UNIVERSITY OF CALIFORNIA,
IRVINE

A Structural and Functional Investigation of the Essential HSV-1 Protein ICP27

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

William K. Hu

Dissertation Committee:
Professor Rozanne M. Sandri-Goldin, Chair
Professor Klemens J. Hertel
Professor Yongsheng Shi

2020

DEDICATION

To

My favorite people in the world, Christina Hong and William Hong Hu

My parents, Angel and Danny Hu

My siblings, Winston and Waresa

And all my family and friends

Thank you for the love and support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank my thesis advisor, Dr. Rozanne M. Sandri-Goldin, for her support and guidance during my graduate career. I would also like to express my appreciation to my committee members, Dr. Klemens J. Hertel and Dr. Yongsheng Shi for their feedback and advice.

I thank the former and current members of the Sandri-Goldin laboratory for their support and assistance. I am especially grateful to Michele Wu for her friendship and help during my graduate career. I would also like to thank Dr. Mark Ou and Brian Kim for sharing their knowledge and expertise and Dr. Gayathri Devi-Rao and Mitchell Schacht for their helpful scientific discussions.

CURRICULUM VITA

William K. Hu

EDUCATION

2020	Ph.D. Biomedical Sciences, University of California, Irvine
2006-12	Staff Research Associate II, Department of Medical Microbiology and Immunology, University of California, Davis
2006	B.A. Biological Sciences, University of California, Davis
2004-06	Research Assistant, Center for Comparative Medicine, University of California, Davis
2003-04	Research Assistant, Plant Sciences Department, University of California, Davis

GRANTS

2015	NIH Virology Training Grant T32 AI 07319

LEADERSHIP

2016-2017	Graduate student representative for the Microbiology and Molecular Genetics Department at the University of California, Irvine
2014-2016	Microbiology and Molecular Genetics Department seminar committee member

PUBLICATIONS

Wang, X., Hennig, T., Whisnant, A.W., Erhard, F., Prusty, B.K., Friedel, C.C., Forouzmand, E., **Hu, W**., Erber, L., Chen, Y., Sandri-Goldin, R.M., Dölken, L., Shi, Y., 2020. Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. Nat. Commun. 11, 1–13.

Tunnicliffe, R.B., **Hu, W.K**., Wu, M.Y., Levy, C., Mould, A.P., McKenzie, E.A., Sandri-Goldin, R.M., Golovanov, A.P. (2019). Molecular Mechanism of SR Protein Kinase 1 Inhibition by the Herpes Virus Protein ICP27. mBio 10.

Klein, R.H., **Hu, W**., Kashgari, G., Lin, Z., Nguyen, T., Doan, M., Andersen, B. (2017). Characterization of enhancers and the role of the transcription factor KLF7 in regulating corneal epithelial differentiation. J. Biol. Chem. 292, 18937–18950.

George, M.D., **Hu, W**., Billingsley, J.M., Reeves, R.K., Sankaran-Walters, S., Johnson, R.P., and Dandekar, S. (2014). Transcriptional profiling of peripheral CD8+T cell responses to SIVΔnef and SIVmac251 challenge reveals a link between protective immunity and induction of systemic immunoregulatory mechanisms. Virology 468–470, 581–591.

Hirao, L.A., Grishina, I., Bourry, O., **Hu, W.K**., Somrit, M., Sankaran-Walters, S., Gaulke, C.A., Fenton, A.N., Li, J.A., Crawford, R.W., et al. (2014). Early mucosal sensing of SIV infection by paneth cells induces IL-1β production and initiates gut epithelial disruption. PLoS Pathog. 10, e1004311.

Verhoeven, D., George, M.D., **Hu, W**., Dang, A.T., Smit-McBride, Z., Reay, E., Macal, M., Fenton, A., Sankaran-Walters, S., and Dandekar, S. (2014). Enhanced innate antiviral gene expression, IFN-α, and cytolytic responses are predictive of mucosal immune recovery during simian immunodeficiency virus infection. J. Immunol. Baltim. Md 1950 192, 3308–3318.

Nagy, L.H., Grishina, I., Macal, M., Hirao, L.A., **Hu, W.K**., Sankaran-Walters, S., Gaulke, C.A., Pollard, R., Brown, J., Suni, M., et al. (2013). Chronic HIV infection enhances the responsiveness of antigen presenting cells to commensal Lactobacillus. PloS One 8, e72789.

Patterson, L.J., Daltabuit-Test, M., Xiao, P., Zhao, J., **Hu, W**., Wille-Reece, U., Brocca- Cofano, E., Kalyanaraman, V.S., Kalisz, I., Whitney, S., et al. (2011). Rapid SIV Env- specific mucosal and serum antibody induction augments cellular immunity in protecting immunized, elite-controller macaques against high dose heterologous SIV challenge. Virology 411, 87–102.

Smythies, L.E., Shen, R., Bimczok, D., Novak, L., Clements, R.H., Eckhoff, D.E., Bouchard, P., George, M.D., **Hu, W.K**., Dandekar, S., et al. (2010). Inflammation anergy in human intestinal macrophages is due to Smad-induced IkappaBalpha expression and NF-kappaB inactivation. J. Biol. Chem. 285, 19593–19604.

Teo, G., Suzuki, Y., Uratsu, S.L., Lampinen, B., Ormonde, N., **Hu, W.K**., DeJong, T.M., and Dandekar, A.M. (2006). Silencing leaf sorbitol synthesis alters long-distance partitioning and apple fruit quality. Proc. Natl. Acad. Sci. U. S. A. 103, 18842–18847.

ABSTRACT OF THE DISSERTATION

A Structural and Functional Investigation of the Essential HSV-1 Protein ICP27

By

William K. Hu

Doctor of Philosophy in Biomedical Sciences

University of California, Irvine, 2020

Professor Rozanne M. Sandri-Goldin, Chair

Herpes simplex virus 1 ICP27 is an essential multifunctional regulatory protein that assumes different roles during infection. ICP27 disrupts multiple aspects of cellular gene expression, contributing to the shutoff of host protein synthesis. ICP27 also recruits RNA polymerase II to viral replication sites and mediates the export of viral mRNA. Furthermore, ICP27 has also been implicated in nuclear protein quality control, cell cycle control, activation of stress signaling pathways and apoptosis. ICP27 performs its activities by interacting with RNA and a myriad of proteins. However, how ICP27 is able to interact with so many different binding partners and how these interactions are regulated during the viral life cycle is not well understood. ICP27 has been reported to form homo-dimers and to undergo an intramolecular N- to C-terminal interaction. It's hypothesized that these different configurations provide the flexibility it needs to interact with various binding partners and events or factors that promote one isomer over the other may serve to regulate different ICP27 activities. At present, how these configurations are related, and their specific functions, have not been well characterized. Moreover, the molecular details as to how ICP27 targets host RNA processing and the wider impacts its activities have on

the host transcriptome have also not been detailed. In these studies, we (1) use a bimolecular fluorescence complementation and western blot-co-immunoprecipitation approach to perform a structural analysis in order to elucidate the functions of ICP27's different conformations, (2) utilize a combination of biophysical and molecular techniques in order to detail the molecular mechanism of how ICP27 inactivates SR protein kinase 1, an important regulator of SR proteins, to inhibit cellular splicing, and (3) perform next-generation sequencing to globally Identify ICP27 modulations to host gene expression.

The results from our studies show: (1) ICP27's conformational isomers utilize different protein domains for their assembly and these isomers can interact with different binding partners, (2) the ICP27 RGG box binds to the SRPK1 substrate docking groove in a similar manner and affinity as SRPK1's cellular substrate SRSF1, and (3) ICP27 promotes accumulation of host antisense transcripts during infection. Together, the information gained in this dissertation expands our understanding of the molecular principles by which ICP27 functions.

CHAPTER 1

Introduction

Herpesviridae

The *Herpesviridae* family represent a large group of enveloped, DNA viruses associated with a variety of disease in a broad range of hosts. Members of this family share several biological properties that distinguish it from other viral families including: expression of enzymes involved in nucleic acid metabolism, DNA synthesis and protein processing; restriction of DNA replication and capsid assembly to the host cell nucleus, with virion maturation and final envelopment occurring in the cytoplasm; destruction of the infected cell during an active productive infection; and ability to persist in the host cell in a latent state indefinitely.

Nine species of *Herpesvirida*e are known to cause disease in humans. Based upon biological similarities, DNA sequence homology and genomic structure, human herpesviruses are further categorized into three subfamilies, alpha-, beta-, and gamma-herpesviruses. Key features that distinguish these subfamilies from each other are the replication cycle and host cell range.

Alpha-herpesviruses are comprised of herpes simplex virus (HSV) -1 and -2 and varicella zoster virus (VZV). HSV-1 and -2 have a short replication cycle with efficient cell destruction and a variable host range. They replicate efficiently in tissue culture and have therefore been used as model systems for other herpesviruses and for eukaryotic cell biology. VZV has a longer replication cycle in tissue culture and *in* vivo. VZV replicates to

relatively low titers in tissue culture and is highly cell-associated. Alpha-herpesviruses typically initiate productive infection in somatic cells prior to establishing latency within the peripheral nervous system[1].

Beta-herpesviruses, consisting of human cytomegalovirus (HCMV) and human herpes virus (HHV) -6A, -6B and -7, have a long reproductive cycle, a more restricted host range, and are less cytopathic than alpha-herpesviruses. HCMV undergoes productive infection in a number of cell types, including smooth muscle, endothelial, epithelial, fibroblasts, macrophages and dendritic cells. HHV-6A and -6B are known to replicate preferentially in human T-lymphocytes. HCMV has been reported to predominantly establish latency in CD34+ myeloid cells in the bone marrow while the site of latency for HHV-6A and -6B have not been well established.

Gamma-herpesviruses are lymphotropic viruses that also have a relatively long replication cycle. These viruses have a very narrow host range, being unable to replicate in organisms not belonging to the family or order of its natural host. Replication occurs primarily in B and T lymphocytes, where they also establish latency. There are two gamma-herpesviruses known to infect humans, Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV). Unlike alpha- and beta-herpesviruses, gamma-herpesvirus infection induce cell proliferation and have been consistently associated with the development of lymphoproliferative diseases and lymphomas in a population of infected individuals[3].

Of the human herpesviruses, at least six - HSV-1 and -2, VZV, HCMV, KSHV and EBV - are extremely widespread among humans. More than 90% of adults in the developed

world are infected with at least one of these species, many of which were acquired during childhood and adolescence.

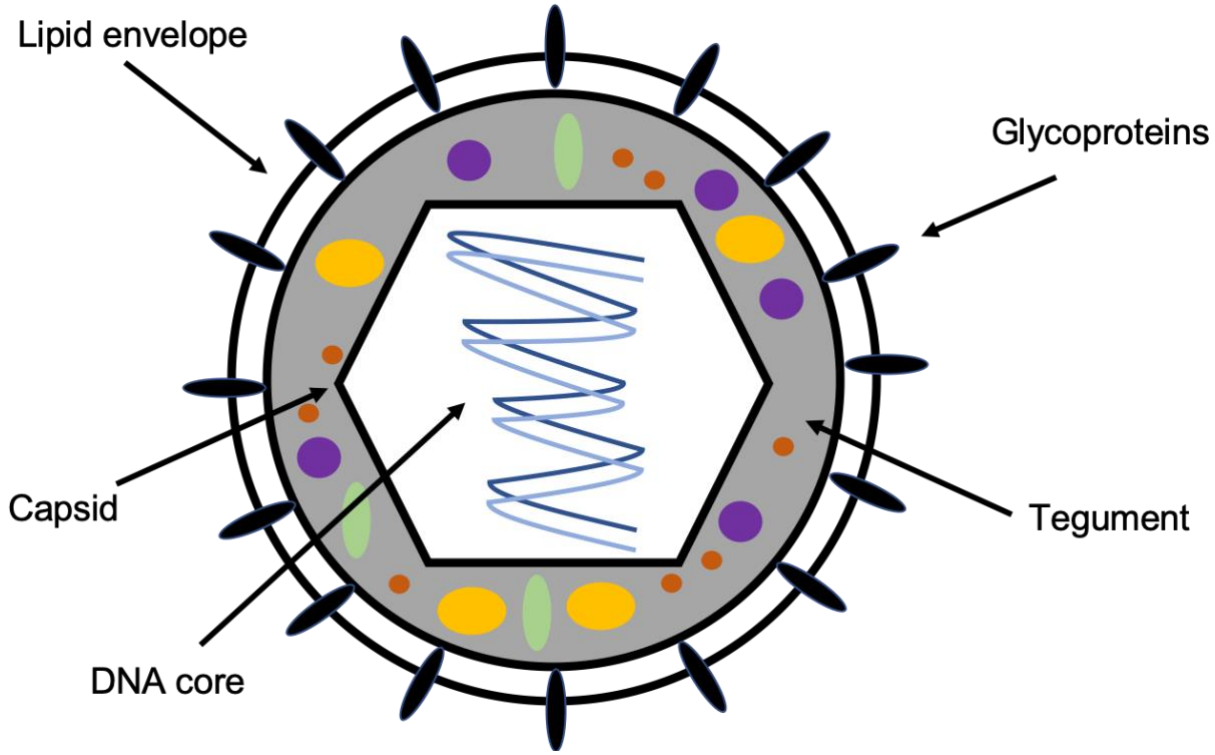
Herpes Simplex Virus 1

The alpha-herpesvirus HSV-1 is the prototype herpesvirus and has served as a powerful model for understanding the molecular details of host cell biology and viral pathogenesis. HSV-1 is common and endemic throughout the world and it's estimated that 70-90% of the adult population are seropositive for HSV-1. Active HSV-1 infections are typically characterized by the development of painful skin lesions or blisters at the mucosal surfaces around the mouth or genital area that resolve without treatment within a period of two to four weeks. However, in more susceptible populations, as in neonates and immunosuppressed individuals, infection with HSV-1 can lead to serious complications including death, lasting neurological disability or encephalitis[4].

The HSV-1 virion shares a common structure with other members of the *Herpesviridae* family (Figure 1.1). The virus particle contains four distinct structural elements: (1) the inner core containing the linear double-stranded DNA genome, (2) an icosahedral capsid approximately 100 nanometers in diameter surrounding the core, (3) an amorphous appearing tegument layer containing several important viral regulatory proteins, (4) and a host derived lipid envelope studded with viral glycoproteins important for receptor binding and entry. Upon fusion of the virus particle with the host cell membrane, the incoming nucleocapsid is rapidly transported along microtubule networks to the nuclear pore complex (NPC) where the viral DNA genome is released into the nucleus for transcription and replication[5].

Figure 1.1. Structure of the HSV-1 virion.

The HSV-1 virion is composed of an icosahedral capsid containing the double-stranded DNA core, a lipid envelope spiked with several glycoproteins and a layer filled with viral regulatory proteins, termed the tegument, which resides between the capsid and envelope.


Lipid envelope
Glycoproteins
Capsid
Tegument
DNA core

The HSV-1 genome is 152 kilobase (Kb) pairs, with a rich GC-content of approximately 68%, coding for more than 80 proteins. All viral transcripts require host RNA polymerase II (RNAP II) and other cellular transcription factors for expression[6]. Consistent with this, viral promoters share many elements common to cellular promoters, including TATA boxes and binding sites for host transcription factors. Additionally, several viral proteins are also required for optimal expression of most viral genes. During lytic infection, expression of viral genes is carefully regulated and proceeds in a temporal cascade of immediate early (IE), early (E), then Late (L) genes. In certain cell lines, progeny virions can be detected in as little as 8 hours post infection (hpi) although peak infection titers usually occur after 18-20 hours.

Transcription of the IE genes begins within 1 hour of infection and peaks between 2-4 hpi; although some expression can still be detected at late times[6]. They require only the viral tegument protein VP16 in conjunction with the cellular transcription factors Oct-1, RNA polymerase II (RNAP II) and Host Cell Factor (HCF) for their expression. This transcription complex binds the consensus TAATGARAT promoter sequence present upstream of all IE genes to activate their expression. Of the five IE genes expressed by HSV-1, four (ICP0, ICP4, ICP22, ICP27) code for regulatory proteins that return to the nucleus after translation to regulate the expression of viral and/or cellular genes while the fifth IE gene (ICP47) encodes a protein that's retained in the cytoplasm where it functions as an immune modulator.

Following the IE phase, the second class of genes expressed are the E genes. This class requires the IE gene products for their expression but not viral DNA replication. The promoters of E genes have served as models for eukaryotic promoters and contain

TATA boxes and binding sites for host transcription factors. Expression of E genes peaks approximately 5-7 hpi and these genes primarily encode proteins involved in DNA replication and nucleic acid metabolism[6].

The final group of viral genes produced during lytic infection are the L genes. This class primarily encodes structural components of the virion and is divided into leaky-late and true-late, depending on whether they require viral DNA replication for their expression. Promoters controlling expression of both classes share some common elements (e.g., TATA boxes) near the transcription start site required for promoter activity, however the locations of other elements (upstream transcription factor binding sites, initiator elements, downstream activator elements) may differ[7]. Leaky-late transcripts can be detected at early stages of infection, however, peak expression of this class is only reached after the initiation of viral DNA replication. In contrast, true-late genes strictly require DNA replication and are not detected prior to viral DNA synthesis.

Upon primary infection of mucosal epithelial cells, progeny virions can infect the sensory neurons innervating the site of initial infection. Through retrograde microtubule-associated transport, the virus gains access to the nerve cell body where it can enter a quiescent state in which the latent viral genome is maintained as a circular episome and the expression of all viral lytic genes are repressed. Following various triggering factors, the viral genome can reactivate to resume virus replication and produce infectious viral particles, which are transported back to the periphery to give rise to secondary infections[8].

Infected Cell Protein 27 (ICP27)

ICP27 is a 63 kilodalton (kDa), 512 amino acid protein expressed during the IE stage of infection. ICP27, along with ICP4, make up the two essential IE proteins required for a productive infection. However, unlike ICP4, which functions as the main viral transcriptional transactivator necessary for all post IE gene expression, ICP27 is a multifunctional regulatory protein that plays a role in many different processes during the viral lifecycle.

*Viral gene expression***.** ICP27 participates in all stages of viral mRNA biogenesis. During the early stages of infection, ICP27 binds directly to the C-terminal domain (CTD) of RNAP II and recruits RNAP II to viral replication sites to promote the transcription of viral E and L genes[9]. This interaction requires both the N- and C-terminal ends of ICP27, as infection with mutant viruses expressing alterations within the N- and C-termini failed to relocate RNAP II, resulting in significantly reduced E and L transcripts[9].

At approximately 5-6 hpi, ICP27 mediates the transport of viral mRNA from the nucleus to the cytoplasm. In the host cell, gene expression is tightly coupled with mRNA maturation. This process is highly regulated, with pre-mRNA splicing playing a critical role, not only for the removal of introns but also for the recruitment of the multi-protein Transcription-Export (TREX) complex, which is required for mRNA export through the TAP/NXF1 pathway[10-11]. Unlike cellular mRNA, the vast majority of HSV-1 transcripts are intronless and therefore unable to acquire the export markers normally deposited through the coupled pathways of transcription and splicing[12]. ICP27 facilitates the efficient export of unspliced viral transcripts by binding viral mRNA, predominantly through an N-terminal RGG motif, and interacting with cellular export adapters, such as ALYREF, a member of the

TREX complex, as well as the cellular mRNA export receptor TAP/NXF1 to guide viral transcripts through the NPC for translation[13-14].

In addition to promoting viral gene expression and export, ICP27 can also directly induce the translation of certain viral mRNAs[15]. In the cytoplasm, ICP27 has been shown to be associated with polyribosomes and interacts with at least three cellular translation initiation factors: poly A binding protein (PABP), eukaryotic initiation factor 3 (eIF3) and eukaryotic initiation factor4 G (eIF4G)[16-17]. In infections with the ICP27 null mutant virus d27-1, translation of viral L transcripts VP16 and ICP5 were reduced. However, translation levels for another L mRNA encoding glycoprotein D (gD) were not affected, indicating that ICP27 promotes the translation of a subset of HSV-1 L genes[16]. At present, the full repertoire of viral transcripts requiring ICP27 for maximal protein expression is unknown.

Host cell shutoff. A common occurrence seen during the progression of many viral infections is the concomitant decrease in host protein synthesis[18]. This phenomenon, termed "host shutoff" is thought to promote virus replication by freeing up host cell machinery and redirecting cellular resources towards viral gene expression. Furthermore, shutoff of host protein synthesis can promote immune evasion and interfere with anti-viral stress responses[18]. During infection with HSV-1, the accumulation of viral proteins is associated with an almost complete suppression of cellular protein production[19-20]. This global shutoff relies, at least in part, on the actions of ICP27, which plays a particularly prominent role in disrupting various facets of host pre-mRNA maturation.

ICP27 interacts with and regulates multiple splicing proteins and splicing associated factors including, the U1 small nuclear ribonucleoprotein (snRNP), spliceosome-association protein 145 (SAP145) and SR proteins - SRSF1, 2, 3, and 7[21-22]. In addition, ICP27 has also

been shown to interact with a key regulator of SR protein function, SR protein kinase 1 (SRPK1), and relocates this normally cytoplasmic host kinase to the nucleus leading to aberrant phosphorylation of SR proteins and disruption of spliceosome assembly, ultimately resulting in increased levels of intron retention[23]. Recent reports have indicated ICP27 mediates splicing inhibition in a subset of virus-like cellular transcripts[20,24]. A study by Tang et al. (2016), utilizing RNA-sequencing of cells transiently transfected with ICP27 or infected with wild-type HSV-1 virus revealed that, in the host cell, ICP27 functions as an alternative splicing regulator, promoting usage of alternative 5' splice sites and inhibiting splicing of certain introns in <1% of cellular genes[24]. Further, the authors also reported ICP27 induced expression of pre-mRNAs prematurely cleaved and polyadenylated from cryptic poly(A) sites (PAS) in approximately 150 cellular genes[24]. These novel transcripts were expressed early during infection, <2 Kb in length and efficiently exported to the cytoplasm. ICP27-mediated disruption of host mRNA maturation likely leads to reduced expression and alterations in mRNA stability for many of the targeted genes, thereby contributing to virus-induced host shutoff. Although the exact molecular mechanism for how ICP27 induces aberrant cellular pre-mRNA processing has yet to be defined, analysis of afflicted transcripts showed high GC-content, similar to viral genes, cytosine-rich sequences near the 5' splice site and suboptimal splice site sequences within the impacted intron[24].

A recent analysis of the temporal changes in host and viral transcriptional and translational activity during lytic HSV-1 infection found widespread defects in RNAP II transcription termination in cellular but not viral genes[20]. This resulted in thousands of host genes displaying varying levels of intergenic reads past their normal termination

points which were not able to associate with ribosomes[20]. In a follow up study by Wang et al. (2020), disruption of transcription termination was directly attributed to the actions of ICP27[25]. ICP27 was found to function as a general inhibitor of mRNA 3' processing by interacting with various components of the essential 3' processing factor CPSF and inducing the formation of a defective 3' processing complex unable to perform its role in cleavage and polyadenylation[25]. However, ICP27 was also reported to be a sequence dependent activator of 3' processing, necessary for efficient RNAP II transcription termination in viral and a subset of host genes[25]. This action required the N-terminal RGG motif (Figure 1.2). The RGG motif is the main RNA-binding domain for ICP27 and has been reported to preferentially bind GC-rich substrates[14,26,]. As the viral genome contains a substantially higher GC-content than the host genome, ICP27 targets viral (and GC-rich host) transcripts by binding to the GC-rich sequences upstream of PAS and recruiting CPSF and other 3' processing factors to the core PAS for efficient 3' processing[25].

Other functions of ICP27. Besides the aforementioned roles in viral and host gene expression, ICP27 also enhances viral infection by: (1) recruiting the cellular chaperone, Heat shock cognate protein 70 (Hsc70) to virus-induced chaperone-enriched (VICE) domains located at the periphery of viral replication compartments resulting in the removal of stalled RNAP II transcription complexes on the viral genome by proteasomal degradation[9,27], (2) depressing the expression of cell cycle regulators, cyclin D1 and cyclin-dependent kinase 4 (cdk4), and blocking phosphorylation of retinoblastoma proteins (pRB) involved in the G1- to S-phase transition, restricting the cell cycle to the G1 phase[28], (3) activating NF-κB by triggering the loss of inhibitory IκBα[29-30] and inhibiting type I interferon expression by downregulating signal transducers and activators of transcription

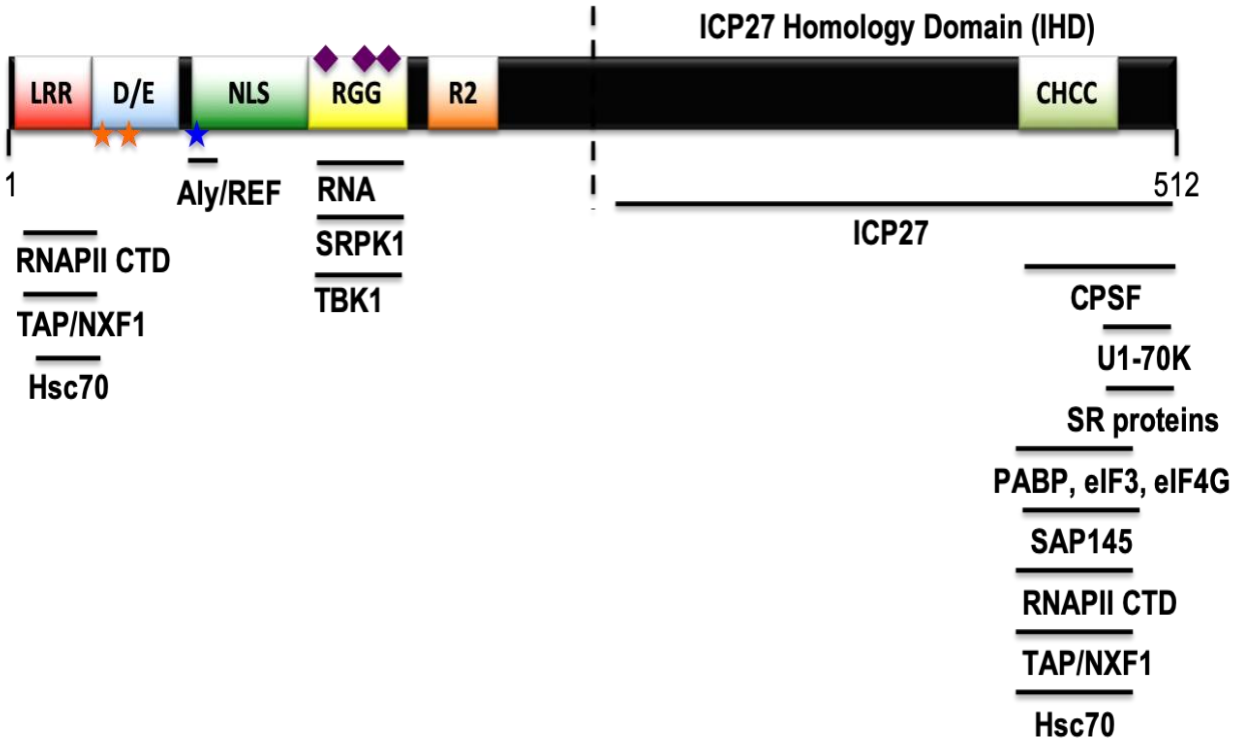
(STAT) 1 phosphorylation and nuclear accumulation in productively infected cells[31] or targeting the stimulator of interferon genes (STING)-TANK-binding kinase 1 (TBK1) signalsome in macrophages[32], and (4) activating the stress-activated protein kinases p38 and JNK[33] and regulating cellular apoptosis[34-35].

