UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Medicina Molecolare

CORSO DI DOTTORATO DI RICERCA IN: Medicina Molecolare
(EVENTUALE) CURRICOLO: Biomedicina
CICLO XXXIV

PART A:

Divide et Impera: An In Silico Screening Targeting HCMV ppUL44 Processivity Factor Homodimerization Identifies Small Molecules Inhibiting Viral Replication

PART B:

Identification of the nuclear proteome from all human viruses by a comprehensive analysis of classical nuclear localizations

Coordinatore: Ch.mo Prof. Riccardo Manganelli
Supervisore: Ch.mo Prof. Gualtiero Alvisi

Dottoranda: Hanieh Ghassabian Gilan

INDEX

Part A:

Divide et impera: An In Silico Screening Targeting HCMV ppUL44 Processivity Factor Homodimerization Identifies Small Molecules Inhibiting Viral Replication

A.1 ABSTRACT

Human cytomegalovirus (HCMV) is a leading cause of severe diseases in immunocompromised individuals, including AIDS patients and transplant recipients, and in congenitally infected newborns. The utility of available drugs is limited by poor bioavailability, toxicity, and emergence of resistant strains. Therefore, it is crucial to identify new targets for therapeutic intervention. Among the latter, viral protein–protein interactions are becoming increasingly attractive. Since dimerization of HCMV DNA polymerase processivity factor ppUL44 plays an essential role in the viral life cycle, being required for oriLyt-dependent DNA replication, it can be considered a potential therapeutic target. We therefore previously performed an in silico screening and selected 18 small molecules (SMs) potentially interfering with ppUL44 homodimerization. Antiviral assays using recombinant HCMV TB4-UL83-YFP in the presence of the selected SMs led to the identification of four active compounds. In this work I have characterized the effect of such compounds on cell viability and growth and began a preliminary analysis of their mode of action. All of them impaired replication of an AD169-GFP reporter virus and its ganciclovir-resistant counterpart to a similar extend. Among the 4 selected SMs compound B3 exhibited the highest selectivity index (SI) and was further investigated. We could show that it also efficiently inhibited HCMV AD169 strain in plaque reduction assays (PRAs). As assessed by qPCR by Western blotting experiments, B3 specifically reduced viral DNA synthesis starting from 72 h post infection, consistent with the inhibition of viral gene expression starting from 48 h post infection by Western blotting experiments. Therefore, our data suggest that inhibition

of ppUL44 dimerization could represent a new class of HCMV inhibitors, complementary to those targeting the DNA polymerase catalytic subunit or the viral terminase complex.

RIASSUNTO

Human cytomegalovirus (HCMV) è un agente patogeno principale di molte malattie in persone immunosoppresse, inclusi pazienti affetti da AIDS e sottoposti a trapianto, e nascituri congenitamente infetti. Le terapie ed i farmaci antivirali utilizzati per il trattamento dell'infezioni da HCMV presentano una serie di limitazioni, tra cui la bassa biodisponibilità, tossicità, e l'insorgenza di ceppi virali farmaco resistenti, rendendo cruciale la necessità di identificare nuovi target terapeutici efficaci. Studi sulle interazioni tra proteine virali (PPI) si sono rivelati alleati importanti per lo sviluppo di nuovi farmaci antivirali, in quanto questi ultimi possono inibire il ciclo vitale del virus interferendo con le attività delle proteine virali. La dimerizzazione del fattore di processività della DNA polimerasi, ppUL44, di HCMV è essenziale per il ciclo vitale del virus infatti necessaria per la replicazione del DNA virale mediata da oriLyt e può essere quindi considerata come un potenziale target terapeutico. Pertanto, in precedenza, tramite uno screening in silico sono state identificate 18 piccole molecole (Small Molecules, SMs) potenzialmente capaci di interferire con la omodimerizazzione di ppUL44. Saggi antivirali delle 18 SMs sul virus ricombinante TB4-UL83-YFP di HCMV hanno permesso l'identificazione dei quattro composti più attivi. In questo lavoro sono riuscita a caratterizzare l'effetto di questi composti sulla viabilità e crescita cellulare, e quindi cominciare un'analisi preliminare del loro meccanismo di azione. Tutte hanno compromesso la replicazione dei virus reporter AD169 di HCMV ed una sua controparte resistente al GCV in maniera simile. Tra le 4 SMs scelte, B3, ha mostrato il Selectivity Index (SI) più alto e quindi l'unico composto ad essere ulteriormente analizzato. Siamo riusciti a dimostrare che B3 efficientemente inibisce lo strain virale AD169 di HCMV in

saggi di Plaque Reduction (PRA). Come misurato tramite qPCR, B3 ha specificatamente inibito la sintesi del DNA virale a partire da 72 ore post infezione, come anche l'espressione dei geni virali a partire da 48 ore post infezione da analisi di Western Blotting, Pertanto, i nostri dati indicano che l'inibizione della dimerizzazione di ppUL44 può rappresentare una nuova classe di inibitori di HCMV, complementarmente a quelli che hanno come target la subunità catalitica della DNA polimerasi o il complesso virale della terminasi.

A.2 INTRODUCTION

A.2.1 The History of the Discovery of Human Cytomegalovirus (HCMV)

In the year of 1881 a German pathologist described still born babies with syphilis-like symptoms which is believed to be the first observation of Cytomegalovirus (CMV). In 1904, Jesionek and Kiolemenoglou described these enlarged cells as 'protozoan-like' finding them also in lung and liver cells. Löwenstein described cytoplasmic and nuclear inclusions in these protozoal-like cells (Ho, 2008);(Riley, 1997);(Weller, 2000)and in 1907 Goodpasture and Talbot termed them as "Cytomegalia". Von Glahn and Pappenheimer observed that cells infected with herpesviruses contained inclusion bodies, suggesting that cytomegalic cells were more likely caused by a virus rather than protozoa (Vonglahn and Pappenheimer, 1925) hypothesis confirmed also by other researchers (Wyatt and Saxton, 1950), agreeing to a virus being likely the cause of the pathology of these cells. In the mid of the 1950s, different laboratories reported to have isolated a virus from tissue cultures of human adenoid and salivary gland and named them AD169 (Rowe *et al.*, 1956; Smith, 1956) and thereafter it was as 'cytomegalovirus'(CMV) (Craig *et al.*, 1957). The isolation and propagation of the virus in cultures enabled further development of understanding of the nature of this virus, such as detection of virus protein, its life cycle and later its molecular pathogenesis.

A.2.2 Human Cytomegalovirus (HCMV)

HCMV is a herpes virus belonging to the ß-Herpesviridae subfamily. All member of the Herpesviridae family share common features where their 120 – 240 kbp linear double-stranded DNA (dsDNA) is enclosed within an icosahedral capsid (100 -130 nm external diameter). The capsid is made of 162 hollows centered capsomeres (12 pentons and 150 hexons) and is surrounded by an asymmetric tegument protein layer, which is further

surrounded by a host-derived lipid bilayer membrane called the envelope. The envelope is covered with viral and host transmembrane proteins on its surface, which mediates binding and entry to the next target cell (Figure A.2.1).

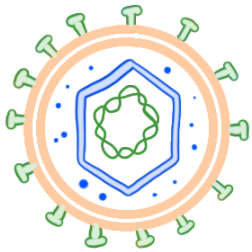


Figure A.2.1 Exemplified Herpesvidae viral structure. All herpesviruses share a similar viral structure. It comprehends a dsDNA within an icosahedral capsid, which is surrounded by viral tegument proteins. An enveloped covers all viral layers, viral and transmembrane proteins are embedded on its surface.

The four most common biological properties shared by all Herpesviridae include:

1. Encoding enzymes involved in nucleic acid metabolism, DNA synthesis and protein processing such as helicases, DNA polymerases, primases, protease, and thymidine kinases, although the viral protein number vary among the family species.

2. DNA synthesis and capsid assembly take place in the nucleus, while maturation of the virion occurs in the cytoplasm of the infected cells.

3. The spread of new viral progeny can occur through multiple mechanisms including lysis of the infected host cell, release of intracellular vesicles containing mature virions, or cytoplasmic bridges from one cell to another.

4. The ability to establish latency in their natural host, from which the virus can be reactivated. The host cell types in which they establish latency vary among the different viruses. For example, Herpes Simplex Virus-1 and -2 (HSV) and Varicella-Zoster Virus (VZV) remain latent in neuronal cells of dorsal root ganglia (Kramer *et al.*, 2003)). EBV

latency is found in B-lymphocytes (Hurley and Thorley-Lawson, 1988) and HCMV establishes latency in myeloid lineage cells (Söderberg-Nauclér *et al.*, 2001).

Although all herpesviruses share common features they also differ in several proprieties, for example they vary from 70 to 250 in the number of open reading frames (ORF) in the genome (VZV - CMV) and in many biological properties such as host cell tropism, latency features, replication cycle, and they also differ in their clinical manifestations.

Herpesviridae family members are further classified into three major subfamilies: α-Herpesvirinae, ß-Herpesvirinae and γ-Herpesvirinae. Among all known herpesviruses, there are nine viruses that are known to primarily infect humans. There are three **α**-herpesviruses that infect humans: HSV-1 and HSV-2, and VZV. The replication cycle of these viruses is very short, about 12-18 h and the host cells lyses is needed for the production of new viral progeny. They all establish a latent infection in sensory ganglia. The ß-Herpesvirinae includes four human viruses: HCMV, Human Herpes virus (HHV) 6A, HHV6B and HHV7. Infected cells frequently become enlarged in size (cytomegalic cells) and those viruses can establish latent infection in lymphoreticular cells, secretory glands, and bone marrow cells. In general, members of this sub-family have a long replication cycle (> 24 h) and productive infection does not necessarily require in cell lysis. The third subfamily is the γ-Herpesvirinae. Members of this subfamily infect and replicate in lymphoblastoid cells such as T- and B- lymphocytes whereby they can establish latency. They can also infect and cause lytic infection of epithelioid/fibroblast cells. There are two human viruses in this subfamily, and they are both endowed with oncogenic potential: Eptesin-Barr Virus/ HHV4, which is associated with different types of lymphomas and, Kaposi's sarcoma-associated herpesvirus/ HHV8 which is associated with Kaposis sarcoma in AIDS patients.

A.2.3 HCMV Genome Organization and Life Cycle

HCMV genome consists of 235 kbp, 252 ORFs and encodes for more than 750 proteins (Stern-Ginossar *et al.*, 2012) suggesting a complex coding capacity, moreover HCMV genome it's the largest among all human herpesviruses and has two major segments named as Unique Long (UL) and Unique Short (US), each flanked by Terminal Repeated Long (TRL) and Terminal Repeated Short (TRS) segments. US and UL segments are linked together with the Internal Repeated Long (IRL) segment. Roizman and Pellett (Davison *et al.*, 2009) classified herpesviruses genome into A to F classes, as shown in Fig. A.2.2, and HCMV is classified as E genome.

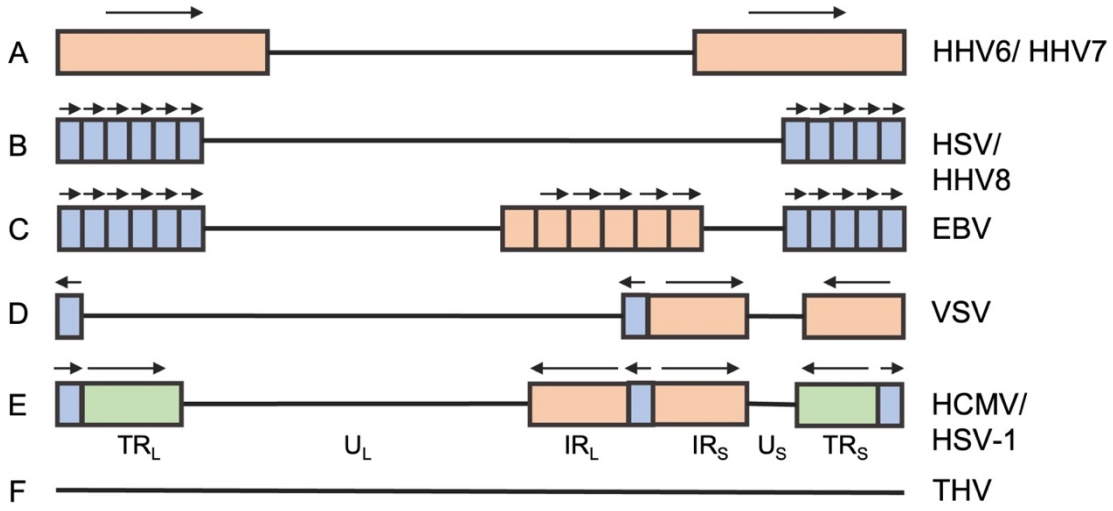


Figure A.2.2 Classes of herpesvirus genome structures (not to scale) defined by Roizman and Pellett (2007). Arrows = Orientations of repeats; Horizontal lines = Unique regions; Rectangles = Repeated regions.

Gene annotations are based on their respective location on the genome segments. In addition to protein-coding genes, HCMV genome also bears genes encoding for polyadenylated non-coding RNAs and non-polyadenylated RNAs such as micro-RNAs, which play a regulatory role during infection (Puente *et al.*, 1989).

HCMV viral life cycle consists of 7 stages: Adhesion, Entry, Uncoating, Genome Transcription, Replication, Assembly and Egress (Fig. A.2.3).

1. HCMV entry is preceded with the interaction of the viral surface to the host cell membrane, by a mechanism involving viral glycoproteins the gH/gL complex and gB, where the latter mediates attachment of cells via binding to cellular receptors including heparan sulphate proteoglycan, integrins and EGFR to promote the entry process.

2. HCMV enters cells by two different entry routes according to the type of infected cell (Compton, Nepomuceno and Nowlin, 1992; Ryckman *et al.*, 2006). In fibroblast cells viral entry is mediated by viral glycoprotein complexes and occurs via direct fusion at the plasma membrane, whereas in monocytes as well as endothelial, epithelial and dendritic cells, the viral complex composed by gH/gL/UL128/UL130/UL131A triggers receptor-mediated endocytosis. The entry is followed by uncoating nucleocapsid tegument proteins access in the cytoplasm.

3. Once the viral capsid enters the cell, some tightly associated tegument proteins (pUL47 and pUL48) mediate the delivery of the capsid along microtubules to the NPC, through which the viral genomic DNA is released in the host cell nucleus where gene expression takes place (Isaacson, Juckem and Compton, 2008).

4. Viral genome transcription is temporary regulated can be exemplified in three stages, even though is more complex. HCMV genes are transcribed from cellular RNA polymerase II upon stimulation by viral trans activators, and the first genes to be transcribed are defined as immediate-early (IE) genes which lead to the expression of IE proteins. IE proteins are mainly trans-activators which regulate host and viral gene

expression. They are required to initiate the transcription of viral genes which express early (E) proteins. E proteins are essential for the initiation of HCMV viral DNA replication and expression of late (L) genes, which encode for L proteins which are devolved in viral morphogenesis, virion assembly and maturation.

5. The replication of HCMV DNA replication initiates form DNA replication origin (oriLyt) and it is mediated by the viral DNA polymerase complex (formed by ppUL44 and pUL54 proteins) to form concatemers by rolling cycle-type replication. The DNA head-to-tail concatemers are cleaved into units by viral proteins including pUL52, prior to nuclear egress and capsid assembly.

6. After viral DNA replication and L gene expression, capsid formation and DNA packaging occurs in the nucleus. DNA containing-capsids acquire a preliminary envelope which they lost by passing through the nuclear membrane and acquire their final envelope by budding into the Golgi-apparatus vesicles.

7.The lytic replication leads to the maturation of new viral particles which are the released. Mature particles are release by transport of Rab3 secretory vesicles to be released by fusion of the vacuole with the plasma membrane, or through cell lysis, resulting in the egress of the enveloped virion.

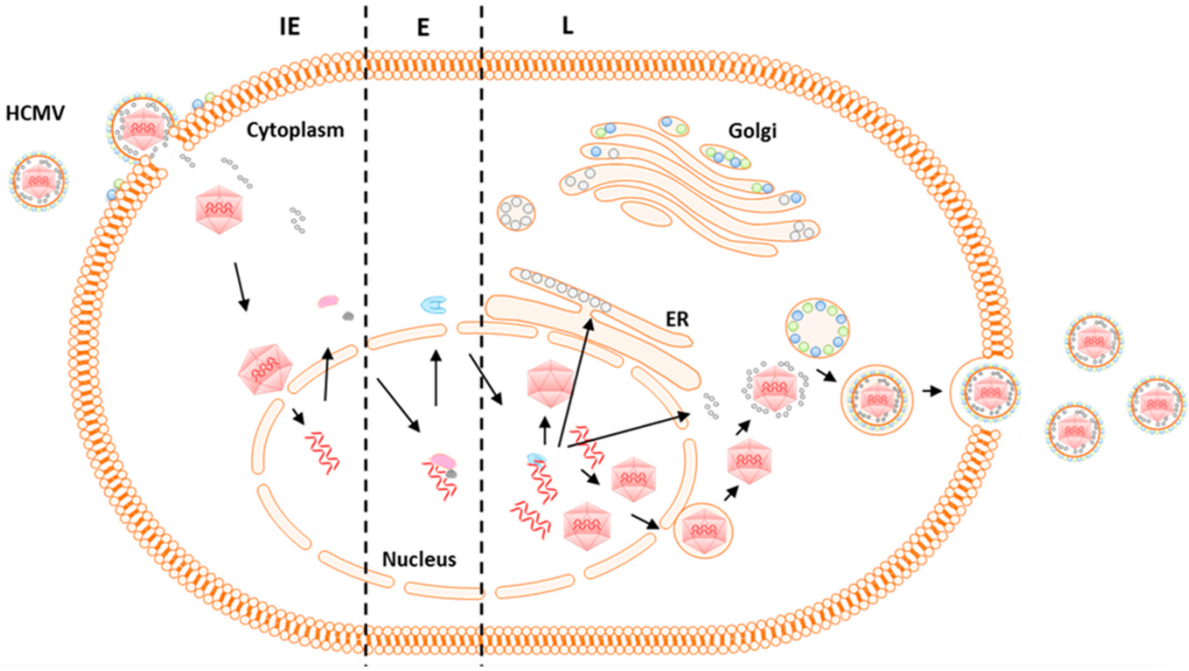


Figure A.2.3 HCMV life cycle and gene expression. HCMV viral life cycle phases are divided in attachment, entry, uncoating, viral genome expression and replication, capsid formation, virion maturation, and mature viral particle egression. Gene expression is temporary organized in three phases where IE proteins are required for E gene expression and E proteins are needed for viral replication and L gene expression. L proteins lead to capsid assembly and maturation, and new infectious viral particle formation. (Manandhar *et al.,* 2019)

A.2.3.1 Immediate-Early Proteins

IE (α -) genes are expressed starting 1 h after the attachment of the viral envelope to the host cell membrane and does not require a de novo protein synthesis. The most important IE genes are the so called major IE proteins, IEUL122/123 (IE1 and IE2) and the auxiliary genes, such as UL36-UL38, UL115-UL119, IRS1/TRS1, and US3. The major IE proteins, alone or in synergy, are potent trans-activators, required for the subsequent viral and cellular gene expression (Fortunato and Spector, 1999). The expression of the major IE genes is regulated by a complex enhancer element, which works in a specific way according to the tissue, the type of cell and its level of differentiation, and which exerts its strong transcriptional activity interacting with different transcription factors. The major IE region 1 encodes 4 proteins (IE-72, IE-19, IE-17.5 and IE-9) and the major

IE region 2 encodes for 5 proteins (IE-86, IE-55, IE-18, L-40 and L-60). The most abundant IE1 protein is the nuclear phosphoprotein pp72 (IE1-72), whose function is required for viral replication at low multiplicity of infection. It cooperates with an IE2 protein, IE2-86. IE1-72 possess kinase activity and may also have effects on cell cycle regulation interacting with the p107 protein of the Retinoblastoma family and with E2F1 (Margolis *et al.*, 1995). IE2-86 is a nuclear phosphoprotein whose function is critical for viral replication, in fact it is a powerful transcriptional regulator that can both stimulate and suppress the expression of HCMV and cellular gene (Song and Stinski, 2002). IE2-86 is believed to be the key regulator of viral transcription transition from the IE to E gene expression phase. IE2-86 is also responsible modulating host cell metabolism to promote viral replication, by blocking the former in early S phase and inhibiting cellular DNA synthesis (Fortunato *et al.*, 2000; Murphy *et al.*, 2000; Kalejta, Bechtel and Shenk, 2003).

A.2.3.2 Early Proteins

As alluded to above, the expression of the E or β genes relies on the expression IE proteins, being stimulated by IE2-86 alone or in cooperation with IE1-72. E genes mostly encode non-structural proteins, including viral DNA replication factors, repair enzymes, and proteins involved in immune response evasion. Six proteins of the replication machinery conserved in members of *herpesviridae* include pUL54 (DNA polymerase), ppUL44 (DNA polymerase processivity factor), pUL57 (single-stranded DNA binding protein), pUL70 (primase), pUL102 (primase-associated factor), and pUL105 (helicase). HCMV proteins pUL79, pUL87, and pUL95 which are recruited into the pre-replication

with ppUL44 and are relevant for HMCV viral genome replication, pp65 tegument protein and protein kinase ppUL97, which also phosphorylates the antiviral drug Ganciclovir, are some other examples of proteins expressed in E phase and important for HCMV life cycle. Through analysis of the viral gene expression profile, it was possible to draw the temporal map of expression of the genes IE, E, and L of the entire viral genome. Expression of 36% of HCMVs ORFs is not affected by treatment with Ganciclovir (GCV), which blocks the replication of viral DNA (Chambers *et al.*, 1999).

A.2.3.3 Late Proteins

The L proteins, which are mainly structural proteins, are the last class of HCMV gene products expressed during its replication and whose transcription begins more than 24 h post infection and requires the successful replication of viral DNA. They are proteins involved in capsid and tegument formation, viral assembly, and egress from the host cell. Three different classes of glycoproteins are expressed at this stage, gCI, gCII and gCIII (Kinzler, Theiler and Compton, 2002; Mach *et al.*, 2005; Isaacson and Compton, 2009). The second most abundant tegument protein, pp150 (pp65 is the first), encoded by UL32 gene, and the highly immunogenic phosphoprotein, pp28 are some of the proteins expressed in L phase of HCMV life cycle, known to be involved in virion maturation and viral egress (Meyer *et al.*, 1988; AuCoin *et al.*, 2006). In particular, pp28 is a component of the virion expressed exclusively in the cytoplasm of infected cells and during the time of maximal virus production.(Sanchez, Sztul and Britt, 2000).

A.2.4 DNA polymerase holoenzyme

All Herpesviruses encode for their own DNA polymerases where their structural and functional domains are conserved in all herpesvirus family members. HCMV DNA polymerase holoenzyme is a multi-functional enzyme that plays a key role during viral infection ensuring replication of the viral genome and consists of the catalytic subunit pUL54 and the processivity factor ppUL44, which physically and functionally interact thought specific residues (Ertl and Powell, 1992; Loregian *et al.*, 2004a, 2004b).

A.2.4.1. pUL54

pUL54 is a 140 kDa protein of 1242 amino acids, where the catalytic core is located in the central part of the protein (Ye and Huang, 1993), the N-terminal domain contains the three regions (Exo I, II and III) responsible for the 5'-3 exonuclease activity (Nishiyama, Maeno and Yoshida, 1983) and the C-terminal domain is devolved in binding ppUL44 and nuclear transport of the protein (Alvisi *et al.*, 2006). pUL54 contains two NLS, which one is an atypical, poorly characterized hydrophobic NLS (hNLS) whose sequence is PRRLHL-1127, and the other is a classical NLS (cNLS) located upstream of ppUL44 binding domain whose sequence is PA**KKR**A**R**-1159. While the first lies within the protein binding domain of UL44, the second is localized upstream, and it is active even when pUL54 is complexed with ppUL44, thus conferring energy- and RAN-dependent nuclear targeting to reporter proteins as well as interacting directly with IMPα in IMPα/β dependent nuclear transport pathway (Alvisi *et al.*, 2006, 2008)

A.2.4.2 ppUL44

ppUL44 is a 52 kDa phosphoprotein of 433 residues which represents HCMV DNA polymerase processivity factor and its role is essential for HCMV genome replication

(Ripalti *et al.*, 1995). ppUL44 is believed to functions as a homodimer and is composed by two identical subunits, each include an N-terminal and a C-terminal domain. The N-terminal domain (residues 1-290) retains all the protein's known biochemical proprieties, such as the ability to bind to HCMV dsDNA and pUL54, thus stimulating the activity of the latter (Weiland *et al.*, 1994) (FigA.2.4). The C-terminal domain (residues 121-433) instead, is largely unstructured and contains five stretches of glycine which confer extreme flexibility and there is a functional cNLS (PNTKKQK-431) recognized by IMPα/β dependent nuclear transport pathway (Alvisi *et al.*, 2005), moreover residues at the C-terminal domain appear to be involved in the regulation of the nuclear import process since there are multiple phosphorylation (S413, S415, S418 and T42T) and sumolyation (K410) sites (Alvisi *et al.*, 2005, 2008, 2011; Sinigalia *et al.*, 2012). While the C-terminal structure is still uncharacterized, the N-terminal of ppUL44 had been successfully crystallized (Appleton *et al.*, 2004). ppUL44 subunits forms two topologically similar domain and the two most N-terminal ß-sheets of each monomer dimerize in a head-to-head dimer linked by a connector loop. The front face of the protein is important for binding other proteins like pUL54 (Appleton *et al.*, 2006) and the back face, which is flanked by four helices and is rich in basic residues, is involved in DNA binding (Komazin-Meredith *et al.*, 2008a). Dimerization of ppUL44 is essential for DNA binding since the basic face of each monomer are involved in the formation of positively charged central cavity where interacts with the DNA via electrostatic interactions (Komazin-Meredith *et al.*, 2008a). Indeed, point mutations impairing dimerization such as L86A/L87A double mutant, which are residues directly involved in the dimer formation present on the dimerization surface of each monomer (Fig.A.2.5), strongly impair dsDNA binding, thus preventing HCMV DNA replication *ori*Lyt dependent,

suggesting that ppUL44 is essential and essential viral genome replication (Sinigalia *et al.*, 2008a). Other examples of mutations that prevents viral DNA replication are reported to be like ppUL44-I135A mutant, a residue substituted in the connector loop, that by disrupting pUL54-ppUL44 interaction impairs DNA replication. A study on living cells by means of FRET and BRET, also demonstrated that the affinity of ppUL44 dimerization is similar to that of pUL54/ppUL44 interaction, and mutation on specific ppUL44 selectively affected its dimerization or its ability to interact with pUL54, thus preventing the formation of DNA polymerase and so impairing binding to HCMV dsDNA (Di Antonio, Palù and Alvisi, 2021a). Therefore, ppUL44 dimerization can be considered a potential target of therapeutic intervention to interfere with HCMV replication.

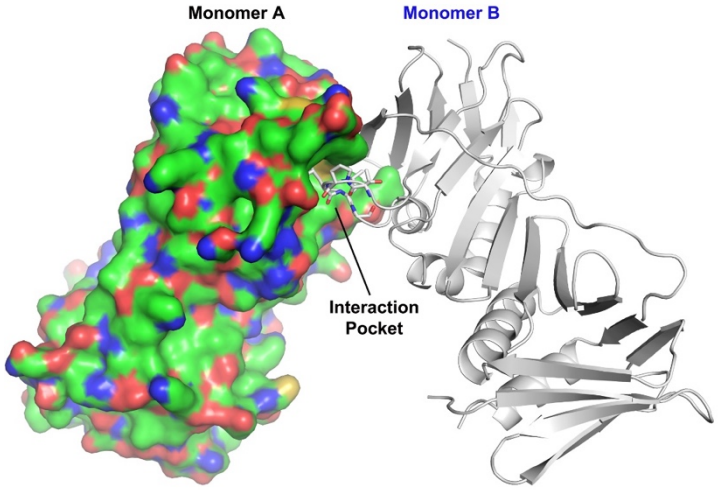


Figure A.2.4 ppUL44 (1-290) monomers graphic representation. One monomer is represented as surface (monomer A) and the other one as ribbons (monomer B). (Ghassabian *et al.*, 2021a)

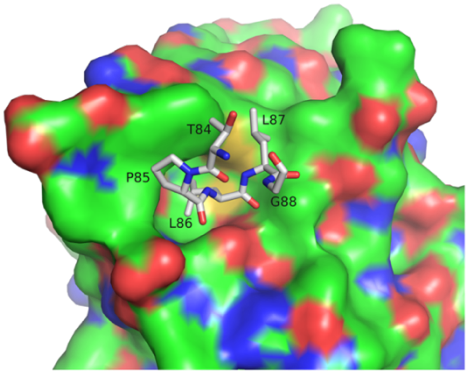


Figure A.2.5 ppUL44 (1-290) monomers dimerization surface. Residues involved in dimerization being shown as sticks located the interaction pocket. (Ghassabian *et al.*, 2021b)

A.2.5. HCMV Pathology

HCMV represents a very common infection worldwide, exhibiting a seroprevalence of 40– 100% depending on geographical location and socioeconomic status, it can be acquired any time during life and the seroprevalence increases with age, and the latter may contribute to general immunosuppression and increased incidence of different diseases such as cancer. Primary HCMV infection in healthy individuals is usually asymptomatic or with mononucleosis-like symptoms, resulting in the establishment of a lifelong latency. It does, however, periodically reactivate, but these reactivation events are quickly shut down by the immune system, generally resulting in a sub-clinical infection. Instead, in immunocompromised individuals, such as acquired immunodeficiency syndrome (AIDS) patients, stem cell and solid organ transplant recipients, as well as in individuals with an underdeveloped immune system such in the case of vertical transmission from mother to fetus, HCMV is a significant cause of morbidity and mortality (Sissons, Bain and Wills, 2002). In AIDS patients, HCMV infection of organs, such as the retina, results in the end-organ disease. As a critical level of viral replication is necessary for the development of the disease, maintaining the viral load below this threshold allows for protection against end-organ manifestation. Maintenance of the viral load in AIDS patients improved with the development of HAART. In solid organ transplants (SOT), approximately 78% of seropositive donors (D+) transmit HCMV to seronegative recipients (R−) in allografts containing lytic or latent virus (Atabani *et al.*, 2012). This combination of D+/R− in SOTs poses the highest risk of severe CMV disease, worsen by the lack of any pre-existing immunity of the recipient. However, similar to congenital CMV, prior exposure to the virus does not preclude disease; around 40% of seropositive SOT recipients (R+) see reactivation of their own

latent reservoirs of HCMV during immunosuppressive treatment, and some are also reinfected by new HCMV strains transmitted by a seropositive donor (Grundy *et al.*, 1988). By contrast, in bone marrow and peripheral hematopoietic stem cell transplantation (HSCT), the most at risk group for CMV disease is seropositive recipients with a seronegative donor (D−/R+) (Krishna, Wills and Sinclair, 2019). In these patients, their own latent virus reactivates, but the donor graft provides no HCMV antigen-specific T cells to control its spread. On the other hand, congenital HCMV infection has an incidence that ranges from 0.2% to 5% of newborns (Naing *et al.*, 2016), and up to 90% of congenitally infected newborns are asymptomatic at birth, although they are at risk to develop serious consequences during their development. The remaining 10/15% present symptoms which lead to microencephaly and lethargy. To date, congenital HCMV infection is the main cause of sensorineural hearing loss during childhood (Dollard, Grosse and Ross, 2007). HCMV is transmitted through body fluids: saliva, blood, urine, breast milk, as well as through sexual contact, organ transplantation and from mother to an unborn child (intrauterine infection) (Sissons, Bain and Wills, 2002).

A.2.6. HCMV Antiviral-drugs

Over the years, the Food and Drug Administration (FDA) has approved several antiviral drugs for the treatment of HCMV infection.

A.2.6.1 Ganciclovir (GCV)

GCV [9-(1,3-dihydroxy-2-propoxy- methyl) guanine], was the first antiviral agent to be approved for clinical treatment of HCMV infection and still remains the gold standard drug for treatment of HCMV infections. GCV is an acyclic nucleoside analog of 2'-deoxyguanosine and is phosphorylated by the viral protein kinase UL97 and cellular

kinases into a biologically active triphosphate form that inhibits viral DNA synthesis by acting as a nucleoside analogue to block the viral DNA polymerase (Littler, Stuart and Chee, 1992). Mutations for GCV resistance have been mapped both to the DNA polymerase pUL54 and the viral protein kinase UL97 genes (Gilbert and Boivin, 2005). GCV is available in intravenous formulation and as a sustained-release intraocular implant approved for the treatment of HCMV retinitis. The main side effects of GCV treatment are hematologic abnormalities (primarily neutropenia, anemia, and thrombocytopenia) as well as kidney and liver toxicity (Markham and Faulds, 1994).

A.2.6.2 Valganciclovir (VGCV)

Valganciclovir is the pro-drug of GCV and acts in similar way although it is only available in enteral formulation. After intake, valganciclovir is rapidly metabolized to the active form in the intestinal wall and liver (Erice, 1999). Valganciclovir, as GCV, has hematologic toxicity with neutropenia and anemia, but also diarrhea due to its direct effects on the intestine.

A.2.6.3 Foscarnet (FOS)

Foscarnet is a pyrophosphate analogue, which reversibly and non-competitively inhibits viral DNA polymerase by blocking the pyrophosphate-binding site of pUL54 and does not require enzyme activation after intake (Mercorelli *et al.*, 2011). FOS is manufactured in intravenous formulation and is considered as a second-line therapy since it is administrated when GCV therapy fails because of viral resistance, or for those who cannot be treated with GCV due to dose-limiting neutropenia or leucopenia (Razonable, Emery, and 11th Annual Meeting of the IHMF (International Herpes Management Forum), 2004). Due to its potential nephrotoxicity, the administration requires slow

infusion, extensive pre-hydration, and frequent monitoring of serum creatinine levels (Naesens and De Clercq, 2001).

A.2.6.4 Cidofovir (CDV)

Cidofovir (CDV) is an acyclic nucleoside phosphonate analogue, and it is converted to the active form by cellular kinases and acts as inhibitor of viral DNA polymerase, causing premature chain termination in viral DNA synthesis (De Clercq and Holý, 2005). Cidofovir is administered only intravenously due to its low oral bioavailability. The main advantage of Cidofovir is the long intracellular half time that allows effective treatments even by infrequent dosage (Aduma *et al.*, 1995). The major limitations of Cidofovir are its side effects such as neutropenia and severe renal toxicity, leading to electrolyte imbalance. In preclinical studies cidofovir has shown carcinogenic and teratogenic effects (Lea and Bryson, 1996). Due to its profile of side effects, cidofovir is used only as a second line therapy.

A.2.6.5 Formivirsen

Fomivirsen inhibits HCMV IE gene expression, which is vital for viral replication. It is produced in intravitreal formulation and is used in clinical settings for treating HCMV retinitis in HIV patients. Due to its intravitreal administration, it has no systemic effects during treatment. It has a half-life of approximately 55 hours, which allows infrequent dosage (Geary, Henry and Grillone, 2002).

A.2.6.6 Letermovir

Letermovir, a novel anti-HCMV antiviral, is a quinazoline, targets pUL56 of the HCMV terminase complex. Letermovir does appear to interact with the immunosuppressants given to patients after HSCT and has recently been approved for the prophylactic

treatment of HCMV disease in HSCT recipients in the USA, Canada and the European Union. Importantly, Letermovir does not seem to have antagonistic effects when combined with currently approved HCMV antivirals *in vitro*, raising the possibility that combinations of these antivirals could be used to treat HCMV. As renal and hepatic impairment affects Letermovir pharmacokinetics, increasing exposure, may affect its use in kidney and liver transplant recipients (Ligat *et al.*, 2018).

A.2.6.7 Maribavir

Maribavir (MBV), developed by ViroPharma, is another promising anti-HCMV compound, which is administered orally and targets the viral kinase UL97. UL97 is required for correct formation of the viral tegument, formation of the viral assembly complex within the cell and virus release. However, co-administration of both MBV and GCV is not advised as the former is an inhibitor of the UL97 enzyme required for anabolism of the latter (Frange and Leruez-Ville, 2018). One benefit to MBV is that it shows reduced hematotoxicity and nephrotoxicity compared to GCV and VGCV and so could eventually replace these older compounds. One Phase III trial is testing MBV (200 mg daily) for transplant recipients with HCMV that already show resistance to GCV, CDV or Foscarnet.

A.2.7 HCMV Infection Available Therapies

There are two main strategies for HCMV infection prevention in transplant recipients and are prophylaxis and preemptive therapy, which are each considered be the preferred prevention approaches after SOT and HCT, respectively. Prophylaxis after SOT

has been shown to reduce not only HCMV infection and disease while patients are receiving antiviral therapy, but also graft loss, mortality, and opportunistic infections (Atabani *et al.*, 2012). All HCMV D+/R– or R+ SOT recipients being treated with prophylaxis usually receive oral valganciclovir for a like 3–6 months, but up to 12 months after lung transplantation. HCT recipients have traditionally been treated with preemptive therapy, whereby they are monitored for the development of asymptomatic HCMV replication and are only given antivirals upon the detection of viremia, with the goal of preventing progression to disease.

Congenitally infected newborns, showing moderate to severe symptomatic disease are treated with VGCV within the first 4 weeks of life for up to 6 months, with monitoring of neutrophil counts and transaminase levels.

A.2.8 Previous Results

As alluded before, our research group confirmed ppUL44 dimerization in living cells by means of FRET and BRET assays (Di Antonio, Palù and Alvisi, 2021b) and in vitro by means of Thermal Shift assay (TSA) and GST-Pulldown assay, thus validating the dimeric protein crystal structure . Since Small Molecules (SMs) have already been identified as capable of inhibiting HCMV replication by mediating disruption of the ppUL44/pUL54 complex (Loregian and Coen, 2006), our research group reasoned that it might be possible to identify compounds hindering viral replication by targeting the dimerization of ppUL44. The crystal structure was then used to perform a virtual screening, employing The Glide software (Schrödinger, NY, USA) to dock molecules to the interface of the two

monomers (Ghassabian *et al.*, 2021c). After three rounds of screening only the top-ranked 500 molecules in term of docking score were selected after visual inspection, cluster analysis, and based on their commercial availability, 18 compounds were purchased and tested for their ability to disrupt impair HCMV replication (Fig. A.2.6).

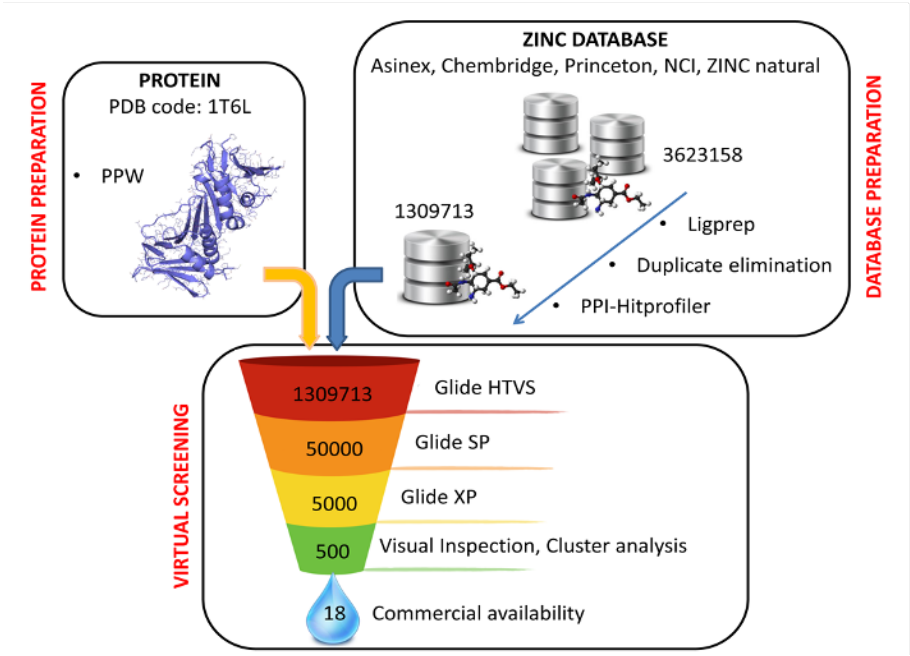


Figure A.2.6 Schematic overview of the virtual screening aimed to identify potential ppUL44 dimerization inhibitors. Glide software was used to dock molecules to the interface of the two monomers (PDB code: 1T6L). Three rounds of screening were performed using the High throughput virtual screening (HTVS), Standard Precision (SP) and Extra Precision (XP) docking settings. After each docking round, the top-ranked molecules in term of docking score were selected for the following round. The resulting 500 molecules were further filtered by visual inspection, cluster analysis and based on their commercial availability, 18 compounds were selected for further studies. (Ghassabian *et al.*, 2021c)

Their cytotoxicity and antiviral activity were preliminary evaluated at the concentrations of 100 and 10 µM on fibroblast cells, using MTT and Fluorescent Reduction Assay (FRA) assays, respectively. Several SMs exhibited poor solubility, precipitates were also observed microscopically, and few of them exhibited cytotoxic effect (Fig.A.2.7). This preliminary experiment identified 4 potential HCMV viral infection inhibitors (B1, B3, B6 and C6). The activity and cytotoxicity of four most active compounds was further

evaluated in dose-response experiments, thus allowing us to calculate CC_{50} and EC_{50} of each compound, by means of FRA and MTT. CC_{50} is defined as the compound's concentration (µg/mL) required for the reduction of cell viability by 50%, and ED_{50} is the dose of a drug that produces, on average, a specified all-or-none response in 50% of a test population. Fluorescent Reduction assay (FRA) was performed on fibroblast cells infected with TB4-UL83-EYFP recombinant HCMV viruses gifted for Michael Winkler (Fig.A.2.8).

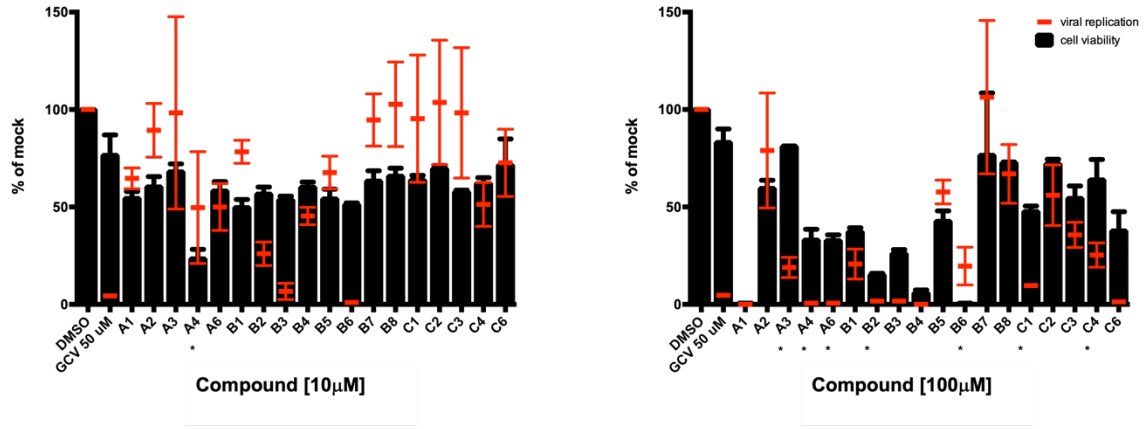


Figure A.2.7 Identification of compounds interfering with HCMV replication. MRC5 cells were infected with TB4-UL83-EYFP at an MOI of 0.03 IU/cell and treated with each SM either at concentration of 10 µM or 100 µM. In parallel, uninfected MRC5 cells were also treated for assessment of SMs cytotoxicity. Seven days post treatment, cells were processed for data acquisition and analysis as described in the Materials and Methods section. Mean YFP values relative to infected cells treated with the indicated SMs are expressed as a percentage of DMSO-treated cells (red bars). The mean + standard error of the mean (SEM) relative to 3 independent experiments is shown. *indicates the presence of precipitates. [Dissertation, PhD student: Martina Timmoneri]

Briefly B6 was the most active compound yet the most toxic with an ED_{50} and CC_{50} of 2.10± 0.6 µM and ~10 µM, respectively, resulting in a poor selectivity index (SI, the ratio CC_{50}/ED_{50}) < 5. B3 inhibited efficiently viral replication (ED_{50} of 4.2 ± 2.4 µM) in absence of evident cytotoxicity, with a SI > 20. Also, C6 exhibited low cytotoxicity but was less efficient compared to the others (ED_{50} of 17.9 ± 9.6 µM) and B1 effects on viral HCMV

replication was observed only at very high concentrations (ED_{50} of 87.7±14.3 µM) resulting in a SI<2.

