytic cycle (Ingram et al., 1996; Jean et al., 2001; Rice & Knipe, 1990). It contains the N-terminal RGG-box motif (residues 138-152) that binds to mRNAs along with the proposed three KH-like domains at the C-terminus (Mears & Rice, 1996; Soliman & Silverstein, 2000a; Soliman & Silverstein, 2000b). Both *in vitro* and *in vivo* studies have indicated that R138, 148 and 150 are important for RNA binding and export function of ICP27 (Corbin-Lickfett et al., 2010). However, there is no particular consensus sequence found that may be binding to it. Unlike several other known RGG-box-containing proteins that preferentially bind to G-quartet structures in RNA, many different studies predict ICP27 to bind to GC-rich regions in mRNA that are flexible and without any secondary structures, making ICP27's RGG-box motif unique in its ability to recognize and bind to its RNA substrates (Corbin-Lickfett et al., 2009; Mears & Rice, 1996; Norseen et al., 2009; Ramos et al., 2003; Sandri-Goldin, 1998; Sandri-Goldin, 2011). This may raise the question whether ICP27 can bind cellular mRNA, rich in GC-content? Even though there do exist cellular RNA with high GC-content, most of these belong to rRNAs, which are highly structured molecules and unlikely to bind ICP27 (Sandri-Goldin, 2011).

EB2 has been reported to enhance accumulation of 60 % of the viral genes of which majority are the late structural genes. Some early genes involved in lytic DNA replication like the EBV primase and polymerase are EB2-dependent (Han et al., 2009; Han et al., 2007). RNA-mobility shift assays recognized an RNA-binding domain (RBD) in EB2 spanning residues 190-223. This RBD encompasses the residues TRKQAR forming the Arginine-rich motif (ARM), which is similar to the

ARM of other RNA-binding proteins like HIV-1 Rev and Tat. There is a second RNA-binding region containing multiple Arginine-X-Proline (RXP) repeats. Unlike the RBD, the RXP repeats are not essential for binding to mRNA. EB2 interacts with a large proportion of freshly transcribed target mRNAs in vivo (Ruvolo et al., 2001). Similar to ICP27, the mechanism by which EB2 recognizes its target mRNAs is not clear. It is hypothesized that EB2 either recognizes a consensus sequence of RNA or secondary structures in RNA.

2.2.5. PROMOTION OF VIRAL mRNA TRANSLATION

Studies in past decade have indicated that ICP27 also plays a role in promoting translation of some viral mRNAs in the cytoplasm by recruiting cellular translation initiation factors like eIF3, eIF4G and poly(A)-binding proteins (PABP), and polyribosomes (Fontaine-Rodriguez & Knipe, 2008; Fontaine-Rodriguez et al., 2004; Larralde et al., 2006; Salaun et al., 2010). The mRNA of a tegument protein of HSV-1, VP16 is translocated into the cytosol without the assistance of this adaptor protein; however, its translation is dependent on ICP27 (Fontaine-Rodriguez & Knipe, 2008).

Poppers et al. (2002) demonstrate that EB2 prevents the inhibition of translation caused by protein kinase R (PKR). dsRNA activates PKR that phosphorylates eIF2. In phosphorylated form, eIF2 stalls translation. An RXP triplet repeat in EB2 binds to PKR and prevents its activation. This repeat and function is not present in ICP27 and other homologues but is the function of another viral protein altogether (Poppers et al., 2003).

2.3. POSTTRANSLATIONAL MODIFICATIONS FOR REGULATION OF ICP27 AND EB2

It is speculated that phosphorylation and arginine-methylation events regulate ICP27. Earlier studies show that ICP27 has stable phosphate groups as well as phosphates that cycle on and off (Wilcox et al., 1980). Zhi and Sandri-Goldin were first to map the phosphorylation sites on the protein (Zhi & Sandri-Goldin, 1999). They reported that S16 and 18 are phosphorylated *in vivo* by casein kinase II (CKII) and S114 by protein kinase A (PKA). They concluded that S114, which is part of the nuclear localization signal (NLS), is an important phosphorylation site as its mutation to

alanine reduced its ability to translocate into the nucleus in a competition assay with ICP27 wild-type protein. However, S114A mutant was predominantly located in the nucleus. S311 and S334 were also detected *in vitro* but not found to be phosphorylated *in vivo*.

Another study illustrated that S16E and S16,18,114E phosphomimetic mutants failed to interact with Aly/REF (Corbin-Lickfett et al., 2010). They used co-localization and co-immunoprecipitation techniques and concluded that phosphorylation of ICP27 at these sites disrupts this interaction. They also probed if these mutants affected ICP27's interaction with TAP/NXF1. They showed that the mutants failed to co-localize with TAP and remained within the nucleus. These studies show that the phosphorylation state of ICP27 affects its ability to bind to its interacting partners, or at least its function of exporting out of the nucleus.

Souki et al. reported that ICP27 undergoes methylation at R138, 148 and 150, all of which are located in the RGG box of the protein (Souki et al., 2009). Mutation of these arginines to lysines demonstrated that ICP27 was hypomethylated compared to wild-type protein. In addition, they discovered that mutant protein had lesser stability and it was exported into the cytoplasm earlier and faster (Souki et al., 2009). Rapid export may affect other functions of ICP27 that are carried on within the nucleus. Methylation of ICP27 is associated to its interaction with SRPK1 and Aly/REF proteins (Souki & Sandri-Goldin, 2009). Hypomethylation of the RGG box leads to loss of interaction between the viral and some host proteins. Thus arginine methylation along with phosphorylation is involved in regulating some of ICP27's activities.

ICP27 is also reported to undergo adenylation and guanylation in the nucleus of Hela S3 cells (Blaho, Mitchell, & Roizman, 1993). It is proposed that early and late gene products of HSV-1 may be needed for the nucleotidylation. However, little is known about how this post-translational modification affects the protein.

EB2 is a phosphoprotein and undergoes phosphorylation by the kinase CK2. *In vitro* studies have shown that CK2 α2β2 heterotetrameric holoenzyme binds efficiently to the NES region of EB2 (Medina-Palazon et al., 2007). The α and β subunits also bind

individually to the C-terminal region of EB2. Similar to ICP27, all the phosphorylation sites reported are mapped to the N-terminal with S55, S56, S57 being the key residues. The triple mutant EB2S3A (S55,56,57A) failed to rescue of BMLF1-knockout virus whereas the phosphomimetic mutant EB2S3E (S55,56,57E) efficiently rescued it. These three serines are positioned close to the NES of EB2. However, export of EB2 was unaffected in the mutants. They also inferred that mRNA export by EB2 is regulated by EB2.

2.4. STRUCTURAL INFORMATION ON ICP27 AND EB2

ICP27 interacts with a multitude of macromolecules including viral mRNA, host nuclear and cytoplasmic proteins (Chen et al., 2005; Ingram et al., 1996; Malik et al., 2012; Mears & Rice, 1996; Sandri-Goldin, 2011; Sokolowski et al., 2003; Zhou & Knipe, 2002). It also interacts with HSV-1 protein, ICP8 (Olesky et al., 2005). ICP27 has recently been proposed to have head-to-tail intramolecular dimer interaction (Hernandez & Sandri-Goldin, 2010a, 2010b). In a nutshell, a single viral protein, performing various tasks in a regulated fashion should have a structure capable of binding to a number of other macromolecules.

Circular Dichroism (CD) analysis and NMR studies on the N-terminal half (residues 1-160) of the protein have shown that this region is predominantly random-coils with no rigid 3D-structure (Corbin-Lickfett et al., 2010; Corbin-Lickfett et al., 2010). Secondary structure predictions using bioinformatics software (e.g. PSIPRED) also indicate that the N-terminal region up to ~300 residues is highly disordered (Buchan et al., 2013). The N-terminal half of ICP27 encompasses interactions with many proteins as well as viral mRNA and it is predicted that the disordered nature of the molecule allows it to have multiple different interactions (Chen et al., 2005; Ingram et al., 1996; Mears & Rice, 1996; Sandri-Goldin, 2011; Tunnicliffe et al., 2011). It is postulated that binding with the interacting partners stabilizes this region. It is seen that many regulatory proteins within the cell (involved in signaling, regulation of transcription, translation, cell-cycle etc.) including many onco-proteins have intrinsically disordered regions, which are involved in multiple interactions (Chouard, 2011; Uversky & Dunker, 2008; Xue et al., 2014).

For the C-terminal half of ICP27, bioinformatics tools predict the presence of multiple α-helices and a few interspersed β-strands (Sandri-Goldin, 2011). The C-terminus of ICP27 also interacts with multiple proteins. This region is predicted to encompass three K-Homology (KH) domains as well as a Zinc-finger-like motif of CCHC type (Soliman & Silverstein, 2000a; Soliman & Silverstein, 2000b; Vaughan et al., 1992).

EB2 has not been studied as extensively as ICP27 and ORF57. Little to no experimental data is available with respect to the structure of the EB2 protein. Secondary structural predictions reveal that the N-terminal, from residues 1 to 150, is disordered, just as in the case of ICP27. The C-terminal half is predicted to have α-helices with few beta strands (Sandri-Goldin, 2011).

There is no 3D-structural information available for ICP27 and EB2, except for an NMR structure of a short region (residues 103-110) of ICP27 that interacts with the RNA recognition motif (RRM) of the Aly/REF protein (Tunnicliffe et al., 2011). Therefore, the molecular mechanisms that facilitate the multiple functions of these proteins still remain largely unknown. Structural analysis of ICP27 and EB2, individually and in complex with their natural interaction partners, will give a better insight into the functional aspects of these key herpesvirus proteins. This thesis strives to achieve these goals by using X-ray crystallography, supported by biophysical and biochemical techniques.

3. MATERIALS AND METHODS

3.1. MATERIALS

All materials used were of at least analytical grade. Chemicals, instruments and miscellaneous were bought from following companies.

3.1.1. BUFFERS AND SALTS

NaCl	Merck
Magnesium sulphate	Sigma
Glycerol	Sigma
Imidazole	Merck
TCEP	Sigma
Other Salts and Buffers	Sigma, Merck, Fluka

3.1.2. CULTIVATION MEDIA AND ANTIBIOTICS

Terrific broth	Formedium
LB media	Formedium
Antifoam	Sigma
Antibiotics	Calbiochem, Sigma, Gibco
Protease Inhibitor	Nacalai Tesque, Inc.

3.1.3. PROTEIN PURIFICATION

Nickel NTA beads	Qiagen
Superdex columns	GE Healthcare
Membrane filters	Sartorius Stedim
Syringe, needle	BD
Concentrator	Sartorius Stedim

3.1.4. MOLECULAR BIOLOGY

KOD Xtreme Hot Start Polymerase	Novagen
Miniprep Plasmid kit	Qiagen
SDS and Native PAGE gels	Life Technologies
Primers	Integrated DNA Technologies
GeneRuler DNA Marker	Life Technologies
SeeBlue Plus2 Protein Standard	Life Technologies

3.1.5. CRYSTALLIZATION

Crystallization screens	Qiagen, Hampton, Molecular Dimensions
Plates	Art Robbins

3.1.6. INSTRUMENTATION

PCR thermocycler	Applied Biosystems
Nano-spectrophotometer	Thermo Scientific
LEX system	Harbinger Biotech
AKTAxpress	GE Healthcare
Static light scattering system	Wyatt, GE Healthcare
Circular dichroism instrument	Applied Biosystems

3.2. CLONING

The entry clone encoding full-length ICP27 (GenBank accession number ABI63515) in pDONR207 was provided by collaborators at the University of Edinburgh. Using the LR reaction of Gateway Recombination Cloning Technology (Invitrogen), the gene was subcloned into another vector pTH27, containing the sequence for an N-terminal hexa-histidine tag (His6-tag) and Tobacco Etch Virus (TEV) protease site (MHHHHHHSSGVDLGTENLYFQSM). Chemically competent MACH1-T1^{R} *Escherichia coli* cells were transformed with the destination vector and cultured on Luria Bertani (LB) agar plates supplemented with 100 μg/ml Ampicillin, grown overnight at 37 °C. Plasmids were extracted and sequenced, before transformation of the expression strain T1^{R}-BL21-DE3 Rosetta of *Escherichia coli*. Cultures were then inoculated in 5 ml LB broth supplemented with 100 μg/ml Ampicillin and 34 μg/ml Chloramphenicol, grown overnight at 37 °C. Glycerol stocks with 50 % (v/v) glycerol were prepared and stored at -80 °C and used for small scale expression screening.

Gene constructs of ICP27 were designed and cloned using Ligation Independent Cloning (LIC) into the vector pNIC28-Bsa4, containing a sequence for an N-terminal His6-tag and TEV protease site. The expression strain T1^{R}-BL21-DE3 Rosetta of *Escherichia coli* was transformed with the plasmid and used for small-scale expression screening.

3.3. SMALL-SCALE EXPRESSION AND PURIFICATION SCREENING

For screening of protein expression at small scale, cultures with full-length gene and different constructs were each cultivated in 1 ml terrific broth (TB) in a 96-deep well block. The broth was supplemented with the appropriate antibiotics. The block was

then sealed with a gas-permeable membrane and incubated at 37 °C till the cultures reached an O.D. of 2 at 600 nm. The temperature was then lowered to 18 °C and expression was induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Overnight expression was allowed. Cells were harvested, resuspended in lysis buffer and lysed by freeze-thaw technique. Lysates were clarified by passing through 1.2 µm membranes.

Small-scale purification of the lysates was performed in 96-well format. 50 µl of Ni-NTA beads equilibrated in Wash 1 buffer was dispensed in each well lined with a 0.65 µm membrane. The clarified lysate was incubated with the beads for 20 min at room temperature with constant shaking. The plate was then centrifuged at 100 rpm for 1 min. 250 µl of Wash 1 buffer was applied to the beads, incubated for 2 min and then centrifuged again. The step was repeated to remove as much unbound protein as possible. The Ni-NTA bound protein was eluted with 50 µl of the Elution buffer incubated for 20 min at room temperature with constant shaking and then centrifuged the eluate. The eluate was then analysed by SDS-PAGE. The protein constructs that expressed well in small-scale screening were then scaled up.

3.4. LARGE-SCALE CULTIVATION

Large-scale cultivation was carried out using the large-scale expression system (LEX; Harbinger Biotech, Canada). 1.5 L of Terrific Broth containing 50 µg/ml Kanamycin, 34 µg/ml Chloramphenicol, 8 g/L (w/v) glycerol and approximately 300 µl of antifoam, was inoculated with the overnight bacterial culture, prepared from the glycerol stocks. The culture was allowed to grow at 37 °C with continuous (filtered) aeration through spargers of LEX system. Once the culture reached an O.D. 600 nm of 1.8 – 2.0, the temperature was lowered from 37 °C to 18°C and expression was induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Protein expression was then allowed to proceed overnight (up to 18 hrs).

3.4.1. SELENO-METHIONINE SUBSTITUTED PROTEIN CULTIVATION

15 ml of overnight bacterial culture was prepared using LB broth supplemented with 50 µg/ml Kanamycin and 34 µg/ml Chloramphenicol and incubated at 37 °C with

shaking at 200 rpm. Once the culture reached an O.D. ≈ 1.0, it was used as inoculum. M9 SeMET minimal growth medium (Shanghai Medicilon Inc., China) was used for expression of seleno-methionine substituted ICP27$^{190-512}$. The overnight culture was prepared for inoculation by first performing exchange of the growth medium. The culture in LB broth was centrifuged at 2000 rpm for 10 min at 15 °C. The supernatant was discarded and the cell pellet was re-suspended completely in the prepared minimal medium. The cells were centrifuged and re-suspended in fresh medium again at least 3 times to ensure that the LB broth was completely replaced by the minimal medium. The re-suspended cells were then used as inoculum for scale-up in minimal medium containing 50 µg/ml Kanamycin, 34 µg/ml Chloramphenicol and 8 g/L (w/v) glycerol. The culture was allowed to grow at 37 °C using the LEX system, until it reached O.D.$_{600}$ = 1.2. The temperature was then reduced from 37 °C to 18 °C and the expression was induced with 0.5 mM of IPTG after 30 min. Protein expression was allowed overnight.

3.4.2. HARVESTING

The cultures were harvested by centrifugation at 4000 g for 20 min at 4 °C. The cell pellets were then re-suspended in Lysis Buffer [100 mM Na-HEPES pH 8.0, 500 mM NaCl, 10 mM Imidazole, 10 % (v/v) glycerol, 0.5 mM TCEP, 1 mM $MgSO_4$, 0.1 mg/ml Lysozyme, 1:1000 Protease Inhibitor Cocktail mix III, 1000 U Benzonase (per 750ml cultivation)]. The cell suspensions were stored at -20 °C.

3.5. PROTEIN PURIFICATION

Protein purification was carried out in two steps with Ion Metal Affinity Chromatography (IMAC) followed by Size Exclusion Chromatography (SEC) using the ÄKTAxpress system. The cell pellets were thawed and subjected to sonication. The lysate was then centrifuged at 47,000 g for 30 min at 4 °C. The supernatant was collected after passing it through a 1.2 µm filter. This supernatant was then used for the two-step purification process.

3.5.1. ION-METAL AFFINITY CHROMATOGRAPHY

1 ml Ni-NTA Superflow columns (Qiagen) were used to trap the His-tagged proteins. The column prior to loading with sample was equilibrated using 10 column volumes (CV) of Wash Buffer 1 [20 mM Na-HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 10 % (v/v) glycerol, 1 mM TCEP]. Once the protein sample was loaded, the column was washed with 20 CV of Wash Buffer 1 followed by 20 CV of Wash Buffer 2 [20 mM Na-HEPES pH 7.5, 500 mM NaCl, 25 mM Imidazole, 10 % (v/v) glycerol, 1 mM TCEP], to wash off any unbound proteins. The bound protein was eluted out of the column by using 5 CV of Elution Buffer [20 mM Na-HEPES pH 7.5, 500 mM NaCl, 100 mM Imidazole, 10 % (v/v) glycerol, 1 mM TCEP].

3.5.2. SIZE EXCLUSION CHROMATOGRAPHY

The IMAC eluate was loaded onto the HiLoad 16/60 Superdex 200 prep grade SEC column. The column was equilibrated using 2 CV Gel Filtration Buffer [20 mM Na-HEPES pH 7.5, 300 mM NaCl, 10 % (v/v) glycerol, 1 mM TCEP]. The protein was eluted from the column using 1 CV of buffer at 1.2 ml/min flow-rate and collected as fractions into 96-well deep well block. Depending on the chromatogram from SEC, protein fractions were analysed by SDS-PAGE. Fractions with high purity were pooled together and concentrated in a centrifugal concentrator (Sartorius Stedim Biotech) with a molecular weight cut-off (MWCO) value of 10 kDa. The concentrated protein sample was measured by UV absorbance at 280 nm using Nanodrop. The protein was then aliquoted in smaller volumes of 60 µl and stored at -80 °C.

3.6. WESTERN BLOT ANALYSIS

5-10 µg of protein was loaded on SDS-PAGE. Electrophoresis was carried out in MES buffer [50 mM MES, 50 mM Tris base, 0.1 % SDS, 1 mM EDTA, pH 7.3] at 200 V for 30 min. Transfer of resolved protein onto polyvinylidene fluoride (PVDF) membrane was carried out using semi-dry iBlot gel transfer device (Invitrogen). The membrane was then washed in 30 ml Tris-Buffered Saline and Tween 20 (TBST) buffer [50 mM Tris buffer pH 7.5, 150 mM NaCl, 0.05 % (v/v) Tween 20] for 3 x 10

min. Blocking was performed with 1 % Bovine serum albumin (BSA) in TBS buffer with shaking overnight. Next day, membrane was washed again in 30 ml TBST buffer for 3 x 10 min to remove traces of unbound BSA. Membrane was incubated in 10 ml TBST containing 1 µg/ml HisProbe-HRP (Thermo Scientific) for 1 hr with constant shaking. Unbound HisProbe-HRP was removed by washing with TBST. The membrane was then developed with equal volumes of Luminol (Enhancer) solution and Stable Peroxidase solution (Thermo Scientific) by incubating for 2-3 min. The solution was drained and the chemiluminescence signal was detected at 428 nm using the ImageQuant LAS 4000 biomolecular imager (GE Healthcare).

3.7. CRYSTALLIZATION TRIALS

3.7.1. INITIAL SCREENING

The protein was screened for crystallization conditions using NeXtal JCSG+ suite screen (Qiagen). The screen consists of 96 different conditions. Screening was done by sitting drop vapour diffusion technique at 4 °C and 20 °C. Using the Phoenix automated dispensing robot (Art Robbins Instruments), 0.3 µl crystallization drops were dispensed in 96-well Intelliplates by mixing protein with reservoir solution in three different ratios, i.e. protein: reservoir solution 2:1, 1:1 and 1:2, and equilibrated with 50 µl of reservoir solution. The plates were incubated and imaged at defined intervals by the RockImager (Formulatrix) using the Rockmaker software. The crystals were verified under UV light, to rule out the possibility of salt crystals.

3.7.2. OPTIMIZATION OF CRYSTALLISATION CONDITION

For some of the conditions that showed crystal or micro-crystal formation, optimization grids were set up. Optimizations included varying the salt concentrations, precipitant concentrations and pH around that of the initial crystallization condition, in order to find a condition that yielded better crystals.

3.7.3. ADDITIVE SCREEN

For certain positive conditions from the initial screen, an additive screen from Hampton was set up in parallel to the optimization screening. 96 different small

molecule additives are included in the crystallization condition to check if addition of such small molecules could help in formation of better crystals. Optimizations of the positive conditions from the additive concentration were also performed.

3.8. FREEZING OF CRYSTALS AND DATA COLLECTION

The crystals were transferred to a cryo-solution for a few seconds and then immediately flash frozen in liquid nitrogen. X-ray diffraction experiments and subsequent data collection were performed for both native crystals and seleno-methionine substituted crystals at National Synchrotron Radiation Research Centre (Taiwan), Australian Synchrotron (Australia) and Diamond Light Source (United Kingdom).

3.9. CIRCULAR DICHROISM

Circular Dichroism is an absorption spectroscopic technique. Proteins are chiral molecules that can absorb circularly polarized light. The difference in the absorption of left and right circularly polarized light is measured as Circular Dichroism (CD) of that protein and is used to evaluate its secondary structure. Chirascan spectropolarimeter (Applied Photophysics) was used for performing CD spectra of the protein samples at 25 °C. Samples were subjected to far UV light ranging from 180 nm to 260 nm in steps of 1 nm with nitrogen gas continuously purged through the system to maintain an inert environment. 60 µl of the sample (1 mg/ml) was applied to the 0.1 mm quartz cells for each run and buffer-blank readings were taken before each sample. Chirascan software was used to subtract the buffer baseline from the protein absorption value.

3.10. MULTI-ANGLE LIGHT SCATTERING

Light scattering is a technique used for characterizing macromolecules like proteins, polymers and some other particles when in solution. It is a non-destructive technique that is dependent only on the molar mass of the sample in use and no other external standards. Basic characteristics like shape, size and oligomeric states can be determined by the intensity of light scattered by the molecules in the sample.

Size exclusion chromatography coupled to a multi angle light scattering system (MALS) with the miniDAWN TREOS and Optilab rEX detectors (Wyatt Technology Corp.) were performed for native ICP27$^{190-512}$ protein. 25 µl of the ICP27$^{190-512}$ at 5 mg/ml was injected on to Superdex 200 15/150 GL SEC column, pre-equilibrated with buffer (20 mM Na-HEPES pH 7.5, 300 mM NaCl and 5 % (v/v) glycerol) filtered through 0.1 µm filter. Separation was carried out at a flow rate of 0.3 ml/min. MALS was performed with a laser at wavelength of 658 nm, on the peak fractions. The differential refractive index (dn / dc) for ICP27$^{190-512}$ was determined to be 0.185 ml/g. Light scattering data was analyzed using ASTRA software (Wyatt Technology Corp.).

3.11. SEDIMENTATION EQUILIBRIUM BY ANALYTICAL ULTRACENTRIFUGATION

A sedimentation equilibrium (SE) experiment was performed on ICP27$^{190-512}$ at 20 °C using Beckman XL-I analytical ultracentrifuge. 130 µl each of homogenized ICP27$^{190-512}$ samples with optical density at 280 nm (OD_{280}) of 0.3, 0.5 and 0.8 were prepared. The buffer system used was 20 mM Na-HEPES pH 7.5, 300 mM NaCl and 10 % (v/v) Glycerol. Centrifugation was carried out in a carbon-epoxy centrepiece with 6 quartz windows for references and samples. The samples were centrifuged till they reached equilibrium (15 hrs per speed) and run at 10500 rpm, 12700 rpm and 15500 rpm. Absorbance data was collected at 280 nm. Partial specific volume and monomer molecular weight were calculated using the program SEDNTERP. The partial specific volume of ICP27$^{190-512}$ was 0.732291 cm^3/g. Scan datasets were analyzed by SEDFIT software. Data fitting was done for checking possibilities of monomeric state and dimeric, trimeric and tetrameric associations.

3.12. PHOSPHORYLATION ASSAY

For phosphorylation with protein kinase A (PKA), 100 µg of ICP27$^{190-512}$ was incubated with 200 U of PKA type I catalytic subunit (New England Biolabs; NEB) in 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 500 µM ATP. For phosphorylation with casein kinase II (CKII), 100 µg of ICP27$^{190-512}$ was incubated with 250 U of CKII (NEB) in 20 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM $MgCl_2$ and 500 µM ATP. For phosphorylation with protein kinase C (PKC), 100 µg of ICP27$^{190-512}$ was

incubated with 160 mU of PKC (Sigma Aldrich) in 20 mM Na-HEPES pH 7.5, 10 mM $MgCl_2$ and 500 μM ATP. All reactions were carried out overnight at 30 °C. To check if phosphorylation had occurred, the sample was incubated with PhosphoProtein purification resin (Qiagen), which binds specifically to the phosphorylated protein. The manufacturer's protocol was followed to elute the bound protein and fractions were analyzed on SDS-PAGE. The protein was analyzed with LC-MS/MS to identify the phosphorylated residues.