ICP27 structure. ICP27's ability to participate in such a diverse array of processes during infection is reflected in its protein structure. Ordered proteins that fold into rigid 3D-structures, such as viral structural proteins, lack the flexibility required for interaction with multiple protein partners. In contrast, many regulatory proteins known to carry out numerous biological functions, including cell signaling, molecular recognition, and interactions with different proteins and nucleic acids have been found to fit into the growing class of intrinsically disordered or unstructured proteins[36-37]. Intrinsically disordered proteins allow for molecular plasticity and binding diversity by adopting different conformations upon interaction with various targets. Secondary structure analysis based on sequence homology predicts ICP27 contains both an intrinsically disordered N-terminal region and a structured, globular C-terminal region[12]. Nuclear magnetic resonance (NMR) studies on the N-terminal 160 amino acids of the protein have confirmed that this region is flexible and contains relatively unstructured loops[38]. The C-terminal, globular region of ICP27 is highly conserved among all herpesviruses and has been termed the ICP27 homology domain (IHD)[39]. The regions of ICP27 required for a number of its interactions have been mapped through mutational analysis (Figure 1.2). ICP27 binds RNA through an RGG motif consisting of 15 arginine and glycine residues from amino acids 138-152[14,40]. The RGG motif is also required for ICP27 to bind the cellular

kinases, SRPK1 and TBK1[23,32]. Binding with TREX complex member ALYREF was mapped to

Figure 1.2. Schematic diagram of ICP27's functional interactions.

Depiction of ICP27 showing the leucine-rich region (LRR) necessary for export, acidic region (D/E) with two CK2 phosphorylation sites (orange stars); the nuclear localization signal (NLS) with a PKA phosphorylation site (blue star); the RGG box with three PRMT1 methylation sites (purple diamonds); R-rich region (R2), and a zinc finger domain (CHCC). The border between the N-terminal disordered region and the C-terminal structured region is delineated by a dashed line and the regions of interaction with proteins and RNA are shown below.


ICP27 Homology Domain (IHD)
LRR
D/E
NLS
RGG
R2
CHCC
1
512
Aly/REF
RNA
SRPK1
TBK1
ICP27
RNAPII CTD
TAP/NXF1
Hsc70
CPSF
U1-70K
SR proteins
PABP, eIF3, eIF4G
SAP145
RNAPII CTD
TAP/NXF1
Hsc70

a short stretch of residues within the nuclear localization signal (NLS) adjacent to the RGG box and interactions with CPSF, splicing factors and translation initiation factors were mapped to a region within the C-terminus from amino acids 450-512[13,17,25,41,42]. What's more, ICP27 interaction with at least three cellular proteins, Hsc70, RNAP II, and TAP/NXF1 requires both the N-terminal leucine-rich region (LRR) and the C-terminal zinc-finger region to be intact[9,27,43]. Bimolecular fluorescence complementation (BiFC) and BiFC based fluorescence resonance energy transfer (FRET) studies have shown that the N- and C-termini of ICP27 undergo an intramolecular self-interaction in vivo and this interaction was necessary for ICP27-TAP/NXF1 binding and export[43,44].

Further, ICP27 was found to dimerize *in-vivo*. The crystal structure of the IHD from residues 241-512 revealed a homo-dimer with each chain comprised of 10 α-helices of varying lengths separated by loop regions[39,45]. The structure of each monomer forms a globular core, which becomes closely intertwined in a yin-yang conformation forming a compact species with numerous intermolecular contacts. Bioinformatic analysis indicated that 47% of residues observed in the structure have atoms within the dimer interface and 88 polar intermolecular contacts (62 hydrogen bonds and 26 salt bridges) were detected[39]. In addition, the crystal structure revealed a novel zinc-binding motif that differed from previously predicted coordination sites. Residues C483, C488, H502 and C508 were initially proposed to be the zinc ligands in ICP27, however, structural analysis revealed each chain actually coordinates zinc in a tetrahedral configuration via residues C400, H479, C483 and C488. The CHCC zinc-binding site is highly conserved among all ICP27 homologs and is thought to have a role in structural stabilization rather than catalytic function[39]. One

of the most striking features of the homo-dimer is a domain swap where the N- and C-terminal regions form extensions surrounding and inserting into the neighboring subunit. The first two N-terminal α-helices adopt an arm-like configuration that wraps around the neighboring subunit while the C-terminal tail from residues 500-512 extends into its dimeric partner and becomes enclosed within the subunit. On the dimer's surface, several potential binding pockets were observed, which may facilitate interactions with other biological molecules: on one face, a shallow cleft with two surface exposed hydrophobic patches, on the opposite face, a larger groove running along the major dimer interface and, positioned on either side of the groove, an extensive cluster of six arginine side chains per monomer (R416, R417, R418, R435, R439, R442)[39]. The function of this RRR motif is unknown, however, it may have the potential to bind negatively charged species such as RNA.

Regulation of ICP27 functions. Currently, the regulatory details governing ICP27's different roles during the viral lifecycle are still unclear. ICP27 is post-translationally modified by phosphorylation and arginine methylation, modifications that regulate some of its interactions. Three arginine residues within the RGG motif (R138, R148, R150) are methylated by the cellular methyltransferase, protein arginine methyltransferase 1 (PRMT1) (Figure 1.2), and methylation was shown to be important for ICP27 export and binding with SRPK1 and ALYREF[46–48]. Phosphorylation of ICP27 by casein kinase 2 (CK2) occurs on serine residues 16 and 18, within the LRR, and a serine at amino acid 114, within the NLS, is phosphorylated by protein kinase A (PKA) (Figure 1.2). Mutation of the phosphorylated residues disrupted ICP27 localization, viral replication compartment formation and gene expression[38,49].

ICP27 also undergoes several structural changes during the course of infection, including adoption of a closed form, conferred when its N- and C-termini come into direct contact, and a homo-dimer, formed upon binding between the CTD of two monomeric subunits. The regulation of these configurations is likely a key means to modulate many ICP27 functions. Control of the open and closed states and dimer formation may conceal/reveal active sites or extend binding sites to increase specificity. Moreover, formation of the dimer likely generates new binding pockets required for specific interaction partners and increases the complexity of possible assemblies involving ICP27, allowing for simultaneous binding of different partners for each subunit.

Significance and Goals

Although much is known about ICP27, important questions remain. The events controlling N- to C-terminal interactions or dimer formation have not been elucidated. The biological significance of dimerization and the relationship between the dimeric and N- to C-terminal configurations have yet to be explored. The breadth of interactions specific to each conformation and the kinetics of multimerization are unknown. Furthermore, details in how ICP27 disrupts host RNA processing, and the full scope of ICP27-mediated alterations within the cellular transcriptome during HSV-1 infection have also not been reported. The experimental data within this dissertation seeks to address some of these questions and build upon the body of ICP27 research.

HSV-1 is a ubiquitous human pathogen correlated with a raft of negative physical and psychosocial consequences. At present, there is no vaccine available for HSV-1 although antivirals, such as Acyclovir and related nucleoside analogues, are available.

However, the continued use of these therapies has led to the emergence of drug resistant strains, especially in immunocompromised individuals, creating an urgent need for alternative treatment strategies. ICP27 is an essential multifunctional viral adapter that assumes numerous roles throughout the course of infection. Additionally, ICP27 is highly conserved, with homologs found in every herpesvirus sequenced so far, making it an enticing target for drug development. By better understanding the molecular principles underpinning key ICP27 interactions, it may be possible to develop novel therapeutics that can disrupt critical processes necessary for viral pathogenesis.

CHAPTER 2

ICP27 protein domains necessary for N- to C-terminal interaction or homo-dimerization

Introduction

A unique feature of ICP27 structure that has not been found among its homologs is its ability to self-associate *in vivo* via its amino and carboxyl ends. Using the BiFC approach, we have shown that ICP27 undergoes an intramolecular N- to C- interaction but not an intermolecular N- to C- interaction (Figure 2.4)[44]. Substitution mutations within the N-terminal LRR (residues 5-16) or the C-terminal zinc-finger (C400, H479, C483, C488) severely reduced the fluorescence complementation of the N- and C-termini whereas mutation of two serine residues adjacent to the LRR had no discernible effect on self-association[44]. This indicated that both the LRR and the zinc-finger domain (ZFD) must be intact for the closed configuration to occur. Furthermore, two publications presenting the IHD crystal structure have shown ICP27 as a tightly intertwined homo-dimer where the N-terminal α-helices and the C-terminal tail extend from the core domain to make extensive interactions with the neighboring subunit[39,45]. The C-terminal tail, consisting of a stretch of highly hydrophobic residues from amino acids 500-512, displays a great deal of conservation among the herpesvirus family and is critical for stable dimer formation[39,45]. Co-immunoprecipitation (co-ip) experiments with a yellow fluorescent protein or green fluorescent protein tagged ICP27 (YFP-ICP27/GFP-ICP27) and ICP27 truncation mutants,

n504 (residues 1-504) or STOP500 (residues 1-500), indicated that mutant proteins lacking the tail motif were unable to dimerize *in vivo*[39].

Thus, mutational experiments have identified the LRR and ZFD to be essential for N- to C-terminal interactions whereas the C-terminal tail is required for dimer interactions. However, how mutations within the LRR and ZFD affect dimerization and what role, if any, the C-terminal tail plays in end-to-end self-association has not been reported. As mentioned earlier, a number of ICP27's binding partners have been identified through mutational analysis. These studies have been invaluable in elucidating ICP27's various functions during infection. Yet, structural data for ICP27 has only recently been available and it's not known how the introduced mutations used in these studies affected overall structure or dimer formation. Here, through BiFC and co-ip-western blot analysis, we re-revaluate these previously used mutants, as well as introduce new, targeted mutations using the ICP27 crystal structure as a guide to better elucidate the roles of the dimeric and N-to C-terminal configurations during infection.

Materials and Methods

Cells, viruses and recombinant plasmids

Rabbit skin fibroblast (RSF) cells were grown on minimal essential medium supplemented with 8% fetal bovine serum and 4% donkey calf serum. HeLa cells were grown on minimal essential medium containing 10% newborn calf serum. HSV-1 wild-type strain KOS, ICP27 viral mutants 27LacZ, N-YFP-ICP27, dLeu, d1-2, d3-4, d4-5, d5-6, WRL, and N504 were previously described[39,42,43,50]. Plasmids pN-Venus-ICP27-C-Venus, pN-

Venus-ICP27, pICP27-C-Venus, pN-Venus-ICP27-C483,488S-C-Venus, pSG130B/S (WT), pGFP-ICP27, pCFP-ICP27-C483, 488S and pS18, were described previously[43-44,51]. Plasmids pN-Venus-ICP27-S18-C-Venus, pN-Venus-ICP27-d4-5-C-Venus and pN-Venus-ICP27-Δ500-512-C-Venus were created by site-directed mutagenesis of pN-Venus-ICP27-C-Venus using the Stratagene QuikChange kit. Plasmid pICP27-Δ500-512-C-Venus was generated from digestion and re-ligation of pN-Venus-ICP27-Δ500-512-C-Venus to cut out the N-Venus fragment and pN-Venus-ICP27-Δ500-512 was generated from pICP27-Δ500-512-C-Venus and pN-Venus-ICP27. Plasmids pGFP-Stop500 and pGFP-ICP27-R416, 417, 418, 435, 439, 442A (GFP-R→A) were generated by site-directed mutagenesis of pGFP-ICP27.

Virus infection, transfection and immunoprecipitation procedures

Cells were infected with wild-type or mutant virus at a multiplicity of infection (MOI) of 10 for single infections and a MOI of 5 for co-infections and incubated at 37°C for 8 hours. Transfection of plasmid DNA was performed by using Lipofectamine 2000 reagent (Invitrogen) according to manufacturer's protocol. Transfected cells were infected with ICP27 null mutant virus 27LacZ 24 hours after transfection to stimulate expression of the native ICP27 promoter and to replicate the conditions of infection. Eight hours after infection, cells were harvested and immunoprecipitations were performed on cell lysates using GFP/YFP antibody Ab290 (Abcam) as described previously[39]. Immunoprecipitated complexes were separated by SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose. Western blot analysis was performed as described previously[39] with anti-

ICP27 antibodies P1119 and P1113 (Virusys), anti-GFP/YFP Ab290 (Abcam), anti-GFP antibody Ab1218 (Abcam) and Beta actin (Abcam).

Bimolecular Fluorescence Complementation

BiFC analysis was performed in transient transfection assays as described previously[43,44]. The BiFC fusion constructs, N-Venus-ICP27-C-Venus; N-Venus-ICP27; ICP27-C-Venus; N-Venus-ICP27-Δ500-512-C-Venus; N-Venus-ICP27-Δ500-512; ICP27-Δ500-512-C-Venus; N-Venus-ICP27-C483, 488S-C-Venus; N-Venus-ICP27-d4-5-C-Venus; N-Venus-ICP27-S18-C-Venus; and N-Venus-ICP27-R→A-C-Venus were transfected into RSF cells grown in glass-bottom culture dishes (MatTek Corporation) using Lipofectamine 2000 by following the manufacturer's protocol. Sixteen hours after transfection, cells were infected with 27LacZ at a MOI of 10. At 4 and 6 hpi, cells were fixed with 4% formaldehyde and viewed directly for Venus fluorescence using a Zeiss Axiovert 200M confocal microscope at a magnification of x10, x20 and x40 with Axiovision software.

Results

ICP27 domains dispensable for dimer formation

Multiple reports have indicated the ICP27 N-terminal domain (NTD) to be hydrophilic and unfolded, while the CTD is highly structured[38–40,45]. Crystallization screens of ICP27 were only successful for the CTD, from residues 241-512, showing this region forms a highly interconnected dimer, while multiple attempts to overexpress and purify the NTD suffered from heavy degradation and difficulties in purification[39,45]. The few NTD

protein constructs that were purified did not crystallize, most likely due to the unstable and flexible nature of the domain[45]. To investigate if the ICP27 NTD plays a role in dimer formation we co-infected HeLa cells with YFP-ICP27, a recombinant virus expressing ICP27 with an N-terminal YFP tag and that replicates as efficiently as KOS[52], and a number of ICP27 mutant viruses bearing deletion or substitution changes spanning the length of the amino portion of the protein, up to residue 173. The addition of the 27 kDa YFP tag allows full length ICP27 to be easily distinguished from the N-terminal mutants. The mutant panel included dLeu (missing the leucine-rich region), d1-2 (lacking the acidic region), d3-4 (deletion of the NLS), d4-5 (deletion of the RGG box), d5-6 (deletion of the R2 region) and WRL (substitution of W105, R107, L108 to alanine). For a full list of viruses used in this section and a detailed description of affected residues and regions see Table 2.1. Following 8 hours of infection, cells were collected, lysed and immunoprecipitations performed with anti-YFP antibody. The Immunoprecipitated samples were separated by SDS-PAGE, transferred to nitrocellulose and western blots probed with anti-ICP27 antibodies. Western blot analysis revealed all of the untagged ICP27 mutants, which are not recognized by anti-YFP antibody, co-precipitated with the YFP tagged protein, indicating all of the N-terminal mutants were able to form dimers and confirming the NTD is dispensable for ICP27 oligomerization (Figure 2.1).

After verifying the domain required for dimerization is localized to the carboxyl end of the protein, we performed a mutational analysis of the ICP27 RRR motif. The arginines 416, 417, 418, 435, 439, 442 form a basic patch on the ICP27 C-terminal surface. Sequence analysis suggests that this basic region is important, with one of the surface exposed arginines (R435) strictly conserved in alphaherpesviruses[39]. Due to its basic properties,

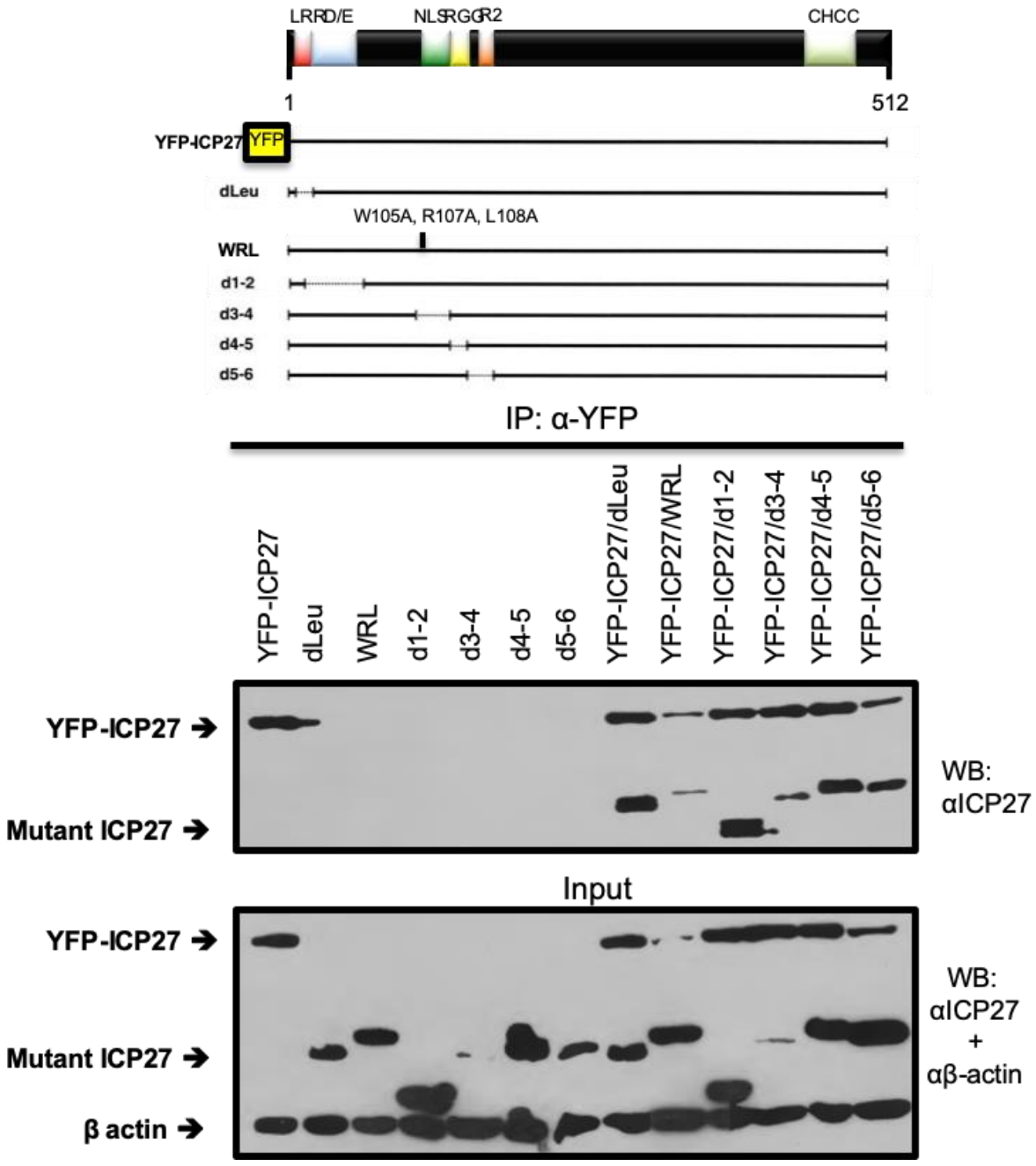
Table 2.1. ICP27 viruses used in this study.

The ICP27 viral mutants used in this study are listed along with the affected regions and the altered residues.

Virus	Residues Altered	Region(s) Affected
YFP-ICP27	N-terminal YFP tag	
27LacZ	Δ1-512 (*lacZ* insertion within the ICP27 locus)	Entire ORF
dLeu	Δ6-19	LRR/NES
d1-2	Δ12-63	Acidic, NES
d3-4	Δ109-138	NLS
d4-5	Δ139-153	RGG box
d5-6	Δ154-173	R2
WRL	W105A, R107A, L108A	NLS

Figure 2.1. ICP27 N-terminal mutants form homo-dimers *in vivo.*

N-terminal YFP-tagged HSV-1 ICP27 (YFP-ICP27) was immunoprecipitated from cellular lysate of HeLa cells co-infected for 8 hours with YFP-ICP27 and the indicated ICP27 N-terminal mutant using anti-YFP antibody (ab290). Western blotting was performed with a combination of ICP27 monoclonal antibodies P1113 and P1119. Blots showing input proteins were also probed with anti-ICP27 antibodies and the cytoskeleton protein Beta actin (ab8229) as a loading control.


LRR
D/E
NLS
RGG
R2
CHCC
1
512
YFP-ICP27
YFP
dLeu
W105A, R107A, L108A
WRL
d1-2
d3-4
d4-5
d5-6
IP: α-YFP
YFP-ICP27
dLeu
WRL
d1-2
d3-4
d4-5
d5-6
YFP-ICP27/dLeu
YFP-ICP27/WRL
YFP-ICP27/d1-2
YFP-ICP27/d3-4
YFP-ICP27/d4-5
YFP-ICP27/d5-6
YFP-ICP27 →
Mutant ICP27 →
WB:
αICP27
Input
YFP-ICP27 →
Mutant ICP27 →
β actin →
WB:
αICP27
+
αβ-actin

this RRR motif may bind to nucleic acids, possibly in tandem with the N-terminal RGG motif. To determine its importance in oligomerization we mutated these arginines along the major dimer interface to alanines (R416, 417, 418, 435, 439, 442A) and cloned it into the GFP background as we have done previously with GFP-ICP27[39]. We then performed co-transfections with plasmids encoding WT ICP27 (WT) and GFP-R416, 418, 419, 435, 439, 442A (GFP-R→A), WT and truncation mutant Stop500 with N-terminal GFP (GFP-Stop500), and WT and GFP-ICP27. For a full list of plasmids used in this section and a detailed description of affected residues and regions see Table 2.2. After 24 hours, cells were infected with ICP27 null mutant virus 27LacZ to stimulate expression of the plasmids, which are under the control of the native ICP27 promoter, as well as to replicate the conditions of infection. At 8 hpi, lysates were immunoprecipitated with anti-GFP antibodies and analyzed by western blot as before. In line with a previously published report[39], WT, which migrates below the tagged proteins, was clearly seen for GFP-ICP27 but not for GFP-Stop500, indicating WT formed dimers with the full length GFP-tagged protein but not for the ICP27 truncation mutant (Figure 2.2). Additionally, WT was also seen following pulldown of GFP-R→A, indicating the arginine to alanine substitutions were not significant enough to disrupt the dimer interaction (Figure 2.2).

Next, we examined two CTD mutants, ICP27-C483, 488S and S18, used in earlier ICP27 studies. BiFC experiments have shown that the C-terminal ZFD is essential for N- to C- formation. When cysteine residues at positions 483 and 488 within the ZFD were replaced with serine (N-Venus-ICP27-C483, 488S-C-Venus), little to no intramolecular interaction was observed (Figure 2.6)[44]. How these substitutions affected dimer formation, however, was not known. Though the cysteine residues are not located within the essential

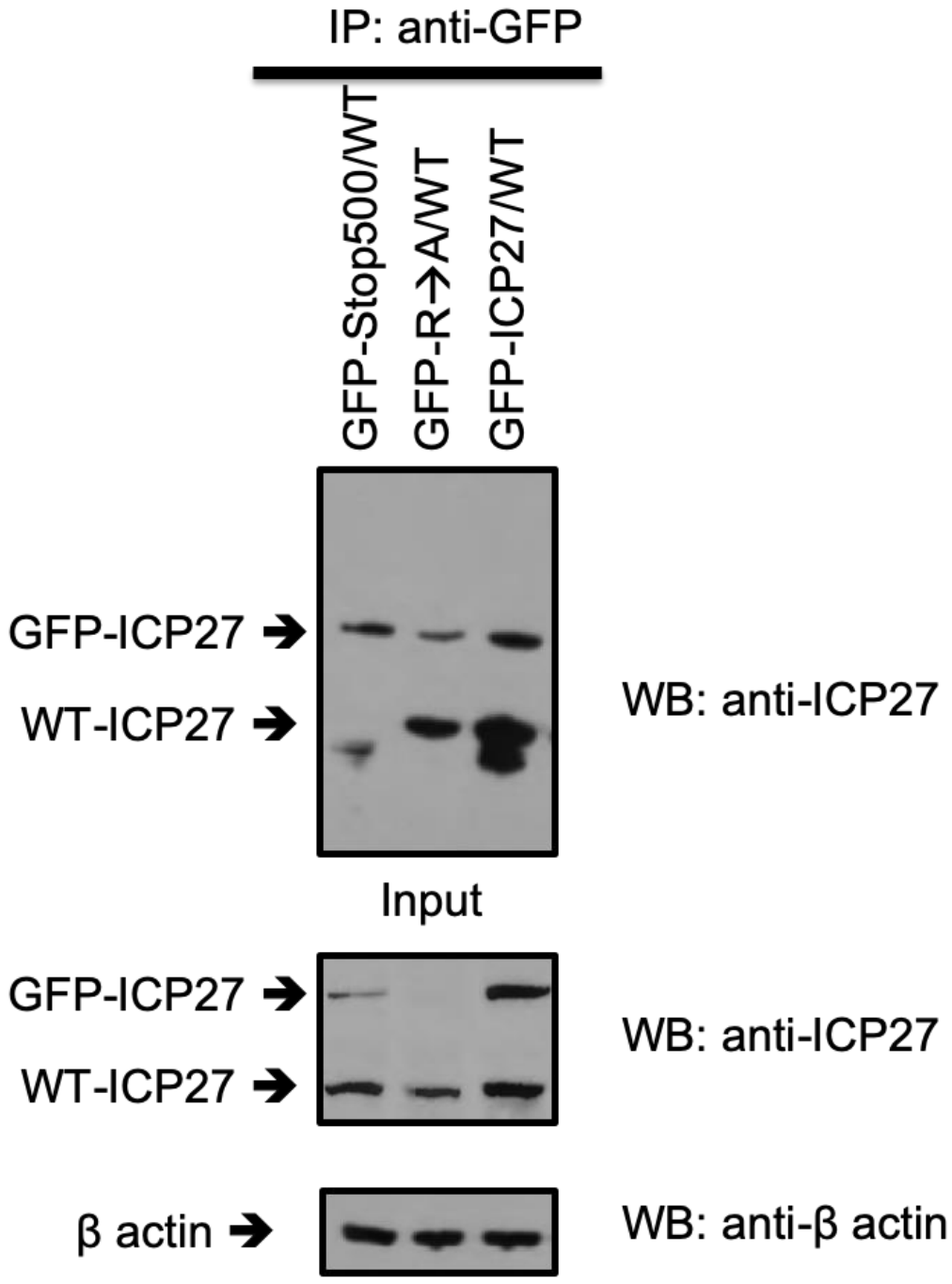
Table 2.2. ICP27 plasmids used in this study.

