# Mechanistic insights into an ancient adenovirus precursor protein VII shows multiple nuclear import receptor pathways

Gualtiero Alvisi[1], Sepehr Nematollahzadeh[1], Ajani Athukorala[2], Camilla M. Donnelly[3], Silvia Pavan[1], Victoria Atelie-Djossou[1], Enzo Di Iorio[1], Babu Nath[3], Karla J. Helbig[2], Brian McSharry[3], Jade K. Forwood[3], and Subir Sarker[2]

[1]Universita degli Studi di Padova Dipartimento di Medicina Molecolare
[2]La Trobe University Department of Microbiology Anatomy Physiology and Pharmacology
[3]Charles Sturt University School of Dentistry and Health Sciences Wagga Wagga Campus

March 19, 2024

**Abstract**

Adenoviral pVII proteins are multifunctional, highly basic, histone-like proteins that can bind to and transport the viral genome into the host cell nucleus. Despite the identification of several nuclear localization signals (NLSs) in the pVII protein of human adenovirus 5 (HAdV5), the mechanistic details of nuclear transport are largely unknown. Here we provide a full characterization of the nuclear import of precursor (Pre-) pVII protein from an ancient siadenovirus, frog siadenovirus 1 (FrAdV1) using a combination of structural, functional and biochemical approaches. Two strong NLSs (termed NLSa and NLSd) interact with importin (IMP)b and IMPa respectively, and are the main drivers of nuclear import. A weaker NLS (termed NLSb) also contributes, together with an additional signal (NLSc) which we found to be important for nucleolar targeting. Expression of Pre-pVII wild-type and NLS defective derivatives in the presence of selective inhibitors of different nuclear import pathways revealed that, unlike its human counterpart, FrAdV1 Pre-pVII nuclear import is dependent on IMPa/b and IMPb1. Clearly, AdVs evolved to maximise the nuclear import pathways for the pVII proteins, whose subcellular localization is the result of a complex process. Therefore, our results pave the way for an evolutionary comparison of the interaction of different AdVs with the host cell nuclear transport machinery.

## Mechanistic insights into an ancient adenovirus precursor protein VII shows multiple nuclear import receptor pathways

**Running title:** Nuclear import of an ancient adenovirus protein VII

Sepehr Nematollahzadeh[1#], Ajani Athukorala[2#], Camilla M. Donnelly[3,4], Silvia Pavan[1], Victoria Atelie-Djossou[1], Enzo Di Iorio[1], Babu Nath[3], Karla J. Helbig[2], Brian P. McSharry[3,4], Jade K. Forwood[3,4], Subir Sarker[2,5*] and Gualtiero Alvisi[1*]

[1] Department of Molecular Medicine, University of Padua, Via Gabelli 63, 35121 Padua, Italy.

[2] Department of Microbiology, Anatomy, Physiology, and Pharmacology, School of Agriculture, Biomedicine and Environment, La Trobe University, Melbourne, VIC 3086, Australia.

[3] School of Dentistry and Medical Sciences Biomedical Sciences, Charles Sturt University, Wagga Wagga, New South Wales, Australia.

[4] Gulbali Institute, Charles Sturt University, Wagga Wagga, NSW, 2678, Australia

[5] Biomedical Sciences & Molecular Biology, College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, QLD 4811, Australia.

# = joint first author

***** Correspondence: e-mail: subir.sarker@jcu.edu.au; phone: +610747814336; gualtiero.alvisi@unipd.it; phone: +39 (0)498272353.

**Synopsis:**

The study explores the nuclear transport mechanism of the precursor protein VII from an ancient siadenovirus that utilizes multiple nuclear localization signals (NLSs) and pathways for nuclear delivery. It identifies two strong NLSs that are key drivers of this process, interacting with specific importins. Additionally, a weaker NLS and another signal are responsible for targeting of the nucleolus. The study reveals that the import process is different between the frog and human virus counterparts, rely on different pathways, and indicates that adenoviruses have evolved diverse strategies to ensure the nuclear import of pVII proteins. These findings offer insights into the evolutionary aspects of adenovirus interactions with host nuclear transport systems.