3.13. THERMAL STABILITY SHIFT ASSAYS

Thermal stability shift assay (TSSA) was performed to assess the effect of nucleic acids (RNA/ssDNA) on ICP27 and BMLF1 protein constructs. The principle of this assay makes use of unfolding of proteins when subjected to gradual thermal stress. As the protein unfolds with an increase in temperature, the hydrophobic regions and patches buried deep in the protein structure will get exposed. Hydrophobic dyes like SYPRO orange do not show any fluorescence when in an aqueous environment. However, upon binding to the exposed hydrophobic regions, they start to fluoresce. When a protein binds to an interacting partner or any ligand, there is shift in the melting temperature (Tm) that can be measured by TSSA. A stabilizing interaction will increase the Tm of the protein (slower unfolding of protein), whereas a destabilizing interaction will decrease the Tm (faster unfolding of protein).

The protein was incubated with 20 μM of RNA/DNA for 30 min on ice. It was subsequently diluted to 0.2 mg/mL in buffer (20 mM Na-HEPES pH 7.5, 300 mM NaCl) containing (1:1000, 5×) SYPRO Orange (Invitrogen). 25 μl of the sample was dispensed in each well of the 96-well PCR plate and immediately covered with optically clear sealing tape to avoid evaporation. Each sample was run in triplicates. An iCycler IQ (Bio-Rad) real-time PCR machine was used for subjecting samples to gradual thermal stress ranging from 25 °C to 80 °C in steps of 1 °C. The instrument is fitted with filters for detection of SYPRO Orange ($\lambda_{excitation}$ = 490 nm, $\lambda_{emission}$ = 575 nm) fluorescence. Control protein samples with buffer in place of nucleic acids were run in the same plate. Changes in the fluorescence signal in each well were detected using a built-in CCD camera. For determining the Tm of each protein sample, the

fluorescent signal data collected by the CCD camera is fed into the XL-fit 4 software, which fits the curves using Boltzmann curve-fitting model.

3.14. SITE-DIRECTED MUTAGENESIS, CULTIVATION AND PURIFICATION OF PHOSPHOMIMETIC MUTANTS

To probe the possibility that PKA-phosphorylated ICP27$^{190-512}$ could undergo any conformational changes, the approach of phospho-mimicking was utilized. Using site-directed mutagenesis, the detected phosphorylation sites (serines and tyrosines) were mutated to glutamate residues that mimic phosphorylation. Forward and reverse primers were designed such that the mutation was centered and the Tm of the primer was in the range of 75°-90° C.

Table 3.1 List of primers for site-directed mutagenesis.

MUTATION	PRIMER SEQUENCES FFROM 5' TO 3'	Tm (°C)
T244E	Fwd: GCGCCCGCCGCCGACGAGATCGACGCCACCACG Rev: CGTGGTGGCGTCGATCTCGTCGGCGGCGGGCGC	86.7
S266E	Fwd: GCGGTCGACCGCATCGAGGAGAGCTTTGGCCGC Rev: GCGGCCAAAGCTCTCCTCGATGCGGTCGACCGC	82.3
S268E	Fwd: GACCGCATCAGCGAGGAGTTTGGCCGCAGCGCA Rev: TGCGCTGCGGCCAAACTCCTCGCTGATGCGGTC	81
S290E	Fwd: TTTCCCGCCGCGAATGAGCCCTGGGCCCCGGTG Rev: CACCGGGGCCCAGGGCTCATTCGCGGCGGGAAA	86.4
S311E	Fwd: GAGACCAGACGGGTCGAGTGGGAAACCTTGGTC Rev: GACCAAGGTTTCCCACTCGACCCGTCTGGTCTC	78.9
T314E	Fwd: CGGGTCTCCTGGGAAGAGTTGGTCGCCCACGGC Rev: GCCGTGGGCGACCAACTCTTCCCAGGAGACCCG	81.4
S493E	Fwd: TGCGAGTTGACGGCCGAGCACATCGTCGCCCCC Rev: GGGGGCGACGATGTGCTCGGCCGTCAACTCGCA	83

The pNIC-Bsa4 vector containing the ICP27$^{190-512}$ gene was used as template DNA. For the polymerase chain reaction (PCR), 50 µl reaction mix was set up as in Table 3.2. Reagents were pipetted into the PCR tubes and mixed by a short centrifugation. The tubes were then placed into PCR thermocycler (Applied Biophysics) and amplified. The program used for amplification is given in Table 3.3.

Table 3.2 Reagents for PCR.

REAGENT	STOCK	FINAL	VOLUME (ml)
KOD Xtreme HS Buffer	2 x	1 x	25
dNTP (mM)	2.5	0.4	8
Primer (fwd and rev) (µM)	10 each	0.3 each	1.5 each
Template DNA Plasmid (ng/µl)	12.5	0.2	0.8
KOD Xtreme HS Polymerase (units/µl)	1	0.02	1
Molecular Biology Grade water		To make 50 µl	12.2

Table 3.3 Two-step gradient PCR protocol.

CYCLE STEP	TEMPERATURE (°C)	TIME	
Pre-denaturation	94	2 min	
Denaturation	98	10 sec	20-30 cycles
Annealing	55-65	30 sec	
Extension	68	9 min	
Final Extension	68	10 min	
Hold	4	∞	

The PCR products were checked for successful amplification by electrophoresis using 1 % agarose gels supplied with 1:20000 SYBR-Safe DNA gel stain. Electrophoresis was carried out at 100 V for 20 min. The bands were observed under transilluminator and size confirmed by comparison to DNA marker. The successfully amplified samples were treated with Dpn1 enzyme at 37 °C for 1 hr. The samples were transformed into MACH1-T1^{R} *E. coli* cells with help of heat-shock treatment, plated onto LB agar plates and incubated overnight at 37 °C. Single colony was picked for each mutant and grown overnight in LB broth at 37 °C. Plasmid DNA was extracted using the miniprep protocol (Qiagen) and the ICP27$^{190\text{-}512}$ gene was amplified in PCR using T7 forward and reverse primers. The final PCR products were sent for verification of site-directed mutagenesis. Successful mutants were used for further analyses.

The plasmids of the verified mutants were transformed into the expression strain of *E. coli*, namely T1^{R}-BL21-DE3 Rosetta, using heat-shock treatment. After overnight growth on LB agar at 37 °C, single colonies were picked for small-scale expression

and purification screening, as discussed in section 3.2. Mutants with well-expressing proteins were scaled up and protein purified at large scale, using same protocols as discussed in sections 3.3 and 3.4.

3.15. CHYMOTRYPSINATION ASSAY

To qualitatively assess if phosphorylation by PKA brought about changes in the oligomeric state of ICP27$^{190-512}$, chymotrypsination assay was performed. ICP27$^{190-512}$ was incubated with 1:5 dilution series of chymotrypsin for 30 min at room temperature. A duplicate set was prepared and incubated on ice (data not shown). After incubation, 10 µl of sample from each mix was aliquoted out and reaction was stopped by denaturing at 95 °C for 5 min, with SDS-PAGE loading dye. All denatured samples were then analysed using SDS-PAGE.

3.16. MULTI-CONSTRUCT APPROACH FOR HERPES VIRAL PROTEIN EXPRESSION

In the study of expressing recombinant herpes viral proteins, 94 herpes viral proteins where subjected to expression trials. The constructs were designed from full-length clones, provided generously by Juergen Haas at the University of Edinburgh, UK. The construct-design was based on information of homologous structures and secondary structure predictions. Cloning was performed as described above in section 3.2, followed by small-scale expression screening, described in section 3.3, to evaluate if the clones will be successful in large-scale production. For C-terminal His6-tagged proteins, pNIC-CH2 plasmid was used instead of pNIC28-Bsa4. The clones that expressed well in small-scale screening were chosen for large-scale protein expression and purification, performed as described above, in sections 3.4 and 3.5, respectively.

4. RESULTS

4.1. Structure and Interactions of the ICP27 C-Terminal Domain from HSV-1

Viral proteins have evolved to perform multiple functions while displaying an array of interactions. Many viral proteins have disordered regions with little secondary structures, which can be moulded to fit with their interacting partners that often have more ordered structures (Uversky & Dunker, 2008; Xue et al., 2014, 2012). In case of herpesviruses, the amount of protein disorder was found to be 17.9 % (Xue et al., 2012). The proteins under study, ICP27 and EB2, both show a relatively high degree of disorder in their predicted structures. This poses a problem while cloning and purifying such viral proteins. The disordered regions are highly susceptible to degradation and aggregation during the protein purification and also during long incubations during crystallization. It is also difficult to crystallize the disordered regions.

4.1.1. Cloning, Expression and Purification of ICP27

Full-length gene of ICP27 was successfully cloned and expressed in *E. coli* cells. However, the protein heavily degraded during the purification process (Figure 4.1). The gel filtration profile of full-length ICP27 showed two broad asymmetrical peaks (Figure 4.1 A). Multiple protein fractions from the two peaks were analyzed on SDS-PAGE (Figure 4.1 B). Majority of the protein was degraded. The secondary structure prediction of ICP27 indicates that out of 512 residues, residues 1-240 are disordered with low secondary structural element content (Figure 2.1 A). The degradation is likely induced by the ease of cleavage of the disordered region in the full-length protein. In spite of increasing protease inhibitor content in buffers and decreasing the time of purification process, the final protein was always found in the degraded form. Hence the approach of designing different constructs of ICP27 was employed to obtain a stably expressing protein (Figure 4.2). The constructs were designed around the predicted helix-rich region, based on PSIPRED and Pfam domain predictions.

Systematic cloning, cultivation and purification of ICP27 constructs were performed with support from the Protein Production Platform, housed in our laboratory at

Nanyang Technological University (Gräslund et al., 2008). Out of a total of 93 constructs designed, only 17 constructs could be successfully purified (Table 4.1).

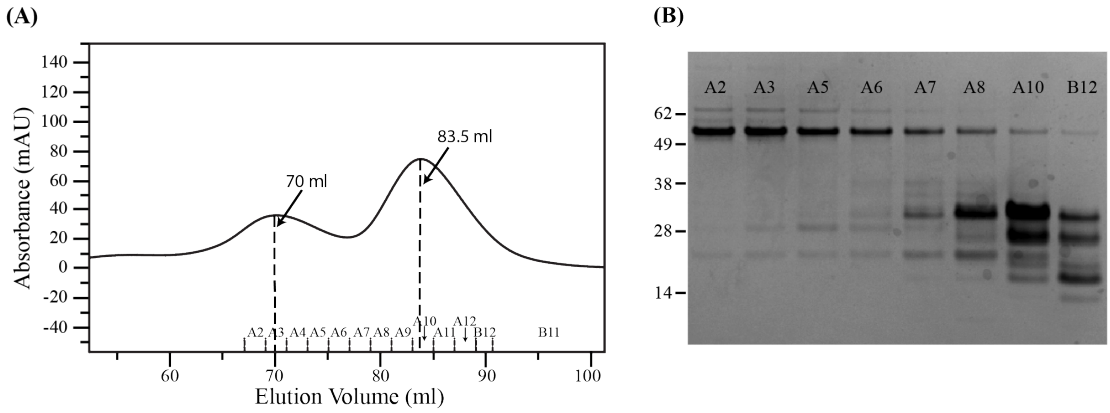


Figure 4.1. ICP27 protein purification profile.

(A) Gel Filtration chromatogram of ICP27 (full-length) protein shows the eluted protein in 20 mM Na-HEPES pH 7.5, 300 mM NaCl, 10 % (v/v) Glycerol and 1 mM TCEP with a flow-rate of 1.2 ml/min. Two elution peaks are obtained at 70 ml and 83.5 ml. (B) The SDS-PAGE profile of the elution fractions shows that ICP27 degrades during the purification process.

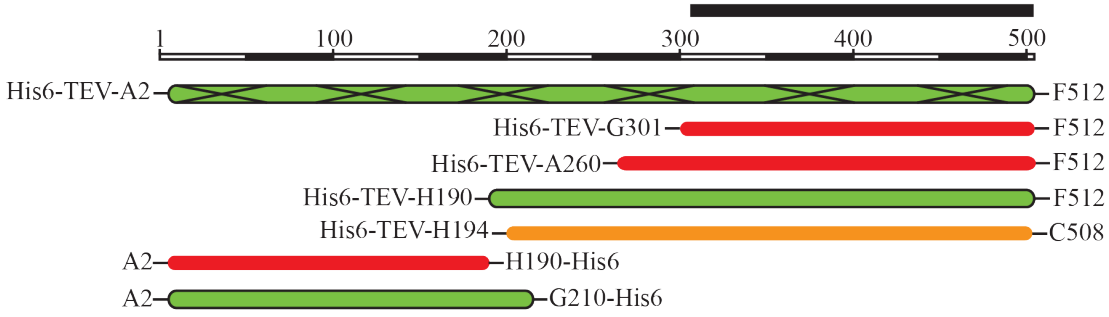


Figure 4.2 ICP27 construct design.

Top black bar represents the predicted Pfam domain. The green bars represent the constructs that succeeded in small-scale expression screen. Solid black border around the bar indicates that the construct could be scaled up and purified by 2-step purification. The green bar with crossed lines represents heavily degraded final product. The red bars indicate clones that failed in expression screening and orange bar indicates low expressing clone.

Figure 4.3 shows the purification profile and purity on SDS-PAGE of the construct ICP27$^{190\text{-}512}$, which was used for further experiments. The observed elution volume for the protein was approximately 77.5 ml, which was in between the calculated

elution volumes for monomeric and dimeric forms of ICP27$^{190-512}$. Slight degradation was always observed, which other researchers also have reported (Zhao et al., 2010; Zhi & Sandri-Goldin, 1999). Final purification yields per 750 ml of bacterial culture are summarized in Table 4.2.

Table 4.1 Statistics for ICP27 construct design

CONSTRUCTS	#
Designed	93
Good Expression in Small-Scale	27
Purified after Scale-up	17

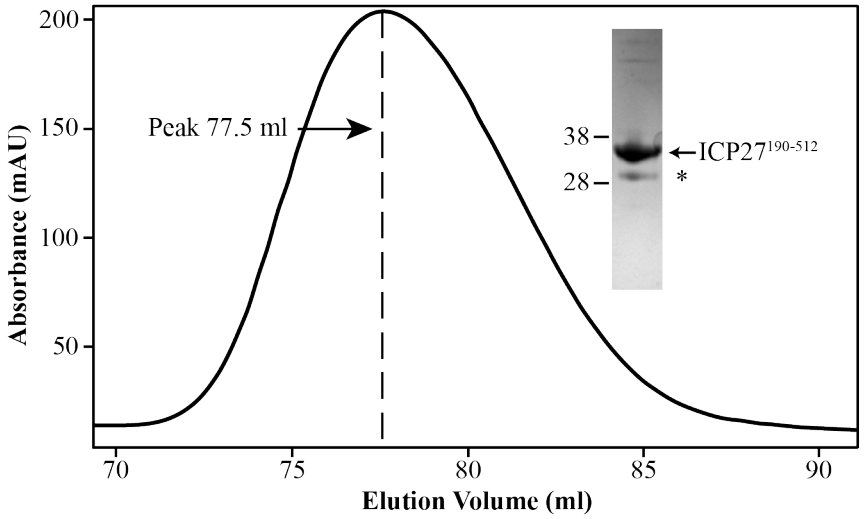


Figure 4.3 ICP27$^{190-512}$ purification profile.

Gel Filtration chromatogram of ICP27$^{190-512}$ shows that the protein eluted in 20 mM Na-HEPES pH 7.5, 300 mM NaCl, 10 % (v/v) Glycerol and 1 mM TCEP. The flow-rate of the column was 1.2 ml/min. Peak elution is at about 77.5 ml. The inserted image shows the SDS-PAGE profile of the final concentrated protein. The arrow shows the band for ICP27$^{190-512}$ at ~38 kDa. Slight degradation of the protein is observed as marked by the asterix (*).

Table 4.2 Purification yields of ICP27 constructs

CONSTRUCT	PROTEIN YIELD	REMARKS
Full-length ICP27	4.5 mg	Heavy degradation
ICP27$^{190-512}$	7 mg	Slight degradation
ICP27$^{190-512}$ (SeMet)	10.2 mg	Slight degradation
ICP2$^{72-210}$	3.7 mg	Multiple bands on PAGE

4.1.2. CRYSTALLIZATION OF ICP27

Full-length ICP27 precipitated in the initial crystallization set-ups. This could be due to the rapid degradation of the protein. In the JCSG+ screen, only ICP27$^{190-512}$ formed crystals, in multiple buffer conditions. Some conditions produced plate-like morphology and some others produced needle clusters, all of which were highly fragile. In order to obtain more structured and rigid crystals, the conditions were optimized. Additive screens were performed for a few of them. A condition containing 0.25 M Sodium thiocyanate, 20 % Polyethylene glycol (PEG) 3500 and the additive Benzamidine hydrochloride (2 %) produced extremely tiny but cubical crystals. After further optimization, the best condition for growing large, 3D crystals, was found to be 0.2 M Sodium thiocyanate and 4 % Benzamidine hydrochloride with 2:1 ratio of protein: reservoir solution (500 µl of 0.2 M Sodium thiocyanate + 100 µl of 20 % Benzamidine hydrochloride). The protein showed polymorphism in crystal morphology, under the same crystallization conditions. It crystallized as thick, tapering needles and as multiple, stacked cube-like crystals. The crystals formed very quickly within 5-6 hours. Figure 4.4 shows the crystals obtained and how the morphology changed with optimization.

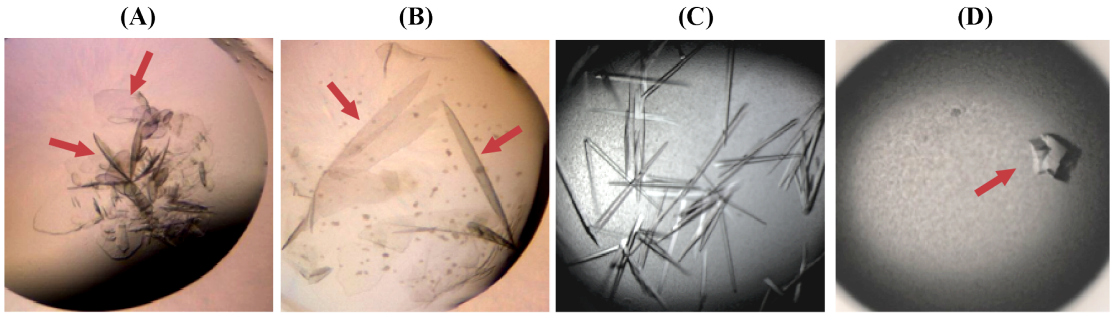


Figure 4.4 Crystals of ICP27$^{190-512}$.

(A) and (B) show crystals of ICP27$^{190-512}$ obtained in initial screening with JCSG+ suite, after 1 day of incubation at 20 °C. Condition (A) contains 0.2 M Ammonium formate, 20 % PEG 3350 and condition (B) contains 0.2 M Sodium thiocyanate, 20 % PEG 3350. The red arrows point to the flaky, thin, dragonfly wing-like plates that are extremely fragile. (C) and (D) show crystals obtained after use of additive screen and optimization. Both the conditions contain 0.2 M Sodium thiocyanate, 4 % Benzamidine hydrochloride, incubated at room temperature. Crystals were obtained in 1 day (overnight). Polymorphism is seen clearly, as the protein forms rod-like crystals (C) as well as stacked cube-like crystals (D).

For the other protein constructs, mainly those expressing N-terminal half of the protein, other screens in addition to the JCSG+ suite were used. These screens included Index, Proplex and Morpheus. However none of the screens and conditions produced any crystal for these proteins. Predictions show that N-terminal half of ICP27 lacks any secondary structural elements (Figure 2.1 A). It is known that it is impossible for intrinsically disordered regions to crystallize.

Once the crystals are formed, generally they are transferred to a buffer solution with cryoprotectant and stored in liquid Nitrogen for further use in X-ray experiments. Mostly such cryo-solutions are prepared by adding 20-25 (v/v) % Glycerol to the crystallizing condition. Other commonly used cryoprotectants include polyethylene glycols (PEG), jeffamine and even some oils like paratone. The crystals of ICP27$^{190-512}$ were highly fragile and easily destroyed while fishing out of the crystallization drops. A thick skin always formed over the drop. When the skin was removed and the crystals exposed to air for more than 5 min while fishing, visible browning and degradation of crystals was observed. Most crystals when put under the X-ray beam did not diffract or diffracted very poorly (~8-10 Å). After trying variations of cryoprotectant and reducing exposure time to minimum, better diffracting crystals were obtained. Only the reservoir solution which was equilibrated in the well when mixed with 50 (v/v) % Glycerol to a final concentration of 28 % could be used as cryoprotectant. No other pre-made cryoprotectant with same composition yielded any results.

4.1.3. DATA PROCESSING AND STRUCTURE REFINEMENT

A Single Anomalous Dispersion (SAD) dataset was collected for ICP27$^{190-512}$ crystals, at peak wavelength for Selenium (0.97893 Å) at NSRRC, Taiwan (13B1 beamline). It was then refined at 2.0 Å. Multiple Anomalous Dispersion (MAD) data were also collected for the same crystals. Native dataset was collected and solved at 2.7 Å at DLS, UK and at 3.0 Å at AS, Australia. All datasets were integrated and scaled using HKL2000 package (Otwinowski and Minor, 1997). Autosol from the PHENIX suite was used for experimental phasing (Terwilliger et al., 2009; Adams et al., 2010). Autosol consists of Hybrid Substructure Search (HySS) for finding heavy atom sites, Phaser for calculation of experimental phases from SAD data, and

RESOLVE for density modification (McCoy et al., 2004; McCoy et al., 2007; Terwilliger, 2003). The initial model was built using Autobuild. Molecular replacement using Phaser solved the SeMet ICP27$^{190\text{-}512}$ structure. Native structure was determined by molecular replacement using Phaser and the SeMet structure as search model. Final model was obtained by refinement using Maximum Likelihood (ML) method of phenix.refine program and manual model building using COOT (Afonine et al., 2012; Emsley et al., 2010). A total of 5% of the structure-factor amplitudes were excluded from refinement for R_{free} calculations. Final model quality was validated using Comprehensive validation software from PHENIX suite and using SFCHECK method from CCP4 suite (Vaguine et al., 1999; Winn et al., 2011). Data statistics are summarized in the Table 4.3. All structural figures of ICP27$^{190\text{-}512}$ were prepared using PyMOL, unless mentioned otherwise (Baker et al., 2001). Electrostatic surfaces were calculated using PDB2PQR and displayed using the Adaptive Poisson-Boltzmann Solver (APBS) plugin in PyMOL (Dolinsky et al., 2004; Dolinsky et al., 2007). The schematic diagram of ICP27$^{190\text{-}512}$ was constructed using TopDraw software from the CCP4 suite (Bond, 2003). The coordinates and structure factors were deposited in the Protein Data Bank (PDB) with the accession code 5BQK.

4.1.4. ICP27$^{190\text{-}512}$ STRUCTURE HAS A NOVEL FOLD

The crystal structure of ICP27$^{190\text{-}512}$ consists of 3 molecules in the asymmetric unit (ASU). Similar overall structure is observed in both crystal morphologies. Each molecule consists of 10 distinct α-helices (α1-α10) of varying lengths separated by 9 loops (L1-L9) (Figure 4.5 A). However, the His6 tag and residues from 190 to 242 are not visible in the structure, probably due to high mobility. Close contacts in the crystal indicate the likelihood of a head-to-tail dimer formation between two monomers. N-terminal helices and the C-terminal region from each monomer form arm-like extensions that make contacts with the other monomer of the ICP27 dimer (Figure 4.5 B, C). The predicted Zinc-finger-like motif is clearly visible along with a tetrahedrally coordinated zinc ion in the centre. Unlike the previous predictions, the residues involved in formation of the motif are C400, H479, C483 and C488. A DALI search on the coordinates for the complete structure as well as the zinc-binding region did not yield any hits, indicating no similarity between them and the known protein

structures in the structure data bank. Hence the overall structure and the zinc-binding region have novel folds.

Table 4.3. Crystallographic Data, Phasing and Refinement Statistics.