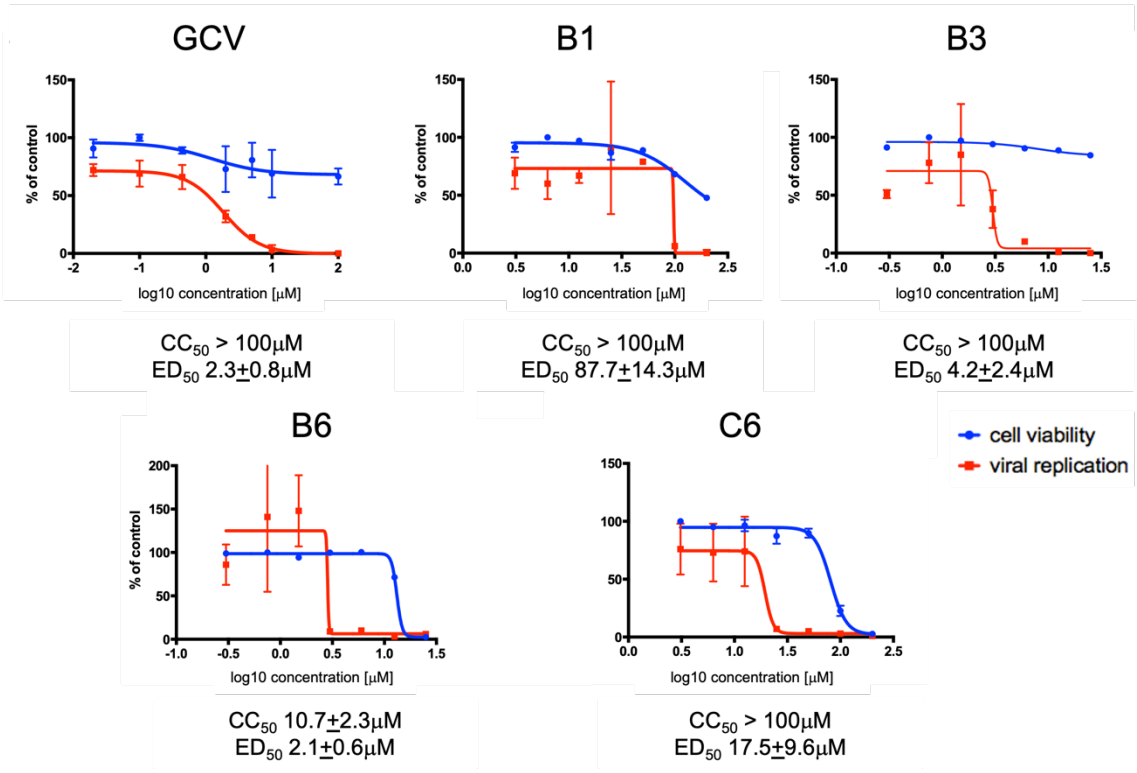


Figure A.2.8 Determination of ED50and CC50values of SMs by FRA and MTT assay. MRC5 cells were infected with TB4-UL83-EYFP at an MOI of 0.03 IU/cell and treated with increasing concentrations of the indicated compounds. In parallel, uninfected MRC5 cells were also treated for assessment of SMs cytotoxicity. Seven days post treatment, cells were processed for data acquisition and analysis as described in the Materials and Methods section. Mean values YFP values relative to infected cells treated with the indicated compounds are expressed as a percentage of DMSO-treated cells (red squares). Cell viability was assessed by MTT assays, and data expressed as a percentage of DMSO treated cells (blue circles). For each compound, representative plots are shown, along with the cytotoxic concentration 50 (CC50) and effective dose 50 (ED50) mean values + standard deviation of the mean relative to at least 4 independent experiments. (Ghassabian *et al.*, 2021b)

A.3 Material and Methods

A.3.1. Cell lines, media and maintenance

Human foreskin fibroblast (HFF) and MRC5 (#CCL171, ATCC) cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (*v*/*v*) fetal bovine serum (FBS), 50 U/mL penicillin, 50 U/mL streptomycin, and 2 mL-glutamine (DMEM cpt) and passaged when reached confluence. For MRC5 and HFF cell lines, cells were passaged when reached > 90% confluence. To this end, cells were briefly washed with Dulbecco's phosphate-buffered saline (D-PBS) and incubated with 0.05% trypsin-EDTA until they started detaching from the flask. Cells where then resuspended in cpt DMEM and an appropriate volume of cells was transferred to a new flask containing required amount of cpt DMEM. All culture reagents were purchased from Gibco, Thermofisher Scientific. PBS, Trypsin and DMEM volumes varied according to the format of the flasks and plate used. Stored at +4°C. For all experiments, MRC5 cells were used between passages 19 and 26, whereas HFFs were used between passages 4-20.

A3.1.1 Mycoplasma Test

Cell were routinely tested for mycoplasma contamination using the N-GARDE Mycoplasma PCR Reagent Set (Euroclone #EMK090020). Briefly, 200 µl or 1ml, based on the flask format, of the cell culture media was centrifuged at 700 rpm for 5 min to discard dead cell debris. Supernatants was placed in a new sterile Eppendorf and centrifuged again for 10 min at max speed. Supernatants were carefully discarded, and pellets resuspended in 50 µl of Buffer solution, boiled at 95°C for 3 min and 5 µl were used as template for the PCR reaction. PCR reaction mix was comprehensive of 10 µl of kit's Mix Solution and the final reaction volume of 50 µl was reached with sterile MilliQ water. A

positive control sample and a blank one was used to compare the template PCR product with. The PCR set up was as follows:

◊ 1x cycle: 94°C for 30 sec

◊ 35x cycles: 94°C for 30 sec – 60°C for 120 sec – 72°C for 60 sec

◊ 1x cycle: 94°C for 30 sec

◊ 1x cycle: 60°C for 120 sec

◊ 1x cycle: 72°C for 5 min

◊ 1x cycle: 4°C ∞

The PCR products were loaded on a 2% agarose gel containing Eurosafe 1:25,000 (Euroclone #EMR440001) and DNA fragments were observed with Uvitec Allinace Image Software (UVITEC Cambridge Alliance).

A.3.2 Cell freezing

For long term storage, cells were stored in liquid nitrogen. For each 25cm^2 of cultured cells corresponded one cryovial (Corning® Cryogenic Vials with Orange Cap, #100-0091). Cells were washed in 1xPBS and Trypsin and resuspended in DMEM cpt. Resuspended cells were then centrifuged for 5 min at 700 rpm, the supernatant discarded, and the pellet cells were finally resuspended on ice in freezing medium (DMSO 10% v/v FBS). Each 25cm^2 of cultured cells were resuspended in 1.8 ml of freezing medium and aliquoted in pre-chilled cryovials, and incubated 10 min on ice, 2 h at -20°C, and over-night (ON) at -80°C before being transferred to liquid nitrogen.

A.3.3 Cell Thawing

For each thawed cryovial one 15ml Falcon® tube was prepared with 6ml of DMEM cpt. Cryovials were rapidly transferred from liquid nitrogen under the hood and thawed by resuspending the cells in warm DMEM cpt medium already present in the 15ml Falcon tube. Cells were centrifuged at 700 rpm for 5 min and the supernatant discarded. Cells were finally resuspended in 6ml DMEM cpt medium, seeded in the appropriate T25 flask and incubated at 37°C and 5% CO_2 in a humidified incubator.

A.3.4 Viruses, Viral Stocks Preparation and Titration

A.3.4.1 Viruses

HCMV laboratory strain AD169 was obtained from ATCC (ATCC, #VR-53). Recombinant virus AD169-GFP, expressing a humanized version of GFP under control of the HCMV IE promoter between open reading frames US9 and US10, as well as its GCV-resistant derivative AD169-GFP26, bearing the UL97 M460I substitution, were a generous gift by Manfred Marschall (Erlangen, Germany). Recombinant virus TB4-UL83-YFP, wherein YFP was fused to the C-terminus of the early-late tegument phosphoprotein pp65 (Straschewski *et al.*, 2010) was kindly provided by Michael Winkler (Gottingen, Germany).

A.3.4.2 Viral Stock preparation

$3x10^6$ MRC5 cells were seeded in a T150 flask. Viral stocks were thawed at 37°C and diluted in 10 ml of DMEM. Cells were infected at low multiplicity of infection (MOI 0.02-0.05) to allow plenty of time for spread of propagating virus. Flasks were gently shaken every 15 min at 37°C during the infection to allow even distribution of the virus on the

cell monolayer. The infection media was removed after 2 h post-infection (p.i.) and replaced with 30 ml of cpt DMEM. Everyday p.i. cells were observed and refed every 3 days with 30 ml of cpt DMEM. When the monolayer displayed approximately 80% cytopathic effect (CPE), cells were refreshed with 20 ml of cpt DMEM in preparation for the harvest. After 2-4 days supernatants were collected from the flask and the cells were refed with 20 ml of cpt DMEM. When possible, another harvest was made 2-3 days after. Supernatants were cleared from cell debris by centrifugation at 1,500 rpm for 5min at 4°C. As a cryopreservative, a mixture of DMSO 10% was added dropwise while gently swirling the tube to the supernatants at 1:10 ratio, in order to reach a final 1% DMSO v/v concentration. Viral aliquotes were subsequently stored at -80°C.

A.3.4.3 Viral Titration

$1.5x10^4$ MRC5 cells/well were seeded into clear flat bottom 96 wells tissue culture plates with low evaporation lids (Falcon, #353072). The day after, an aliquot of frozen virus stock was quickly thawed and was serially diluted in DMEM, 100 µl or each dilution were added to each well in triplicate. The infection was carried out for 1 h at 37°C and 5% CO_2 in a humidified incubator, cells were subsequently refed with 200 µl of fresh media. 24 h p.i., cells were washed three times with 200 µl of PBS/well. After the final wash, the cell monolayer was fixed with 200 µl/well of 96% ethanol for 15 min at room temperature (RT). The ethanol was removed, the plate was washed twice with PBS and finally aspecific binding sites were blocked by incubating the cells with 50 µl/well of PBS-FBS 5% either for 1 h at 37°C or overnight at 4°C. Cell were incubated with primary monoclonal anti-CMV IE1&2 antibody (P1215 – Virusys Corporation) 1:100 (10 µg/ml, in PBS-FBS 5%) for 1 h at 37°C. Following this incubation, wells were washed four times

with 200 µl of PBS and were incubated with 50 µl/well of appropriate secondary antibodies [Alexa Fluor 555 Goat Anti-mouse IgG (A21424-Life Technologies), or Alexa Fluor 488 Goat Anti-mouse IgG (A11001)] 1:1000 (2 µg/ml) for 1 h at 37°C. The secondary antibody was removed washing two times with 200 µl of PBS. Occasionally, for recombinant HCMV viruses TB4-IE2-EYFP and TB4-UL83-EYFP encoding for spontaneously fluorescent viral proteins, the immunostaining procedure was avoided, and infected plates were analyzed either 24 (TB4-IE2-EYFP) or 48-72 (TB4-UL83-EYFP) h p.i.. Infected cells were counted using an inverted fluorescence microscope (Leica, DFC420 C) equipped with a 20x objective and appropriate fluorescence filters. Each plate was initially scanned to identify wells containing a number between 50 and 100 positive cells. The titer of the sample was calculated using the following formula: infectious units/ml= (number of positive cells x10)/ dilution of inoculum).

A.3.5. Preparation of Small Molecules (SMs) and Ganciclovir (GCV) Stocks

Ganciclovir (GCV; S1878, Selleckchem, Houston, TX, USA) and small molecules (SMs), with a >90% purity assessed by liquid chromatography mass spectrometry and high-performance liquid chromatography (Vitas-M Laboratory, Radio City, Hong Kong), were resuspended in 100% DMSO to obtain 50 mM and 20 mM stocks, respectively, and stored at −20 °C protected from light, until required. (Table A.3.1)

ZINC_ID	NAME	Molecular Weight	Formula
ZINC04503235	A1	409.202	C15H16BrF3N2O3
ZINC06494010	A2	214.232	C10H10N6
ZINC01112200	A3	364.401	C21H20N2O4
ZINC01245101	A4	371.418	C18H17N3O4S
ZINC04085376	A6	312.321	C18H16O5
ZINC18198322	B1	415.449	C24H21N3O4
ZINC02434635	B2	383.407	C23H17N3O3
ZINC27696609	B3	332.403	C21H20N2O2
ZINC02789051	B4	499.592	C28H25N3O4S
ZINC18266646	B5	337.427	C19H23N5O
ZINC09019436	B6	529.666	C29H31N5O3S
ZINC00793356	B7	357.337	C19H16FNO5
ZINC01441541	B8	367.38	C20H18FN3O3
ZINC00798328	C1	356.809	C19H17ClN2O3
ZINC02269985	C2	409.438	C23H23NO6
ZINC04524596	C3	350.356	C14H14N4O5S
ZINC01070326	C4	350.356	C14H14N4O5S
ZINC09762871	C6	467.934	C23H18ClN3O4S

Table A.3.1 SMs list and their relative ZINC-ID. Table with all the 18 SMs potentially disrupting ppUL44 homodimerization. Molecular Weight, ZINC database ID and the respective Molecular Formula for each compound is shown.

A.3.6. Antiviral Compounds Testing

A.3.6.1 Fluorescent Reduction Assay (FRA)s using AD169-GFP

To set up the Fluorescent Reduction Assays enabling rapid quantification of the inhibition of SMs and to evaluate the DMSO effects on viral replication, initial experiments were performed to establish the proper assay conditions in terms of plate format, infection time, MOI and DMSO concentration. After optimization of the protocol, to calculate the dose that give around 50% of the maximum possible drug effect (Effective Dose, ED_{50}) of each compound relative to AD169-GFP and its GCV resistant AD169- GFP26 counter apart by means of FRA, $1.8x10^5$ MRC5 cells were seeded in 12 well plates in 1 ml DMEM cpt and incubated ON at 37°C, 5% CO_2 and 95% humidity. The day after, cells were infected with 1ml/well of either AD169-GFP or AD169- GFP at MOI of 0.1 for 2 h at 37 °C. Subsequently, cells were washed with 2 ml of PBS, and 1 ml of media containing either DMSO 0.5% or increasing concentrations (0.001 to 100 µM) of each compound with a 0.5% DMSO final concentration were added. Mock infected cells served as a reference for calculation of background fluorescence. Every day plates were observed under an optical microscope to evaluate cells confluence and morphology as well as CPE and the presence of precipitates. Fluorescence signals were visualized on an inverted fluorescent microscope (Leica, DFC420 C). Seven days p.i., if required, supernatants were collected, cleared from cells and debris by centrifugation for 5 min at 700 rpm, and used for virus yield assays as described below. Subsequently, cells were washed with 2 ml of ice cold PBS and lysed in 400 µl of GFP lysis buffer (25 mM Tris [pH 7.8], 2 mM DTT, 2 mM trans-1,2-diaminocyclohexane-N,N,N,N-tetra-acetic acid, 1% Triton X-100, 10% glycerol). Plates were incubated 10 min at 37°C in a humidified incubator, before being incubated for 30 min at RT orbital shaking at 225

rpm. Samples were centrifuged for 5 min at 4°C at 13,0000 rpm and 100 µl of cleared lysates were transferred to black bottomed 96-well plate (Costar®, REF 3916, 2018-04-12). Fluorescent signals were acquired with fluorometric plate reader (VICTOR X2 Multilabel Plate Reader, PerkinElmer), using a fluorometric excitation filter (band pass 485± 14 nm) and a fluorometric emission filter (band pass 535 ± 25 nm). Data were exported to excel and analyzed to calculate the mean fluorescence relative to each condition. To this end, the average signal relative to the mock infected samples was subtracted to the signal relative to each well. Subsequently, data were normalized to solvent-treated controls and analyzed with Graphpad Prism (Graphpad Software Inc., San Diego, CA, USA) to measure ED_{50}. The experiment was performed seven times, and each plate included at least two wells treated with the same compound, as well as at least 12 wells treated with solvent only, in order to calculate the p-value relative to the Student's T-test between appropriate groups and DMSO treated cells.

A.3.6.2 Virus Yield Reduction Assays (VYRAs)

To determine the ED_{50} of each compound against AD169-GFP virus and its GCV- resistant AD169-GFP26 counterpart by means of VYRAs, 1.5×10^4 cells/well MRC5 cells were seeded in clear flat bottom 96-well tissue culture plates with low evaporation lids (#353072, Corning). The next day, the medium was replaced with 200 µl/well of serial dilutions of supernatants containing AD169-GFP or AD169-GFP26 virus grown in the presence of inhibitory compounds. Every day plates were observed under optical microscope to evaluate cells fluorescent signal, confluence, and morphology as well as CPE and the presence of precipitates. One week later, virus yield relative to each

condition was calculated based fluorescence signals count for each well with inverted fluorescent microscope (Leica, DFC420 C) and expressed as 50% Tissue Culture Infectious Dose ($TCID_{50}$)/mL using the Spearman and Karber algorithm and subsequently to calculate the ED_{50}. Data was statistically analyzed using Graphpad Prism (Graphpad Software Inc.), to calculate the *p*-value relative to the Student's T-test between each condition and DMSO treated cells.

A.3.6.3 Plaque Reduction Assays (PRA)

The effect of B3 and GCV on AD169 replication in HFF cells was investigated by plaque reduction assays (PRA) as previously described (Mercorelli *et al.*, 2018). Briefly, 2×10^5 cells/well HFF cells were seeded in 24-well plates. The following day cells were infected at 37 °C with 70 Plaque Forming Unit (PFU) of HCMV AD169 per well in DMEM containing FBS 5%. At 2 h p.i., the inocula were removed, cells were washed, and media containing various concentrations of each compound, 5% FBS, and 0.6% methylcellulose (SIGMA, n. M7027) were added. All compound concentrations were tested at least in triplicate. After incubation at 37 °C for 10–11 days, cell monolayers were fixed, stained with crystal violet, and viral plaques were counted. Data were normalized to solvent-treated controls and analyzed with Graphpad Prism (Graphpad Software Inc., San Diego, CA, USA) to measure ED_{50}.

A.3.7 Cell Cytotoxicity Assays

To evaluate the effect of SMs on cell viability and proliferative potential, three separate assays were performed. MTT and MTS assays, which measure the activity NAD(P)H-dependent oxidoreductase enzymes, and CellTiter-Glo® Luminescent Cell Viability Assay assays, which measures intracellular ATP.

A.3.7.1 MTT

For MTT assays, three time points, 24, 48 and 72 h post treatment were evaluated and for each, different number of cells were seeded. $1.5x10^4$ cells/well, $4x10^3$ cells/well and $1x10^3$ cells/well of MRC5 were seeded in clear flat bottom 96-well tissue culture plates with low evaporation lids (#353072, Corning), in triplicate, for analysis at 24, 48 and 72h post treatment, respectively. After 24 h, cells were treated with different concentrations of GCV or SMs, or solvent (DMEM-DMSO 0,5%) only. Several wells containing ed only DMEM and no cells were also included for background correction. At 24, 48 and 72 h post-treatment, 10µl/well MTT (5mg/ml, SIGMA 1Y1228) were added to the wells and incubated at 37°C, 5% CO_2 and 95% humidity. After 2 h and 30 min, cells were lysed with 100 µl of an appropriate Lysis buffer (10%SDS HCl 0.01M). The plate was covered with foil and incubated at 37°C ON. Subsequently, plates were read with a plate reader at 570nm. The measurement of cell metabolic activity using 3-(4,5-Dimethyl-2-thiazolyl)-2,5-diphenyl- 2H-tetrazoliumbromid (MTT; #A2231, Applichem, Darmstadt, Germany) was following the manufacturer's recommendations. After background subtraction, data were normalized to solvent-treated controls and analyzed with Graphpad Prism (Graphpad Software Inc.) to calculate the cell cytotoxicity (CC_{50}) value which is defined

as the extract concentration that reduced the cell viability by 50% when compared to untreated controls. The Selectivity Index (SI) relative to selected compounds was subsequently calculated as the ratio between the CC_{50} and the ED_{50} values.

A.3.7.2 The CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay

For CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay (Promega, # G5421) 1.5 x 10^4 cell/well MRC5 cells were seeded in clear flat bottom 96-well tissue culture plates with low evaporation lids (#353072, Corning) in duplicate. After 24 h, cells were treated with two different concentrations of GCV or SMs (10 and 100 µM), or solvent only, by first removing the culture media and then adding 100 µl of either GCV, SMs or DMEM-DMSO 0.5%. Several wells containing only DMEM-DMSO 0.5%, and no cells were also included for background correction. Seven days post-treatment cells were processed for measurement of cell metabolic activity. To this end, 20 µl/well of MTS/PTS kit solution (G3582, Promega, Madison, WI, USA) were added in the dark, plates were incubated for 3 h at 37°C, 5% CO_2 and 95% humidity and signals were acquired with a plate reader. After background subtraction, data were normalized to solvent-treated controls and analyzed with Graphpad Prism (Graphpad Software Inc.)

A.3.7.3 The CellTiter-Glo® Luminescent Cell Viability Assay

For measurement of intracellular ATP levels by means of CellTiter-Glo® assays (G7570, Promega, # G7570), 1.5 x 10^4 cell/well MRC5 cells were seeded in clear flat bottom 96-well tissue culture plates with low evaporation lids (#353072, Corning) in duplicate. After

24 h, cells were treated with two different concentrations of GCV or SMs (10 and 100 µM), or solvent only, by first removing the culture media and then adding 100 µl of either GCV, SMs or DMEM-DMSO 0.5%. Several wells containing only DMEM-DMSO 0.5%, and no cells were also included for background correction. Seven days post-treatment cells were processed for measurement of cell ATP levels. Instead, for studying the effect of our compounds on cell growthas well as on their viability three time points, 24, 48 and 72 h post treatment were evaluated and for each a different number of cells were seeded: 1.5×10^4 cells/well, $4x10^3$ cells/well and $1x10^3$ cells/well of MRC5 were seeded in clear flat bottom 96-Well Treated Multiwell Tissue Culture Plates, Opaque White plates (#353296, Corning) for 24, 48 and 72 h respectively. As for CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay After 24 h, cells were treated with increasing concentrations of GCV or SMs, or solvent only by first removing the culture media and then adding 100 µl of either GCV, SMs or DMEM-DMSO 0.5%. Several wells containing only DMEM-DMSO 0.5% and no cells were also included for background correction. Plates were incubated at 37°C, 5% CO_2 and 95% humidity. At the desired time point post-treatment, cells were processed with CellTiter-Glo® kit Reagent for the measurement of ATP levels as follows: 100 µl of CellTiter-Glo® kit Reagent was added to each well, the plate was orbitally shaken for 2 min at RT and then incubated for 10 min at RT. Signals were acquired using a plate reader compatible with luminometric measurements (VICTOR X2 Multilabel Plate Reader, PerkinElmer). After background subtraction, data were normalized to solvent-treated controls and analyzed with Graphpad Prism (Graphpad Software Inc.) to calculate the mean for each compound and the standard error of mean.

A.3.8 Analysis of HCMV DNA Replication

A.3.8.1 Infection and Treatment

The effect of B3 and GCV on HCMV viral replication was evaluated by Real-Time PCR. To this end, 4×10^4 cells/well MRC5 cells were seeded on 24-well flat bottom plates with low evaporation lid (#353047, Corning). The following day, cells were either mock infected or infected with HCMV (TB4-UL83-EYFP) at MOI of 0.01 or 0.02 IU/cell in DMEM at 37 °C. Two h p.i., cells were washed twice with PBS and medium containing either solvent only (0.5% DMSO), GCV, or B3 at a concentration of 6 times the ED_{50} calculated for AD169-GFP in FRAs (16 µM for GCV and 50 µM for B3), was added to each well and incubated at 37°C, 5% CO_2 and 95% humidity. At 24, 72, and 144 h p.i., cells were detached by incubation with 100 µl trypsin for 5 min at 37°C, 5% CO_2 and 95% humidity. Following tryspin inactivation by addition of 400 µl DMEM 10% FBS (*v*/*v*), cells were centrifugated for 5 min at 1400 rpm, the supernatant discarded, and the pellet stored at −80°C until further analyses. Total DNA was extracted using the GenElute™ Mammalian Genomic DNA Miniprep Kits (G1N70, Merck Millipore).

A.3.8.2 Extraction of Mammalian Genomic DNA

DNA extraction was performed with GenElute™ Mammalian Genomic DNA Miniprep Kits (G1N70, Merck Millipore). Briefly, frozen cells were thawed, the pellet resuspended in 200 µl of Resuspension Buffer with the addition of 20 µl of Rnase-A solution and incubated at RT for 2 min. Then 20 µl of Proteinase K was added to each sample prior to lysis with 200 µl of Lysis Buffer solution. Samples were vortexed roughly for 15 sec and

incubated for 10 min at 70°C. In the meanwhile, pre-assembled Binding Column were activated with 500 µl of Column preparation solution, centrifuged for 1 min at max speed and the flow-through discarded. Subsequently, 200 µl of EtOH 96% were added to cell lysates and roughly mixed by vortexing, to obtain a homogeneous mixture prior to being loaded on the binding column. Columns were centrifuged for 1 min at 12,000 x g, the flow-through discarded and the binding column filter placed on in a new collection tube. Subsequently, two washing steps with 500 µl of Washing Solution allowed to further purify the DNA from salt contaminations. A void centrifuge was performed to remove excess washing solution, prior to incubation for 5 min with 200 µl of Elution Buffer. DNA was eluted by centrifugation for 1 min at 12,000 x and 10 µl of eluted material were loaded on a 0.8% agarose gel to check for DNA integrity. Purified DNA was quantified with NanoDrop® ND-1000 Spectrophotometer (Nanodrop Technologies).

B.3.8.2 Real Time PCR

The levels of viral DNA contained in each elution were determined by quantitative real-time PCR (qPCR) using an ABI Prism 7000 Sequence Detection System (Thermofisher Scientific) and Power SYBR Green PCR Master Mix (#4309155, Thermofisher Scientific). The reaction mix included 5 µl of DNA template, 12.5 µl of SYBR Green PCR Master Mix, 0.5 µl of each primer 10 µM, and MilliQ water up to 25 µl of final reaction volume. Primers used were B2.7-FWD (5′-TGTTCTTCTTGGTTCATTTCC-3′) and B2.7-REV (5′-CGTGTCCGGTCCTGATTC-3′), or BAF-FWD (5′-CGGGACCTGACTGACTACCTC-3′) and BAR-REV (5′CCATCTCTTGCTCGAAGTCCAG-3′) for detection of HCMV genomic region corresponding to the non-coding β2.7 RNA or of the human β-actin gene, respectively (Collins-McMillen *et al.*, 2019). Samples were loaded on MicroAmp® Optical 96-well

Reaction Plate (Applied Biosystem®, #N8010560) in triplicate and sealed with MicroAmp®Optical Adhesive Film (Applied Biosystem®, #4311971). Real-Time PCR set up was as follow:

◊ 1x cycle 50°C for 2 min

◊ 1x cycle: 95°C per 10 min

◊ 40x cycles: 95°C for 15 sec – 60°C for 1 min

◊ 1x cycle 95°C for 15 sec – 60°C for 20 sec and 95°C for 15 sec

Reactions with only MilliQ water as template and a dissociation curve step were also included to discriminate possible contamination and primer-dimers formation. Levels of viral DNA were normalized to the cellular β-actin gene copies. Raw data were used to calculate the HCMV genomes fold change, relative to each experimental condition using the ΔΔct method (Mercorelli *et al.*, 2020). Values relative to each condition were further normalized to that obtained for DMSO-treated cells at 24 h p.i. and analyzed with Graphpad Prism (Graphpad Software Inc.) to calculate the *p*-value relative to each condition by means of two-way ANOVA, followed by ad hoc post-test analysis for multiple comparison and Tukey correction.

A.3.9 Analysis of HCMV Gene Expression by Western Blotting

A.3.9.1 Infection and Treatment

To evaluate of B3 and GCV on HCMV protein expression, 6 × 10^5/well MRC5 cells were seeded on 6-well flat bottom plates with low evaporation lid (#353046, Corning). The following day, cells were either mock non-infected or infected with AD169-HCMV strain at MOI of 2 IU/cell in DMEM, and incubated at 37°C, 5% CO_2 and 95% humidity. One

h.p.i., cells were washed twice with 1 ml of PBS. DMEM cpt containing either solvent only (0.5% DMSO), 16 µM GCV or 50 µM B3 corresponding to a concentration of 6 times the ED_{50} as calculated for AD169-GFP in FRAs, was added to each well. At different times p.i. (6-12-24-48-72 and 96 h p.i), cells were washed twice with 1 ml of PBS ice-cold and lysed on ice with 250 µL of RIPA buffer containing protease inhibitors (Tris-HCl 50 mM, pH 7.4, 150 mM NaCl, 1% Triton X-100 (*v*/*v*), 1% sodium de- oxycholate, 0.1% SDS, 1 mM EDTA, 17.4 µg/mL phenylmethylsulfonyl fluoride, 2 µg/mL aprotinin, and 4 µg/mL leupeptin). Cells were carefully detached from the plate with a sterile scraper, transferred into Eppendorf microcentrifuge tubes and stored at -80°C until further required.

A.3.9.2 Biochinconinic Acid Assay (BCA)

Cell lysates were quantified using Micro BCA™ Protein Assay Kit (Thermofisher Scientific #23235). To this end, serial Bovine Serum Albumin (BSA) dilutions in PBS1x served as standard. Cells lysates were thawed on ice and the tubes in which they were contained, were inverted several times during thawing for 30 min. After 1 h centrifugation at 4°C at 13,000 rpm, the supernatants were transferred in new Eppendorfs and each sample was diluted 1:50 and 1:100 times in a final volume of 250 µl. 100 µl of each standard and sample were loaded on the plate in duplicate. 100 µl of Working Solution, provided with the commercial kit, were then added to each well. After an incubation of 2 h at 37°C, the plate was analyzed with spectrophotometer Plate Reader at 492 nm (Sunrise, Tecan). Data were exported to excel, the mean absorbance calculated between the two dilutions and correlated with the standard values to evaluate the concentration of each protein.

A.3.9.3 Western Blotting (WB)

The protein concentration for each sample was quantified using the Micro BCA Protein Kit assay (#23235, ThermoFisher Scientific) (§A3.9.2). Subsequently, 30 µg of cell lysates were diluted in Laemmli sample buffer (0.05 M Tris-HCl, pH 6.8, 0.05% Bromophenol blue, 0.1 M DTT, 10% Glycerol (*v*/*v*), 2% SDS) and boiled 5 min at 95 °C before being loaded on 8% or 10% bis-tris polyacrylamide gels (the polyacrylamide concentration is based on the molecular weight of the samples) and being electrophoretically separated using a Biorad vertical gel system apparatus for 2 h at 100 V in Running Buffer (25mM Tris Base, 192mM glycine, 0.1% SDS). Separated proteins were blotted on polyvinylidene fluoride (PVDF) membranes (RPN303F, GE Healthcare, Chicago, IL, USA) for 1 h in Transfer Buffer (25mM Tris Base, 150mM glycine, 20% MetOH) at 350 mA. Following Ponceau staining (Amplichem, #A1405,0025, solution 0.1% Ponceau-5% Acetic Acid) the membranes were cut in the correspondence of the desired molecular weight ranges and were saturated with PBS containing 0.2% Tween20 and 5% milk (*w*/*v*) for at least 1 h at RT. Subsequently membranes were incubated with the appropriate primary diluted in PBS containing 0.2% Tween20 and 5% milk (*w*/*v*) orbitally shaking at 4°C ON. Membranes were then washed from exceeding primary antibodies, and the latter were conserved at -20°C and recycled for a maximum of three times before being discarded. After 6 washing steps with PBS containing 0.2% Tween20 and 5% milk (*w*/*v*), membranes were incubated for 1 h, shaking horizontally at RT, with appropriate secondary antibodies diluted in PBS containing 0.2% Tween20 and 5% milk (*w*/*v*). After washing out the exceeding secondary antibodies with two washes with PBS containing 0.2% Tween20 and 5% milk (*w*/*v*), two with PBS containing 0.1% Tween20 and 5% milk (*w*/*v*),

and three with PBS only, membranes were incubated with an enhanced chemiluminescence substrate (ECL Prime Western Blotting Detection Reagent, #RPN2236, GE Healthcare) and proteins were detected with Uvitec thanks to the software Aliance Q9 software (UVITEC Cambridge). Acquired signals were quantified using Image J (NIH), where the mean intensity for each protein expression signal was calculated by first subtracting the background signal and then by normalizing the values upon the expression of the loading control (β-Actin or GADPH). The following antibodies, diluted in PBS containing 0.2% Tween20 and 5% milk (*w*/*v*), were used: α-IE1&2 mAb (P1251, Virusys Corporation, Randallstown, MD, USA; 1:10,000), α-UL44 mAb (P1202-1, Virusys Corporation; 1:100); α-pp65 mAb (CA003-1, Virusys Corporation; 1:2,000); α-pp28 mAb (ab6502, Abcam, Lysates of mock infected cells were analyzed to verify antibody specificity. Cambridge, UK; 1:10,000), rabbit α-GADPH pAb (sc-25778, Santa Cruz Biotech, Dallas, TX, USA; 1:5,000); mouse α-β-Actin mAb (A5316, Merck Millipore, Burlington, MA, USA; 1:5000); goat α-mouse (Santa Cruz Biotech, sc-2055; 1:10,000) and α-rabbit (A6154, Merck Millipore; 1:10,000) immunoglobulin Abs conjugated to horseradish peroxidase.

A.4 Results

A.4.1 Activity of selected SMs on a HCMV-GCV resistant strain

A desirable characteristic of new HCMV antiviral drugs would be the ability to inhibit replication of viral strains resistant to the currently use antivirals. To verify if the SMs identified in previous studies (Fig. A.2.8) were endowed with the ability to efficiently inhibit drug resistant HCMV strains, consistently with the activity through interference with ppUL44 dimerization, we compared their effective dose (ED_{50}), defined as the dose that give around 50% of the maximum possible drug effect, against a recombinant reporter virus AD169-GFP and its GCV-resistant counterpart AD169-GFP-26, bearing the UL97 M4601 substitution (Marschall *et al.*, 2000). To this end, MRC5 cells were seeded and infected with either AD169-GFP or AD169-GFP-26 virus at a MOI of 0.1 IU/cell. Two h post infection, exceeding viral load was washed and cells were treated with either different concentration of compounds (0.001- 100 µM) as described in (§A.3.6.1) or with solvent only. Mock cells were used as control for background correction. One week post infection, viral replication was assessed by cell lysis and fluorescent signal was detected with a fluorimetric plate reader. In parallel, infectious virus titers in cell culture supernatants were collected and used to quantify the ED_{50} of the progeny virus release by Virus Yield Reduction Assay (VYRA) (§A.3.6.2), using $TCID_{50}$ method. FRA assays revealed (Fig A.4.1.A) that GCV inhibited more efficiently the replication of AD169-GFP virus compared to inhibition of AD169-GFp-26 viral strain (ED_{50} of 2.3 ± 1.9 µM versus 21.7 ± 9.6 µM, respectively; n=6) with an ED_{50} ratio (between AD169-GFP-26 and AD169-GFP) of 17 ± 19.9 µM, as expected. On the contrary, all other tested compounds inhibited both viruses with similar potency (Table A.4.1). B3 appeared to efficiently impair replication of both GCV-sensitive and GCV-resistant HCMV (ED_{50} of 7.9 ± 2.7 µM versus 17.6 ± 12.8 µM, respectively; n=6) resulting in a ED_{50} ratio of 2.0 ± 1.6 µM. B6 and

C6 were more efficient in inhibiting replication of both viruses compared to B3 with a ED_{50} ratio of 0.8 ± 0.5 µM for both compounds. A similar pattern was observed after quantification of viral progeny in VYRA (Fig A.4.1.B). However, despite GCV inhibited more efficiently viral production of AD169-GFP (ED_{50} of 0.7 ± 0.4 µM; n=4) as compared to AD169-GFP-26 (ED_{50} of 4.7 ± 4.8 µM; n=4), such difference was not statistically significant, probably due to high variations in quantification of progeny virus release. Overall, our compounds inhibit viral HCMV replication of both AD169-GFP and AD169-GFP-26 with similar potency, however both B6 and C6 strongly impaired cell morphology, implying potential effects on cell viability which at least in in the case of C6 (Fig.A.2.8) had been underestimated under previous experimental conditions (Dissertation). Considering that analysis of cell cytotoxicity was exclusively performed by MTT assays 7 days post treatment, after seeding high number of cells (1 x 10^5 cells/ 0.32 cm^2), we decided to evaluate the effect of our compounds using several different assays, and by varying cell number and incubation time. This would offer the possibility to obtain a better understanding of the specificity of our compounds.

A.4.2 Further Evaluation of the effect on cell viability of potential inhibitors of ppUL44 dimerization

Previously, the cytotoxicity of our 18 compounds was evaluated through MTT (Fig.A.2.7) assay and in parallel their antiviral activity was also tested on MRC5 cells infected with TB4—UL83-EYFPE HCMV strain (Fig.A.2.7). These results were obtained by seeding 1 x 10^5 cells/well in 96 well plate and measuring viral replication and cytopathic effect at 7 days post incubation. Based on the new observations, we decided to corroborate MTT cell viability assays with additional ones. To this end the effect of all 18 SMs as well as

GCV, on cell viability was tested by MTS and Cell Titer Glo® assays (see §A.3.7). Cells treated with solvent only (DMEM-DMSO 0.5%) were used for background corrections. Each assay evaluates cell viability in different ways, while MTT and MTS assays dependent on mitochondrial respiration and indirectly serve to assess the metabolic activity of a cell, Cell Titer Glo® relies on the quantitation of ATP levels in cells and is directly proportional to the number of cells present in the culture. MRC5 cells were seeded and the day after were treated with two different concentrations of SMs and GCV (10 or 100 μM). Plates were microscopically monitored daily for 7 days post treatment for the presence of precipitates. 7 days post treatment signals relative to individual wells were acquired as described in (§A.3.7) After background subtraction, data were normalized to solvent-treated controls and analyzed with GraphPad Prism (GraphPad Software Inc.).

Although MTT and MTS share the ability to evaluate cell viability by measuring mitochondrial metabolic rate, the latter's product is completely soluble, making its solubilization unnecessary before data acquisition. Therefore, reading resulted more robust and reliable as compared to MTT assay. On the other hand, Cell Titer Glo® assay gave similar results to MTS assay, but it was easier to perform and had much higher signal to noise ratio. As shown in Figure A.4.2 at the lowest concentration tested (10 μM), only A4 compound was not soluble and caused evident cytotoxicity. On the other hand, at the highest concentration tested (100 μM) several of the 18 SMs formed visible precipitates and caused cell death, indicating poor solubility and toxicity. Moreover, since we already performed dose response experiments at 7 days post infection with 4 of the most active compounds on both HCMV infected fibroblast (to calculate ED_{50} by means of Fluorescent Reduction Assay) and on non-infected cells (to calculate the CC_{50}

values by means of MTT), we decided to use Cell Titer Glo assay to calculate the CC_{50} of such 4 SMs at different time points post treatment. Since we seeded an appropriate number of cells with respect to the treatment time ($1.5x10^4$ cells/well for 24h, $4.5x10^3$ cells/well for 72h and $1.5x10^3$ cells/well for 144h) we could estimate their effect on both cell viability and growth (Fig. A.4.3). Results confirmed high toxicity of B6 compound, which reduced ATP intracellular levels by more than 50% at 100 µM already at 24h post treatment and by almost 100% at 72h post treatment. Importantly, cell treatment with C6 at 100 µM similarly reduced ATP content by more than 30% already at 24 h, whereas compounds B1 and B3 showed minimal effects at all time points analyzed. Based on these data, we concluded that the effects observed on HCMV replication obtained by B6 and C6 were likely due to cell toxicity and therefore we decided to further investigate the mode of action of B3, whose activity is more likely to be due to specific inhibition of HCMV function.

A.4.3 Effect of 3 SMs on HCMV replication by Plaque Reduction Assay (PRA)

Indeed, although B6 was the most potent compound (ED_{50} of 2.1 ± 0.6 µM, Fig. A.2.8) it was also extremely toxic (CC_{50} of ~10 µM), resulting in a poor selectively index (SI) < 5. On the other hand, C6 was also relatively toxic starting from 24h post treatment (Fig. A.2.8) while B1 effects on HCMV replication were evident only at very high concentrations (Fig.A.2.8). B3 was the sole compound with no evident cytotoxicity up to 100 µM and showed to efficiently inhibit HCMV viral replication with an ED_{50} of 7.9 ± 2.7 µM (Table. A.1). Therefore, we wished to confirm B3 antiviral activity also for non-recombinant HCMV. To this end we performed PRA assays, where HFF cells were seeded, infected with AD169 HCMV strain and after 2 were treated with different

concentration of B3 and GCV. After 10-11 days cells were treated as described in §A3.6.3 and viral plaques were counted. In such experimental settings, B3 inhibited HCMV replication with ED_{50} of 7.9 ± 3.5 µM (fig.A.4.4) suggesting that this compound also interferes with infection mediated by non-recombinant HCMV.

A.4.4 B3 impairs HCMV genome replication

To begin the characterization of B3 mode of action, we decided to investigate its effect on HCMV genome replication. Indeed, we would expect, that a compound disrupting ppUL44 homodimerization would inhibit the virus life cycle by impairing viral genome replication. MRC5 cells were infected with TB4-UL83-EYFP at different MOI (0.01 and 0.002 IU/cell) for 2 h and treated with either solvent only or concentrations of GCV or B3 corresponding to 6 times the ED_{50} as calculated for AD169-GFP in FRAs (16 µM for GCV and 50 µM for B3). Mock cells were used as negative control. At different time points (24, 72 and 120h) the infection was stopped, DNA was extracted and purified from cell lysates, and the viral genomes were quantified by qPCR thanks to specific primers targeting HCMV genomic region corresponding to non-codifying β2.7 RNA. Primers targeting human β-actin gene were used to evaluate the expression of MRC5 cells (§A.3.10).