The ICP27 plasmids used in this study are listed along with the affected regions and the altered residues.

Plasmid	Residues Altered	Region(s) Affected
Wild-type		
GFP-ICP27	N-terminal GFP tag	
S18	SFIPR insert between residue 504 and 505	C-terminal tail
CFP-C483, 488S	N-terminal CFP tag; C483, 488S	ZFD
GFP-Stop500	N-terminal GFP tag; Δ500-512 (Stop codon at residue 500)	C-terminal tail
GFP-ICP27-R416, 417, 418, 435, 439, 442A (GFP-R→A)	N-terminal GFP tag; R416A, R417A, R418A, R435A, R439A, R442A	RRR motif
N-Venus-ICP27-C-Venus	N-terminal Venus 1-154 tag; C-terminal Venus 155-238 tag	
N-Venus-ICP27	N-terminal Venus 1-154 tag	
ICP27-C-Venus	C-terminal Venus 155-238 tag	
N-Venus-ICP27-Δ500-512-C-Venus	N-terminal Venus 1-154 tag; C-terminal Venus 155-238 tag; Δ500-512	C-terminal tail
N-Venus-ICP27-Δ500-512	N-terminal Venus 1-154 tag; Δ500-512	C-terminal tail
ICP27-Δ500-512-C-Venus	C-terminal Venus 155-238 tag; Δ500-512	C-terminal tail
N-Venus-ICP27-C483, 488S-C-Venus	N-terminal Venus 1-154 tag; C-terminal Venus 155-238 tag; C483S, C488S	ZFD
N-Venus-ICP27-d4-5-C-Venus	N-terminal Venus 1-154 tag; C-terminal Venus 155-238 tag; Δ139-153	RGG box
N-Venus-ICP27-S18-C-Venus	N-terminal Venus 1-154 tag; C-terminal Venus 155-238 tag; SFIPR Insert between residue 504 and 505	C-terminal tail
N-Venus-ICP27-R→A-C-Venus	N-terminal Venus 1-154 tag; C-terminal Venus 155-238 tag; R416A, R417A, R418A, R435A, R439A, R442A	C-terminal tail

Figure 2.2. The ICP27 arginine residues within the CTD are not required for dimer formation.

HeLa cells were co-transfected with pSG130 encoding WT ICP27 and pGFP-ICP27 plasmids with the indicated C-terminal mutations. ICP27 residues R416, R417, R418, R435, R439, R442 were substituted with alanine. GFP-tagged ICP27 plasmid Stop500 has a stop codon in all three reading frames at amino acid 500. At 24 hours post transfection, cells were infected with 27-LacZ for 8 hours at a MOI of 5. Cell lysates were immunoprecipitated with anti-GFP antibody (Ab290). Western blots were probed with anti-ICP27 antibodies P113 and P119 and anti-Beta actin as indicated.


IP: anti-GFP
GFP-Stop500/WT
GFP-R→A/WT
GFP-ICP27/WT
GFP-ICP27 →
WT-ICP27 →
WB: anti-ICP27
Input
GFP-ICP27 →
WT-ICP27 →
WB: anti-ICP27
β actin →
WB: anti-β actin

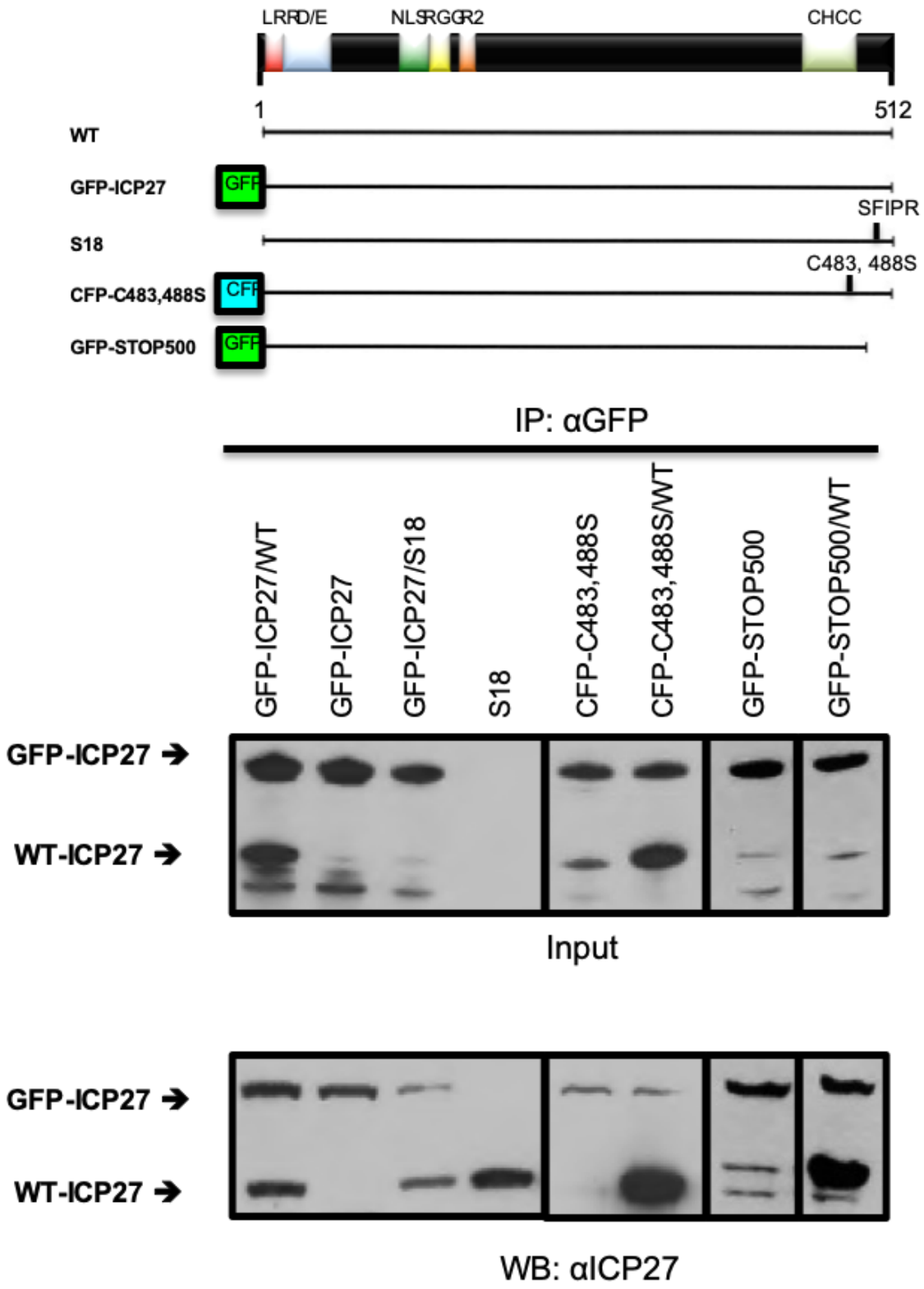
tail motif, the zinc-finger region is believed to serve a structural function and mutation of these residues may perturb ICP27 domain structure. The ICP27 S18 mutant is a tail insertion mutant, with wild-type residues 1-504, followed by SFIPR insert, then wild-type residues FYCNSLF[39]. This mutant has been used in past interaction studies and has been reported to have a repressive effect on self-interaction and transcription of a subset of late genes[51,53,54]. HeLa cells were transfected with plasmids expressing GFP-ICP27 and S18, WT and ICP27-C483, 488S with an N-terminal cyan fluorescent protein tag (CFP-C483, 488S), along with the controls GFP-ICP27 and WT, and GFP-Stop500 and WT. Twenty-four hours after transfection, cells were infected with 27LacZ and immunoprecipitations and western blot analysis was performed in a similar fashion as earlier. Results from the western blot revealed that whereas pulldown of CFP-C483, 488S efficiently co-precipitated WT (Figure 2.3), the ICP27 S18 mutant was not detected following immunoprecipitation of GFP-ICP27 (Figure 2.3). Suggesting that, while the ZFD may play a critical role for the intramolecular self-association, this region is not required for homo-dimerization. On the other hand, the C-terminal tail has been repeatedly shown to be vital for oligomerization, with every introduced mutation tested in this study - deletion of the tail (Stop500), truncation of the last 8 residues (N504), or a 5 amino acid insertion within the tail (S18) – significantly abrogating formation of the dimer (Figures 2.2 and 2.3).

Role of the C-terminal tail in N- to C- terminal self-association

We then sought to determine whether the closed configuration formed upon N- to C-terminal interaction requires ICP27 dimerization. BiFC analysis was performed as before[43,44] with two C-terminal tail mutants, ICP27-Δ500-512 (missing the tail domain) and

Figure 2.3. ICP27 C-terminal mutations disrupting dimer formation.

HeLa cells were co-transfected with plasmids pSG130 encoding WT ICP27 and pGFP-ICP27, pGFP-Stop500 (stop codon at residue 500) or pCFP-C483, 488S (substitution of cysteine 483 and 488 with serine). pS18 (SFIPR insertion between residue 504 and 504) was co-transfected with pGFP-ICP27. At 24 hours post transfection, cells were infected with 27-LacZ for 8 hours. Cell lysates were immunoprecipitated with anti-GFP antibody (Ab290). Western blots were probed with anti-ICP27 antibodies P1113 and P1119.

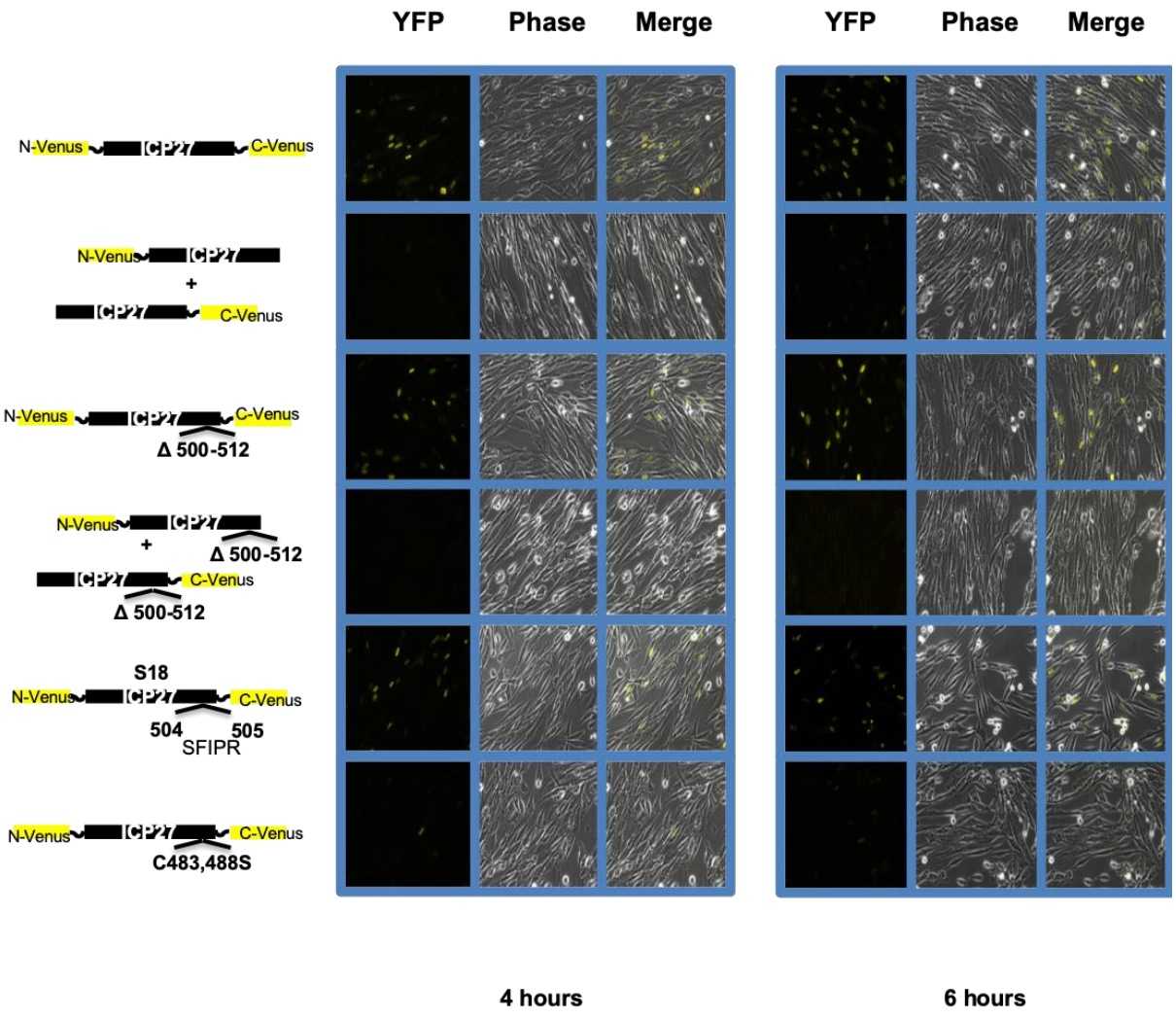

LRRD/E
NLSRGGR2
CHCC
1
512
WT
GFP-ICP27
GFP
SFIPR
S18
C483, 488S
CFP-C483,488S
CFP
GFP-STOP500
GFP
IP: αGFP
GFP-ICP27/WT
GFP-ICP27
GFP-ICP27/S18
S18
CFP-C483,488S
CFP-C483,488S/WT
GFP-STOP500
GFP-STOP500/WT
GFP-ICP27
WT-ICP27
Input
GFP-ICP27
WT-ICP27
WB: αICP27

S18 (Wu and Sandri-Goldin, unpublished). The N-terminal half of YFP Venus (residues 1-154) was cloned onto the N-terminus of ICP27-Δ500-512 (N-Venus-ICP27-Δ500-512-C-Venus) and S18 (N-Venus-ICP27-S18-C-Venus), and the C-terminal half of Venus (residues 155-308) was fused to the C-terminal end. In addition, constructs expressing ICP27-Δ500-512 with only an N-terminal Venus tag (N-Venus-ICP27-Δ500-512) or a C-terminal Venus tag (ICP27-Δ500-512-C-Venus) were also created. If the two tagged ends come into close, nanometer-scale contact, the two halves of Venus recombine, and fluorescence occurs. Plasmid DNA encoding the above constructs along with control plasmids N-Venus-ICP27-C-Venus (positive control), N-Venus-ICP27-C483,488S-C-Venus (negative control), and N-Venus-ICP27 and ICP27-C-Venus (negative control), used in earlier BiFC studies, were transfected into RSF cells. Twenty-four hours after transfection, cells were infected with 27LacZ to activate expression of the plasmids. At 4 and 6 hpi, RSF cells were viewed for fluorescence (Figure 2.4). Venus fluorescence was clearly seen at 4 and 6 hpi for the positive control, N-Venus-ICP27-C-Venus, as well as the two ICP27 tail mutants, N-Venus-ICP27-Δ500-512-C-Venus and N-Venus-ICP27-S18-C-Venus (Figure 2.4), signifying the tail, and by extension the homo-dimer, is not required for the N- to C- configuration. Conversely, fluorescence was not detected when the ZFD was mutated or when cells were co-transfected with plasmids expressing N-Venus on one molecule of ICP27 and C-Venus on another molecule of ICP27 (Figure 2.4).

We next analyzed BiFC with the ICP27-R416, 418, 419, 435, 439, 442A (R→A) mutant (Wu and Sandri-Goldin, unpublished). This R→A mutant was found to dimerize efficiently *in vivo*, yet, how might the alanine substitutions affect the N- to C- interaction was unclear. The N-Venus (1-154) and C-Venus (155-308) tags were fused to either ends

Figure 2.4. BiFC analysis shows the C-terminal tail is not required for N- to C-terminal interactions.

RSF cells were transfected with plasmid DNA from ICP27-Venus fusion constructs N-Venus-ICP27-C-Venus, N-Venus-ICP27-Δ500-512-C-Venus, N-Venus-ICP27-S18-C-Venus, or N-Venus-ICP27-C483, 488S-C-Venus or were co-transfected with N-Venus-ICP27 and ICP27-C-Venus, or N-Venus-ICP27-Δ500-512 and ICP27-Δ500-512-C-Venus as indicated. Twenty-four hours after transfection, cells were infected with 27LacZ at a MOI of 10. Venus fluorescence was visualized directly at 4 and 6 hpi (Wu and Sandri-Goldin, unpublished).


YFP
Phase
Merge
YFP
Phase
Merge
N-Venus
C-Venus
N-Venus
+
C-Venus
N-Venus
C-Venus
Δ 500-512
N-Venus
+
Δ 500-512
C-Venus
Δ 500-512
S18
N-Venus
C-Venus
504
505
SFIPR
N-Venus
C-Venus
C483,488S
4 hours
6 hours

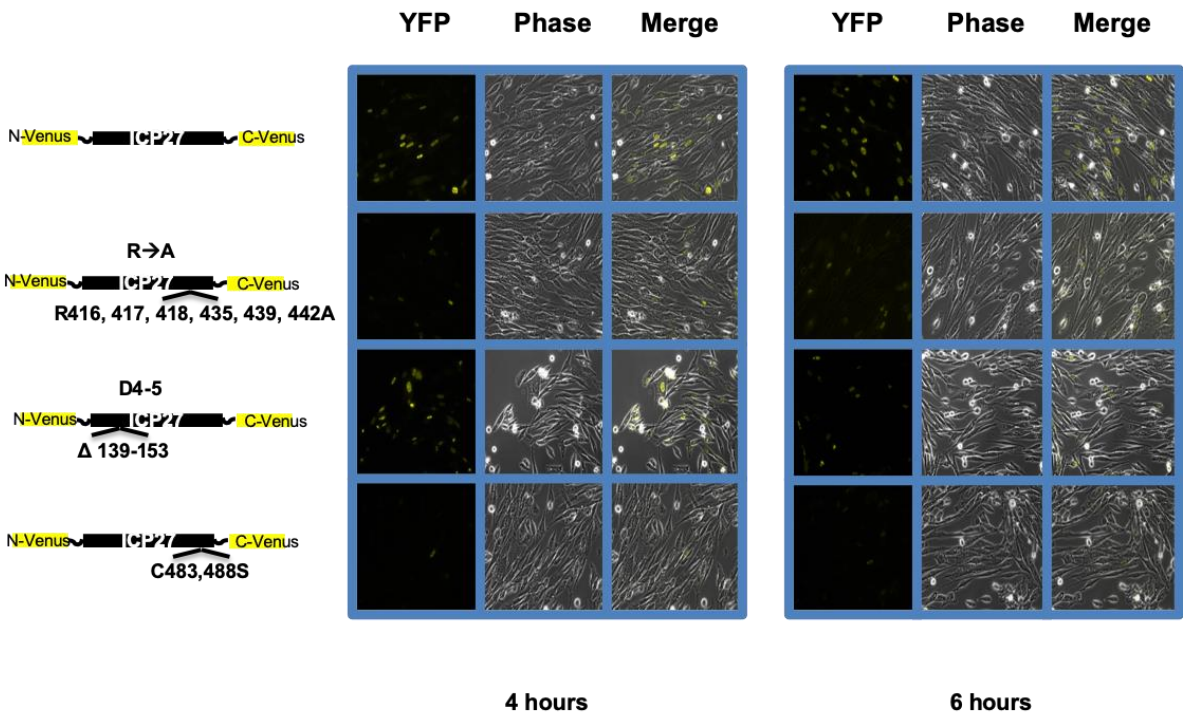
of ICP27-R416, 418, 419, 435, 439, 442A in a similar fashion as N-Venus-ICP27-C-Venus. RSF cells were transfected with plasmids encoding N-Venus-ICP27-R416, 418, 419, 435, 439, 442A-C-Venus, N-Venus-ICP27-C-Venus, N-Venus-ICP27-C483,488S-C-Venus, and N-Venus-ICP27-d4-5-C-Venus (RGG box deletion mutant with N- and C-terminal Venus) and BiFC experiments were carried out with the same protocol as above (Figure 2.5). Here, fluorescence complementation looked to be considerably diminished for the R→A mutant at 4 and 6 hpi compared to N-Venus-ICP27-C-Venus and another mutant with changes outside of the CTD, N-Venus-ICP27-d4-5-C-Venus; and only slightly higher signal intensity than N-Venus-ICP27- C483,488S-C-Venus, which was previously shown to not undergo BiFC at 4 and 8 hpi in RSF cells[44]. This result shows that mutation of the arginine residues that lie between the first cysteine (C400) and the histidine (C479) within the ZFD negatively impacts intramolecular self-association and suggests a possible role for these residues in N- to C- interactions.

Discussion

ICP27's role as a versatile viral adapter required for a successful infection is contingent upon its ability to participate in a wide array of protein-protein and protein-RNA interactions. Although the details for how ICP27 performs its many functions and how these functions are regulated remain elusive, clues can be ascertained through examination of its conformational isomers. Our past studies have demonstrated that, during infection, ICP27 adopts two distinct configurations that influence its binding partners: (1) an intramolecular folded state, where the N- and C-termini come into direct contact[43,44], and

Figure 2.5. BiFC analysis shows arginine-to-alanine mutations within the CTD weaken N- to C-terminal interactions.

RSF cells were transfected with plasmid DNA from ICP27-Venus fusion constructs N-Venus-ICP27-C-Venus, N-Venus-ICP27-R→AC-Venus, N-Venus-ICP27-d4-5-C-Venus, or N-Venus-ICP27-C483, 488S-C-Venus. Twenty-four hours after transfection, cells were infected with 27LacZ at a MOI of 10. Venus fluorescence was visualized directly at 4 and 6 hpi (Wu and Sandri-Goldin, unpublished).


YFP
Phase
Merge
YFP
Phase
Merge
N-Venus
C-Venus
R→A
N-Venus
C-Venus
R416, 417, 418, 435, 439, 442A
D4-5
N-Venus
C-Venus
Δ 139-153
N-Venus
C-Venus
C483,488S
4 hours
6 hours

(2) a homo-dimer, formed upon binding of the IHD in a reverse coffee bean-like orientation[39]. In this chapter, we've evaluated the dimerization and intramolecular binding capabilities for a number of ICP27 mutants. The results show non-overlapping essential regions between the two configurations. The domain necessary for dimerization (the C-terminal tail), is not required for N- to C- binding, while the regions with significant contributions to the intramolecular N- to C- interaction (LRR, RRR motif, ZFD) is expendable for dimer formation. Hence, one possible scenario is that the dimer and N- to C-terminal interactions occur separately and that the protein exists in two forms in vivo: as a monomer that undergoes a reversible intramolecular self-association and as a homo-dimer in which the C-terminal ends form an intertwined dimer with two unstructured N-terminal halves. Due to the differing configurations between the monomeric and dimeric forms, each species may interact with distinct targets and perform specific roles during infection.

Reanalysis of past interaction studies have identified several ICP27 binding partners that may be specific for each conformation. ICP27 binding to TAP/NXF1, RNAP II and Hsc70 require both the N- and C-termini of ICP27, and possibly the monomeric form, with BiFC based FRET analysis showing interaction with TAP/NXF1 dependent on ICP27 N- to C- terminal contacts[9,27,43]. Alternatively, binding between ICP27 and cellular proteins SRSF3, SRSF7, SAP145, PABP, eIF3 and eIF4G require only the C-terminal end of ICP27[17,41]. Pulldown experiments with an ICP27 tail mutant unable to homo-dimerize, S18, reported defects in binding with SRSF3 but did not affect interaction with RNAP II[9,23]. Thus, the dimer form of ICP27 may be required for interaction with some of these cellular partners. As we've now identified several ICP27 mutants specific for each conformational state,

current work in our lab is focused on identifying partners specific to each configuration in order to better elucidate the roles of the monomeric and dimeric forms during the course of infection.

All species of herpesviruses express a multifunctional regulatory protein typified by ICP27[12]. The region of sequence conservation within this family of proteins resides within the globular IHD, which is responsible for dimerization[39]. Every member of this family analyzed so far has been found to oligomerize, with dimerization detected for VZV IE4[55] and KSHV ORF57[56,57]; while UL69 in HCMV has been reported to form stable tetramers composed of a dimer of dimers[58]. A better understanding of the significance and function of the ICP27 dimer could provide new insights into the actions of the ICP27 homologs as well.