	SELENOMETHIONINE DATASET	NATIVE DATASET
DATA COLLECTION		
Radiation source	NSRRC 13B1	DLS
Wavelength (Å)	0.97893	0.9795
Space group	C2	C2
Unit cell dimensions		
a, b, c (Å)	156.98 120.33 79.04	157.88 121.4 79.76
α, β, γ (°)	90.0 100.46 90.0	90 100.46 90
Resolution range (Å)	23.80 - 2.0 (2.07 - 1.997)	39.16 - 2.73 (2.83 - 2.73)
R_{merge}	9.8 (67.7)	17.2 (78.4)
Mean I/σ(I)	8.5 (1.46)	8.7 (2.5)
Completeness (%)	91.5 (93.91)	99.9 (99.8)
Redundancy	2.6 (2.5)	7.5 (7.6)
PHASING		
No. of sites	20	
Initial FOM	0.312	
FOM after DM	0.721	
Phasing power (acentric/centric)	1.382/1.053	
Anomalous phasing power	0.715	
R_{cullis} (acentric/centric)	0.848/0.831	
Anomalous R_{cullis}	0.859	
REFINEMENT		
R_{work} / R_{free} (%)	21.68 / 24.62	22.7 / 27.66
No. atoms (non-H)	6383	6375
Protein	6090	6264
Ligand/Ion	3 Zn	3 Zn
Water	290	108
MODEL QUALITY		
R.M.S. deviations	0.022	0.01
Bond lengths (Å)	1.749	1.2
Bond angles (°)		
RAMACHANDRAN PLOT		
Favoured Region (%)	98.0	96.0
Allowed Region (%)	1.87	3.38
Outliers (%)	0.13	0.62

Highest resolution values are written in parenthesis

FOM: Figure of Merit; DM: Density Modification

R_{merge} = [Σ Σ $|I_i$ - <I>| / Σ Σ I_i] x 100; where I_i is the intensity measurement of reflection h and <I> is the average intensity from multiple observations

Phasing power = F_H / E, where FH is the calculated structure factor contributed by the heavy atom and E is the root mean square lack of closure

R_{cullis} = Σ $||F_{PH} - F_P| - F_H|$ / Σ$|F_{PH} - F_P|$, where F_{PH} and F_P are the observed structure factor of the native and derivative datasets respectively and F_H is the calculated structure factor contributed by the heavy atom (Se)

R_{work} = [Σ $||F_o| - |F_c||$ / Σ $|F_o|$] x 100; where F_o and F_c are the observed and calculated structure factors respectively

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

The core structure encloses a hydrophobic pocket formed by the helices 3, 4 and 5 and the long loop L2 connecting helices 2 and 3. Sequence alignment of ICP27 with its homologues from α-, β- and γ-herpes sub-families, shows that the residues lining this pocket are either partially or completely conserved (Figure 4.6 and 4.7). The loop forms the base of the pocket region and is predicted to be highly flexible as can be understood from the high B-factor values of its residues (Figure 4.8). This flexible loop is called 'Cap' since it covers the pocket from the bottom. Electron densities for 4 residues in this loop, 298 to 301 forming GQGG motif, are missing in the structure, indicating high mobility. A few charged residues are present in this pocket, which are well conserved. The hydrophobic pocket of the 1st molecule houses the C-terminal end residues from 501 to 512 of the 2nd molecule, bringing the two monomers in close contact. We have termed this C-terminal end as the C-tail. In the N-terminal end, residues 242 to 278 in helices 1 and 2 of the structure together form a hinged arm-like structure. This arm wraps around the neighbouring subunit. Figure 4.5 B shows the proximity of the arm to the second monomer in the structure.

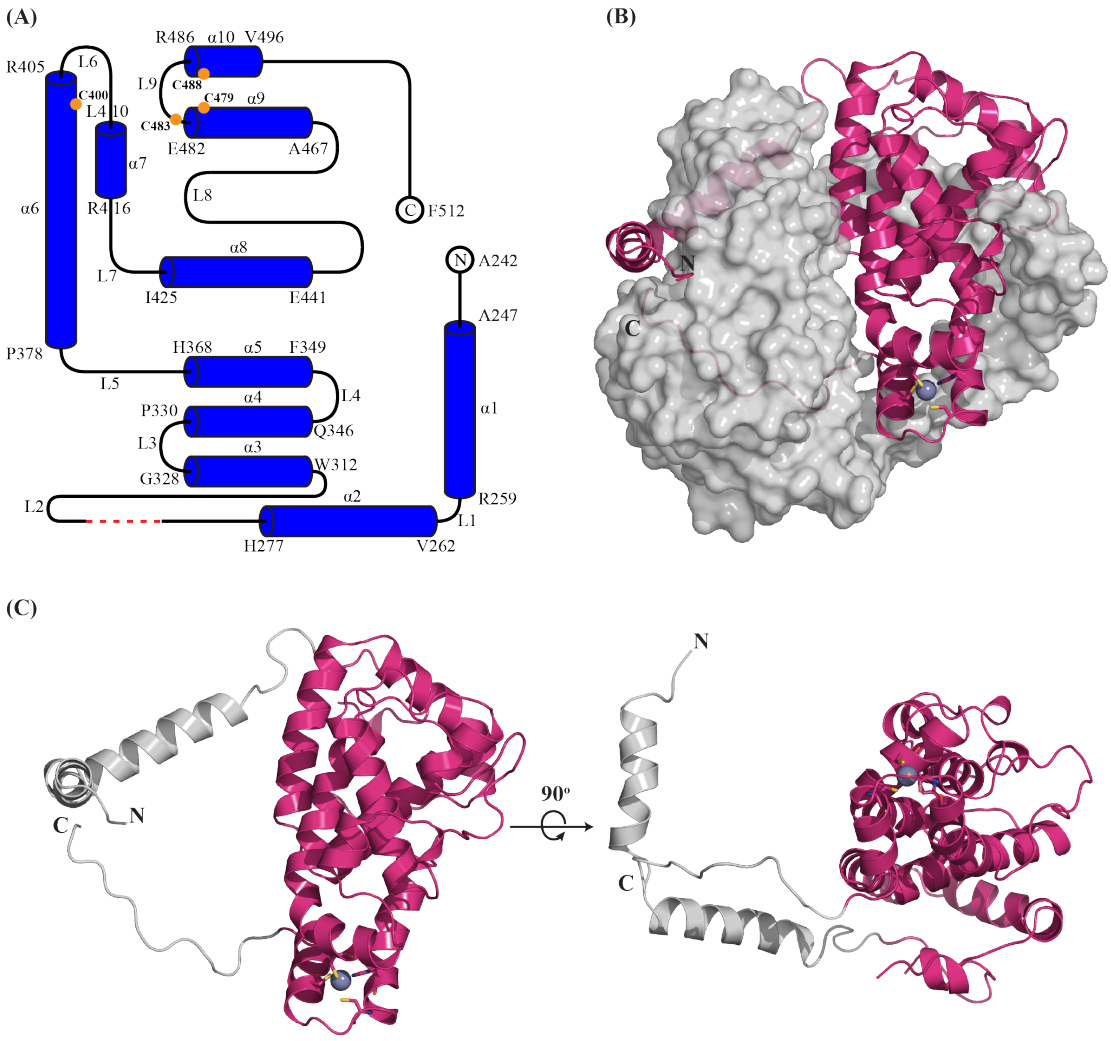


Figure 4.5 Overall structure of ICP27$^{190\text{-}512}$.

(A) The schematic representation shows that ICP27$^{190\text{-}512}$ monomer consists of 10 α-helices (α1-α10), separated by 9 loops (L1-L9) of varied lengths. The structure starts at A242 and ends at F512 with the residues from 190 to 241 and His6-TEV tag missing in the structure due to high mobility. The first and the last residues of each helix are indicated. The orange circles (●) indicate positions of the zinc-binding residues. The red-dashed line (- - -) indicates the 4 residues (G298-G301) with missing electron densities. The schematic is prepared using Topdraw software in the CCP4 suite. (B) The magenta cartoon representation of ICP27$^{190\text{-}512}$ monomer is in close contact with the second monomer showed as grey surface representation. The N-terminal arm of first monomer is wrapped around the second monomer, whereas the C-terminal region is buried. The zinc ion is represented as purple-grey sphere and is tetrahedrally coordinated by the cysteine and histidine residues shown as sticks. (C) The cartoon representation of ICP27$^{190\text{-}512}$ monomer is shown in two views rotated 90° along X-axis. The helical core is coloured in magenta and the N- and C-termini are coloured in grey. The rotated figure shows the hollow region between the globular domain and the N-terminal arm.

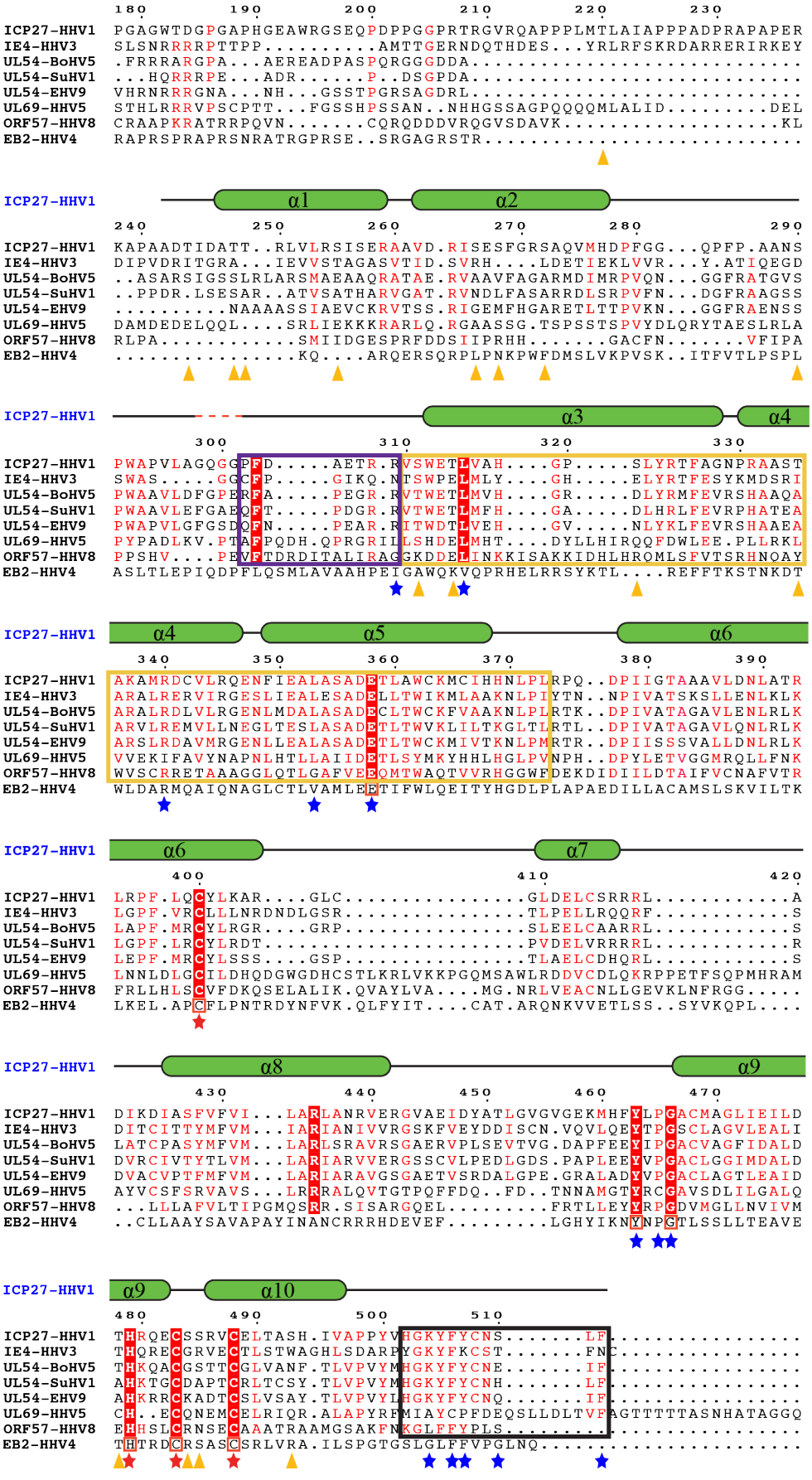

ICP27-HHV1 PGAGWTDGPGAPHGEAWRGSEQPDPPGGPRTRGVRQAPPPLMTLAIAPPPADPRAPAPER
IE4-HHV3 SLSNRRRPTTPP.....AMTTGERNDQTHDES...YRLRFSKRDARRERIRKEY
UL54-BoHV5 .FRRRARGPA...AEREADPASPQRGGGDDA.....
UL54-SuHV1 ...HQRRRPE...ADR.....P..DSGPDA.....
UL54-EHV9 VHRNRRRGNA...NH...GSSTPGRSAGDRL.....
UL69-HHV5 STHLRRRVPSCPTT...FGSSHPSSAN..NHHGSSAGPQQQQMLALID.....DEL
ORF57-HHV8 CRAAPKRATRRPQVN.....CQRQDDDVRQGVSDAVK.....KL
EB2-HHV4 RAPRSPRAPRSNRATRGPRSE..SRGAGRSTR.....
ICP27-HHV1
α1
α2
ICP27-HHV1 KAPAADTIDATT..RLVLRSISERAAVD.RISESFGRSAQVMHDPFGG...QPFP.AANS
IE4-HHV3 DIPVDRITGRA...IEVVSTAGASVTID.SVRH....LDETIEKLVVR...Y.ATIQEGD
UL54-BoHV5 ..ASARSIGSSLRLARSMAEAAQRATAE.RVAAVFAGARMDIMRPVQN...GGFRATGVS
UL54-SuHV1 ..PPDR.LSESAR..ATVSATHARVGAT.RVNDLFASARRDLSRPVFN...DGFRAAGSS
UL54-EHV9 .....NAAAASSIAEVCKRVTSS.RIGEMFHGARETLTTPVKN...GGFRAENSS
UL69-HHV5 DAMDEDELQQL....SRLIEKKKRARLQ.RGAASSG.TSPSSTSPVYDLQRYTAESLRLA
ORF57-HHV8 RLPA.....SMIIDGESPRFDDSIIPRHH.....GACFN.....VFIPA
EB2-HHV4 .....KQ....ARQERSQRPLPNKPWFDMSLVKPVSK..ITFVTLPSPL
ICP27-HHV1
α3
α4
ICP27-HHV1 PWAPVLAGQGGPFD.....AETR.RVSWETLVAH....GP.....SLYRTFAGNPRAAST
IE4-HHV3 SWAS.....GGCFP.....GIKQ.NTSWPELMLY....GH.....ELYRTFESYKMDSRI
UL54-BoHV5 PWAAVLDFGPERFA.....PEGR.RVTWETLMVH....GR.....DLYRMFEVRSHAAQA
UL54-SuHV1 PWAAVLEFGAEQFT.....PDGR.RVTWETLMFH....GA.....DLHRLFEVRPHATEA
UL54-EHV9 PWAPVLGFGSDQFN.....PEAR.RITWDTLVEH....GV.....NLYKLFEVRSHAAEA
UL69-HHV5 PYPADLKV.PTAFPQDH.QPRGRILLSHDELMHT....DYLLHIRQQFDWLEE.PLLRKL
ORF57-HHV8 PPSHV....PEVFTDRDITALIRAGGKDDELINKKISAKKIDHLHRQMLSFVTSRHNQAY
EB2-HHV4 ASLTLEPIQDPFLQSMLAVAAHPEIGAWQKVQPRHELRRSYKTL....REFFTKSTNKDT
ICP27-HHV1
α4
α5
α6
ICP27-HHV1 AKAMRDCVLRQENFIEALASADETLAWCKMCIHHNLPLRPQ..DPIIGTAAAVLDNLATR
IE4-HHV3 ARALRERVIRGESLIEALESADELLTWIKMLAAKNLPIYTN..NPIVATSKSLLENLKLK
UL54-BoHV5 ARALRDLVLRGENLMDALASADECLTWCKFVAAKNLPLRTK..DPIVATAGAVLENLRLK
UL54-SuHV1 ARVLREMVLLNEGLTESLASADETLTWVKLILTKGLTLRTL..DPIVATAGAVLQNLRLK
UL54-EHV9 ARSLRDAVMRGENLLEALASADETLTWCKMIVTKNLPMRTR..DPIISSSVALLDNLRLK
UL69-HHV5 VVEKIFAVYNAPNLHTLLAIIDETLSYMKYHHLHGLPVNPH..DPYLETVGGMRQLLFNK
ORF57-HHV8 WVSCRRETAAAGGLQTLGAFVEEQMTWAQTVVRHGGWFDEKDIDIILDTAIFVCNAFVTR
EB2-HHV4 WLDARMQAIQNAGLCTLVAMLEETIFWLQEITYHGDLPLAPAEDILLACAMSLSKVILTK
ICP27-HHV1
α6
α7
ICP27-HHV1 LRPF.LQCYLKAR....GLC.....GLDELCSRRRL.....A
IE4-HHV3 LGPF.VRCLLLNRDNDLGSR.....TLPELLRQQRF.....S
UL54-BoHV5 LAPF.MRCYLRGR....GRP.....SLEELCAARRL.....S
UL54-SuHV1 LGPF.LRCYLRDT.....PVDELVRRRRL.....R
UL54-EHV9 LEPF.MRCYLSSS....GSP.....TLAELCDHQRL.....S
UL69-HHV5 LNNLDLGCILDHQDGWGDHCSTLKRLVKKPGQMSAWLRDDVCDLQKRPPETFSQPMHRAM
ORF57-HHV8 FRLLHLSCVFDKQSELALIK.QVAYLVA....MG.NRLVEACNLLGEVKLNFRGG.....
EB2-HHV4 LKEL.APCFLPNTRDYNFVK.QLFYIT....CAT.ARQNKVVETLSS..SYVKQPL....
ICP27-HHV1
α8
α9
ICP27-HHV1 DIKDIASFVFVI...LARLANRVERGVAEIDYATLGVGVGEKMHFYLPGACMAGLIEILD
IE4-HHV3 DITCITTYMFVM...IARIANIVVRGSKFVEYDDISCN.VQVLQEYTPGSCLAGVLEALI
UL54-BoHV5 LATCPASYMFVM...LARLSRAVRSGAERVPLSEVTVG.DAPFEEYIPGACVAGFIDALD
UL54-SuHV1 DVRCIVTYTLVM...LARIARVVERGSSCVLPEDLGDS.PAPLEEYVPGACLGGIMDALD
UL54-EHV9 DVACVPTFMFVM...LARIARAVGSGAETVSRDALGPE.GRALADYVPGACLAGTLEAID
UL69-HHV5 AYVCSFSRVAVS...LRRRALQVTGTPQFFDQ..FD..TNNAMGTYRCGAVSDLILGALQ
ORF57-HHV8 ...LLLAFVLTIPGMQSRR.SISARGQEL.....FRTLLEYYRPGDVMGLLNVIVM
EB2-HHV4 ..CLLAAYSAVAPAYINANCRRRHDEVEF.....LGHYIKNYNPGTLSSLLTEAVE
ICP27-HHV1
α9
α10
ICP27-HHV1 THRQECSSRVCELTASH.IVAPPYVHGKYFYCNS.....LF.....
IE4-HHV3 THQRECGRVECTLSTWAGHLSDARPYGKYFKCST.....FNC.....
UL54-BoHV5 THKQACGSTTCGLVANF.TLVPVYMHGKYFYCNE.....IF.....
UL54-SuHV1 AHKTGCDAPTCRLTCSY.TLVPVYMHGKYFYCNH.....LF.....
UL54-EHV9 AHKRRCKADTCSLVSAY.TLVPVYLHGKYFYCNQ.....IF.....
UL69-HHV5 CH..ECQNEMCELRIQR.ALAPYRFMIAYCPFDEQSLLDLTVFAGTTTTTASNHATAGGQ
ORF57-HHV8 EHHSLCRNSECAAATRAAMGSAKFNKGLFFYPLS.....
EB2-HHV4 THTRDCRSASCSRLVRA.ILSPGTGSLGLFFVPGLNQ.....

Figure 4.6 Comparison of ICP27 with its homologues.

ICP27$^{190\text{-}512}$ (HHV1) sequence is aligned with its various homologues from α-herpes family: IE4 from HHV3, UL54 from Bovine Herpes Virus 5 (BoHV5), UL54 from Suid Herpes Virus 1 (SuHV1), UL54 from Equine Herpes Virus 9 (EHV9); β-herpes family: UL69 from HHV5; and γ-herpes family: ORF57 from HHV8. Secondary structural elements of ICP27$^{190\text{-}512}$ are drawn above the sequences as green α-helices, black lined loops with red dashed-line showing position of the missing 4 residues. Residues conserved in all sequences are highlighted in red boxes with white font; residues with similar sequences are highlighted in red font. Residues involved in the zinc-binding motif of ICP27$^{190\text{-}512}$ are marked with red star (★). Black box (□) indicates residues forming the C-tail, yellow box (□) indicates the hydrophobic pocket region and purple box (□) indicates the flexible cap region as observed in ICP27$^{190\text{-}512}$ structure. Residues involved in key interactions between the hydrophobic pocket and the C-tail are marked with a blue star (★). Orange triangle (▲) indicates the position of all residues phosphorylated in the in vitro assay and analyzed by mass-spectrophotometry. Multiple sequence alignment was performed using Clustal Omega and sequence similarities rendered using ESPript3. The sequence of ICP27 homolog from HHV4, i.e. EB2 protein, is manually aligned to the above multiple sequence alignment. The conserved residues are boxed in red (□).

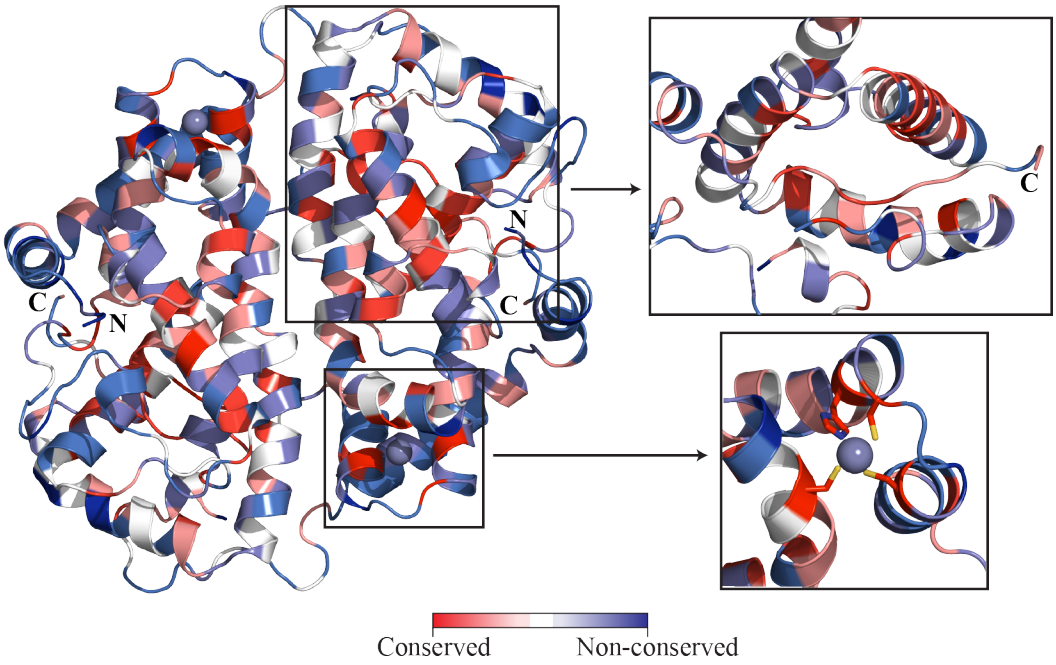


Figure 4.7 Mapping of amino-acid sequence conservation on ICP27-CTD structure.

Based on multiple sequence alignment of ICP27 and its homologs, the conserved amino-acid residues are mapped on the ICP27-CTD structure. Depending on the extent of conservation, the residues are colored from red, for highest conservation, to blue, for non-conserved residues. The upper inset shows the hydrophobic pocket, lined with conserved residues, along with the C-tail inserted in it. The lower

inset displays the highly conserved zinc-binding residues. UCSF Chimera was used for rendering by sequence-conservation. The final images were prepared in PyMOL.

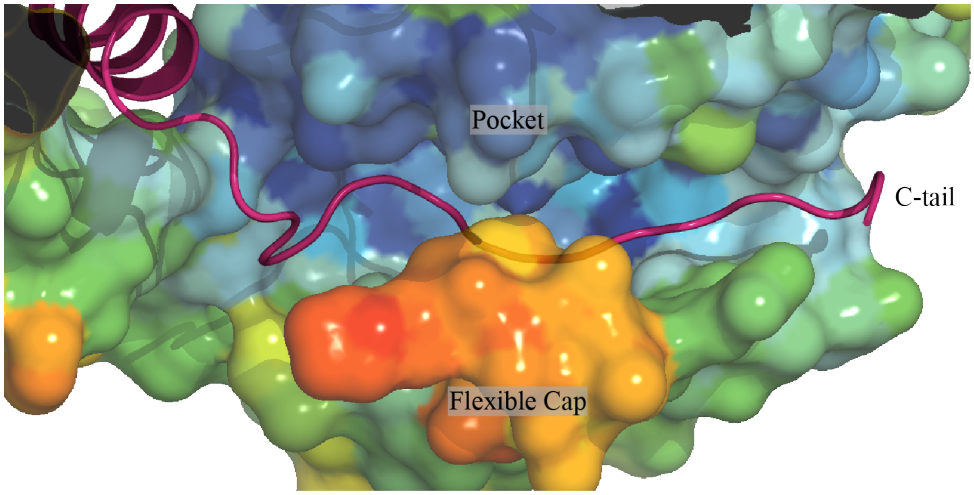


Figure 4.8 Hydrophobic pocket of ICP27$^{190\text{-}512}$.

ICP27$^{190\text{-}512}$ C-tail of one molecule (magenta cartoon) is lodged in the hydrophobic pocket of 2nd molecule (surface representation). The lower wall of the pocket forms a flexible cap (as seen by high B-factor value shown in orange-red colours).

4.1.5. VARIOUS INTERACTIONS OF ICP27$^{190\text{-}512}$

The C-tail consists mostly of hydrophobic residues involved in some key interactions and some conserved charged residues. From the sequence alignment of ICP27 homologues and mapping of the conserved residues on structure, we can see that the C-tail region is highly conserved (Figure 4.6 and 4.7) and may be of importance for its function. It makes extensive interactions with the residues of the hydrophobic pocket. The C-tail conserved residues F506 and Y507 form hydrophobic interactions with the conserved L315 in the pocket. The extended backbone structure of the key C-tail residues F506 and Y507 is stabilized by the buried E358, which makes hydrogen bonds to the amino groups of F506 and Y507 (Figure 4.9 B). E358 is also highly conserved in all the homologues; it forms a polar interaction with the conserved R340 in helix 4. Another C-tail residue K504, which is conserved in all α-herpes homologues, makes polar contacts with E358. E347 is buried with its carboxylate group only 4 Å away from K504 (Figure 4.9 C). The series of buried charged residues, being unusual, are indicative of importance of the pocket region in

ICP27 structure and/or function. The N-terminal arm formed by helices 1 and 2 of one monomer lie in close proximity to the globular region of the second monomer. The hydrophobic and aromatic residues in the arm bring about the interaction between these two regions.