For both the tested MOI, we observed that in the absence of HCMV-antivirals, there is a sharp increase of viral genome starting from 72 h post infection consistent with the typical replication pattern of HCMV, whereas treatment with either GCV or B3 significantly impaired viral genome replication. In particular, in non-treated infection, HCMV viral genome copy number increases by ~4 folds starting 72 h p.i. compared to the ~ 1.5 folds increase when treated with our compounds. (Fig. A.4.5). Although both

MOI gave similar results, infecting cells with TB4-UL83-EYFP at MOI of 0.002 data were highly variable and not very reliable, therefore we decided to repeat the experiments exploiting TB4-UL83-EYFP at MOI of 0.01. When we compare viral genome replication starting 72 h p.i. and normalize signal upon non-treated infected cells (Fig. A.4.6) it can be noticed that B3 treatment displayed extremely similar results as GCV wherein a ~4-fold and 6-fold decrease in viral genome copy number was quantified at 72 and 120 h.p.i., respectively. Therefore, B3 impairs HCMV life cycle by interfering with viral replication. Overall, our results are consistent with the possibility of B3 in interfering with HCMV genome replication by disrupting UL44 homodimerization.

A.4.5. B3 impairs HCMV early and late gene expression

In order to characterize the mode of action of B3 on HCMV life cycle further, we decided to investigate its effect on HCMV gene expression by Western Blotting. Indeed, our experiment set up for antiviral activity assessment allowed us to exclude activity on viral entry, since the infection was prior to SMs treatment and Real Time PCR data also implied activity before viral assembly (Fig.A.4.6).

However, the decrease in viral DNA replication observed starting 72 h p.i. (Fig.A.4.6) could depend on inhibition at several stages of HCMV life cycle, including expression and activity if Immediate Early (IE) and Early (E) gene products. Expression of antigens characteristic of each pahse was detected by specific antibodies against IE (IE1/2), early-late (ppUL44 and pp65) and late (pp28) HCMV gene products. MRC5 cells were infected with AD169 at MOI of 2 IU/cell for 2 h and treated with solvent only, GCV or B3, in concentration corresponding to of 6 times the ED_{50} as calculated for AD169-GFP in FRAs (16 μM and 50 μM, respectively). At different time points the infection was stopped,

cells were lysed, proteins were separated by SDS-PAGE Western Blotting and relative protein expression levels were analyzed with ImageJ as described in §A.3.8. Infected cells treated with solvent only, displayed the typical HCMV protein expression pattern, with IE1/2 being readily detectable starting from 6 h p.i., ppUL44 together with pp65 from 12 h p.i. and pp28 starting 48 h p.i. At 72 and 96 h p.i. the expression levels of pp65, ppUL44 and pp28 were constantly increasing (Fig.A.4.7.A). When looking at the protein expression pattern of infected cells treated with either GCV or B3 it can be observed that as expected, neither GCV or B3 had effect on the expression levels of IE1/2 proteins at 6 and 12 h p.i. (Fig. A.4.7.B e C, respectively) suggesting that these compounds have no role in the inhibition of activity on the major IE promoter. Drug treatment did not affect expression of early gene products ppUL44 and pp65 (Fig. A.4.7.B e C, GCV and B3 treatment respectively) at 24h p.i., confirming that IE function was not compromised. These observations, taken together implied that B3 similarly to GCV did not affect viral phases prior to genome replication. Importantly starting 48 h p.i. a decrease in the expression of ppUL44 and to a greater extend of pp65 could be observed. Moreover, both B3 and GCV inhibited expression of late gene product pp28 at 72 and 96 h p.i.. Indeed, pp28 expression was highly inhibited (Fig. A.4.7). Densitometric analysis confirmed that GCV (Fig.A.4.8.G) and B3 (Fig.A.8.H) inhibited HCMV gene expression with similar kinetics. These observations together with qPCR data are compatible with the ability of B3 to interfere with early HCMV functions, possibly disrupting ppUL44 homodimerization.

A.5 Discussion

HCMV is the leading cause of severe disease in immunocompromised individuals and in congenitally infected newborns, it's a member of the beta Herpesviridae subfamily and its ds DNA is duplicated by a number of viral encoded protein, including a DNA polymerase holoenzyme, composed by a catalytic subunit and a processivity factor. FDA over the years approved many drugs for the HCMV infection treatment and the target of most of them is the DNA polymerase, but then we had emergence of resistant viral strain beside the associated toxicity and recently the focus is to find new targets for the development of new anti HCMV drugs. For example, Letermovir that inhibits the virus terminase and impairing viral particle formation, approved as prophylaxis treatment for HCMV infection in HSC transplants. But overall is still important to find new target for treating HCMV infection. An alternative target which is characterized in our lab, is the HCM DNA polymerase processivity factor ppUL44. This 52 kDa protein binds to the dsDNA and directly interact with pUL54, stimulating its holoenzyme activity. To play its role ppUL44 forms head-to-head dimers and each domain form two topologically similar units. its dimerization is absolutely required for DNA binding and OriLyt-dependent DNA replication. Therefore, the dimerization of the UL44 can be a potential target for HCMV infection treatment considering the interaction interface shown in the crystal structure of ppUL44(1-290), and the fact that single amino acid substitutions affecting dimerization in vitro also impaired dsDNA binding (Komazin-Meredith *et al.*, 2008b)and prevented oriLyt dependent DNA replication in trans-complementation assays (Sinigalia *et al.*, 2008b). Moreover, disruption of PPI interaction between viral proteins is becoming an increasingly attractive strategy for the antiviral drugs development. In this context, several studies identified peptides and SMs disrupting the interaction between herpesvirus DNA polymerase holoenzymes and their respective processivity

factors (Loregian, Marsden and Palù, 2002; Palù and Loregian, 2013; Chen *et al.*, 2017). This is the first study exploring the possibility to directly target the dimerization of HCMV DNA polymerase accessory subunit ppUL44 as an antiviral strategy. Therefore, we performed a virtual screening aimed to identify SMs inhibiting ppUL44 dimerization. We identified 18 SMs potentially interfering with ppUL44 homodimerization and were initially screened for their ability to inhibit viral replication using a spontaneously fluorescent recombinant CMV derivative (TB4- UL83 EYFP). In parallel and in parallel MTT, MTS and Cell Titer Glo assays evaluated the cytotoxic profile of these SMs where for each SMs 2 different concentrations were tested (Fig. A.2.7, A.4.2). Overall, 4 SMs reduced viral replication in the absence of precipitates and evident cell cytotoxicity (Fig.A.2.8). As assessed by FRA assay only 3 of these SMs inhibited HCMV replication in a dose dependent fashion with ED_{50} in a low micromolar range, using GCV as control; while B1 failed to reproducibly inhibit HCMV replication, therefore was not considered for further experiments (Fig. A.2.8). The remaining 3 SMs were tested for their ability to inhibit GCV resistant strain. To this end we measured their ED50 towards the recombinant reporter virus AD169-GFP and its GCV-resistant counterpart bearing a mutation upon the UL97 gene by means of FRA and Viral Yield Assay (Fig. A.4.1, A.4.3). Our analysis revealed that, in both assays GCV inhibited replication of the *wt* virus 10 times more efficiently than GCV resistant strain, while both viruses appeared equally sensitive to our compounds. Unfortunately, treatment with two compounds resulted in high toxicity therefore were excluded from further characterization. We finally characterized the mode of action of B3 by evaluating its effect on HCMV genome replication and gene expression by means of PRA, Real-Time PCR and Western Blotting. The effect of B3 on viral genome replication was comparable to that of GCV, which is

known to inhibit the DNA polymerase and act chain terminator (Fig. A.4.4). Importantly as assessed by Western Blotting, neither B3 nor GCV inhibited IE gene expression until 24 h p.i., but both inhibited ppUL44 and pp65genome expression starting 48 h p.i (Fig. A.4.7/8) and viral genome copy number was drastically reduced starting from 72h.p.i. (Fig.A.4.5). These observations are consistent with a specific inhibition of HCMV life cycle by acting on viral DNA replication.

Overall, B3, was able capable of inhibiting the replication of different HCMV strains at concentrations not affecting cell growth and viability. ED_{50} values ranged from 4.2 μM (FRA with TB4-UL83-EYFP) to 7.9 μM (FRA with AD169-GFP), while those determined for GCV were comprised ranged between 0.7 μM (VYRA) and 2.3 μM (FRA with AD169-GFP). The ED_{50}values calculated here for GCV are compatible with those reported previously in the literature, with some variance being attributable to intrinsic differences between the different assays and viruses tested. For example, FRAs with the TB4-UL83-EYFP virus rely on the measurement of pp65 expression, which is expressed with an early-late kinetic, whereas the expression of the reporter gene in the AD169-GFP virus is under control of the IE promoter (Marschall *et al.*, 2000; Straschewski *et al.*, 2010). Importantly, B3 also retained antiviral activity against a GCV-resistant strain, suggesting that its mechanism of action against HCMV differs from viral DNA polymerase inhibitors. Although we did not formally prove here that B3 acts by disrupting the ppUL44 homodimer during viral infection, the fact that it inhibits HCMV genome replication as well as early and late gene expression starting from 48 h.p.i., without affecting production of IE viral antigens at earlier time points, in a very similar fashion to GCV is compatible with inhibition of ppUL44 homodimerization (He *et al.*, 2013; Weekes *et al.*, 2014; Hahn *et al.*, 2018).Importantly, the effect of B3 reported here on HCMV

replication is specific, since we recently demonstrated it did not affect either protein expression nor hepatitis C virus replication (HCV) in Huh7-Lunet cells (Elbadawy *et al.*, 2020). Even though B3 was less potent and more toxic than GCV, it might be useful as a starting platform for hit-to-lead optimization to develop more effective compounds as it has been performed with other PPI inhibitors endowed with antiviral activity against influenza virus (Lepri *et al.*, 2014) . Future work in our laboratory is currently focusing on characterizing in more detail B3 mechanism of action against HCMV, on the isolation of B3-resistant viral strains and on the identification of more potent analogs. Therefore, our results raise hopes in terms of potential use of ppUL44 dimerization inhibitors for the treatment of patients infected with drug resistant HCMVs.

A.6 Figures, Tables and Legends

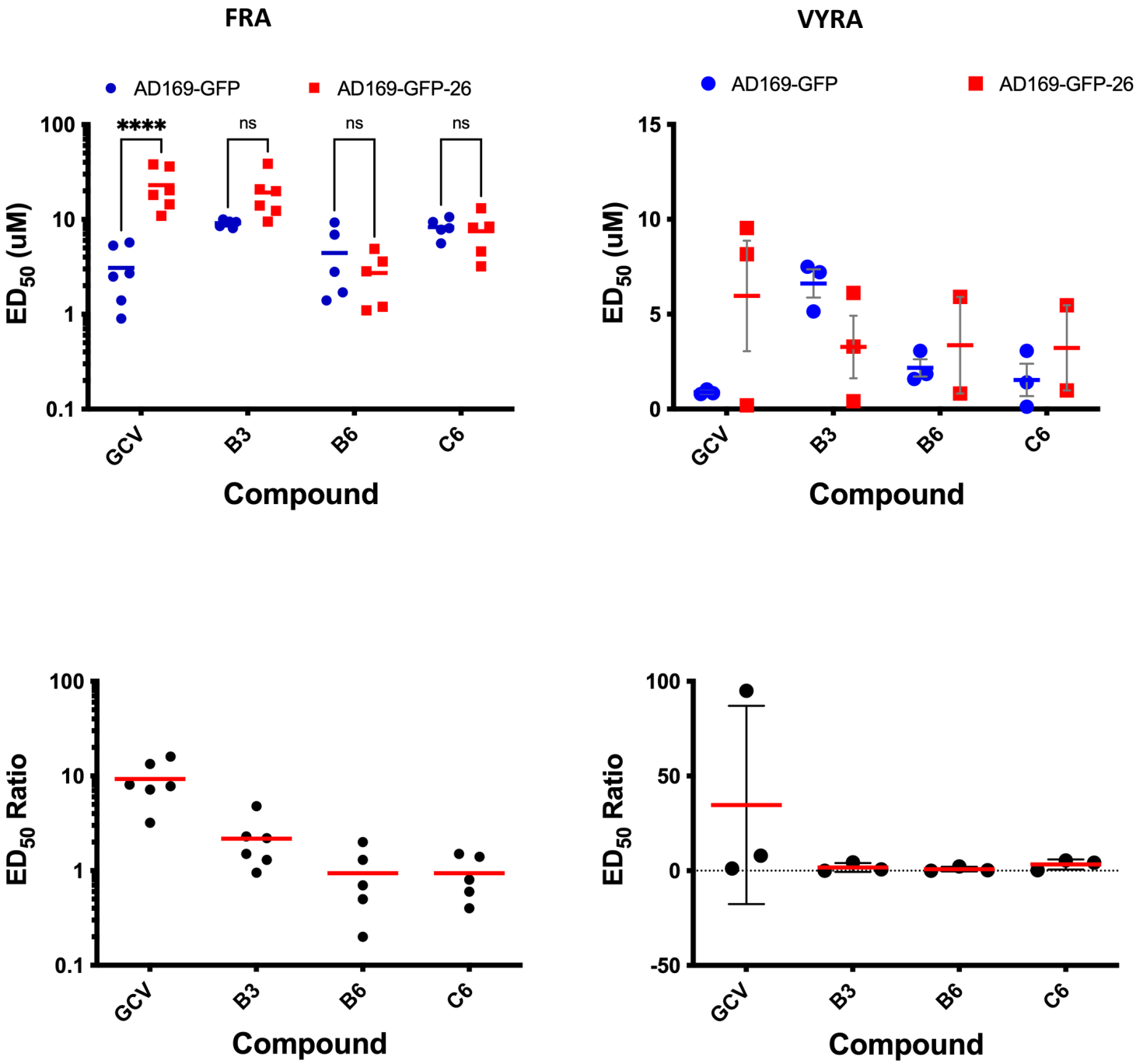


Figure A.4.1 SMs effect on inhibition of replication of GCV-resistant AD169-GFP26 virus. MRC5 were infected with either AD169-GFP virus or its GCV-resistant counterpart AD169-GFP26 at a MOI of 0.05 IU/cells and treated with increasing concentrations of the indicated compounds. At 7 days p.i., cells were lysed, and plates processed for FRAs, while supernatants were collected and used for VYRAs. For the latter experiments, MRC5 cells were infected with serial dilutions of supernatants derived from infected cells. At 7 days p.i., viral titers were calculated using the $TCID_{50}$ method. Data from both assays were used to calculate the ED_{50} relative to the two viruses as well as the ratio between the ED_{50} calculated for AD169-GFP26 and AD169-GFP, for every compound. Data shown are single measurements, means, and standard deviation of the mean relative to at least three independent experiments (Table A.4.1), along with the *p*-value relative to the Student's *t*-test the indicated groups; *: $p \leq 0.05$.

Table A.4.1

	FRA (n=6) ED_{50} (μM)[1]			VYRA(n=4) ED_{50} (μM)		
	AD169-GFP	AD169-GFP26	Ratio[2]	AD169-GFP	AD169-GFP26	Ratio[2]
GCV	2.3± 1.9	21.2± 8.7	17± 19.9	0.7± 0.4	4.7± 4.8	5.6± 5.2
B3	7.8± 2.6	16.8± 12.6	2.0± 1.6	5.1± 3.3	9.8± 15.6	1.4± 2.0
B6	4.4± 3.2	2.7± 1.6	0.8± 0.5	2.2± 0.8	3.5± 2.6	2.0± 1.7
C6	8.3± 1.9	7.4± 3.9	0.8± 0.5	1.5± 1.5	2.4± 2.7	3.2± 2.6

[1] ED_{50}, effective dose 50, the dose of compound that reduces by 50% the fluorescence (FRA) or virus titers (VYRA). [2] Ratio between values obtained for AD169-GFP26 and AD169-GFP. Data shown are mean and standard deviation of the mean relative to FRA (*left panel*) and VYRA (*right panels*) shown in Figure 6. The numbers between brackets indicate the number of independent experiments for each assay.

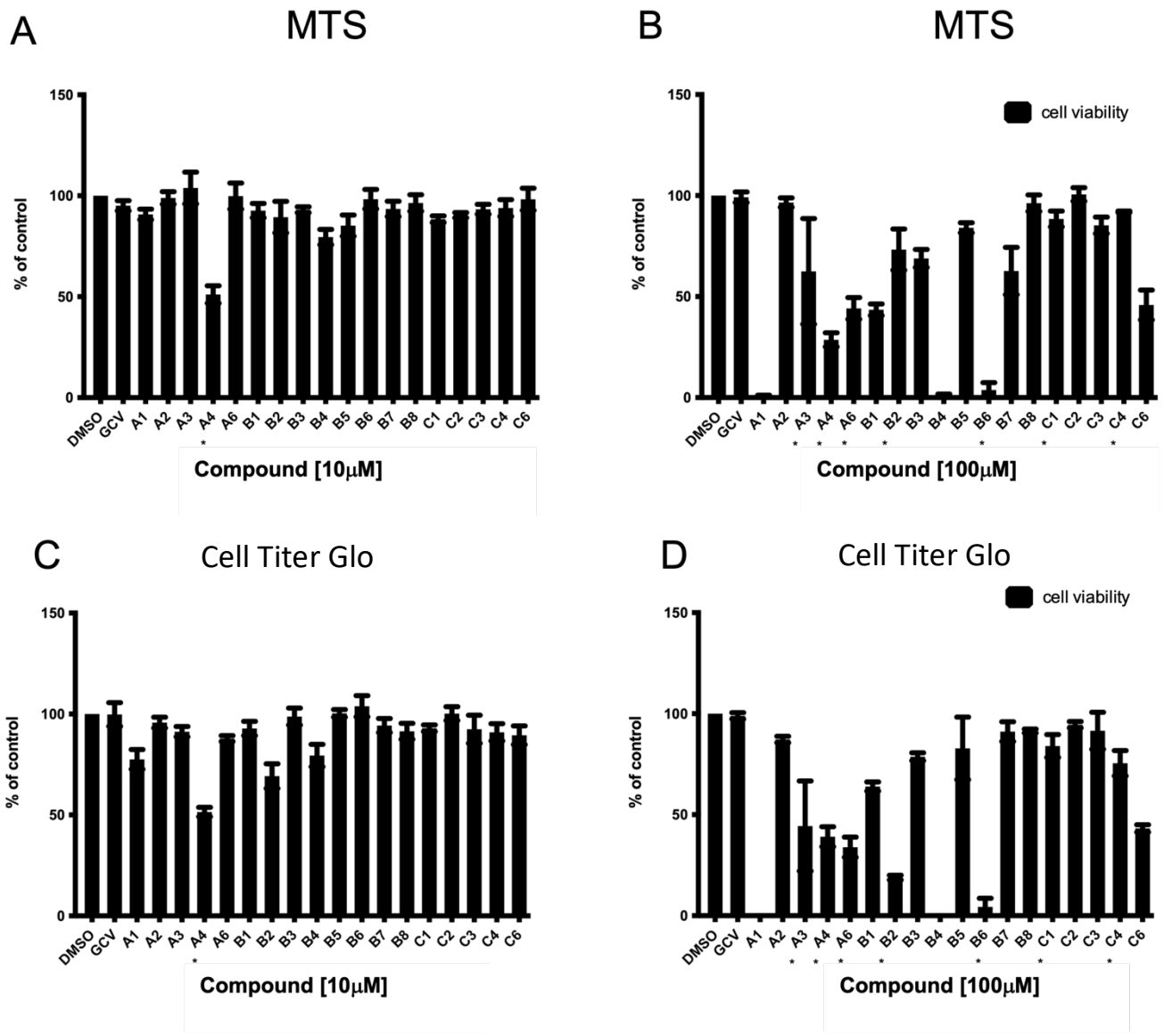


Figure A.4.2 Effect of SMs on cell viability. Cell viability was assessed by MTS (A,B) or Cell Titer Glo® assays (C,D), and data expressed as a percentage of DMSO-treated cells (black columns). The mean + standard error of the mean (SEM) relative to 3 independent experiments is shown. * indicates the presence of precipitates

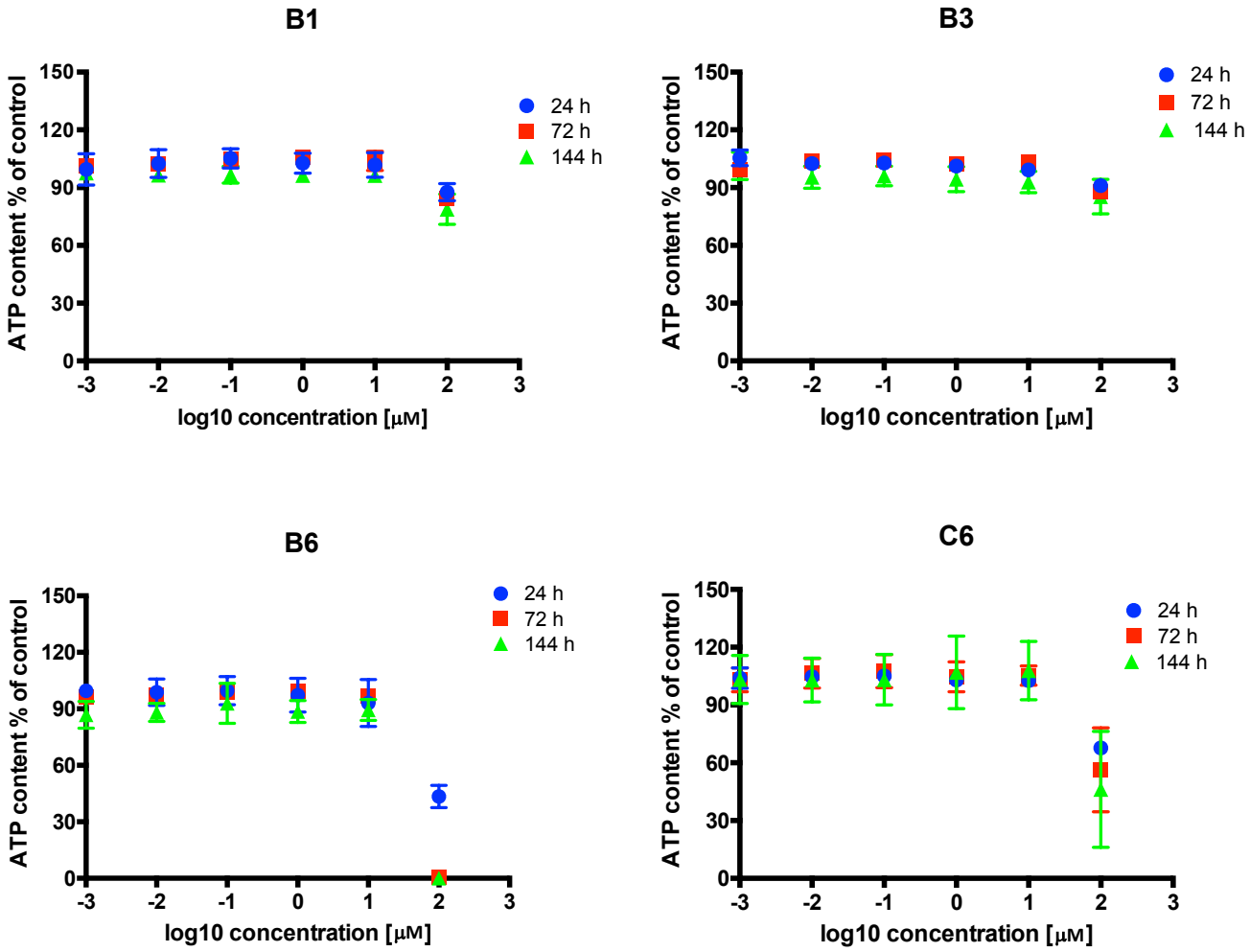


Figure A.4.3 Effect of selected 4 SMs on Cell viability and growth. MRC5 cells were treated with increasing concentrations of indicated compounds or solvent only, as described in Material and Methods section. At the indicated time point post treatments cells were processed for intracellular ATP quantification. Data shown are the mean + standard error of the mean relative to 3 independent experiments.

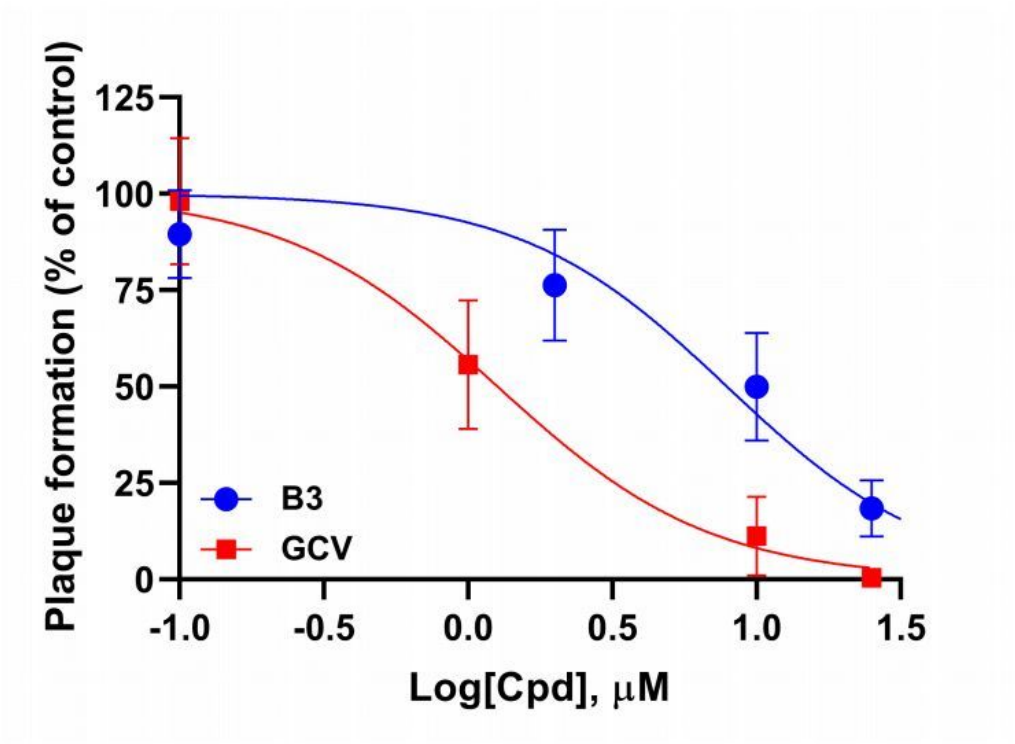
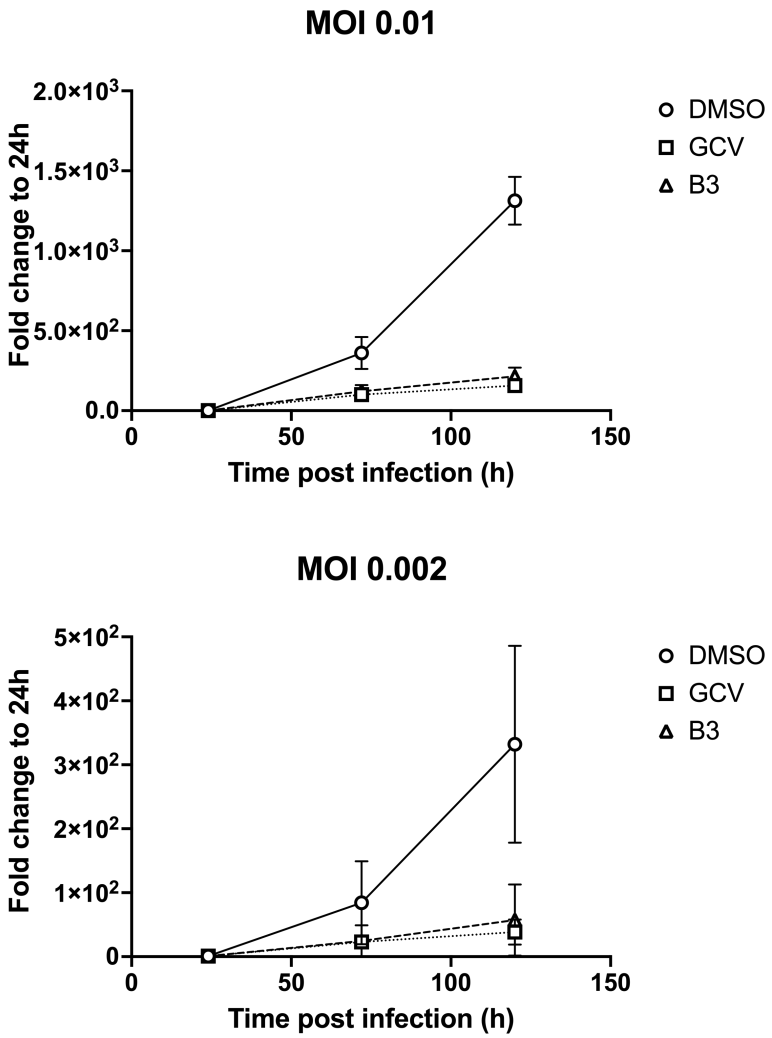
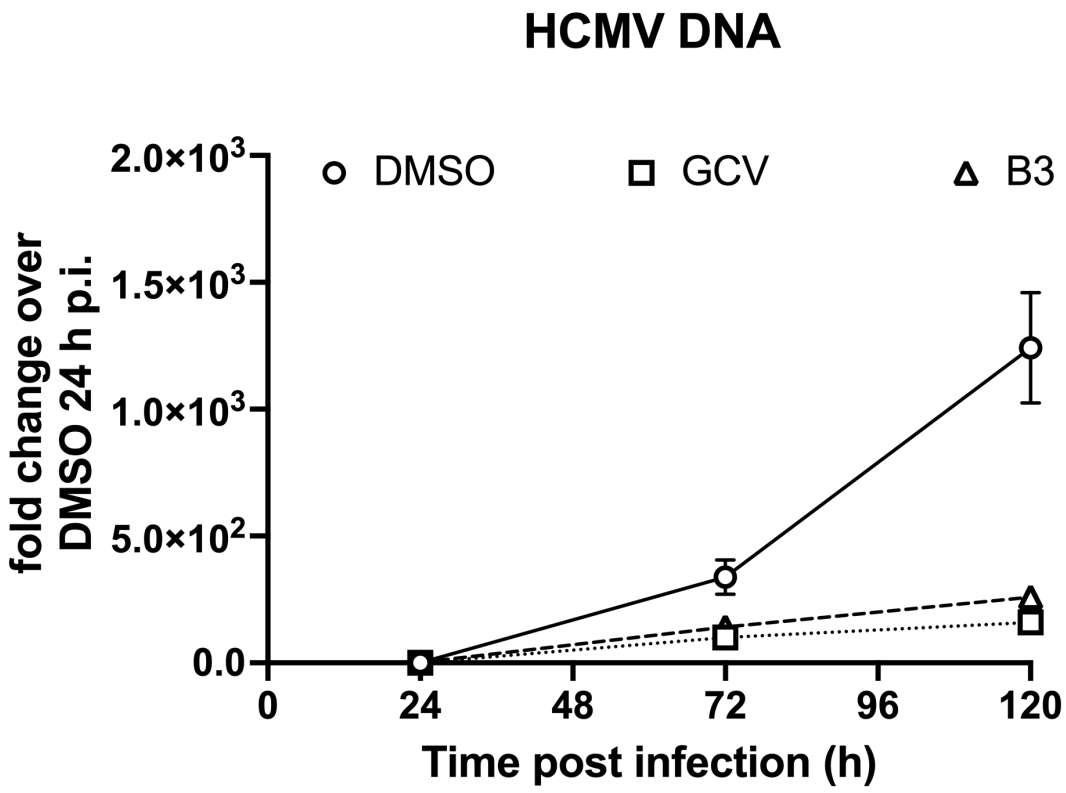


Figure A.4.4 Inhibition of AD169 replication by B3 in PRA. Dose-response curves for B3 (blue) or GCV (red) were obtained by infecting HFF cells with HCMV AD169 and then treating them with different concentrations of the indicated compounds. Data shown are the means ± standard deviation of the mean relative to four independent experiments performed in duplicate.



	MOI 0.01 (n=4)			MOI 0.002 (n=4)		
	24h	72h	120h	24h	72h	120h
DMSO	1	360.5± 100.1	1312± 149.2	1	84.4± 64.9	332.2± 153.9
GCV	1	100.3± 22.8	156.7± 29.5	1	23.02±4.2	38.4± 19.6
B3	1	120.6± 39.3	215.1± 54.7	1	24.6± 24.5	57.3± 55.6

Figure A.4.5 Effect of B3 on viral genome replication. B3 impairs HCMV genome replication. MRC5 cells were infected with HCMV TB4-UL83-EYFP and treated with the indicated compounds or vehicle alone (DMSO) as described in the Materials and Methods section. At the indicated time points p.i., cells were lysed and processed for qPCR to detect the presence of viral and host cell genomic DNAs as described in the Materials and Methods section. Data are shown as HCMV DNA foldchanges with respect to the respective -treated cell at 24 h p.i. Means +standard deviation of the mean relative from to two independent experiments performed in quadruplicate.



MOI 0.01 (n=4) DMSO normalized			
	24h	72h	120h
DMSO	1	360.5± 100.1	1312± 149.2
GCV	1.12± 0.2	100.1± 15.4	159.6± 10.0
B3	1.3 ±0.3	141.7± 19.1	259.3± 18.6

Figure A.4.6 Effect of B3 on viral genome replication. B3 impairs HCMV genome replication. MRC5 cells were infected with HCMV TB4-UL83-EYFP and treated with the indicated compounds or vehicle alone (DMSO) as described in the Materials and Methods section. At the indicated time points p.i., cells were lysed and processed for qPCR to detect the presence of viral and host cell genomic DNAs as described in the Materials and Methods section. Data are shown as HCMV DNA foldchanges with respect to the DMSO-treated cell at 24 h p.i. Means +standard deviation of the mean relative from to two independent experiments performed in quadruplicate.

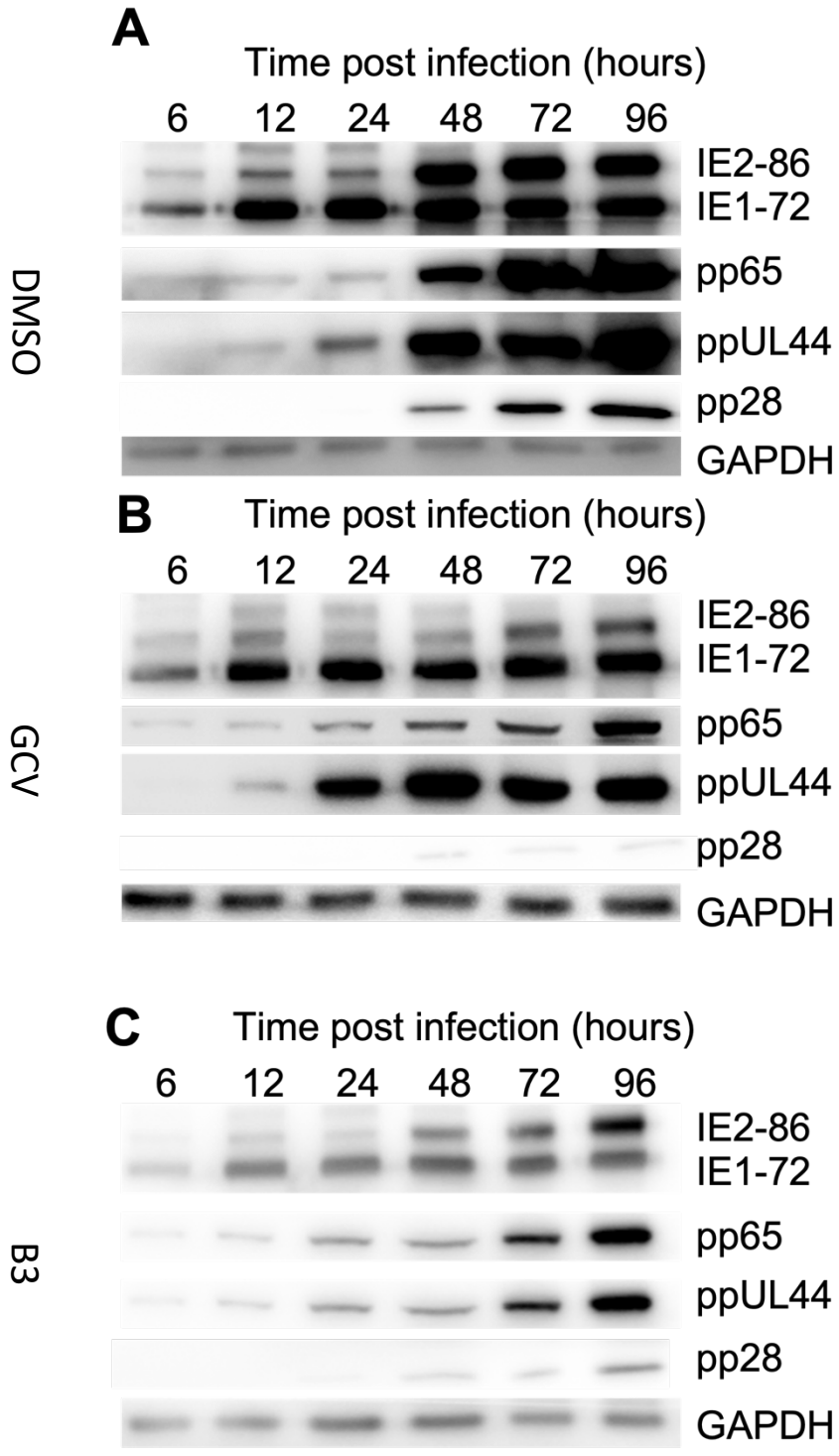


Figure A.4.7 B3 specifically impairs early and late HCMV AD169 gene expression. MRC5 were infected with HCMV AD169 and treated as described in the Materials and Methods section with either vehicle alone (A, DMSO 0.5%), GCV (B, 16 uM in DMSO 0.5%) or B3 (C, 50 uM in DMSO 0.5%). At the indicated time points p.i., cells were lysed and processed for Western Blotting to detect the expression of the immediate early IE1/2 antigens, the early-late antigens ppUL44 and pp65 as well as the late antigen pp28. GAPDH was also detected as loading control.

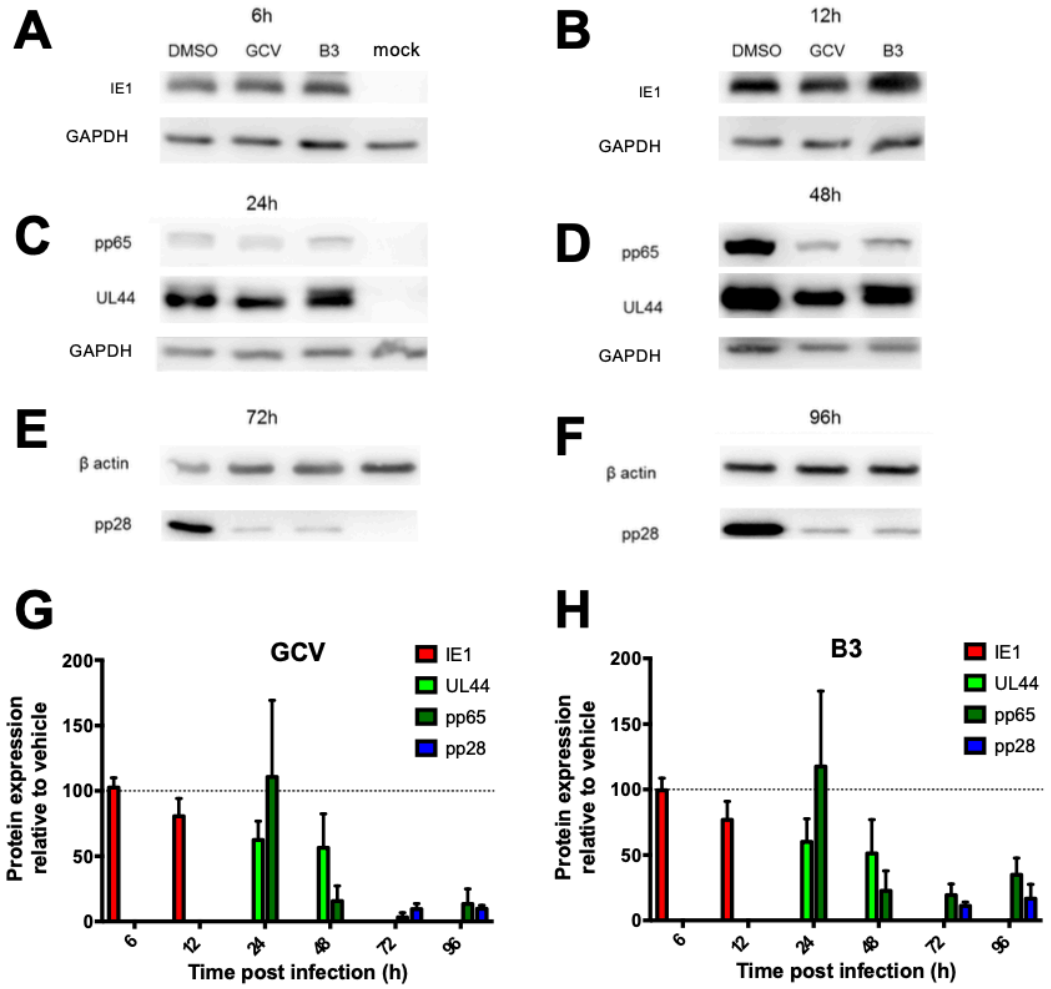


Figure A.4.8 B3 specifically impairs early and late HCMV AD169 gene expression. MRC5 were infected with HCMV AD169 and treated as described in the Materials and Methods section. At the indicated time points p.i., cells were lysed and processed for Western blotting to detect the expression of the immediate early IE1 antigen (**A**,**B**; at 6 and 12 h p.i.), the early-late antigens ppUL44 and pp65 (**C**,**D**; at 24 and 48 h p.i.), and the late antigen pp28 (**E**,**F**; at 72 and 96 h p.i.). GAPDH or β-actin were also detected as loading controls. (**G**,**H**): Loading controls were used to normalize signal intensity relative to each antigen after treatment with GCV (**G**) or B3 (**H**). Lysates of mock infected cells were analyzed to verify antibody specificity. Data shown are the mean + standard deviation of the mean relative to three independent experiments

PART B:

Identification of the nuclear proteome from all human viruses by a comprehensive analysis of classical nuclear localizations

B.1 ABSTRACT

Our research group previously defined the nuclear proteome of all human viruses, discriminating between viral proteins translocated in an IMPα/β1 dependent or independent process by combining bioinformatics analysis with extensive functional characterization of viral cNLSs. This study represents an unprecedented opportunity to compare how viruses differently interact with the host cell nuclear transport machinery, with important implications for the development of broad-range host targeted antivirals. In depth functional validation of identified putative classical nuclear localization signals (cNLSs) led to the discovery of more than 500 novel viral cNLS. We also report the first characterization of the nuclear import process of Human Polyomaviruses (HPyVs) Large T antigens (LT) as well as of the cNLS involved. Although LT from all 14 HPyVs bear a functional cNLS, the latter are extremely heterogenous, both in terms of activity and structural organization. Importantly, cNLS activity mirrored the levels of nuclear accumulation of full-length proteins, with lowest activity associated to HPyV7. Surprisingly, while most HPyVs bear one or more monopartite cNLS, four of them bear a bipartite cNLS. Clearly, such structural differences suggest an important role in conferring binding abilities to specific IMPα isoforms with potential implication for viral tropism determination. Furthermore, among the 26 top ranked cNLS based on cNLS mapper score, two extremely well conserved cNLS in orthologues of Vaccinia Virus proteins A19 and N2 were identified. Both proteins localized in the cell nucleus via energy and IMPα/β-dependent process, and their nuclear import could be abolished by site specific mutagenesis of the cNLSs, thus A19 and N2 mutant derivatives failed to localize in the nucleus.