**Abstract**

Adenoviral pVII proteins are multifunctional, highly basic, histone-like proteins that can bind to and transport the viral genome into the host cell nucleus. Despite the identification of several nuclear localization signals (NLSs) in the pVII protein of human adenovirus 5 (HAdV5), the mechanistic details of nuclear transport are largely unknown. Here we provide a full characterization of the nuclear import of precursor (Pre-) pVII protein from an ancient siadenovirus, frog siadenovirus 1 (FrAdV1) using a combination of structural, functional and biochemical approaches. Two strong NLSs (termed NLSa and NLSd) interact with importin (IMP)β and IMPα respectively, and are the main drivers of nuclear import. A weaker NLS (termed NLSb) also contributes, together with an additional signal (NLSc) which we found to be important for nucleolar targeting. Expression of Pre-pVII wild-type and NLS defective derivatives in the presence of selective inhibitors of different nuclear import pathways revealed that, unlike its human counterpart, FrAdV1 Pre-pVII nuclear import is dependent on IMPα/β and IMPβ1. Clearly, AdVs evolved to maximise the nuclear import pathways for the pVII proteins, whose subcellular localization is the result of a complex process. Therefore, our results pave the way for an evolutionary comparison of the interaction of different AdVs with the host cell nuclear transport machinery.

**Keywords**

pVII, preVII, NLS, genome packaging, nuclear entry, transportin, nucleolar localization signal (NoLS), siadenovirus

**Abbreviations:**

| | | | |
|---|---|---|---|
| Å | Angstrom | MX2 | High-flux undulator microfocus beamline |
| $^0$C | Celsius | Ng | Nanograms |
| ANOVA | Analysis of Variance | nM | Nanomolar |
| DNA | Deoxyribonucleic acid | Nm | Nanometre |
| dsDNA | Double-stranded DNA | PBS | Phosphate-buffered saline |
| DTT | Dithiothreitol | PDB | Protein Data Bank |
| FITC | Fluorescein isothiocyanate | pH | potential of hydrogen |
| FPLC | Fast protein liquid chromatography | Rpm | Revolutions Per Minute |
| GFP | Green fluorescent protein | SD | Standard deviation |
| GTPase | GTP-binding proteins | SDS-PAGE | Sodium dodecyl-sulfate polyacrylamide gel electrophoresis |
| His-Tag | Poly histidine-tag | U/mL | Units per millilitre |
| kDa | Kilo Dalton | V | Volt |
| Mg | Milligram | v/v | volume per volume |
| Min | Minute | µL | Microlitre |
| mL | Millilitre | Mm | Micrometre |

| | | | |
|---|---|---|---|
| Å | Angstrom | MX2 | High-flux undulator microfocus beamline |
| mM | Millimolar | µM | Micromolar |
| MWCO | Molecular Weight Cut-Off | % | Percent sign |



## Introduction

Adenoviruses (AdVs) are an intensively studied group of viruses due to their remarkable diversity, pathogenic relevance and prominent role as leading candidates as viral vectors in gene therapy. AdVs are classified into six genera: *Mastadenovirus* , *Aviadenovirus* ,*Atadenovirus* , *Siadenovirus* , *Ichtadenovirus* , and *Testadenovirus* , and their infections have been reported in almost all vertebrates. *Siadenovirus* , in particular, is known for its distinctive genomic characteristics and has been identified from reptiles and avian species [1-5]. Genus *Siadenovirus* is proposed to have an amphibian origin. Frog siadenovirus 1 (FrAdV1), identified in the northern leopard frog (*Lithobates pipiens* ), was the first siadenovirus to be sequenced [1] and serves as the common ancestor for all siadenoviruses. However, despite extensive investigations, no further evidence of FrAdV1 infections in frogs or any other animal has been uncovered [6]. The evolution of adenoviruses with an amphibian origin to infect distantly related host species including, critically endangered bird species, emphasises potential host-switching[7]. Their initial detection in a cell line of reptilian origin, and the absence of further evidence concerning frog siadenoviruses presents them as a compelling area of study.