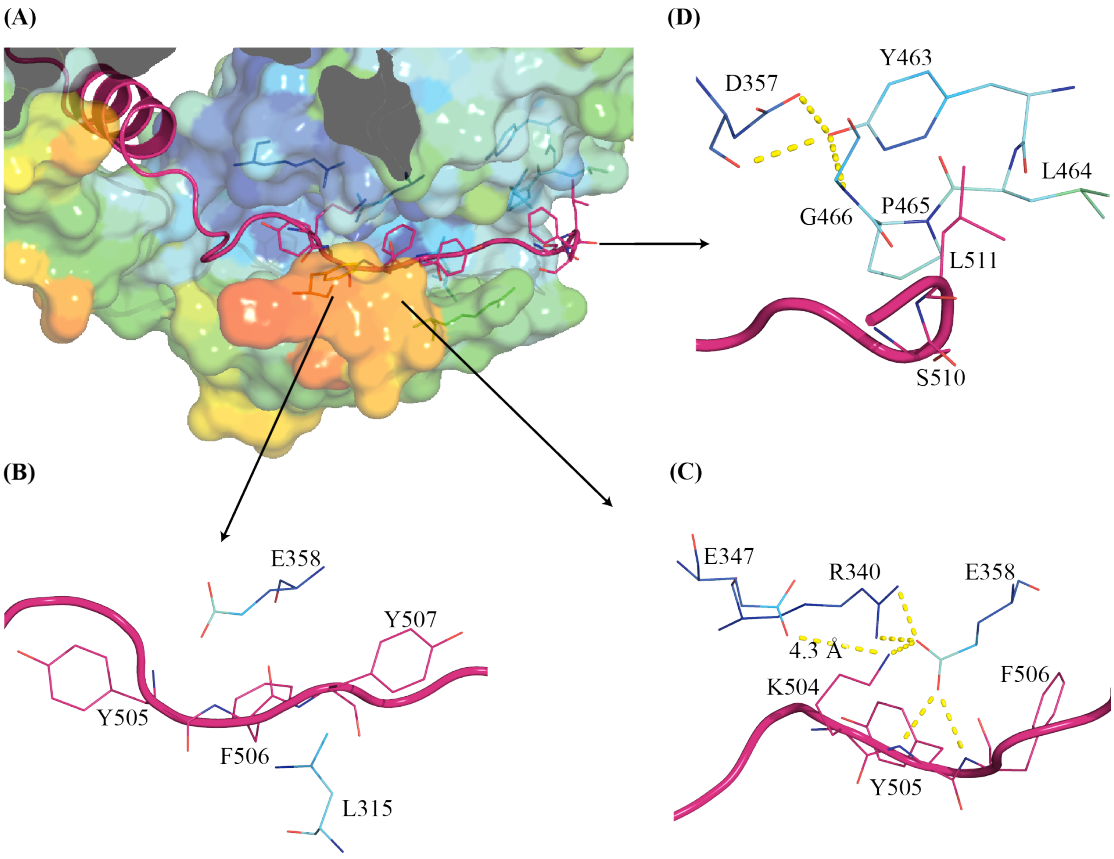


Figure 4.9 Interactions of the C-tail residues with the conserved hydrophobic pocket region.

(A) ICP27$^{190\text{-}512}$ C-tail of one molecule (magenta cartoon representation) is lodged in the hydrophobic pocket of 2nd molecule (surface representation coloured by B-factor values). The lower wall of the pocket forms a flexible cap (as seen by high B-factor value shown in orange-red colours). (B) Conserved C-tail residues (magenta cartoon and stick representation) Y505, F506 and Y507 sandwiched via hydrophobic interactions with conserved residues L315 and E358 from the pocket region. (C) Several buried charges are conserved around the C-tail interaction region, making polar interactions (yellow dashes). E358 forms salt-bridge with Y505 and F506 of C-tail. R340 from the pocket and K504 from the C-tail make stabilizing interactions with E358. E347 has its carboxylate group only 4.3 Å away from K504. (D) C-tail residues S510 and L511 make hydrophobic interactions with the conserved YXPG motif. The motif is, in turn, stabilized by the conserved D357.

The end residues of the C-tail extend just beyond the pocket and are involved in interactions with residues from the same monomer. L511 and F512 form hydrophobic interactions with the loop L2, which is part of the N-terminal arm. S510 and L511 interact with a conserved motif YXPG in loop L8. The buried, charged residue E357 stabilizes this motif by forming a charge-compensating side chain to the N-termini of helix 8 (Figure 4.9 D). The extended loop L2 forms a cap beneath the C-tail. The side-chain of F303 in L2 interacts with the main chain of F506 and side-chain of L504 in the C-tail. The aromatic region of Y507 in the C-tail is in close proximity to the aliphatic region of R309. Both these residues are conserved in the ICP27 homologues.

ICP27 has been noted to interact with various proteins within the host cell nucleus and cytoplasm (Sandri-Goldin, 2011). Though most studies focus on the interactions with the N-terminal region, some interactions have been mapped to the C-terminal domain of ICP27. The electrostatic potential surface of ICP27$^{190-512}$ reveals multiple charged patches that may have potential to bind to different interacting partners (Figure 4.10). Two large positively charged surfaces are formed on the top surface (Figure 4.10 A). Similarly, along the bottom surface, two negatively charged patches are formed in close proximity and close to the flexible cap region (Figure 4.10 B). A closer analysis of these regions reveals that multiple conserved residues line these patches. One asparagine and six arginine residues form the positively charged patch, of which five residues (R417, R418, R435, N438 and R442) are either conserved or semi-conserved (Figure 4.10 A, inset). The residues D304 and E306 in the flexible cap loop, and the semi-conserved residues E347 and E351 form the negatively charged surface (Figure 4.10 B, inset). Hydrophobic regions are also visualized on the surface of the protein (Figure 4.10 B) as well as along the contact interface between the two molecules. The charged and uncharged regions have potential interactions with other macromolecules.

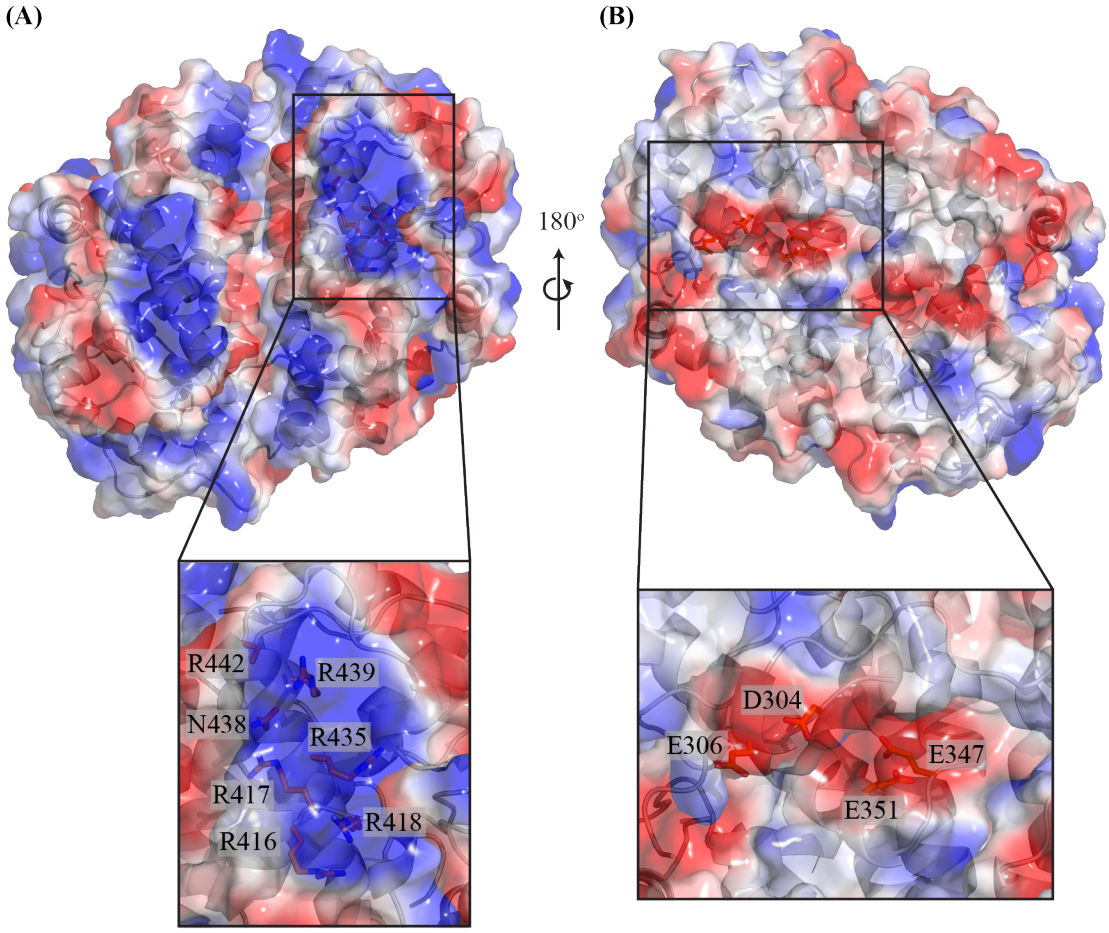


Figure 4.10 Electrostatic potential surface map of ICP27$^{190\text{-}512}$.

The negative, positive and uncharged surfaces are coloured in red, blue and white respectively. Overall, the protein exhibits multiple charged patches scattered over the surface. (A) The orientation of the molecule is the same as that seen in Figure 4.6. The top surface consists of two large positively charged regions. The inset shows the underlying residues (in sticks). Of the seven residues forming the charged surface region, R417, R418, R435, N438 and R442 are either conserved or semi-conserved in other herpes families. (B) A negatively charged patch is visible on the reverse side of the protein. D304, E306, E347 and E351 form this charged patch. D304 and E306 are part of the flexible cap region that interacts with the C-tail, while E347 and E351 are semi-conserved. Large uncharged, hydrophobic regions are also visible on the surface. The charged and uncharged regions of ICP27$^{190\text{-}512}$ may be potential regions for bonded and non-bonded interactions, respectively. The surface charges were calculated using PDB2PQR and rendered using APBS plugin in PyMOL.

4.1.6. ICP27 HAS A NOVEL ZINC-BINDING MOTIF

The structure reveals one zinc-binding site per molecule. The presence of zinc in the crystal was confirmed by sharp absorption spectrum when the crystal was irradiated

with X-ray at wavelength 1.2837 Å, i.e. at 9.6586 keV, the X-ray absorption K-edge energy for zinc. The occupancy of the zinc in each monomer is 1.0. The residues C400, H479, C483 and C488 tetrahedrally coordinate each zinc ion (Figure 4.11, inset). Previous studies had predicted that a zinc-finger motif was formed by the residues C483, C488, H502 and C508 based on sequence alignment and conservation of some residues in ICP27 and its homologues. As seen from the structure, H502 and C508 are buried within the hydrophobic pocket with little scope of coordinating the zinc ion. H479, C483 and C488 form a short helix-loop-helix motif. H479 and C488 are found in the C- and N-terminal loop of the helices 9 and 10 respectively, while C483 is found in the interconnecting loop 9. C400, in helix 6, is 78 residues away from the other zinc-binding residues and coordinates with the zinc ion completing the tetrahedral coordination of the motif. This residue helps in bringing together the residues otherwise far apart. All the 4 residues are conserved in ICP27 homologues from other hosts as well as homologues from other human herpes viruses, as seen from the sequence alignment (Figure 4.6).

A closer look at the zinc-binding region reveals that the 4 residues involved in direct coordination are stabilized by various interactions. The coordination sphere of the Zn is tetrahedral with coordination distances in range of 1.9-2.5 Å for the cysteine sulphur and 2.0 Å for the histidine nitrogen. H479 interacts with E482 that forms a salt-bridge with R405 in helix 6 (Figure 4.11). This R405 is in close proximity to C400, thus helping to make the zinc-motif more compact and sturdy. C483 forms a hydrogen bond with R480, via its peptide backbone. The partial positive charge provided by this interaction indicates that C483 might predominantly be in its deprotonated form. C400 and C488 are both stabilized by several well-ordered water molecules. The overall structure of this zinc-binding motif is very compact. Though there is not a clear indication in any of the previous studies whether the zinc-binding region of ICP27 is catalytic or not, the compactness of the motif as seen in the structure suggests that it probably has a structural function rather than a catalytic one. Most zinc-finger motifs are compact and involve only coordinating residues that are in close proximity to each other sequentially. The zinc-binding motif of ICP27 differs in its structure as it includes a sequentially distant residue (78 amino acids away) that is highly conserved. When, as mentioned above, the comparison of this structure

coordinates to other zinc-finger containing proteins using DALI search did not yield any results, the zinc-binding motif of ICP27 has a novel structure.

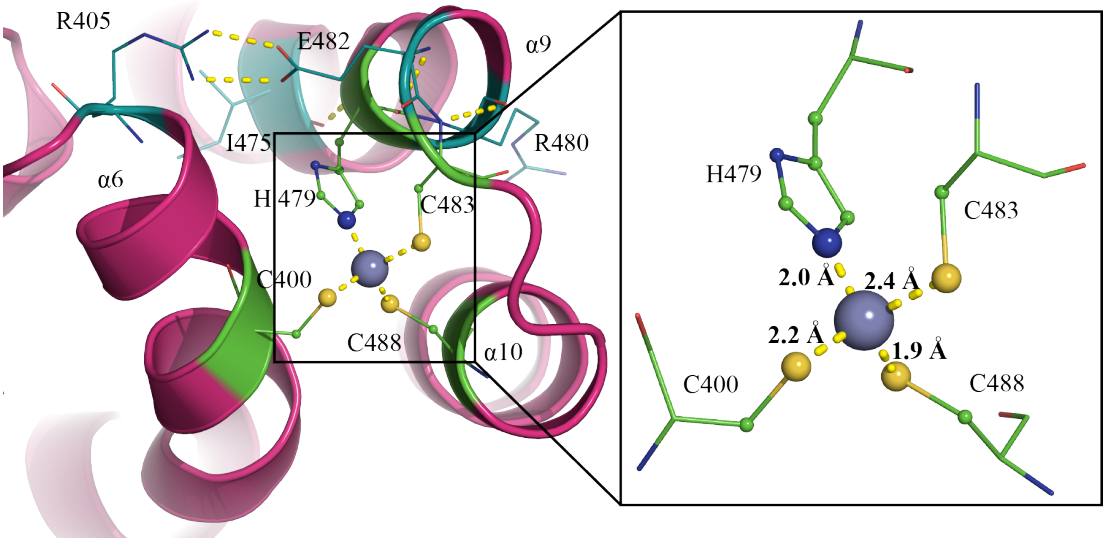


Figure 4.11 Zinc-binding motif of ICP27$^{190-512}$.

The cartoon representation of the zinc-binding motif shows the protein in magenta and zinc ion as grey sphere. The zinc ion is tetrahedrally coordinated via C400, H479, C483 and C488 as shown by yellow-dashed lines. H479, C483 and C488 form helix-loop-helix structure. H479 is back-coordinated by E482, which in turn is stabilized by I475 and R405. R405 is in close proximity to the zinc-binding residue C400, hence providing more stability between helix and the helix-loop-helix. C483 is coordinated by R480.

4.1.7. ICP27$^{190-512}$ CAN UNDERGO DIMERIZATION

From examining the crystal structure, it is seen that of the three molecules of ICP27$^{190-512}$ in the asymmetric unit (ASU), two are in close proximity in a probable dimeric form and the 3rd molecule pairs up with a molecule from the adjoining ASU in a similar dimeric interaction. The two molecules of the dimer are positioned in a yin-yang orientation giving the dimer a coffee bean like shape. The dimensions of this homodimer are 63.4 x 59.4 x 40.7 Å. A prominent central groove is formed at the interface of the dimer, which is about 11-13 Å wide (Figure 4.12). Although potentially functionally interesting, the residues lining the groove do not show conservation indicating any key structural or functional roles, neither do they show any charges suggesting nucleic acid binding is found (Figure 4.7).

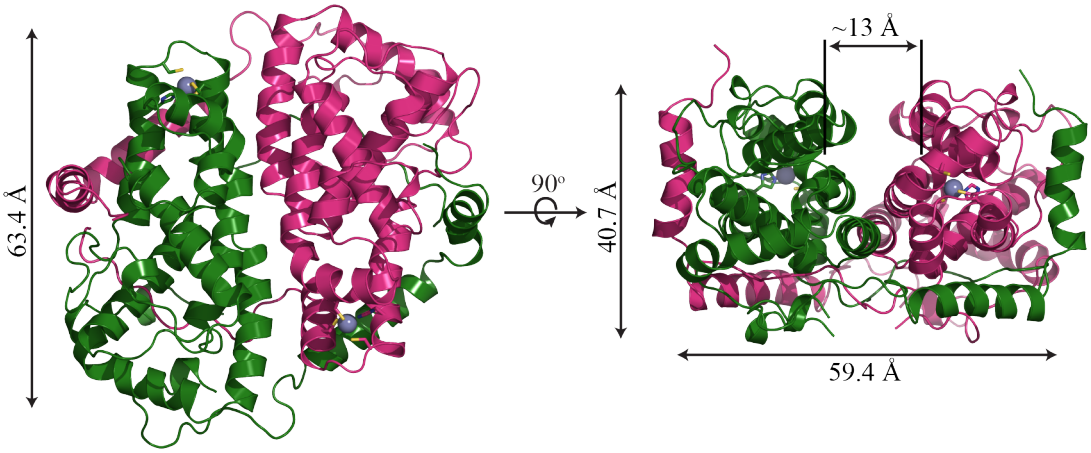


Figure 4.12 ICP27$^{190\text{-}512}$ in dimeric conformation.

The cartoon representation of ICP27$^{190\text{-}512}$ is shown in its dimeric conformation. The two monomers (in magenta and green) are positioned in a yin-yang orientation. The dimensions of the homo-dimer are measured in PyMol software. The 90° rotated view shows a prominent groove running along the length of the dimer interface.

Multi-angle laser light scattering (MALS) measurements can be used as a non-destructive technique to determine various properties of proteins including molecular weight and hence the oligomeric state of protein. MALS coupled with size-exclusion chromatography were employed to investigate if purified ICP27$^{190\text{-}512}$ exists as dimer in solution. A single major UV absorbance peak was obtained that was highly homogenous as indicated by light-scattering signal (Figure 4.13). The weighted-average molar mass of this sample was 75.91 ± 1.5 kDa that is exactly twice the molecular mass of ICP27$^{190\text{-}512}$ (i.e. 37.87 kDa). This result supports that at higher concentrations ICP27$^{190\text{-}512}$ is primarily present as a dimer in solution.

A MALS experiment requires protein concentrations to be in a range of 3-8 mg/ml. However in the cell, the ICP27 concentrations will be much lower and hence to confirm the oligomeric state of ICP27$^{190\text{-}512}$ at lower concentration analytical ultracentrifugation sedimentation equilibrium experiment was performed. The data could be best fitted to a monomer-dimer self-association model (Figure 4.14). The association constant (Ka) for this model is 1.8113 x 10^6 M^{-1} with an RMS deviation of 0.005 and chi-square value of 0.76828. The values for the other models are tabulated in Table 4.4.

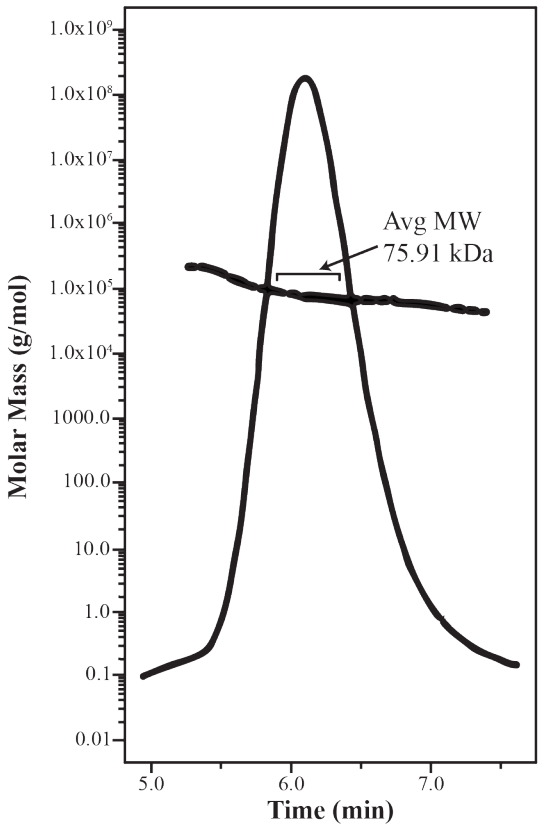


Figure 4.13 Elution profile of ICP27$^{190\text{-}512}$ with MALS.

ICP27$^{190\text{-}512}$ elutes in 6 min from the SEC coupled with MALS. The molar mass corresponding to the elution peak is calculated to be 75.91 ± 1.5 kDa indicating that ICP27$^{190\text{-}512}$ is dimeric in solution. Buffer used was 20 mM Na-HEPES pH 7.5, 300 mM NaCl and 5 % (v/v) glycerol filtered though 0.1 µm filter. A flow-rate of 0.3 ml/min was maintained.

Table 4.4 Analytical centrifugation data of ICP27$^{190\text{-}512}$ after fitting in various monomer:n-mer self-association models.

ICP27$^{190\text{-}512}$	N	Log Ka	Ka (M^{-1})	Kd (M)	R.M.S.D.	CHI-SQUARE
Monomer-Monomer	1	-	-	-	0.0271	16.313
Monomer-Dimer	2	6.258	1.8113×10^{5}	0.552×10^{-5}	0.0050	0.7683
Monomer-Trimer	3	9.092	1.2359×10^{7}	0.809×10^{-7}	0.0070	1.5265
Monomer-Tetramer	4	13.021	1.0495×10^{10}	0.953×10^{-10}	0.0081	3.1062

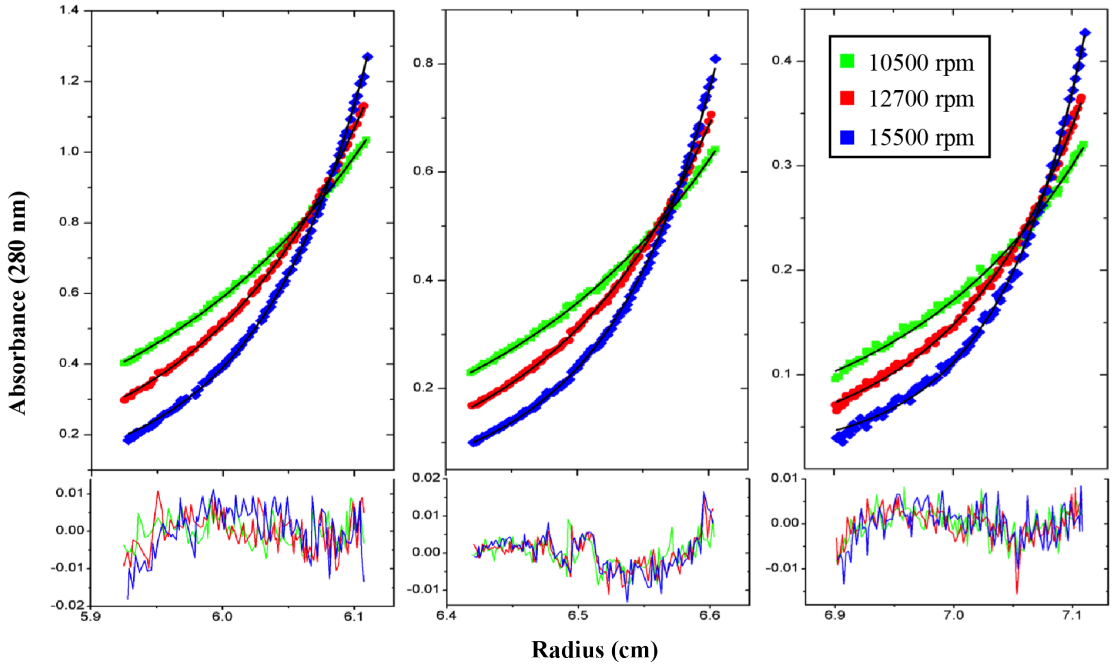


Figure 4.14 Analytical ultracentrifugation profile of ICP27$^{190\text{-}512}$.

The sedimentation equilibrium analysis of ICP27$^{190\text{-}512}$ is shown with experimental data in the top graph and the residuals of fit in the lower graph. Data was collected for 3 different concentrations of protein (0.8, 0.5 and 0.3 A.U.) at 10500, 12700 and 15500 rpm. The data is fit to a monomer:dimer model thus showing that ICP27$^{190\text{-}512}$ is capable of forming self-associating dimers even at low concentrations.

4.1.8. ICP27$^{190\text{-}512}$ CAN UNDERGO PHOSPHORYLATION

Considering the number of functions ICP27 performs and the many proteins it interacts with, it is obvious that ICP27 undergoes some form of post-translational modification that helps it switch between different functions. Phosphopeptide mapping technique was employed for detection of phosphorylation in ICP27. This study, which was performed *in vitro* and *in vivo,* reported that ICP27 is phosphorylated in multiple regions with protein kinase A (PKA) and casein kinase II (CKII) (Zhi & Sandri-Goldin, 1999). However phosphorylation sites in the C-terminal domain of ICP27, for which we have determined the structure, have not previously been reported.