CHAPTER 3

Mechanism of SR protein kinase 1 inhibition by the HSV-1 protein ICP27

Introduction

A fundamental step in eukaryotic gene expression is the alternative splicing (AS) of nascent precursor messenger RNA (pre-mRNA) transcripts to remove noncoding introns and adjoin exons in different arrangements to encode alternative protein isoforms. AS is a widespread phenomenon, providing an important mechanism to increase proteomic diversity and introduce additional layers of gene regulation, with current estimates based on transcriptome wide gene analyses suggesting up to 95% of genes within the human genome are alternatively spliced[59,60]. In human cells, splicing is predominantly catalyzed by the major spliceosome, a large multi-molecular complex comprised primarily of 5 small nuclear ribonuclear proteins (snRNPs) - U1, U2, U4, U5, and U6 - and regulated by a variety of *cis*-acting features and *trans*-acting proteins. One of the key protein families involved in splicing regulation are members of the highly conserved serine-arginine rich (SR) group of proteins. SR proteins play significant roles in constitutive and alternative pre-mRNA splicing, mediating spliceosome association with different splice sites, and also function in various post splicing activities such as mRNA export, nonsense-mediated decay and translation[61–63]. In humans, there are 12 SR proteins, designated SR splicing factor (SRSF) 1-12, with SRSF1 (also known as ASF/SF2) being the first identified and prototypical SR protein[61]. SR proteins have a modular structure characterized by two domains: an N-

terminus containing one or more RNA-recognition motifs (RRMs) that determine RNA-binding specificity, and a arginine-serine (RS) rich domain within the C-terminus, which mediate interactions of SR proteins with each other and other spliceosome components[63]. Residues within the RS domain are reversibly phosphorylated and the level of phosphorylation regulates their localization as well as their ability to participate in spliceosome assembly[63]. Two protein kinase families have been reported to specifically phosphorylate SR proteins: the serine-arginine protein kinases (SRPKs), typified by SRPK1, strictly phosphorylate arginine-serine dipeptides, while the CDC2-like kinases (CLKs) can phosphorylate both arginine-serine and serine-proline dipeptides found within the RS domains of SR proteins[64]. All SRPKs are constitutively active kinases and SRPK1 phosphorylation of SR proteins has been shown to promote both their nuclear import as well as their splicing activity[65]. The structure of SRPK1 is comprised of an N-terminal lobe of mostly β strands, harboring the active site, linked by a spacer sequence to a larger C-terminal lobe, containing mostly α helices, that forms a docking groove specific for the RS repeats within SR proteins. Studies investigating SRPK1-SRSF1 interactions report SRPK1 binds to its native substrate with unusually high affinity, with a dissociation constant (K_d) on the order of 50 nanomolar (nM), and rapidly phosphorylates approximately half the sites within its RS domain using a processive mechanism that incorporates sequential C- to N-terminal phosphorylation[65,66].

In mammalian cells, SRPK1 is primarily localized within the cytoplasm. This is due to the presence of a strong cytoplasmic retention signal within the spacer domain as well as its interactions with the molecular chaperones HSP70 and HSP90, which in complex, anchors the kinase to the cytoplasm[67]. SRPK1-mediated phosphorylation of SR proteins in

the cytoplasm is required for their nuclear import as it enhances their interaction with the nuclear import receptor, transportin SR[65]. However, under certain cellular conditions, such as osmotic stress, activation of the epidermal growth factor (EGF) signaling pathway, or during the G2/M phase of the cell cycle, SRPK1 can be recruited into the nucleus where it can participate in the release of SR proteins from nuclear speckles[64,68].

Due to their importance in host gene expression, a number of human viruses have been shown to target various components of the spliceosome and its auxiliary factors in order to shift cellular conditions to favor viral growth[18,69]. RNA-sequencing of HSV-1 infected cells revealed significant changes to host AS[70] with many of these changes mediated by the HSV-1 protein ICP27[24]. ICP27 associates with an array of proteins involved in splicing and its presence has been reported to redistribute snRNPs to a more punctate distribution within the nucleus[71] and inhibit spliceosome formation[72,73]. ICP27 interacts with and relocalizes a number of SR proteins and SRPK1[22,23]. ICP27 binding to SRPK1 recruits this predominantly cytoplasmic kinase to the nucleus, resulting in aberrant phosphorylation of SR proteins and disrupting their ability to participate in spliceosome assembly[23]. Consequently, spliceosome complex formation is stalled at the pre-spliceosomal complex A stage, before the first catalytic step[23]. Through mutational analysis, ICP27 interactions with SR proteins and SRPK1 have been mapped to two distinct regions of the protein. ICP27 binding to SR proteins, snRNP components and other splicing associated factors has been localized to the C-terminal half of ICP27, within the conserved IHD, while ICP27 binding to SRPK1 has been mapped to the RGG box, located within the intrinsically disordered region of the protein[21–23]. The RGG box, from residues 138 to 152, is rich in arginines and glycines and has been reported to be post-translationally modified

by methylation. Arginines 138, 148 and 150 are methylated by the major cellular methyltransferase PRMT1[47,48]; and this modification has been shown to modulate ICP27-SRPK1 interactions[46]. However, the exact details of ICP27-SRPK1 binding and the mechanism of methylation modulation have not been previously described. Here, in collaboration with the lab of Dr. Alexander Golovanov, a structural biologist from the University of Manchester, we performed an in-depth analysis of the ICP27-SRPK1 complex in order to fully ascertain the molecular details of ICP27-SRPK1 interaction and further characterize the mechanism of ICP27-mediated disruption of host RNA processing[74].

Materials and Methods

Protein expression and purification

The proteins glutathione S-transferase (GST)-ICP27$_{103–155}$, GB1-SRPK1, and SRPK1 ΔNS1 were expressed in *Escherichia coli.* Proteins were purified by standard GST or His tag purification methods followed by size exclusion chromatography.

Synthetic peptides

Peptides for the ICP27 RGG box used in crystallization experiments were produced by EZBiolab (Carmel, IN, USA) comprised of ICP27 residues 137 to 152 (sequence ARGGRRGRRRGRGRGG), in two forms: unmodified standard L-peptide and methylated with asymmetric dimethylation on R138, R148, and R150. Peptides used for ITC comprising ICP27 residues 103 to 155 in unmodified form and dimethylated on R138, R148, and R150 as described were produced by Peptide Protein Research (Hampshire, United Kingdom).

Crystallization

To generate complexes of SRPK1 ΔNS1 with RGG box peptide ICP27$_{137–152}$ (ICP27 residues 137 to 152) for crystallography, purified 2 μM SRPK1 ΔNS1 was combined with a 6 μM peptide and incubated for 16 h at 4°C. Complex was purified using a Superdex 75 10/300 column equilibrated with buffer supplemented with 1 μM ICP27$_{137–152}$ peptide. Protein concentrated to 260 μM grew crystals by sitting drop vapor diffusion at 4°C when mixed 1:1 with a reservoir solution of 0.09 M [$NaNO_3$, 0.3 Na_2HPO_4, 0.3 M $(NH_4)_2SO_4$], 0.1 M (sodium HEPES, MOPS [morpholinepropanesulfonic acid]) buffer system (pH 7.5) with 50% vol/vol GOL_P4K mix (Morpheus HT96 C7; Molecular Dimensions). Crystals were flash frozen by plunge freezing in liquid nitrogen.

Data collection, structure determination, model building and refinement

Data were collected from single cryo-frozen crystal SRPK1 ΔNS1-RGG box at beamline i04-1 (Diamond Light Source). Data were indexed, scaled and integrated with Xia2[75]. Phases were determined by molecular replacement using phaser with coordinates from Protein Data Bank code 1WAK[76]. An automated model built against the phased map in Phenix AutoBuild was used as the basis for iterative cycles of rebuilding and refinement in COOT and Phenix.refine[77,78], with validation with MolProbity and PDB_REDO[79,80]. ICP27 RGG box peptides were initially built as polyalanine to ensure unbiased identification of the correct register of residues. Co-crystallization of SRPK1 ΔNS1-RGG box was only successful with the unmodified ICP27$_{137–152}$ peptide, which diffracted to 2.8 Å. Coordinates were deposited in the Protein Data Bank under accession code 6FAD.

Nuclear magnetic resonance

Uniformly ^{13}C, ^{15}N-labeled ICP27$_{103–155}$ backbone amide signal (BMRB accession no. 27341[81]) perturbations were monitored upon addition of unlabeled GB1-SRPK1. Similarly, uniformly ^{15}N-labeled SRSF1 backbone amide signals signal perturbations in ^{1}H-^{15}N HSQC spectra of SRSF1 were monitored upon addition of unlabeled GB1-SRPK1. For IDIS-HSQC-NMR experiments[82], a sample was prepared containing a 1:1 mixture of differentially labeled proteins, specifically 50 μM ^{13}C, ^{15}N-labeled ICP27$_{103–155}$ and 50 μM ^{15}N-labeled SRSF1. IDIS-HSQC spectra were acquired of this 1:1 mixture and then with unlabeled GB1-SRPK1 added to 25 μM and then to 50 μM. Data were processed in Topspin (Bruker) and analyzed in Sparky[83].

Isothermal titration calorimetry

GB1-SRPK1 and ICP27 peptides were placed into Slide-a-Lyzer G2 cassettes (Thermo Fisher) and dialyzed within identical buffer consisting of 20 mM HEPES, 150 mM NaCl, and 1 mM TCEP [Tris(2-carboxyethyl)phosphine hydrochloride] (pH 7.4). Protein concentrations were measured by absorbance at 280 nm using extinction coefficients[84] of 86,580 M^{-1} cm^{-1} and 5,500 M^{-1} cm^{-1} for GB1-SRPK1 and ICP27$_{103–155}$, respectively. Experiments were performed at 25°C on a PEAQ-ITC (Malven instruments), using 10 μM GB1-SRPK1 in the sample cell and 120 μM ICP27 peptide in the syringe. Each experiment consisted of 19 injections: the first at 0.4 μl and the rest at 2 μl. Data were processed and fitted to a single-site binding model with Microcal PEAQ-ITC analysis (V1.00.1262).

***In vivo* studies**

For immunoprecipitation experiments, HeLa cells grown on minimal essential medium containing 10% newborn calf serum were infected with HSV-1 wild-type strain KOS or null mutant 27LacZ as described previously[39]. Transfection of plasmid DNA was performed with Lipofectamine 2000 (Invitrogen). Cells were infected with 27LacZ 24 hours after transfection to stimulate expression of the native ICP27 promoter by the virion tegument protein VP16 as previously described[39,43]. Cells were harvested 8 hours later, and immunoprecipitation was performed using GFP antibody Ab290 (Abcam) or ICP27 monoclonal antibody P1119 (Virusys) as described previously[39]. Western blots were probed with anti-ICP27 antibody P1119, anti-GFP antibody JL-8 (TaKaRa Bio), or anti-β-actin antibody.

In immunofluorescence experiments, rabbit skin fibroblasts (RSF) were transfected and/or infected as indicated in the legend to Fig. 5. At 8 hpi, cells were fixed in 3.7% formaldehyde, and immunofluorescent staining was performed as previously described[41] with ICP27 antibody P1119 or ICP4 P1114 (Virusys). GFP fluorescence was viewed directly. Cells were viewed with a Zeiss Axiovert S100 microscope.

For *in vivo* phosphorylation analysis, HeLa cells were transfected with GFP-SRPK1 and ICP27 RGG box mutants, as indicated in Fig. 5D, and 24 hours after transfection, cells were infected with 27LacZ for 8 hours. To monitor SR protein phosphorylation, 5% of each cellular lysate was resolved by SDS-PAGE. Western blot analysis was performed with

antiphosphoepitope SR antibody MABE50 (Millipore), anti-SF2 (SRSF1) antibody ab38017 (Abcam), anti-ICP27 antibody P1119, and anti-lamin B1 antibody.

Data availability

Co-crystallization of SRPK1 ΔNS1-RGG box coordinates have been deposited in the Protein Data Bank, under accession code 6FAD.

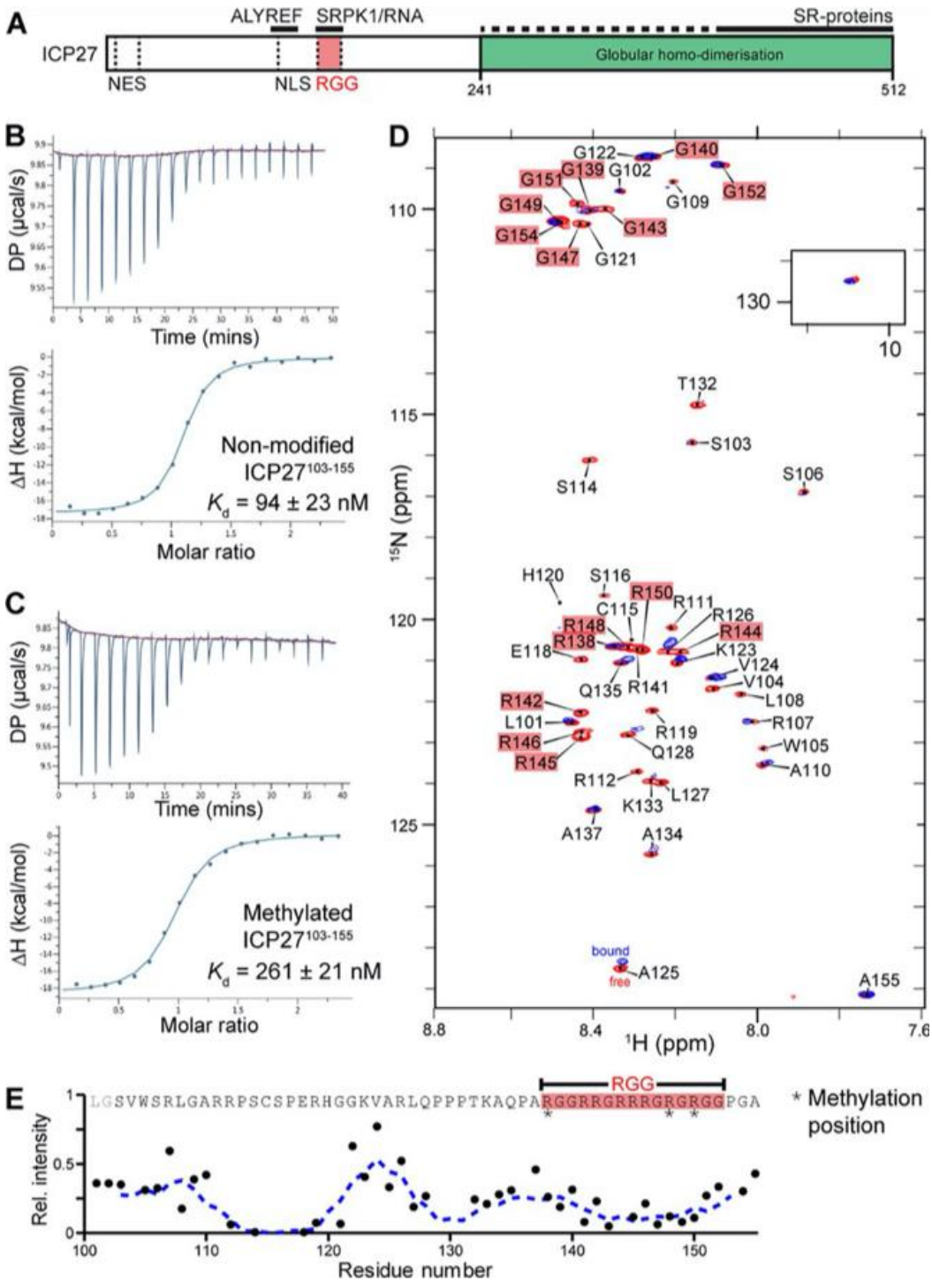
Results

Mapping and quantification of SRPK1 binding to ICP27 peptide fragments containing the RGG box motif

Previous *in vitro* binding experiments have indicated ICP27 binds SRPK1 via its N-terminal RGG box, with binding modulated by methylation of arginine residues 138, 148 and 150. To verify the isolated RGG motif is sufficient for SRPK1 binding and to measure the binding affinity between the RGG box and SRPK1, an isothermal titration calorimeter was used to measure the interaction between a synthetic ICP27 peptide ($ICP27_{103\text{-}155}$), comprised of the RGG box residues plus the preceding NLS and ALYREF-binding regions (Figure 3.1A), and purified full-length SRPK1 expressing an N-terminal GB1 solubility tag (GB1-SRPK1). Furthermore, the binding affinity of a methylated ICP27 peptide (methylated at arginines 138, 148 and 150) and GB1-SRPK1 was also measured. Results from the isothermal titration calorimetry (ITC) measurements revealed $ICP27_{103\text{-}155}$ binds GB1-SRPK1 with high affinity, with a K_d of 94 ± 23 nM (Figure 3.1B). However, methylation of the RGG box negatively affected $ICP27_{103\text{-}155}$-GB1-SRPK1 binding, weakening the

Figure 3.1. SRPK1 binds to the RGG box of ICP27 *in vitro.*

(A) Schematic of subdomains and interaction sites within HSV-1 ICP27. (B) Affinity quantification of SRPK1 interaction with unmodified ICP27$_{103–155}$. (C) As in panel B but with methylated ICP27$_{103–155}$. (D) ^{1}H-^{15}N correlation spectra of [^{15}N]ICP27$_{103–155}$ in the absence (red) and presence (blue) of a stoichiometric equivalent of SRPK1. (E) Binding site mapped by analysis of relative signal intensity change upon binding, IB/IF. The decrease in signal intensity of a protein occurs upon interaction with SPRK1.


A
ALYREF
SRPK1/RNA
SR-proteins
ICP27
Globular homo-dimerisation
NES
NLS
RGG
241
512
B
DP (µcal/s)
Time (mins)
ΔH (kcal/mol)
Non-modified
ICP27103-155
Kd = 94 ± 23 nM
Molar ratio
C
DP (µcal/s)
Time (mins)
ΔH (kcal/mol)
Methylated
ICP27103-155
Kd = 261 ± 21 nM
Molar ratio
D
15N (ppm)
1H (ppm)
bound
free
E
RGG
* Methylation position
Rel. intensity
Residue number

interaction approximately 3-fold, to a K_d of 262 ± 21 nM (Figure 3.1C). Next, we mapped the ICP27 residues directly responsible for SRPK1 contact by performing nuclear magnetic resonance (NMR) signal perturbation mapping with a purified ^{15}N-labeled ICP27$_{103-155}$ peptide. Using the previously established backbone assignment[81], we compared the ^{1}H-^{15}N correlation spectra in the presence and absence of unlabeled GB1-SRPK1 (Figure 3.1D). ICP27 residues that interact with SRPK1 exhibit signal broadening, resulting in lower signal intensity. Upon addition of GB1-SRPK1, we saw a significant decrease in signal intensity for ^{15}N-labeled ICP27$_{103-155}$, particularly for residues 112-121 and 138-152, the latter of which contains the RGG motif (Figure 3.1E). Taken together, these data confirmed that ICP27 binds SRPK1 with high affinity through its RGG motif with additional contributions from upstream residues within the NLS region.

NMR directly confirms competition between SRSF1 and ICP27$_{103–155}$ for SRPK1 binding

Our ITC results suggests a similar binding affinity between ICP27's RGG region and the RS domain of SRSF1 with SRPK1; with a K_d of 94 ± 23 nM versus 50 ± 25 nM for SRSF1, previously determined through *in vitro* phosphorylation assays. To identify whether these two proteins bound SRPK1 in a competitive manner and to determine which had the higher affinity we mixed all three proteins (in differentially labeled forms) in various stoichiometric amounts and performed isotopically discriminated (IDIS)-NMR. Initial ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) and transverse relaxation-optimized spectroscopy (TROSY) experiments established fingerprint spectra of uniformly ^{15}N-labeled full-length SRSF1 in the presence and absence of unlabeled GB1-SRPK1. These

spectral patterns allowed us to identify the SRSF1-SRPK1 complex in solution. The high contour view of the SRSF1 spectra showed sharp, poorly dispersed amide signals from the RS residues within SRSF1 (Figure 3.2A), while the low contour view revealed lower-intensity, broad amide signals from the globular RRM domains as well as the sharp signals from the RS region (Figure 3.2B). Upon introduction of GB1-SRPK1, the RS repeat signals were broadened, significantly weakening their signal intensity. For ICP27$_{103\text{-}155}$, the previous mapping experiment (Figure 3.1D) provided the reference spectral patterns that allowed us to recognize the presence of the ICP27 RGG-SRPK1 complex. In the IDIS-NMR experiments, baseline spectra and spectral intensities were first recorded for [^{13}C,^{15}N]ICP27$_{103–155}$ and [^{15}N]SRSF1 in isolation (Figure 3.3Ai and 3.3Bi, 3.3C and 3.3D) and upon addition of half a stoichiometric amount(25μM) of unlabeled GB1-SRPK1 or an equivalent amount (50 μM) of [^{15}N]SRSF1 (for [^{13}C,^{15}N]ICP27$_{103–155}$) or [^{13}C,^{15}N]ICP27$_{103–155}$ (for [^{15}N]SRSF1). Addition of SRPK1 resulted in substantial signal broadening for ICP27$_{103–155}$ and SRSF1 (Figure 3.3Aii and 3.3Bii, 3.3C and 3.23), confirming SRPK1 binding between the two proteins. On the other hand, spectra from the 1:1 mixture of ICP27$_{103\text{-}155}$:SRSF1 revealed only minor differences in signal patterns compared to baseline (Figure 3.3Aiii and 3.3Biii, 3.3C and 3.3D), indicating ICP27 residues 103-155 do not interact with SRSF1. Next, we measured the NMR spectra and signal intensities of the 1:1 mixture of ICP27$_{103\text{-}155}$:SRSF1 following addition of a substoichiometric amount of non-labelled GB1-SRPK1. Introduction of 25 μM GB1-SRPK1 resulted in marked spectral changes and a strong decrease in signal intensity within the SRSF1 RS residues, whereas in ICP27, the RGG box signals were relatively undisturbed (Figure 3.3Aiv and 3.3Biv, 3.3C and 3.3D), demonstrating preferential binding of SRPK1 to SRSF1. However, upon increasing the

Figure 3.2. TROSY ^{1}H-^{15}N NMR correlation spectrum of uniformly [^{15}N]-labelled full-length SRSF1.

The same spectrum is shown with contour levels adjusted to illustrate differential peak heights. (A) High contour level only reveals sharp poorly dispersed signals from intrinsically disordered termini including the RS-repeats labelled with black dashed boxes. (B) Low contour view where broad and dispersed signals from globular RRM domains can be observed in addition to sharp signals.

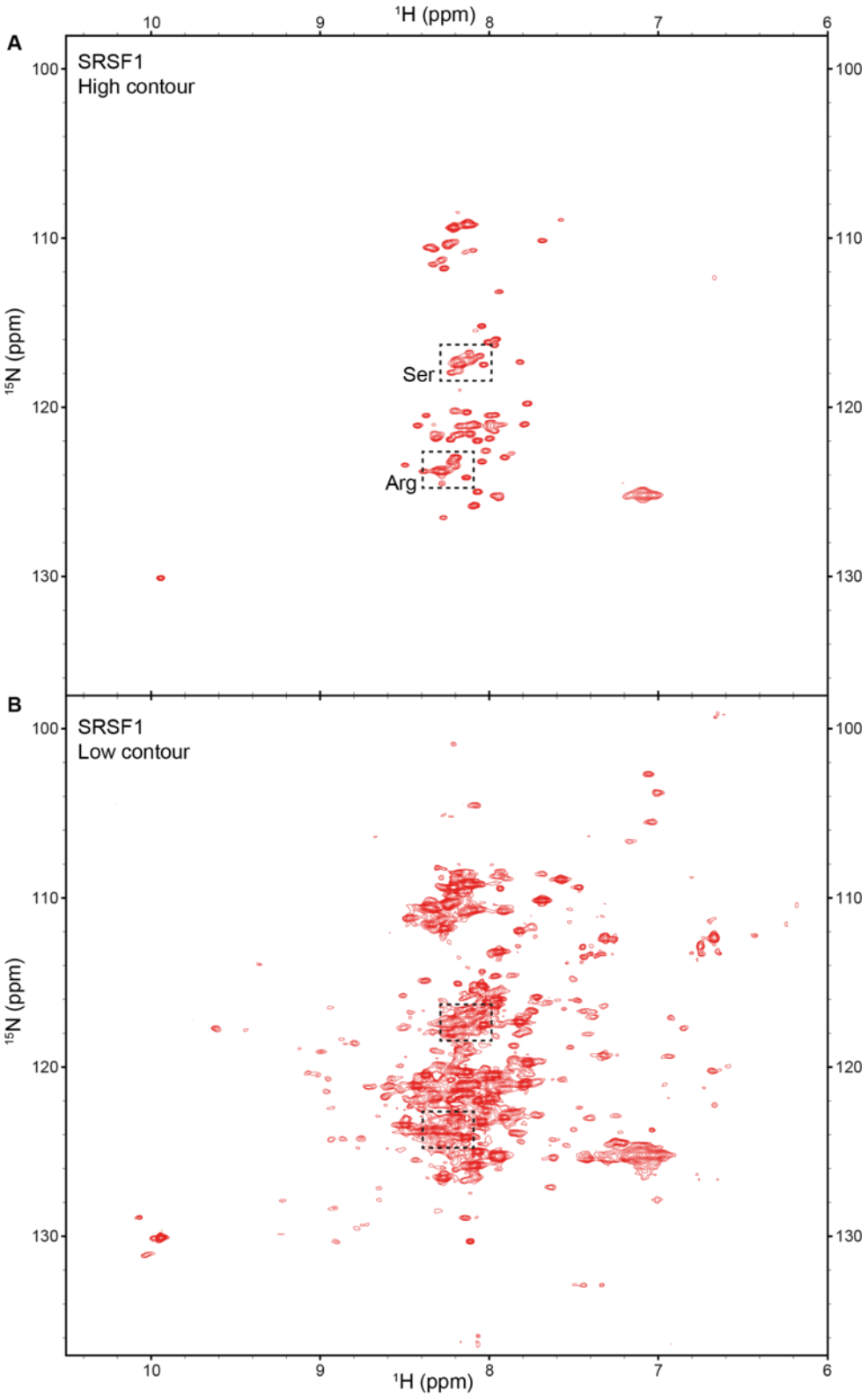
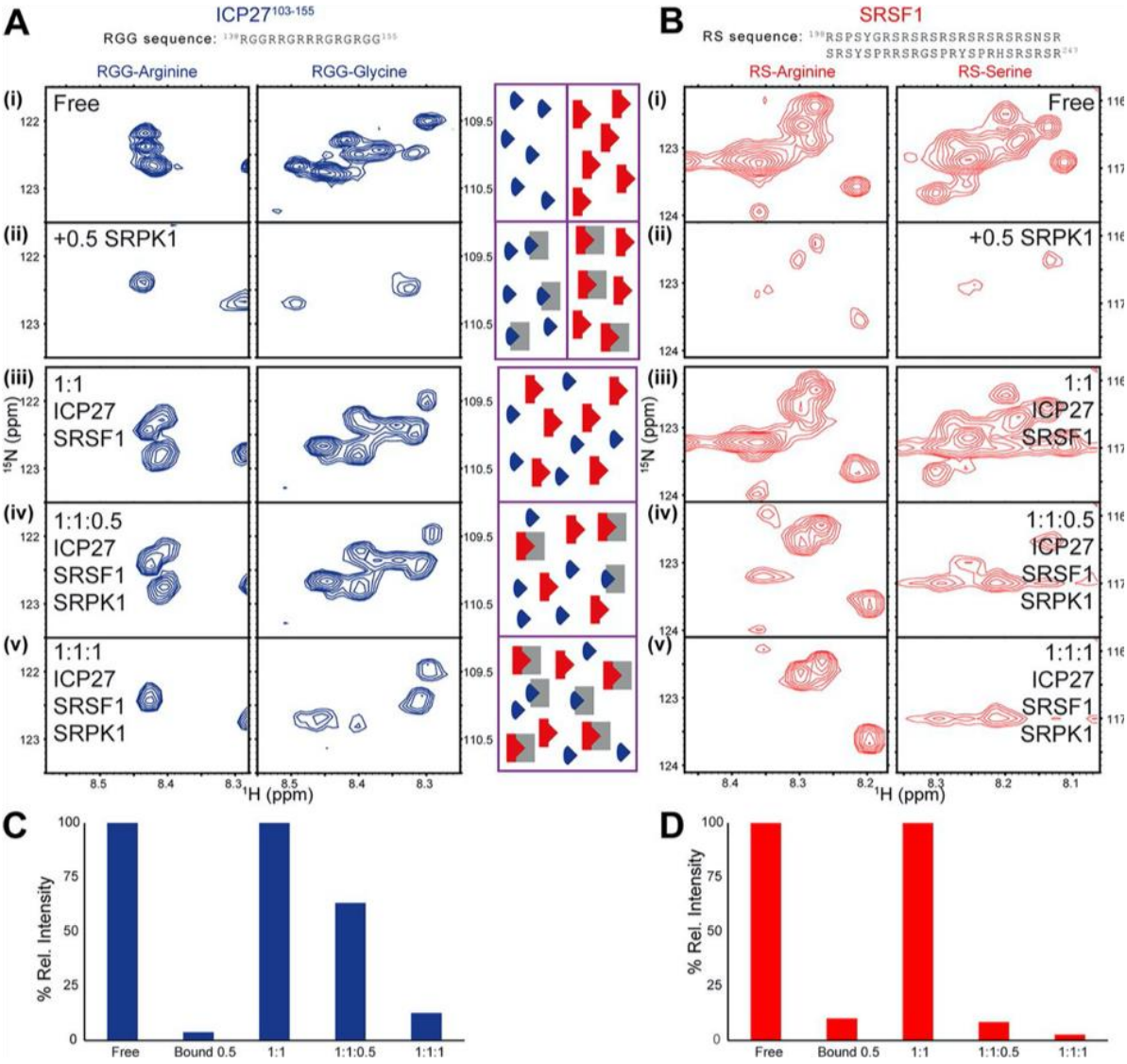

A
SRSF1
High contour
Ser
Arg
B
SRSF1
Low contour
^{1}H (ppm)
^{15}N (ppm)
10
9
8
7
6
100
110
120
130

Figure 3.3. IDIS-NMR experiments indicate competition for SRPK1 binding between SRSF1 and the RGG box region of ICP27.