During AdV infection, viral fiber and penton proteins bind and interact with the coxsackie adenoviral receptor and integrin V respectively, triggering virus uptake into clathrin-coated early endosomes (EE)[7-10]. In the EE, a drop in the pH induces conformational changes in the capsid and facilitates the release of the proteins VI and penton, disrupting the endosomal membrane, thus allowing the release and translocation of nucleocapsid to the nucleus in a microtubule and dynein dependent fashion [11]. The capsid consists of three cationic proteins termed protein V, VII, and mu, with protein VII (pVII) being the most abundant. Shortly after viral infection, pVII is expressed as a precursor (Pre-pVII) and transported into the cell nucleus where it tightly interacts with the viral genome, preparing it for assembly [11]. Before genome packaging, Pre-pVII is cleaved by the adenoviral protease, which removes 24 amino acid residues from the N-terminus, resulting in mature pVII, which will be packaged into the new viral particle. Both proteins are highly basic due to the presence of arginine and lysine residues and can be considered as functional analogues of histones, condensing the viral genome to be positioned inside the viral particle, with mature pVII being involved in nuclear delivery of viral DNA upon infection of new cells [10]. Accordingly, both pVII and Pre-pVII are endowed with nuclear targeting abilities during viral infection[8,10,12].

Molecules smaller than 70 kDa are believed to passively diffuse through the nuclear pore complex (NPC); on the other hand, larger ones and those which need to quickly accumulate in the cell nucleus require energy dependent transport, mediated by cellular transporters belonging to the importin (IMP) superfamily, which recognize nuclear localization signals (NLS) on cargoes [13,14]. The best characterized NLS are termed "classical" (c)NLS and are short sequences enriched in basic amino acids such as arginine and lysine. cNLSs can either be monopartite, consisting of a single stretch of basic amino acids, or consist of two basic sequences separated by a short linker in a bipartite arrangement [15-17]. cNLSs directly interact with IMPα, which functions as an adapter bridging the cargo to IMPβ1, which in turn mediates translocation of the complex across the NPC. Non-classical (nc)NLSs, on the other hand, are able to bind directly to IMPβ1 or one of its several homologues such as IMPβ2, IMP7 and many others, without the need for IMPα [18-20]. In recent years, several studies have investigated the structural, functional, and cellular properties of HAdV pVII, and began the characterization of its nuclear targeting abilities. HAdV pVII proteins are able to interact with cellular components such as IMP7, IMPβ2, IMPα/β1, hsp70 and histone H1, so that pVII and Pre-pVII are believed to be able to use multiple nuclear import pathways. Furthermore, HAdV Pre-VII and pVII appear to have different nuclear import preferences, with cleavage believed to alter its IMPs interaction properties. Indeed, while pVII preferentially interacts with IMPβ2, Pre-pVII has a higher affinity for IMPα/β. Unfortunately,

despite several NLSs have been described within pVII, their contribution to nuclear transport is still unclear, and the molecular determinants of nuclear import are not well characterized[8,10,12]. Additionally, nothing is known regarding nuclear import of adenoviruses of animal origin, including siadenoviruses, with the exception of a recent report from psittacine siadenovirus F (PsSiAdV). To address this issue, we thoroughly characterized the nuclear import process of Pre-pVII from FrAdV1, which can be considered the most ancient member of the *Siadenovirus*genus. We identified four putative NLSs and characterized their structural, functional and biochemical properties utilising a wide range of approaches and dissected the contribution of each to Pre-VII subcellular localization. Upon expression in mammalian cells in the absence of any other viral protein, a GFP Pre-pVII fusion protein strongly accumulated in the nucleolus. The subcellular localization of several full-length Pre-pVII derivatives carrying amino acid substitutions within key NLSs basic residues revealed that a total impairment on nuclear localization was detected only upon inactivation of all NLSs, while co-expression with selective nuclear transport inhibitors highlighted an important role for IMPα/β and IMPβ1 but not IMPβ2 for nuclear targeting. By combining quantitative confocal laser scanning microscopy (CLSM) with biochemical and crystallographic assays we could further show that specific NLSs selectively bind different IMPs, and differently contribute to Pre-pVII subcellular localization. Indeed, NLSd (PP**RKRRR** VA-149) binds with high affinity to IMPα, in analogous fashion to several cNLSs, while NLSa (GYW**RRKR** S**KK** A-53) preferentially binds to IMPβ1. On the other hand, NLSc (G**RK** I**KK** A**R** AP-120) is crucial in mediating nucleolar targeting and can, therefore, be considered a nucleolar localization sequence (NoLS), while NLSb plays a minor but significant role by contributing to nucleolar and nuclear targeting.