To explore the possibility of ICP27$^{190-512}$ being phosphorylated and to identify the potential residues, the purified protein was phosphorylated *in vitro* using 3 different kinases namely PKA, CKII and PKC. The protein was run on SDS-PAGE and the band corresponding to molecular weight of ICP27$^{190-512}$ was sent for phosphorylation detection using mass spectrometry. The residues, confirmed to be phosphorylated by the respective kinases, are shown in Table 4.5 and are marked in the Figure 4.6. Interestingly, a number of sites are found in the region surrounding the C-tail interaction site, including S290, S311, T314 and S321 in the loop (L2) region, which caps the C-tail (Figure 4.15 A). This region capping the C-tail has a high B-factor value indicating higher flexibility, as discussed before (Figure 4.8). Similarly, multiple phosphorylation sites are found in close proximity of the zinc-binding motif, including S266, S268, S272, T478, S484, S485 and S493 (Figure 4.15 B). These may be involved in phosphorylation-induced local conformational changes. The phosphorylation of residues in the cap region could be of some functional importance.

Table 4.5 Phosphorylated residues of ICP27$^{190-512}$

PKA	CKII	PKC
T244	T244	T220
S266	T249	T244
S268	S266	T248
S290	S272	S255
S311	S290	S268
T314	T478	S272
S493	S484	S290
	S493	T314
		S321
		T335
		S484
		S485
		S493

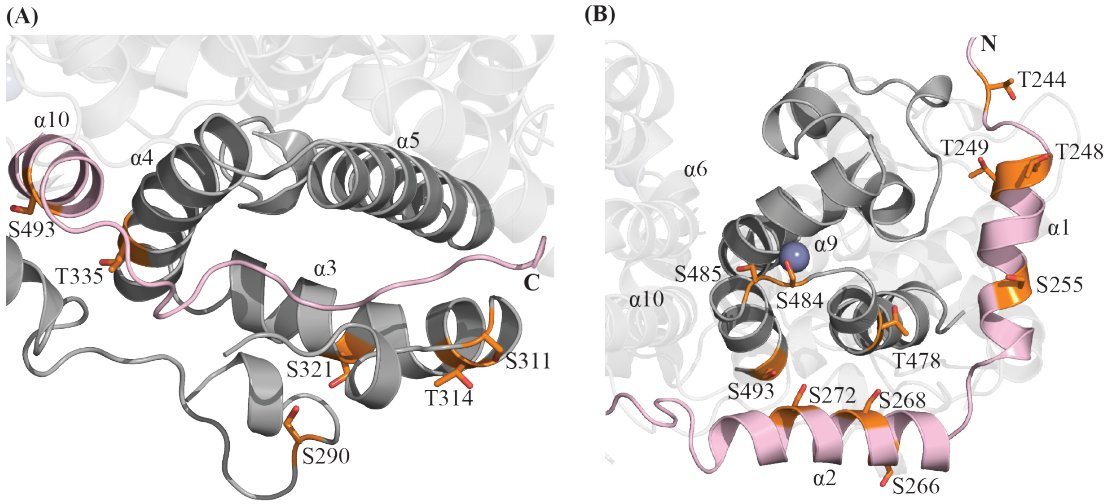


Figure 4.15 Phosphorylation sites of ICP27-CTD.

In vitro phosphorylation assays using PKA, CKII and PKC kinases, followed by analysis using mass-spectrophotometry reveals several phosphorylated sites in ICP27-CTD (shown in orange). Molecule one is depicted in grey and molecule two in light pink. (A) Many of the observed sites were concentrated in the flexible loop capping the C-tail, mainly S290, S311, T314 and S321. Considering the high flexibility of the capping loop, there exists a possibility of phosphorylation-induced local conformation changes. (B) Most other sites are concentrated in the region in close to the zinc-binding motif, mainly S266, S268, S272, T478, S484, S485 and S493.

To study the effect of phosphorylation on ICP27$^{190\text{-}512}$, partial-digestion assay was performed on PKA-phosphorylated ICP27$^{190\text{-}512}$. The initial hypothesis was that ICP27$^{190\text{-}512}$ exists as dimer and on phosphorylation, may undergo monomerization that would expose the N-terminal arm and C-tail, making them available for its interacting partners. It was expected that if the protein became monomeric on phosphorylation, both the termini of the protein would be more sensitive to chymotrypsin. However, in preliminary experiments, no significant difference between control and phosphorylated samples were seen (Figure 4.16). The lane with highest chymotrypsin concentration in control protein (Figure 4.16 A) shows higher degree of proteolytic activity than unphosphorylated one (Figure 4.16 B). However such high protease concentrations are unlikely to occur in cellular environment and hence should not be considered in isolation. All the other samples showed little difference in their proteolytic profiles.

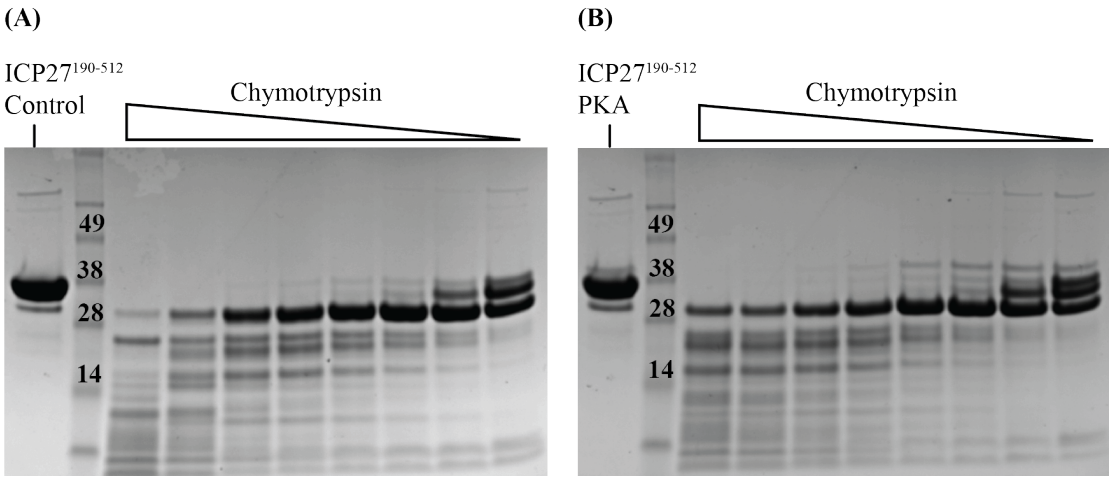


Figure 4.16 Chymotrypsin partial-digestion assay of PKA-phosphorylated ICP27$^{190\text{-}512}$.

SDS-PAGE profiles of (A) control and (B) PKA-phosphorylated ICP27$^{190\text{-}512}$ are shown. Each protein sample was incubated with decreasing amounts of chymotrypsin (1:5 dilution series) for 30 min at room temperature. It was hypothesized that the N-terminal arm and the C-tail may be exposed on phosphorylation and hence the protein would be more susceptible to digestion by protease. However, the results form chymotrypsin partial-digestion assay did not show any significant differences between control and phosphorylated ICP27$^{190\text{-}512}$.

In order to study the effect of phosphorylation on ICP27 in solution, the next approach used was to generate phosphomimick mutants. Earlier study by Zhi and Sandri-Goldin had demonstrated that ICP27 is phosphorylated *in vitro* by the PKA, CKII and PKC in decreasing levels (Zhi & Sandri-Goldin, 1999). Based on this study and on the earlier mass spectrometry results, the seven PKA phosphorylation sites of ICP27 (Table 4.5) were individually mutated from serine/threonine to glutamine. In spite of multiple trials and optimizations of site-directed mutagenesis, S290 and T311 failed to produce mutants. Figure 4.17 shows the expression profile of the mutants that succeeded in transformation and expression. S266E, S268E and T314E expressed well whereas S493E expressed at comparatively lower levels. S266E, S268E and T314E were scaled-up and purified for analysis by MALS (data not shown).

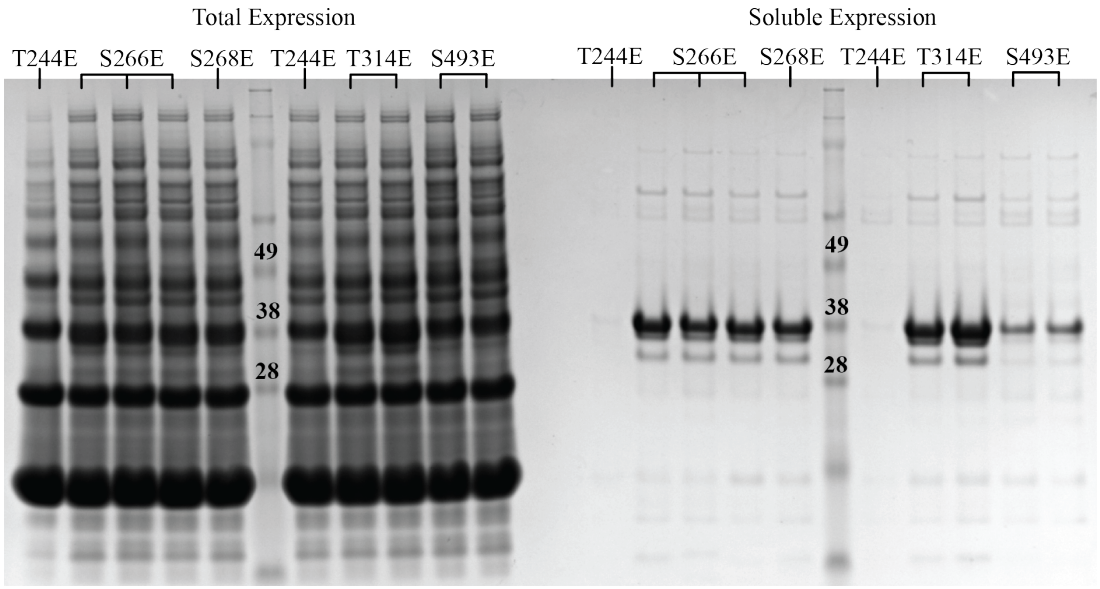


Figure 4.17 Small-scale expression screening of phosphomimetic mutants.

Phosphomimetic mutants with serine or threonine mutated to glutamate were cloned and expressed. Of the seven possible PKA phosphorylation sites, five sites were successfully cloned, but only four of them expressed the soluble protein in small-scale screening. The figure shows total expression of the mutants on the left and soluble expression (after IMAC) on the right.

4.1.9. ICP27$^{190-512}$ INTERACTION WITH RNA/ SSDNA

ICP27 is a regulator of gene expression of viral genes. It is an important protein that helps export intronless viral mRNA out of the nucleus for protein synthesis. This interaction of ICP27 with mRNA is extremely important for virus replication cycle. It is also involved in recruiting host export factors to the bound mRNA. ICP27 is implicated in the export of mRNAs of the immediate early and the late genes, but not the early genes. This indicates that the protein should not only be recognizing single-stranded nucleic acid, but should also be recognizing a consensus sequence or a specific fold to differentiate between the mRNAs. Soliman and Silverstein have suggested that ICP27 consists of 3 KH domains within the C-terminal half of the protein and that they are extremely important for the protein to bind to RNA (Soliman & Silverstein, 2000; Soliman & Silverstein, 2000). Structures of KH domains have been determined for other proteins and have a typical structure consisting of a mixture of alpha helices and beta sheets and a GXXG loop that is conserved (Grishin, 2001; Valverde et al., Edwards, & Regan, 2008). It is the GXXG loop that is responsible for binding to the single-stranded nucleic acids. However, in structure of

$ICP27^{190-512}$, no KH domains are present and we can conclude the prediction strategy used gave a false prediction.

To investigate whether $ICP27^{190-512}$ is capable of binding to the single-stranded nucleic acids, in spite of the absence of the canonical KH domains, thermal stability shift assays were performed for $ICP27^{190-512}$ with several different RNA and DNA oligomers. No study has so far identified a consensus sequence that ICP27 may recognize. However, Corbin-Lickfett et al., using HSV-1 glycoprotein C (gC) DNA, demonstrated that ICP27 shows preference to GC-rich sequences, containing guanine in tandem repeats (Corbin-Lickfett et al., 2009). Additionally, Massimelli et al. demonstrated that ICP27 homolog in KSHV, ORF57, possibly recognizes and interacts with a 9 nucleotide sequence in polyadenylated nuclear RNA (PAN), which is called Mta-responsive element II (MRE-II) (Massimelli et al., 2011). Based on these studies, poly-G, poly C and MRE-II ssRNA sequences were used. Random ssDNA sequences were also used for the assay. Table 4.6 lists the single stranded nucleic acid sequences.

Table 4.6 Nucleic acid sequences used for TSSA.

ID	NUCLEIC ACID SEQUENCE	MW (Da)
RNA 1	5'-GUA AAG UAU GGA UUU AAA AAA-3'	6779
RNA 2	5'-GGG GGG GGG-3'	3045
RNA 3	5'-AAA AAA AAA-3'	2902
RNA 4	5'-CCC CCC CCC-3'	2685
DNA 1	5'-/5Phos/TGC AT-3'	1622.3
DNA 2	5'-ATG GGA C-3'	2201.8
DNA 3	5'-GCC GAG GCT GTA GTC CCC CCA GGC CCC TAG-3'	9129.9

As seen from the Figure 4.18 A, the melting temperature (Tm) of the protein shows a negative shift with increase in RNA concentration. Negative thermal shifts can, in special cases, support ligand binding; but from unpublished experience in our laboratory for metal containing proteins, it is often likely that the negative shift is due to the chelation of a stabilizing metal. We, therefore, hypothesized that the

negatively charged RNA may chelate the zinc ions in the protein leading to a lower melting temperature. To compensate for the probable zinc-chelation, zinc was externally added to the reaction mixture in the form of zinc chloride (1 to 10 mM). This resulted in a tremendous positive shift in the Tm (Figure 4.18 B). This demonstrates that there is a possibility of interaction between the RNA and ICP27$^{190-512}$.

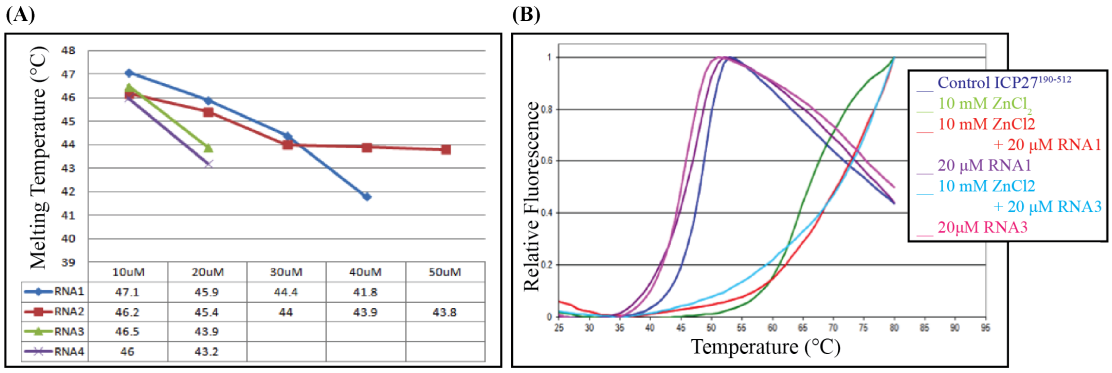


Figure 4.18 Interaction between ICP27$^{190-512}$ and RNA.

(A) The graph shows the melting temperatures of ICP27$^{190-512}$ in presence of increasing amounts of RNA oligomers. A trend of decrease in the melting temperature was observed with an increase in the RNA concentration, indicating destabilization of ICP27$^{190-512}$ with RNA. This may be due to chelation of the zinc ion of ICP27$^{190-512}$ by the negatively charged RNA molecules. The missing temperatures are a result of precipitation of protein with addition of RNA. (B) The graph shows the change in melting temperature of ICP27$^{190-512}$ from 48 °C of the reference protein (with no additives) to about 45 °C with RNA added to 72 °C with RNA and zinc ions added. The chelation effect of RNA was negated by addition of zinc in the reaction.

4.2. EB2 C-Terminal Domain from EBV may be Similar to ICP27 C-Terminal Domain

EB2 protein from Epstein-Barr virus (EBV) is a functional homologue of ICP27 from HSV-1. The overall sequence similarity of EB2 and ICP27 is about 19.65 %, with conservation seen mainly in the C-terminal domain; specifically in the zinc-binding motifs, the C-tail and the hydrophobic pocket. Experiments on ICP27 gave an insight on how to approach expression and purification of the homologous proteins. However, EB2 also had other problems different from those of ICP27. This chapter collates the information obtained from various trials with EB2.

4.2.1. Cloning, Expression and Purification of EB2

Full-length gene of BMLF1 encoding EB2 protein was successfully cloned and expressed in *E. coli* cells. Similar to ICP27, EB2 also suffered from heavy degradation. Its secondary structure prediction indicates that out of 438 residues, the region encompassing residues 1-200 is mostly disordered with no secondary structural elements (Figure 2.1 B). Hence constructs of the EB2 were designed to obtain a stably expressing protein (Figure 4.19). The constructs were designed around the predicted helix-rich region, based on PSIPRED and Pfam domain predictions.

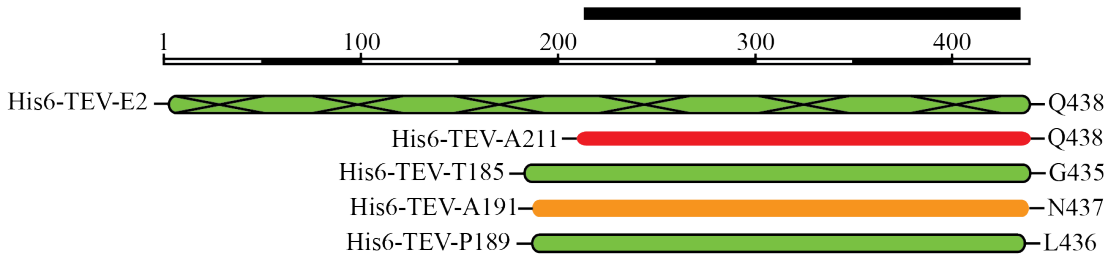


Figure 4.19 EB2 construct design.

Top black bar represents the predicted Pfam domain. The green bars represent the constructs that succeeded in small-scale expression screen. Solid black border around the bar indicates that the construct could be scaled up and purified by two-step purification. The green bar with crossed lines represents heavily degraded final product. The red bars indicate clones that failed in expression screening and orange bar indicates low expressing clone.

Cloning, cultivation and purification of EB2 constructs were performed. Of the 53 designed constructs, only 5 could be purified in large scale (Table 4.7). Only constructs designed around the stable C-terminal half of the protein expressed well. Figure 4.20 shows the purification profile and purity on SDS-PAGE of the construct $EB2^{189-436}$, which was used for further experiments. The observed elution volume for the protein is approximately 92 ml.

Table. 4.7 Statistics for EB2 constructs.

CONSTRUCTS	#
Designed	53
Good Expression in Small-Scale	8
Purified after Scale-up	5

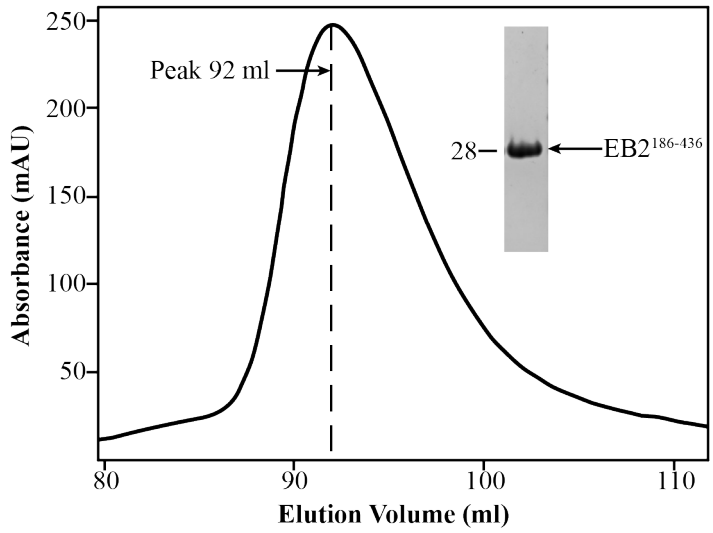


Figure 4.20 $EB2^{186-436}$ purification profile.

Gel Filtration chromatogram of $EB2^{189-436}$ using Superdex 200 column eluted in 20 mM Na-HEPES pH 7.5, 300 mM NaCl, 10 % Glycerol and 1 mM TCEP. Flow-rate of elution was 1.2 ml/min. Peak elution is at about 92 ml. The inserted image shows the SDS-PAGE profile of the final concentrated protein. The arrow shows the $EB2^{189-436}$ band at ~28 kDa.

Though the yield of $EB2^{189-436}$ is decent (4.5-6 mg per 750 ml of culture), it is not soluble in buffer beyond 1 mg/ml concentrations. All attempts to concentrate the protein further resulted in heavy precipitation and loss of protein. Trials to solubilise the protein like changing buffer systems, using buffers with high TCEP, addition of

detergents and increase in salt concentration did not help. This challenge was posed for all the constructs of EB2 that were purified.

4.2.2. CRYSTALLIZATION OF $EB2^{189-436}$

The initial screening was performed using JCSG+ suite (Qiagen). No hits were obtained in the screen. Other crystallization screens, namely Proplex, Index and Morpheus were used for initial screenings. However, due to very low concentration of the protein, the trials did not produce any result. Most of the wells had drops that were either clear or with very light precipitation, indicating that the protein concentration needed to be increased.

4.2.3. CIRCULAR DICHROISM OF $EB2^{189-436}$ SHOWS SIMILARITY TO $ICP27^{190-512}$

To check if $EB2^{189-436}$ was correctly folded after purification, a circular dichroism (CD) spectrum was measured and analyzed. $ICP27^{190-512}$ was used as reference protein. CD spectrum for $EB2^{189-436}$ was collected from 180-260 nm. The spectrum showed the curve characteristic of alpha helical structured proteins (Figure 4.21) with a maxima at 195 nm and 2 minima at 210 nm and 209 nm. The profile is similar to that of $ICP27^{190-512}$, supporting a similarity in 3D-structure between the two proteins.

Sequence alignment of EB2 to ICP27 family proteins shows that though the sequence similarity ranges between only 18 to 20 %, the C-terminal half of the protein shows a significant degree of conservation of key residues identified in the structure. The residues involved in the zinc-binding motif and the hydrophobic pocket of ICP27 are also conserved in EB2 (Figure 4.6).

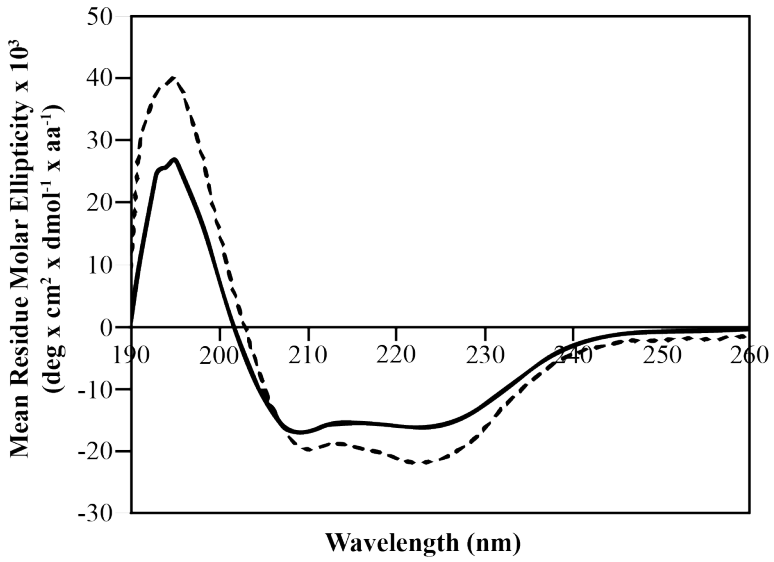


Figure 4.21 Secondary structural analysis of EB2$^{189-436}$ and ICP27$^{190-512}$.

Circular dichroism profile of EB2$^{189-436}$ shown as black solid line with ICP27$^{190-512}$ profile in black-dashed line as reference. The CD data was collected from 180-260 nm at room temperature. The maxima and minima of both profiles are comparable: ICP27$^{190-512}$ max- 194 nm, min- 209, 223 nm and EB2$^{189-436}$ max-195 nm, min- 210, 223 nm.

4.3. Multi-Construct Approach has Potential to Boost Herpes Protein Expression

Viral protein expression and purification have always been a challenge (Bowman et al., 2006; Fogg et al., 2006; Henson et al., 2011; Le Sage et al., 2013; Sam et al., 2009; Tait & Straus, 2011; Tarbouriech et al., 2006). Their properties differ somewhat from the eukaryotic and prokaryotic proteins. Based on the success of multi-construct approach for human proteins, the same strategy of designing constructs of proteins around their predicted structural domains was employed (Gräslund et al., 2008). For many herpesviral proteins, however, there are still no homologous structures, making domain predictions much more difficult than, for example, for human proteins where homologous domain structures more often exist. The aim of this study is to assess how fruitful the multi-construct approach has been for production of herpes viral proteins.

4.3.1. Construct-Design is a Critical Step in Over-Expression of Viral Proteins

A total of 94 herpes viral proteins were subjected to the multi-construct approach. 155 human proteins were also subjected to similar strategy and a comparison was done between the two sets to understand how viral proteins differ from human proteins in terms of expression in bacterial systems. Depending on the level of soluble expression, the proteins are scored on a scale of 0 to 4, 0 being no expression to 4 being highest amount of expression (Figure 4.22 A). The scores of 3 and 4 are considered good expression. Longest constructs of herpes proteins, corresponding to the full-length proteins, expressed poorly as compared to those of human proteins. Only 18 % of herpes proteins expressed well as opposed to 31 % of human proteins (Figure 4.22 B). When N- and C- truncations were introduced, the expression statistics increased to 45 % for herpes proteins and 60 % for human. Additionally, the longest constructs of herpes proteins that do express have difficulties in yielding enough amounts of protein as needed for structural and biophysical studies.