RIASSUNTO

Il nostro Gruppo di ricerca ha precedentemente identificato il proteoma nucleare di tutti i virus infettanti l'essere umano, distinguendo tra proteine virali che traslocano all'interno del nucleo della cellula infetta in modo IMPα/β1 dipendente o meno, combinando analisi bioinformatiche estese anche alla caratterizzazione funzionale delle sequenze di localizzazione nucleare (NLS) virali. Questo studio presenta l'opportunità senza precedenti di comparare la diversa interazione tra virus differenti con l'apparato di trasporto al nucleo della cellula infetta, con importanti implicazioni sullo sviluppo di nuovi target terapeutici antivirali a largo spettro. Una profonda analisi funzionale sui classici NLS (cNLS) putative identificati ci ha portato alla scoperta di più di 500 proteine codificanti cNLS. Siamo riusciti anche a fare una prima caratterizzazione del processo di import nucleare delle proteine Large T antigen (LTA) dei Polyomavirus (HPyV) e delle cNLS coinvolte. Nonostante i LT di 14 HPyV presentavano cNLS funzionali, queste erano molto diverse tra di loro sia in termini di struttura che di attività. Le attività delle cNLS hanno riflesso I livelli di accumulo nucleare delle proteine full-length, con l'attività più bassa associata alla cNLS di HPyV7. Nonostante molti HPyV codificano per una o più cNLS monopartite, 4 di esse presentavano cNLS bipartita. Chiaramente queste differenze strutturali influenzano affinità verso l'apparato IMPα/β1 dipendente ed il tropismo del virus. Inoltre, 2 tra le 26 cNLS con il punteggio di cNLS mapper più alto identificate con i nostri studi, sono altamente conservate e presentano vari ortologhi di due proteine, A19 e N2, della famiglia dei Poxivirdae. Entrambe le proteine localizzano nel nucleo della cellula ospite, in un pathway attivo IMPα/β1 dipendente, e la loro traslocazione nucleare viene inibita in presenza di mutazioni sito-specifico.

B.2. INTRODUCTION

B.2.1 Nuclear Pore Complexes (NPC) and Nuclear Transport

In Eucaryotic cells, the nuclear and the cytoplasmic compartments are separated by a double membrane named the nuclear envelope (NE), formed by an outer (ONM) and an inner (INM) phospholipid bilayer. While the ONM is contiguous with the endoplasmic reticulum (ER) membrane, the INM surfaces the nucleoplasm (Burns and Wente, 2012), with several INM proteins having important functions in chromosome organization, DNA repair, nuclear structure, and transcriptional control (Katta, Smoyer and Jaspersen, 2014). The selective exchange of macromolecules between the two compartments occurs through Nuclear Pore Complexes (NPCs), large multiprotein assemblies which consist of multiples of ~30 different nucleoporins (Nups) (Kabachinski and Schwartz, 2015), embedded where the ONM and the INM fuse . The central NPC channel is lined with Nups carrying phenylalanine-glycine (FG)-rich repeats that promote receptor-mediated transport of soluble cargoes containing specific localization signals and limit diffusion of inert macromolecules. Passage across the NPC can occur either by passive diffusion or via an active process. Indeed, proteins up to 240 kDa can passively diffuse across the NPC, with a diffusion rate inversely proportional to their size. On the other hand, active nucleocytoplasmic transport is a directional, signal- and energy-dependent process. Specific nuclear transporter receptors (NTRs) belonging to the Karyopherin superfamily can recognize specific localization sequences on their cargoes and translocate them either from cytoplasm into the nucleus (importins; IMPs) or vice-versa (exportins; EXPs), facilitated by the GTPase Ran. From a structural point of view, karyopherins are superhelices with an inherent flexibility, which is functionally important for resolving the appropriate conformational changes associated with cargo binding and release (Fig. B.2.1).

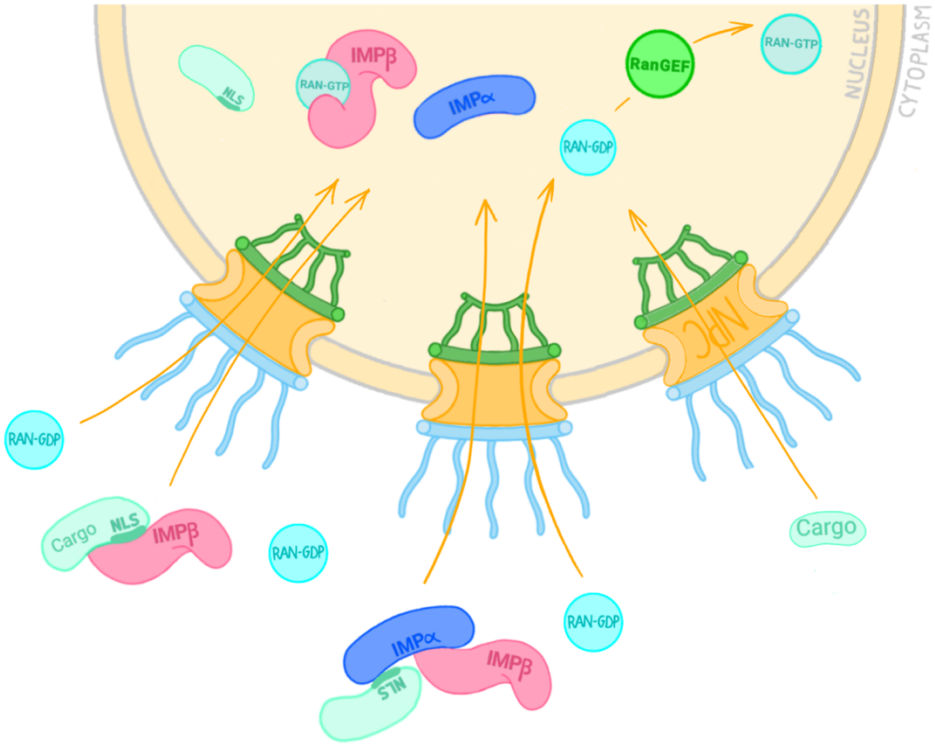


Figure B.2.1 Generic view of Nuclear Import Pathways. Cargo translocation is shown via passive diffusion (right), via IMPα/β1 pathway (centre) or directly via IMPβ pathway (left). Active process is mediate by RAN protein cycle, and cargo are transported in the nucleus through recognition of Nuclear Localization Signals (NLS). Different NLS have different affinities to different transporters.

B.2.2 Transporters

B.2.2.1 Importin-β (IMPβ)

IMPβ is the major component of nuclear transport protein within the karyopherin superfamily.The human genome encodes at least 20 importin β isoforms, including 10 IMPs, 7 EXPs and 2 bidirectional receptors . All β karyopherin share common features and are endowed with a similar architecture. These super-helical structures are formed from modular folding units and are composed by 19-21 HEAT tandem sequence repeats each comprising ~40-45 amino acids that form two amphiphilic αhelices (A helix and B helix) connected by a short loop . The intramolecular helix-helix interaction is supported by a hydrophobic core that is extended along the superhelix and confers structural flexibility. Indeed, their flexibility is the key to their versatility in binding cargoes of different shape and size (Conti, Müller and Stewart, 2006). The protein coils into a short

superhelix, with extensive interaction surfaces both on the inside and the outside of the superhelix. IMPβs functional domains comprehend an N-terminal Ran-GTP binding domain, a central zone that interacts with the nuclear pore, and a C-terminal domain interacting either with IMPα or a cargo (Fig.B.2.2). IMPβs transport their cargoes upon recognition of a localization signal localized in the latter that determines the direction of the translocation. Such recognition can occur either directly or through adaptors that link the cargo to IMPβs. The best studied adaptors are the IMPα proteins, also known as karyopherins-α (KPNAs), which allow IMPβ1 to interact with a subset of cargoes bearing a specific class of NLSs also known as classical NLSs (cNLSs) (Görlich *et al*., 1994) .The energy required for this process is provided through a small Ras-related nuclear protein, RAN, which is highly enriched in its GTP-bound form in the nucleus and has a high binding affinities to IMPβs, whereas its GDP-bound form is more concentrated in the cytoplasm, with very low binding affinities to IMPβs. By binding to Ran-GTP, IMPβ dissociates from the import complex, releasing the NLS bearing cargo, thus, the IMPβ-Ran-GTP complex is directed to the cytoplasm.

Figure B.2.2 IMPβ 3D structure. 3D structure of IMPβ from Protein Data Bank (PDB, code 1O6O)

B.2.2.2 Importin-α (IMPα)

The 7 human IMPαs identified so far can be classified into three subfamilies of IMPα-1/S, IMPα-2/P and IMPα-3/Q. All IMPα isoforms share important common features and are functionally and structurally divided into two domains: an N-terminal and a C-terminal domain. The approximately 70 amino acids long N-terminal domain is responsible for binding to IMPβ1 as well as auto inhibition of cargo recognition in the absence of the latter and it has been therefore dubbed the IMPβ1 binding (IBB) domain. The C-terminal domain is responsible for binding to the cargo proteins and is formed by a helical core containing 10 stack of Armadillo repeats. Each Armadillo repeats is a relatively hydrophobic sequence of 40-43 amino acids, reminiscent of to the HEAT repeats of IMPβs with stacks of Arm repeats generating super helical solenoids, which are highly specialized to participate in protein-protein interaction. Each Arm repeats forms three helices (named H1, H2 and H3), with the H1 helix splitting in two helices (H1-A and H1-B), conferring a more rigid structure as compared to the other 2 helices of the HEAT repeats. However, in all IMPα isoforms 2 Arm repeats deviate from the standard three helices structure: Arm1 lacking the first helix and Arm5 with H1 and H2 fused together. Stacking of Arm repeats results in an extended concave surface, thus creates two NLS-binding sites, with Arms repeats 2-4 form the major and Arms repeats 6-8 form the minor NLS binding sites. Although the major NLS binding site is conserved among all known IMPα isoforms, the minor NLS binding site shows significant differences. The IBB domain contains a bipartite basic sequence ('RRRR$(X)_{17}$KRR') that in the absence of IMPβ, it folds back to occupy the NLS-binding site of IMPα. The first basic sequence binds the minor binding site, and the second motif occupies the major binding site. This interaction between the Arm repeats and the IBB domain prevents IMPβ1 to

bind to an empty IMPα, representing a further regulation mechanism that halts futile nuclear transport. (Fig. B.2.3).

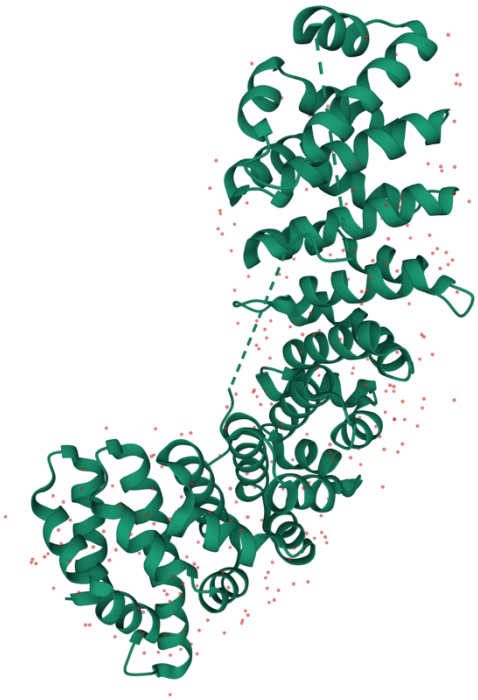


Figure B.2.3 IMPα 3D structure. 3D structure of IMPα from Protein Data Bank (PDB), code 4B8J

B.2.3 Localization signals

Active nucleocytoplasmic transport relies on recognition of specific localization signals on cargoes called nuclear localization signals (NLSs). There are different classes of NLSs, such as Proline-tyrosin NLSs (PY-NLSs), Arginine-rich NLSs (R-rich NLSs) and Classical NLSs (cNLS). cNLSs are recognized by IMPα for nuclear protein import and consist of either one (monopartite) or two (bipartite) stretches of basic amino acids (Dingwall and Laskey, 1991). The best characterized cNLS is the one of the Simian Virus 40 (SV40) large T antigen NLS (^{126}P**KKK**R**R**V^{132}), a highly basic sequence that exemplifies the monopartite cNLS while the nucleoplasmin NLS sequence (^{155}**KR**PAAT**KK**AGQA**KKKK**^{170}) exemplifies the bipartite cNLSs. Both monopartite and bipartite NLSs are recognized by an IMPα

adapter, which, when bound by IMPβ1, can mediate transport through the nuclear pore complex into the nucleoplasm (Loveland *et al.*, 2015).Monopartite cNLSs and the longer C-terminal stretch of basic residues in bipartite cNLSs bind preferentially to the major NLS binding site (Fontes, Teh and Kobe, 2000) of IMPα which lies nearer the N-terminal domain. The minor binding site binds the shorter stretch of basic residues in bipartite cNLSs. A class of monopartite NLSs binding preferentially to the minor NLS binding site of IMP has been described by Kosugi *et. al* (Shunichi Kosugi, Hasebe, Matsumura, *et al.*, 2009). The efficiency of nuclear transport depends on NLS: IMPα binding affinity (Hodel *et al.*, 2006; Smith *et al.*, 2018a) .At the major binding site, side chains of the residues of monopartite cNLSs, ^{26}P**KKK**R**R**V^{132} in the case of T-ag, accommodate in the IMPα pockets termed P1–P6, where the contribution of the residues in position P1 and P6 appear to be minimal (Hodel *et al.*, 2006) compared to the ones accommodating in positions P2-P5, with residue in P2 dominating the energy profile of the interaction. Structural (Conti *et al.*, 1998) and thermodynamic (Hodel, Corbett and Hodel, 2001) studies have shown that a monopartite cNLS requires a lysine in position P2 position, followed by basic residues in positions P3 and P5 to yield a loose consensus sequence of K(K/R) X (K/R) (P2-P5) (Smith *et al.*, 2018a) IMPα recognizes cNLSs with a range of different residues in P4 position, including K, R, S, Y, L, P, A and V, suggesting that the residue at P4 position has less impact in respect to the other residues present on the cNLS consensus sequence (Shunichi Kosugi, Hasebe, Matsumura, *et al.*, 2009). Several studies in mammalian systems, however, suggest that certain cNLSs, although matching the K-(R/K)-X-(R/K) consensus, are not functional in nuclear targeting. Indeed, a recent study showed that the nature of residues at P4 position modulates cNLS function with residue G being associated with low activity (Smith *et al.*, 2018b), consistent with another study in

Saccharomyces cerevisiae, which showed that the residues with the biggest negative impact on nuclear targeting at the P4 position are D, E and G (S. Kosugi *et al.*, 2009).

In contrast, PY-NLSs have diverse sequence and larger structure compared to cNLSs and are directly recognized by IMPβ2 by multiple interactions between the weak NLS and IMPβ2. PY-NLSs sequence is structurally disordered, overall positive charged and composed of and binding epitopes composed of a loose N-terminal hydrophobic or basic motif and a C-terminal R-X_{2-5}P-Y motif, thus are necessary and sufficient for nuclear import mediated by IMPβ2, while also being required for receptor binding and protein function, respectively (Lange *et al.*, 2008). Another non-canonical NLS class is represented by Arginine-rich (R-rich) NLSs which is directly recognized by IMPβ1. R-rich domains present in Human Immunodeficiency Virus Type 1 Tat and Rev. IMPβ also binds various other NLSs such as CREB, ribosomal proteins, the human T-cell leukemia virus type 1 protein Rex, PTHrP, cyclin B1, Smad3, SREBP-2, and TRF NLSs (Chook and Blobel, 1999; Mach *et al.*, 2005). The accumulation and import of cNLS bearing cargoes in the nucleus are both affected by the concentration of the IMPα receptor and by the affinity of the cNLS cargo for IMPα itself (Riddick and Macara, 2005)

B.2.4 RAN Cycle

Ran GTPase belongs to the Ras superfamily of small GTPases. It possesses a distinctive acidic C-terminal DEDDDL motif, and it is known to regulate nucleocytoplasmic trafficking as well as mitotic spindle and nuclear envelope formation. Ran and its associated proteins are constantly shuttling into and out of the nucleus, but GTP-bound Ran is predominantly distributed in the nucleus, while GDP-bound Ran is predominantly cytoplasmic. Ran constantly cycles between GDP- and GTP-bound states, as controlled

by Ran-Guanine nucleotide exchange factors (Ran-GEFs) and Ran-GTPase-activating proteins (Ran-GAPs) proteins enabling it to shuttle cargoes in an accurate spatial and temporal manner. When IMP-cargo complex is imported into the nucleoplasm, free nuclear Ran-GTP binds to the Ran-binding domain of IMPβ, thus triggering a conformational change in IMPβ which destabilizes the interaction between IMPβ and IMPα/cargo complex. Subsequently, the IBB domain then releases the cargo via a self-inhibitory mechanism. On the other hand, the Ran-GTP: IMPβ complex is recycled to the cytoplasm across the NPC, whereas the free IMPα binds to exportin cellular apoptosis susceptibility (CAS) and Ran-GTP for recycling (Moroianu, 1999). In the cytoplasm, Ran-GAP activates the GTPase activity, thus resulting in the formation of Ran-GDP, which is endowed with low affinity for IMPβ. This event triggers the release of IMPβ from Ran-GDP. The free IMPβ can then initiate another round of nuclear import, while free Ran-GDP is recycled back into the nucleus by forming a complex with the specific nuclear importer of Ran, nuclear transporter factor 2 (NTF2) (also known as p10 or pp15) (Ribbeck *et al.*, 1998). It has been shown that NTF2 forms a homodimer, which creates two reactive hydrophobic pockets to bind with Ran-GDP while the hydrophobic ends of NTF2 dimer are able to interact with FXFG motifs from nucleoporins to induce the nuclear translocation (Bullock *et al.*, 1996; Stewart, Kent and McCoy, 1998; Bayliss *et al.*, 2002). Once this complex reaches the nucleus, nuclear RCC1 (Ran-GEF) activates the exchange of Ran bound-GDP with GTP, thus decreasing the affinity of Ran for NTF2 and increasing the affinity its affinity for IMPβ. Then, the free Ran bound-GTP binds another IMPβ-cargo complex.

Figure B.2.4 RAN-GFP protein binding to IMPβ. 3D interaction of RAN-GTP protein to IMPβ from PDB, code 1IBR.

B.2.5 Viruses and Nuclear Trafficking

As obligate intracellular parasites, viruses are critically dependent on numerous host cell proteins and pathways for their replicative cycle. A common theme among many viruses is their limited coding capacity and, therefore, their absolute dependence on host proteins and pathways for productive infection. Nucleocytoplasmic shuttling often plays a critical role during viral infection. Most DNA viruses and even some RNA viruses replicate their genomes within the host nucleus, and crossing the NE represents a barrier to infection. To overcome the barrier presented by the nuclear membrane and gain access to the nucleus, virally encoded proteins have evolved ways to exploit components of the nuclear transport machinery. Each component of the nuclear transport system potentially represents a viable target that can be routed by viruses during infection to allow and regulate entry of viral genomic information, export of viral

mRNA, and passage of viral proteins bidirectionally across the NE. The most direct approach for viral proteins to traverse the NE is to target the NPC itself. Generally, this phenomenon is reserved for capsid interactions to bring viral genomic information into the nucleus. Herpes simplex virus type 1 (HSV-1) UL36 is a preformed tegument protein, which aids in docking the viral capsid to the NPC by bridging the capsid with Nup358 of the NPC (Copeland, Newcomb and Brown, 2009). Similarly, the capsid protein of HIV-1 interacts with Nup153 to mediate import of the pre-integration complex (PIC) (Matreyek *et al.*, 2013). Aside from viral capsids reaching the nucleus, many viral proteins themselves must be transported into and out of the nucleus by employing a variety of nucleoporins or karyopherins. Indeed, some viral proteins interact with the NPC directly such as BGLF4 from Epstein–Barr virus (EBV) and HIV-1 Vpr (Fouchier *et al.*, 1998; Jenkins *et al.*, 1998; Chang *et al.*, 2012). A simple yet effective approach to target cellular IMPs is through molecular mimicry of a cNLS, which would allow viral proteins to interact with IMPα (Tessier *et al.*, 2019). Because of the sequence characteristics and predictability of these peptide motifs, many viral cNLSs have been discovered in a diverse range of viruses. For example, influenza A virus (IAV) NP and PB2, HIV-1 integrase and Vpr, HAdV E1A, Human Papilloma virus (HPV) E2, HSV-1 pUL30, and many more all contain viral cNLSs (Köhler *et al.*, 2001; Nitahara-Kasahara *et al.*, 2007; Ao *et al.*, 2010; Bian and Wilson, 2010; Cohen *et al.*, 2014). As a result of their low sequence complexity, changing a single amino acid can easily destroy or create a novel interaction (Davey, Cyert and Moses, 2015). An alternative approach for viral proteins to target the nuclear import machinery is to directly bind IMPβ for entry into the nucleus (Tessier *et al.*, 2019). Multiple viruses have been shown to bind IMPβ1 directly, surpassing the need for an adaptor protein. HIV-1 Rev was the first identified example of this, although similar

examples can be seen with HIV-1 Tat, human T-cell leukaemia virus (HTLV) Rex, HSV-1 capsid protein, hepatitis B virus (HBV) core protein, and HAdV VII (Palmeri and Malim, 1999; Ojala *et al.*, 2000; Arnold *et al.*, 2006; Chen *et al.*, 2016).

Together, these examples demonstrate the need for continued mechanistic investigation into nucleocytoplasmic transport, as novel strategies relating to transport, and possible therapeutic interventions.

B.2.6 IMPα/β inhibitors

B.2.6.1 Ivermectin (IVM)

In the last years several small molecules were developed, targeting the two best characterized nuclear import pathways, i.e the IMPα/β and the IMPβ dependent pathways, yet the simple fact that the inhibitors target transporters essential for cell function means that toxicity is an inevitable corollary of their use, limiting clinical application. The first FDA-approved compound was the anti-parasitic drug Ivermectin (IVM) (Jans and Wagstaff, 2020). IVM is a semisynthetic derivative of the naturally produced Avermectin B_1, developed by Merk & Co. and frequently used in veterinary medicine since the 80s, due to its broad-spectrum anti-parasitic activity, high efficacy and wide margin of safety. The first formula was launched in 1987 against human onchocerciasis (González Canga *et al.*, 2008). IVM is composed by 22,23-dihydro-avermectin B1a and 22,23-dihydro-avermectin B1b, and it's available for both oral and parenteral administration.

In 2011 IVM was identified as an inhibitor of Human Immunodeficiency Virus-1 (HIV-1) integrase (IN) recognition by IMPα/β. HIV-1 IN nuclear import is a potential therapeutic target since it plays a crucial role in viral genome nuclear delivery in quiescent cells

(Wagstaff *et al.*, 2012), thus demonstrating the anti-viral activity of IVM. Recently, it has been shown that IVM's effect on IMPα/β nuclear import dependent is related to its ability to dissociate the preformed IMPα/β heterodimer, as well as prevent its formation, through binding to the IMPα armadillo (ARM) repeat domain to affect the protein's thermal stability and α -helicity. Since its discovery, IVM has been shown to reduce nuclear import of a number of viral proteins and to inhibit viral replication of several viruses see Table B.2.3, thus emerging as a promising broad-range antiviral molecule.

Viruses targeted by IVM include both nuclear replicating viruses (such as HIV, Adenoviruses and PRV1) and cytoplasmic replicating viruses (such as Zika, Dengue, yellow fever, West Nile, Hendra and SARS-COV2).

B.2.6.2 Other IMPs inhibitors

Like IVM, other molecules have been identified as inhibitors of nuclear import. For example, several small molecules with anti-tumour potential that target IMPβ1 have been characterized, like Importazole and INI-43.

In 2011 Importazole is a 2,4-diaminoquinazoline was discovered as inhibitor of the transport receptor IMPβ as disrupts RanGTP binding to IMPβ1(Soderholm *et al.*, 2011).

INI-43, 3-(1H-benzimidazol-2-yl)-1-(3-dimethylaminopropyl)pyrrolo[5,4-b]quinoxalin-2-amine, targets sites of IMPα and RanGTP binding to IMPβ1 and shows anticancer therapeutic potential against esophageal and cervical cancer in a mouse xenograft model (van der Watt *et al.*, 2016).Other inhibitors targeting IMPβs are Peptide inhibitor M9M, which inhibits nuclear import of IMPβ 2 cargoes (e.g. hnRNP A1 and M and HuR),

but not IMPα/β1 cargoes, and Karyostatin 1Ae, that as Importazole, disrupts RanGTP binding to IMPβ1, shown to inhibit IMPα/β mediated nuclear import of NFAT in cancer cell lines (Hintersteiner *et al.*, 2010). Instead, Bimax1/2, nucleartransportmodifier (NTM) cSN50.1 and Gossypol are proved to target IMPα, with NTM capable of targeting also IMPα5 and IMPα1 (Kosugi *et al.*, 2008; Gronewold, Horn and Neundorf, 2018; Lopez-Denman *et al.*, 2018). In particular Bimax1/2 has been reported to Inhibits IMPα/β mediated nuclear import and growth of yeast, NIH 3T3 cells (Kosugi *et al.*, 2008), NTM has been showed to reduce nuclear import of pro-inflammatory NF-k B, STAT1a and AP-1, and when fused to a cell-penetrating peptide, can target to the nucleus of cancer lines to increase efficacy of co-administered doxoru-bicin in cell killing (Gronewold, Horn and Neundorf, 2018). Gossypol has been reported to inhibits WNVb protein NS5 nuclear import, and virus production (Lopez-Denman *et al.*, 2018), Hendra Virus V protein nuclear import and virus production (Atkinson *et al.*, 2018) and to Reduce viability of prostate cancer cell lines and induces apoptosis/activates p53 (Volate *et al.*, 2010).

B.2.6.3 Cargo-specific nuclear import inhibitors

This class of inhibitors target the host-pathogen interface and are of great interest since in contrast to IMP inhibitors their use could not be intrinsically linked with toxicity, thus targeting host-pathogen interface would also avoid issues of resistance insurgence. The first cargo-specific nuclear transport inhibitor described was Mifepristone as a specific inhibitor of recognition by IMPα/β1 of HIV-1 IN but not other IMPα/β1 -recognised cargoes (Wagstaff *et al.*, 2011). Even though was approved for human use (in combination with prostaglandin analog) for medical abortion and emergency contraception, or to treat hyperglycemia in Cushing's disease, uterine leiomyomata,

endometriosis and unresectable or malignant meningioma (Chen and Creinin, 2015) it did not pass Phase I/II trials for HIV-1 treatment (Para *et al.*, 2010). Other examples of inhibitors are Budesonide and analogues, Flunisolide and 4-HPR Fenretinide. Budenoside and its analogues target HIV IN- IMPα/β1 interaction and inhibits nuclear accumulation of IN but no other cargoes in an in vitro nuclear import assay (Wagstaff *et al.*, 2019) and 4-HPR targets the interaction between Flavivirus NS5 and IMPα/β1, shown to inhibit infection with DENV including severe disease form, as well as ZIKV and WNV in cell culture (Pitts *et al.*, 2017; Wang *et al.*, 2017).

B.2.7 Previous Results

Nuclear transport is crucial for viral life cycle and the development of NLS to modulate their viral replication, infection and host cell immune response highlight the fact that the interaction of viral proteins with the host nuclear transport apparatus can be a potential therapeutic target. To this end we retrieved the proteome of all human viruses from Viralzone database, and we performed bioinformatic analyses quoting from Uniprot all the functional annotations of these proteins, we combined three algorithms (Psort II, cNLS mapper and Deep Loc) to predict the localization of all these proteins, thus predicting if they are endowed with NLSs. These analyses allowed us to to classify all viral proteins into three main categories: Confirmed Nuclear, Putative Nuclear and Cytosolic proteins. This study then allowed us to perform several comparations regarding the different requirements for nuclear import of different viral proteins. For example, we compared what percentage of proteins was classified either confirmed nuclear or putative nuclear in all human infecting viruses and compare such percentage in cytoplasmic and nuclear replicating viruses. Our data showed that 15% (532 proteins)

of all viral proteins were confirmed nuclear, of which 80% belonged to nuclear replicating viruses (corresponding to 40% of their proteome) and the remaining 20% belonged to cytoplasmic replicating ones (corresponding to 3% of their proteome). Interestingly, by considering all confirmed and putative viral nuclear proteins, we saw that the percentage of nuclear proteins increased up to 30%. Also, the percentage of nuclear proteins in nuclear and cytoplasmic replicating viruses increased up to 50 and 20%, respectively, suggesting that several nuclear were wrongly annotated on Uniprot (Fig. B.2.5). By investigating the mechanism by which these nuclear proteins translocate in the host cell nucleus, our data revealed that 65% of all confirmed viral nuclear proteins were endowed with cNLSs, this percentage was also observed for nuclear proteins in nuclear replicating viruses, while the percentage of nuclear proteins bearing cNLS reached 70% in cytosolic replicating ones (Fig. B.2.6).

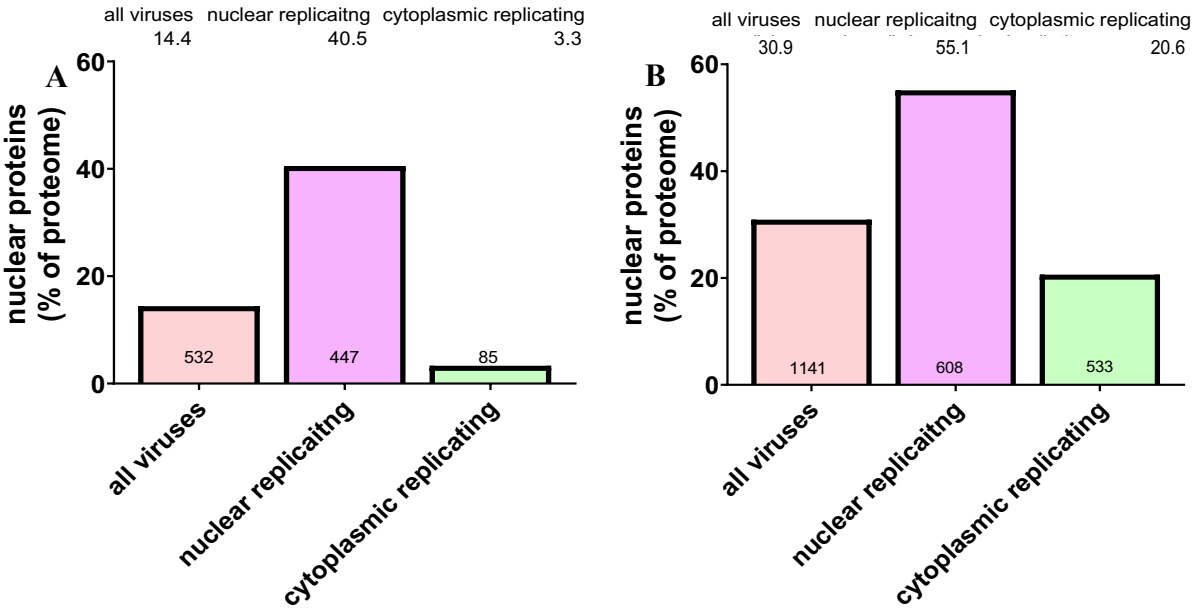


Figure B.2.5. Percentage of Nuclear Viral Proteins increases if all hypothetical and confirmed nuclear proteins are considered. Bioinformatic analyses highlighted that 14.4% of all viral proteins localizes in the host cell nucleus, of which 40.5% belong to nuclear replicating virus and 3.3% to cytoplasmic replicating ones (A). Interestingly, the percentages of viral nuclear proteins increases if hypothetical nuclear viral proteins are included, reaching 55.1% nuclear proteins in nuclear replicating viruses and up to 20.6% for cytoplasmic replicating viruses (B).

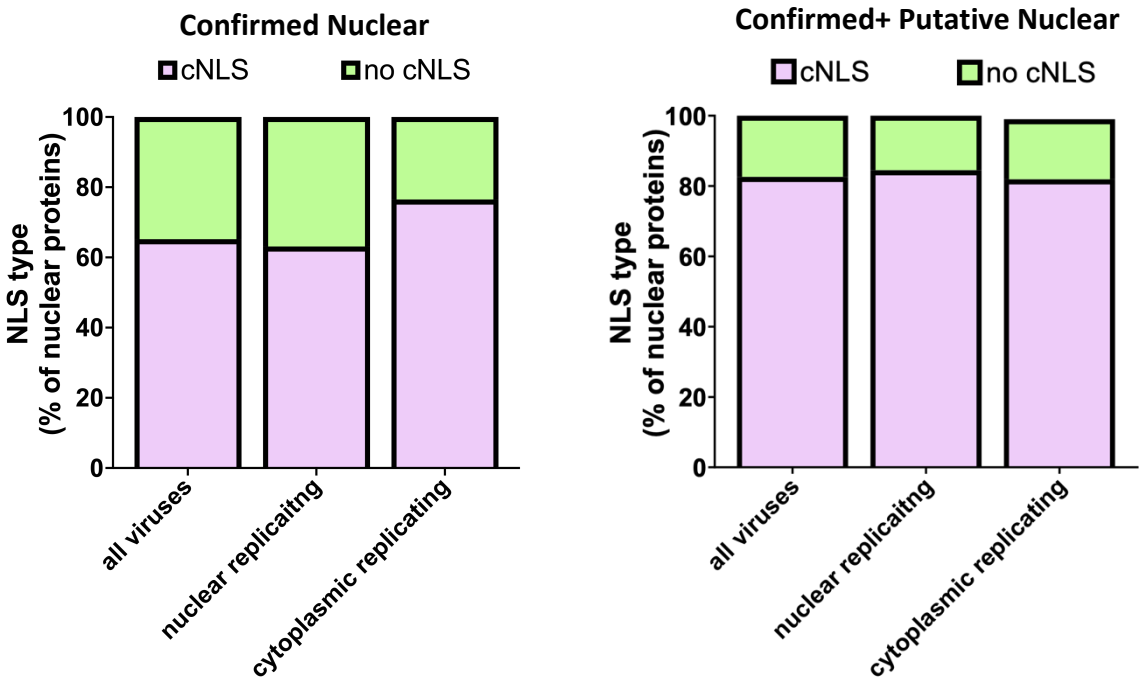


Figure B.2.6. Viral nuclear proteins bearing NLSs. Bioinformatic analyses shown that more than 80% of all viral nuclear protein are endowed with classical NLS (cNLS).

Moreover, these percentages all raised to more than 80% for all confirmed and putative nuclear proteins in all the classes considered, underlining the fact that viruses heavily rely on IMPα/β1 pathway (Fig B.2.6) to target their product into the host cell nucleus, thus suggesting that these protein-protein interactions can be target for therapeutic interventions. Indeed, these data suggest that several viral proteins are wrongly annotated on Uniprot.

Such database can also be used to study how different viral families interact with the host cell nuclear transport apparatus. If we consider the Polyomaviridae family, for example, such high percentage of nuclear proteins is not surprising. Indeed, Polyomaviruses are small non-enveloped viruses replicating in the host cell nucleus and therefore, most of their proteins can be expected to have a nuclear localization during

viral life cycle. In fact, our analysis revealed that 68% ± 21.7% of all *Polyomaviridae* encoded proteins were classified as nuclear. Importantly, if we considered even putative nuclear proteins, the percentage leaded to a potentially 91 ± 12.5% proteins proposed to be able to localize to the host cell nucleus (Figure B.2.7). Intriguingly, if we consider the results obtained for each of the five individual *Polyomaviridae* members considered in the bioinformatic analysis, it is clear that the percentage of confirmed nuclear proteins are highly variable among them. For example, the percentage of nuclear proteins over the whole viral proteome ranged from 100% in KIPyV (4) to 50% for MCPyV (2/4). This most likely reflects the different accuracy of Uniprot annotations. Consistent with this, if all hypothetical and confirmed nuclear proteins are considered the differences in percentage of nuclear proteins decreases, ranging from 100% for KI, WU and JCPyVs and 80% for MCPyV (Fig.B.2.7)

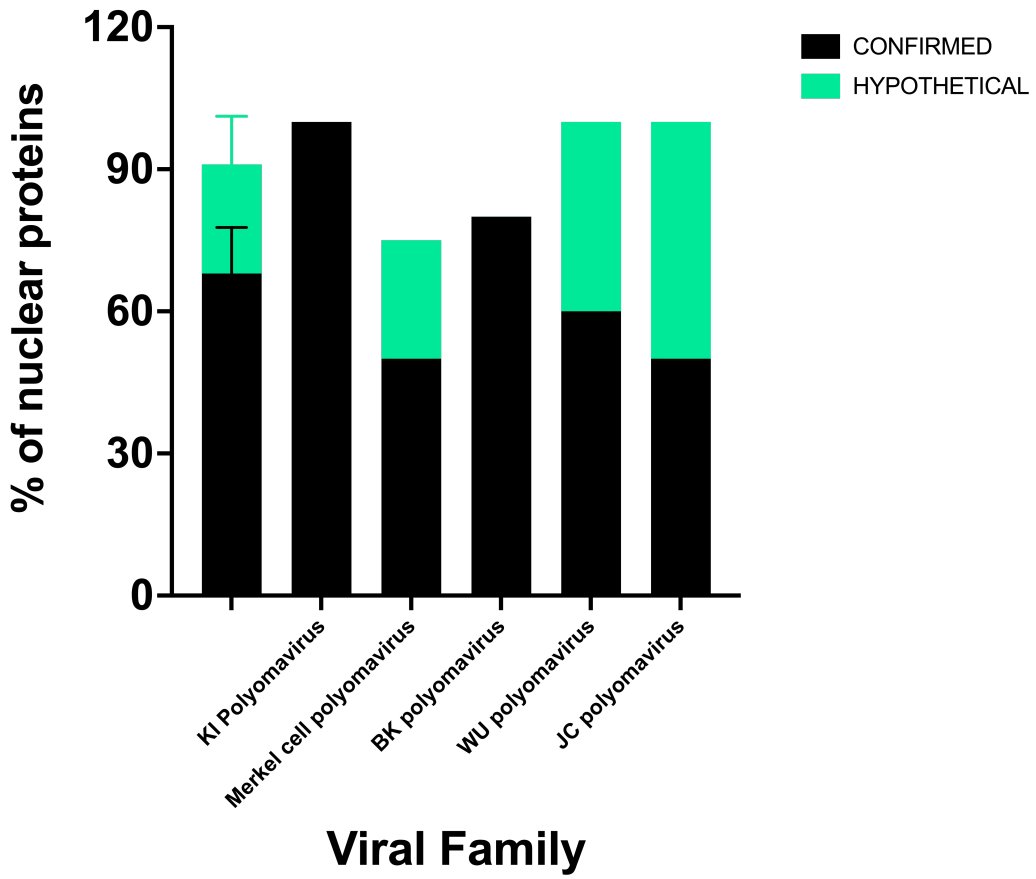


Figure B.2.7. Percentage of nuclear proteins in Polyomaviridae family. A preliminary Bioinformatic analysis on 5 members of the Polyomaviridae family showed that 68 ± 21.68 % of the overall viral proteins, localized in the nucleus (confirmed, black bars) and the percentage of nuclear proteins increases up to 91 ± 12,5% when hypothetical nuclear proteins are included (green bars).

Our bioinformatic analyses allowed us to classify the proteins encoded by these five members of Polyomaviruses: confirmed nuclear proteins with or without cNLS, hypothetical nuclear proteins bearing a cNLS and cytosolic proteins. Some proteins were not found in some family members. The comparative analyses revealed for example, that Large T antigen (LTAs) protein were confirmed nuclear bearing cNLSs only in two of the families, instead in the remaining three families the LTAs were considered hypothetical nuclear proteins bearing cNLSs. Our pipeline identified putative cNLSs for all of them, so we decided to extend our analysis to all the HPyVs LTAs and a large number of cNLSs could be mapped in their sequences (Fig.B.2.9).

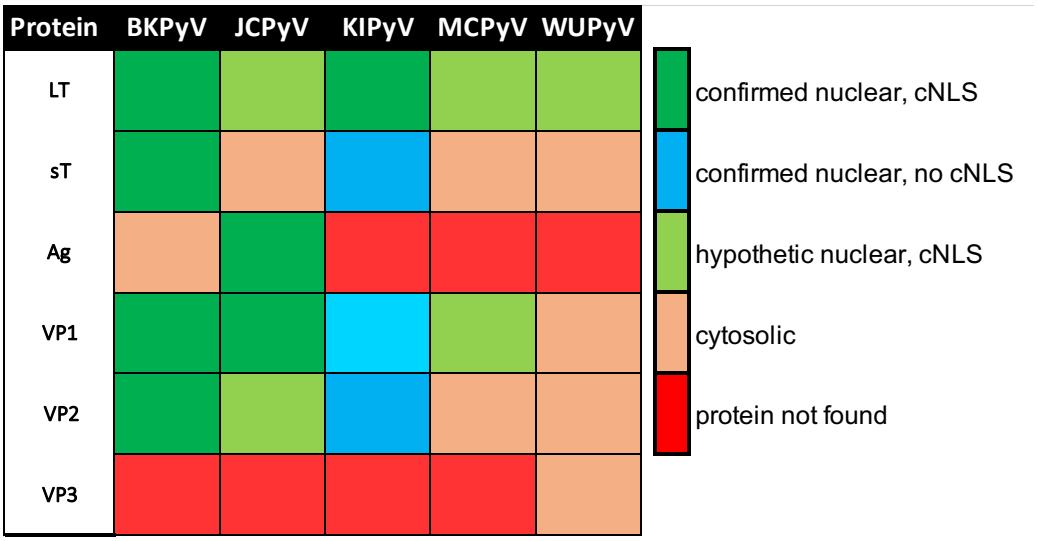


Figure B.2.8. Conserved orthologs proteins in Polyomaviruses are differently classified on Uniprot. A first glimpse of some conserved proteins among five members of Polyomaviruses: LT, sT, Ag, VP1, VP2 and VP3.