Reporter signals for SRPK1 binding are shown for (A) RGG box within [$_{13}$C, $_{15}$N]ICP27$_{103–155}$ and (B) RS region from full-length [$_{15}$N]SRSF1. (i) ICP27 or SRSF1 in isolation, (ii) ICP27 or SRSF1 with half a stoichiometric equivalent of unlabeled GB1-SRPK1 added, (iii) 1:1 mixture of ICP27 and SRSF1, (iv) 1:1:0.5 mixture of ICP27, SRSF1 and GB1-SRPK1, (v) 1:1:1 mixture of ICP27, SRSF1, and GB1-SRPK1. The normalized relative intensities of reporter signals I_{rel} measured from the spectra shown in panels A and B are plotted in panel C for ICP27 and panel D for SRSF1. The decrease in signal intensity of a protein occurs upon interaction with SPRK1, and comparison of signals from ICP27 and SFSR1 within the same sample tube therefore provides an indication of binding preference. The central scheme illustrates the binding scenarios for each sample, with GB1-SRPK1, ICP27$_{103–155}$, and SRSF1 molecules colored gray, blue, and red, respectively.


A
ICP27103-155
RGG sequence: 139RGGRRGRRRGRGRGG155
RGG-Arginine
RGG-Glycine
(i) Free
(ii) +0.5 SRPK1
(iii) 1:1 ICP27 SRSF1
(iv) 1:1:0.5 ICP27 SRSF1 SRPK1
(v) 1:1:1 ICP27 SRSF1 SRPK1
15N (ppm)
1H (ppm)
B
SRSF1
RS sequence: 199RSPSYGRSRSRSRSRSRSRSRSNSR SRSYSPRRSRGSPRYSPRHSRSRSR247
RS-Arginine
RS-Serine
(i) Free
(ii) +0.5 SRPK1
(iii) 1:1 ICP27 SRSF1
(iv) 1:1:0.5 ICP27 SRSF1 SRPK1
(v) 1:1:1 ICP27 SRSF1 SRPK1
15N (ppm)
1H (ppm)
C
% Rel. Intensity
Free
Bound 0.5
1:1
1:1:0.5
1:1:1
D
% Rel. Intensity
Free
Bound 0.5
1:1
1:1:0.5
1:1:1

concentration of GB1-SRPK1 (final concentration of 50 μM) to an equivalent stoichiometric amount of 1:1:1 for all three components within the ternary mixture, we observed significant signal broadening for the ICP27$_{103\text{-}155}$ peptide, as well as some additional broadening for SRSF1 (Figure 3.3Av and 3.3Bv, 3.3C and 3.3D). These results indicate that, although SRPK1 has a slightly higher preference for the RS regions of SRSF1 over a short ICP27$_{103–155}$ peptide, the latter is still competitive for SRPK1 binding.

The crystal structure of ICP27 RGG box peptide bound to the docking groove of SRPK1

Following the competition experiments, we then sought to determine the structure of the ICP27 RGG box-SRPK1 complex by X-ray crystallography. Crystal trials were attempted with the SRPK1 ΔNS1 protein construct (used in earlier crystallization studies) containing the two kinase domains (Figure 3.4) and methylated and unmethylated ICP27$_{137\text{-}155}$ RGG box peptides. Successful co-crystallization was only achieved with the unmethylated ICP27$_{137\text{-}155}$-SRPK1 complex, which diffracted to 2.8 Å. The crystal structure identified a 7-residue palindrome sequence - RGRRRGR - located within the center of the RGG motif, from positions 142 to 148, forms the crux of the interactions with SRPK1 (Figure 3.4). Comparison of the structure with the SRPK1-SRSF1 crystal structure reveal a highly similar pattern of recognition (Figure 3.5). Notably, ICP27 arginines R142, R144, R146 and R148 and the arginine residues R204, R206, R208 and R210 within the RS domain of SRSF1 form almost identical intermolecular contacts with the SRPK1 docking groove (Figure 3.5). Thus, the structural data obtained in this section visually illustrates

Figure 3.4. Structure of SRPK1 with RGG box peptide ICP27$_{137–152}$ bound in the substrate docking groove.

(A) Schematic of SRPK1 domains. (B) Crystal structure of the SRPK1 in complex with ICP27$_{137–152}$. SRPK1 is colored as panel A, ICP27 is yellow. (C) As in panel A but view rotated 90°. (D) $2F_o - F_c$ electron density map of RGG peptide colored blue, contoured to 1σ orientated as in panel B.

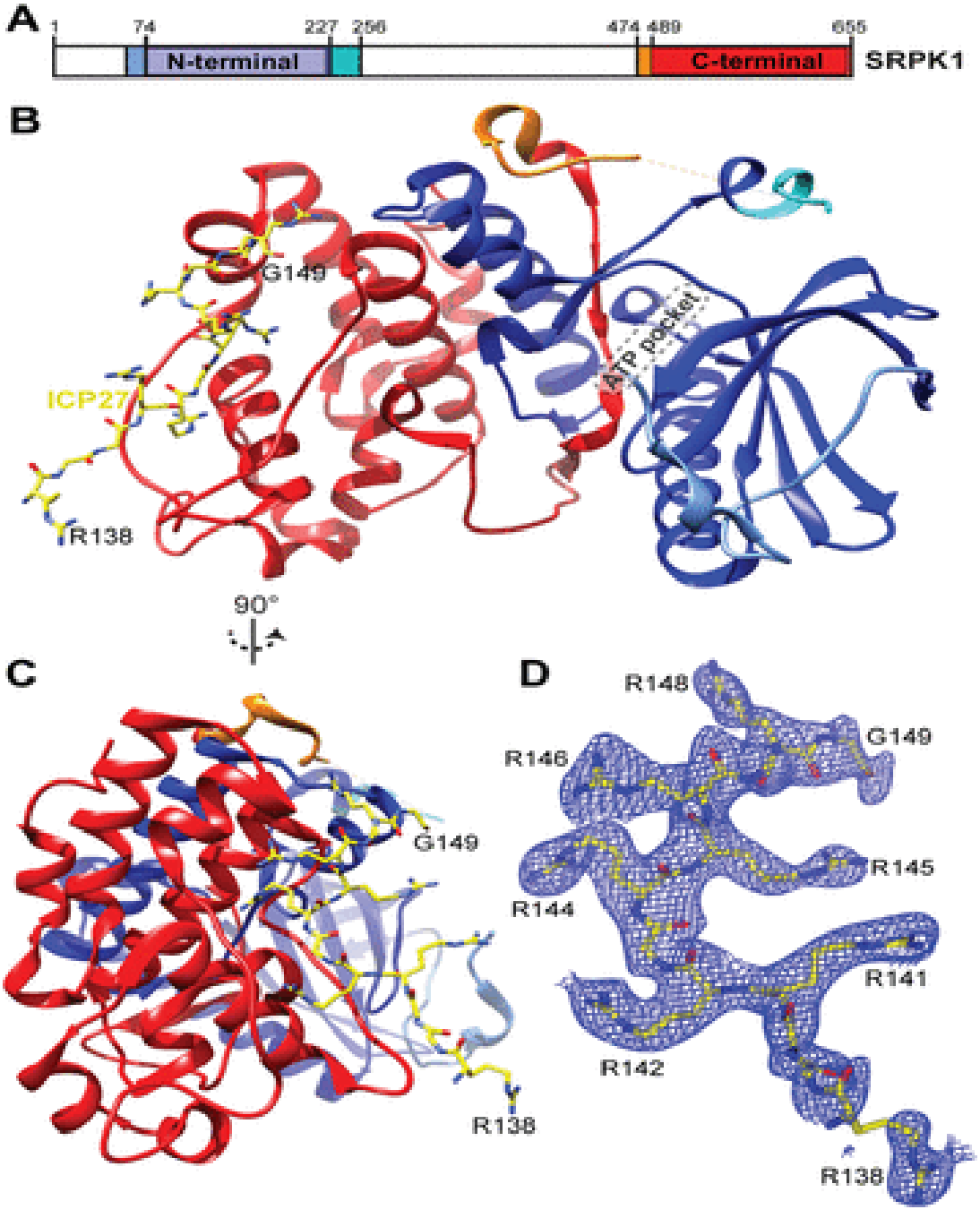
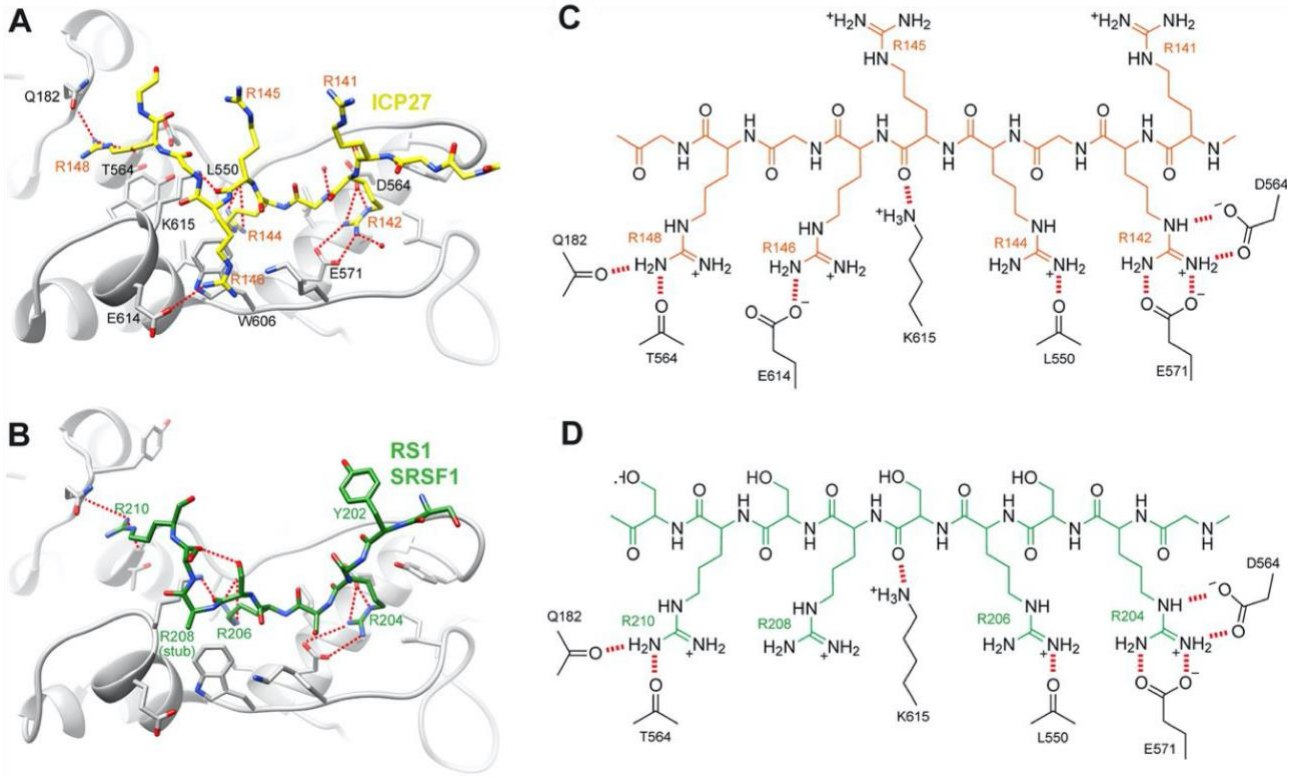

A
1
74
227
256
474
489
655
N-terminal
C-terminal
SRPK1
B
G149
ICP27
R138
ATP pocket
90°
C
G149
R138
D
R148
G149
R146
R145
R144
R141
R142
R138

Figure 3.5. Comparison of SRPK1 docking groove interactions.

(A) Detail of the RGG box peptide bound within the SRPK1 docking groove, plus comparative view of SRPK1 with (B) SRSF1 RS1 region. SRPK1 is shaded gray, and bound peptides are colored. Intermolecular hydrogen bonds or salt bridges are represented by red dashes. Chemical schematics of the interactions shown in panels A and B are shown in panels C and D, respectively.


A
Q182
R148
T564
L550
R145
R141
ICP27
D564
K615
R144
R142
R148
E571
E614
W606
B
R210
RS1
SRSF1
Y202
R208
(stub)
R206
R204
C
R145
R141
Q182
R148
R146
R144
R142
D564
T564
E614
K615
L550
E571
D
Q182
R210
R208
R206
R204
D564
T564
K615
L550
E571

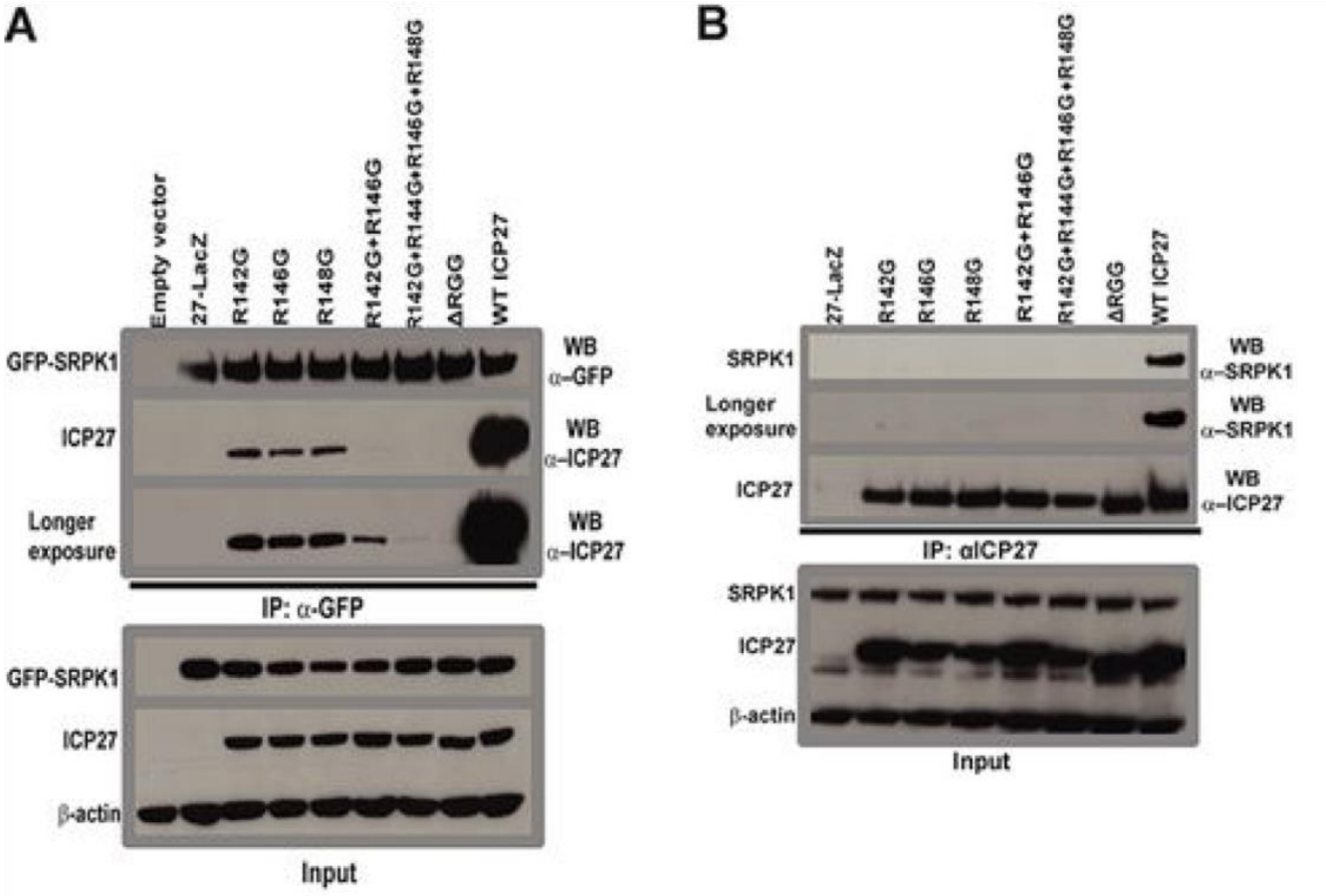
the similarities between ICP27 RGG box and SRSF1 binding to SRPK1 and corroborates their competitive interactions observed during the NMR experiments.

ICP27 RGG box mutants do not interact with or change cellular localization of SRPK1 *in vivo*

We next performed a series of experiments to further characterize the function of ICP27-SRPK1 binding *in vivo.* First, we constructed a series of ICP27 mutants targeting the arginine residues that form the majority of SRPK1 contacts. ICP27 residues R142, R144, R146 and R148 were mutated to glycine as single (R142G, R146G or R148G), double (R142G R146G) or quadruple (R142G R144G R146G R148G) substitutions. Plasmids encoding either the arginine mutants, WT ICP27 or an RGG box deletion mutant were co-transfected with a plasmid expressing GFP-SRPK1 and subsequently infected with ICP27 null mutant virus 27LacZ to stimulate expression of the plasmids. At 8 hpi, lysates were immunoprecipitated with anti-GFP antibody, separated by SDS-PAGE, transferred to nitrocellulose and analyzed by western blot with anti-ICP27 antibodies. As expected, the pulldown experiment showed substantial binding between WT ICP27 and GFP-SRPK1, as immunoprecipitations with anti-GFP antibody efficiently co-immunoprecipitated untagged WT ICP27 (Figure 3.6A). However, interaction with GFP-SRPK1 was considerably reduced for ICP27 single (R142G, R146G or R148G) and double (R142G R146G) arginine substitution mutants and was essentially abolished for the quadruple (R142G R144G R146G R148G) mutant and the RGG box deletion mutant (ΔRGG) (Figure 3.6A). ICP27 interaction with endogenous SRPK1 was also analyzed. Cells were transfected with WT ICP27, ΔRGG or ICP27 mutant plasmids and infected with 27LacZ as described earlier.

Figure 3.6. ICP27 RGG box mutants do not interact with SRPK1 *in vivo*.

HeLa cells cotransfected with GFP-SRPK1 and WT ICP27 or RGG box mutants were infected for 8 hours with 27LacZ. Immunoprecipitation was performed using anti-GFP antibody. Western blots were probed with anti-GFP or anti-ICP27 antibody and anti-β-actin as a loading control. (B) HeLa cells transfected with WT and ICP27 RGG box mutants were infected with 27LacZ for 8 hours. Immunoprecipitation was performed with anti-ICP27 antibody. Western blots were probed with anti-SRPK1 or anti-ICP27 antibody as indicated and anti-β-actin as a loading control.


A
Empty vector
27-LacZ
R142G
R146G
R148G
R142G+R146G
R142G+R144G+R146G+R148G
ΔRGG
WT ICP27
GFP-SRPK1
ICP27
Longer exposure
WB α-GFP
WB α-ICP27
WB α-ICP27
IP: α-GFP
GFP-SRPK1
ICP27
β-actin
Input
B
27-LacZ
R142G
R146G
R148G
R142G+R146G
R142G+R144G+R146G+R148G
ΔRGG
WT ICP27
SRPK1
Longer exposure
ICP27
WB α-SRPK1
WB α-SRPK1
WB α-ICP27
IP: αICP27
SRPK1
ICP27
β-actin
Input

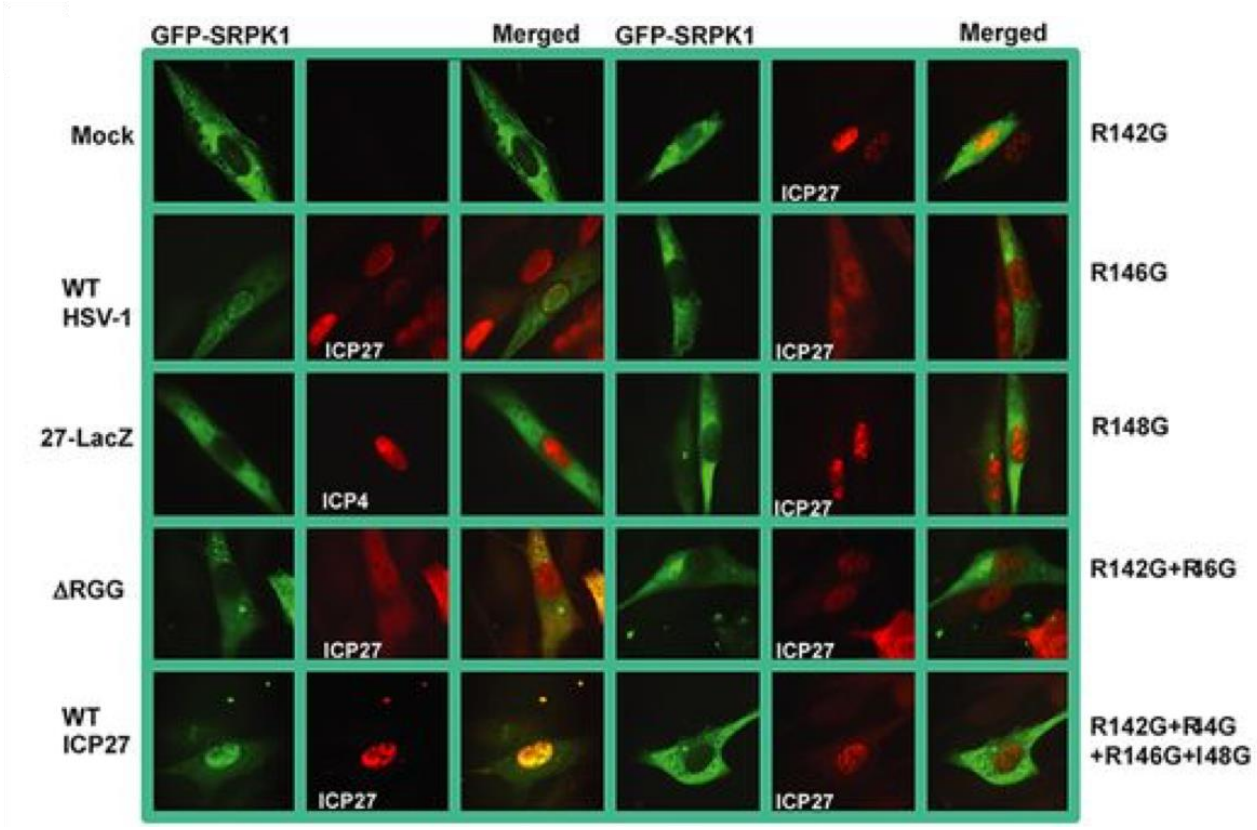
Immunoprecipitations were carried out with anti-ICP27 antibody and the blots were probed with antibody to SRPK1. In this instance, endogenous SRPK1 was only detected with WT ICP27 and did not co-ip with any of the mutants (Figure 3.6B).

We've previously shown that during HSV-1 infection, ICP27 relocalizes SRPK1 from the cytoplasm to the nucleus and this relocalization requires an intact RGG motif[23,46]. To determine if the ICP27 arginine-to-glycine mutants are able to redistribute SRPK1 to the nucleus, we co-transfected cells with GFP-SRPK1 and WT or ICP27 mutant plasmids, infected with 27LacZ for 6 hours, then fixed and stained the cells for immunofluorescent (IF) analysis. Further, cells were also transfected with GFP-SRPK1 and mock infected or infected with WT HSV-1, 27LacZ, or ΔRGG mutant virus as controls. In agreement with our earlier reports, IF analysis of mock, ΔRGG and 27LacZ infected cells showed SRPK1 to be predominantly cytoplasmic, whereas in cells infected with WT HSV-1 or transfected with WT ICP27 plasmid, SRPK1 accumulated in the nucleus (Figure 3.7). For cells expressing the RGG box substitution mutants (R142G, R146G, R142G R146G, or R142G R144G R146G R148G), SRPK1 was, again, largely confined to the cytoplasm; indicating the reduced interactions between these mutants and SRPK1, illustrated by our earlier co-ip experiments, prevented ICP27 from relocating the cellular kinase to the nucleus (Figure 3.7).