## Materials and methods

**Phylogenetic analysis.** Phylogenetic analysis based on adenovirus pVII was performed to determine the evolutionary affiliation of the genus *Siadenovirus* with the protein of interest. For that amino acid sequences of pVII available in GenBank for genus*Siadenovirus* members were extracted and aligned with MAFTT (version 7.450), using Geneious (version 22.1.1, Biomatters, Ltd., Auckland, New Zealand) (scoring matrix BLOSUM62; gap open penalty 1.53; offset value 0.123). Following the amino acids sequence alignments, maximum likelihood (ML)-based phylogenetic analysis was performed with 500 non-parametric bootstrap replicates implemented in Geneious (version 22.1.1, Biomatters, Ltd., Auckland, New Zealand).

**Identification of putative NLSs on pVII.** The primary sequence from FrAdV1 Pre-pVII (GenBank accession no; NC_002501[1], was analysed with the cNLS mapper software[21], and by visual inspection to identify putative NLSs.

**Tertiary structure prediction** . AlphaFold2 [22]was utilized to predict a model for pVII in response to the absence of structural information. A comprehensive model for the full-length pVII of FrAdV1 [1] was created, emphasising the identified NLS regions. Additionally, for the purpose of structural comparison, we employed AlphaFold2 and generated a structural model for HAdV5 Pre-pVII[10].

**Plasmids** . Bacterial expression plasmids mediating the expression of truncated version of mouse (m)IMPα2 lacking the autoinhibitory IMPβ1 binding (IBB) domain, and mIMPβ1 from a pET30a backbone were described previously [23,24]. Mammalian expression plasmids pcDNA3.1-NT-GFP-TOPO and pcDNA3.1-NT-GFP-TOPO-SV40-NLS, mediating the expression of GFP cycle3, or of a fusion protein between GFP cycle3 and Simian vacuolating virus (SV) 40 large tumor antigen NLS (PKKKRKV-132), respectively, were described previously [25]. Plasmid pEPI-GFP-UL44, encoding a fusion protein between GFP and human cytomegalovirus (HCMV) DNA polymerase UL44, localizing to the nucleus via the IMPα/β dependent pathway, was also previously described [26]. Plasmid GST-GFP-FUS, encoding a fusion protein translocated via IMPβ2[27] was kindly provided from Dorothee Dormann (Mainz, Germany), plasmid pDsRed-C1-Fibrillarin [28], was kindly provided by Denis Archambault (University of Québec, Canada). Plasmid RFP-M9M, encoding from a potent and selective IMPβ2 inhibitor[29] was provided by Yoshihiro Kino (Tokyo, Japan), plasmid mcherry-Bimax2, encoding for a competitive inhibitor of the IMPα/β nuclear import pathway [30] was a generous gift from Yoshihiro Yoneda and Masahiro Oka (Osaka, Japan). Plasmid DsRed-RanQ69L, encoding for a transdominant negative, GTPase deficient mutant of the Ran GTPase, which impairs Ran-dependent nu-

clear transport[31], was a generous gift from Michael Green (University of Massachusetts, USA). Mammalian expression plasmids encoding for GFP-FrAdV1 Pre-pVII (1-149) and substitution derivatives thereof, as well as for GFP fused to FrAdV1 Pre-pVII individual NLS sequences, were synthesized by BioFab research (Rome, Italy) and GeneScript (New Jersey, USA). A list of all plasmids used in this study is available in Supplementary Table S1.