Indeed, LTAs are large multifunctional proteins involved both in viral gene expression and replication - by binding the viral promoters and origin of replication, as well as for modulation of host cell function, by interacting with a plethora of cellular targets, including the oncosoppressor proteins p53 and pRb. According to the need to localize to the cell nucleus to mediate viral transcription and replication, the very first cNLS was described on the LTA from SV40. Despite the SV40 LTA (PKKKRKV-132) has since then become the best characterized NLS to date, surprisingly, little is known regarding the nuclear transport process of LTAs from other Polyomaviridae members. Early studies identified two NLSs in the mouse polyomavirus, while SV40 LTA cNLS is completely conserved in JC and BK PyVs. Besides that, a similar cNLS has been described for the MCPyV LTA (RKRK-280), which has been proposed to be required and sufficient for nuclear targeting (Nakamura and Katano, 2009). Intriguingly, nuclear targeting MCPyV LT is not absolutely required for tumor formation, since from at least two patients with MCC, sequencing of the viral genomes revealed integration upstream of the LT-NLS

coding region (Kalderon and Smith, 1984).Therefore, the cNLSs responsible for nuclear import of LTAs from most HPyVs and the pathways involved are currently unknown. We set out to shed light on this issue. To this end, we downloaded the coding sequence of LTAs from the reference proteome of all the 14 known HPyVs and combined phylogenetic and bioinformatics analyses to identify putative cNLSs on their sequence. We could identify at least one putative cNLS on the sequence of each HPyV LTA, but strikingly almost 50% of the LTA HPyVs possessed more than one (Fig. B.2.9). A putative cdc2 T phosphorylation site immediately upstream of the P2-P5 basic consensus residues, shown to inhibit IMPα/β binding and nuclear import on SV40 LTA, was also conserved in all of identified cNLS. Similarly, in most cases CK2 phosphorylation enhancing IMPα/β binding and nuclear targeting was also conserved.

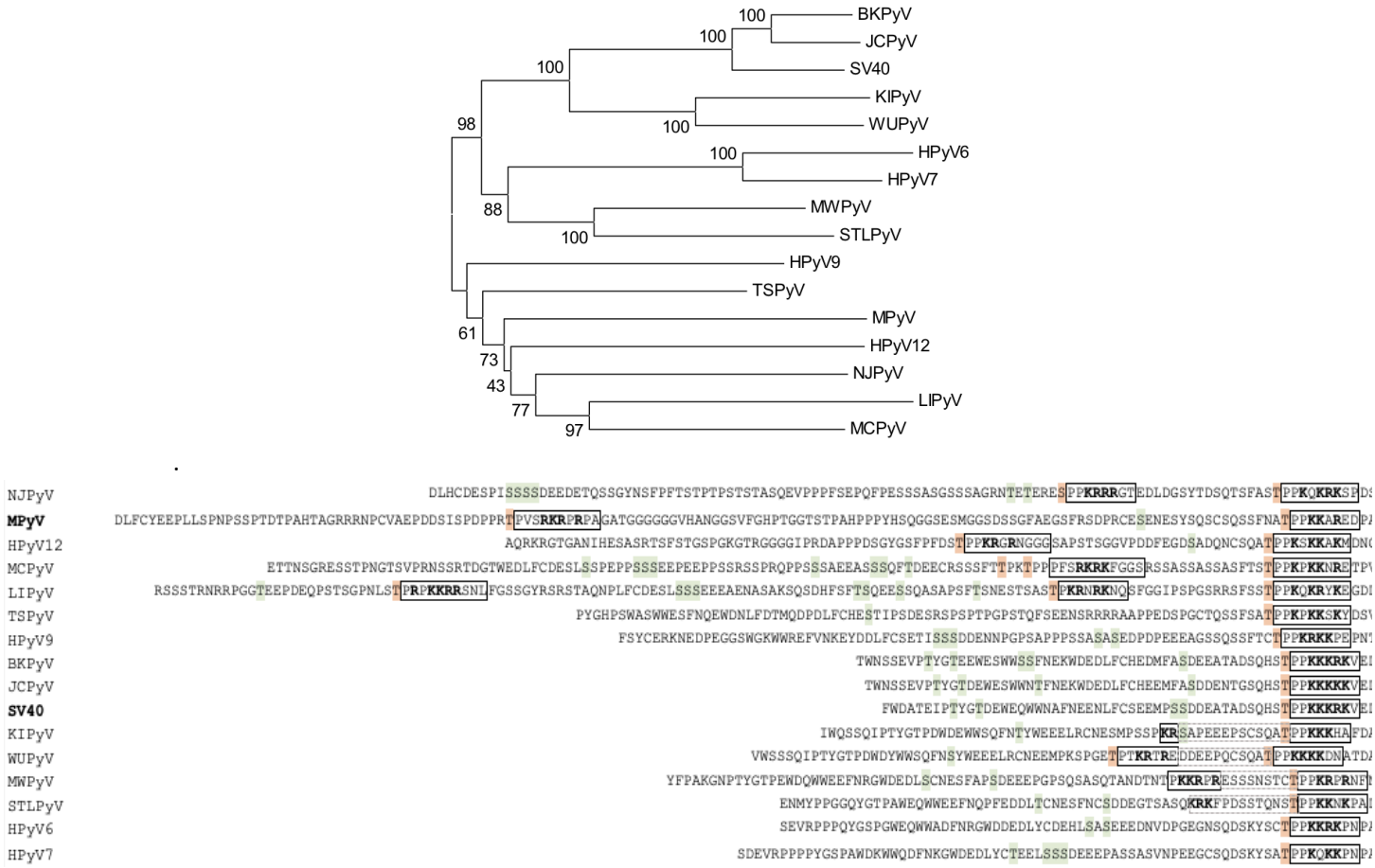


Figure B.2.9. Combined phylogenetic and bioinformatics analyses identified putative cNLSs on Polyomaviridae LTAs. Several putative cNLSs were identified in LTAs sequences. Nearly 50% of LTA proteins were endowed with more than cNLSs.

Alongside the different comparative bioinformatic analyses performed in our study, we eventually investigated the correlation between the size of viral proteome and the percentage of nuclear proteins. This correlation, considering for example the first Baltimore class of viruses, highlighted that between the proteome size and percentage of nuclear proteins, there is an inverse proportionality, where bigger the virus, less the percentage of nuclear proteins, just as expected. Big complex viruses with large genome, are capable to encode for viral protein necessary for their life cycle resulting in a less relying on host cell proteins. Instead, small and simple viruses highly need host cell proteins to ensure their life cycle. However, it can be noticed that the percentage of

nuclear proteins increases if all confirmed and putative nuclear proteins are considered (Fig.B.2.10). In particular, large cytoplasmic replicating Poxviridae family exhibit an increase from nearly no nuclear protein to a 40% of nuclear proteins if all confirmed and hypothetical nuclear proteins are considered That is comparable with the fact that Poxviruses co-evolve with their host where a very large portion of their genome interfere with antiviral defences, intrinsic, innate and adaptative cell defences.

Overall, in our studies we identified nearly 200/300 new viral nuclear proteins from different viruses and 11 out of 26 proteins with the highest cNLS mapper score, belonged to the Poxviridae family, and we decided to functionally characterize them all.

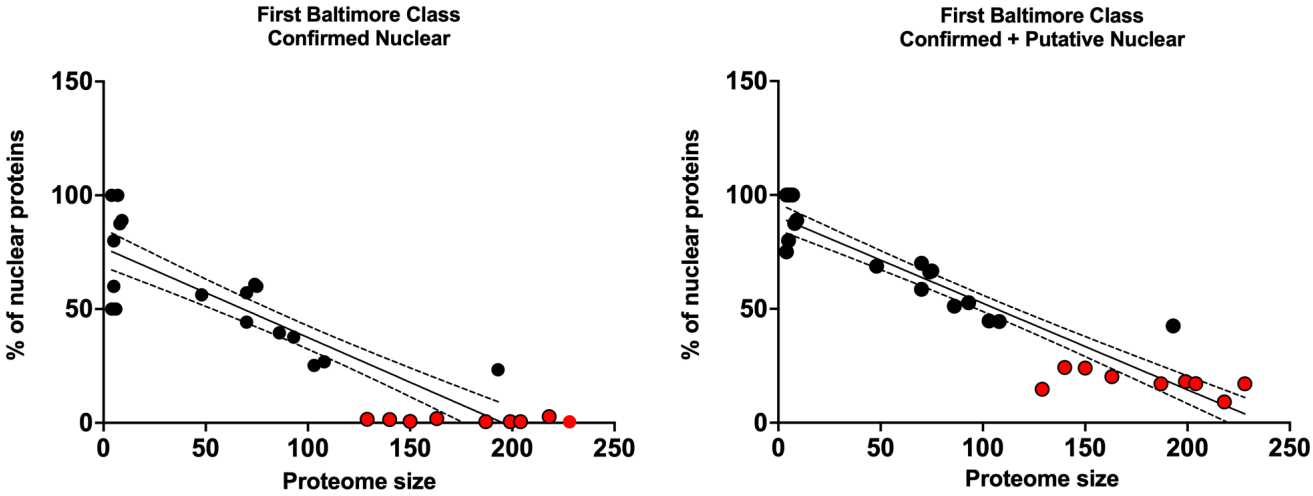


Figure B.2.10 Correlation between proteome size and percentage of nuclear proteins in first Baltimore class of viruses. The percentage of nuclear proteins is inversely proportional to viral proteome size. Poxviruses confirmed nuclear proteins are almost zero, but the percentage increases if all hypothetical and confirmed nuclear proteins are considered.

B.3 Material and Methods

B.3.1 Primer resuspension

Lyophilizate primers were centrifuged for 10 min at 4°C and 13,000 rpm, and then resuspended in an appropriate volume with Tris-HCl pH7,8 10mM, under aPCR hood flow, to obtain final concentration of 100 µM, vortexed and briefly centrifuged at max speed. Resuspended primers were further diluted (1:100) in the same buffer to a concentration of either 1 µM [used for the duplex formation in TOPO® reactions (§3.4)] or 5 µM [for PCR amplification for Gateway® reactions (§3.3)]. All primers were labeled and subsequently stored at - 20°C until needed.

B.3.2 Plasmid elution from 3M paper

The portion of the paper where the plasmid was spotted, was cut, transferred into a 1.5 ml Eppendorf tube and incubated for 30 min at room temperature (RT) with 100 ul of TE buffer (10 mM Tris-HCl pH8.0 and 0.1 mM EDTA). The Eppendorf was then vortexed for 20 sec and centrifuged for 5 min at 13,000 rpm. The supernatant was transferred into a new 1.5 ml Eppendorf tube and stored at -20°C. (Table B.3.1)

B.3.3 DNA cloning with Gateway® Technology

B.3.3.1 Gateway template productions

Some cNLS sequences flanking *attB* sites were generated from already available plasmids expressing the sequence of interest lacking the *attB* Gateway recombinant sites, by means of PCR, using FWD and REV primers embedded with *attB* sequence (*attB1* 3'-CAAAAAA- and *attB2* -CTTTCTT-5, respectively). The PCR reaction mix included

1x Gold Buffer (Applied Biosystem™ #4311806), 10mM dNTPs mix, 25 mM $MgCl_2$, 0.25 µM FWD primer, 0.25 µM REV primer, and U/1000 AmpliTaq Gold™ (Applied Biosystem™ #4311806). 100 ng of the template is then added to the reaction mix and the PCR setup was as follows:

◊ 1x cycle: 95°C for 5 min

◊ 35x cycles: 95°C for 30 sec – 56°C for 30 sec – 72°C for 1 min

◊ 1x cycle: 72°C for 1 min

◊ 1x cycle 4°C ∞

Then the PCR product was separated from nonspecific amplifications on a 2% agarose gel, cut at the corresponding kDa band and purified from the exceeding gel with GenElute™ Extraction Kit (SIGMA # NA1111-1KT). The purified DNA was quantified with Nanodrop.

B.3.3.2 BP recombination reactions

The entry vector was generated with 1µl pDONR-207 (75 ng/µl), 1µl *attB*-plasmid (75 ng/µl) (Fig. B.3.4) and TE buffer up to 4µl being mixed at Room Temperature (RT) in a 1,5 ml microcentrifuge Eppendorf tube. The BP clonase ™ enzyme mix (Thermofisher Scientific, #11789020) was briefly vortexed and 1 µl was added to reaction tube. The reaction was briefly vortexed again and incubated overnight (ON) at 25°C. 0.5 µl of Proteinase K (2 µg/ µl) solution was added to the reaction and incubated 10 min at 37°C. Chemically competent E. Coli DH5α were then transformed (§ 3.5.1) and the plasmid DNA was then purified (§3.8), the sequence further verified by Sanger Sequencing

(§3.10) and for each plasmid a bacterial glycerol stock (§3.7) was stored at -20°C. (Table B.3.2)

B.3.3.3 LR recombination reaction

LR reactions generate the expression vector. 1µl pDEST vector of choice (75 ng/µl) (Fig. B.3.5), 1µl Entry clone (75 ng/µl) and TE buffer up to 4µl were mixed at RT in a 1.5 ml microcentrifuge Eppendorf tube. 1 µl of LR Clonase™ enzyme mix (Thermofisher Scientific, #11701020) was added to the reaction after being briefly vortexed. The reaction was incubated ON at 25°C. 0.5 µl of Proteinase K (2 µg/ µl) was added to the reaction and incubated 10 min at 37°C. Chemically competent E. Coli DH5α were then transformed (§ 3.5.1) and the plasmid DNA was then purified (§3.8), the sequence further verified by Restriction analyses and for each plasmid a bacterial glycerol stock (§3.7) was prepared and stored at -20°C. (Table.B.3.3)

B.3.4 TOPO cloning

B.3.4.1 Duplex formation

For the duplex formation reaction, 1 µl of Forward (FWD) primer 1 µM, 1 µl of Reverse (REV) primer 1 µM (§3.1) were added to 122 µl of TE buffer in a sterile environment and mixed for a final primer concentration of 8.2 nM. The duplex subsequently incubated at 95°C for 5 min and then gradually cooled to 25°C for the primers to anneal correctly. Primer duplexes were used for TOPO® reaction, and subsequently stored at -20°C.

B.3.4.2 TOPO® reaction

TOPO® reaction allowed the integration of DNA sequences within the vector pcDNA3.1-NT-GFP-TOPO (Thermofisher Scientific, #K4810-01) (Fig.B.3.6). For each TOPO reaction, in one 1.5 ml Eppendorf microcentrifuge tube 1 µl Duplex (§3.4.1), 1 µl Salt Solution diluted 1:2 (v/v) in TE buffer and 1 µl TOPO vector diluted 1:3 (v/v) in TE buffer were mixed and incubated for 30 min at RT.

Then 1.6 µl of the reaction was used to transform 16 µl of One Shot® TOP10 Chemically Competent E. coli as described in section 3.5.2. For each TOPO® reaction, 5 colonies were used to set up liquid cultures (§3.6), recombinant plasmid DNA was then purified by affinity chromatography (§3.8), the sequence verified by Sanger Sequencing (§3.10) and for each plasmid a bacterial glycerol stock was prepared (§3.7). The salt solution and TOPO vectors come with the commercial kit. The Salt Solution (200mM NaCl, 10 mM $MgCl_2$).

B.3.5 Heat Shock transformation of competent E. coli

B.3.5.1 Inoue protocol Heat Shock transformation of competent E. coli

50 µl of chemically competent DH5α *E. coli* (Life Technologies, #C404010) were incubated with 1ng/µl of the plasmid for 30 min on ice. Then for 45 sec at 45°C and again in ice for other 2 min. 600 µl of sterile LB were then added to the reaction and incubated at 37°C for 1 h in shaking before being plated on a LB agar with the right antibiotic: Ampicillin (SIGMA ALDRICH, #A9518-25G), Kanamycin (Applichem, #A1493,0005) or

Gentamicin (SIGMA ALDRICH, #G1272-10ML). Plates were incubated at 37°C and 16 h later single bacteria colonies were sought.

B.3.5.2 Transformation of One Shot® TOP10 Chemically Competent E. coli

The commercial kit provides 50 µl bacteria aliquots (Thermofisher Scientific #C404003) which were thawed in ice and 16 µl were transferred in a new 1.5 ml Eppendorf tube in a sterile environment. 1.6 µl of the TOPO reaction product (§3.4.2.) was added to the TOP10 Chemically Competent *E. coli*. After and incubation of 15 min in ice, followed by 30 sec at 42°C and 2 min in ice, 250 µl of SOC medium was added to the TOPO reaction which was then incubated for 1h at 37°C shaking horizontally (180-200 rpm) before being completely plated on LB agar plates with the appropriate antibiotic, in our case Ampicillin (SIGMA ALDRICH, #A9518-25G).

B.3.6 Bacteria inoculum preparation

Single bacteria colonies from *E. Coli* transformation were inoculated in 10 ml of LB with the supplement of and appropriate antibiotic and incubated at 37°C ON horizontally shaking at 130 rpm.

B.3.7 Miniprep Plasmid DNA purification

Purification of plasmid DNA from *E. coli* cultures was performed with GenElute™ Plasmid Miniprep (SIGMA, #PLN70). 10 ml of bacterial culture holding the appropriate plasmid was grown at 37°C under shaking at 150 rpm ON. Subsequently, 1 ml was stored at 4°C for Glycerol stock preparation (§3.11), and the remaining was centrifuged for 10

min at 4000 rpm at 4°C. The supernatant was discarded, and the pellet resuspended in 200 µl of Resuspension Solution containing RNase A (100 µg/ml). 200 µl of Lysis Solution and 350 µl of Neutralizing/Binding Solution were then added to each sample. The tubes' content was mixed by inverting the tube at each step of the protocol. Debris were pelleted with a 13,000 rpm speed centrifugation for 10 min at 4°C and the debris-cleaned lysate supernatant was poured in the Binding Column containing the silica membrane which had been previously activated with 500 µl Columns Preparation Solution, and centrifuged at 13,000 rpm for 1 min at 4°C. The columns were centrifuged at 13,000 rpm for 1 min at 4°C, the flow-through was discarded and the column was washed 2 times with 500 µl of washing solutions with different saline concentration. A void centrifuge of 2 min at max speed at 4°C was done to ensure that undesired material was left within the microcentrifuge tube and the silica membrane. The elution of the plasmid DNA was accomplished by adding 55 µl of the Elution Buffer and centrifuging the tube at max speed for 1 min at 4°C. The highly purified DNA was quantified (§3.9), and the sequence confirmed by Sanger sequencing (§3.10), then stored at -20°C.

B.3.8 DNA quantification

The DNA plasmid concentration obtained with the GenElute™ Plasmid Miniprep was determined spectrophotometrically using NanoDrop® ND-1000 Spectrophotometer (Nanodrop Technologies) using 1µl from the sample. The plasmid DNA purity was obtained by determining A260/280 and A260/230 ratio which must have an optical density up to 1.8 and a range of 2.0-2.2, respectively. Quantification was performed using as blank kit's Elution Buffer in which the purified DNA was eluted.

B.3.9 Enzymatic digestion

In some cases, the integrity of recombinant was verified by enzymatic digestion. Based on the designed theoretical plasmid map with ApE software, specific restriction enzymes were selected for the enzymatic digestion. Within a 1.5ml microcentrifuge tube 0.5 µg x µl DNA, 2µl of 10x restriction enzyme buffer, 2µl of BSA 10x, 0.5 µl of Restriction enzyme (Promega; 10U/ µl) and MilliQ water up to 20 µl (final reaction volume) were mixed and incubated for 2h at 37°C. Subsequently, 4 µl of Gel Loading Dye 10x were added to the reaction mix and electrophoretically separated on agarose gel (0.8-2%, according to fragment size). If the restriction pattern was correct, DNA plasmids were either considered ok (for LR reactions or purchased ones) or confirmed by sequencing (in case of BP or TOPO reactions) (§3.10).

B.3.10 Sequencing

The nucleotide sequence of the clones generated in this study was determined with Sanger sequencing using two main approaches and services. Some DNA sequencing were done using ABI Big Dye 3.1Cycle Sequencing Terminator Reactions (Thermofisher Scientific, #4337455) in a volume of 20 µl. 300 ng/µl x µl of DNA template, 2 µl BigDye Terminator, 4 µl BigDye Sequencing Buffer 5X, 3,2 µl of Primer BGH Reverse (5'-TAGAAGGCACAGTCGAGG-3') 1 µM and MilliQ water up to 20 µl were mixed. The PCR reaction setup was as follows:

◊ 1x cycle: 95°C for 5 min

◊ 30x cycles: 96°C for 10 sec – 50°C for 5 sec – 60°C for 4 min

◊ 1x cycle 4°C ∞

The sequencing products were removed from the thermocycler and transferred in labelled 1.5 ml microcentrifuge tubes where they were purified by adding 2 µl of NaAc 3M pH5.2 and 50 µl of EtOH 96%. The reaction was vortexed, spinned, and incubated at -20°C for at least 20 min. Samples were subsequently centrifuged at 13,000 rpm for 20 min at 4°C, the supernatant carefully removed, and the pellet washed with 150 µl of EtOH 70% and centrifuged at 13,000 rpm for 5 min at 4°C. The pellet was dried for 5-10 min at 40°C. Finally, samples were resuspended in 15 µl of MilliQ water and taken to the Sequencing Service of the Azienda Ospedaliera of Padova.

Other sequences were screened thanks to BMR Genomics S.r.l sequencing service. In order to be sequenced, 700ng of the sample was loaded in a 0.5-microcentrifuge tube, spinned and the pellet dried at 65°C with open lid. The 0.5 microcentrifuge tubes were then labeled according to the code given by BMR Genomics software. Since their service includes also the use of their own primers, we choose the BGH Reverse primer available on their website. The tubes were sealed and taken directly to the company.

Sequences were downloaded from the respective sequencing service servers, analyzed with ApE software and were compared to the theoretic reference sequence. If the theoretic and the effective sequence matched, samples were assigned an appropriate label code and stored at -20°C for research application. Bacteria from which verified DNA sequences were extracted were further propagated to generate glycerol stocks (see §B.3.11)

B.3.11 Glycerol stock preparation

1 ml of bacteria culture was taken and temporary stored at 4°C until the GOI sequence was confirmed to be correct by either sequencing (§3.10) or restriction analysis (§3.19). Subsequently, the sample was centrifuged for 10 min at 2,000 rpm, the supernatant removed, and the pellet gently resuspended in 500 µl of Glycerol-LB (30%v/v). Samples were labelled and stored at -80°C.

B.3.12 Cell lines, media and maintenance

HEK 293A (ATCC® CRL-1573™) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) which was supplemented (DMEM cpt) with 10% Fetal Bovine Serum (FBS), non-essential aminoacids, L-glutamine (2mM), Penicillin and Streptomycin (100U/ml) al 37°C and 5% CO_2 in a humidified incubator. Cells were passaged when reached >90% confluence. To this end, cells were briefly washed with Dulbecco's phosphate-buffered saline (D-PBS). HEK 293A, cells incubated with 0.05% trypsin-EDTA until they started detaching from the flask. Cells where then resuspended in cpt DMEM and an appropriate volume of cells was transferred to a new flask containing required amount of cpt DMEM. All culture reagents were purchased from Gibco, Thermofisher Scientific. PBS, Trypsin and DMEM volumes varied according to the format of the flasks and plate used. Stored at +4°C.

B.3.12.1 Mycoplasma Test

Cell were routinely tested for mycoplasma contamination using the N-GARDE Mycoplasma PCR Reagent Set (Euroclone #EMK090020). Briefly, 200 µl/ 1ml of the cell culture media was centrifuged at 700 rpm for 5 min to discard dead cell debris. Supernatants was placed in a new sterile Eppendorf and centrifuged again for 10 min at max speed. Supernatants were carefully discarded, and pellets resuspended in 50 µl of Buffer solution, boiled at 95°C for 3 min and 5 µl were used as template for the PCR reaction. PCR reaction mix was comprehensive of 10 µl of kit's Mix Solution and the final reaction volume of 50 µl was reached with sterile MilliQ water. A control sample and a blank one was used to compare the template PCR product with. The PCR set up was as follows:

- ◊ 1x cycle: 94°C for 30 sec
- ◊ 35x cycles: 94°C for 30 sec – 60°C for 120 sec – 72°C for 60 sec
- ◊ 1x cycle: 94°C for 30 sec
- ◊ 1x cycle: 60°C for 120 sec
- ◊ 1x cycle: 72°C for 5 min
- ◊ 1x cycle: 4°C ∞

The PCR products were loaded on a 2% agarose gel containing Eurosafe 1:25,000 (Euroclone #EMR440001) and DNA fragments were observed with Uvitec Allinace Image Software (UVITEC Cambridge Alliance).

B.3.13 Cell freezing

For long term storage, cells were stored in liquid nitrogen. For each 25cm^2 of cultured cells corresponded one cryovial (Corning® Cryogenic Vials with Orange Cap, #100-0091). Cells were washed in 1xPBS and Trypsin and resuspended in DMEM cpt. Resuspended cells were then centrifuged for 5 min at 700 rpm, the supernatant discarded, and the pellet cells were finally resuspended on ice in freezing medium (DMSO 10% v/v FBS). Each 25cm^2 of cultured cells were resuspended in 1.8 ml of freezing medium and aliquoted in pre-chilled cryovials, and incubated 10 min on ice, 2 h at -20°C, and ON at -80°C before being transferred to liquid nitrogen

B.3.14 Cell Thawing

For each thawed cryovial one 15ml Falcon® tube was prepared with 6ml of DMEM cpt. Cryovials were rapidly transferred from liquid nitrogen under the hood and thawed by resuspending the cells in warm DMEM cpt medium already present in the 15ml Falcon tube. Cells were centrifuged at 700 rpm for 5 min and the supernatant discarded. Cells were finally resuspended in 6ml DMEM cpt medium, seeded in the appropriate T25 flask and incubated at 37°C and 5% CO_2 in a humidified incubator.

B.3.15 Treatment of Coverslips with poly-Lysine

Prior cell seeding, in a 24 well plate sterile glass coverslips (12 mm round, 1001/12 BIOSCIENTIFICA) were treated with poly-Lysine to promote cell adhesion. Poly-Lysine (poly-L) solution was prepared by resuspending 5mg of poly-lysine (SIGMA 6282) in 50 ml of sterile MilliQ water in a complete sterile environment and stored ad +4°C. After placing the sterile coverslips in the 24 well plate, 500 µl of poly-L were added to each

well and the coverslips were pressed to the bottom of the well with the tip of a p1000 tip, and incubated at 37°C and 5% CO_2 in a humidified incubator for at least 30 min. Under the hood, the poly-L was carefully removed, recycled for a maximum of three times and the coverslips were washed with sterile MilliQ water and with the help of a needle and a plier, were placed in an oblique way to optimize the drying and avoid their sticking to the plate.

B.3.16 Cellular Transfection with Lipofectamine 2000

Cell were transfected using Lipofectamine 2000 (Thermofisher Scientific, #11668030). To this end, $5x10^4$ cell/well HEK 293A cells were seeded into a 24well plate on poly-L pre-treated coverslips (§3.15) and incubated at 37°C ON. In 1.5 ml Eppendorf microcentrifuge tube the amount of DNA needed to be transfected (1 µl for 250-500 ng/ µl) was added to 50 µl of DMEM medium lacking antibiotics. For each transfection well 1 µl Lipofectamine 2000 was diluted in 50 µl of OPTIMEM (Opti-MEM Reduced Serum Media # 11058021Gibco, Thermofischer) and the mixture was added to DNA-DMEM solution and incubated for 20 min. The reaction mix was subsequently added dropwise to the cells and the plate was further incubated at 37°C and 5% CO_2 in a humidified incubator. 24h post transfection the medium was replaced with DMEM cpt and the transfection efficiency was monitored with inverted fluorescence microscope Leica (DFC420C).

B.3.16.1 Energy Depletion Assays

To understand the energy dependency of viral proteins during nuclear translocation, we generated a cell culture media without the glucose support. No phenol-red glucose-free

DMEM (Gibco ™ ThermoFischer Scientific, #A1443001) was enriched with 10% Fetal Bovine Serum (FBS), L-glutamine (2mM), Penicillin and Streptomycin (100U/ml) and HEPES 1M. Sodium Azide 3M and 2-deoxy-D-Glucose were added fresh to the media each time. Stored at +4°C. Transfection procedure was as described in §3.16, where samples were transfected in duplicate. After 48 h post transfection and prior to sample staining and fixation on coverslips, one well x sample was treated with ATP depletion media. To this end, media of such wells were replaced with energy depletion media and incubated or 1 h at 37°C and 5% CO_2 in a humidified incubator. The effect on transfected cells were monitored via inversed fluorescent microscope Leica and cells were washed and processed for Immune Fluorescence as described in §B.3.17.

B.3.17 Sample staining and processing for Immune Fluorescence (IF)

At 48h post transfection cells were incubated for 30 min with DRAQ5 (1:10,000 in DMEM no phenol red, see APPENDIX) washed with 500 µl PHEM 1x (§3.19). Cells were fixed with 350 µl in Paraphormaledehyde 3% in PHEM (§APPENDIX) for 15 min at RT, before being washed 2 times with PHEM 1x. After a brief rinse in MilliQ water and being dried on a tissue paper, coverslips were mounted on objector holder slide using 7 µl of Fluoromount-G® (FG) mounting medium (eBioscience #00-4958-02).

B.3.18 Confocal Laser Scanning Microscopy and image analysis

Subcellular localization of fusion proteins was analyzed using a Leica Nikon A1 confocal laser scanning microscope (Leica) with 60x magnification, an oil immersion objective. Images were acquired using all the three lasers (488, 561 and 640 nm) available and transmitted light. Images of 4 fields of each coverslip were acquired, and for each, 4

additional areas were further analyzed with a magnification of 2.5x and 1024x1024 px resolution. Once the images were acquired, the Fn/c values were determined using the NIH ImageJ 1.62 public domain software from single cell measurements for each of the nuclear (Fn), nucleolar (Fnuc) and cytoplasmic (Fc) fluorescence. To normalize the signal each fluorescence value was subtracted the autofluorescence/background fluorescence (Fb). Data were collected and statistically analyzed using Graphpad Prism 9 (GraphPad Software Inc., San Diego, CA, USA) software. The ratio between F(n-b)/F(c-b) quantified the nuclear accumulation and so the activity of the studied viral protein NLSs. Data significancy was obtained via Student's T-test and Anova test.

B.4 Results

B.4.1. Polyomaviruses

As mentioned in section §B.2.7, our analysis revealed that several viral proteins are likely mis-annotated when it comes to subcellular localization and offered the unparallel opportunity to experimentally address this issue (Fig.B.2.8). One notable is Large T antigen (LTA) from Polyomavirus, which was annotated as nuclear on Uniprot, and therefore termed as confirmed nuclear protein only in two of the 5 family members considered in our study. Therefore, we extended our analysis to all Human Polyomaviruses (HPyV) and a large number of cNLSs could be mapped in their sequences (Fig.B.2.9). In order to discriminate functional from non-functional newly identified cNLSs we first cloned them to generate recombinant plasmid expressing our putative cNLS fused to a reporter fluorescent protein. Based on the length of cNLS sequence we adopted different cloning techniques. Sequences shorter than 30 amino acids were cloned via TOPO® reaction (§3.4), where the cNLS sequences were cloned in frame with Cycle3-GFP within the vector pcDNA3 NT-GFP-TOPO (Thermofisher Scientific, #K4810-01) (Fig. B.3.6), while larger sequences were cloned with Gateway® Technology (Life Technologies) (§3.3) which generated pDEST-YFP vectors with protein fused either at C-terminal of YFP (fig B.3.4). These newly generated plasmids (Table B.4.1) were then used to transfect HEK 293A cells which 48 hours post transfection were treated with DRAQ5, fixed, mounted on coverslips in order to observe and quantify the localization of the viral proteins by Confocal Laser Scanning Microscopy (CLSM) (§3.18). The localization of Cycle3-GFP protein was used as control which is shown in Figure B.4.2 to localize ubiquitously in transfected HEK293A cells.

According to the need to localize into the cell nucleus to mediate viral transcription and replication, the very first cNLS was described on the LTA from SV40. Despite the SV40

LTA (PKKKRKV-132) has since then become the best characterized NLS to date, surprisingly, little is known about the nuclear transport process of LTAs from other Polyomaviridae members. By extending our analyses to all the members of human infecting Polyomavirus family and comparing the sequence of the LTA proteins of all these viruses we could identify at least one putative cNLS on the sequence of each HPyV (Fig.B.2.9). We realized that SV40 LTA cNLS is completely conserved in Human Polyomavirus 2 (JCPyV) and Human polyomavirus 1 (BKPyV), which were also the ones annotated as confirmed nuclear in Uniprot. Besides that, a similar cNLS has been described for the Merkel cell polyomavirus (MCPyV) LTA (RKRK-280), which has been proposed to be required and sufficient for nuclear targeting (Nakamura and Katano, 2009). Strikingly almost 50% of the LTA HPyVs possessed more than one presumed cNLS, like Lyon IARC Polyomavirus (LIPyV) for which there have been estimated three. A putative cdc2 T phosphorylation site immediately upstream of the P2-P5 basic consensus residues, shown to inhibit IMPα/β1 binding and nuclear import on SV40 LTA, was also conserved in all identified cNLS. Similarly, in most cases CK2 phosphorylation enhancing IMPα/β1 binding, and nuclear targeting were also conserved (Fig B.4.1). While some of the tested proteins' LTAs has at least one strong active cNLS accumulating (Fn/c >2) in the nucleus compared to the cycle3-GFP, some of the cNLS seemed to be very weak, like the case of WU, LIPyVs and HPyV7 (Fn/c <2) or not to work, like the case of KI and Saint Louis (STL) PyVs (Fig. B.4.3).

By looking in detail at the sequences (Fig.B.2.9), it can be noticed that the non-functional or weak putative monopartite cNLSs have few basic residues upstream of the sequence, resulting in two close stretches of basic amino acids thus potentially forming a bipartite cNLS. Therefore, we hypothesized that STL, WU, MW and KIPyVs might not have a

monopartite cNLS but a bipartite one (Fig. B.4.4.A). To confirm such idea, we generated the bipartite version of STL, WU and KIPyV cNLS which activity was compared also to a mutant version were the basic residues at the N-term of this bipartite cNLS were substituted with Alanine, which is known to abolish the NLS activity (Weber *et al.*, 1998; Eberhard, Onder and Moroianu, 2013) (Fig. B.4.4.B). Although MWPyV had two functional cNLSs (with Fn/c> 5 and 2<Fn/c<5, respectively), we also combined them into a bipartite cNLS and analyzed it. These plasmid combinations would allow us to understand if the bipartite NLSs are functional, thus allowing us to discriminate between a possible contribution of the non-basic residues upstream the monopartite cNLS and the basic residues of the potential bipartite cNLS. Indeed, if the bipartite form is not functional but the mutant is, it might not be a bipartite cNLS and the upstream residues may contribute to IMPα/β1 binding affinity. The recombinant plasmids expressing either the monopartite, the bipartite or the mutant bipartite cNLS of these four viral proteins were used to transfect HEK293A cells and their respective localization were observed and quantified as already described (§3.16-18). By analyzing the data, STL, KI, WU and MWPyVs had all possible bipartite cNLSs (Fig. B.4.5.B). The Fn/c ratio of the bipartite cNLS was markedly increased for all protein constructs and the mutant cNLSs had barely any activity consistent with the idea that in order to be functional these cNLSs must be in the bipartite form (Fig B.4.5.A).

HPyV7 NLS was one curious case. It's putative cNLS showed a various localization pattern (Fig. B.4.6) compared to the localization of MCPyV and SV40 NLSs, suggesting that the nuclear accumulation depends on the strength of the cNLS. Moreover, the cdc2 T phosphorylation sites may play a role in affecting its nuclear translocation as well as CK2 phosphorylation which may enhance IMPα-cNLS interaction. We still have to uncover

the reason behind this localization pattern and investigate also the role of upstream residues in stabilizing IMPα-cNLS.

Another interesting LTA protein was the one belonging to MCPyV (Fig. B.4.7) which has two putative cNLS where one has a strong activity in accumulating in the nucleus and the other one is relatively weak, their Fn/c values were 8.6 and 1.7 respectively. We questioned about the reason for having two functional cNLS within 20 amino acids, and we hypothesized that these two NLSs might work synergically or as a bipartite cNLS. As for the other plasmids, we generated different versions of MCPyV cNLS sequence where we compared the cNLS activity of the C-term monopartite cNLS, with the bipartite cNLS and the mutated bipartite cNLS, where the upstream basic residues were substituted with Alanines. These data are still ongoing.

To further evaluate the role of these newly identified either monopartite or bipartite cNLSs, these sequences will be crystallized and their binding affinity to different IMPα/β1 isoform will be tested.

B.4.2 Top Hits

In order to evaluate the localization pattern of the hypothetical nuclear proteins identified in our bioinformatic analyses, we focused on proteins bearing putative cNLSs with an NLSmapper score higher than 9, which we considered more likely to be translocated into the host cell nucleus, given the strength of their cNLS (Shunichi Kosugi, Hasebe, Tomita, *et al.*, 2009). Our analysis identified 26 viral proteins bearing putative cNLS which were therefore likely to be endowed with nuclear transport abilities via IMPα/β1 pathway. (Table B.4.2).

Of the 26 identified viral proteins surprisingly 11 belonged to the Poxviridae family members, 10 to Herpesviridae family members and two to Anelloviridae family member. The remaining three proteins belonged to Phabdoviridae, Phenuiviridae and Orthomyxoviridae family members (Fig.B.4.8). These observations suggested that proteins from certain viruses might be wrongly annotated on Uniprot, and this hypothesis was further strengthened by the evidence that 4 out of 26 identified viral proteins were encoded by Human Herpes Virus-7 and 3 from Molluscum Contagiosum virus.

To experimentally test the nuclear targeting ability of such sequences, as for Polyomaviruses, based on the length of cNLS sequence we adopted different cloning techniques that allowed us to generate recombinant plasmid expressing our putative cNLS fused to a reporter fluorescent protein. We generated recombinant plasmid mediating the expression of Cycle3-GFP C-terminal fusion via TOPO® cloning reaction for sequences shorter than 30 amino acids (§3.4), whereas we generated pDEST-YFP vectors with protein fused either at C- terminal of YFP with Gateway® Technology (Life Technologies) (§3.3) for sequences longer than 30 amino acids. HEK293A cells were then transfected with plasmids expressing cNLSs of interest and their localization were observed and quantified via CLSM. Cells were also transfected with a plasmid mediating the expression of either Cycle3-GFP of -YFP alone as a negative control.

Importantly all proteins, accumulated in the nucleus to higher levels as compared to GFP alone (Fig. B.4.9). As for GFP-Q9DUB7, the protein also markedly accumulated in host cells nucleoli. Nucleolar localization was also detected for 6 additional proteins: Q98187, Q6TVJ0, Q6TVM4, Q98291, Q6956 and Q69514, (Fig.B.4.9.A) which implies the possibility that the observed nuclear accumulation is due to passive diffusion across the

NPC into the nucleus, followed by intracellular binding to nuclear structures such as RNA, rather than an active nuclear import.

To discriminate between such possibilities the subcellular localization of GFP fusions was also investigated upon ATP depletion (§3.16) To this end transfected cells were treated one hour before fixation with energy depletion media and incubated at 37°C and 5% CO2 in a humidified incubator. Plasmids mediating the expression of either Cycle3-GFP alone and GFP-SV40TagNLS were used as a negative control. Results showed that as expected incubation with energy depletion media significantly impaired nuclear targeting of GFP-SV40Tag-NLS but did not affect the subcellular localization of GFP alone (Fig. B.4.10). Importantly the treatment also significantly decreased the Fn/c of all nuclear accumulating proteins, demonstrating that nuclear localization was also due to active nuclear import (Fig B.4.10).

B.4.4 Poxviruses

By previous already described analyses, in the Poxviridae family we can notice that three of these proteins are orthologs of Vaccinia A19 viral protein, and four are orthologs of N2 proteins, which are known to go in the nucleus and to interfere with innate cell defense but their NLSs are yet to be characterized (Fig.B.4.8) (Ferguson *et al.*, 2013; Satheshkumar *et al.*, 2013). These data suggested that the ability to encode proteins that go in the nucleus is conserved in Poxviridae family and that the function of these proteins and orthologs must be important in the viral life cycle and modulation of viral-host immune response. Moreover, preliminary data showed that several viral nuclear proteins belonging to this cytoplasmic replicating family are wrongly annotated on

Uniprot since confirmed nuclear proteins were considered to be only 1.13± 0.81% of the whole family members proteome, instead this number increased if hypothetical viral nuclear proteins are also considered (16.85±5.04%) (Fig. B.2.10).

B.4.4.1 N2 protein

Previous studies suggested that N2 proteins is encoded in the early phases of the viral life cycle of Vaccina virus (Morgan and Roberts, 1984) and it's a very small protein of ~175 amino acids for an estimated mass of 21 kDa. Its 3-dimentional structure is very similar to N1 protein (Jacobs *et al.*, 2008; Ferguson *et al.*, 2013; Maluquer de Motes *et al.*, 2014), where they share a Bcl-2 homology, but the main difference relies on a loop present between the αhelices of the N2 protein.

Our analyzes revealed that all these proteins from vaccinia virus orthologs contain potential strong monopartite cNLSs, like the case of COWPX, which Fn/c was 9.6 with a predicted cNLS mapper score of 10.3. On the other hand other 3 proteins, from Horse Pox virus (HSPV), MonkeyPox Virus (MNPZ) and VAR67 had very weak NLSs with an Fn/c of ~1 compared to the predicted cNLS mapper score of 9.5 (Fig B.4.12).

Sequence alignment of the residues present in the extra sequence of the N2 proteins in respect to N1 highlighted 102-RKR-104 basic residues upstream the predicted monopartite NLS for HSPV, MNPZ and VAR67, suggesting a potential bipartite cNLS for N2 and all its orthologs (Fig B.4.13).

We tested this hypothesis by comparing the activity of the monopartite and bipartite cNLS for the sequences of HSPV and MONPZ (since VAR67 presented the exact sequence

of HSPV was not included in the experiment). We cloned them into recombinant vector and as for T-antigens we used them to transfect HEK293A cells, where at 48 h post transfection the localization of their cNLSs were seen by CLSM and their activity was quantified (Fig. B.4.14). Plasmid mediating the expression of Cycle3-GFP only was used as negative control. Surely the activity of the bipartite NLSs was much stronger than the monopartite ones, consistent with cNLSmapper score and with the hypothesis that these NLSs are functional only as bipartite cNLSs. To validate these results, we analyzed mutant versions of these bipartite cNLS and compared their nuclear accumulation with the non-mutant ones. The Fn/c values drastically decreased for the sequences with mutation either on the C-term, N-term or both, compared to the wild type bipartite cNLS, confirming our theories (Fig. B.4.15). In the end we successfully characterized Vaccina N2 bipartite cNLS. In addition to our approach, we investigated whether the nuclear import is via passive diffusion or energy mediated, since vaccinia N2 is a very small protein it might localize in the nucleus either by passive diffusion or with an IMPα/β1 independent manner. To this end we treated transfect cell expressing wild type VACC-N2 cNLS either with an IMPα/β1 transport inhibitor or depleting energy by removing ATP (§3.16) from the culture media as described in the 4.4 section and quantified the cNLS activity (Fig.B.4.16).

Intriguingly, upon ATP-depletion the protein seems to accumulate even better in the nucleus raising questions about the role of the energy in the export pathway, but the nuclear accumulation is impaired when transfected cells are treated with IMPα/β1 transport inhibitor Bimax2, suggesting that the nuclear import is IMPα/β1 dependent (Fig.B.4.16).