ICP27 has also been demonstrated to decrease the phosphorylation state of SR proteins by interfering with SRPK1's kinase activity[23]. To determine if the ICP27 RGG box mutants could alter SR protein phosphorylation *in vivo*, cells were co-transfected with GFP-SRPK1 and WT ICP27 or ΔRGG, R142G R144G R146G R148G, and R142G R146G mutants and infected with 27LacZ. 8 hours after infection, western blots were probed with an anti-

Figure 3.7. ICP27 RGG box mutants do not relocalize SRPK1 into the nucleus.

RSF cells were transfected with GFP-SRPK1 and 16 hours later were mock infected, infected with WT HSV-1 or 27LacZ, or were cotransfected with GFP-SRPK1 and WT ICP27 or ICP27 RGG box mutants and infected 16 hours later with 27LacZ. Cells were fixed 6 hours later and immunofluorescent staining was performed with anti-ICP27 antibody. ICP4 antibody was used to stain the 27LacZ sample to mark the nucleus. GFP fluorescence was visualized directly. The panels shown are representative of ~80% of cells visualized.


GFP-SRPK1
Merged
GFP-SRPK1
Merged
Mock
WT HSV-1
27-LacZ
ΔRGG
WT ICP27
R142G
R146G
R148G
R142G+R146G
R142G+R146G +R146G+I48G
ICP27
ICP4

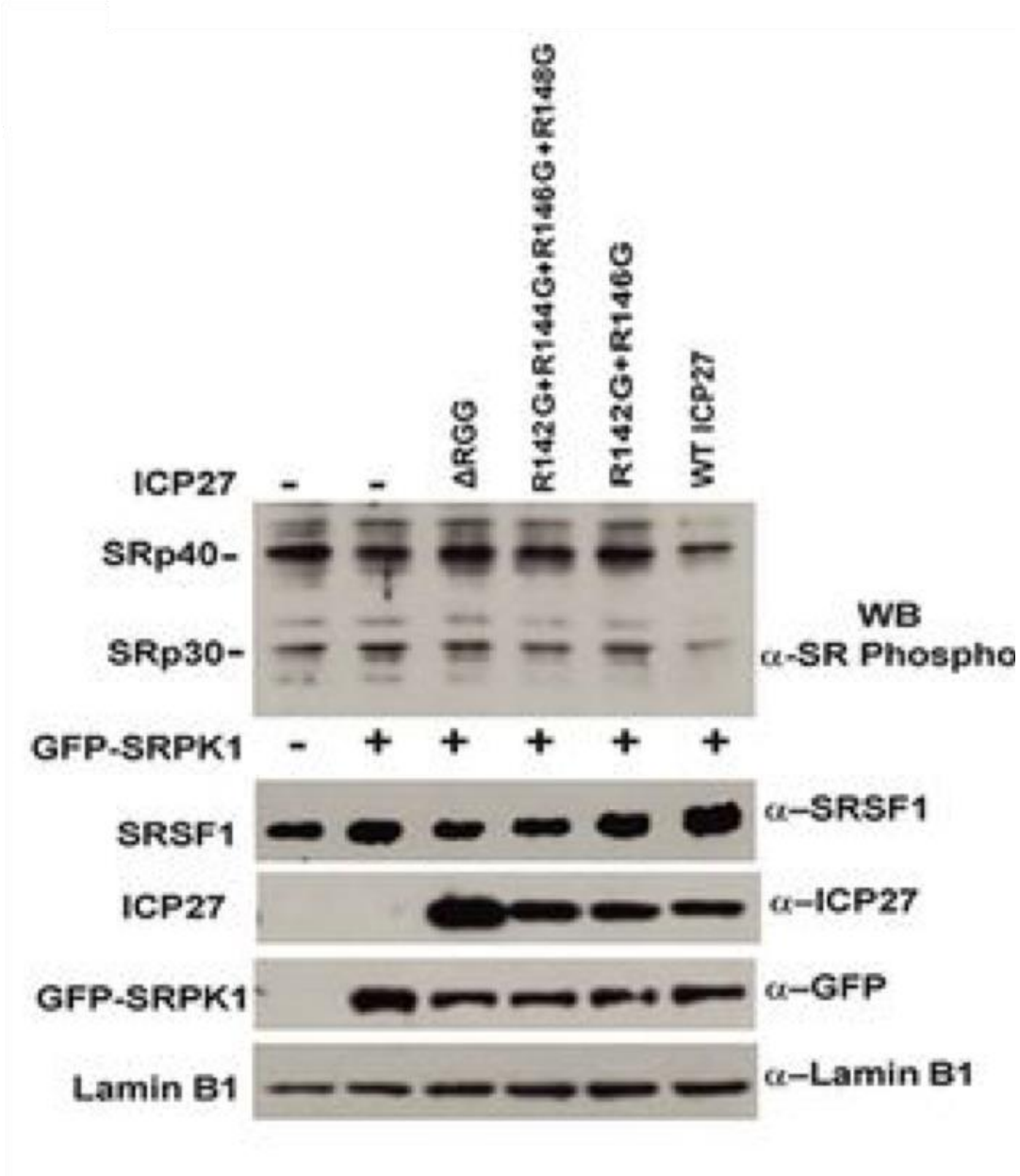
phosphoepitope antibody that recognizes phosphorylated serines within the RS repeats of SR proteins. Analysis of the western blot revealed reduced phosphorylation for SRSF1 (SRp30) and SRSF5 (SRp40) in WT ICP27 transfected cells but not in the ICP27 RGG box mutants (Figure 3.8). To verify that reduced phosphorylation was not due to lower protein levels, the blot was probed with anti-SRSF1 antibody which indicated equivalent amounts of SRSF1 among the samples (Figure 3.8). These results further affirm ICP27 induces hypophosphorylation of SR proteins during HSV-1 infection. This action requires ICP27 to interact with SRPK1 through its RGG motif; and mutation of the RGG residues disrupts SRPK1 binding and does not lead to aberrant SR protein phosphorylation.

Discussion

SR proteins are an essential class of non-snRNP splicing factors with important roles in regulating constitutive and alternative splicing, mRNA export, stability and translation[61–63]. SR proteins are phosphorylated by SR specific kinases and their phosphorylation state determines their localization as well as their function[63]. A major regulator of SR protein activity in humans is the ubiquitously expressed cellular kinase SRPK1, which has been shown to phosphorylate the N-terminus of the SRSF1 RS domain, promoting its translocation into the nucleus and spliceosome assembly[65,66,85,86]. During infection with HSV-1, the virus expresses the multifunctional adaptor protein ICP27, which has been shown to disrupt various facets of host pre-mRNA maturation in order to facilitate viral replication[12]. ICP27 interacts with SRPK1 and redistributes the kinase into the nucleus, resulting in aberrant hypophosphorylation of SR proteins and stalled splicing complexes[23].

Figure 3.8. ICP27 RGG box mutants do not inhibit SRPK1 activity *in vivo*.

HeLa cells were co-transfected with GFP-SRPK1 and WT ICP27 or RGG box mutants and were infected with 27LacZ 16 hours after transfection for 8 hours. Portions of cell lysates were fractionated by SDS-PAGE, and Western blots were probed with anti-phosphoepitope SR antibody, anti-SF2 (SRSF1), anti-ICP27 antibody, and anti-lamin B1 antibody as a loading control.


ICP27
-
-
ΔRGG
R142G+R144G+R146G+R148G
R142G+R146G
WT ICP27
SRp40-
SRp30-
WB
α-SR Phospho
GFP-SRPK1
-
+
+
+
+
+
SRSF1
α–SRSF1
ICP27
α–ICP27
GFP-SRPK1
α–GFP
Lamin B1
α–Lamin B1

Binding between SRPK1 and ICP27 was previously shown to require the ICP27 RGG box and arginine methylation was reported to modulate the binding[46]. However, it was not known whether the RGG box residues by themselves were sufficient for binding or whether other ICP27 residues were necessary for interaction. Further, the binding affinity between the two proteins had not been characterized and it was not known whether binding between ICP27 and SRPK1 was competitive with SRPK1 binding to its native substrate SRSF1. Additionally, it was also not clear how methylation of the ICP27 RGG box affected the interface between the two proteins.

The results from this study have provided key details regarding the ICP27-SRPK1 interaction. ITC and NMR experiments revealed the ICP27$_{103\text{-}155}$ peptide, which contains the RGG motif and its surrounding residues, can independently bind SRPK1 in a competitive manner and with an affinity comparable to that of SRSF1. The crystal structure of the ICP27 RGG-SRPK1 complex illustrated the RGG residues bind to the same SRPK1 docking groove, and in a similar fashion, as the RS repeats within SRSF1. Examination of the structure also provides an explanation for why methylation of the RGG box weakens ICP27 affinity for SRPK1, seen from our ITC results. Within the RGG box, the side chains of arginines 142, 144 and 148 become buried when bound to SRPK1. The addition of methyl groups to the arginine residues surrounding this region is likely to introduce steric clashes between the two proteins and, consequently, interfere with binding. Through our *in vivo* experiments, we further established the importance of the RGG motif for SRPK1 disruption. Western blot and IF analysis showed that ICP27 RGG box mutants exhibited greatly diminished binding with SRPK1, was not able to recruit SRPK1

into the nucleus and did not induce a lower phosphorylation state for SR proteins as was seen with WT ICP27.

In sum, the data here elucidates the molecular mechanism of ICP27-mediated inhibition of SRPK1 function. ICP27 RGG box residues bind directly to the SRPK1 docking groove, which ordinarily recognizes the RS domains of SR proteins, thus displacing SRPK1's cellular substrates and preventing their phosphorylation. In addition, binding between the two proteins resulted in redistribution of SRPK1 to the nucleus, likely preventing contact between the kinase and cytoplasmic SR proteins, resulting in restriction of SR proteins to the cytoplasm and disrupting their ability to assemble the host splicing machinery.

CHAPTER 4

ICP27 promotes host antisense transcript levels during HSV-1 infection

Introduction

HSV-1 lytic infection significantly impacts the cellular transcriptome. These impacts include alterations to gene isoform expression, alternative splicing, 3' end formation and differential gene expression[20,25,70]. Moreover, in a recent investigation by Wyler et al. (2017), relying on stranded RNA-sequencing data as well as published sequencing data of 4-thiouridine (4sU) pulse-labeled RNA[20] from HSV-1 infected fibroblasts, HSV-1 was shown to specifically upregulate the expression of approximately 1000 antisense transcripts from the cellular genome[87]. Examination of these transcripts revealed most were not included in existing genome annotations, were expressed as early as 1 hpi, unspliced and polyadenylated[87]. In experiments with the viral DNA polymerase inhibitor phosphonoacetic acid (PAA), mutant ICP0 (ΔICP0) or ICP4 (ΔICP4) deletion viruses, and transient transfections with an ICP4 encoding plasmid, the authors demonstrated that activation of a subset of antisense transcripts was dependent on the IE protein ICP4, the major viral transcription factor[87]. Notably, it was observed that, in all cases, levels of antisense transcripts were markedly higher with infection as compared to ICP4 overexpression, alluding to the possibility that other viral factors may contribute to this phenomenon[87]. Due to differences in expression kinetics as well as sensitivity to ICP4

deletion, it was hypothesized that disparate antisense transcripts were induced by different mechanisms; with ICP4 required for expression of a subset of these transcripts.

Upon infection with HSV-1, one of the first proteins expressed by the virus is the multifunctional adaptor ICP27. As discussed in preceding chapters, ICP27 has been reported to interact with a multitude of protein partners, bind GC-rich RNA and specifically disrupt various aspects of cellular mRNA biogenesis in order to advance virus production[12,88,25,89,91]. Although the functions of ICP27 have long been the subject of active investigation by numerous research groups, to date, the molecular details and extent of ICP27-mediated disruptions to the host transcriptome are still not well defined. Results disclosed by Wyler and colleagues indicate other factors are necessary for maximum accumulation of host antisense transcripts. However, it's currently not known what other factor(s) are involved and the authors did not specifically examine the actions of ICP27 in their study. In this chapter, we sought to determine if ICP27 contributes to the increased levels of cellular antisense RNA observed during infection. Utilizing both transcriptome-wide sequencing (RNA-seq) and individual-nucleotide resolution UV crosslinking and immunoprecipitation followed by high-throughput sequencing (iCLIP-seq) of wild-type and ICP27 mutant infected cells we show that ICP27 binds to induced host antisense transcripts *in vivo* and that accumulation of these transcripts require both ICP27 and its RNA-binding RGG motif.

Materials and Methods

Cells, viruses and recombinant plasmids

HeLa cells were grown on minimal essential medium containing 10% newborn calf serum. HSV-1 wild-type strain KOS and ICP27 viral mutants 27LacZ and ΔRGG were previously described[40,47]. Plasmid pSG130B/S (WT) was described previously[51] and plasmid pMd4-5 was a generous gift from Steve Rice[90].

Virus infection and transfection

Cells were infected with wild-type or mutant virus at a multiplicity of infection (MOI) of 5 and incubated at 37°C in a 5% (v/v) CO_2-enriched incubator for 8 hours. Transfection of plasmid DNA was performed by using Lipofectamine 2000 reagent (Invitrogen) according to manufacturer's protocol. Twenty-four hours after transfection, cells were infected with ICP27 null mutant virus 27LacZ at a MOI of 5 for 8 hours to activate expression of the plasmids and to replicate the conditions of infection.

RNA extraction and sequencing

HeLa cells were infected with HSV-1 KOS, 27LacZ or ΔRGG. Eight hours after infection, cells were harvested and nuclei fractionated from the cytoplasm as described previously[42]. Briefly, cells were lysed in low-salt lysis buffer (10 mM Tris [pH 7.4], 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.5% NP-40, and protease inhibitor cocktail [Roche]). The cell lysate was passed through a syringe with a 25-gauge needle 10 times. The nuclei were pelleted by centrifugation at 14,000 × g for 30 seconds. The supernatant was transferred to a new tube and represented the cytoplasmic fraction. The nuclear pellet was resuspended in high-salt extraction buffer consisting of phosphate-buffered saline (PBS) containing 250 mM NaCl, 0.5% NP-40, and protease inhibitor cocktail. Total RNA was isolated with TRIzol

(Invitrogen) from the nuclear and cytoplasmic fractions. RNA was extracted with chloroform and precipitated in ethanol. Poly(A) RNA was selected using poly-T oligos on magnetic beads and sequencing libraries were generated with the TruSeq mRNA stranded kit (Illumina) according to manufacturer's instructions. The Agilent 2100 Bioanalyzer and Qubit fluorometer were used in quantification and qualification of the sample library before being sequenced using the Illumina HiSeq4000.

iCLIP-sequencing

HeLa cells were transfected with plasmids expressing WT-ICP27, ΔRGG or empty vector pUC18 using Lipofectamine 2000 reagent (Invitrogen) according to manufacturer's protocol. Twenty-four hours after transfection, cells were infected with ICP27 null mutant virus 27LacZ to induce expression of the plasmids. At 8 hpi, cells were exposed to UV light at 254 nm with 400mj/cm$_2$ and then collected for iCLIP library preparation. The iCLIP protocol was carried out as described by Huppertz et al. (2014)[91] with minimal modifications. Briefly, lysates were incubated with anti-ICP27 antibody P1119 (Virusys) bound to 100 µl protein G Dynabeads (ThermoFisher) at 4 °C under rotation with subsequent washing. Following immunopurification, RNA 3′ end dephosphorylation, ligation of the adapter 5′-rAppAGATCGGAAGAGCGGTTCAG/ddC/-3′ to the 3′ end and 5′ end radiolabeling, protein–RNA complexes were size-separated by SDS-PAGE and transferred to nitrocellulose membrane. The regions corresponding to 70-100 kDa were excised from the membrane to isolate the bound RNA by proteinase K treatment. RNAs were reverse-transcribed in all experiments using SuperScript III reverse transcriptase (ThermoFisher) and custom indexed primers. Resulting cDNAs were subjected to

electrophoresis on a 6% TBE-urea gel (ThermoFisher) for size selection. Purified cDNAs were circularized, linearized and amplified for high-throughput sequencing. The Agilent 2100 Bioanalyzer and Qubit fluorometer were used in quantification and qualification of the sample library before being sequenced on the Illumina NovaSeq.

Data analysis

The 100-cycle single-end sequencing libraries were aligned to human reference genomes hg19/GRCh37 and hg38/GRCh38 using STAR version 2.5.2a. No multimapping was allowed in alignment. Bigwig files were then generated using deepTools v3.02 with RPKM normalization.

For iCLIP samples, identification of significant crosslinked clusters and enriched K-mers was performed as described by Modic et al. (2019)[92] using the iCount python package (https://github.com/tomazc/iCount). Briefly, fastq files were demultiplexed and barcodes removed via the 5' experimental barcodes. Unique molecular identifiers (UMIs) were used to remove PCR duplicates and reads were mapped to the human reference genome. After the mapping, a BAM file containing both unique and multi-mapped reads was generated for each sample. Two BED files showing (unique or unique and multi-mapped) crosslink sites on the genome and a BedGraph file giving coordinate information and count was generated from each BAM file. Paraclu (http://cbrc3.cbrc.jp/~martin/paraclu/) was used to aggregate crosslink sites. Thresholded crosslinks, representing high read counts present within peaks, were obtained from Paraclu peaks. BED files were used to generate RNA maps and motif analysis, which identified enriched k-mers and their distribution around

thresholded crosslink positions. Sample clustering was performed with deepTools plotCorrelation[93] (https://github.com/deeptools/deepTools/blob/master/deeptools/plotCorrelation.py).

Results

Optimum accumulation of cellular antisense transcripts is dependent on ICP27 and its RGG motif

A previous RNA-seq study reported HSV-1 induced the expression of approximately 1000 cellular antisense transcripts, of which 12 were validated and selected for further analysis[87]. Based on their locations, antisense transcripts were categorized as: 1) divergent, if the sense and antisense transcripts start from the same promoter region, but don't overlap; 2) convergent, if the 5' ends of the antisense and sense transcripts overlap; or 3) internal, if several exons of the sense transcript overlap with the antisense transcript. Examination of these transcripts show regions of high GC-content and expression kinetics coincident with ICP27 expression.

HSV-1 ICP27, is an RNA-binding regulatory protein demonstrated to modulate host gene expression. Binding of RNA has been shown to occur primarily through an N-terminal RGG box[14,40] and multiple studies have reported the RGG box participates in altering cellular pre-mRNA processing[25,46,89]. To further characterize the role of ICP27 and its RGG box in regulating host transcriptional changes during HSV-1 infection, we sequenced mRNA purified from nuclear and cytoplasmic fractions of HeLa cells mock infected or infected with WT HSV-1, ΔRGG or 27LacZ for 8 hours. Analysis of the resulting expression profiles

identified several, previously reported, ICP27-mediated phenomena; including changes to host cell splicing[23,73,94–96] (Figure 4.1Aiv, viii and 4.1Biv, viii) and premature cleavage and polyadenylation[24] (Figure 4.1Civ, viii and 4.1Div, viii). In agreement with published results, many of these impacted regions were rich in GC sequences and deletion of the RGG box diminished disruptions within the targeted transcripts (Figure 4.1Aii, iii, vi, vii; 4.1Bii, iii, vi, vii; 4.1Cii, iii, vi, vii; 4.1Dii, iii, vi, vii). In addition, we also observed the above mentioned accumulation of antisense transcripts in our RNA-seq data. Despite inherent differences among disparate cell lines (HeLa vs fibroblasts; immortalized vs primary), comparison of sequencing data from our WT HSV-1 infected HeLa cells and published human foreskin fibroblasts (HFF) or primary lung fibroblasts (WI-38) revealed remarkably similar expression of host antisense RNAs (Figure 4.2). Of the 12 virus induced antisense transcripts validated by Wyler and colleagues in HFF and WI-38 cells[87], almost all were found to be analogously induced in HeLa cells. Henceforth, we focus our investigation to these transcripts. For a full list of antisense RNAs examined in this chapter and a brief characterization see Table 4.1. Among the list of activated transcripts are antisense transcripts to BCL2 Binding Component 3 (BBC3as), a potent inducer of apoptosis[97]; Inhibitor of Growth Family Member 1 (ING1as), a tumor suppressor protein also reported to induce apoptosis and cell growth[98]; and Regulatory Factor X1 (RFX1as), a transcription factor involved in immunity and cancer[99,100] (Figure 4.2). Of note, although levels of antisense transcripts were greater in both nuclear and cytoplasmic portions of WT HSV-1 infected cells as compared to mock infected controls, differences between the two conditions were dramatically more pronounced within the nuclear compartment (Figure 4.2Biv vs viii; 4.2Div vs viii; 4.2Eiv vs viii). In contrast to WT infection, cells infected with

Figure 4.1. Differential effects on host pre-mRNA processing in WT and ICP27 mutant infections.

Read counts mapping to four representative ICP27-targeted genes. Examples show induced retention of specific introns (Aiv, viii; Biv, viii) and expression of prematurely cleaved and polyadenylated pre-mRNAs (Civ, viii; Div, viii) in some host genes upon WT HSV-1 infection. These alterations were noticeably weakened in infections with ΔRGG (Aiii, vii; Biii, vii; Ciii, vii; Diii, vii) or 27LacZ (Aii, vi; Bii, vi; Cii, vi; Dii, vi). Line charts plotting GC-content for whole transcript or impacted regions are at top. (C and D) Cryptic PASs denoted in green below.

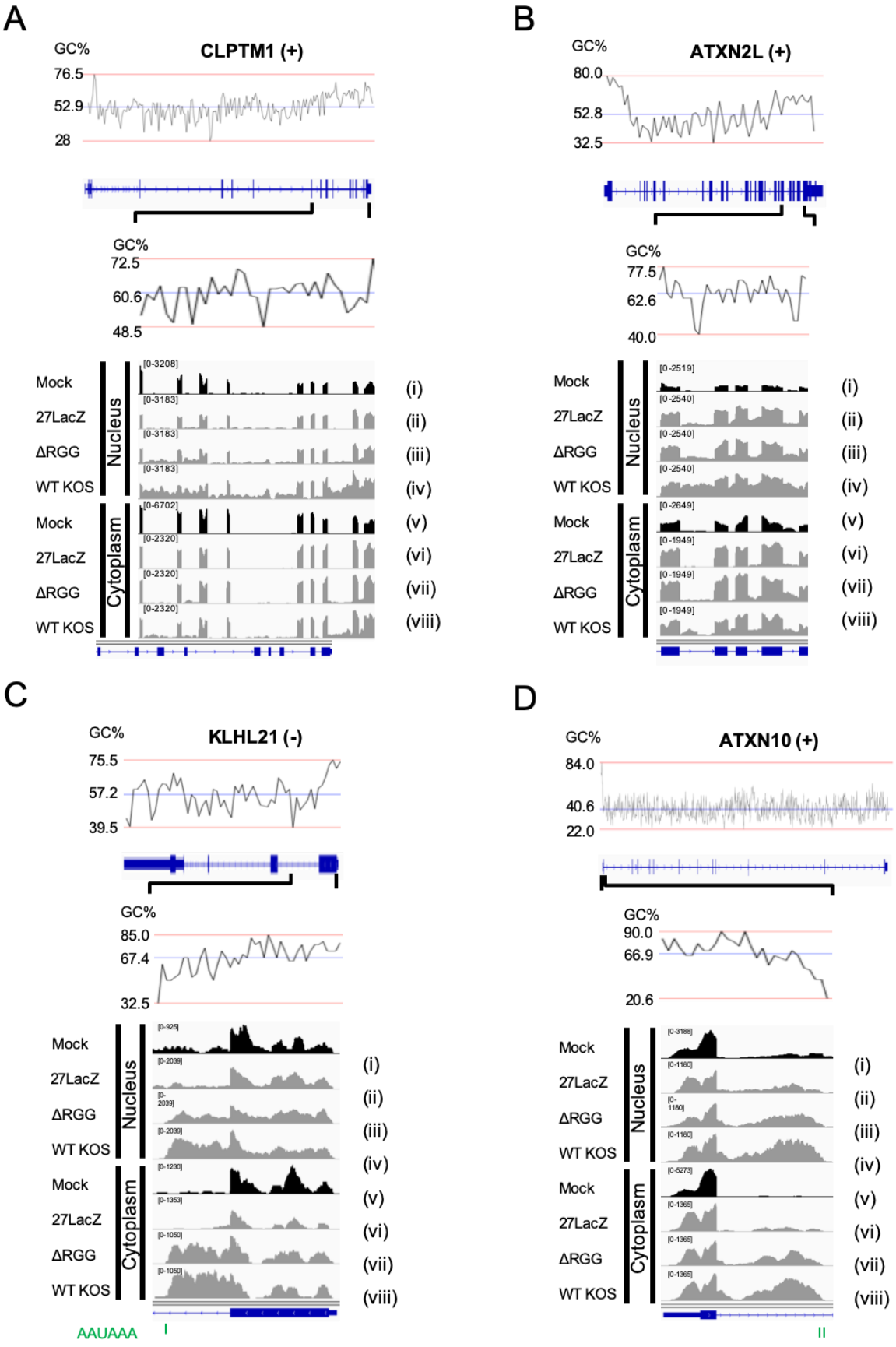
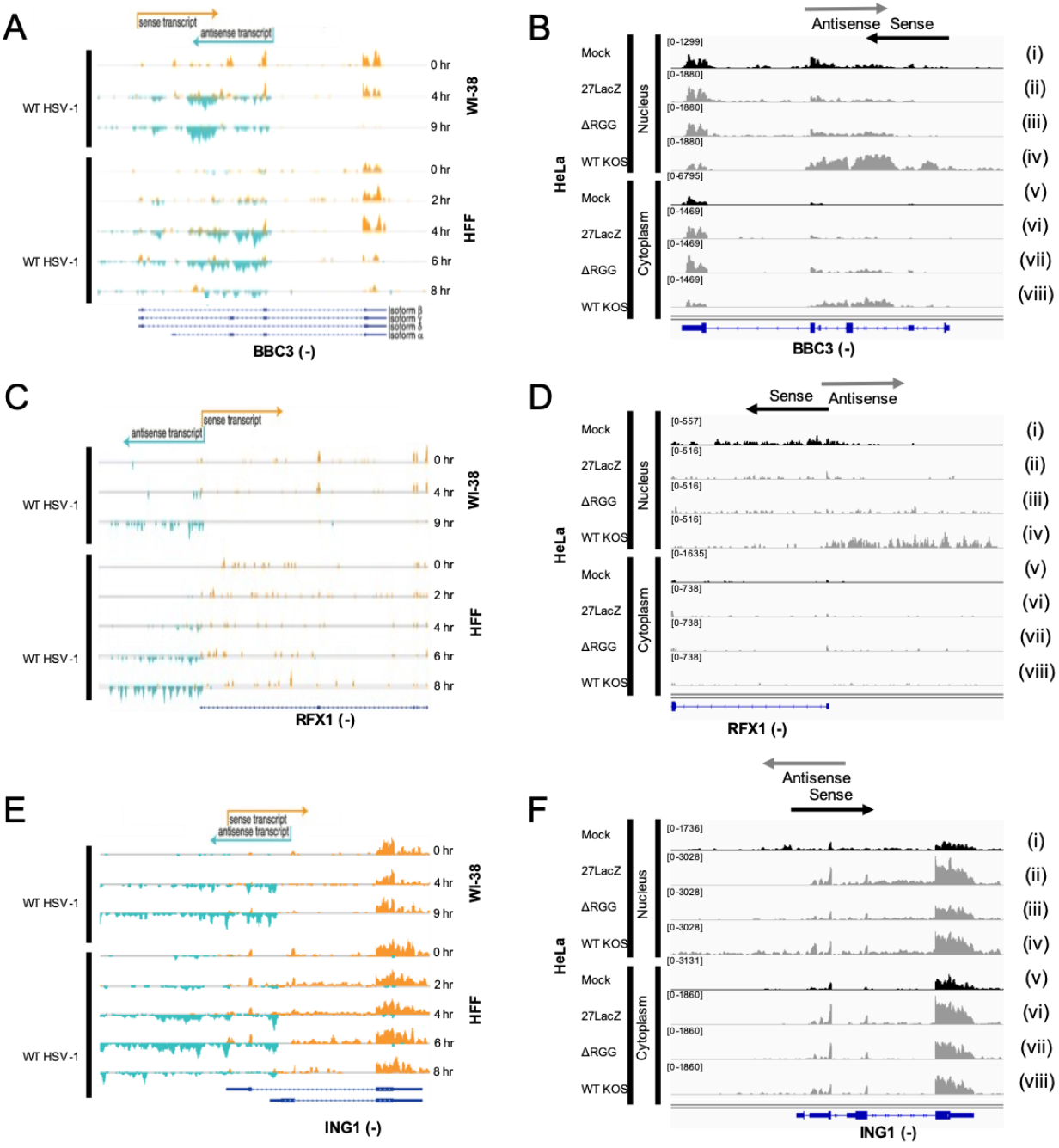

A
GC%
CLPTM1 (+)
76.5
52.9
28
GC%
72.5
60.6
48.5
Mock
27LacZ
ΔRGG
WT KOS
Nucleus
Cytoplasm
[0-3208]
[0-3183]
[0-3183]
[0-3183]
[0-6702]
[0-2320]
[0-2320]
[0-2320]
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
(viii)
B
GC%
ATXN2L (+)
80.0
52.8
32.5
GC%
77.5
62.6
40.0
[0-2519]
[0-2540]
[0-2540]
[0-2540]
[0-2649]
[0-1949]
[0-1949]
[0-1949]
C
GC%
KLHL21 (-)
75.5
57.2
39.5
GC%
85.0
67.4
32.5
[0-925]
[0-2039]
[0-2039]
[0-1230]
[0-1353]
[0-1050]
[0-1050]
AAUAAA
I
D
GC%
ATXN10 (+)
84.0
40.6
22.0
GC%
90.0
66.9
20.6
[0-3188]
[0-1180]
[0-1180]
[0-5273]
[0-1365]
[0-1365]
[0-1365]
II

Figure 4.2 Accumulation of host antisense transcripts during HSV-1 infection require ICP27 and its RGG box.