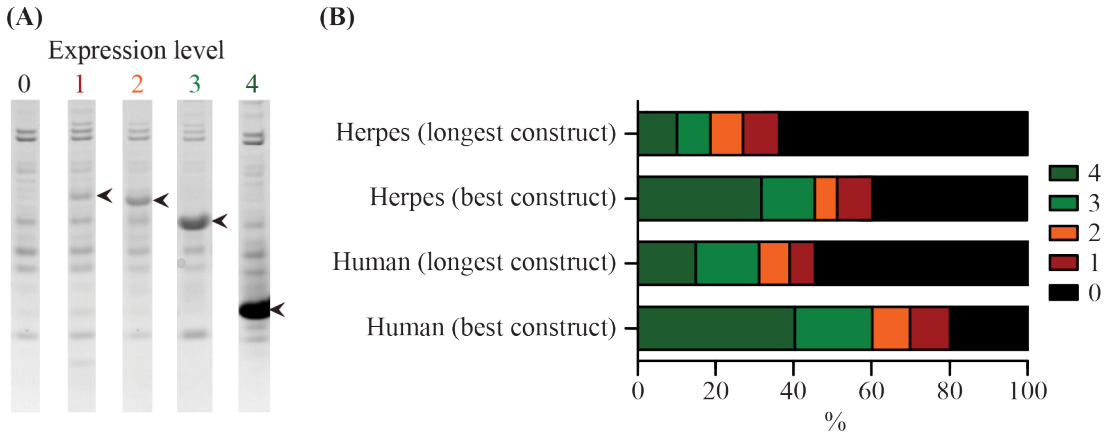


Figure 4.22 SDS-PAGE after affinity purifications and success rates after the multi construct approach.

(A) To investigate the soluble expression level, small-scale affinity purifications are made of each target and analyzed on SDS-PAGE. The expression level is scored on a subjective scale from 0-4 where 0 is no soluble expression and 4 is the highest. An arrow indicates the target band. (B) To investigate the effectiveness of the multi-construct approach, a control set corresponding to the longest construct was chosen for both the human and herpes viral proteins. After creating multiple constructs for each target, there is marked increase in success rate of constructs expressing soluble protein at score 3 or 4. The number of targets/constructs for Herpes n=88 and for human n=155.

Constructs were designed using sequence-based predictions like that of secondary structural elements, Pfam domains, regions of low complexity and structural similarity to known protein structures (Gräslund et al., 2008; Savitsky et al., 2010). However, similar approaches yielded less beneficial results in case of herpes proteins. From our data, we see that domain boundary predictions using bioinformatic tools often do not correspond to the good constructs, as shown in the three examples in Figure 4.23. In case of ORF29b from Kaposi's Sarcoma-associated Herpesvirus (KSHV), Pfam prediction encompassed the entire length of the protein as a single domain (Figure 4.23 A). No low complexity regions (LCR) were predicted. Secondary structural predictions displayed multiple alpha helices and beta strands throughout the protein. However, the full-length construct of ORF29b expressed extremely poorly as seen in Figure 4.23 D lane 1. K9 is a viral homologue of the well-documented cellular interferon regulatory factor (IRF). Pfam prediction showed two distinct domains (Figure 4.23 B). In the case of K9, bioinformatic predictions correlated well with the successful construct designs. Also, structures for both the domains from the homologues are available, adding to the reliability of the prediction results, and K9 DNA binding domain structure has been solved in our lab to a very

resolution of less than 2 Å. The main observation in this case was the difference in levels of expression depending on position of translational start site (Figure 4.23 D lanes 4, 5). Constructs 4 and 5 had only a difference of four residues, but the effect on expression was very significant.

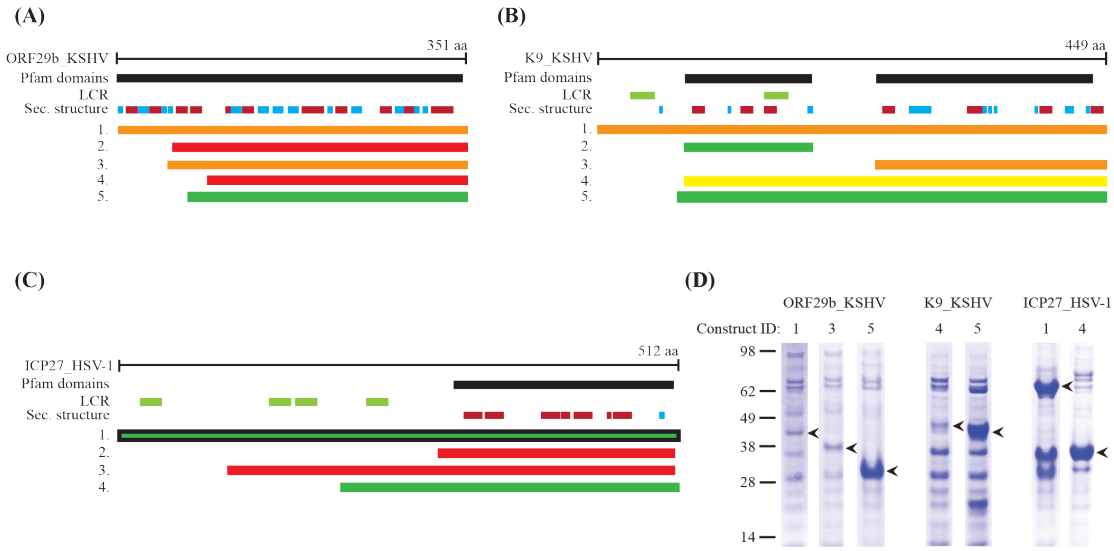


Figure 4.23 Typical examples of rational construct design for three viral proteins.

(A), (B) and (C) depict the typical construct designs. The black lines show the span of the protein along with their name and total number of residues. The black boxes depict predicted Pfam domains. The light green bars (■) are the predicted low complexity regions (LCR). The red (α-helixes ■) and blue (β-strands ■) boxes represent secondary structure predictions, where the majority of the prediction methods were in agreement. Only predicted elements of >3 residues are shown. The lower bars depict the longest construct, as well as some truncated constructs where red (■) indicates a score 0, orange (■) score 1, yellow (■) score 2, green (■) score 3 or 4 after the small-scale expression screen. A) ORF29b from KSHV: Constructs 2-5 were designed around bigger gaps in the secondary structure predictions as well as in between stable and separated helical elements. Constructs that started between the only two neighbouring helical predictions yielded in stable and pure protein samples of sufficient yields for structural studies. B) K9 from KSHV: In this case all prediction methods were in very good agreement and all shown data was used in construct design as well as domain data as reported in the literature. In addition, protein structure models exist for both predicted domains. The difference between construct 4 and 5 is only four amino acid residues in the beginning. C) ICP27 from HSV-1: The full-length construct suffered from degradation both during purification and storage. As can be seen, no expression could be detected from constructs corresponding to the conserved domain. New constructs were designed to start in between the low complexity regions. Only constructs that had a translational start between low complexity region 3 and 4 yielded in stable protein samples suitable for crystallization trials. D) SDS-PAGE showing some of the different constructs after small-scale IMAC, from above the mentioned proteins.

In case of ICP27 from HSV-1, the Pfam domain prediction lies in the C-terminal region, encompassing a little more than a third of the protein (Figure 4.23 C). However, every construct designed around this predicted region failed. The full-length construct did express successfully, however, the protein suffered from heavy degradation (Figure 4.23 D, lanes 1, 4). Interestingly, the degraded product was of similar size as ICP27$^{190\text{-}512}$, the construct that finally expressed successfully and the structure of which was solved to a resolution of 2 Å, as discussed in the section 4.1. In that final ICP27$^{190\text{-}512}$ structure, it was observed that residues 190 to 240 were missing, further validating the prediction that residues 1 to 240 are disordered.

4.3.2. SUCCESS IN SCALING-UP

The herpes protein constructs that expressed decently (scoring 3 or 4 in soluble fraction) were further up-scaled and purified. The chromatographic profile from size exclusion chromatography (SEC) can indicate homogeneity and quality of the final protein. Purity is checked using SDS-PAGE. A symmetrical peak with clean PAGE profile often indicates that the protein may concentrate better. This is important for techniques like crystallization that require highly concentrated samples. Stable protein constructs with scores of 3 or 4, on scaling up, were found to behave just as any other protein, regardless of their origin. Also, the final yield and purity were found to be comparable to that of human protein constructs (Figure 4.24).

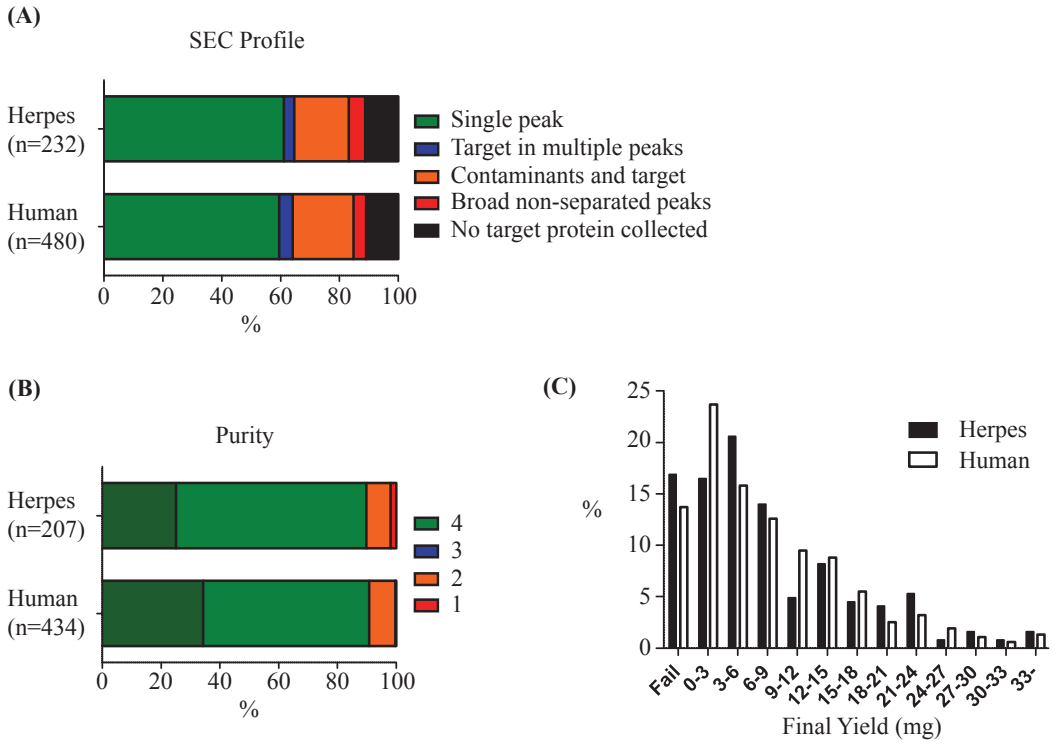


Figure 4.24 Analysis of the results from large-scale expression and purification of herpes and human targets.

(A) The chromatographic profile after size exclusion chromatography was analyzed for all targets. The profile was classified into four categories, single peak, target protein in multiple peaks, the peak contains both target protein and contaminants and broad non-separated peaks. (B) The purity score of the Herpes protein batches was compared to the purity score of the human (C) The final protein yield of the Herpes protein batches was compared to the yield of the human targets.

5. DISCUSSION

Herpes viral infections have been observed and documented since centuries and are, presently, widespread around the globe. They have evolved to circumvent and evade the immune systems of various organisms, including humans. They can thrive in the host without making their presence felt for a long time in an otherwise healthy looking individual. A great deal is known about herpes viruses in general. However, there is still relatively limited knowledge about molecular mechanisms of herpes proteins. Some of the well-known strategies these viruses use for evading the host immune responses include establishment of latency, down-regulation of the host protein synthesis (e.g. of anti-viral antibodies) and inhibition of functions of host immune response elements (e.g. cytokines, interferons). Detailed molecular studies of such strategies in detail can give an insight to the mechanisms by which herpes viruses invade hosts and provide concepts for how to inhibit them by designing specific drugs.

Herpes Simplex Viruses (HSVs) are the causative agents of the skin and genital herpes, and the most prevalent ones, worldwide, amongst the human herpesviruses (HHVs). Some genes present in the HSVs are highly conserved in all HHVs. ICP27 is one such multi-functional protein with homologues present across the family. The other functional homologues include IE4 in VZV, UL69 in HCMV, EB2 protein in EBV and ORF57 in KSHV. Although ICP27 is known to be indispensible for viral replication in the latent state, there are many grey areas in our knowledge regarding its mechanism. This thesis provides information about the distinct structural features of ICP27.

ICP27 helps in viral transcription by recruiting RNAP II to viral replication centres, is involved in virion host shut-off and affects the splicing of cellular pre-RNAs (Dai-Ju et al., 2006; Hardy & Sandri-Goldin, 1994; Kuss et al., 2013). The other homologues of ICP27 have not been reported to interact with RNAP II. ICP27 also acts as a shuttling protein by binding specifically to intronless mRNAs and exports them out of nucleus (Ingram et al., 1996; Mears & Rice, 1996). However, there are no reports on any consensus sequence identified by ICP27 or most of its homologues. $ORF57_{KSHV}$ has been found to bind a response element in the KSHV PAN mRNA (Colgan et al. 2009). In addition, ICP27 is expected to help initiate translation of certain viral

transcripts (Ellison et al., 2005; Fontaine-Rodriguez & Knipe, 2008; Fontaine-Rodriguez et al., 2004; Larralde et al., 2006). $ORF57_{KSHV}$, $EB2_{EBV}$ and $UL69_{HCMV}$ have also been reported to interact with translation initiation factors or other proteins needed to form translation complex (Aoyagi et al., 2010; Boyne et al., 2010; Ricci et al., 2009). All these functions involve interaction with either proteins or nucleic acids. It is therefore evident that ICP27 family proteins should have a structure capable of handling all these interactions and also of being regulated in a defined fashion. The past 25 yrs have seen numerous studies conducted on ICP27 and its homologues. However, the structures of none of the ICP27 family proteins have been solved hitherto, barring for the NMR structure of an 8 amino acid peptide of ICP27 N-terminal region (Tunnicliffe et al., 2011). In this thesis, I present the first crystal structure of $ICP27^{190-512}$ at 2.0 Å showing multiple distinctive features.

5.1. $ICP27^{190-512}$ HAS A NOVEL FOLD

The $ICP27^{190-512}$ structure consists of a compact α-helical core (Figure 4.5). The N- and the C-termini of $ICP27^{190-512}$ form extensions beyond this core and are involved in protein-protein interactions, forming dimers. The helical core of the protein encompasses a hydrophobic pocket with a flexible cap region (Figure 4.8). Though ICP27 has functions similar to several cellular export factors, the structure did not match any of known protein structures.

ICP27 has been predicted to have a CCHC type of zinc-finger-like motif at the C-terminal end of the protein (Vaughan et al., 1992). The residues involved were predicted to be C483, C488, H502 and C508. $ICP27^{190-512}$ structure reveals a distinct zinc-binding motif (Figure 4.11). The zinc ion is tetrahedrally coordinated via C400, H479, C483 and C488 residues. All four residues are highly conserved across the ICP27 family members including those from β- and γ-herpes viruses (Figure 4.6). Unlike most known zinc-binding motifs, this motif, represented as C400-$X_{(79)}$-H479-$X_{(2)}$-C483-$X_{(4)}$-C488, is unique in that the distance between 1st and the 2nd coordinating residues is very large. The exact function of the zinc-binding motif has not yet been reported; although it most likely acts as a structural or a protein-protein interaction motif.

Though the RNA-binding RGG box at the N-terminal is required for binding ICP27 to RNA targets, it has been speculated that the C-terminal half of the protein helps in secondary binding to the RNA (Mears & Rice, 1996). ICP27 C-terminus was predicted to have three K-homology (KH) domains, overlapping with the earlier predicted zinc-finger-like domain (Soliman & Silverstein, 2000a; Soliman & Silverstein, 2000b). The KH domains have not been reported in any of the other homologues. Surprisingly, none of the three KH domains was identified in the structure of ICP27$^{190\text{-}512}$. The typical KH domain motif and the conserved GXXG sequence were absent in ICP27$^{190\text{-}512}$ (Valverde et al., 2008). To identify the RNA-binding region within the C-terminal region, Soliman & Silverstein (2000a) firstly aligned the sequence of ICP27 and some of its homologues along with the proteins containing known KH domains (Soliman and Silverstein, 2000a). They visually analysed the alignment and proposed that KH domains are present in the ICP27 homologues. However, the alignment provided is far from convincing and the motif GXXG that is important for KH domain function was not conserved in ICP27 family proteins.

To investigate further, they mutated F303→N in the proposed KH1 domain, the effect of it being lethal for virus growth. From the structure, we can see that the well-conserved F303 is part of the flexible cap region of ICP27. The conserved, hydrophobic C-tail region of second molecule in the dimer surrounds F303, forming non-bonded interactions. The F303→N mutation leads to loss of the aromatic ring of phenylalanine. It may have led to altered interactions with the C-tail that probably is involved in dimer formation. Another mutation, L387→N in the proposed KH2 domain had no effect on ICP27 function. They also mutated R480→H in the proposed KH3 domain that resulted in loss of binding to RNA and also loss of the shuttling function. From the ICP27$^{190\text{-}512}$ structure, we can see that R480 is adjoining to the zinc-binding residue H479. It also forms hydrogen bond with the peptide backbone of another zinc-binding residue C483 (Figure 4.11). The R480→H mutation leads to clashes in the surrounding region. This may result in distortion of the zinc-binding motif.

ICP27 has been reported to interact with the C-terminal Domain (CTD) of RNA Polymerase II (RNAP II) (Zhou and Knipe, 2001; Dai-Ju et al., 2006). This

interaction is postulated to occur via both the N- and the C-terminal regions. Dai-Ju et al. (2006) performed a series of analyses to narrow down the region on ICP27 that is involved in interaction with RNAP II-CTD (Dai-Ju et al., 2006). They found that deletion of ICP27 residues from 406 to 512 and 450 to 512 resulted in loss of interaction. As seen from the structure, this region encompasses the bulk of ICP27 CTD including the zinc-binding motif, C-tail, the conserved YXPG motif and many other semi-conserved residues (Figure 4.6). Deletion of 406 to 512 also excludes the large positively charged surface that is lined with multiple conserved arginines (Figure 4.10 A). They showed that inserting PERF between the residues P465 and G466 led to drastic reduction in the interaction between ICP27 and RNAP II-CTD. Similarly, mutating PG→LE led to loss of function. These residues are part of the YXPG motif described earlier. From the structure, it is visible that this motif interacts with the C-tail via hydrophobic interactions and also with the conserved D357 via polar interactions through its protein backbone. The YXPG motif is part of a hydrophobic patch on ICP27 surface. The insertion or mutation at the PG motif introduced charged residues in an uncharged region. This may have resulted in distortion of the region around the YXPG motif, leading to loss of interaction between the RNAP II-CTD and ICP27. The residue C488 was mutated to leucine. This cysteine is part of the zinc-binding motif as seen from the structure (Figure 4.11). This mutation did not result in loss of contact with RNAP II-CTD, thus indicating that the zinc-binding motif may not be important for this interaction. They also concluded that within the CTD of ICP27, mutations or deletion of residues 504 to 512 did not affect the interaction between the two proteins. These residues form the C-tail region.

ICP27 is proposed to inhibit cellular splicing process by binding to and altering the phosphorylation pattern on the essential SR proteins, and also by interacting with SR Protein Kinase 1 (SRPK1) (Escudero-Paunetto et al., 2010, Sciabica et al., 2003; Souki & Sandri-Goldin, 2009). $ORF57_{KSHV}$ on the other hand is shown to promote splicing by binding to some SR proteins (Majerciak et al., 2008). It does so to enhance splicing of some KSHV RNAs. $EB2_{EBV}$ too has been reported to mediate alternative splicing of cellular RNA by interacting with SRp20 (Jumaa & Neilson, 1997; Verma et al., 2010).

Sciabica et al. (2003) reported that SRp20 and SRp40 might be interacting with ICP27 via its zinc-finger region within the KH3 domain (Sciabica et al., 2003). However, it should be noted that the earlier predicted zinc-binding residues (C483, C488, H502, C508) are different from those seen in the structure (C400, H479, C483, C488). Deletion of ICP27 residues form 434 to 512 resulted in loss of interaction between ICP27 and SRp20. As seen from the structure, this deleted region encompasses the novel zinc-binding motif, the C-tail, the conserved YXPG motif and several semi-conserved residues. To narrow down the interacting region, they inserted FIPR sequence between positions 504 and 505 as well as mutated Y505→S. They intended to disrupt the zinc-finger domain using this mutant as the position of C508 gets affected. They observed that the mutant did not interact with SRp20 and concluded that an intact zinc-finger of ICP27 is needed for the interaction. Similarly, in case of SRp40, they concluded that the interaction is mediated via the zinc-finger domain of ICP27. However, these mutations are actually mapped in the C-tail region of the $ICP^{190\text{-}512}$ structure and not the zinc-binding motif. The said mutations are far from the zinc-binding motif and unlikely to disrupt it. Combining the results from these mutational studies along with the now known structure of $ICP27^{190\text{-}512}$, it is more plausible that the interaction between ICP27 and SR proteins (SRp20, SRp40) is mediated directly via an exposed C-tail. Another possibility is that SRp20 interacts with ICP27 only in the dimer form, which in turn may need the intact C-tail interactions (Figure 4.9).

Verma et al. (2010) studied the $EB2_{EBV}$ interaction with SRp20. After performing pull-down assays with multiple deletion mutants, they mapped the interacting region to residues 185 to 281 on the viral protein. This region aligns with the flexible cap and the hydrophobic pocket regions of $ICP27^{190\text{-}512}$ (Figure 4.6). This eliminates the involvement of the conserved zinc-binding motif or the C-tail regions within $EB2_{EBV}$. Further mutational studies on $EB2_{EBV}$ and corresponding studies on ICP27 might help in revealing the true nature of this interaction with SR proteins.

5.2. ICP27 CAN UNDERGO DIMERIZATION VIA ITS C-TERMINAL DOMAIN

Few studies have proposed that ICP27 is capable of undergoing dimerization via its C-terminal half of the protein (Hernandez & Sandri-Goldin, 2010a, 2010b, 2011). Though the importance of dimerization has not been explored in detail, it is speculated that in its dimeric form the protein can be imported into the nucleus (Zhi et al., 1999). They also inferred that the zinc-finger-like motif (importantly residues 480 to 508) is essential for forming dimers while the N-terminal half is not. However there was no clear picture as to how the two protein molecules in the dimer interact with each other. $ORF57_{KSHV}$, $IE4_{VZV}$ and $UL69_{HCMV}$ have been reported to form homodimers via their ICP27-homology domains that corresponds to the C-terminal domain of ICP27 (Nekorchuk et al., 2007; Taylor et al., 2011; Baudoux et al., 2000; Lischka et al., 2007).

The crystal structure of $ICP27^{190-512}$ suggested that protein formed dimer with such that the monomers are positioned in a yin-yang orientation, giving the dimer a coffee bean-like shape (Figure 4.12). The intriguing features of this structure are the N- and C-termini, both of which appear to be involved in homodimerization. The hydrophobic C-tail region spanning residues 501 to 512 of one monomer is inserted into the hydrophobic pocket of the second (Figure 4.8). Some key interactions between the pocket and the C-tail are revealed in the structure (Figure 4.9). In addition to hydrophobic interactions, contacts between the buried charges can also be seen. These charged residues in both the pocket and the C-tail region are conserved. The N-terminal arm of one protomer wraps around the second protomer via hydrophobic interactions (Figure 4.5 B). Biophysical assays using multi-angle light scattering (MALS) (Figure 4.13) and analytical ultracentrifugation (AUC) (Figure 4.14) confirm that $ICP27^{190-512}$ exists as dimer in solution. The AUC result confirms the hypothesis that the C-terminal domain of ICP27 favours dimerization over other higher order oligomerizations as seen from the Table 4.4.

Taylor et al. (2011) speculated that the RGG2 motif within the C-terminal region of $ORF57_{KSHV}$ might be responsible for the homodimerization. They performed mutational studies to confirm their claim. Contrasting to them were the results of Nekorchuk et al. (2007) who demonstrated that deletion of the same RGG2 motif

resulted in no effect on dimer formation. This RGG2 motif maps very close to the RRR motif at position 416 to 418 on $ICP27^{190-512}$. This motif is present on the top surface of the protein and is not involved in dimer formation (Figure 4.10 A). In the case of $IE4_{VZV}$, it was shown that the C-terminal domain, especially the residues from 442 to 447 (GKYFKC) were extremely essential for the homodimerization (Baudoux et al., 2000). This region corresponds to the highly conserved C-tail of the ICP27 with the sequence GKYFYC. They further demonstrated that the C-terminal domain of $IE4_{VZV}$ and each of the conserved GKYFKC residues were important for gene activation and shuttling of the protein to the cytoplasmic. This indicates that the C-tail of ICP27 that is located in the hydrophobic pocket might be involved in functions other than dimerization. $UL69_{HCMV}$ is the largest amongst the ICP27 family proteins. Lischka et al. (2007) performed a series of N- and C- termini deletions of $UL69_{HCMV}$ and demonstrated that the residues from 269 to 574 are essential for dimer formation (Lischka et al., 2007). This region corresponds to the conserved ICP27 homology region. In other words, this region covers all the structural elements of ICP27 from the flexible cap in loop L2 all the way to the C-tail at the end of the ICP27 protein. A mutant with deletion from 478 to 527, corresponding to ICP27 conserved regions of YXPG motif, three of the four zinc-binding residues and the entire C-tail, resulted in loss of dimer formation. A point mutation C495→A also disrupted dimer formation. This C495 corresponds to the conserved C483 in ICP27 and one of the zinc-binding residues. The mutation might have caused distortion of the zinc-binding motif and dislocation of the downstream C-tail out of the hydrophobic pocket. Thus, all of the above studies support the idea that C-tail motif of ICP27 is extremely essential and involved in dimer formation of ICP27 and it might also have other functions.