**Peptides.** Synthetic peptides representing the predicted amino acid residues, with an N-terminal FITC tag, were synthesized and obtained from GenScript (Supplementary Table S2).

**Expression and purification of recombinant proteins.** The overexpression of IMPs was carried out in *E. coli* pLysS BL21(DE3) as described previously [32] using the auto-induction method [33]. After induction, the cultures were centrifuged at 6000 rpm for 20 minutes at 4°C, and the resulting bacterial pellets were resuspended in HIS buffer A (pH 8), which consisted of 50 mM phosphate buffer, 300 mM NaCl, and 20 mM imidazole. To lyse the cells, two freeze-thaw cycles were performed, followed by the addition of lysozyme (1 mL of 20 mg/mL) (Sigma-Aldrich, St. Louis, MI, USA) and DNase (10 µL of 50 mg/mL) (Sigma-Aldrich, St. Louis, MI, USA), and incubation at room temperature for 1 hour. The supernatants containing soluble proteins were collected by centrifugation at 12,000 rpm for 30 minutes at 4°C. The extracts were then filtered through 0.45 µm low protein affinity filters and injected into a 5 mL HisTrap HP column (GE Healthcare, Chicago, IL, USA) that had been pre-equilibrated with His buffer A, in an AKTA purifier FPLC system (GE Healthcare, USA). Followed by washes of 20 column volumes with His buffer A, the proteins of interest were eluted using a gradually increasing gradient of imidazole (ranging from 20 mM to 500 mM) (ChemSupply, Gillman, SA, Australia). The eluted protein fractions were combined and loaded onto a pre-equilibrated HiLoad 26/60 Superdex 200 column (GE Healthcare, USA) in GST buffer A (50 mM Tris and 125 mM NaCl) for further purification using size-exclusion chromatography. The fractions corresponding to the eluted volumes at the respective protein sizes were collected, and the samples were concentrated using an Amicon MWCO 10 kDa filter (Merck Millipore, Burlington, MA, USA). Prior to experimental use, the purity of the samples was assessed by performing SDS-PAGE at 165 V for 30 minutes on a 4–12% Bis-Tris Plus gel (Thermo Fisher Scientific, Waltham, MA, USA).

**Fluorescence polarization (FP) assays.** FrAdV1 Pre-pVII NLSs were incubated with IMPs in a two-fold serial dilution scheme, essentially as described previously [34,35]. The experiment involved testing IMPα2ΔIBB isoform 2 and IMPβ1, separately. Starting with a low concentration of 5 µM, the dilution series was performed across 23 wells in a 96-well, black Fluotrac microplate (Greiner Bio-One, Austria) each plate containing an appropriate negative control lacking the IMP binding partner. GST buffer A (50 mM Tris and 125 mM NaCl) was added to bring the total volume of 200 µL per well and fluorescence polarization measurements were recorded immediately using a CLARIOstar Plus plate reader (BMG Labtech, Germany). The assays were repeated in triplicate to ensure consistency. The data from the three independent experiments were analysed using GraphPad Prism software (version 9.3.1) to determine the dissociation constant (Kd). A summary of FP assays results is shown in Supplementary Table S3.

**Crystallization, data collection, and structure determination.**The hanging drop vapor diffusion method was employed to crystallize the IMPα2ΔIBB:FrAdV1 pVII NLS protein complex. The complex was attained in a 1:1 molar ratio, with each hanging drop having a total volume of 3 µL over a well of 300 µL precipitant solution containing 650 mM sodium citrate (pH 6.5) and 10 mM DTT at a temperature of 23°C. Crystals were formed after 3 days of incubation. Formed crystals were collected and cryoprotected in a precipitant solution containing 20% glycerol, before being rapidly frozen in liquid nitrogen. X-ray diffraction data were obtained at the Australian Synchrotron using the MX2 macromolecular beam lines, utilizing an Eiger 16M detector [36]. The data obtained were subjected to indexing and integration using MOSFLM[37]. Subsequent steps including merging, space group assignment, scaling, and Rfree calculations were carried out using AIMLESS within CCP4, a suite of programs for crystallographic data processing [38]. The final model building and refinement were performed using software tools that are commonly used for model building and refinement in structural biology, COOT[39], and Phenix [40]. Phasing was performed using molecular replacement in Phaser [41] and PDB code 3UKX was used as the search model for IMPα2. The finalized

model was subjected to validation and subsequently deposited to the Protein Data Bank (PDB) with an assigned accession number 8U36; refinement statics are detailed in Supplementary Table S3.