Examples of antisense transcripts. (A, C, E) For reference. Adapted from Wyler et al. (2017)[87]. Coverage profiles for BBC3as, RFX1as and ING1as from poly(A)-selected WI-38 cells infected with WT-HSV-1 for 0, 4 or 9 hours or total RNA-seq data from Ribozero treated HFF cells infected for 0, 2, 4, 6, 8 hours. Sense genes shown in orange, running from left to right. Antisense transcripts in cyan, running from right to left. (B, D, F) RNA-seq data from poly(A)-selected transcripts isolated from nuclear (i-iv) or cytoplasmic fractions (v-viii) of HeLa cells infected with WT HSV-1 (iv, viii), ΔRGG (iii, vii) or 27LacZ (ii, v) for 8 hours. Gray and black arrows denote directionality of antisense and sense transcripts, respectively, and begin at transcription start sites. For BBC3 (B) and RFX1 (D) sense transcripts proceed from right to left while antisense transcripts proceed from left to right. For ING1 (F), sense transcripts run from left to right and antisense transcripts run from right to left.


A
sense transcript
antisense transcript
WT HSV-1
0 hr
4 hr
9 hr
WI-38
0 hr
2 hr
4 hr
6 hr
8 hr
HFF
isoform β
isoform γ
isoform δ
isoform α
BBC3 (-)
B
Antisense
Sense
HeLa
Nucleus
Cytoplasm
Mock
27LacZ
ΔRGG
WT KOS
[0-1299]
[0-1880]
[0-6795]
[0-1469]
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
(viii)
BBC3 (-)
C
sense transcript
antisense transcript
RFX1 (-)
D
Sense
Antisense
[0-557]
[0-516]
[0-1635]
[0-738]
RFX1 (-)
E
ING1 (-)
F
Antisense
Sense
[0-1736]
[0-3028]
[0-3131]
[0-1860]
ING1 (-)

ICP27 null mutant virus 27LacZ displayed substantial decreases in antisense transcript levels (Figure 4.2Bii, vi; 4.2Dii, vi; 4.2Eii, vi). Further, unlike ICP4 deletion, which was reported to negatively affect only a subset of antisense transcripts[87], ICP27's effects appear to be more comprehensive, as deletion of ICP27 abridged expression levels for almost every antisense transcript we examined. Additionally, antisense RNA levels were markedly reduced upon infection with the RGG box deletion mutant (ΔRGG) as well (Figure 4.2Biii, vii; 4.2Diii, vii; 4.2Eiii, vii). The RGG box is important for ICP27-RNA binding and these results, taken together, confirm HSV-1 infection induces expression of antisense transcripts from the human host cell genome, indicate ICP27 is critical for maximum accumulation of antisense transcripts and suggest this function is mediated by ICP27's RNA-binding activity.

ICP27 selectively binds to GC-rich sequences within antisense transcripts

Data from the previous section identify the RGG box is important for accumulation of host antisense transcripts during infection. As discussed previously, the RGG box is the site of SRPK1 interaction and the main ICP27 RNA-binding domain, demonstrated to selectively bind GC-rich transcripts *in vitro*[14,101] and *in vivo*[25,102]. Interestingly, during our examination of the expression coverage profiles, we observed regions significantly high in GC-content for many of the antisense transcripts (Table 4.1 and Figure 4.5). Given ICP27's preference for GC-rich sequences we reasoned that ICP27's RGG box may bind to these regions. To test this hypothesis, we profiled genome-wide ICP27-RNA interactions by iCLIP-seq. iCLIP-seq libraries were prepared from uninfected HeLa cells or cells transfected with plasmids encoding WT ICP27 or ΔRGG and then infected with 27LacZ for 8

Table 4.1. Transcripts selected for detailed analysis.

The antisense transcripts investigated in this study are listed along with their classification, position and GC-content.

Transcript	Type	Chr	Start	End	Strand	GC%
BBC3as	Internal	19	47729547	47735593	+	66
C1orf159as	Convergent	1	1043123	1066056	+	59
EFNB1as	Divergent	X	68044923	68048747	-	55
FOXO3as	Convergent	6	108874346	108883564	-	57
IER2as	Convergent	19	13261198	13264876	-	65
IFFO2as	Divergent	1	19233363	19283746	+	55
ING1as	Convergent	13	111365951	111367813	-	67
MEGF6as	Internal	1	3420963	3455048	+	63
NFKB2as	Convergent	10	104150343	104155306	-	51
RFX1as	Convergent	19	14116734	14138959	+	47
SLC27A4as	Divergent	9	131101098	131102612	-	51
SNX6as	Read-through	14	35016928	35024457	+	50

hours. Two replicate iCLIP experiments were performed for WT ICP27 and ΔRGG transfected samples. In order to control for the different steps during library preparation, several negative controls were maintained throughout the experiment. These included the uninfected sample, a sample transfected with empty vector (pUC18) then infected with 27LacZ and a sample where UV cross-linking was omitted (Figure 4.3). During the course of the protocol, the progress and quality of library preparation was monitored at two steps: the autoradiograph of the protein-RNA complex after membrane transfer (Figure 4.3A) and the gel image of the PCR products (Figure 4.3C). High-throughput sequencing and bioinformatic analysis of the resulting libraries identified over 35 million reads uniquely mapped to the human genome for WT ICP27 samples and approximately 18 million reads for ΔRGG and uninfected samples (Figure 4.4A). A heatmap correlation matrix of our libraries showed WT ICP27 samples were highly reproducible and clustered separately from the negative controls (Figure 4.4B).

The iCLIP data identified ICP27 cross-linking events in almost all antisense transcripts examined in this chapter (Table 4.1), with preference shown for especially GC-rich locations (Figure 4.5). As such, the number of reads within individual antisense transcripts varied widely, with transcripts containing higher GC-content exhibiting greater cDNA counts. For example, BBC3as, ING1as and RFX1as were previously demonstrated to be elevated in WT HSV-1 infected cells (Figure 4.2). However, the average GC-content for BBC3as (66%) and ING1as (67%) is conspicuously higher than RFX1as (47%) (Table 4.1), with RFX1as containing only a short GC-rich stretch near its 5' end. Consequently, ICP27 cross-linking to RFX1as was only observed at the 5' end of the transcript while cross-linking to BBC3as and ING1as was distributed extensively throughout the respective

Figure 4.3. iCLIP experiments.

Analysis of cross-linked ICP27–RNA complexes using (A) denaturing gel electrophoresis and (B) western blotting. (A) Immunoprecipitated ICP27-RNA complexes were size separated using PAGE and visualized by autoradiography. The white boxes mark regions of the membrane isolated for subsequent purification steps. (B) Western blot from 10% of immunoprecipitated samples was performed to monitor pulldown efficiency. (C) Agarose gel of the iCLIP libraries. White boxes indicate samples submitted for sequencing.

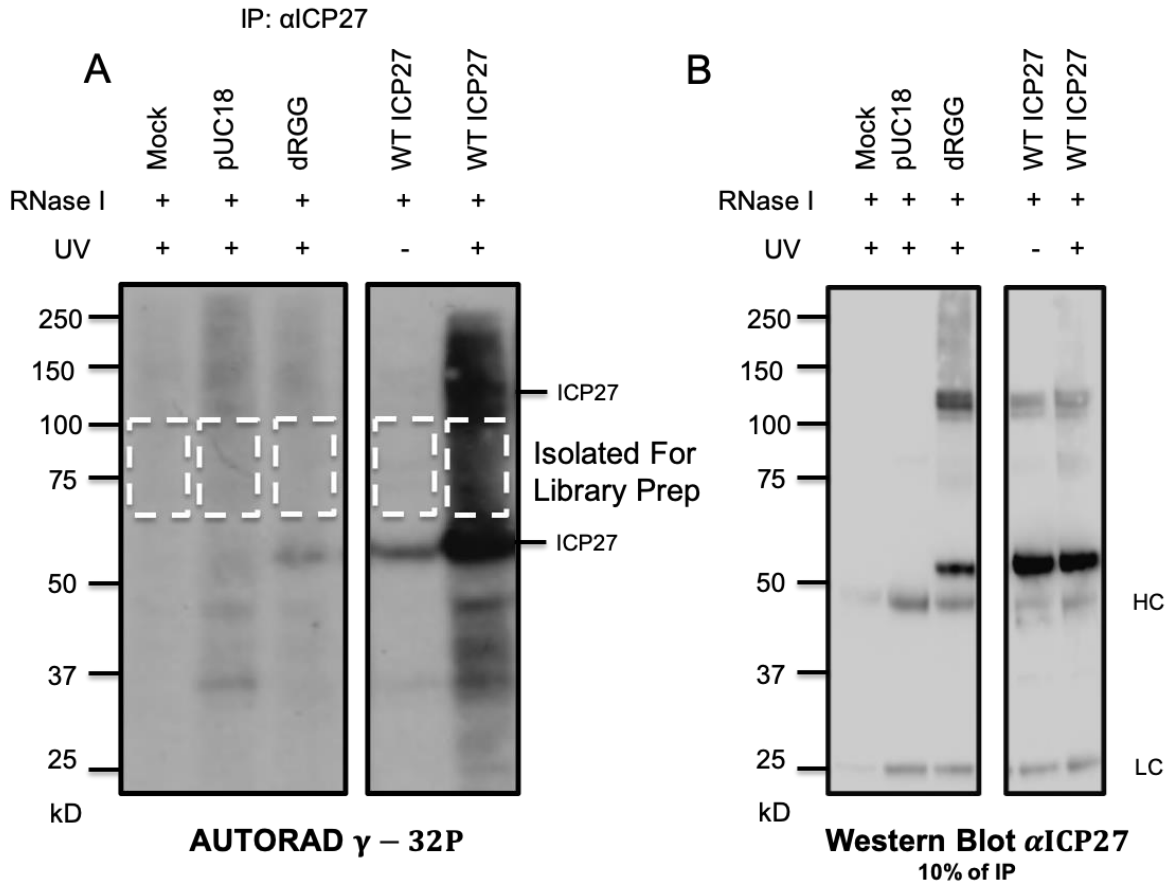

IP: αICP27
A
Mock
pUC18
dRGG
WT ICP27
WT ICP27
RNase I + + + + +
UV + + + - +
250
150
100
75
50
37
25
kD
ICP27
Isolated For Library Prep
ICP27
AUTORAD γ − 32P
B
Mock
pUC18
dRGG
WT ICP27
WT ICP27
RNase I + + + + +
UV + + + - +
250
150
100
75
50
37
25
kD
HC
LC
Western Blot αICP27
10% of IP
C
Mock
pUC18
dRGG
-XL
WT ICP27
bp
500
400
300
200
100
iCLIP Analysis
2.5% GTG Agarose
PCR Library

Figure 4.4. iCLIP-seq libraries.

(A) Illumina sequencing results showing total number of reads and number of reads uniquely mapped to the human genome (hg38/GRCh38) for each sample. (B) Sample clustering results with deepTools plotCorrelation[93].

A

	Mock	dRGG 1	dRGG 2	WT 1	WT 2
Total Reads	43,384,596	58,788,338	57,553,440	102,432,961	70,555,284
Uniquely Mapped Reads (Human)	17,401,870 (40.11%)	19,077,372 (32.45%)	18,834,266 (32.72%)	43,977,435 (42.93%)	35,447,179 (50.24%)

B

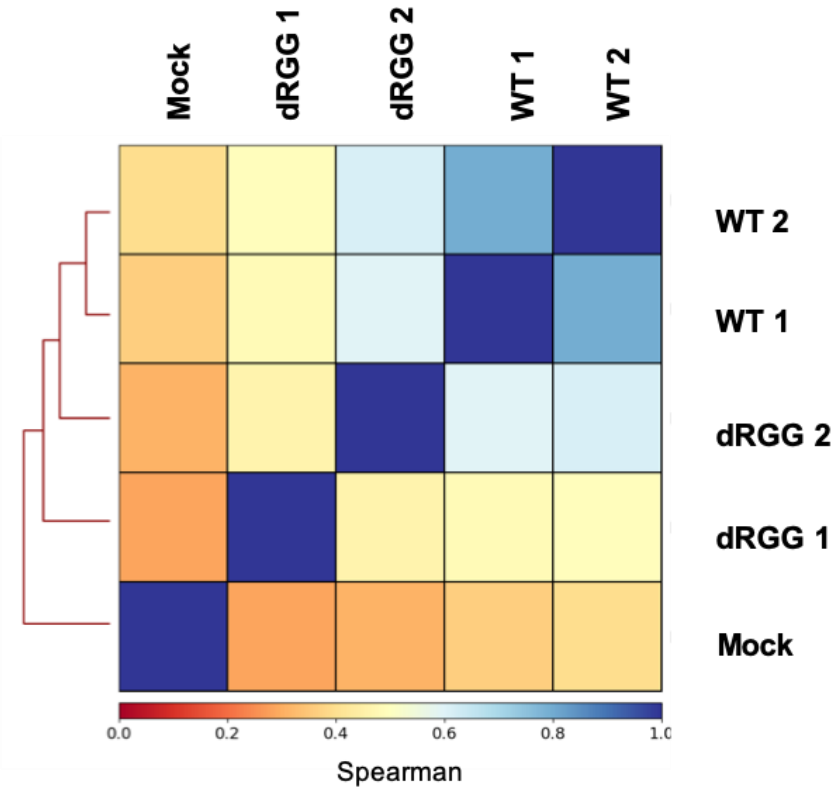
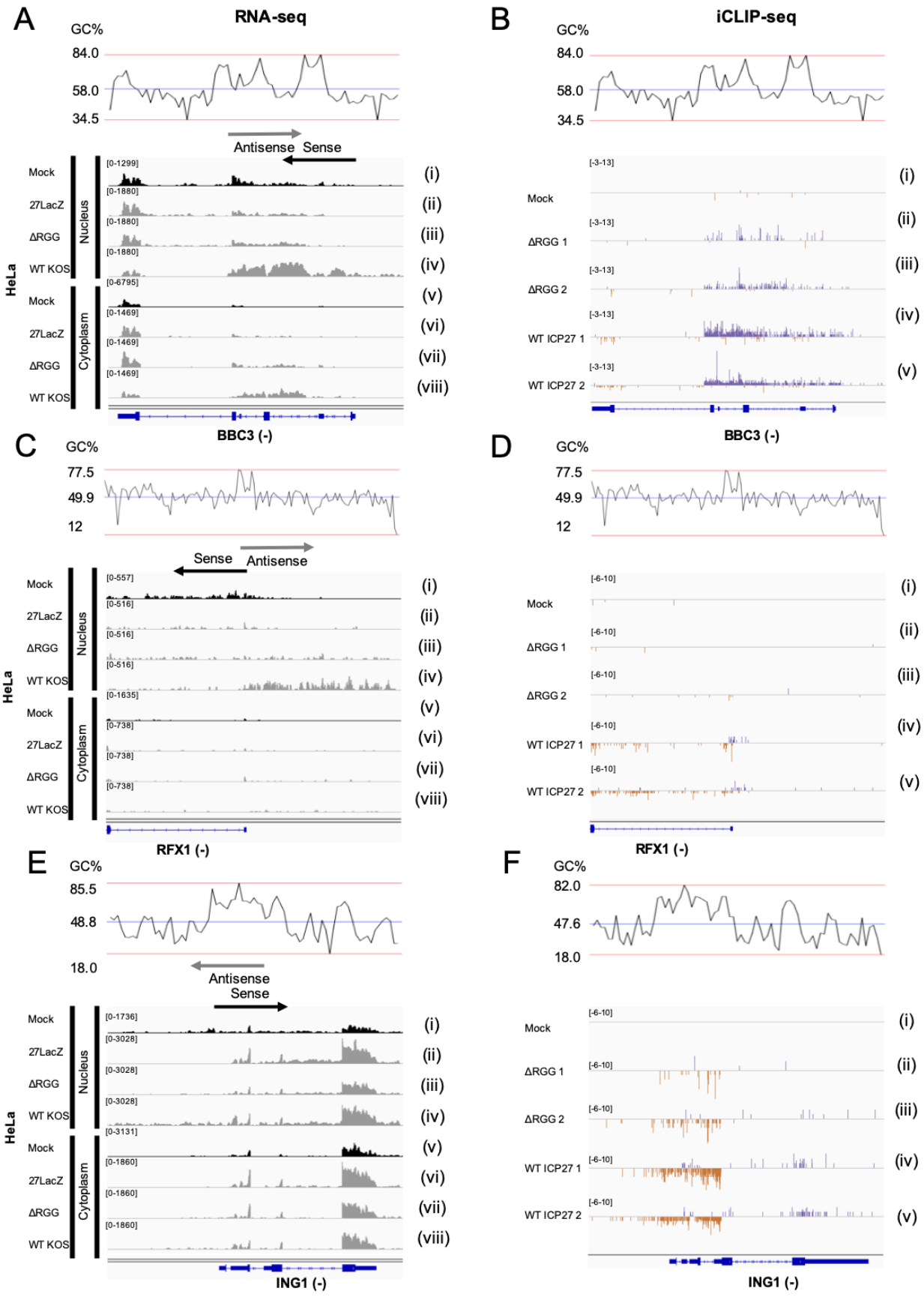


Figure 4.5. ICP27 specifically binds GC-rich regions within host antisense transcripts. BBC3as, RFX1as and ING1as were selected to illustrate ICP27 binding. (A, C, E) RNA-seq coverage profiles for these transcripts were included for reference. Gray arrows show directionality of antisense transcripts and black arrows show directionality of sense transcripts. Beginning of arrows coincide with transcription starts. (B, D, F) ICP27-RNA interaction profiles from uninfected cells (i) or cells transfected with plasmids expressing WT ICP27 (iv, v) or ΔRGG (ii, iii), then infected with 27LacZ for 8 hours. ICP27 cross-linking events on the (-) strand are in orange, below, and cross-linking events on the (+) strand are in purple, above. Charts plotting GC content within the indicated regions are located at the top of each profile.


A
RNA-seq
GC%
84.0
58.0
34.5
Antisense
Sense
HeLa
Nucleus
Cytoplasm
Mock
27LacZ
ΔRGG
WT KOS
[0-1299]
[0-1880]
[0-1880]
[0-1880]
[0-6795]
[0-1469]
[0-1469]
[0-1469]
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
(viii)
BBC3 (-)
B
iCLIP-seq
GC%
84.0
58.0
34.5
Mock
ΔRGG 1
ΔRGG 2
WT ICP27 1
WT ICP27 2
[-3-13]
(i)
(ii)
(iii)
(iv)
(v)
BBC3 (-)
C
GC%
77.5
49.9
12
Sense
Antisense
[0-557]
[0-516]
[0-516]
[0-516]
[0-1635]
[0-738]
[0-738]
[0-738]
RFX1 (-)
D
GC%
77.5
49.9
12
[-6-10]
RFX1 (-)
E
GC%
85.5
48.8
18.0
Antisense
Sense
[0-1736]
[0-3028]
[0-3028]
[0-3028]
[0-3131]
[0-1860]
[0-1860]
[0-1860]
ING1 (-)
F
GC%
82.0
47.6
18.0
[-6-10]
ING1 (-)

transcripts (Figure 4.5Biv, v; 4.5Div, v; 4.5Fiv, v). Thus, the results here reveal cellular antisense transcripts induced by HSV-1 infection contain GC-rich regions targeted by ICP27's RGG box for binding.

Discussion

Recent advancements in high-throughput sequencing technologies have allowed for the identification of widespread transcription outside of delineated gene boundaries among all kingdoms of life. This phenomenon, termed "pervasive transcription", results in the production of several types of non-coding RNAs (ncRNAs), including antisense RNAs[103]. Antisense RNAs belong to the class of long ncRNAs (lncRNAs) (>200 nucleotides in length) and are prevalent throughout the genomes of a number of species[103]. In humans, approximately 70% of the genome has been reported to generate antisense transcripts[104]. However, in general, antisense transcripts are expressed at much lower rates, usually more than 10-fold lower, than the canonical sense transcript[105]. Further, the RNA surveillance machinery has also been reported to modulate antisense transcript levels down to non-existent or physiologically tolerable levels[106]. Generated from independent, bidirectional or cryptic promoters, antisense transcripts have been reported to primarily function in gene regulation; regulating expression of target genes at almost every stage of mRNA biogenesis, from transcription and splicing to RNA degradation and translation[103]. Able to act *in cis* or *in trans*, and demonstrated to integrate diverse regulatory signals, antisense transcripts have been described as hubs for gene regulation[103]. Antisense transcripts are implicated in numerous cellular processes and their dysregulation has been blamed for a

number of human disorders, including cancer, diabetes, cardiovascular, muscular and neurodegenerative disorders[107].

A recent study examining cellular gene expression changes during HSV-1 infection identified widespread activation of host antisense RNAs[87]. This was determined to occur very early during infection and to involve the actions of the viral transcriptional transactivator ICP4, which was found to bind within the promoter regions of the antisense transcripts. Examination of histone marks within these promoter regions revealed marks of active promoters in both uninfected and infected cells; indicating these regions were poised for transcription prior to infection[87]. This finding was later corroborated in a study by Dremel and DeLuca (2019), which similarly reported ICP4 binding to cellular promoters free of histones and adjacent to euchromatin marks early during infection[108]. Gene ontology analysis revealed ICP4 bound to promoters of genes with common housekeeping functions, such as those involved in chromatin, transcription and metabolism[108]. However, the consequences of ICP4 binding to these specific genes were not investigated further.

In this chapter, we provide evidence for a previously uncharacterized function of ICP27. Our high-throughput sequencing results indicate ICP27 contributes significantly to a general increase in antisense transcript abundance. Although the mechanism for this phenomenon has not been elucidated, we observed that efficient accumulation of antisense transcripts required the ICP27 RGG box and that almost all of the induced antisense transcripts displayed varying levels of ICP27 cross-linking, specifically within the GC-rich regions. We speculate that ICP27 binding to antisense transcripts may alter their turnover rate and that ICP27 likely works in tandem with ICP4 to maximize accumulation of antisense transcripts that could be beneficial to the virus. In this instance, we envision a

scenario where, upon HSV-1 infection, the virus immediately begins to express the IE class of proteins, of which ICP27 and ICP4 are both members of. Following translation, ICP27 and ICP4 return back to the nucleus, where ICP4 binds to the promoter regions of antisense transcripts, which are already poised for transcription. This allows for the rapid expression of antisense RNAs containing regions of high GC-content, which are then targeted by ICP27's RGG box for binding. Binding by ICP27 stabilizes the transcripts and prevents their degradation. This hypothesis addresses several observations from the data and the literature; however, several important questions remain. How does ICP27 binding to antisense RNAs promote their escape from the cellular RNA degradation machinery? As illustrated in Figure 4.5, GC-content and ICP27 binding patterns fluctuate widely among antisense transcripts. Does higher GC-content or greater ICP27 binding affinity correlate with higher antisense RNA levels? What is the underlying molecular mechanism leading to the accumulation of antisense transcripts? What other viral or host factors are required? And, lastly, what role do these transcripts play during infection? It was previously demonstrated that expression of an antisense transcript to BBC3 inhibited this potent inducer of apoptosis *in cis*[87], however, the functions of other antisense transcripts were not examined and a number of other cellular genes with significant roles in critical pathways have also been determined to generate antisense transcripts upon HSV-1 infection. While many may not be biologically relevant, it's likely at least some could alter the cellular state to favor virus production. Future work will be directed at addressing these questions in order to better characterize ICP27 contributions to viral pathogenesis.