B.4.4.2 A19

As for the case of N2 proteins, A19 is another Vaccinia Virus protein for which there is little knowledge about. This 9kDa protein it's thought to have a role the virion morphogenesis but its specifical function is still unclear (Satheshkumar *et al.*, 2013). The protein cNLS sequence is conserved among Poxviruses (Fig B.4.17) and since this protein and its orthologs are annotated as possible nuclear proteins, for a better comprehension of its role we investigated the mechanisms of its nuclear import and localization. As for previous experiments we fused the full-length protein sequence of A19 and its orthologs, with Cycle3-GFP and transfect HEK293A cells with this newly generated recombinant vector, using the plasmid expressing Cycle3-GFP only as negative control. After 48 hours post transfection, we sought for the protein nuclear localization. Compared to the ubiquitous Cycle3-GFP localization, the proteins accumulated in higher levels into the nucleus and mutation of specific residues on the predicted 10-MKSRKKKPKTT-20 monopartite NLS sequence impaired nuclear accumulation of A19. When introducing Bimax2, an inhibitor of nuclear import IMPα/β1 dependent, the protein fails also to accumulate into the nucleus. Same results were obtained with ATP depletion treatment, where the protein nuclear accumulation is impaired suggesting that the protein has a monopartite cNLS which is actively transported into the nucleus relying on IMPα/β1 complex for nuclear import.

B.5 Discussion

Nuclear transport is an active process in eucaryotic cells, where the nucleus, so the genetic information, is separated from the cytoplasm and other cell organelles by the Nuclear Envelope (NE), thus passage through the NE is possible thanks to Nuclear Pore Complexes (NPCs). Small proteins and molecules up to 60 kDa pass through the NPC via passive diffusion. Since the rate of passing is inversely proportional to the protein size, large proteins exploit nuclear transporters for a rapid and efficient translocation.

These proteins recognize specific sequences on their cargo called nuclear localization (NLS) and rely on Ran proteins for cargo release.

There are different kinds of NLSs, and all transporters belong to the importin family. The best characterized NLSs are the Classic NLSs that are highly basic sequences recognized by IMPα/β1. cNLSs can be either monopartite or bi-partite if they bind to the major (in some cases also minor) or both NLS binding sites of the IMPα, the adapter for IMPβ1 which is the real transporter.

Other NLSs are directly recognized by IMPβ, some examples are the R-rich NLSs, recognized by IMPβ1 and PY-NLSs recognized by IMPβ2.

Viruses exploit this machinery for the nuclear transport of their proteins because crucial for the viral life cycle. Cytoplasmic replicating virus mainly use cell transporters to modulate host cell functions and in addition to that, nuclear replicating viruses use nuclear transport also for the delivery of viral genome to the nucleus, expression, replication, and the encapsidation of the newly formed virions.

Viral protein-MPα/β interaction inhibitors impair viral infection of a variety of viruses. Therefore, it's far-reaching to characterize the nucleocytoplasmic trafficking pathway of human infecting viruses. Therefore, to investigate the relationship of human infecting viral proteins with the transport apparatus we performed bioinformatic and

experimental analyses. We retrieved the proteome of all known human viruses from Viralzone. Then we performed a bioinformatic analysis where we quoted from Uniprot all the functional annotations of these proteins and combined them with NLS and localization prediction softwares independently. We classified all viral proteins into three main categories: Confirmed Nuclear, Hypothetical Nuclear and Cytosolic Proteins. And thanks to the data acquired we could perform several analyses.

For example, comparative bioinformatic analyses showed that nuclear replicating viruses have higher amount of nuclear proteins compared to the cytoplasmic replicating ones (Fig. B.2.5) as expected, and the percentage of nuclear proteins increases if hypothetical nuclear proteins are also included (Fig. B.2.5).

Moreover, a great percentage of all considered nuclear proteins bear a cNLS consensus sequence (Fig.B.2.6). Therefore, underlining the heavily relying of viral nuclear proteins on IMPα/β1 pathway.

Our analysis revealed that the small Polyomaviruses heavily rely on nuclear transport to guarantee their life cycle, since the percentage of nuclear proteins were about 80% if all confirmed and putative nuclear proteins were considered (Fig.B2.7). Among the conserved proteins in this family Large T antigen (LTA) was annotated in Uniprot as nuclear only in two family members, highlighting once again that several proteins are likely to be misannotated. We focused our attention on the characterization of the LTA of all human infecting polyomaviruses. Surprisingly, more than 50% of the LTAs bared more than one putative cNLS and the localization of all HPyV LTAs' cNLSs was observed and quantified as a ratio between Nuclear fluorescent signal and cytoplasmic fluorescent signal (Fn/c) by Confocal Laser Scanning Miscroscopy (CLSM) and ImageJ software.

Our analyses revealed that the LTA of Human Polyomavirus 1 (BKPyV), Human Polyomavirus 2 (JCPyV), Trichodysplasia spinulosa polyomavirus (TSPyV), Murine Polyomavirus (MPyV) and NUPyVs as well as HPyV6, 9 and HPyV12, were bearing a strong functionally active monopartite cNLS, whereas Saint Louis Polyomavirus (STLPyV), KI, WU and MWPyVs are more likely to have a bipartite cNLS. HPyV7 and MCPyV were two curious cases where the first showed a variable localization pattern and the second has two functional cNLS which can either work as a bipartite cNLS or synergically.

In other setting, we correlated the proteome size and the amount of viral nuclear proteins, and the data showed that they were inversely proportional. If considering the first Baltimore class of viruses, in large cytoplasmic replicating Poxviridae family the percentage of viral nuclear proteins increased if all confirmed and hypothetical nuclear proteins were considered. Therefore, Poxviridae were a good starting point to identify new viral nuclear proteins.

Overall, our studies identified nearly 200/300 new viral nuclear proteins from different viruses and 11 out of 26 proteins with the highest cNLS mapper score, belonged to the Poxviridae family. Nevertheless, we decided to functionally characterize all the 26 proteins. As for all the other proteins investigated in this study, we fused the NLS signal to a reporter protein and transfected HEK -A cells. We quantified their localization by calculating the Fn/c signal and all of them localized into the nucleus. Surprisingly, four of these proteins were orthologs of vaccinia N2 protein, and three of Vaccinia A19. These data suggested that the ability to encode proteins that go in the nucleus is conserved in Poxviridae family, and that A19 and N2 and their orthologs need to go in the nucleus.

Both are known to interfere with innate cell defense, but their NLS and specific functions are yet to be characterized. By sequence alignment we first predicted that N2 and its orthologs has a bipartite NLS and A19 and its orthologs has a classical NLS.

We fused full length protein of A19 with GFP and the protein localized into the nucleus, but mutation of specific residues on the NLS sequence impaired nuclear accumulation of A19.When introducing Bimax2, an inhibitor of nuclear import IMPα/β1 dependent, the protein failed to accumulate into the nucleus and similar localization pattern was observed with ATP depletion treatment, suggesting that the protein was actively transported into the nucleus and relied on IMPα/β1 complex to do so. We obtained similar results for A19 orthologs

For the case of N2 protein similar localization pattern as A19 was observed even with mutations of residues in the bipartite sequence and Bimax treatment. As for A19, we obtained similar results also with N2 orthologs.

B.6 Figures, Tables and Legends

Plasmids ordered on Biofab , Vector Builder or supplied		
Code	**Name**	**Description**
561	pUC57-VACCWA19	Plasmid purchased by BioFab
562	pUC57-VACCWN2	Plasmid purchased by BioFab
749	pUC-57_VACC-N2-NLSmut	Plasmid purchased by BioFab
750	pUC-57_VACC-A19-NLSmut	Plasmid purchased by BioFab
724	HPyV7-EGFP	Purchased by Vector Builder
742	EGFP-CAPSD-TTVZ1-NL	Purchased by Vector Builder
747	Nucl-pRed	Supplied Plasmid
721	pEGFP-N1-SV40.LT	Supplied from Patrick Moore
722	pcLT206-eGFP	Suppied from Patrick Moore

Table B.3.1 Plasmid and sequences purchased or gifted

BP REACTION	
Reaction	**Resulting Entry Vector**
GW5 pDNR207 x **561** pUC57-VACCW-A19	**637** pDNR207-VACCW-A19
GW5 pDNR207 x **562** pUC57-VACCW-N2	**638** pDNR207-VACCW-N2
GW5 pDNR207 x **749** pUC57-VACCW-N2-NLS mut	**751** pDNR207-VACCW-N2 -NLS mut
GW5 pDNR207 x **750** pUC57-VACCW-A19-NLSmut	**752** pDNR207-VACCW-A19 – NLS mut
GW5 pDNR207 x PCR product [**724** x p **123/124**]	**763** pDNR207-HPyV7-NLS ext
GW5 pDNR207 x PCR product [**779** x p **124/125**]	**764** pDNR207-HPyV7-NLS ext-mut

Table B.3.2 BP reactions performed in this study and the relative generated vector

LR REACTION	
Reaction	**Resulting Expression Vector**
GW22 pDEST-N-YFP x **637** pDNR207-VACCW-A19	**697** pDEST-N-YFP-VACCW-A19
GW22 pDEST-N-YFP x **638** pDNR207-VACCW-N2	**715** pDEST-N-YFP-VACCW-N2
GW22 pDEST-N-YFP x **751** pDNR207-VACCW-N2 -NLS mut	**753** pDEST-N-YFP-VACCW-A19-NLS mut
GW22 pDEST-N-YFP x **752** pDNR207-VACCW-A19 – NLS mut	**754** pDEST-N-YFP-VACCW-N2-NLS mut
GW22 pDEST-N-YFP x **763** pDNR207-HPyV7-NLS ext	**779** pDEST-N-YFP-HPyV-NLS ext
GW22 pDEST-N-YFP x **764** pDNR207-HPyV7-NLS ext-mut	**786** pDEST-N-YFP-HPyV-NLS ext-mut

Table B.3.2 LR reactions performed in this study and the relative generated vector

TOPO recombinant plasmids		
Code	**Name**	**Codifying Protein sequence**
588	pcDNA3.1 NT-GFP- TOPO	/
575	pcDNA3.1 NT-GFP- TOPO SV40 TagNLS	PPKKKRKV*
606	pcDNA3.1 NT-GFP- TOPO- KIPyVtagC	PPKKKHA*
645	pcDNA3.1NT GFP-TOPO KIP_tag_Cb	KRSAPEEEPSCSQATPPKKKHA*
608	pcDNA3.1 NT-GFP- TOPO- WUPyVtagM	PTKRTRE*
609	pcDNA3.1 NT-GFP- TOPO- WUPyVtagC	PPKKKKDNA*
695	pcDNA3.1 NT–GFP–WU_MC_NLS	KRTREDDEEPQCSQATPPKKKD*
576	pcDNA3.1 NT-GFP- TOPO- STLPytagC	PPKKNKPA*
593	pcDNA3.1 NT-GFP- TOPO- STLPyVtagCb	KRKFPDSSTQNSTPPKKNKPA*
610	pcDNA3.1 NT-GFP- TOPO- HPyV7tagC	PPKQKKPN*
592	cDNA3.1 NT-GFP- TOPO- LipyVtagM	PKRNRKNQ*
600	pcDNA3.1 NT-GFP- TOPO- LipyVtagN	PRPKKRRSNL*
601	pcDNA3.1 NT-GFP- TOPO- LipyVtagC	PPKQKRYKE*
647	pcDNA3.1 NT GFP-TOPO-MCP_tag_M	PFSRKRKFGGS*
648	pcDNA3.1 NT GFP-TOPO-MCP_tag_C	PPKPKKNRE*
579	pcDNA3.1 NT-GFP- TOPO- NJPyVtagC	PPKQKRKSP*
595	pcDNA3.1 NT-GFP- TOPO- NJPyVtagM	PPKRRRGT*
607	pcDNA3.1 NT-GFP- TOPO- TSPyVtagC	PPKPKKSKY*
611	pcDNA3.1 NT-GFP-TOPO- HPyV12tagM	PPKRGRNGGG*
612	pcDNA3.1 NT-GFP-TOPO-HPyV12tagC	PPKSKKAKM*
578	pcDNA3.1 NT-GFP-TOPO-HPyV6tagC	PPKKRKPN*
590	pcDNA3.1 NT-GFP-TOPO-HPyV9tagC	PPKRKKPE*
591	pcDNA3.1 NT-GFP-TOPO-MPyVtagN	PVSRKRPRPA*
602	pcDNA3.1 NT-GFP-TOPO-MPyVtagC	PPKKARED*
577	pcDNA3.1 NT-GFP-TOPO-MWPyVtagC	PPKRPRNF*
594	pcDNA3.1 NT-GFP-TOPO-MWPyVtagM	PKKRPRES*
646	pcDNA3.1 NT-GFP-TOPO-MWPyV_tag_MC	PKKRPRESSSNSTCTPPKRPRNF*
755	pcDNA3.1 NT-GFP-TOPO-KIPyV_tag_bip_Mut	aaSAPEEEPSCSQATPPKKKHA*
757	pcDNA3.1 NT-GFP-GFP-WU_tag_bip_mut	aaTaEDDEEPQCSQATPPKKKKD*
756	pcDNA3.1NT GFP-TOPO STLP_tag_bip_mut	aaaFPDSSTQNSTPPKKNKPA*
613	pcDNA3.1 NT-GFP-TOPO Q7M6G6 (US34A_HCMV)-NLS	KFRKRRRPVVV*
614	pcDNA3.1 NT-GFP-TOPO Q69567 (IE1_HHV6U) -NLS	KRVAKRKHVSSKSPKNKKIKTD*
657	pcDNA3.1 NT GFP-TOPO Q98263_MCV1-NLS	RPSAKRRRCSR*
615	pcDNA3.1 NT-GFP-TOPO Q9DUB8 (ORF2_TTVZ1) -NLS	PPPPKKRRPWC*
659	pcDNA3.1 NT GFP-TOPO YLDV_Q9DHK5-NLS	LKWLRKKRKIALQTY*
660	pcDNA3.1 NT GFP-TOPO MCV1_Q98187-NLS	RTHKRKGTPLPLRPRSKRVRAR*
676	pcDNA3.1 NT GFP-TOPO D2CRM8_DUVV-NLS	RFRKRRKSKP*
677	pcDNA3.1 NT GFP-TOPO P28284 (ICP0_HHV2H) -NLS	RPRKRRGSDS*
661	pcDNA3.1 NT GFP-TOPO P03199 (BLLF2_EBVB9)-NLS	RPPVAKRRRFPR*
678	pcDNA3.1 NT GFP-TOPO P52352 (GB_HHV7J) -NLS	ASRKRRKREL*
679	pcDNA3.1 NT GFP-TOPO Q6TVJ0_ORFSA-NLS	RRKRKRKTPNC*
681	pcDNA3.1 NT GFP-TOPO U5TCT3_COWPX-NLS	IRKRPNQHHTIDLFKRMKR*
685	pcDNA3.1 NT GFP-TOPO Q98291_MCV1-NLS	RRRKRKPRTT*
682	pcDNA3.1 NT GFP-TOPO D6PT84_ORTO-NLS	RRANKRRLEEL*
683	pcDNA3.1 NT GFP-TOPO P16738 (UL109_HCMVA) -NLS	RRGKRRKLI*
684	pcDNA3.1 NT GFP-TOPO P52344_HHV7J-NLS	RPCKVKRKLFGSENIRPNKKIPL*
658	pcDNA3.1 NT GFP-TOPO Q69518_HHV7J-NLS	KRSFPEICPEHFKKRRFI*

687	pcDNA3.1 NT GFP-TOPO Q69514_HHV7J-NLS	KKQLKRKSESKLKTSKAKKKKLI*
723	pcDNA3.1 NT-GFP-TOPO Q9DUB7(CAPSD_TTVZ1) -NLS	RWRRRPRRRRRPYRRRPYRRYGRRRKVRRR*
686	pcDNA3.1 NT GFP-TOPO Q0GP58_HSPV/N2_VAR67-NLS	LFKKIKRTRYDTF*
772	pcDNA3.1 NT GFP-TOPO Q8V556_MONPZ-NLS	LFKRIKRTRYDTF*
719	pcDNA3.1 NT GFP-TOPO Q6TVM4_ORFSA-NLS	GNKKRRRRRV*
717	pcDNA3.1 NT GFP-TOPO A0A2N9DYY9_HHV6U-NLS	EYTKKRRRHRV*
720	pcDNA3.1 NT GFP-YOPO P22026 (NSS_UUKS) -NLS	RLRRKKRSRVS*
680	pcDNA3.1 NT GFP-TOPO Q6TUQ8_YMTV5-NLS	GGAKRKKRKPK*
718	pcDNA3.1 NT GFP-TOPO Q8V556_MONPZ-NLSbip	IRKRPNQHHTIDLFKRIKRTRYDTF*
731	pcDNA3.1 NT GFP-TOPO Q8V556_MONPZ-NLSbip-short	IRKRPNQHHTIDLFKRIK
733	pcDNA3.1 NT GFP-TOPO Q8V556_MONPZ-NLSbip-shortNm	IaaaPNQHHITDLFKIRK*
734	pcDNA3.1 NT GFP-TOPO Q8V556_MONPZ-NLSbip-shortCm	IRKRPNQHHTIDLFaaIa*
735	pcDNA3.1 NT GFP-TOPO Q8V556_MONPZ-NLSbip-shortNCm	IaaaPNQHHTIDLFaaIa
732	pcDNA3.1 NT GFP-TOPO Q0GP58_HSPV/N2_VAR67-NLSbip	IRKRPNQHHTIDLFKKIKRT*
736	pcDNA 3.1-NT-GFP-TOPO 5UTCT3-COWPX-Nm	IaaaPNQHHTIDLFKRMKR*
737	pcDNA 3.1-NT-GFP-TOPO 5UTCT3-COWPX-Cm	IRKRPNQHHTIDLFaaMaa*
738	pcDNA 3.1-NT-GFP-TOPO 5UTCT3-COWPX-NCm	IaaaPNQHHTIDLFaaMaa*
805	pcDNA 3.1 GFP-TOPO-Q9DUB7-cNLSCt	LPPPEKRARWGFP*
806	pcDNA 3.1 GFP-TOPO-D6PT84-cNLSct	DEARRKRLKRVCLM*
807	pCDNA3.1 GFP-TOPO-MCPyV NLSm278	PFSRtRKFGGS*
808	pCDNA3.1 GFP-TOPO- MCPyV NLSm280	PFSRKRTFGGS*

Table. B.4.1 Recombinant clones obtained by TOPO reaction. cNLs of putative nuclear viral proteins fused in frame with Cycle-3-GFP and their coding sequence are shown.

GATEWAY recombinant plasmids		
Code	**Name**	**Codifying Protein sequence**
697	pDEST-N-YFP-VACCW-A19	MKSRKKKPKTT
715	pDEST-N-YFP-VACCW-N2	IRKRPNQHHTIDLFKRIKR*
753	pDEST-N-YFP-VACCW-A19-NLS mut	MKSRaaaPaTT
754	pDEST-N-YFP-VACCW-N2-NLS mut	IaaaPNQHHTIDLFaaIaa
779	pDEST-N-YFP-HPyV-NLS ext	SSSDEEEPASSASVNPEEGCSQDSKYSATPPKQKKPNP
786	pDEST-N-YFP-HPyV-NLS ext-mut	aaaDEEEPASSASVNPEEGCSQDSKYSATPPKQKKPNP
824	pDESTntYFP_MCPyV_NLSbip	KRKFGGSRSSASSASSASFTSTPPKPKKNRE

Table. B.4.1 Recombinant clones obtained by GATEWAY reaction. cNLs of putative nuclear viral proteins fused in frame with EYFP, and their coding sequence are shown.

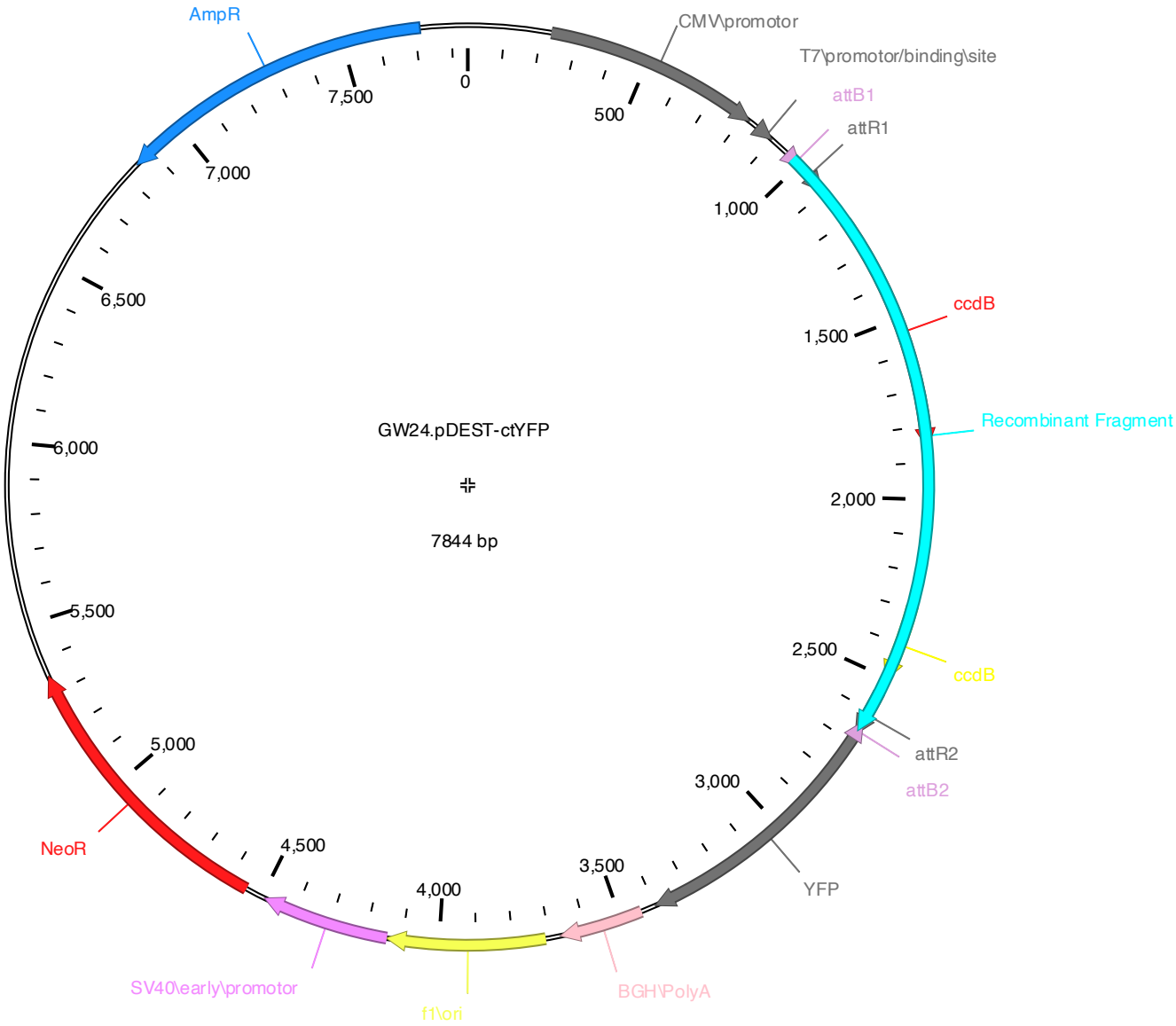


Figure B.3.4. pDEST-CT-YFP. This recombination vector was used in Gateway reaction to generate expression clones from sequences bearing *attB* recombination sites through BP and LR reactions. Sequences lacking STOP codon were fused in frame with YFP at the N-terminal under T7 promoter, in order to detect their subcellular localization by CLSM

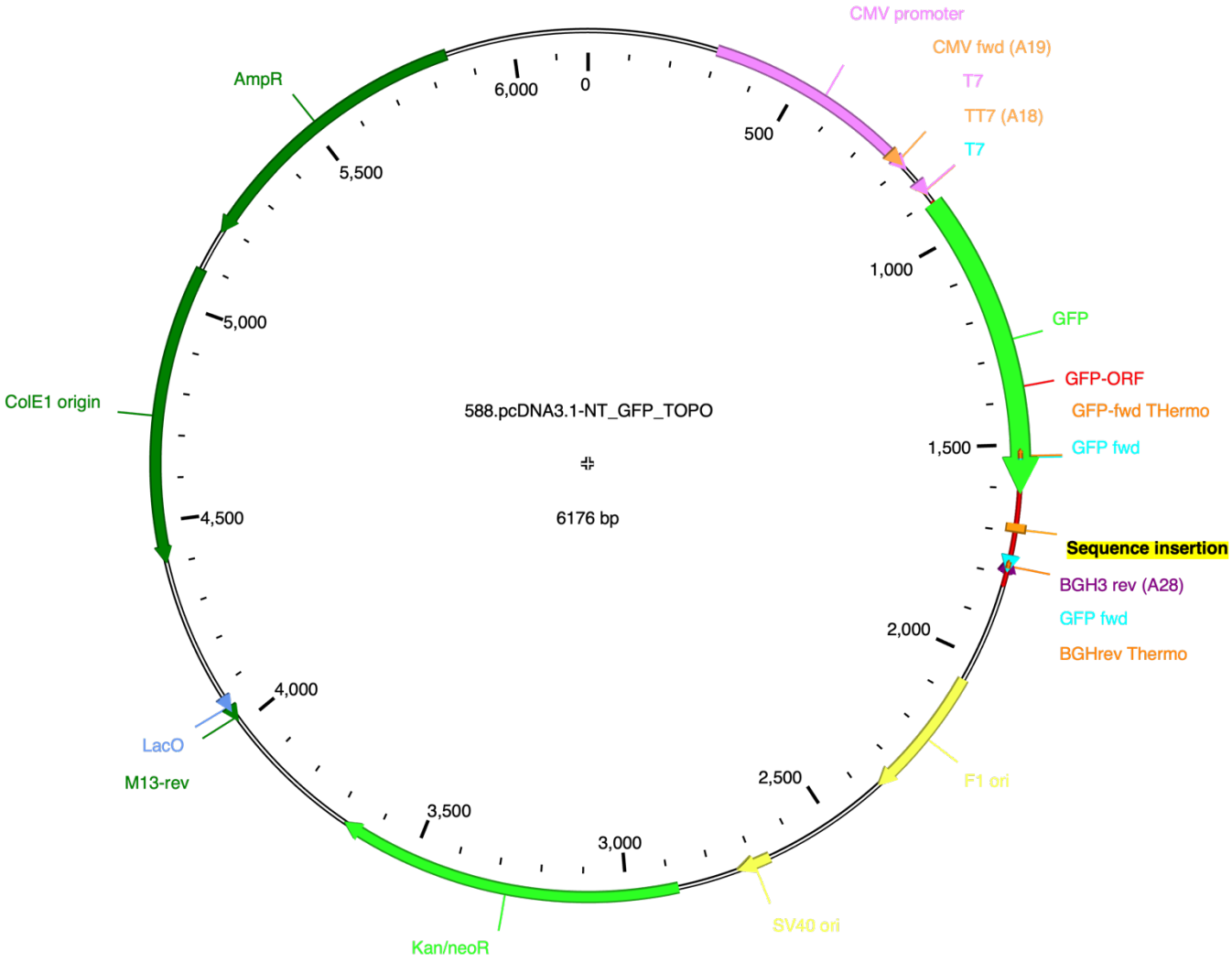


Figure B.3.6. pcDNA3.1 NT-GFP-TOPO. This recombination vector was used to clone cNLSs sequences less than 30 aa via TOPO® reaction in frame with Cycle3-GFP under the promoter T7. cNLSs subcellular localizations were detected thanks to Cycle-3-GFP fluorescent signal by CLSM. Empty vector was used as control since it localizes ubiquitously within the host cell.

Figure B.4.1. Confirmed and Putative cNLS sequences on Large T antigen proteins of all human infecting Polyomaviruses. By sequence analyses it was clear that all PyV LT bear at least on cNLS on their sequence. Surprisingly SV40LTag cNLS was conserved in JC and BK PyVs. Almost 50% of all these proteins have more than one putative cNLS, for LYPyV there have been estimated three. In green Serine residues upstream of cNLSs are highlighted since potential target for CK2 phosphorylation. It can also be notable that there is a putative cdc2 T phosphorylation site immediately upstream of the P2-P5 basic cNLSs consensus residues

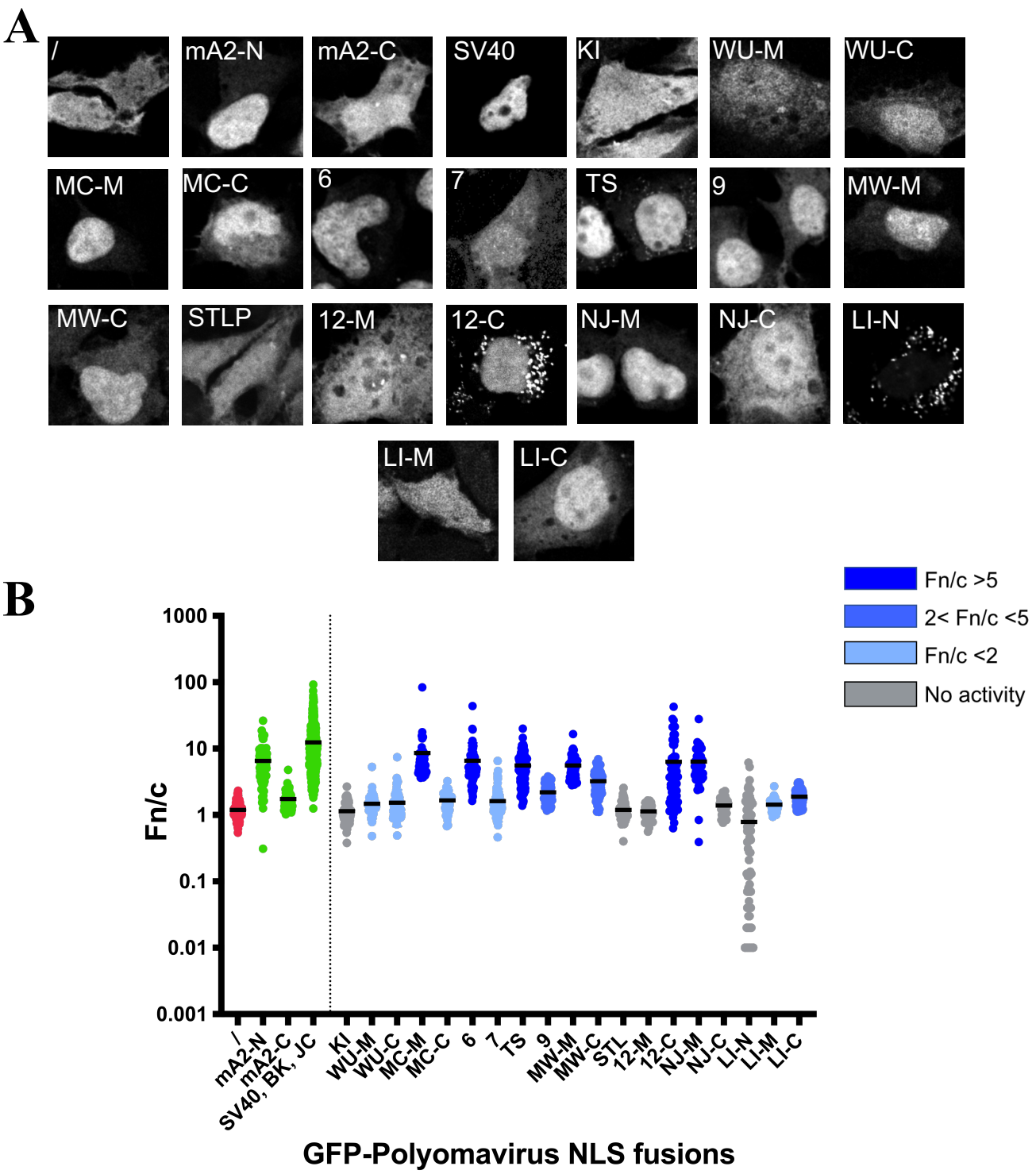


Figure B.4.2. Subcellular localization and quantification of all cNLSs present on Human infecting Polyomaviruses LTs. Panel A shows the subcellular localization of Cycle3-GFP fused putative cNLSs. HEK293A cells were seeded and transfected with recombinant vector expressing cNLSs fused in frame with Cycle3-GFP. 48hrs post transfection cells were fixed, mounted on coverslips and the proteins subcellular localization was observed by CLSM and the mean signal quantified (B). Compared to GFP ubiquitous localization, several proteins accumulated in the host cell nucleus. Interestingly some cNLSs seemed to be very weak or non-functional

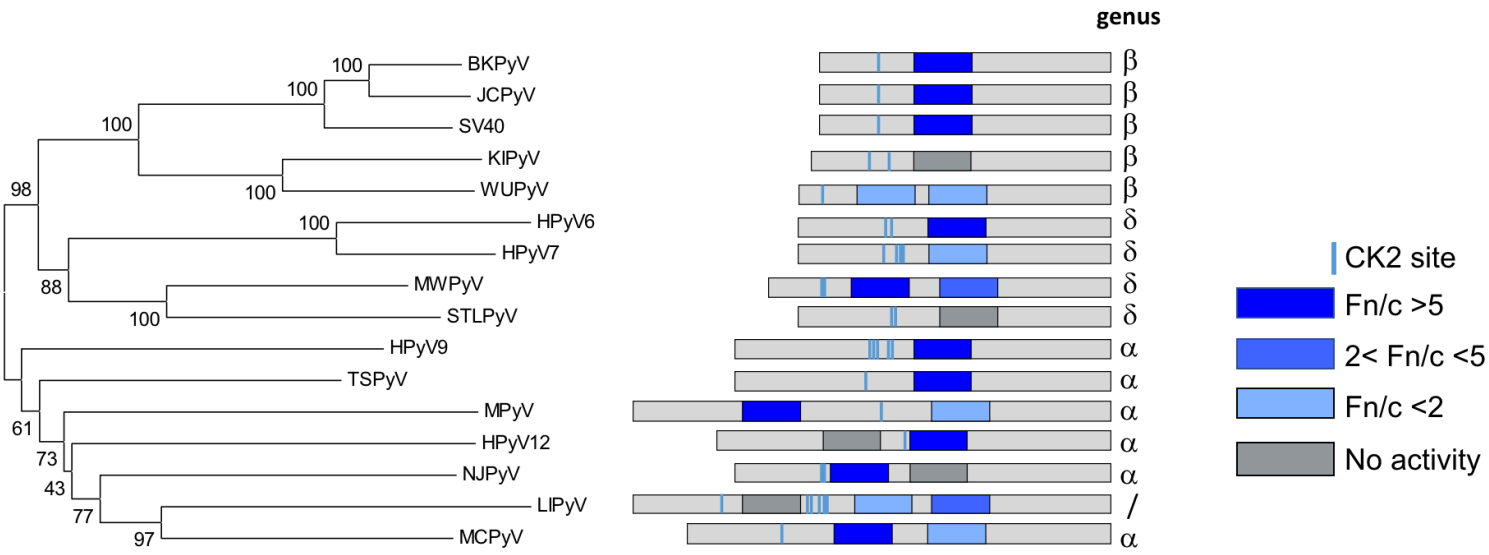


Figure B.4.3. Polyomaviruses LT cNLSs functionality. This panel summed up the correlation between different Polyomavirus LT proteins and classifies the functionality of each tested respective putative cNLS sequences and the potential CK2 phosphorylation S residues.

A

Virus	Sequence	Linker	Monopartite NLS
MWPyV	KKRPRESSSNSTCTPPKRPR	11	good; fair
STLPyV	KRKFPDSSTQNSTPPKKNK	12	none
KIPyV	KRSAPEEEPSCSQATPPKKKH	15	none
WUPyV	KRTREDDEEPQCSQATPPKKKK	16	Poor; poor
MCPyV	KRKFGGSRSSASSASSASFTSTPPKPKKNR	20	good; poor
LIPyV	KRNRKNQSFGGIPSPGSRRSFSSTPPKQKRYK	21	poor; fair
NJPyV	KRRRGTEDLDGSYTDSQTSFASTPPKQKRK	21	good; none
HPyV12	KRGRNGGGSAPSTSGGVPDDFEGDSADQNCSQATPPKSKKAK	32	none; good

B

GFP-fusion	Sequence
STLPyV-NLS	PP**KK**N**K**PA-144
STLPyV-NLS_bip	**KRK**FPDSSTQNSTPP**KK**N**K**PA-144
STLPyV-NLS_bip_m	aaaFPDSSTQNSTPP**KK**N**K**PA-144
KIPyV-NLS	PP**KKK**HA-140
KIPyV-NLS_bip	**KR**SAPEEEPSCSQATPP**KKKH**A-140
KIPyV-NLS_bip_m	aaSAPEEEPSCSQATPP**KKKH**A-140
WUPyV-NLS_C	PP**KKKK**DNA-148
WUPyV-NLS_bip	**KR**TREDDEEPQCSQATPP**KKKK**D-146
WUPyV-NLS_bip_m	aaTaEDDEEPQCSQATPP**KKKK**D-146
MWPyV-NLS-M	P**KKR**P**R**ES-142
MWPyV-NLS-C	PP**KR**P**R**NF-157
MWPyV-NLS-bip	P**KKR**P**R**ESSSNSTCTPP**KR**P**R**NF-157

Figure B.4.4 cNLSs sequences examination. The weak and non-functional cNLS sequences relative to MW, STL, KI, WU, MC, LI, NJPyVs and HPyV12 were carefully examinated. It can be noticed basic residues upstream resulting in two close stretches of basic aminoacids resembling a bipartite cNLS. Therefore, we assumed that some cNLS function as bipartite cNLS (A). To this end, recombinant vector encoding either bipartite cNLSs or their mutant version were generated in order to evaluate their cNLS activity (B).

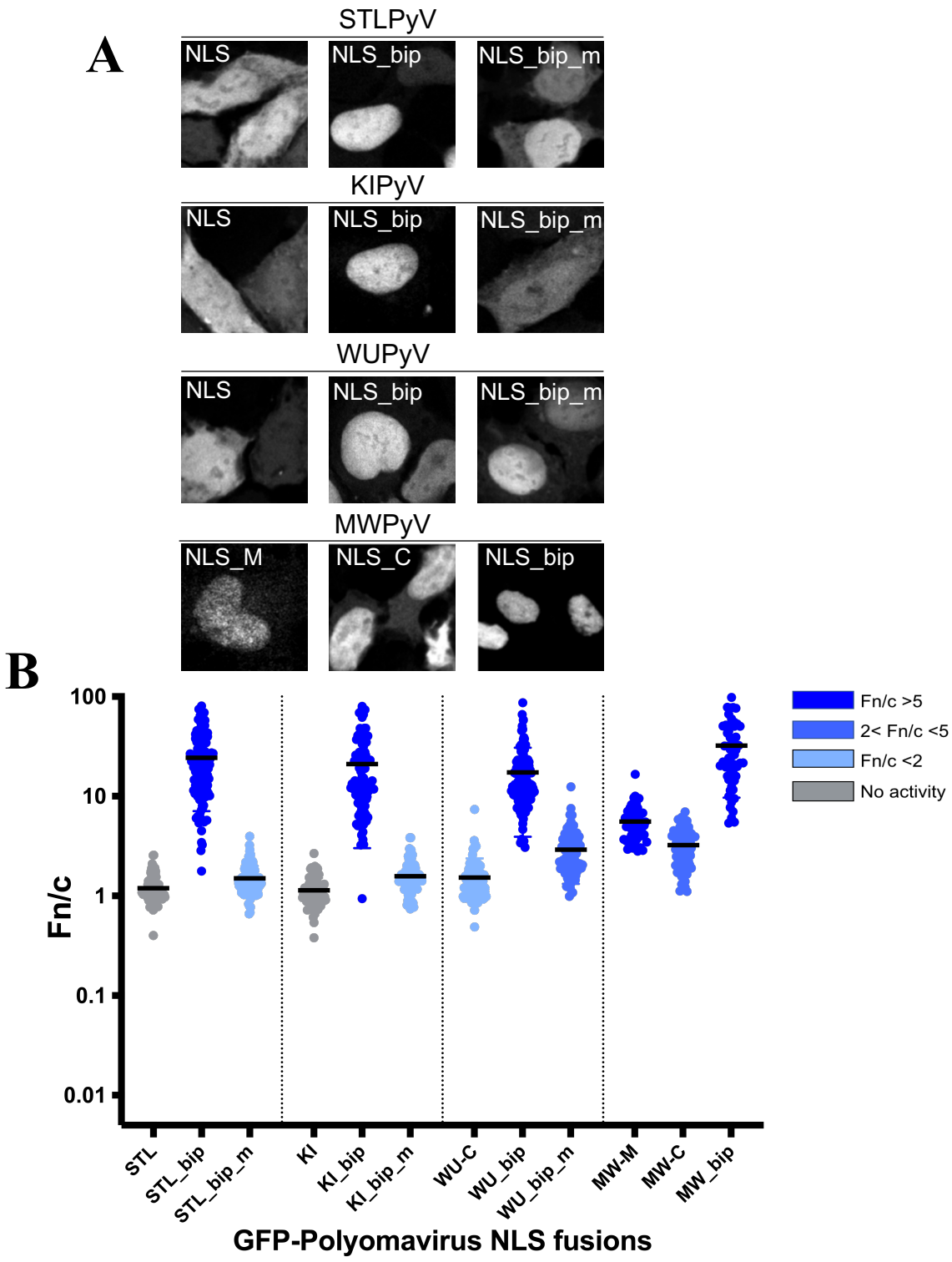


Figure B.4.5 Subcellular localization and quantification of wildtype and mutant bipartite cNLS of STL, KI,WU and MWVPyVs. Recombinant vectors expressing either bipartite cNLS, mutant bipartite cNLS or monopartite cNLS of STL, KI, Wu and MWVPyV were fused to Cycle3-GFP and used to transfect HEK293A cells. 48h.p.t. cells were fixed and mounted on coverslips and the subcellular localization of Cycle3-GPF fused proteins were observed by CLSM (A) and quantified (B). By microscopy analyses it's clear that bipartite cNLSs are more active compared to their mutant and monopartite counterpart. These observations were confirmed also by quantification f their nuclear accumulation (B).