The features exhibited by the N- and C-termini of $ICP27^{190-512}$ are similar to the domain-swapping dimer seen in the case of C-terminal domain (CTD) of SARS coronavirus (SARS-CoV) main protease M^{pro} (Zhong et al., 2009). The α1 helix of each monomer of M^{pro} exchanges positions and interacts with the other monomer. They observe that full-length M^{pro} does not form dimers and hint that the dimer formation may lead to an inactive state of the protein. Similarly, the CTD of Infectious Bronchitis Virus (IBV) nucleocapsid (N) protein also forms the domain-swapping dimer (Jayaram et al., 2006). One α helix (α4) and two β strands (β1 and β2) from one monomer extend and interact with the other monomer. They propose

that the full-length N protein forms dimers to perform its functions. They also suggest that the CTD of IBV N protein may be conserved in the N proteins of other coronaviruses as well as those of some ssRNA-viruses. Bennett et al. (1995) summarize various examples of domain swapping dimers (Bennett et al., 1995). They suggest that such dimers may help switch between the oligomeric states of proteins that may be needed for specific functions. In the case of ICP27, it is known that the protein performs many functions in the cell, while interacting with a number of proteins as well as RNA. Forming such dimers may be one of the mechanisms by which ICP27 switches between functions or interacts with its partners.

5.3. ICP27$^{190\text{-}512}$ IS PHOSPHORYLATED BY KINASES AT KEY POSITIONS

ICP27 is a multifunctional protein, interacting with a variety of proteins and nucleic acids. For its effective functioning, there must be some kind of a switch or trigger that helps ICP27 to swap between its different functions. In eukaryotic cells, post-translational modifications like ubiquitination, arginine-methylation and phosphorylation are a few common forms of such triggers. Phosphorylation and arginine-methylation of proteins can affect many functions like relocalization in the cell, protein-protein interactions or regulation of enzymatic reactions, amongst others. Effects of both these modifications on ICP27 have been studied briefly. Methylation of R138, 148 and 150 residues in the RGG box is linked to its interaction with Aly/REF and SRPK1 proteins (Mears & Rice, 1996; Souki & Sandri-Goldin, 2009). Phosphorylation sites of ICP27 are mapped at S16, 18 mediated by CK2, and at S114 within the NLS region, mediated by PKA (Zhi & Sandri-Goldin, 1999). Phosphorylation site mutants were shown to have a negative effect on the interaction of ICP27 with TAP/NXF1 and Aly/REF (Rojas, Corbin-Lickfett, Escudero-Paunetto, & Sandri-Goldin, 2010). All these sites have been mapped to the N-terminal region of ICP27 whereas not a single site has been mapped to the C-terminal domain. It is surprising that this structurally stable domain does not have any modification site. Similarly, ORF57$_{KSHV}$, EB2$_{EBV}$ and UL69$_{HCMV}$ have been reported to undergo phosphorylation by different cellular kinases at their N-terminal regions but none in the conserved C-terminal domain (Malik and Clements, 2004; Medina-Palazon et al., 2007; Winkler et al., 1994).

In order to probe the possibility of phosphorylation at the C-terminal domain, $ICP27^{190-512}$ was phosphorylated *in vitro* using the common cellular kinases PKA, CK2 and PKC. Mass spectrometry of the phosphorylated protein revealed that multiple sites on $ICP27^{190-512}$ could be targeted by kinases (Table 4.5). When these sites are mapped to the structure of $ICP27^{190-512}$, it is observed that some residues are in interesting places. S311 and T314, phosphorylated by PKA, are present in the region at the end of loop L2 and the N-terminal of the α3 helix, which forms the base of the hydrophobic pocket.

Based on these findings, we propose a model wherein phosphorylation of the Cap region acts as a regulatory mechanism for ICP27 function. As discussed earlier, it is observed that both, the C-tail residues and the hydrophobic pocket in the core are conserved within the ICP27 family. The C-tail forms an interesting structure as it inserts itself into the hydrophobic pocket of the 2^{nd} subunit. We hypothesize that this tail region is not only involved in dimerization but may also have a role in protein-protein interactions. According to our hypothesis, on phosphorylation, S311 and T314 gain a negative charge. In their proximity is E313 with its carboxyl group, facing towards S311 side chain. Phosphorylation of S311 or T314 or both leads to formation of large negative charge in that region which is repelled by the carboxyl group of E313. This leads to conformational changes in the already flexible Cap region, moving it away from the hydrophobic pocket and exposing the C-tail from the base. This exposed C-tail may then bind to its interacting partners via hydrophobic interactions.

As mentioned earlier, none of the other homologues has been reported to harbour a phosphorylation site within its ICP27 homology domain. From the multiple sequence alignment (Figure 4.6), it is visible that S311 as well as T314 of ICP27 are conserved in most homologues. Phosphorylation studies specifically to the C-terminal domains of the homologues might provide more information on the post-translational modification status of the proteins.

5.4. ICP27$^{190-512}$ INTERACTION WITH NUCLEIC ACIDS

It is known that ICP27 family members are essential for the efficient export of intronless viral mRNAs. ICP27 is specifically responsible for the regulation of mRNAs of the immediate early and the late genes (Ingram et al., 1996; Rice & Knipe, 1990). The capacity of the protein to bind to target mRNAs has been mapped to the N-terminal region, specifically the RGG box (Mears & Rice, 1996). However, some studies have predicted involvement of the C-terminal region of ICP27 in RNA binding (Soliman & Silverstein, 2000a; Soliman & Silverstein, 2000b). Three KH domains were proposed to carry out the RNA binding function within the C-terminal half. As discussed earlier, the crystal structure of ICP27$^{190-512}$ does not support the presence of any of these predicted KH domains. ORF57$_{KSHV}$, IE4$_{VZV}$ and UL69$_{HCMV}$ contain their primary RNA binding domains within the N-terminal region of the protein, outside of the ICP27 homology domain (Majerciak and Zheng, 2009; Toth et al., 2006; Sandri-Goldin, 2011). ORF57$_{KSHV}$ contains a second non-canonical RNA binding motif called the RGG2 box in its C-terminal domain, which is dispensible for function. The RNA-binding domain of EB2$_{EBV}$ overlaps with the C-terminal domain of ICP27 (Hiriart et al., 2003; Sergeant et al., 2008). However, no definite information is available about the nature of RNA-protein interaction.

To investigate whether ICP27$^{190-512}$ is capable of interacting with RNA despite the absence of the KH domains, thermal stability shift assays were performed using short RNA sequences. There is little information about the consensus RNA sequence that may be recognized by ICP27. So far, there has been only one study proposing that ICP27 binds to GC-rich regions of the RNA that lacks any secondary structures or folds (Corbin-Lickfett et al., 2009). It was hypothesized that if ICP27$^{190-512}$ binds to RNA, then it will be stabilized and the melting temperature (Tm) of the protein will increase. However, to our surprise, addition of RNA to the protein led to decrease in the Tm, indicating that the protein was rather destabilized (Figure 4.18 A). We attributed this effect to the chelation of the zinc ion from the zinc-binding motif by the negatively charged RNA. Providing zinc externally resulted in a huge positive shift in Tm indicating probable stabilization. From the structure, the possible regions that might interact with RNA are either the central groove formed along the dimer (Figure 4.12) or the large positively charged surfaces seen on the top of the dimer

(Figure 4.10 A). The positive surface provides a good area for non-specific binding to the RNA backbone.

5.5. EB2$^{189-436}$ HAS A SECONDARY STRUCTURE PROFILE SIMILAR TO ICP27$^{190-512}$

EB2 belongs to the ICP27 family of proteins. Although not all functions overlap, the two proteins share some similar functions. Like ICP27, EB2 helps in efficient translocation of intronless viral mRNAs from the host nucleus to the cytoplasm (Semmes et al., 1998; Sergeant et al., 2008). EB2 has two nuclear localization signals (NLS) and probably two nuclear export signals (NES), all of which are important for shuttling of the protein in and out of the host nucleus (Boyle et al., 1999; Chen L. et al., 2001; Hiriart et al., 2003a; Hiriart et al., 2003b). Both NLS and NES regions lie within the N-terminal half of EB2. It has a RNA-binding domain (RBD), also called Arginine rich motif (ARM), positioned at the centre of the protein length (Hiriart et al., 2003; Sergeant et al., 2008). A second RNA-binding region consisting of Arginine-X-Proline (RXP) repeats is present upstream of the RBD (Buisson et al.,1999; Ruvolo et al., 2001; Semmes et al., 1998). However, no information is available regarding the recognition mechanism of target mRNA by EB2. EB2 is responsible for translocation of intronless as well as unspliced RNAs. Similar to ICP27, EB2 has been found to interact with various export factors like Aly/REF and OTT3/RBM15b (Hiriart et al., 2003b; Hiriart et al., 2005). Similar to ICP27, the binding of the mRNA takes place via an RNA-binding domain and the export requires an intact C-terminal domain. The protein is phosphorylated by enzyme CKII mostly in the N-terminal region but it interacts with the enzyme even via C-terminal domain (Medina-Palazon et al., 2007).

EB2 probably has a structure that supports its interactions with the different proteins and RNAs. By secondary structure predictions we know that like ICP27, the N-terminal half of EB2 lacks the secondary structural elements, whereas the C-terminal half consists of α-helices. Also, from the sequence alignment of EB2 protein with ICP27, we see that many regions that have been defined in the ICP27 structure are conserved in EB2 (Figure 4.6). Hence, despite the low sequence similarity between the two proteins, there is a probability of them having similar structure. Due to constraints on protein concentration of EB2, crystallization trials didn't yield any

crystals. However, to determine secondary structural organization of EB2, a circular dichroism experiment with ICP27$^{190\text{-}512}$ as reference was performed. The CD spectrum is similar to that of ICP27$^{190\text{-}512}$ (Figure 4.21). This further emphasizes that there may be structural similarity between the two proteins.

5.6. SYSTEMATIC PRODUCTION OF HERPES VIRAL PROTEINS USING A MULTI-CONSTRUCT APPROACH

Enormous amounts of data are available on different aspects of herpes viruses such as life cycle, virion structure, pathology, epidemiology, etc. Multiple studies have defined various protein-protein interactions between viral and host proteins as well as their cellular localizations in the host cell. However, many herpes protein still remain structurally uncharacterized. In fact, the mechanisms by which many key viral proteins function are largely unknown. One of the major hurdles in structural studies of proteins is the challenge involved in expression of recombinant viral proteins. It has been proposed that viral proteins possess different biophysical properties than eukaryotic and bacterial proteins (Tokuriki et al., 2009; Xue et al., 2012). They may be difficult to express due to strict codon usage as well as high susceptibility to degradation due to their loose, flexible nature (Bahir et al., 2009; Xue et al., 2014).

Based on the multi-construct strategy, which was successfully implemented for expressing human proteins (Gräslund et al., 2008), we tried to achieve expression and purification of herpes viral proteins for uses in structural and biophysical studies. 94 herpes proteins and 155 human proteins were subjected to the multi-construct approach and analyzed critically to give a comprehensive view on the success of this approach. From our findings, we observe that longest (full-length) constructs of herpes proteins are much more difficult to express than those of human proteins, with only 18 % success rate as compared to 31 % in the latter (Figure 4.22 B). Even after designing N- and C- terminal truncations, the success-rate did increase to 45 % for herpes proteins, but it was much higher for human proteins (60 %). This indicates that more number of iterations in construct-design is needed for herpes proteins.

Gräslund et al. (2008) as well as Savitsky et al. (2010) made use of multiple bioinformatic tools like Pfam domains, secondary structure predictions, low complexity region predictions as well as known PDB structures of homologous proteins (Gräslund et al., 2008; Savitsky et al., 2010). However, it was challenging to explore such a strategy for herpes proteins, due to lack of homologous structures, poor domain predictions and presence of numerous low complexity regions. Domain predictions were often not useable since the protein would be predicted to have a single large domain. Hence constructs were designed based on secondary structure predictions along with sequence alignments with orthologs of the target protein, and using available literature to give indication of translational start and stop sites. For most proteins, multiple rounds of iterations had to be made to obtain a well-expressing construct (Figure 4.23). A comparison of the well-expressing herpes protein constructs to those from human proteins showed that once a stable construct was obtained, the downstream processing remained unaffected regardless of origin of the protein (Figure 4.24).

6. CONCLUSIONS

ICP27 has a novel fold compared to the currently known protein structures. The ICP27 family proteins show significant sequence homology in certain regions of the protein, namely, the zinc-binding motif, the hydrophobic pocket and the C-tail. It is plausible that the protein structure is conserved across the ICP27 family. Taylor et al. (2011) used a 3D-structure prediction algorithm, I-TASSER, and reported a predicted model for the ICP27 homolog, ORF57 from Kaposi's Sarcoma-associated Herpesvirus (KSHV) (Taylor et al., 2011). This model differs significantly from the ICP27 crystal structure. Similarly, Soliman and Silverstein (2000) predicted presence of three KH domains in the C-terminal domain of ICP27 (Soliman & Silverstein, 2000). Though computational predictions give some insight to protein structures, they should be treated with precautions.

The extended arms of ICP27 C-terminal domain are intriguing structures. They help in formation of the dimer via multiple interactions. At the same time, the C-tail lodged in a hydrophobic pocket point to a function more than just dimerization. The high conservation of both these regions in all the ICP27 homologues indicates towards a conserved function. The cap region has a high B-factor value and is flexible to move and maybe release the C-tail housed in the pocket. Presence of phosphorylation sites on the cap region of the hydrophobic pocket could be interesting to study from regulation point of view.

ICP27 is one of the key regulatory proteins of HSV and is central to survival of the virus. Both N- and C-terminal domains of ICP27 are predicted to bind to several host proteins. The zinc-binding motif in the C-terminal region is conserved throughout the ICP27 family and the structure reveals it to be distinct from the known zinc-finger-like motifs. However, there is lack of studies in the C-terminal domain of ICP27. The zinc-binding motif has hardly been characterized and no information is available regarding its function. We predict that this motif is important to ICP27 and its homologues to maintain their structures. It may also be involved in protein-protein interactions as it is positioned on the surface. However, unlike some zinc-finger motifs, it probably is not involved in binding to the nucleic acids as can be seen from the surface charges around the motif.

A comprehensive understanding of the mechanism of these immediate-early herpes proteins is still a few steps away. However, this thesis provides a good starting point for studies targeted at the same. Crystallization of the full-length ICP27 and EB2 may be a hurdle since they undergo rapid degradation and also because crystallization of disordered regions is difficult if not implausible. However, *in vivo* and *in vitro* studies designed based on structure of $ICP27^{190-512}$ might provide more insight to the mechanism of the protein. Future studies could be directed towards understanding the role of the novel zinc-binding motif in protein-protein interactions. The importance of dimer configuration as well as the C-tail of ICP27 should be probed further as they may be involved in recruiting interacting proteins. Phosphorylation of ICP27 C-terminal domain could be probed further using *in vivo* studies to elucidate the exact nature of the post-translational modification in regulating its function. A deep understanding of the key regulator protein of HSV-1 and EBV could pave the way to a better understanding of the mechanisms of its homologues in other human herpes viruses. ORF57 is one of the other homologues that has been studied and can add to the understanding of the ICP27 family of proteins.

A rational multi-construct approach to obtain well-expressing herpes viral proteins will help drive the viral structural biology ahead. The challenge of expressing difficult proteins may be circumvented by investing time and resources in construct-design. The multi-construct approach is more challenging, when there are often no homologous proteins with structural information, and rationally designed domain definitions remain poorly determined. In such cases, more innovative approaches such as domain hunting in construct libraries can be used (Cornvik et al., 2005).

Overall, a detailed knowledge of the molecular mechanisms of herpes viral proteins will help us understand the survival strategies of these viruses. Structural information of these proteins is necessary and more efforts should be directed for the same. Not only will it provide information of the mechanisms, but also give a more detailed view about the evolutionary processes that the herpes viruses underwent. Epidemiologically, herpes viral infections are prevalent worldwide. They pose a greater threat in immuno-compromised patients (e.g. HIV or cancer patients). Though a few drugs are available against some herpes infections, drug-resistance is on

increase in the immuno-compromised patients. Rational drug development projects can be undertaken with the knowledge of herpes protein structures.

REFERENCES

Books

Davison, A.J. (2007). Overview of Classification. In A. Arvin, G. Campadelli-Fiume, E. Mocarski, P.S. Moore, B. Roizman, R.J. Whitley, & K. Yamanishi (editors), *Human Herpesviruses: Biology, Therapy and Immunoprophalaxis* (pp. 3-9). New York, NY: Cambridge University Press.

Smiley, J., Elgadi, M.M., & Saffran, H.A. (2001). Herpes Simplex Virus vhs Protein. In A. Nicholson (editor), *Methods of Enzymology - Ribonucleases, Part B: Artificial and Engineered Ribonucleases and Specific Applications* (pp 440-451). Academic Press.

Whitley, R.J. (1996). Herpes simplex virus. In B.N. Fields., D.M. Knipe, & P.M. Howley (editors), *Fields Virology, 3rd Edition* (pp 2297-2342). Philadelphia, PA: Lippincott-Raven.

Otwinowski Z. & Minor W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In C.W. Carter Jr., & R.M. Sweet (editors), In *Methods in Enzymology - Macromolecular Crystallography, Part A, 276* (pp 307-326). New York, NY: Academic Press.

Journal Articles

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C., Zwart P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D, Biological Crystallography,* **66,** 213-221.

Afonine P.V., Grosse-Kunstleve R.W., Echols N., Headd J.J., Moriarty N.W., Mustyakimov M., Terwilliger T.C., Urzhumtsev A., Zwart P.H., and Adams P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D, Biological Crystallography, 68,* 352-367.

Aoyagi M., Gaspar M., & Shenk T. (2010). Human cytomegalovirus UL69 protein facilitates translation by associating with the mRNA cap-binding complex and excluding 4EBP1. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 2640–2645.

Ariyoshi, M., & Schwabe, J.W.R. (2003). A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes and Development*. *17*, 1909-1920.

Bahir, I., Fromer, M., Prat, Y., & Linial, M. (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Molecular Systems Biology*, *5*(311), 1-14.

Baker N.A., Sept D., Joseph S., Holst M.J., McCammon J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 10037-10041.

Baudoux L., Defechereux P., Rentier B., & Piette J. (2000). Gene activation by Varicella-zoster virus IE4 protein requires its dimerization and involves both the arginine-rich sequence, the central part, and the carboxyl-terminal cysteine-rich region. *Journal of Biological Chemistry, 275* (42), 32822–32831.

Blaho, J.A., Mitchell, C., & Roizman, B. (1993). Guanylylation and adenylylation of the alpha regulatory proteins of herpes simplex virus require a viral beta or gamma function. *Journal of Virology*, *67*(7), 3891-3900.

Bond C.S. (2003). TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics, 19*(2), 311-312.

Bowman, B.R., Welschhans, R.L., Jayaram, H., Stow, N.D., Preston, V.G., & Quiocho, F.A. (2006). Structural Characterization of the UL25 DNA-Packaging Protein from Herpes Simplex Virus Type 1. *Journal of Virology*, *80*(5), 2309–2317.

Boyne J.R., Jackson B.R., Taylor A., Macnab S.A., & Whitehouse A. (2010). Kaposi's sarcoma-associated herpesvirus ORF57 protein interacts with PYM to enhance translation of viral intronless mRNAs. *EMBO Journal, 29* (11), 1851-1864.

Bryant, H.E., Wadd, S.E., Lamond, A.I., & Silverstein, S.J. (2001). Herpes Simplex Virus IE63 (ICP27) Protein Interacts with Spliceosome-Associated Protein 145 and Inhibits Splicing prior to the First Catalytic Step. *Journal of Virology*, *75*(9), 4376–4385.

Buchan, D.W.A., Minneci F., Nugent T.C.O., Bryson K., & Jones D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research, 41* (W1), W340-W348.

Buisson, M., Hans, F., Kusters, I., Duran, N., & Sergeant, A. (1999). The C-terminal region but not the Arg-X-Pro repeat of Epstein-Barr virus protein EB2 is required for its effect on RNA splicing and transport. *Journal of Virology*, *73*(5), 4090–4100.

Carmody, S.R., & Wente, S.R. (2009). mRNA nuclear export at a glance. *Journal of Cell Science*, *122*(Pt 12), 1933–1937.

Chen, I.B., Li, L., Silva, L., & Sandri-Goldin, R.M. (2005). ICP27 Recruits Aly / REF but Not TAP / NXF1 to Herpes Simplex Virus Type 1 Transcription Sites although TAP / NXF1 Is Required for ICP27 Export. *Journal of Virology*, *79*(7), 3949–3961.

Chen, I. B., Sciabica, K. S., & Sandri-Goldin, R. M. (2002). ICP27 Interacts with the RNA Export Factor Aly / REF To Direct Herpes Simplex Virus Type 1 Intronless mRNAs to the TAP Export Pathway. *Journal of Virology*, *76*(24), 12877–12889.

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J.L., Dias, A., & Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, *127*(7), 1389–1400.

Chevallier-Greco, A., Manet, E., Chavrier, P., Mosnier, C., Daillie, J., & Sergeant, A. (1986). Both Epstein -Barr virus (EBV)-encoded trans-acting factors, EB1 and EB2,. *EMBO Journal*, *5*(12), 3243–3249.

Chouard, T. (2011). Breaking the Protein Rules. *Nature*, *471*, 151–153.

Coen, N., Duraffour, S., Snoeck, R., & Andrei, G. (2014). KSHV targeted therapy: an update on inhibitors of viral lytic replication. *Viruses*, *6*(11), 4731–4759.

Cohen, J.I., Mocarski, E.S., Raab-traub, N., Corey, L., & Nabel, G.J. (2013). The need and challenges for development of an Epstein-Barr virus vaccine. *Vaccine*, *31*, B194–B196.

Corbin-Lickfett, K.A., Chen, I.H.B., Cocco, M.J., & Sandri-Goldin, R.M. (2009). The HSV-1 ICP27 RGG box specifically binds flexible, GC-rich sequences but not G-quartet structures. *Nucleic Acids Research*, *37*(21), 7290–7301.

Corbin-Lickfett, K.A., Rojas, S., Li, L., Cocco, M.J., & Sandri-Goldin, R.M. (2010). ICP27 phosphorylation site mutants display altered functional interactions with cellular export factors Aly/REF and TAP/NXF1 but are able to bind herpes simplex virus 1 RNA. *Journal of Virology*, *84*(5), 2212–2222.

Corbin-Lickfett, K.A., Souki, S.K., Cocco, M.J., & Sandri-Goldin, R.M. (2010). Three arginine residues within the RGG box are crucial for ICP27 binding to herpes simplex virus 1 GC-rich sequences and for efficient viral RNA export. *Journal of Virology*, *84*(13), 6367–6376.

Cornvik, T., Dahlroth, S., Magnusdottir, A., Herman, M.D., Knaust, R., Ekberg, M., & Nordlund, P. (2005). Colony filtration blot : a new screening method for soluble protein expression in Escherichia coli. *Nature Methods*, *2*(7), 507–509.

Cui, X., Cao, Z., Sen, G., Chattopadhyay, G., Fuller, D.H., Fuller, J.T., Snapper, D. M., Snow, A.L., Mond, J. J., & Snapper, C.M. (2013). A novel tetrameric gp350 1-470 as a potential Epstein-Barr virus vaccine. *Vaccine*, *31*(30), 3039–3045.

Dai-ju, J.Q., Li, L., Johnson, L.A., & Sandri-Goldin, R.M. (2006). ICP27 Interacts with the C-Terminal Domain of RNA Polymerase II and Facilitates Its Recruitment to Herpes Simplex Virus 1 Transcription Sites, Where It Undergoes Proteasomal Degradation during Infection. *Journal of Virology*, *80*(7), 3567-3581.

Davison, A.J. (2002). Evolution of the herpesviruses. *Veterinary Microbiology*, *86*(1-2), 69–88.

Delaney, A.S., Thomas, W., & Balfour, H.H. Jr. (2014). Coprevalence of Epstein-Barr Virus, Cytomegalovirus, and Herpes Simplex Virus Type-1 Antibodies Among United States Children and Factors Associated With Their Acquisition. *Journal of the Pediatric Infectious Diseases Society*. doi:10.1093/jpids/piu076

Deluca, N.A., Mccarthy, A.M., & Schaffer, P.A. (1985). Isolation and Characterization of Deletion Mutants of Herpes Simplex Virus Type 1 in the Gene Encoding Immediate-Early Regulatory Protein ICP4. *Journal of Virology*, *56*(2), 558–570.

Dolinsky T.J., Nielsen J.E., McCammon J.A., & Baker N.A. (2004). PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research, 32,* W665-W667.

Dolinsky T.J., Czodrowski P., Li H., Nielsen J.E., Jensen J.H., Klebe G., & Baker N.A. (2007). PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research, 35,* W522-W525.

Dufu, K., Livingstone, M.J., Seebacher, J., Gygi, S.P., Wilson, S.A., & Reed, R. (2010). ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes & Development*, *24*(18), 2043–2053.

Ellison, K.S., Maranchuk, R.A., Mottet, K.L., & Smiley, J.R. (2005). Control of VP16 Translation by the Herpes Simplex Virus Type 1 Immediate-Early Protein ICP27 Control of VP16 Translation by the Herpes Simplex Virus Type 1 Immediate-Early Protein ICP27. *Journal of Virology*, *79*(7), 4120–4131.

Emsley P., Lohkamp B., Scott W.G., & Cowtan K. (2010). Features and Development of Coot. *Acta Crystallographica. Section D, Biological Crystallography, 66*, 486-501.