Interestingly, as a side note, for ΔRGG infections, we observed binding to antisense transcripts was considerably reduced but not completely abrogated, indicating other

regions of the protein likely contribute to RNA interactions as well, possibly the aforementioned RRR motif within the CTD. However, this RNA binding activity was insufficient to increase antisense transcript levels. This is in line with our previous reports, which demonstrated ΔRGG mutants exhibited substantial defects in RNA binding and was insufficient for viral mRNA export[40,109].

CHAPTER 5

Discussion

The studies in this dissertation have addressed several unresolved questions regarding ICP27 structure and activity. In our first study (Chapter 2), we probed the relationship between ICP27's conformational isomers. Through the analysis of targeted mutants by BiFC and co-ip-western blot experiments, we determined protein domains essential for assembly of the ICP27 homo-dimer are distinct from the domains necessary for N- to C-terminal contacts. ICP27 dimer formation require the C- terminal tail residues while the N-terminal LRR and C-terminal ZFD must be intact for the intramolecular interaction. We hypothesize that the reversible N- to C-terminal interactions likely occur as a monomer and that the interchange between the dimeric and N- to C-conformations could serve as a means of regulating some of ICP27's functional interactions. Upon reexamination of previous studies, we identified several protein partners specific for each conformation and are currently working to chart the full array of interaction partners by mass spectrometry in order to better understand the significance and activity of both configurations.

A large number of studies have characterized ICP27's various activities and interactions, however, relatively little is known about how these activities are regulated and carried out. We've shown that phosphorylation and arginine methylation regulate some of ICP27's protein interactions[38,47,49]. Both modifications cause local structural changes which alters ICP27's affinity for different binding partners. Similarly,

intramolecular N- to C-terminal interaction and dimer formation would also create structural changes, creating binding pockets or interaction sites that would likely be specific for different interaction partners. Thus, factors that regulate the transition between the conformations would regulate ICP27 activity. Protein oligomerization has been demonstrated to be a significant regulatory mechanism[110,111]. As dimerization occurs only when the protein concentration is higher than the K_d for dimerization, this would allow for a simple mechanism for sensing protein concentration; and possibly provide a regulatory switch in activity[111]. Additionally, oligomeric proteins can be allosterically regulated, and allosteric control of the ICP27 dimer may allow for further regulatory flexibility[110]. The vast amount of protein interactions that ICP27 undergoes is undeniably linked to its structure and its ability to both interact with itself intramolecularly and dimerize. Future studies will investigate specific factors responsible for mediating these changes.

In our second study (Chapter 3), we addressed the mechanism of SRPK1 inactivation by ICP27. During HSV-1 infection, we previously reported that SRPK1 is relocalized from the cytoplasm to the nucleus and its activity is inhibited through its interactions with ICP27[23]. However, the details for how this occurs and the exact role of the ICP27 RGG box was unclear. In a collaboration with Dr. Alexander Golovanov, we performed a series of biophysical and *in vivo* experiments to demonstrate that an isolated peptide containing the ICP27 RGG box is able to bind to SRPK1 with an affinity comparable to that of its native substrate, SRSF1[74]. Analysis of the ICP27 RGG box-SRPK1 crystal structure revealed a short palindromic RGRRRGR sequence within the center of the RGG box binds to the SRPK1 substrate docking groove in a manner exceedingly similar to the

SRSF1 RS repeats. Targeted arginine mutations within the palindrome residues confirmed their importance for interaction, relocalization and inhibition of SRPK1.

Several studies have reported ICP27 participates in alternative splicing during infection, preventing removal of select cellular introns in a gene-specific manner[24,94,112]. RNA-sequencing of cells infected with WT HSV-1 or transiently transfected with WT ICP27 protein revealed ICP27 promotes usage of alternative 5' splice sites and inhibits splicing of certain introns in approximately 100 cellular genes[24]. Due to the similarity of the targeted genes to viral transcripts, (high GC-content, C-rich sequences near 5' splice site) it was speculated that these host mRNAs may escape the selective degradation of the viral ribonuclease, virion host shutoff (vhs). Thus, by specifically targeting virus-like cellular transcripts likely spared by the vhs protein, ICP27 contributes to virus-induced host shutoff necessary for efficient virus production. Although the molecular details for this action has, as yet, not been reported, the ICP27-SRPK1 interaction is expected to contribute to ICP27's role as an alternative splicing regulator. Moreover, ICP27-mediated disruption of host spliceosome assembly may be important for proper viral gene expression as well. A recent high-throughput sequencing analysis of the viral transcriptome in HSV-1 infected cells identified hundreds of previously unknown alternative splice junctions, the majority of which alter the coding potential of viral genes[89]. However, splicing of these novel splice sites was only efficient during infection with an HSV-1 ICP27 deletion mutant, indicating ICP27 is required for splicing inhibition and identifying an important role for the protein in viral gene expression[89].

In our last study (Chapter 4), we report ICP27 expression is required for accumulation of cellular antisense transcripts during infection; a novel finding. ICP27 has

been widely reported to alter the cellular gene expression program, disrupting multiple aspects of host mRNA biogenesis in order to promote virus production[24,25]. Many of these activities have been shown to involve the ICP27 RGG box. Yet, the impact of ICP27 and its RGG box in regulating host gene expression is still not well understood. Furthermore, a recent report revealed increased host antisense RNA levels due to the activity of ICP4[87]. To determine if host antisense RNAs were similarly elevated within our conditions and to clarify ICP27's activities within the host transcriptome, we performed RNA-seq and iCLIP-seq for WT HSV-1 and ΔRGG mutant infected cells. Our results confirmed activation of host antisense RNAs during WT HSV-1 infection. In contrast, upon deletion of ICP27 or the RGG box, levels of antisense transcripts were greatly diminished. Moreover, iCLIP-seq data indicate ICP27 selectively interacts with the antisense transcripts. Given these results, we hypothesize ICP27 may bind to these antisense RNAs and promotes their stability. However, the molecular mechanism underlying this action is, presently, unclear.

Antisense transcripts are key regulators of protein-coding transcripts; modulating expression of overlapping or neighboring genes through multiple mechanisms[103]. Antisense transcripts have been shown to function at all levels of gene expression: they can affect transcription initiation through transcriptional interference, DNA methylation or histone modifications; they can act co-transcriptionally, by mediating RNAP II collisions and formation of secondary DNA structures or affecting mRNA splicing by masking specific splice sites; and they can function at the post-transcriptional level, masking miRNA binding sites and altering translation efficiency[106,103]. Given the biological significance of antisense RNAs to host cell function, it's likely that HSV-1 targets this phenomenon to promote its own life cycle. ICP27 has been widely reported to disrupt host gene expression, targeting

many aspects of mRNA maturation; including splicing inhibition. Yet, a number of studies have indicated that, in some instances, ICP27 inhibits splicing in a gene- or sequence-specific manner that cannot be explained by current, known mechanisms. It's likely that, for some of these cases, regulatory activity by host antisense transcripts may be involved.

1. Kramer, T. & Enquist, L. W. Directional Spread of Alphaherpesviruses in the Nervous System. *Viruses* **5**, 678–707 (2013).

2. Davison, A. J. Herpesviruses: General Features☆. in *Reference Module in Biomedical Sciences* (Elsevier, 2014). doi:10.1016/B978-0-12-801238-3.04812-1.

3. Jung, J. U. & Speck, S. H. Insights into chronic gamma-herpesvirus infections. *Curr. Opin. Virol.* **3**, 225–226 (2013).

4. Wald, A. & Corey, L. Persistence in the population: epidemiology, transmission. in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis* (eds. Arvin, A. et al.) (Cambridge University Press, 2007).

5. Sodeik, B., Ebersold, M. W. & Helenius, A. Microtubule-mediated Transport of Incoming Herpes Simplex Virus 1 Capsids to the Nucleus. *J. Cell Biol.* **136**, 1007–1021 (1997).

6. Boehmer, P. & Nimonkar, A. Herpes Virus Replication. *IUBMB Life* **55**, 13–22 (2003).

7. Kibler, P. K., Duncan, J., Keith, B. D., Hupel, T. & Smiley, J. R. Regulation of herpes simplex virus true late gene expression: sequences downstream from the US11 TATA box inhibit expression from an unreplicated template. *J. Virol.* **65**, 6749–6760 (1991).

8. Nicoll, M. P., Proença, J. T. & Efstathiou, S. The molecular basis of herpes simplex virus latency. *FEMS Microbiol. Rev.* **36**, 684–705 (2012).

9. Dai-Ju, J. Q., Li, L., Johnson, L. A. & Sandri-Goldin, R. M. 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. *J. Virol.* **80**, 3567–3581 (2006).

10. Exporting RNA from the nucleus to the cytoplasm | Nature Reviews Molecular Cell Biology. https://www.nature.com/articles/nrm2255.

11. Carmody, S. R. & Wente, S. R. mRNA nuclear export at a glance. *J. Cell Sci.* **122**, 1933–1937 (2009).

12. Sandri-Goldin, R. M. The many roles of the highly interactive HSV protein ICP27, a key regulator of infection. *Future Microbiol.* **6**, 1261–1277 (2011).

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

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

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

16. Fontaine-Rodriguez, E. C. & Knipe, D. M. Herpes Simplex Virus ICP27 Increases Translation of a Subset of Viral Late mRNAs. *J. Virol.* **82**, 3538–3545 (2008).

17. Larralde, O. *et al.* Direct Stimulation of Translation by the Multifunctional Herpesvirus ICP27 Protein. *J. Virol.* **80**, 1588–1591 (2006).

18. Rivas, H. G., Schmaling, S. K. & Gaglia, M. M. Shutoff of Host Gene Expression in Influenza A Virus and Herpesviruses: Similar Mechanisms and Common Themes. *Viruses* **8**, (2016).

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

20. Rutkowski, A. J. *et al.* Widespread disruption of host transcription termination in HSV-1 infection. *Nat. Commun.* **6**, 1–15 (2015).

21. Sandri-Goldin, R. M. & Hibbard, M. K. The herpes simplex virus type 1 regulatory protein ICP27 coimmunoprecipitates with anti-Sm antiserum, and the C terminus appears to be required for this interaction. *J. Virol.* **70**, 108–118 (1996).

22. Sandri-Goldin, R. M., Hibbard, M. K. & Hardwicke, M. A. The C-terminal repressor region of herpes simplex virus type 1 ICP27 is required for the redistribution of small nuclear ribonucleoprotein particles and splicing factor SC35; however, these alterations are not sufficient to inhibit host cell splicing. *J. Virol.* **69**, 6063–6076 (1995).

23. Sciabica, K. S., Dai, Q. J. & Sandri-Goldin, R. M. ICP27 interacts with SRPK1 to mediate HSV splicing inhibition by altering SR protein phosphorylation. *EMBO J.* **22**, 1608–1619 (2003).

24. Tang, S., Patel, A. & Krause, P. R. Herpes simplex virus ICP27 regulates alternative pre-mRNA polyadenylation and splicing in a sequence-dependent manner. *Proc. Natl. Acad. Sci.* **113**, 12256–12261 (2016).

25. Wang, X. *et al.* Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nat. Commun.* **11**, 1–13 (2020).

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

27. Li, L., Johnson, L. A., Dai-Ju, J. Q. & Sandri-Goldin, R. M. Hsc70 Focus Formation at the Periphery of HSV-1 Transcription Sites Requires ICP27. *PLOS ONE* **3**, e1491 (2008).

28. Herpes Simplex Virus Gene Products Required for Viral Inhibition of Expression of G1-Phase Functions | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0042682201911754?token=41C6409D9BE2D3242CCAFEDA173CFD3FB723C2F6AA62045F13AAD5A29A8DCA569493B0BD628BB1AB490312CBD58E4F00 doi:10.1006/viro.2001.1175.

29. Hargett, D., Rice, S. & Bachenheimer, S. L. Herpes Simplex Virus Type 1 ICP27-Dependent Activation of NF-κB. *J. Virol.* **80**, 10565–10578 (2006).

30. HSV-1 ICP27 suppresses NF-κB activity by stabilizing IκBα - Kim - 2008 - FEBS Letters - Wiley Online Library. https://febs.onlinelibrary.wiley.com/doi/full/10.1016/j.febslet.2008.05.044.

31. Johnson, K. E., Song, B. & Knipe, D. M. Role for Herpes Simplex Virus 1 ICP27 in the Inhibition of Type I Interferon Signaling. *Virology* **374**, 487–494 (2008).

32. Christensen, M. H. *et al.* HSV-1 ICP27 targets the TBK1-activated STING signalsome to inhibit virus-induced type I IFN expression. *EMBO J.* **35**, 1385–1399 (2016).

33. Hargett, D., McLean, T. & Bachenheimer, S. L. Herpes Simplex Virus ICP27 Activation of Stress Kinases JNK and p38. *J. Virol.* **79**, 8348–8360 (2005).

34. Aubert, M. & Blaho, J. A. The Herpes Simplex Virus Type 1 Regulatory Protein ICP27 Is Required for the Prevention of Apoptosis in Infected Human Cells. *J. Virol.* **73**, 2803–2813 (1999).

35. Gillis, P. A., Okagaki, L. H. & Rice, S. A. Herpes Simplex Virus Type 1 ICP27 Induces p38 Mitogen-Activated Protein Kinase Signaling and Apoptosis in HeLa Cells. *J. Virol.* **83**, 1767–1777 (2009).

36. Singh, V. K. *et al.* Intrinsically disordered human C/EBP homologous protein regulates biological activity of colon cancer cells during calcium stress. *J. Mol. Biol.* **380**, 313–326 (2008).

37. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).

38. Corbin-Lickfett, K. A., Rojas, S., Li, L., Cocco, M. J. & Sandri-Goldin, R. M. 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. *J. Virol.* **84**, 2212–2222 (2010).

39. Tunnicliffe, R. B. *et al.* The structure of the folded domain from the signature multifunctional protein ICP27 from herpes simplex virus-1 reveals an intertwined dimer. *Sci. Rep.* **5**, 1–15 (2015).

40. Corbin-Lickfett, K. A., Souki, S. K., Cocco, M. J. & Sandri-Goldin, R. M. 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. *J. Virol.* **84**, 6367–6376 (2010).

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

42. Tian, X., Devi-Rao, G., Golovanov, A. P. & Sandri-Goldin, R. M. The Interaction of the Cellular Export Adaptor Protein Aly/REF with ICP27 Contributes to the Efficiency of Herpes Simplex Virus 1 mRNA Export. *J. Virol.* **87**, 7210–7217 (2013).

43. Hernandez, F. P. & Sandri-Goldin, R. M. 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**, (2010).

44. Hernandez, F. P. & Sandri-Goldin, R. M. Herpes Simplex Virus 1 Regulatory Protein ICP27 Undergoes a Head-to-Tail Intramolecular Interaction. *J. Virol.* **84**, 4124–4135 (2010).

45. Patel, V. *et al.* Structure of the C-Terminal Domain of the Multifunctional ICP27 Protein from Herpes Simplex Virus 1. *J. Virol.* **89**, 8828–8839 (2015).

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

47. Souki, S. K., Gershon, P. D. & Sandri-Goldin, R. M. Arginine Methylation of the ICP27 RGG Box Regulates ICP27 Export and Is Required for Efficient Herpes Simplex Virus 1 Replication. *J. Virol.* **83**, 5309–5320 (2009).

48. Yu, J. *et al.* Protein arginine methyltransferase 1 regulates herpes simplex virus replication through ICP27 RGG-box methylation. *Biochem. Biophys. Res. Commun.* **391**, 322–328 (2010).

49. Rojas, S., Corbin-Lickfett, K. A., Escudero-Paunetto, L. & Sandri-Goldin, R. M. ICP27 Phosphorylation Site Mutants Are Defective in Herpes Simplex Virus 1 Replication and Gene Expression. *J. Virol.* **84**, 2200–2211 (2010).

50. Sedlackova, L. & Rice, S. A. Herpes Simplex Virus Type 1 Immediate-Early Protein ICP27 Is Required for Efficient Incorporation of ICP0 and ICP4 into Virions. *J. Virol.* **82**, 268–277 (2008).

51. Self-Interaction of the Herpes Simplex Virus Type 1 Regulatory Protein ICP27 | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S004268229999698 8?token=94A99B23CF3822F6D495CBCC5CD3E00CF65119898AD57DB7EFD4C985638845A549C5E72962BC331477B3497F611A45E2.

52. Johnson, L. A., Li, L. & Sandri-Goldin, R. M. The Cellular RNA Export Receptor TAP/NXF1 Is Required for ICP27-Mediated Export of Herpes Simplex Virus 1 RNA, but the TREX Complex Adaptor Protein Aly/REF Appears To Be Dispensable. *J. Virol.* **83**, 6335–6346 (2009).

53. Hibbard, M. K. & Sandri-Goldin, R. M. Arginine-rich regions succeeding the nuclear localization region of the herpes simplex virus type 1 regulatory protein ICP27 are required for efficient nuclear localization and late gene expression. *J. Virol.* **69**, 4656–4667 (1995).

54. Hardwicke, M. A., Vaughan, P. J., Sekulovich, R. E., O'Conner, R. & Sandri-Goldin, R. M. The regions important for the activator and repressor functions of herpes simplex virus type 1 alpha protein ICP27 map to the C-terminal half of the molecule. *J. Virol.* **63**, 4590–4602 (1989).

55. 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. https://www.jbc.org/content/275/42/32822.

56. Taylor, A. *et al.* Mutation of a C-Terminal Motif Affects Kaposi's Sarcoma-Associated Herpesvirus ORF57 RNA Binding, Nuclear Trafficking, and Multimerization. *J. Virol.* **85**, 7881–7891 (2011).

57. Yuan, F. *et al.* The crystal structure of KSHV ORF57 reveals dimeric active sites important for protein stability and function. *PLOS Pathog.* **14**, e1007232 (2018).

58. Tunnicliffe, R. B., Collins, R. F., Nivia, H. D. R., Sandri-Goldin, R. M. & Golovanov, A. P. The ICP27 Homology Domain of the Human Cytomegalovirus Protein UL69 Adopts a Dimer-of-Dimers Structure. *mBio* **9**, (2018).

59. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).

60. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

61. Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**, 15–27 (2009).

62. Shepard, P. J. & Hertel, K. J. The SR protein family. *Genome Biol.* **10**, 242 (2009).

63. Zhou, Z. & Fu, X.-D. Regulation of Splicing by SR proteins and SR Protein-Specific Kinases. *Chromosoma* **122**, 191–207 (2013).

64. Aubol, B. E. *et al.* Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol. Cell* **63**, 218–228 (2016).

65. Ghosh, G. & Adams, J. A. Phosphorylation Mechanism and Structure of Serine-Arginine Protein Kinases. *FEBS J.* **278**, 587–597 (2011).

66. Velazquez-Dones, A. *et al.* Mass Spectrometric and Kinetic Analysis of ASF/SF2 Phosphorylation by SRPK1 and Clk/Sty. *J. Biol. Chem.* **280**, 41761–41768 (2005).

67. Zhong, X.-Y., Ding, J.-H., Adams, J. A., Ghosh, G. & Fu, X.-D. Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. *Genes Dev.* **23**, 482–495 (2009).

68. Wang, H.-Y. *et al.* SRPK2: A Differentially Expressed SR Protein-specific Kinase Involved in Mediating the Interaction and Localization of Pre-mRNA Splicing Factors in Mammalian Cells. *J. Cell Biol.* **140**, 737–750 (1998).

69. Boudreault, S., Roy, P., Lemay, G. & Bisaillon, M. Viral modulation of cellular RNA alternative splicing: A new key player in virus–host interactions? *WIREs RNA* **10**, e1543 (2019).

70. Hu, B. *et al.* Cellular responses to HSV-1 infection are linked to specific types of alterations in the host transcriptome. *Sci. Rep.* **6**, 1–14 (2016).

71. Phelan, A., Carmo-Fonseca, M., McLaughlan, J., Lamond, A. I. & Clements, J. B. A herpes simplex virus type 1 immediate-early gene product, IE63, regulates small nuclear ribonucleoprotein distribution. *Proc. Natl. Acad. Sci.* **90**, 9056–9060 (1993).

72. Splicing Inhibition at the Level of Spliceosome Assembly in the Presence of Herpes Simplex Virus Protein ICP27 | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0042682201913017?token=AC198E02FB41A50FA83AEC6167C303B4AB09C287C6EF49E756BBCBBC6E68CF1A93B25D76438097A8BBFA370D5CD10513 doi:10.1006/viro.2001.1301.

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

74. Tunnicliffe, R. B. *et al.* Molecular Mechanism of SR Protein Kinase 1 Inhibition by the Herpes Virus Protein ICP27. *mBio* **10**, (2019).

75. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).

76. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

77. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

78. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

79. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

80. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).

81. Tunnicliffe, R. B., Tian, X., Storer, J., Sandri-Goldin, R. M. & Golovanov, A. P. Overlapping motifs on the herpes viral proteins ICP27 and ORF57 mediate interactions with the mRNA export adaptors ALYREF and UIF. *Sci. Rep.* **8**, 1–12 (2018).

82. Golovanov, A. P., Blankley, R. T., Avis, J. M. & Bermel, W. Isotopically Discriminated NMR Spectroscopy: A Tool for Investigating Complex Protein Interactions in Vitro. *J. Am. Chem. Soc.* **129**, 6528–6535 (2007).

83. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

84. Wilkins, M. R. *et al.* Protein Identification and Analysis Tools in the ExPASy Server. in *2-D Proteome Analysis Protocols* (ed. Link, A. J.) 531–552 (Humana Press, 1999). doi:10.1385/1-59259-584-7:531.

85. Giannakouros, T., Nikolakaki, E., Mylonis, I. & Georgatsou, E. Serine-arginine protein kinases: a small protein kinase family with a large cellular presence. *FEBS J.* **278**, 570–586 (2011).

86. Patel, M., Sachidanandan, M. & Adnan, M. Serine arginine protein kinase 1 (SRPK1): a moonlighting protein with theranostic ability in cancer prevention. *Mol. Biol. Rep.* **46**, 1487–1497 (2019).

87. Wyler, E. *et al.* Widespread activation of antisense transcription of the host genome during herpes simplex virus 1 infection. *Genome Biol.* **18**, (2017).

88. Sandri-Goldin, R. M. The many roles of the regulatory protein ICP27 during herpes simplex virus infection. *Front. Biosci. J. Virtual Libr.* **13**, 5241–5256 (2008).

89. Tang, S., Patel, A. & Krause, P. R. Hidden regulation of herpes simplex virus 1 pre-mRNA splicing and polyadenylation by virally encoded immediate early gene ICP27. *PLoS Pathog.* **15**, (2019).

90. Sedlackova, L. *et al.* Identification of an ICP27-Responsive Element in the Coding Region of a Herpes Simplex Virus Type 1 Late Gene. *J. Virol.* **84**, 2707–2718 (2010).

91. Huppertz, I. *et al.* iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).

92. Modic, M. *et al.* Cross-Regulation between TDP-43 and Paraspeckles Promotes Pluripotency-Differentiation Transition. *Mol. Cell* **74**, 951-965.e13 (2019).

93. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

94. Nojima, T. *et al.* Herpesvirus protein ICP27 switches PML isoform by altering mRNA splicing. *Nucleic Acids Res.* **37**, 6515–6527 (2009).

95. Hardwicke, M. A. & Sandri-Goldin, R. M. The herpes simplex virus regulatory protein ICP27 contributes to the decrease in cellular mRNA levels during infection. *J. Virol.* **68**, 4797–4810 (1994).

96. Hardy, W. R. & Sandri-Goldin, R. M. Herpes simplex virus inhibits host cell splicing, and regulatory protein ICP27 is required for this effect. *J. Virol.* **68**, 7790–7799 (1994).

97. Hikisz, P. & Kiliańska, Z. M. PUMA, a critical mediator of cell death--one decade on from its discovery. *Cell. Mol. Biol. Lett.* **17**, 646–669 (2012).

98. Guérillon, C., Larrieu, D. & Pedeux, R. ING1 and ING2: multifaceted tumor suppressor genes. *Cell. Mol. Life Sci.* **70**, 3753–3772 (2013).

99. Aftab, S., Semenec, L., Chu, J. S.-C. & Chen, N. Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evol. Biol.* **8**, 226 (2008).

100. Feng, C. *et al.* Regulatory factor X1 is a new tumor suppressive transcription factor that acts via direct downregulation of CD44 in glioblastoma. *Neuro-Oncol.* **16**, 1078–1085 (2014).

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

102. Sokolowski, M., Scott, J. E., Heaney, R. P., Patel, A. H. & Clements, J. B. Identification of Herpes Simplex Virus RNAs That Interact Specifically with Regulatory Protein ICP27 in Vivo. *J. Biol. Chem.* **278**, 33540–33549 (2003).

103. Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893 (2013).

104. Barman, P., Reddy, D. & Bhaumik, S. R. Mechanisms of Antisense Transcription Initiation with Implications in Gene Expression, Genomic Integrity and Disease Pathogenesis. *Non-Coding RNA* **5**, (2019).

105. Ozsolak, F. *et al.* Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation. *Cell* **143**, 1018–1029 (2010).

106. Nair, L., Chung, H. & Basu, U. Regulation of long non-coding RNAs and genome dynamics by the RNA surveillance machinery. *Nat. Rev. Mol. Cell Biol.* **21**, 123–136 (2020).

107. Wanowska, E., Kubiak, M. R., Rosikiewicz, W., Makałowska, I. & Szcześniak, M. W. Natural antisense transcripts in diseases: From modes of action to targeted therapies. *Wiley Interdiscip. Rev. RNA* **9**, (2018).

108. Dremel, S. E. & DeLuca, N. A. Herpes simplex viral nucleoprotein creates a competitive transcriptional environment facilitating robust viral transcription and host shut off. *eLife* **8**, e51109 (2019).

109. Johnson, L. A. & Sandri-Goldin, R. M. Efficient Nuclear Export of Herpes Simplex Virus 1 Transcripts Requires both RNA Binding by ICP27 and ICP27 Interaction with TAP/NXF1. *J. Virol.* **83**, 1184–1192 (2009).

110. Ali, M. H. & Imperiali, B. Protein oligomerization: How and why. *Bioorg. Med. Chem.* **13**, 5013–5020 (2005).

111. Marianayagam, N. J., Sunde, M. & Matthews, J. M. The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **29**, 618–625 (2004).

112. Ellison, K. S., Rice, S. A., Verity, R. & Smiley, J. R. Processing of α-Globin and ICP0 mRNA in Cells Infected with Herpes Simplex Virus Type 1 ICP27 Mutants. *J. Virol.* **74**, 7307–7319 (2000).