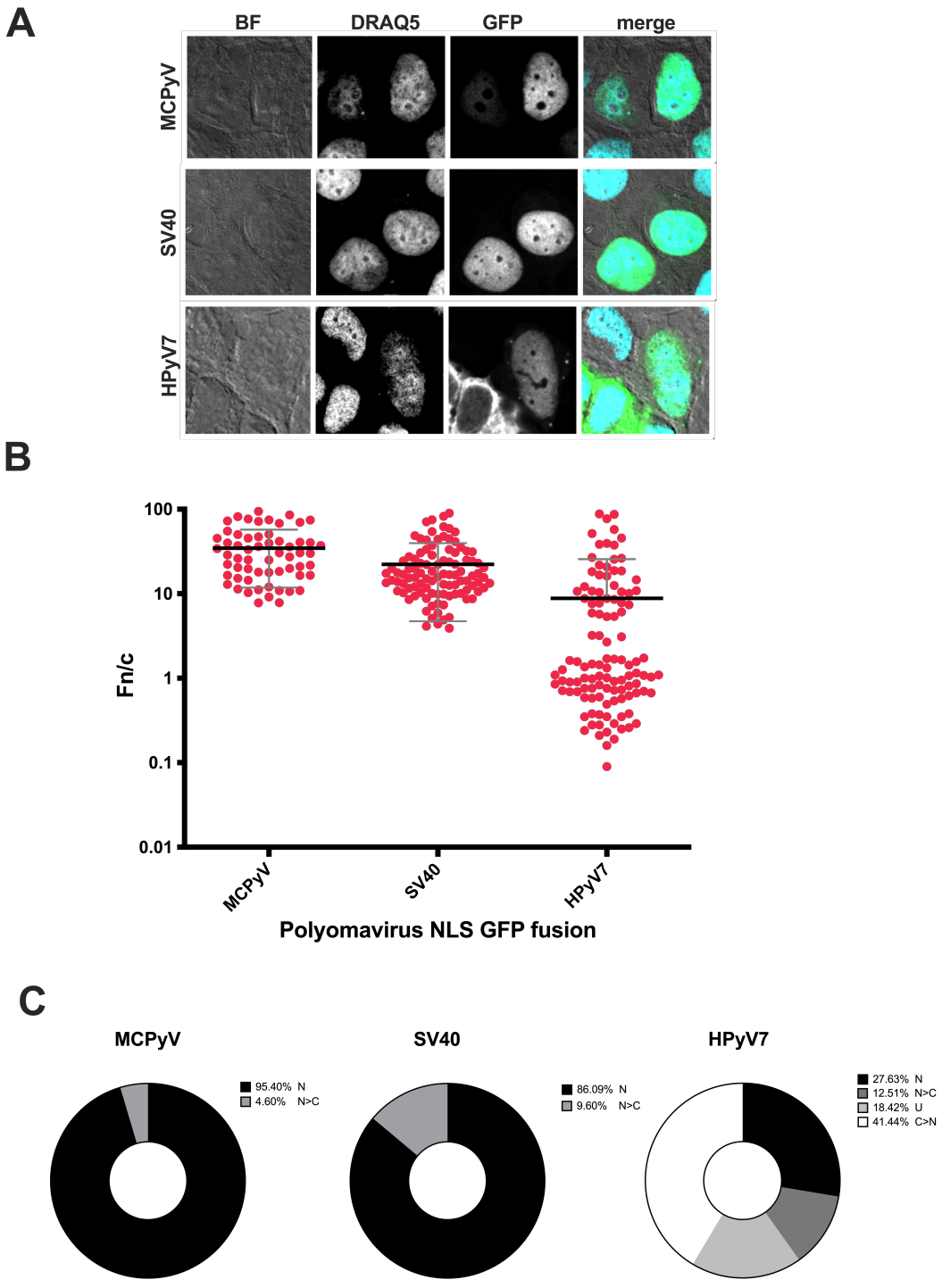
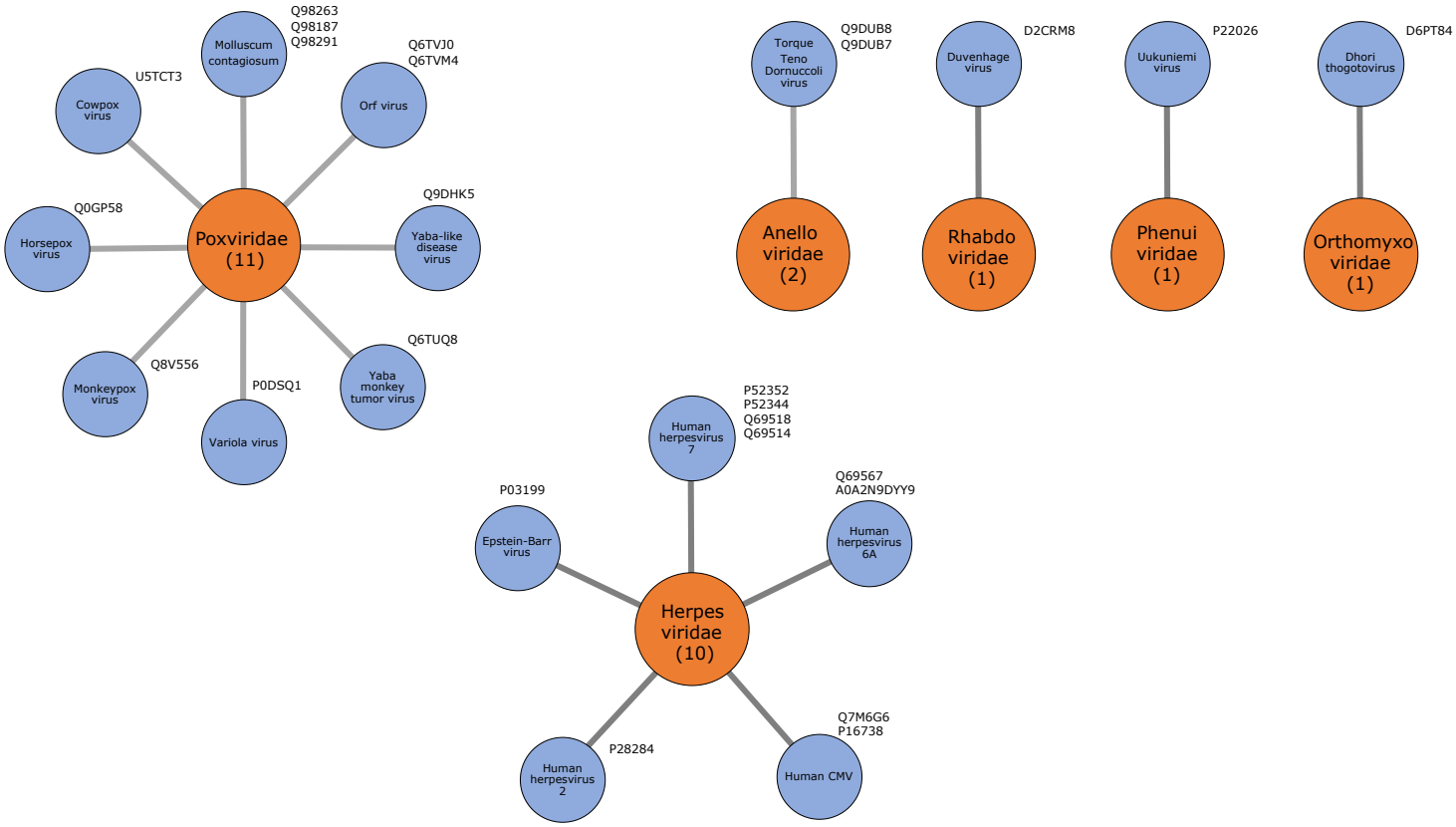


Figure B.4.6. HPyV7 cNLS presents a unique localization pattern. Compared to the localization of MCPyV and SV40 NLSs, HPyV7 cNLS subcellular localization is various (A). Its localization pattern was quantified, and the wide range of Fn/c values (B) suggest that the nuclear accumulation depends on the strength of the cNLS. cdc2 T phosphorylation sites may play a role in affecting its nuclear translocation as well as CK2 phosphorylation which may enhance IMPα-cNLS interaction. The redundancy of different HPyV7 cNLS subcellular signal localizations was also compared with SV40 expressed in percentage of the overall analyses cells (C).

GFP-fusion	Sequence
MCPyV-NLS-C	----------PPKPKKNRE
MCPyV-NLS-bip	KRKFGGSRSSASSASSASFTSTPPKPKKNRE
MCPyV-NLS-bip_m	aaaFGGSRSSASSASSASFTSTPPKPKKNRE

Figure B.4.7. The case of the two active MCPyV cNLSs. MCPyV has two putative cNLS where one has a strong activity in accumulating in the nucleus and the other one is weak, with Fn/c values of 8.6 and 1.7 respectively. We hypothesize that these two NLSs might work synergically or as a bipartite cNLS. Further investigations are still on going.



Poxviridae	Herpersviridae	Anelloviridae
Q98263	Q7M6G6	Q9DUB8
Q9DHK5	Q69567	Q9DUB7
Q98187	P28284	Orthomyxoviridae
Q6TVJ0	P03199	D6PT84
U5TCT3	P52352	Phenuiviridae
Q0GP58	P52344	P22026
P0DSQ1	Q69518	Rhabdoviridae
Q8V556	Q69514	D2CRM8
Q6TVM4	P16738	
Q6TUQ8	A0A2N9DYY9	
Q98291		

Figure B.4.8. Top HITS cNLSs and the viral family they belong to. Based on cNLS mapper score, several viral proteins were identified as potentially nuclear. We decided to investigate the proteins with cNLSmapper score more than 9, that are more likely to have a nuclear localization. 26 viral proteins were identified and surprisingly most of them belong to Poxviridae family. The second most represented family was the Herpesviridae one. Two potential nuclear proteins were also spotted in Anelloviridae family, whereas one protein belonging to Rhabdoviridae, Pheniviridae and Orthomyxoviridae family, respectively.

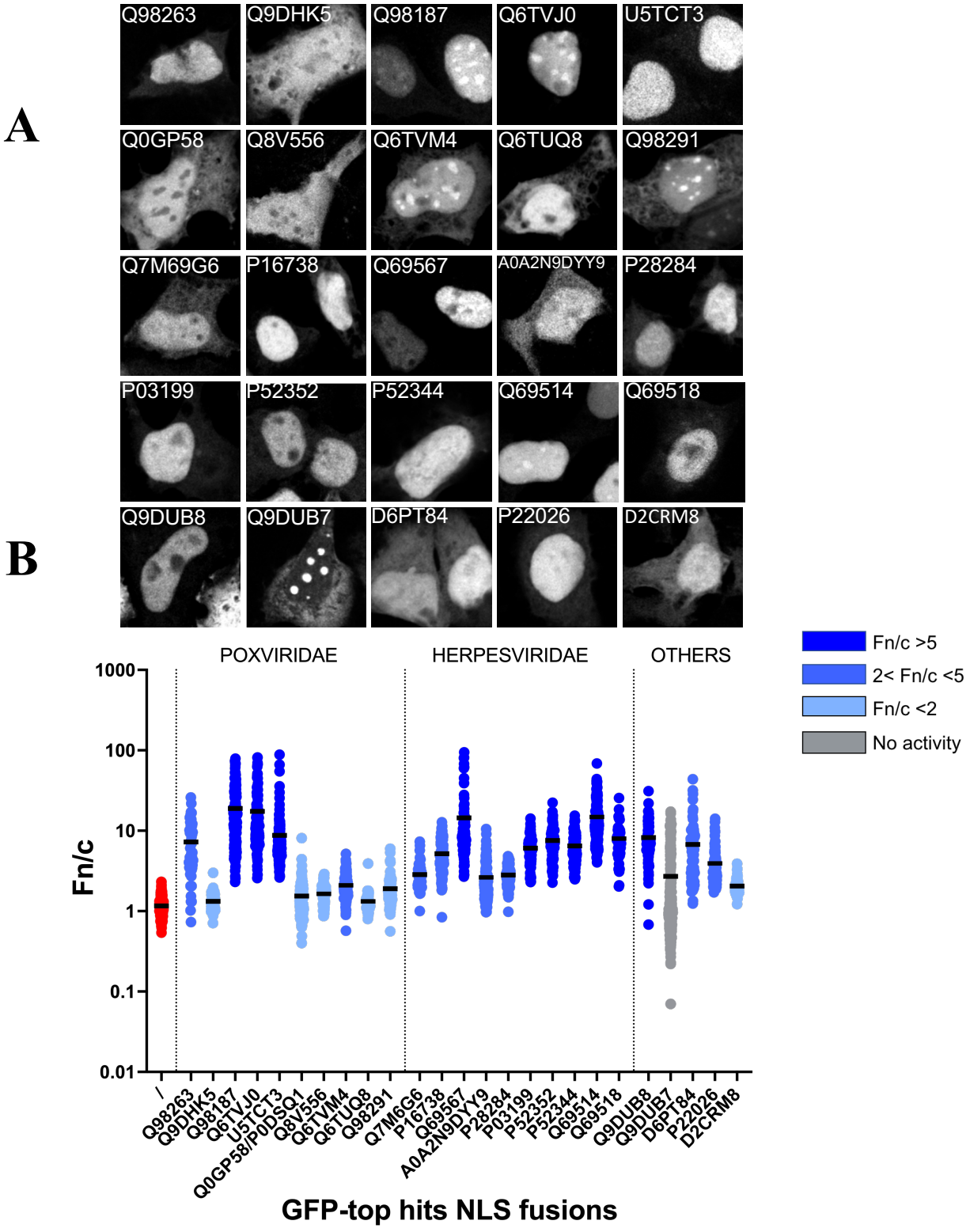


Figure B.4.9. Subcellular localization and quantification of TOP HITS cNLS. Recombinant vectors expressing TOP HITS cNLS were fused to Cycle3-GFP and used to transfect HEK293A cells. 48h.p.t. cells were fixed and mounted on coverslips and the subcellular localization of Cycle3-GPF fused proteins were observed by CLSM (A) and quantified (B). Microscopy analyses revealed that almost all the fused proteins accumulate in the host cell nucleus underling that the cNLSs are functionally active compared to Cycle3-GFP localization alone. These observations were confirmed also by quantification of their nuclear accumulation (B). Interestingly only one fused protein (GFP-Q9DUB7)

failed to accumulate in the nucleus whereas its localization was distributed within the nucleoli and the cytoplasm. Nucleolar accumulation was spotted also for other proteins.

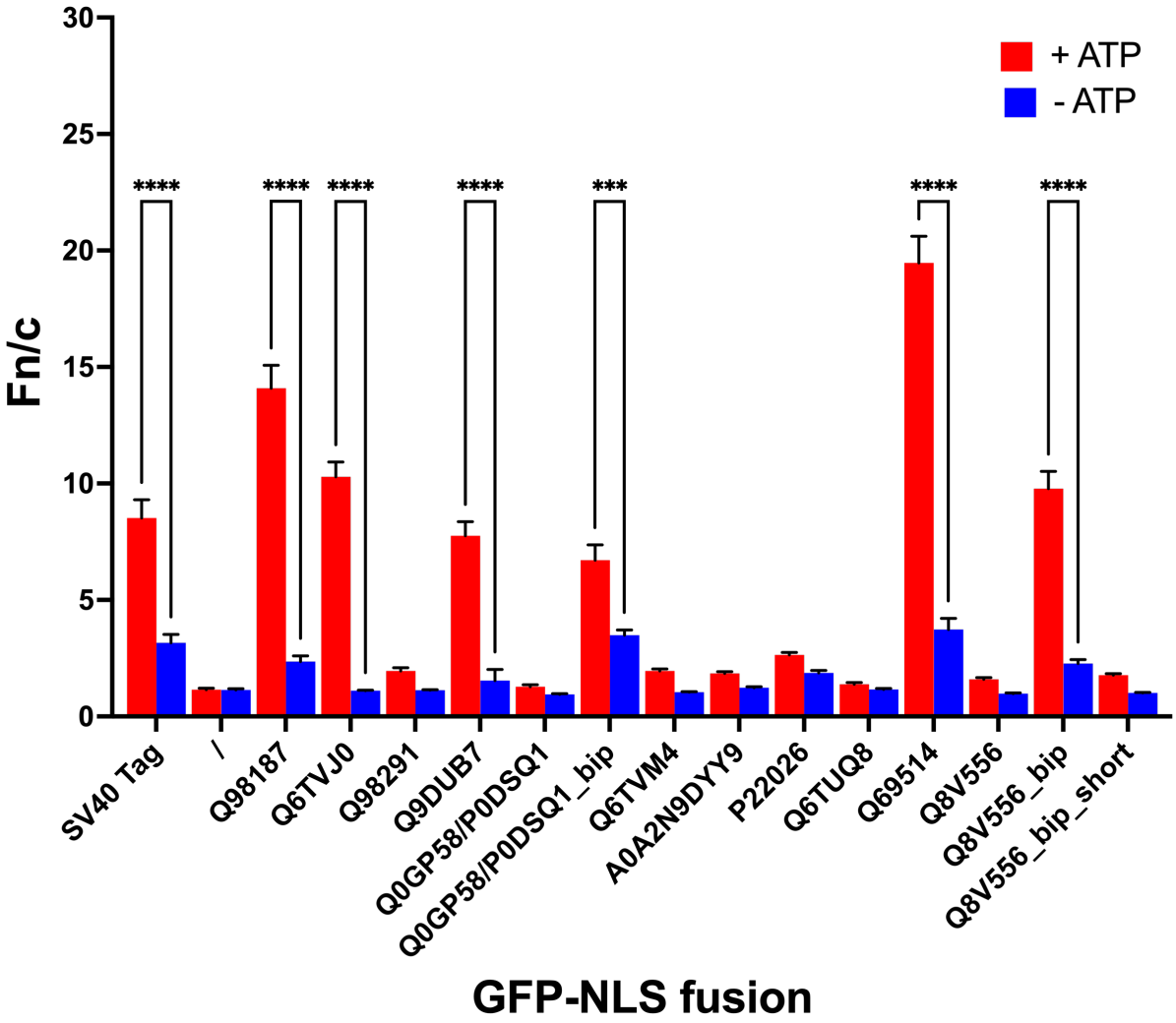


Figure B.4.10. Evaluation of active transport of TOP HITS cNLS by means of ATP depletion. To confirm that these Cycle3-GFP fused proteins are actively transported in the nucleus and that these proteins don't diffuse passively through the NPC, one hour before cell fixation, we treated the cell with energy depletion media. Cells were then fixed and mounted, and their subcellular localization was evaluated with CLSM and quantified. The red bars are relative to nuclear accumulation in normal energy conditions whereas the black bars are relative to cells treated with ATP depletion media. As shown by the graph, removing energy from the media significantly decreased nuclear accumulation of Cycle3-GFP fused proteins, confirming that their nuclear localization is due to active transport

IMPa/b mediated and not by passive diffusion. Mean, STD and Student-T test were calculated on mean fluorescent signals.

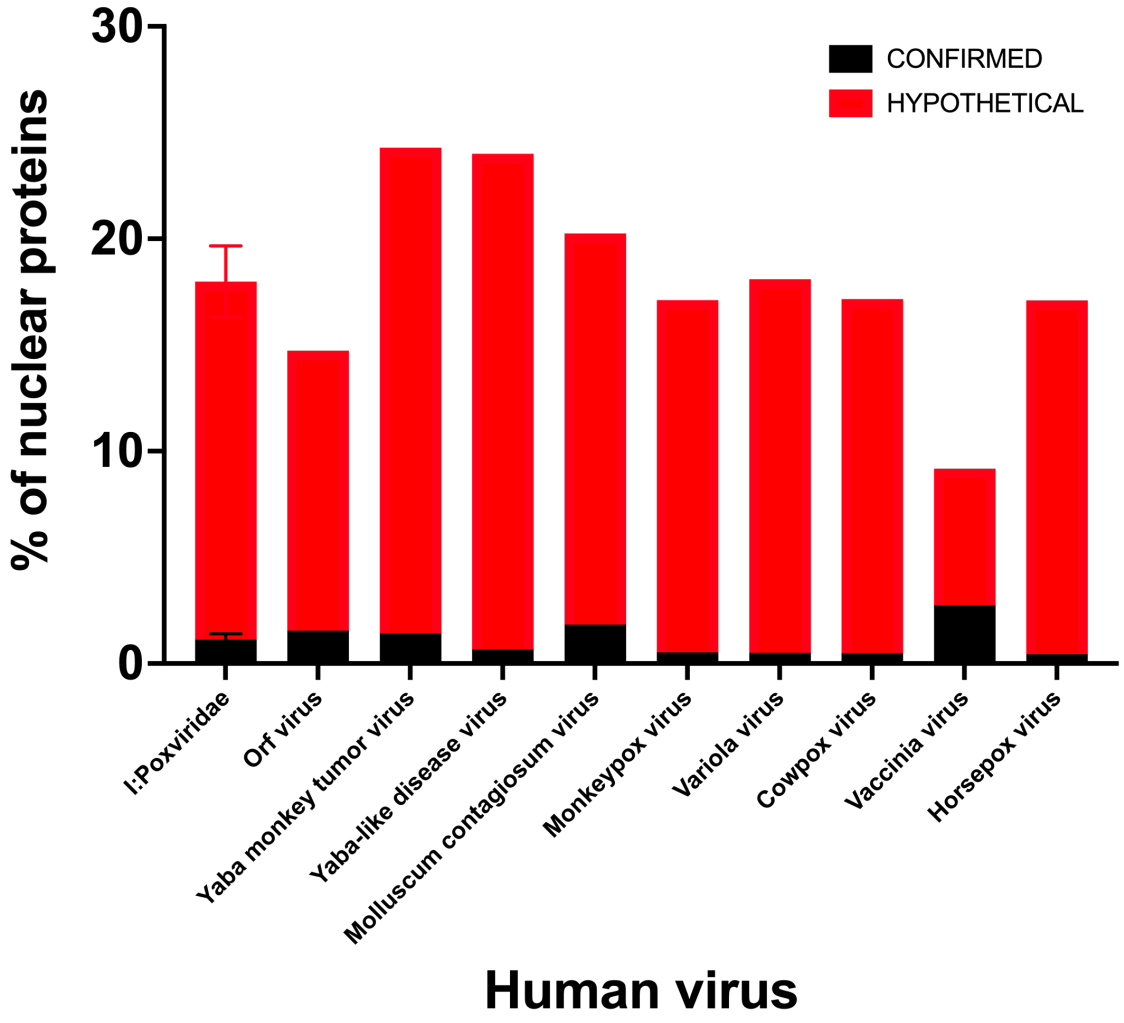


Figure B.4.11. Confirmed and Hypothetical nuclear proteins in Poxviridae family. Poxviruses are large cytoplasmic replicating viruses, and a great percentage of their nuclear proteins are aimed to interfere with hose cell immune responses. Surprisingly only few nuclear proteins are annotated as nuclear in literature. Our bioinformatic analyses revealed that the percentage of nuclear proteins significantly increases if all the hypothetical nuclear proteins are also considered, suggesting that several proteins are still unknown or are wrongly annotated on Uniprot.

#UniProt_ACC	Protein	Organism	NLS	cNLS mapper	Fn/c
U5TC3_COWPX	CPXV038	CPV	IRKRPNQHHTIDLFKRMKR	10.3	9.6
Q0GP58_HSPV	HSPV032	HSPV/VARV	LFKKIKRTYDTF	9.5	1.6
Q8V556_MONPZ	P2L	MPX	LFKRIKRTYDTF	9.5	1.7
N2_VAR67	N2L, P2L	VARV	LFKKIKRTYDTF	9.5	1.6

Figure B.4.12. cNLS sequence and score of 4 Poxviridae proteins that showed poor nuclear accumulation. While cNLS mapper scored valued like 9.5 for potential nuclear proteins for Q0GP58_HSPV, Q8V556_MONPZ and N2_VAR67, our analyses revealed a very low nuclear accumulation. Instead U5TC3_COWPX nuclear accumulation was as expected. By looking at the sequences, it can be noticed that CPXV038 is annotated as bipartite instead the other, as monopartite.

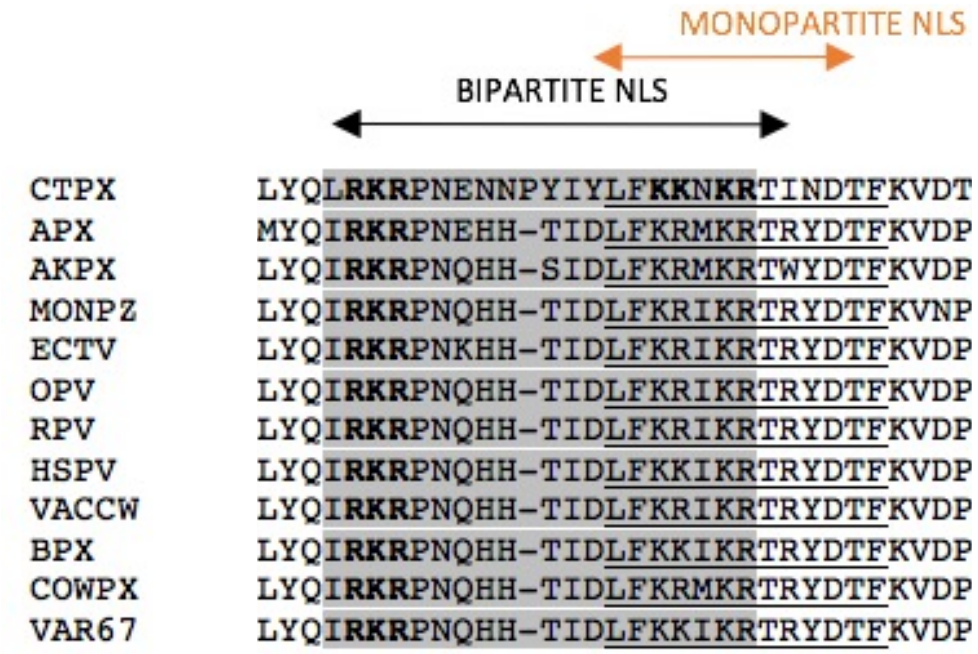
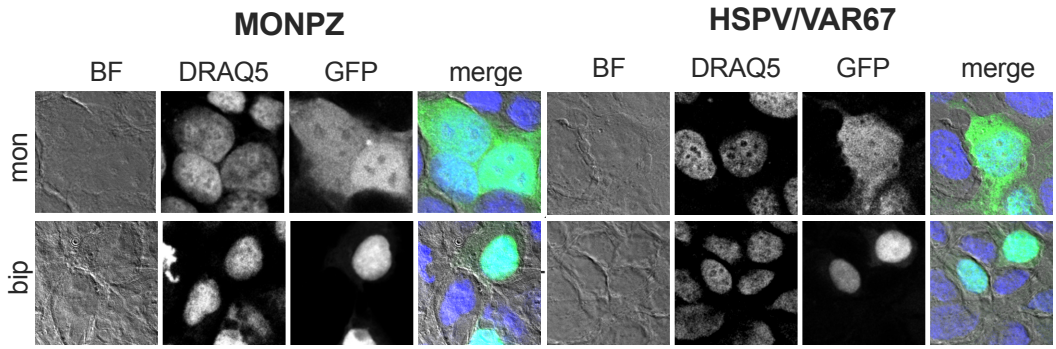


Figure B.4.13 Sequence alignment spotted bipartite cNLS for VACCINIA N2 orthologs. Sequence alignment revealed that all potential nuclear proteins identified in Poxviridae family have a bipartite cNLS sequence overlapping a monopartite one, suggesting that these proteins may have a functional bipartite cNLS rather than monopartite.

A

Uniprot	Name	Sequence	Fn/c
Q8V556	MONPZ mono	LFKRIKRTRYDTF	1.7
Q8V556	MONZ bip	IRKRPNQHHTIDLFKRIKRTRYDTF	10.1
Q0GP58/P0DSQ1	HSPV/VAR67	LFKKIKRTRYDTF	1.6
Q0GP58/P0DSQ1	HSPV/VAR67bip	IRKRPNQHHTIDLFKKIKRT	7.5

B



C

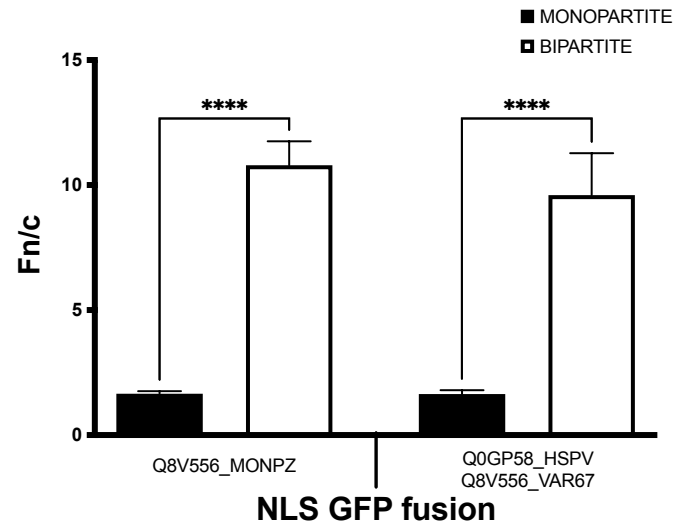


Figure B.4.14. Bipartite cNLS of MONPZ, HSPV and VAR67 proteins are stronger than their monopartite counterpart. As microscopy analysis has shown, the bipartite cNLSs accumulated significantly more in the nucleus compared to the monopartite cNLSs (B. Microscopy observations were confirmed by fluorescent signal quantifications (A, C).

Uniprot	Name	Sequence	Fn/c
Q8V556	MONZ_bip_s	IRKRPNQHHTIDLFKRIK	2.2
Q8V556	MONZ_bip_sNLSmN	IaaaPNQHHITDLFKIRK	1.0
Q8V556	MONZ_bip_sNLSmC	IRKRPNQHHTIDLFaaIa	1.1
Q8V556	MONZ_bip_sNLSmNC	IaaaPNQHHTIDLFaaIa	1.0
U5TCT3	COWPX-NLS	IRKRPNQHHTIDLFKRMKR	7.5
U5TCT3	COWPX-NLS_NLSmN	IaaaPNQHHTIDLFKRMKR	1.0
U5TCT3	COWPX-NLS_NLSmC	IRKRPNQHHTIDLFaaMaa	1.2
U5TCT3	COWPX-NLS_NLSmNC	IaaaPNQHHTIDLFaaMaa	1.1

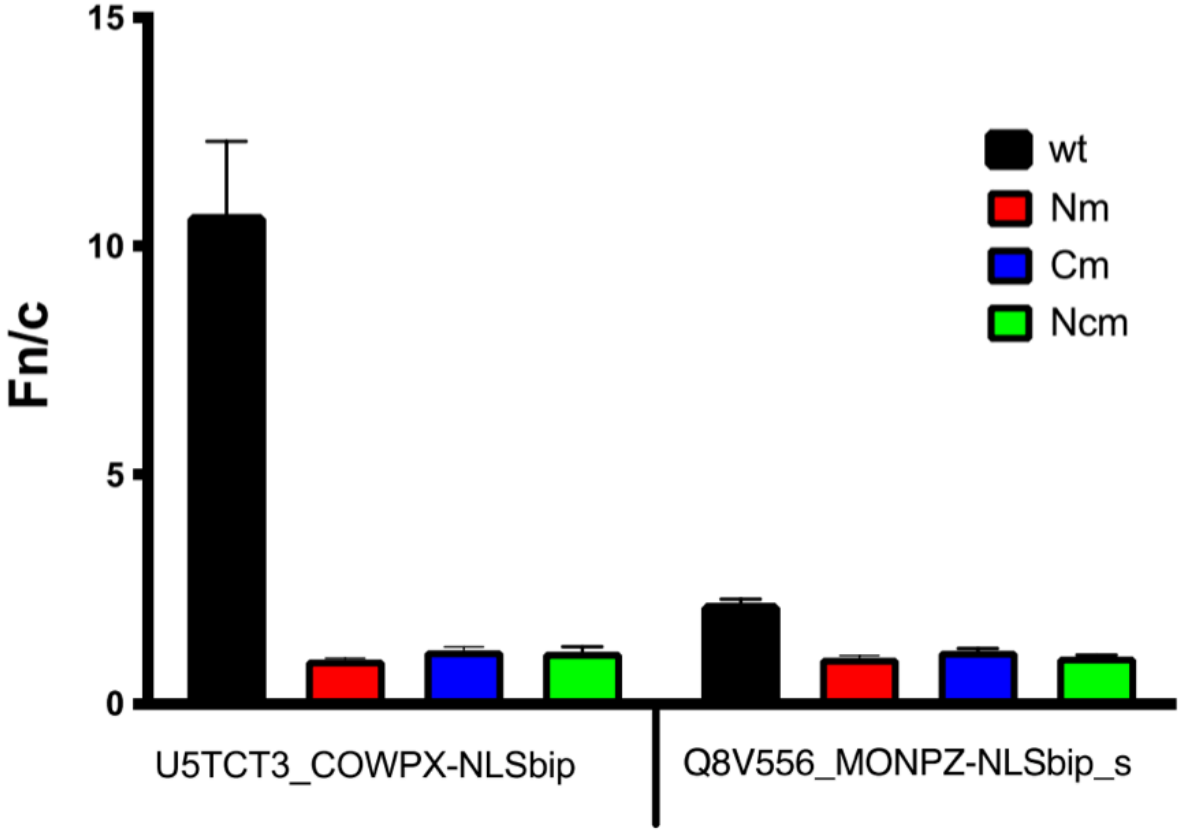


Figure B.4.15. Different nuclear accumulation upon cNLS mutations. Nuclear accumulation and subcellular localization of Cycle3-GFP fused to sequences bearing mutations on the C- and N-terminal of the bipartite cNLSs of U5TC3_COWPX and Q8V556_MONPZ were analyzed. Indeed, only wild type bipartite cNLS is functional.

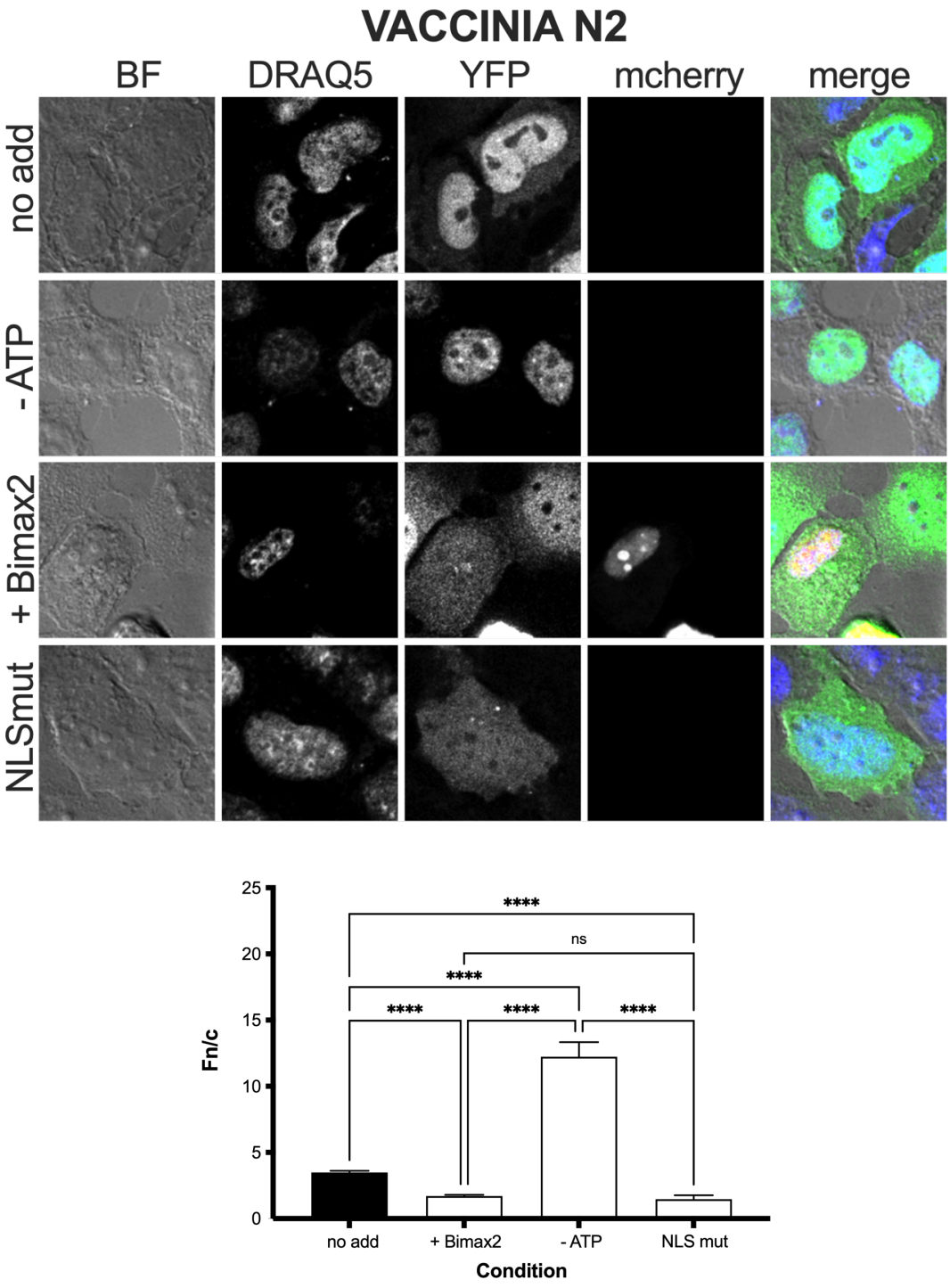


Figure B.4.16 VACCINIA N2 nuclear localization. VACCINIA N2 cNLS was fused in frame with Cycle3-GFP and used to transfect HEK293A cells. It accumulates in the nucleus (no add) yet treatment with Bimax2, an IMPa/b inhibitor, impairs its nuclear localization. Same results is obtained if upstream

basic sequences of the bipartite cNLS are mutated. Surprisingly, the protein localizes strongly in the nucleus upon ATP depletion treatment.

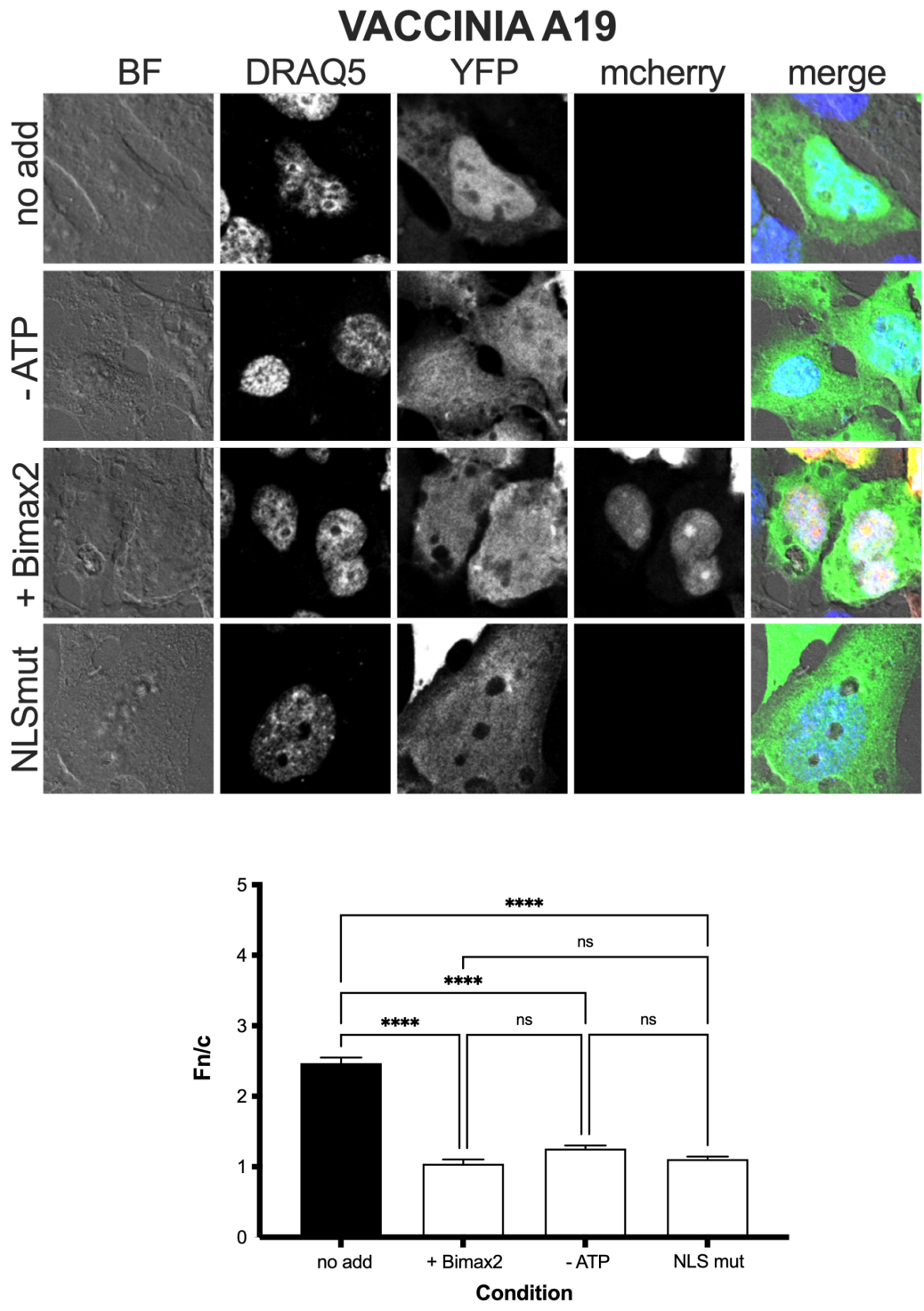
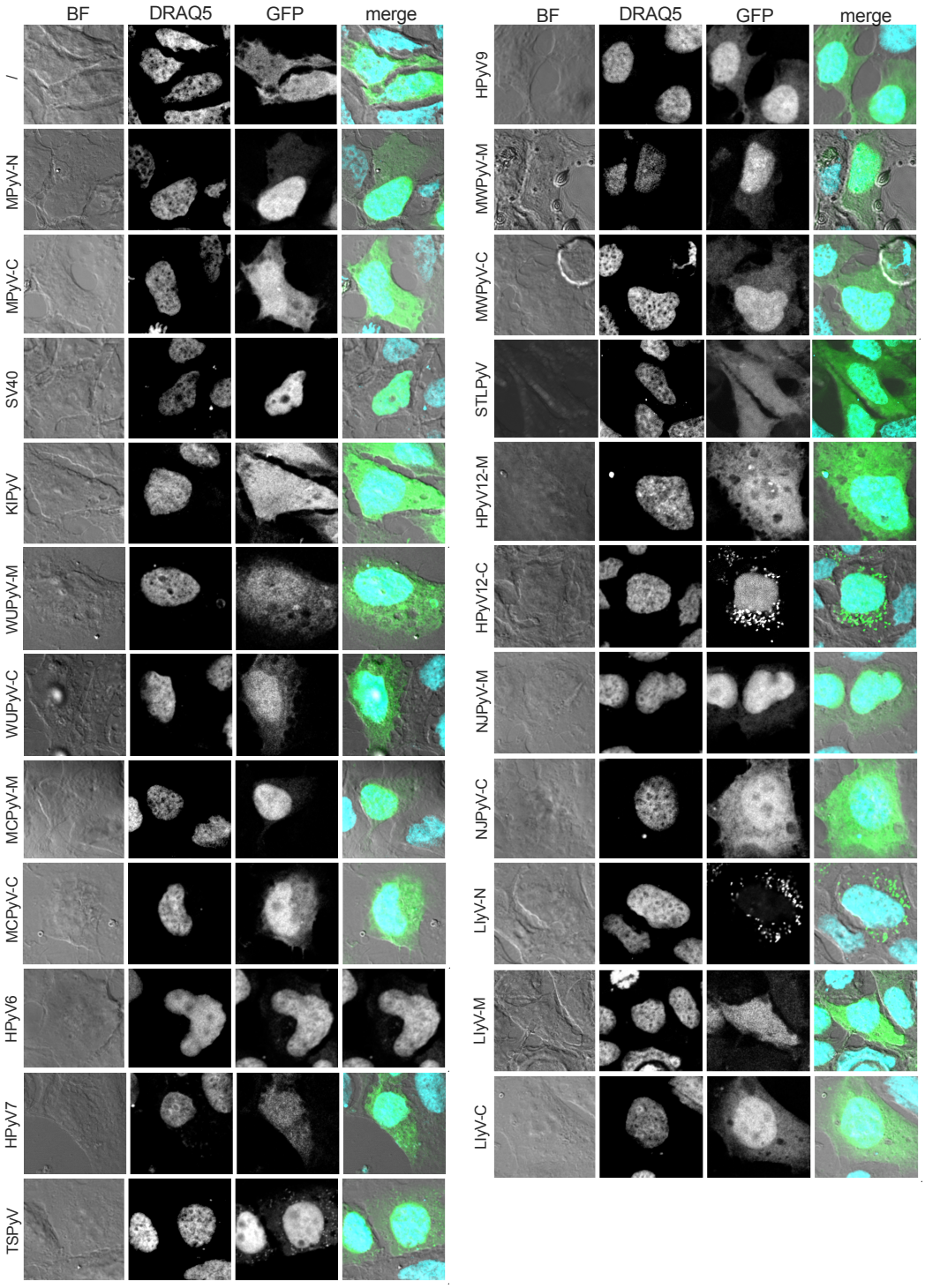


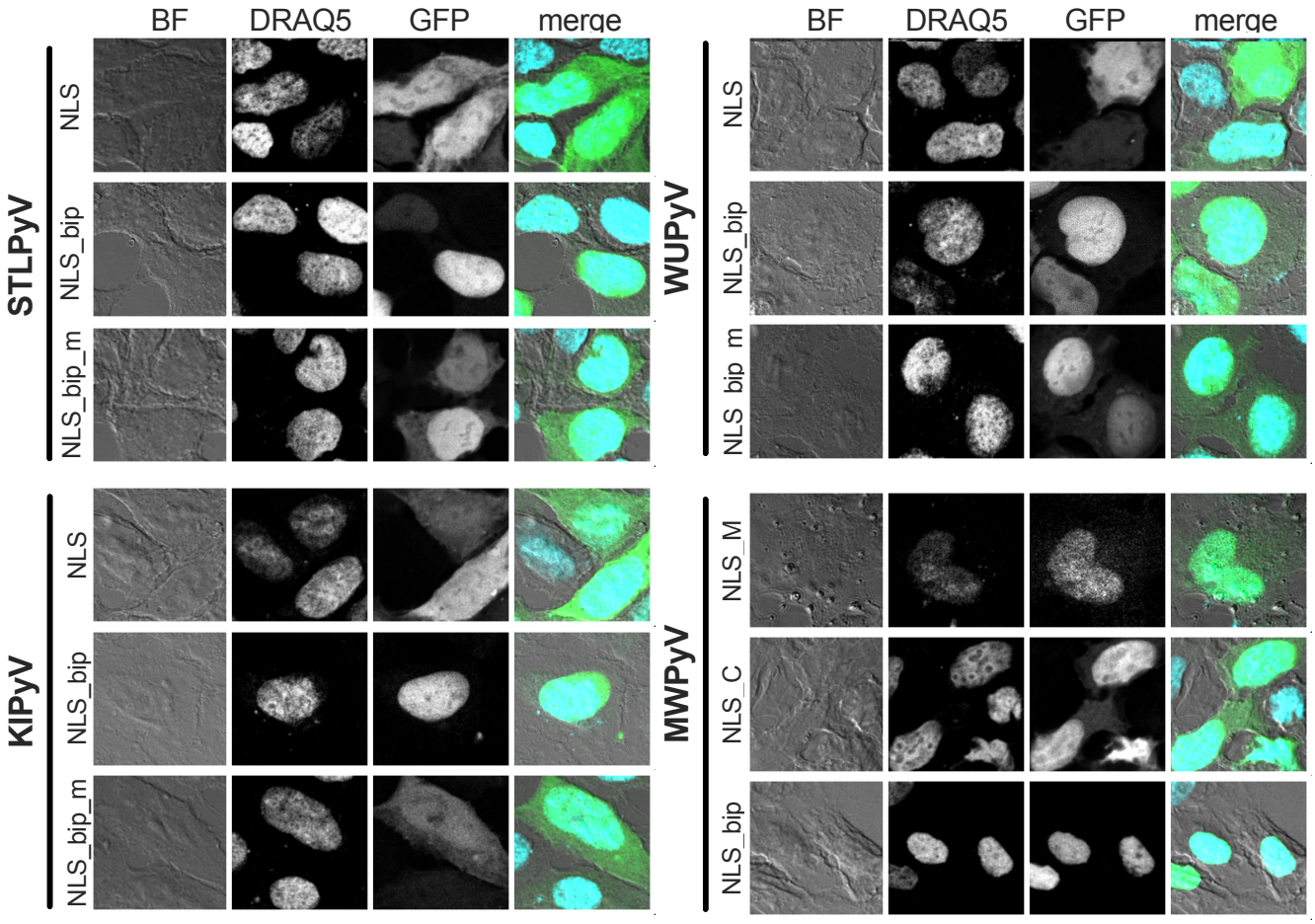
Figure B.4.17. VACCINIA A19 nuclear localization VACCINIA A19 cNLS was fused in frame with Cycle3-GFP and used to transfect HEK293A cells. It accumulates in the nucleus (no add) yet treatment with Bimax2, an IMPa/b inhibitor, impairs its nuclear localization. Same results is obtained if basic

sequences of the monopartite cNLS are mutated. Surprisingly, the protein localizes strongly in the nucleus upon ATP depletion treatment

Supplementary Figures



Supplementary Fig. B.1 Subcellular localization and quantification of all cNLSs present on Human infecting Polyomaviruses LTs. Panel A shows the subcellular localization of Cycle3-GFP fused putative cNLSs. HEK293A cells were seeded and transfected with recombinant vector expressing cNLSs fused in frame with Cycle3-GFP. 48hrs post transfection cells were fixed, mounted on coverslips and the proteins subcellular localization was observed by CLSM.