Escudero-Paunetto, L., Li, L., Hernandez, F.P., & Sandri-Goldin, R.M. (2010). SR proteins SRp20 and 9G8 contribute to efficient export of herpes simplex virus 1 mRNAs. *Virology*, *401*(2), 155–164.

Fogg, M.J., Alzari, P., Bahar, M., Bertini, I., Betton, J.M., Burmeister, W.P., Cambillau, C., Canard, B., Corrondo, M.A., Carrondo, M., Coll, M., Daenke, S., Dym, O., Egloff, M.P., Enguita, F.J., Geerlof, A., Haouz, A., Jones, T.A., Ma, Q., Manicka, S.N., Migliardi, M., Nordlund, P., Owens, R.J., Peleg, Y., Schneider, G., Schnell, R., Stuart, D.I., Tarbouriech, N., Unge, T., Wilkinson, A.J., Wilmanns, M., Wilson, K.S., Zimhony, O., & Grimes, J.M. (2006). Application of the use of high-throughput technologies to the determination of protein structures of bacterial and viral pathogens. *Acta Crystallographica. Section D, Biological Crystallography*, *62*(Pt 10), 1196–1207.

Fontaine-Rodriguez, E.C., & Knipe, D.M. (2008). Herpes simplex virus ICP27 increases translation of a subset of viral late mRNAs. *Journal of Virology*, *82*(7), 3538–3545.

Fontaine-Rodriguez, E.C., Taylor, T.J., Olesky, M., & Knipe, D.M. (2004). Proteomics of herpes simplex virus infected cell protein 27: association with translation initiation factors. *Virology*, *330*(2), 487–492.

Gräslund, S., Sagemark, J., Berglund, H., Dahlgren, L.G., Flores, A., Hammarström, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P., & Weigelt, J. (2008). The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expression and Purification*, *58*(2), 210–221.

Griffiths, P., Plotkin, S., Mocarski, E., Pass, R., Schleiss, M., Krause, P., & Bialek, S. (2013). Desirability and feasibility of a vaccine against cytomegalovirus. *Vaccine*, *31*, B197–B203.

Grishin, N.V. (2001). KH domain : one motif , two folds. *Nucleic Acids Research*, *29*(3), 638–643.

Han, Z., Marendy, E., Wang, Y., Yuan, J., Sample, J.T., & Swaminathan, S. (2007). Multiple Roles of Epstein-Barr Virus SM Protein in Lytic Replication. *Journal of Virology*, *81*(8), 4058–4069.

Han, Z., Verma, D., Hilscher, C., Dittmer, D.P., & Swaminathan, S. (2009). General and Target-Specific RNA Binding Properties of Epstein-Barr Virus SM Posttranscriptional Regulatory Protein. *Journal of Virology*, *83*(22), 11635–11644.

Hardy, W.R., & Sandri-Goldin, R.M. (1994). Herpes Simplex Virus Inhibits Host Cell Splicing, and Regulatory Protein ICP27 Is Required for This Effect. *Journal of Virology*, *68*(12), 7790–7799.

Henson, B.W., Johnson, N., Bera, A., Okoye, M.E., Desai, K.V., & Desai, P.J. (2011). Expression of the HSV-1 capsid protein VP19C in E. coli: A single amino acid change overcomes an expression block of the full-length polypeptide. *Protein Expression and Purification*, *77*(1), 80–85.

Hernandez, F.P., & Sandri-Goldin, R.M. (2010a). Head-to-Tail Intramolecular Interaction of Herpes Simplex Virus Type 1 Regulatory Protein ICP27 Is Important for Its Interaction with Cellular mRNA Export Receptor TAP / NXF1. *mBio*, *1*(5).

Hernandez, F.P., & Sandri-Goldin, R.M. (2010b). Herpes simplex virus 1 regulatory protein ICP27 undergoes a head-to-tail intramolecular interaction. *Journal of Virology*, *84*(9), 4124–35.

Hernandez, F.P., & Sandri-Goldin, R.M. (2011). Bimolecular Fluorescence Complementation analysis to reveal protein interactions in herpes virus infected cells. *Methods*, *55*(2), 182–187.

Hiriart, E., Bardouillet, L., Manet, E., Gruffat, H., Penin, F., Montserret, R., Farjot, G., & Sergeant, A. (2003). A region of the Epstein-Barr virus (EBV) mRNA export factor EB2 containing an arginine-rich motif mediates direct binding to RNA. *The Journal of Biological Chemistry*, *278*(39), 37790–37798.

Hiriart, E., Gruffat, H., Buisson, M., Mikaelian, I., Keppler, S., Meresse, P., Mercher, T., Bernard, O.A., Sergeant, A., & Manet, E. (2005). Interaction of the Epstein-Barr virus mRNA export factor EB2 with human Spen proteins SHARP, OTT1, and a novel member of the family, OTT3, links Spen proteins with splicing regulation and mRNA export. *The Journal of Biological Chemistry*, *280*(44), 36935–36945.

Huang, Y., & Steitz, J.A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Molecular Cell*, *7*(4), 899–905.

Huang, Y., & Steitz, J.A. (2005). SRprises along a messenger's journey. *Molecular Cell*, *17*(5), 613–615.

Ingram, A., Phelan, A., Dunlop, J., & Clements, J.B. (1996). Immediate early protein IE63 of herpes simplex virus type 1 binds RNA directly. *Journal of General Virology*, *77*, 1847–1851.

Jean, S., LeVan, K.M., Song, B., Levine, M., & Knipe, D.M. (2001). Herpes simplex virus 1 ICP27 is required for transcription of two viral late (gamma 2) genes in infected cells. *Virology*, *283*(2), 273–284.

Johnson, L.A., & Sandri-Goldin, R.M. (2009). Efficient nuclear export of herpes simplex virus 1 transcripts requires both RNA binding by ICP27 and ICP27 interaction with TAP/NXF1. *Journal of Virology*, *83*(3), 1184–1192.

Jones DT. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology, 292*, 195-202.

Juillard, F., Bazot, Q., Mure, F., Tafforeau, L., Macri, C., Rabourdin-Combe, C., Lotteau, V., Manet, E., & Gruffat, H. (2012). Epstein-Barr virus protein EB2 stimulates cytoplasmic mRNA accumulation by counteracting the deleterious effects of SRp20 on viral mRNAs. *Nucleic Acids Research*, *40*(14), 6834–6849.

Jumaa H. & Nielsen P.J. (1997). The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO Journal, 16* (16), 5077-5085.

Kalamvoki, M., & Roizman, B. (2011). The histone acetyltransferase CLOCK is an essential component of the herpes simplex virus 1 transcriptome that includes TFIID, ICP4, ICP27, and ICP22. *Journal of Virology*, *85*(18), 9472–9427.

Karasneh, G.A., & Shukla, D. (2011). Herpes simplex virus infects most cell types in vitro: clues to its success. *Virology Journal*, *8*(1), 481-491.

Kieff, E., Hoyer, B., Bachenheimer, S., & Roizman, B. (1972). Genetic Relatedness of Type 1 and Type 2 Herpes Simplex Viruses. *Journal of Virology*, *9*(5), 738–745.

Koffa, M.D., Clements, J.B., Izaurralde, E., Wadd, S., Wilson, S.A, Mattaj, I.W., & Kuersten, S. (2001). Herpes simplex virus ICP27 protein provides viral mRNAs with access to the cellular mRNA export pathway. *The EMBO Journal*, *20*(20), 5769–5778.

Köhler, A., & Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews. Molecular Cell Biology*, *8*(10), 761–773.

Kuss, S.K., Mata, M.A., Zhang, L., & Fontoura, B.M.A. (2013). Nuclear imprisonment: viral strategies to arrest host mRNA nuclear export. *Viruses*, *5*(7), 1824–1849.

Larralde, O., Smith, R.W.P., Wilkie, G.S., Malik, P., Gray, N.K., & Clements, J.B. (2006). Direct Stimulation of Translation by the Multifunctional Herpesvirus ICP27 Protein †. *Journal of Virology*, *80*(3), 1588–1591.

Le Sage, V., Jung, M., Alter, J.D., Wills, E.G., Johnston, S.M., Kawaguchi, Y., Baines, J.D., & Banfield, B.W. (2013). The herpes simplex virus 2 UL21 protein is essential for virus propagation. *Journal of Virology*, *87*(10), 5904–5015.

Lindberg, A., & Kreivi, J.P. (2002). Splicing inhibition at the level of spliceosome assembly in the presence of herpes simplex virus protein ICP27. *Virology*, *294*(1), 189–198.

Lindtner, S., Zolotukhin, A.S., Uranishi, H., Bear, J., Kulkarni, V., Smulevitch, S., Samiotaki, M., Panayotou, G., Felber, B.K., & Pavlakis, G.N. (2006). RNA-binding motif protein 15 binds to the RNA transport element RTE and provides a direct link to the NXF1 export pathway. *The Journal of Biological Chemistry*, *281*(48), 36915–36928.

Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., & Reed, R. (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature*, *413*(6856), 644–647.

Luo, M., & Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(26), 14937–14942.

Majerciak V., Yamanegi K., Allemand E., Kruhlak M., Krainer A.R., & Zheng Z.M. (2008). Kaposi's sarcoma-associated herpesvirus ORF57 functions as a viral splicing factor and promotes expression of intron-containing viral lytic genes in spliceosome-mediated RNA splicing. *Journal of Virology*, *82* (6), 2792-2801.

Majerciak V., & Zheng Z.M. (2009). Kaposi's sarcoma associated herpesvirus ORF57 in viral RNA processing. *Frontiers in Biosciences, 14*, 1516-1528.

Makarova, O.V., Makarov, E.M., & Lu, R. (2001). The 65 and 110 kDa SR-related proteins of the U4 / U6 ´ U5 tri-snRNP are essential for the assembly of mature spliceosomes. *The EMBO Journal*, *20*(10), 2553–2563.

Malik P. & Clements J.B. (2004). Protein kinase CK2 phosphorylation regulates the interaction of Kaposi's sarcoma-associated herpesvirus regulatory protein ORF57 with its multifunctional partner hnRNP K. *Nucleic Acids Research, 32,* 5553–5569.

Malik, P., Tabarraei, A., Kehlenbach, R.H., Korfali, N., Iwasawa, R., Graham, S.V., & Schirmer, E.C. (2012). Herpes simplex virus ICP27 protein directly interacts with the nuclear pore complex through Nup62, inhibiting host nucleocytoplasmic transport pathways. *The Journal of Biological Chemistry*, *287*(15), 12277–12292.

Massimelli, M.J., Kang, J.G., Majerciak, V., Le, S.Y., Liewehr, D.J., Steinberg, S. M., & Zheng, Z.M. (2011). Stability of a long noncoding viral RNA depends on a 9-nt core element at the RNA 5' end to interact with viral ORF57 and cellular PABPC1. *International Journal of Biological Sciences*, *7*(8), 1145–1160.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., & Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. *Genes & Development*, *19*(13), 1512–1517.

McCoy A.J., Storoni L.C., & Read R.J. (2004). Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallographica Section D, Biological Crystallography, 60,* 1220-1228.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., & Read R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography, 40* (Pt 4), 658-674.

McGeoch D.J., Cook S., Dolan A., Jamieson F.E., Telford E.A. (1995). Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *Journal of Molecular Biology, 247* (3), 443-458.

Mears, W.E., & Rice, S.A. (1996). The RGG box motif of the herpes simplex virus ICP27 protein mediates an RNA-binding activity and determines in vivo methylation. *Journal of Virology*, *70*(11), 7445–7453.

Medina-Palazon, C., Gruffat, H., Mure, F., Filhol, O., Vingtdeux-Didier, V., Drobecq, H., Cochet, C., Sergeant, N., Sergeant, A., & Manet, E. (2007). Protein kinase CK2 phosphorylation of EB2 regulates its function in the production of Epstein-Barr virus infectious viral particles. *Journal of Virology*, *81*(21), 11850–11860.

Misteli, T., Cáceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., & Spector, D.L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *The Journal of Cell Biology*, *143*(2), 297–307.

Norseen, J., Johnson, F.B., & Lieberman, P.M. (2009). Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *Journal of Virology*, *83*(20), 10336–10346.

Olesky, M., McNamee, E.E., Zhou, C., Taylor, T.J., & Knipe, D.M. (2005). Evidence for a direct interaction between HSV-1 ICP27 and ICP8 proteins. *Virology*, *331*(1), 94-105.

Poppers, J., Mulvey, M., Perez, C., Khoo, D., & Mohr, I. (2003). Identification of a Lytic-Cycle Epstein-Barr Virus Gene Product That Can Regulate PKR Activation. *Journal of Virology*, *77*(1), 228-236.

Ramos, A., Hollingworth, D., & Pastore, A. (2003). G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA*, *9*, 1198–1207.

Ricci E.P., Mure F., Gruffat H., Decimo D., Humaine D.V., & Supe E.N. (2009). Translation of intronless RNAs is strongly stimulated by the Epstein–Barr virus mRNA export factor EB2. *Nucleic Acids Research, 37*, 4932-4943.

Rice, S.A, & Knipe, D.M. (1990). Genetic evidence for two distinct transactivation functions of the herpes simplex virus alpha protein ICP27. *Journal of Virology*, *64*(4), 1704-1715.

Rojas, S., Corbin-Lickfett, K.A., Escudero-Paunetto, L., & Sandri-Goldin, R.M. (2010). ICP27 phosphorylation site mutants are defective in herpes simplex virus 1 replication and gene expression. *Journal of Virology*, *84*(5), 2200–2211.

Rondón, A.G., Jimeno, S., & Aguilera, A. (2010). The interface between transcription and mRNP export: from THO to THSC/TREX-2. *Biochimica et Biophysica Acta*, *1799*(8), 533–538.

Ruvolo, V., Gupta, A.K., & Swaminathan, S. (2001). Epstein-Barr Virus SM Protein Interacts with mRNA In Vivo and Mediates a Gene-Specific Increase in Cytoplasmic mRNA. *Journal of Virology*, *75*(13), 6033–6041.

Salaun, C., Macdonald, A.I., Larralde, O., Howard, L., Lochtie, K., Burgess, H.M., Brook, M., Malik, P., Gray, N.K., & Graham, S.V. (2010). Poly (A) -Binding Protein 1 Partially Relocalizes to the Nucleus during Herpes Simplex Virus Type 1 Infection in an ICP27-Independent Manner and Does Not Inhibit Virus Replication. *Journal of Virology*, *84*(17), 8539–8548.

Sam, M.D., Evans, B.T., Coen, D.M., & Hogle, J.M. (2009). Biochemical, biophysical, and mutational analyses of subunit interactions of the human cytomegalovirus nuclear egress complex. *Journal of Virology*, *83*(7), 2996–3006.

Sandri-Goldin, R.M. (1998). ICP27 mediates HSV RNA export by shuttling through a leucine-rich nuclear export signal and binding viral intronless RNAs through an RGG motif. *Genes & Development*, *12*(6), 868–879.

Sandri-Goldin, R.M. (2011). The Many Roles of the Highly Interactive HSV protein ICP27, a Key Regulator of Infection. *Future Microbiology*, *6*(11), 1261–1277.

Savitsky, P., Bray, J., Cooper, C.D.O., Marsden, B.D., Mahajan, P., Burgess-Brown, N.A., & Gileadi, O. (2010). High-throughput production of human proteins for crystallization: the SGC experience. *Journal of Structural Biology*, *172*(1), 3–13.

Schmader, K.E., Levin, M.J., Gnann J.W. Jr., Mcneil, S.A., Vesikari, T., Betts, R.F., Keay, S., Stek, J.E., Bundick, N.D., Su, S.C., Zhao, Y., Li, X., Chan, I.S.F., Annunziato, P.W., & Parrino, J. (2012). Efficacy , Safety , and Tolerability of Herpes Zoster Vaccine in Persons Aged 50 – 59 Years. *Clinical Infectious Diseases*, *54*, 922–928.

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. *The EMBO Journal*, *22*(7), 1608–1619.

Semmes, O.J., Chen, L., Sarisky, R.T., Gao, Z., Zhong, L., & Hayward, S.D. (1998). Mta has properties of an RNA export protein and increases cytoplasmic accumulation of Epstein-Barr virus replication gene mRNA. *Journal of Virology*, *72*(12), 9526–9534.

Sergeant, A., Gruffat, H., & Manet, E. (2008). The Ebstein-Barr Virus (EBV) protein EB2 is an mRNA export factor essential for virus production. *Frontiers in Bioscience*, *13*, 3798–3813.

Smiley, J. (2004). Herpes Simplex Virus Virion Host Shutoff Protein: Immune Evasion Mediated by a Viral RNase? *Journal of Virology*, *78*(3), 1063–1068.

Sokolowski, M., Scott, J.E., Heaney, R.P., Patel, A.H., & Clements, J.B. (2003). Identification of herpes simplex virus RNAs that interact specifically with regulatory protein ICP27 in vivo. *The Journal of Biological Chemistry*, *278*(35), 33540–33549.

Soliman, T.M., & Silverstein, S.J. (2000a). Herpesvirus mRNAs Are Sorted for Export via Crm1-Dependent and -Independent Pathways. *Journal of Virology*, *74*(6), 2814–2825.

Soliman, T.M., & Silverstein, S.J. (2000b). Identification of an export control sequence and a requirement for the KH domains in ICP27 from herpes simplex virus type 1. *Journal of Virology*, *74*(16), 7600–7609.

Souki, S.K., Gershon, P.D., & Sandri-Goldin, R.M. (2009). Arginine methylation of the ICP27 RGG box regulates ICP27 export and is required for efficient herpes simplex virus 1 replication. *Journal of Virology*, *83*(11), 5309–5320.

Souki, S.K., & Sandri-Goldin, R.M. (2009). Arginine Methylation of the ICP27 RGG Box Regulates the Functional Interactions of ICP27 with SRPK1 and Aly / REF during Herpes Simplex Virus 1 Infection. *Journal of Virology*, *83*(17), 8970–8975.

Tait, A.R., & Straus, S.K. (2011). Overexpression and purification of U24 from human herpesvirus type-6 in E. coli: unconventional use of oxidizing environments with a maltose binding protein-hexahistine dual tag to enhance membrane protein yield. *Microbial Cell Factories*, *10*(1), 51–62.

Taniguchi, I., & Ohno, M. (2008). ATP-dependent recruitment of export factor Aly/REF onto intronless mRNAs by RNA helicase UAP56. *Molecular and Cellular Biology*, *28*(2), 601–608.

Tarbouriech, N., Daenke, S., & Cusack, S. (2006). Structural genomics of the Epstein – Barr virus. *Acta Crystallographica Section D*, *D62*, 1276–1285.

Taylor, A., Jackson, B.R., Noerenberg, M., Hughes, D.J., Boyne, J.R., Verow, M., Harris, M., & Whitehouse, A. (2011). Mutation of a C-terminal motif affects Kaposi's sarcoma-associated herpesvirus ORF57 RNA binding, nuclear trafficking, and multimerization. *Journal of Virology*, *85*(15), 7881–7891.

Terwilliger T. (2003). SOLVE and RESOLVE: automated structure solution, density modification and model building. *Journal of Synchrotron Radiation, 11,* 49-52.

Terwilliger T.C., Adams P.D., Read R.J., McCoy A.J., Moriarty N.W., Grosse-Kunstleve R.W., Afonine P.V., Zwart P.H., & Hung L.W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D, Biological Crystallography, 65,* 582-601.

Tokuriki, N., Oldfield, C.J., Uversky, V.N., Berezovsky, I.N., & Tawfik, D.S. (2009). Do viral proteins possess unique biophysical features? *Trends in Biochemical Sciences*, *34*(2), 53–59.

Toth Z., Lischka P., & Stamminger T. (2006). RNA-binding of the human cytomegalovirus transactivator protein UL69, mediated by arginine-rich motifs, is not required for nuclear export of unspliced RNA. *Nucleic Acids Research, 34*, 1237-1249.

Tunnicliffe, R.B., Hautbergue, G.M., Kalra, P., Jackson, B.R., Whitehouse, A., Wilson, S.A., & Golovanov, A.P. (2011). Structural basis for the recognition of cellular mRNA export factor REF by herpes viral proteins HSV-1 ICP27 and HVS ORF57. *PLoS Pathogens*, *7*(1), e1001244, 1-13.

Uversky, V.N., & Dunker, A.K. (2008). Controlled Chaos. *Science*, *322*(November), 1340–1341.

Vaguine A.A., Richelle J., & Wodak S.J. (1999). SFCHECK: A unified set of procedures for evaluating the quality of macromolecular structure factors and their agreement with the atomic model. *Acta Crystallographica, D55*, 191-305.

Valverde, R., Edwards, L., & Regan, L. (2008). Structure and function of KH domains. *The FEBS Journal*, *275*(11), 2712–2726.

Vaughan, P.J., Thibault, K.J., Hardwicke, M.A., & Sandri-Goldin, R.M. (1992). The Herpes Simplex Virus Immediate early protein ICP27 encodes a potential metal binding domain and binds Zinc in vitro. *Virology*, *384*, 377–384.

Verma, D., & Swaminathan, S. (2008). Epstein-Barr virus SM protein functions as an alternative splicing factor. *Journal of Virology*, *82*(14), 7180–7188.

Verma D., Bais S., Gaillard M., & Swaminathan S. (2010). Epstein-Barr Virus SM protein utilizes cellular splicing factor SRp20 to mediate alternative splicing. *Journal of Virology*, *84* (22), 11781-11789.

Wilcox, K., Kohn, A., Sklyanskaya, E., & Roizman, B. (1980). Herpes simplex virus Phosphoproteins: Phosphate cycles on and off some viral polypeptides and can alter their affinity for DNA. *Journal of Virology*, *33*(1), 167-182.

Winkler M., Rice S. A., & Stammingerl T. (1994). UL69 of Human Cytomegalovirus, an Open Reading Frame with Homology to ICP27 of Herpes Simplex Virus, Encodes a Transactivator of Gene Expression. *Journal of Virology, 68* (6), 3943-3954.

Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G., McCoy A., McNicholas S.J., Murshudov G.N., Pannu N.S., Potterton E.A., Powell H.R., Read R.J., Vagin A., Wilson K.S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica. Section D, Biological Crystallography, 67*(Pt 4), 235-242.

Wu, T.T., Qian, J., Ang, J., & Sun, R. (2012). Vaccine prospect of Kaposi sarcoma-associated herpesvirus. *Current Opinion in Virology*, *2*(4), 482-488.

Xu, F., Schillinger, J.A., Sternberg, M.R., Johnson, R.E., Lee, F.K., Nahmias, A.J., & Markowitz, L.E. (2002). Seroprevalence and Coinfection with Herpes Simplex Virus Type 1 and Type 2 in the United States , 1988 – 1994. *The Journal of Infectious Diseases*, *185*, 1019-1024.

Xue, B., Blocquel, D., Habchi, J., Uversky, A.V, Kurgan, L., Uversky, V.N., & Longhi, S. (2014). Structural Disorder in Viral Proteins. *Chemical Reviews*, *114*, 6880–6911.

Xue, B., Dunker, A. K., & Uversky, V. N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure & Dynamics*, *30*(2), 137–49.

Zhao, L., Ren, X., & Zheng, A.C. (2010). Herpes simplex virus type 1 ICP27 protein: its expression, purification and specific antiserum production. *Virologica Sinica*, *25*(3), 199–205.

Zhi, Y., & Sandri-Goldin, R.M. (1999). Analysis of the phosphorylation sites of herpes simplex virus type 1 regulatory protein ICP27. *Journal of Virology*, *73*(4), 3246–3257.

Zhi, Y., Sciabica, K.S., & Sandri-Goldin, R.M. (1999). Self-interaction of the herpes simplex virus type 1 regulatory protein ICP27. *Virology*, *257*(2), 341–351.

Zhou, C., & Knipe, D.M. (2002). Association of Herpes Simplex Virus Type 1 ICP8 and ICP27 Proteins with Cellular RNA Polymerase II Holoenzyme. *Journal of Virology*, *76*(12), 5893–5904.

Zhou, Z., Luo, M., Straesser, K., Katahira, J., Hurt, E., & Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature*, *407*(September), 401–405.

PUBLICATIONS AND MANUSCRIPTS

1. **Patel, V.**, Dahlroth, S.L., Venkatachalam, R., Ho, H.T., Cornvik, T. and Nordlund, P. (2015). Structure of C-Terminal Domain of the Multifunctional ICP27 Protein from Herpes Simplex Virus-1. *Journal of Virology*, 89 (17).

2. Working Manuscript:
   Dahlroth, S. L., Nilsson, M., Dahlgren, L. G., Hew, K., **Patel, V.**, Chen, Y., Darwis, D., Ho, H. T., Jong, A., Karlsson, J., Lim, B. T., Nasertorabi, F., Pan, L., Yeung, K., Verappan, S., Zhang, V., Cornvik, T. and Nordlund, P. *Obtaining recombinant expression for herpesviral proteins in E. coli – strategies, results and resources.*

CONFERENCE/ WORKSHOP ATTENDANCE

1. 'Frontiers in Structural Biology' and the joint meeting on 'G Protein-Coupled Receptors: Structural Dynamics and Functional Implications', Snowbird, Utah, USA (March 30th to April 4th 2014).

2. 'CCP4 School on Advanced X-ray crystal structure analysis: Inaugural Australian Advanced Crystallography Methods Workshop', Australian Synchrotron, Clayton, Victoria, Australia (February 12th to 15th 2012).

POSTERS PRESENTED

1. SBS Graduate Students' Retreat, Nanyang Technological University, Singapore. October 5th 2012.
   Title: The Centre for Biomedical Structural Biology

2. Frontiers in Structural Biology, Snowbird, Utah, USA. March 30th to April 4th 2014.
   Title: Crystal Structure of the Core Domain of Regulatory Protein ICP27 from Herpes Simplex Virus Type - 1

SCHOLARSHIP

NTU Research Scholarship 2010 - 2014