Supplementary Figure B.2 Subcellular localization and quantification of wildtype and mutant bipartite cNLS of STL, KI, WU and MWVPyVs. Recombinant vectors expressing either bipartite cNLS, mutant bipartite cNLS or monopartite cNLS of STL, KI, Wu and MWVPyV were fused to Cycle3-GFP and used to transfect HEK293A cells. 48h.p.t. cells were fixed and mounted on coverslips and the subcellular localization of Cycle3-GPF fused proteins were observed by CLSM.

References

Aduma, P. *et al.* (1995) 'Metabolic diversity and antiviral activities of acyclic nucleoside phosphonates', *Molecular Pharmacology*, 47(4), pp. 816–822.

Alvisi, G. *et al.* (2005) 'A protein kinase CK2 site flanking the nuclear targeting signal enhances nuclear transport of human cytomegalovirus ppUL44', *Traffic (Copenhagen, Denmark)*, 6(11), pp. 1002–1013. doi:10.1111/j.1600-0854.2005.00331.x.

Alvisi, G. *et al.* (2006) 'Human cytomegalovirus DNA polymerase catalytic subunit pUL54 possesses independently acting nuclear localization and ppUL44 binding motifs', *Traffic (Copenhagen, Denmark)*, 7(10), pp. 1322–1332. doi:10.1111/j.1600-0854.2006.00477.x.

Alvisi, G. *et al.* (2008) 'Nuclear import of HSV-1 DNA polymerase processivity factor UL42 is mediated by a C-terminally located bipartite nuclear localization signal', *Biochemistry*, 47(52), pp. 13764–13777. doi:10.1021/bi800869y.

Alvisi, G. *et al.* (2011) 'Multiple phosphorylation sites at the C-terminus regulate nuclear import of HCMV DNA polymerase processivity factor ppUL44', *Virology*, 417(2), pp. 259–267. doi:10.1016/j.virol.2011.06.015.

Ao, Z. *et al.* (2010) 'Importin alpha3 interacts with HIV-1 integrase and contributes to HIV-1 nuclear import and replication', *Journal of Virology*, 84(17), pp. 8650–8663. doi:10.1128/JVI.00508-10.

Appleton, B.A. *et al.* (2004) 'The cytomegalovirus DNA polymerase subunit UL44 forms a C clamp-shaped dimer', *Molecular Cell*, 15(2), pp. 233–244. doi:10.1016/j.molcel.2004.06.018.

Appleton, B.A. *et al.* (2006) 'Crystal structure of the cytomegalovirus DNA polymerase subunit UL44 in complex with the C terminus from the catalytic subunit. Differences in structure and function relative to unliganded UL44', *The Journal of Biological Chemistry*, 281(8), pp. 5224–5232. doi:10.1074/jbc.M506900200.

Arnold, M. *et al.* (2006) 'Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1', *The Journal of Biological Chemistry*, 281(30), pp. 20883–20890. doi:10.1074/jbc.M602189200.

Atabani, S.F. *et al.* (2012) 'Cytomegalovirus replication kinetics in solid organ transplant recipients managed by preemptive therapy', *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 12(9), pp. 2457–2464. doi:10.1111/j.1600-6143.2012.04087.x.

Atkinson, S.C. *et al.* (2018) 'Recognition by host nuclear transport proteins drives disorder-to-order transition in Hendra virus V', *Scientific Reports*, 8(1), p. 358. doi:10.1038/s41598-017-18742-8.

AuCoin, D.P. *et al.* (2006) 'Betaherpesvirus-conserved cytomegalovirus tegument protein ppUL32 (pp150) controls cytoplasmic events during virion maturation', *Journal of Virology*, 80(16), pp. 8199–8210. doi:10.1128/JVI.00457-06.

Bayliss, R. *et al.* (2002) 'Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats', *The EMBO journal*, 21(12), pp. 2843–2853. doi:10.1093/emboj/cdf305.

Bian, X.-L. and Wilson, V.G. (2010) 'Common importin alpha specificity for papillomavirus E2 proteins', *Virus Research*, 150(1–2), pp. 135–137. doi:10.1016/j.virusres.2010.02.011.

Bullock, T.L. *et al.* (1996) 'The 1.6 angstroms resolution crystal structure of nuclear transport factor 2 (NTF2)', *Journal of Molecular Biology*, 260(3), pp. 422–431. doi:10.1006/jmbi.1996.0411.

Burns, L.T. and Wente, S.R. (2012) 'Trafficking to uncharted territory of the nuclear envelope', *Current Opinion in Cell Biology*, 24(3), pp. 341–349. doi:10.1016/j.ceb.2012.01.009.

Chambers, J. *et al.* (1999) 'DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression', *Journal of Virology*, 73(7), pp. 5757–5766. doi:10.1128/JVI.73.7.5757-5766.1999.

Chang, C.-W. *et al.* (2012) 'Epstein-Barr virus protein kinase BGLF4 targets the nucleus through interaction with nucleoporins', *Journal of Virology*, 86(15), pp. 8072–8085. doi:10.1128/JVI.01058-12.

Chen, C. *et al.* (2016) 'Importin β Can Bind Hepatitis B Virus Core Protein and Empty Core-Like Particles and Induce Structural Changes', *PLoS pathogens*, 12(8), p. e1005802. doi:10.1371/journal.ppat.1005802.

Chen, H. *et al.* (2017) 'A Small Covalent Allosteric Inhibitor of Human Cytomegalovirus DNA Polymerase Subunit Interactions', *ACS infectious diseases*, 3(2), pp. 112–118. doi:10.1021/acsinfecdis.6b00079.

Chen, M.J. and Creinin, M.D. (2015) 'Mifepristone With Buccal Misoprostol for Medical Abortion: A Systematic Review', *Obstetrics and Gynecology*, 126(1), pp. 12–21. doi:10.1097/AOG.0000000000000897.

Chook, Y.M. and Blobel, G. (1999) ‘Structure of the nuclear transport complex karyopherin-β2–Ran˙GppNHp’, *Nature*, 399(6733), pp. 230–237. doi:10.1038/20375.

Cohen, M.J. *et al.* (2014) ‘Functional analysis of the C-terminal region of human adenovirus E1A reveals a misidentified nuclear localization signal’, *Virology*, 468–470, pp. 238–243. doi:10.1016/j.virol.2014.08.014.

Collins-McMillen, D. *et al.* (2019) ‘Alternative promoters drive human cytomegalovirus reactivation from latency’, *Proceedings of the National Academy of Sciences of the United States of America*, 116(35), pp. 17492–17497. doi:10.1073/pnas.1900783116.

Compton, T., Nepomuceno, R.R. and Nowlin, D.M. (1992) ‘Human cytomegalovirus penetrates host cells by pH-independent fusion at the cell surface’, *Virology*, 191(1), pp. 387–395. doi:10.1016/0042-6822(92)90200-9.

Conti, E. *et al.* (1998) ‘Crystallographic Analysis of the Recognition of a Nuclear Localization Signal by the Nuclear Import Factor Karyopherin α’, *Cell*, 94(2), pp. 193–204. doi:10.1016/S0092-8674(00)81419-1.

Conti, E., Müller, C.W. and Stewart, M. (2006) ‘Karyopherin flexibility in nucleocytoplasmic transport’, *Current Opinion in Structural Biology*, 16(2), pp. 237–244. doi:10.1016/j.sbi.2006.03.010.

Copeland, A.M., Newcomb, W.W. and Brown, J.C. (2009) ‘Herpes simplex virus replication: roles of viral proteins and nucleoporins in capsid-nucleus attachment’, *Journal of Virology*, 83(4), pp. 1660–1668. doi:10.1128/JVI.01139-08.

Craig, J.M. *et al.* (1957) ‘Isolation of intranuclear inclusion producing agents from infants with illnesses resembling cytomegalic inclusion disease’, *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, 94(1), pp. 4–12. doi:10.3181/00379727-94-22841.

Davey, N.E., Cyert, M.S. and Moses, A.M. (2015) ‘Short linear motifs - ex nihilo evolution of protein regulation’, *Cell communication and signaling: CCS*, 13, p. 43. doi:10.1186/s12964-015-0120-z.

Davison, A.J. *et al.* (2009) ‘The order Herpesvirales’, *Archives of Virology*, 154(1), pp. 171–177. doi:10.1007/s00705-008-0278-4.

Di Antonio, V., Palù, G. and Alvisi, G. (2021a) ‘Live-Cell Analysis of Human Cytomegalovirus DNA Polymerase Holoenzyme Assembly by Resonance Energy Transfer Methods’, *Microorganisms*, 9(5), p. 928. doi:10.3390/microorganisms9050928.

Di Antonio, V., Palù, G. and Alvisi, G. (2021b) ‘Live-Cell Analysis of Human

Cytomegalovirus DNA Polymerase Holoenzyme Assembly by Resonance Energy Transfer Methods', *Microorganisms*, 9(5), p. 928. doi:10.3390/microorganisms9050928.

Dingwall, C. and Laskey, R.A. (1991) 'Nuclear targeting sequences — a consensus?', *Trends in Biochemical Sciences*, 16, pp. 478–481. doi:10.1016/0968-0004(91)90184-W.

Dollard, S.C., Grosse, S.D. and Ross, D.S. (2007) 'New estimates of the prevalence of neurological and sensory sequelae and mortality associated with congenital cytomegalovirus infection', *Reviews in Medical Virology*, 17(5), pp. 355–363. doi:10.1002/rmv.544.

Eberhard, J., Onder, Z. and Moroianu, J. (2013) 'Nuclear import of high risk HPV16 E7 oncoprotein is mediated by its zinc-binding domain via hydrophobic interactions with Nup62', *Virology*, 446(1–2), pp. 334–345. doi:10.1016/j.virol.2013.08.017.

Elbadawy, H.M. *et al.* (2020) 'Generation of Combinatorial Lentiviral Vectors Expressing Multiple Anti-Hepatitis C Virus shRNAs and Their Validation on a Novel HCV Replicon Double Reporter Cell Line', *Viruses*, 12(9), p. E1044. doi:10.3390/v12091044.

Erice, A. (1999) 'Resistance of human cytomegalovirus to antiviral drugs', *Clinical Microbiology Reviews*, 12(2), pp. 286–297. doi:10.1128/CMR.12.2.286.

Ertl, P.F. and Powell, K.L. (1992) 'Physical and functional interaction of human cytomegalovirus DNA polymerase and its accessory protein (ICP36) expressed in insect cells', *Journal of Virology*, 66(7), pp. 4126–4133. doi:10.1128/JVI.66.7.4126-4133.1992.

Ferguson, B.J. *et al.* (2013) 'Vaccinia virus protein N2 is a nuclear IRF3 inhibitor that promotes virulence', *The Journal of General Virology*, 94(Pt 9), pp. 2070–2081. doi:10.1099/vir.0.054114-0.

Fontes, M.R.M., Teh, T. and Kobe, B. (2000) 'Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-α11Edited by K. Nagai', *Journal of Molecular Biology*, 297(5), pp. 1183–1194. doi:10.1006/jmbi.2000.3642.

Fortunato, E.A. *et al.* (2000) 'Exploitation of cellular signaling and regulatory pathways by human cytomegalovirus', *Trends in Microbiology*, 8(3), pp. 111–119. doi:10.1016/s0966-842x(00)01699-1.

Fortunato, E.A. and Spector, D.H. (1999) 'Regulation of human cytomegalovirus gene expression', *Advances in Virus Research*, 54, pp. 61–128. doi:10.1016/s0065-3527(08)60366-8.

Fouchier, R.A. *et al.* (1998) 'Interaction of the human immunodeficiency virus type 1 Vpr protein with the nuclear pore complex', *Journal of Virology*, 72(7), pp. 6004–6013. doi:10.1128/JVI.72.7.6004-6013.1998.

Frange, P. and Leruez-Ville, M. (2018) 'Maribavir, brincidofovir and letermovir: Efficacy and safety of new antiviral drugs for treating cytomegalovirus infections', *Medecine Et Maladies Infectieuses*, 48(8), pp. 495–502. doi:10.1016/j.medmal.2018.03.006.

Geary, R.S., Henry, S.P. and Grillone, L.R. (2002) 'Fomivirsen: clinical pharmacology and potential drug interactions', *Clinical Pharmacokinetics*, 41(4), pp. 255–260. doi:10.2165/00003088-200241040-00002.

Ghassabian, H. *et al.* (2021a) 'Divide et impera: An In Silico Screening Targeting HCMV ppUL44 Processivity Factor Homodimerization Identifies Small Molecules Inhibiting Viral Replication', *Viruses*, 13(5), p. 941. doi:10.3390/v13050941.

Ghassabian, H. *et al.* (2021b) 'Divide et impera: An In Silico Screening Targeting HCMV ppUL44 Processivity Factor Homodimerization Identifies Small Molecules Inhibiting Viral Replication', *Viruses*, 13(5), p. 941. doi:10.3390/v13050941.

Ghassabian, H. *et al.* (2021c) 'Divide et impera: An In Silico Screening Targeting HCMV ppUL44 Processivity Factor Homodimerization Identifies Small Molecules Inhibiting Viral Replication', *Viruses*, 13(5), p. 941. doi:10.3390/v13050941.

Gilbert, C. and Boivin, G. (2005) 'Human cytomegalovirus resistance to antiviral drugs', *Antimicrobial Agents and Chemotherapy*, 49(3), pp. 873–883. doi:10.1128/AAC.49.3.873-883.2005.

González Canga, A. *et al.* (2008) 'The pharmacokinetics and interactions of ivermectin in humans--a mini-review', *The AAPS journal*, 10(1), pp. 42–46. doi:10.1208/s12248-007-9000-9.

Gronewold, A., Horn, M. and Neundorf, I. (2018) 'Design and biological characterization of novel cell-penetrating peptides preferentially targeting cell nuclei and subnuclear regions', *Beilstein Journal of Organic Chemistry*, 14, pp. 1378–1388. doi:10.3762/bjoc.14.116.

Grundy, J.E. *et al.* (1988) 'Symptomatic cytomegalovirus infection in seropositive kidney recipients: reinfection with donor virus rather than reactivation of recipient virus', *Lancet (London, England)*, 2(8603), pp. 132–135. doi:10.1016/s0140-6736(88)90685-x.

Hahn, F. *et al.* (2018) 'Novel cytomegalovirus-inhibitory compounds of the class pyrrolopyridines show a complex pattern of target binding that suggests an unusual

mechanism of antiviral activity', *Antiviral Research*, 159, pp. 84–94. doi:10.1016/j.antiviral.2018.09.012.

He, R. *et al.* (2013) 'Unique and highly selective anticytomegalovirus activities of artemisinin-derived dimer diphenyl phosphate stem from combination of dimer unit and a diphenyl phosphate moiety', *Antimicrobial Agents and Chemotherapy*, 57(9), pp. 4208–4214. doi:10.1128/AAC.00893-13.

Hintersteiner, M. *et al.* (2010) 'Identification of a small molecule inhibitor of importin β mediated nuclear import by confocal on-bead screening of tagged one-bead one-compound libraries', *ACS chemical biology*, 5(10), pp. 967–979. doi:10.1021/cb100094k.
Ho, M. (2008) 'The history of cytomegalovirus and its diseases', *Medical Microbiology and Immunology*, 197(2), pp. 65–73. doi:10.1007/s00430-007-0066-x.

Hodel, A.E. *et al.* (2006) 'Nuclear Localization Signal Receptor Affinity Correlates with in Vivo Localization in Saccharomyces cerevisiae', *Journal of Biological Chemistry*, 281(33), pp. 23545–23556. doi:10.1074/jbc.M601718200.

Hodel, M.R., Corbett, A.H. and Hodel, A.E. (2001) 'Dissection of a Nuclear Localization Signal', *Journal of Biological Chemistry*, 276(2), pp. 1317–1325. doi:10.1074/jbc.M008522200.

Hurley, E.A. and Thorley-Lawson, D.A. (1988) 'B cell activation and the establishment of Epstein-Barr virus latency', *The Journal of Experimental Medicine*, 168(6), pp. 2059–2075. doi:10.1084/jem.168.6.2059.

Isaacson, M.K. and Compton, T. (2009) 'Human cytomegalovirus glycoprotein B is required for virus entry and cell-to-cell spread but not for virion attachment, assembly, or egress', *Journal of Virology*, 83(8), pp. 3891–3903. doi:10.1128/JVI.01251-08.

Isaacson, M.K., Juckem, L.K. and Compton, T. (2008) 'Virus entry and innate immune activation', *Current Topics in Microbiology and Immunology*, 325, pp. 85–100. doi:10.1007/978-3-540-77349-8_5.

Jacobs, N. *et al.* (2008) 'Vaccinia virus lacking the Bcl-2-like protein N1 induces a stronger natural killer cell response to infection', *The Journal of General Virology*, 89(Pt 11), pp. 2877–2881. doi:10.1099/vir.0.2008/004119-0.

Jans, D.A. and Wagstaff, K.M. (2020) 'Ivermectin as a Broad-Spectrum Host-Directed Antiviral: The Real Deal?', *Cells*, 9(9), p. E2100. doi:10.3390/cells9092100.

Jenkins, Y. *et al.* (1998) 'Characterization of HIV-1 vpr nuclear import: analysis of signals and pathways', *The Journal of Cell Biology*, 143(4), pp. 875–885. doi:10.1083/jcb.143.4.875.

Kabachinski, G. and Schwartz, T.U. (2015) 'The nuclear pore complex--structure and function at a glance', *Journal of Cell Science*, 128(3), pp. 423–429. doi:10.1242/jcs.083246.

Kalderon, D. and Smith, A.E. (1984) 'In vitro mutagenesis of a putative DNA binding domain of SV40 large-T', *Virology*, 139(1), pp. 109–137. doi:10.1016/0042-6822(84)90334-9.

Kalejta, R.F., Bechtel, J.T. and Shenk, T. (2003) 'Human cytomegalovirus pp71 stimulates cell cycle progression by inducing the proteasome-dependent degradation of the retinoblastoma family of tumor suppressors', *Molecular and Cellular Biology*, 23(6), pp. 1885–1895. doi:10.1128/MCB.23.6.1885-1895.2003.

Katta, S.S., Smoyer, C.J. and Jaspersen, S.L. (2014) 'Destination: inner nuclear membrane', *Trends in Cell Biology*, 24(4), pp. 221–229. doi:10.1016/j.tcb.2013.10.006.

Kinzler, E.R., Theiler, R.N. and Compton, T. (2002) 'Expression and reconstitution of the gH/gL/gO complex of human cytomegalovirus', *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, 25 Suppl 2, pp. S87-95. doi:10.1016/s1386-6532(02)00098-7.

Köhler, M. *et al.* (2001) 'Adenoviral E1A protein nuclear import is preferentially mediated by importin alpha3 in vitro', *Virology*, 289(2), pp. 186–191. doi:10.1006/viro.2001.1151.

Komazin-Meredith, G. *et al.* (2008a) 'The human cytomegalovirus UL44 C clamp wraps around DNA', *Structure (London, England: 1993)*, 16(8), pp. 1214–1225. doi:10.1016/j.str.2008.05.008.

Komazin-Meredith, G. *et al.* (2008b) 'The human cytomegalovirus UL44 C clamp wraps around DNA', *Structure (London, England: 1993)*, 16(8), pp. 1214–1225. doi:10.1016/j.str.2008.05.008.

Kosugi, S. *et al.* (2008) 'Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling', *Chemistry & Biology*, 15(9), pp. 940–949. doi:10.1016/j.chembiol.2008.07.019.

Kosugi, Shunichi, Hasebe, M., Matsumura, N., *et al.* (2009) 'Six classes of nuclear localization signals specific to different binding grooves of importin alpha', *The Journal of Biological Chemistry*, 284(1), pp. 478–485. doi:10.1074/jbc.M807017200.

Kosugi, S. *et al.* (2009) 'Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs', *Proceedings of the National Academy of Sciences*, 106(25), pp. 10171–10176.

doi:10.1073/pnas.0900604106.

Kosugi, Shunichi, Hasebe, M., Tomita, M., *et al.* (2009) 'Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs', *Proceedings of the National Academy of Sciences of the United States of America*, 106(25), pp. 10171–10176. doi:10.1073/pnas.0900604106.

Kramer, M.F. *et al.* (2003) 'Latent herpes simplex virus infection of sensory neurons alters neuronal gene expression', *Journal of Virology*, 77(17), pp. 9533–9541. doi:10.1128/jvi.77.17.9533-9541.2003.

Krishna, B.A., Wills, M.R. and Sinclair, J.H. (2019) 'Advances in the treatment of cytomegalovirus', *British Medical Bulletin*, 131(1), pp. 5–17. doi:10.1093/bmb/ldz031.
Lange, A. *et al.* (2008) 'A PY-NLS nuclear targeting signal is required for nuclear localization and function of the Saccharomyces cerevisiae mRNA-binding protein Hrp1', *The Journal of Biological Chemistry*, 283(19), pp. 12926–12934. doi:10.1074/jbc.M800898200.

Lea, A.P. and Bryson, H.M. (1996) 'Cidofovir', *Drugs*, 52(2), pp. 225–230; discussion 231. doi:10.2165/00003495-199652020-00006.

Lepri, S. *et al.* (2014) 'Optimization of small-molecule inhibitors of influenza virus polymerase: from thiophene-3-carboxamide to polyamido scaffolds', *Journal of Medicinal Chemistry*, 57(10), pp. 4337–4350. doi:10.1021/jm500300r.

Ligat, G. *et al.* (2018) 'The human cytomegalovirus terminase complex as an antiviral target: a close-up view', *FEMS microbiology reviews*, 42(2), pp. 137–145. doi:10.1093/femsre/fuy004.

Littler, E., Stuart, A.D. and Chee, M.S. (1992) 'Human cytomegalovirus UL97 open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir', *Nature*, 358(6382), pp. 160–162. doi:10.1038/358160a0.

Lopez-Denman, A.J. *et al.* (2018) 'Nucleocytoplasmic shuttling of the West Nile virus RNA-dependent RNA polymerase NS5 is critical to infection', *Cellular Microbiology*, 20(8), p. e12848. doi:10.1111/cmi.12848.

Loregian, A. *et al.* (2004a) 'Residues of human cytomegalovirus DNA polymerase catalytic subunit UL54 that are necessary and sufficient for interaction with the accessory protein UL44', *Journal of Virology*, 78(1), pp. 158–167. doi:10.1128/jvi.78.1.158-167.2004.

Loregian, A. *et al.* (2004b) 'Specific residues in the connector loop of the human cytomegalovirus DNA polymerase accessory protein UL44 are crucial for interaction

with the UL54 catalytic subunit', *Journal of Virology*, 78(17), pp. 9084–9092. doi:10.1128/JVI.78.17.9084-9092.2004.

Loregian, A. and Coen, D.M. (2006) 'Selective anti-cytomegalovirus compounds discovered by screening for inhibitors of subunit interactions of the viral polymerase', *Chemistry & Biology*, 13(2), pp. 191–200. doi:10.1016/j.chembiol.2005.12.002.

Loregian, A., Marsden, H.S. and Palù, G. (2002) 'Protein-protein interactions as targets for antiviral chemotherapy', *Reviews in Medical Virology*, 12(4), pp. 239–262. doi:10.1002/rmv.356.

Loveland, KateL. *et al.* (2015) 'Putting things in place for fertilization: discovering roles for importin proteins in cell fate and spermatogenesis', *Asian Journal of Andrology*, 17(4), p. 537. doi:10.4103/1008-682X.154310.

Mach, M. *et al.* (2005) 'Complex formation by glycoproteins M and N of human cytomegalovirus: structural and functional aspects', *Journal of Virology*, 79(4), pp. 2160–2170. doi:10.1128/JVI.79.4.2160-2170.2005.

Maluquer de Motes, C. *et al.* (2014) 'Vaccinia virus virulence factor N1 can be ubiquitylated on multiple lysine residues', *The Journal of General Virology*, 95(Pt 9), pp. 2038–2049. doi:10.1099/vir.0.065664-0.

Margolis, M.J. *et al.* (1995) 'Interaction of the 72-kilodalton human cytomegalovirus IE1 gene product with E2F1 coincides with E2F-dependent activation of dihydrofolate reductase transcription', *Journal of Virology*, 69(12), pp. 7759–7767. doi:10.1128/JVI.69.12.7759-7767.1995.

Markham, A. and Faulds, D. (1994) 'Ganciclovir. An update of its therapeutic use in cytomegalovirus infection', *Drugs*, 48(3), pp. 455–484. doi:10.2165/00003495-199448030-00009.

Marschall, M. *et al.* (2000) 'Recombinant green fluorescent protein-expressing human cytomegalovirus as a tool for screening antiviral agents', *Antimicrobial Agents and Chemotherapy*, 44(6), pp. 1588–1597. doi:10.1128/AAC.44.6.1588-1597.2000.

Matreyek, K.A. *et al.* (2013) 'Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity', *PLoS pathogens*, 9(10), p. e1003693. doi:10.1371/journal.ppat.1003693.

Mercorelli, B. *et al.* (2011) 'Early inhibitors of human cytomegalovirus: state-of-art and therapeutic perspectives', *Pharmacology & Therapeutics*, 131(3), pp. 309–329. doi:10.1016/j.pharmthera.2011.04.007.

Mercorelli, B. *et al.* (2018) 'Repurposing the clinically approved calcium antagonist manidipine dihydrochloride as a new early inhibitor of human cytomegalovirus targeting the Immediate-Early 2 (IE2) protein', *Antiviral Research*, 150, pp. 130–136. doi:10.1016/j.antiviral.2017.12.014.

Mercorelli, B. *et al.* (2020) 'The Clinically Approved Antifungal Drug Posaconazole Inhibits Human Cytomegalovirus Replication', *Antimicrobial Agents and Chemotherapy*, 64(10), pp. e00056-20. doi:10.1128/AAC.00056-20.

Meyer, H. *et al.* (1988) 'Identification and procaryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus', *Journal of Virology*, 62(7), pp. 2243–2250. doi:10.1128/JVI.62.7.2243-2250.1988.

Morgan, J.R. and Roberts, B.E. (1984) 'Organization of RNA transcripts from a vaccinia virus early gene cluster', *Journal of Virology*, 51(2), pp. 283–297. doi:10.1128/JVI.51.2.283-297.1984.

Moroianu, J. (1999) 'Nuclear import and export pathways', *Journal of Cellular Biochemistry*, Suppl 32-33, pp. 76–83. doi:10.1002/(sici)1097-4644(1999)75:32+<76::aid-jcb10>3.3.co;2-h.

Murphy, E.A. *et al.* (2000) 'The human cytomegalovirus IE86 protein can block cell cycle progression after inducing transition into the S phase of permissive cells', *Journal of Virology*, 74(15), pp. 7108–7118. doi:10.1128/jvi.74.15.7108-7118.2000.

Naesens, L. and De Clercq, E. (2001) 'Recent developments in herpesvirus therapy', *Herpes: the journal of the IHMF*, 8(1), pp. 12–16.

Naing, Z.W. *et al.* (2016) 'Congenital cytomegalovirus infection in pregnancy: a review of prevalence, clinical features, diagnosis and prevention', *The Australian & New Zealand Journal of Obstetrics & Gynaecology*, 56(1), pp. 9–18. doi:10.1111/ajo.12408.
Nakamura, T. and Katano, H. (2009) '[Merkel cell polyomavirus and Merkel cell carcinoma]', *Uirusu*, 59(1), pp. 37–42. doi:10.2222/jsv.59.37.

Nishiyama, Y., Maeno, K. and Yoshida, S. (1983) 'Characterization of human cytomegalovirus-induced DNA polymerase and the associated 3'-to-5', exonuclease', *Virology*, 124(2), pp. 221–231. doi:10.1016/0042-6822(83)90339-2.

Nitahara-Kasahara, Y. *et al.* (2007) 'Novel nuclear import of Vpr promoted by importin alpha is crucial for human immunodeficiency virus type 1 replication in macrophages', *Journal of Virology*, 81(10), pp. 5284–5293. doi:10.1128/JVI.01928-06.

Ojala, P.M. *et al.* (2000) 'Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro', *Molecular and Cellular Biology*, 20(13), pp. 4922–4931. doi:10.1128/MCB.20.13.4922-4931.2000.

Palmeri, D. and Malim, M.H. (1999) 'Importin beta can mediate the nuclear import of an arginine-rich nuclear localization signal in the absence of importin alpha', *Molecular and Cellular Biology*, 19(2), pp. 1218–1225. doi:10.1128/MCB.19.2.1218.

Palù, G. and Loregian, A. (2013) 'Inhibition of herpesvirus and influenza virus replication by blocking polymerase subunit interactions', *Antiviral Research*, 99(3), pp. 318–327. doi:10.1016/j.antiviral.2013.05.014.

Para, M.F. *et al.* (2010) 'Phase I/II trial of the anti-HIV activity of mifepristone in HIV-infected subjects ACTG 5200', *Journal of Acquired Immune Deficiency Syndromes (1999)*, 53(4), pp. 491–495. doi:10.1097/QAI.0b013e3181d142cb.

Pitts, J.D. *et al.* (2017) 'Antiviral activity of N-(4-hydroxyphenyl) retinamide (4-HPR) against Zika virus', *Antiviral Research*, 147, pp. 124–130. doi:10.1016/j.antiviral.2017.10.014.

Puente, S. *et al.* (1989) '[Brucellar cutaneous abscess]', *Enfermedades Infecciosas Y Microbiologia Clinica*, 7(6), pp. 336–337.

Razonable, R.R., Emery, V.C., and 11th Annual Meeting of the IHMF (International Herpes Management Forum) (2004) 'Management of CMV infection and disease in transplant patients. 27-29 February 2004', *Herpes: the journal of the IHMF*, 11(3), pp. 77–86.

Ribbeck, K. *et al.* (1998) 'NTF2 mediates nuclear import of Ran', *The EMBO journal*, 17(22), pp. 6587–6598. doi:10.1093/emboj/17.22.6587.

Riddick, G. and Macara, I.G. (2005) 'A systems analysis of importin-α–β mediated nuclear protein import', *Journal of Cell Biology*, 168(7), pp. 1027–1038. doi:10.1083/jcb.200409024.

Riley, H.D. (1997) 'History of the cytomegalovirus', *Southern Medical Journal*, 90(2), pp. 184–190. doi:10.1097/00007611-199702000-00004.

Ripalti, A. *et al.* (1995) 'Cytomegalovirus-mediated induction of antisense mRNA expression to UL44 inhibits virus replication in an astrocytoma cell line: identification of an essential gene', *Journal of Virology*, 69(4), pp. 2047–2057. doi:10.1128/JVI.69.4.2047-2057.1995.

Rowe, W.P. *et al.* (1956) 'Cytopathogenic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids', *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, 92(2), pp. 418–424.

Ryckman, B.J. *et al.* (2006) 'Human cytomegalovirus entry into epithelial and endothelial cells depends on genes UL128 to UL150 and occurs by endocytosis and low-pH fusion', *Journal of Virology*, 80(2), pp. 710–722. doi:10.1128/JVI.80.2.710-722.2006.

Sanchez, V., Sztul, E. and Britt, W.J. (2000) 'Human cytomegalovirus pp28 (UL99) localizes to a cytoplasmic compartment which overlaps the endoplasmic reticulum-golgi-intermediate compartment', *Journal of Virology*, 74(8), pp. 3842–3851. doi:10.1128/jvi.74.8.3842-3851.2000.

Satheshkumar, P.S. *et al.* (2013) 'Interactions of the vaccinia virus A19 protein', *Journal of Virology*, 87(19), pp. 10710–10720. doi:10.1128/JVI.01261-13.

Sinigalia, E. *et al.* (2008a) 'Role of homodimerization of human cytomegalovirus DNA polymerase accessory protein UL44 in origin-dependent DNA replication in cells', *Journal of Virology*, 82(24), pp. 12574–12579. doi:10.1128/JVI.01193-08.

Sinigalia, E. *et al.* (2008b) 'Role of homodimerization of human cytomegalovirus DNA polymerase accessory protein UL44 in origin-dependent DNA replication in cells', *Journal of Virology*, 82(24), pp. 12574–12579. doi:10.1128/JVI.01193-08.

Sinigalia, E. *et al.* (2012) 'The human cytomegalovirus DNA polymerase processivity factor UL44 is modified by SUMO in a DNA-dependent manner', *PloS One*, 7(11), p. e49630. doi:10.1371/journal.pone.0049630.

Sissons, J.G.P., Bain, M. and Wills, M.R. (2002) 'Latency and reactivation of human cytomegalovirus', *The Journal of Infection*, 44(2), pp. 73–77. doi:10.1053/jinf.2001.0948.

Smith, K.M. *et al.* (2018a) 'Contribution of the residue at position 4 within classical nuclear localization signals to modulating interaction with importins and nuclear targeting', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1865(8), pp. 1114–1129. doi:10.1016/j.bbamcr.2018.05.006.

Smith, K.M. *et al.* (2018b) 'Contribution of the residue at position 4 within classical nuclear localization signals to modulating interaction with importins and nuclear targeting', *Biochimica Et Biophysica Acta. Molecular Cell Research*, 1865(8), pp. 1114–1129. doi:10.1016/j.bbamcr.2018.05.006.

Smith, M.G. (1956) ‘Propagation in tissue cultures of a cytopathogenic virus from human salivary gland virus (SGV) disease’, *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, 92(2), pp. 424–430. doi:10.3181/00379727-92-22498.

Söderberg-Nauclér, C. *et al.* (2001) ‘Reactivation of latent human cytomegalovirus in CD14(+) monocytes is differentiation dependent’, *Journal of Virology*, 75(16), pp. 7543–7554. doi:10.1128/JVI.75.16.7543-7554.2001.

Soderholm, J.F. *et al.* (2011) ‘Importazole, a small molecule inhibitor of the transport receptor importin-β’, *ACS chemical biology*, 6(7), pp. 700–708. doi:10.1021/cb2000296.

Song, Y.-J. and Stinski, M.F. (2002) ‘Effect of the human cytomegalovirus IE86 protein on expression of E2F-responsive genes: a DNA microarray analysis’, *Proceedings of the National Academy of Sciences of the United States of America*, 99(5), pp. 2836–2841. doi:10.1073/pnas.052010099.

Stern-Ginossar, N. *et al.* (2012) ‘Decoding human cytomegalovirus’, *Science (New York, N.Y.)*, 338(6110), pp. 1088–1093. doi:10.1126/science.1227919.

Stewart, M., Kent, H.M. and McCoy, A.J. (1998) ‘Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran’, *Journal of Molecular Biology*, 277(3), pp. 635–646. doi:10.1006/jmbi.1997.1602.

Straschewski, S. *et al.* (2010) ‘Human cytomegaloviruses expressing yellow fluorescent fusion proteins--characterization and use in antiviral screening’, *PloS One*, 5(2), p. e9174. doi:10.1371/journal.pone.0009174.

Tessier, T.M. *et al.* (2019) ‘Viral Appropriation: Laying Claim to Host Nuclear Transport Machinery’, *Cells*, 8(6), p. E559. doi:10.3390/cells8060559.

Volate, S.R. *et al.* (2010) ‘Gossypol induces apoptosis by activating p53 in prostate cancer cells and prostate tumor-initiating cells’, *Molecular Cancer Therapeutics*, 9(2), pp. 461–470. doi:10.1158/1535-7163.MCT-09-0507.

Vonglahn, W.C. and Pappenheimer, A.M. (1925) ‘Intranuclear Inclusions in Visceral Disease’, *The American Journal of Pathology*, 1(5), pp. 445-466.3.

Wagstaff, K.M. *et al.* (2011) ‘An AlphaScreen®-based assay for high-throughput screening for specific inhibitors of nuclear import’, *Journal of Biomolecular Screening*, 16(2), pp. 192–200. doi:10.1177/1087057110390360.

Wagstaff, K.M. *et al.* (2012) 'Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus', *The Biochemical Journal*, 443(3), pp. 851–856. doi:10.1042/BJ20120150.

Wagstaff, K.M. *et al.* (2019) 'Molecular dissection of an inhibitor targeting the HIV integrase dependent preintegration complex nuclear import', *Cellular Microbiology*, 21(1), p. e12953. doi:10.1111/cmi.12953.

Wang, C. *et al.* (2017) 'Nuclear import inhibitor N-(4-hydroxyphenyl) retinamide targets Zika virus (ZIKV) nonstructural protein 5 to inhibit ZIKV infection', *Biochemical and Biophysical Research Communications*, 493(4), pp. 1555–1559. doi:10.1016/j.bbrc.2017.10.016.

van der Watt, P.J. *et al.* (2016) 'Targeting the Nuclear Import Receptor Kpnβ1 as an Anticancer Therapeutic', *Molecular Cancer Therapeutics*, 15(4), pp. 560–573. doi:10.1158/1535-7163.MCT-15-0052.

Weber, F. *et al.* (1998) 'A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins', *Virology*, 250(1), pp. 9–18. doi:10.1006/viro.1998.9329.

Weekes, M.P. *et al.* (2014) 'Quantitative temporal viromics: an approach to investigate host-pathogen interaction', *Cell*, 157(6), pp. 1460–1472. doi:10.1016/j.cell.2014.04.028.

Weiland, K.L. *et al.* (1994) 'Functional analysis of human cytomegalovirus polymerase accessory protein', *Virus Research*, 34(3), pp. 191–206. doi:10.1016/0168-1702(94)90124-4.

Weller, T.H. (2000) 'Cytomegaloviruses: a Historical Perspective', *Herpes: the journal of the IHMF*, 7(3), pp. 66–69.

Wyatt, J.P. and Saxton, J. (1950) 'Generalized cytomegalic inclusion disease', *The Journal of Pediatrics*, 36(3), pp. 271–294, illust. doi:10.1016/s0022-3476(50)80097-5.

Ye, L.B. and Huang, E.S. (1993) 'In vitro expression of the human cytomegalovirus DNA polymerase gene: effects of sequence alterations on enzyme activity', *Journal of Virology*, 67(11), pp. 6339–6347. doi:10.1128/JVI.67.11.6339-6347.1993.