**Cells** . HEK293A cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) foetal bovine serum (FBS), 50 U/ml penicillin, 50 U/ml streptomycin, and 2 mM L-glutamine in a humidified incubator at 37°C in the presence of 5% $CO_2$ and passaged when reached confluence as described in[42].

**Transfections.** HEK293A cells were seeded in a 24-well plate onto glass coverslips ($5\times10^4$ cells/well). The next day, cells were transfected with appropriate amounts of expression constructs (range 5-250 ng), using Lipofectamine 2000 (ThermoFisher Scientific, Monza, Italy), following the manufacturer's recommendations and further incubated at 37 °C and 5% $CO_2$ in complete medium as described [43], until being processed for CLSM.

**CLSM and image analysis.** Cells were transfected and incubated for 24 hours to allow expression of spontaneously fluorescent proteins. Afterward, the cells were treated with DRAQ5 (#62251, ThermoFisher Scientific, Monza, Italy) at a dilution of 1:5000 in DMEM without phenol red for 30 minutes [44]. Following incubation, the cells were washed twice with PHEM 1x solution (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, and 4 mM $MgSO_4$) and fixed with 4% paraformaldehyde (v/v) for 10 minutes at room temperature (RT). After three washes with PBS 1x, coverslips were mounted on glass slides using Fluoromount G (Southern Biotech, Birmingham, AL, USA). The subcellular localization of fusion proteins was examined using a Nikon A1 confocal laser scanning microscope (Nikon, Tokyo, Japan) equipped with a 60x oil immersion objective, following the established protocol outlined in[45,46]. To determine the levels of nuclear accumulation of the proteins of interest, the FiJi public domain software (https://doi.org/10.1038/nmeth.2019) was utilized, and single-cell measurements were taken for nuclear (Fn), nucleolar (Fno), and cytoplasmic (Fc) fluorescence. DRAQ5 and fibrillarin were used to define nuclear and nucleolar masks, respectively, while a small area close to DRAQ5 was used to define a cytosolic mask as previously[47]. The fluorescence attributed to autofluorescence/background (Fb) was subtracted from the measurements, to calculate the Fn/c and Fno/Fn ratios according to the formulas Fn/c = (Fn-Fb)/(Fc-Fb) and Fno/Fn = (Fno-Fb)/(Fn-Fb). Cells with oversaturated signals were excluded from analysis. To allow easier detection of nucleoli, cells were co-transfected with DsRed-fibrillarin expression plasmid, and rgb profile plots calculated with FiJi. Statistical analysis was performed using GraphPad Prism 9 software (GraphPad, San Diego, CA, USA) applying either Student's t-test, one-way ANOVA, or two-way ANOVA, as appropriate.

## Results

**Evolution of Pre-pVII in siadenoviruses and identification of NLSs within Frog siadenovirus 1 (FrAdV1).** Nuclear transport of AdV pVII has been partially characterized for HAdV. Several studies reported that HAdV5 pVII utilizes several different NLSs, enabling transportation to the host cell nucleus via various import pathways, whereas very little is known for other AdVs. While HAdV2 Pre-pVII localized to the nucleolus, we have recently shown that the psittacine siadenovirus F (PsSiAdV) Pre-pVII mainly localized in the nucleoplasm, when expressed in the absence of any other viral proteins, implying important differences in nuclear localization among diverse AdVs[47]. To gain insights into the evolutionary aspects of pVII among siadenovirus members, we performed a phylogenetic analysis using their primary amino acids sequences and identified FrAdV1 pVII as the most ancient form within the genus *Siadenovirus* (Fig. 1A). In order to better understand the nuclear transport of FrAdV1 Pre-pVII, we analyzed its primary sequence bioinformatically with cNLSmapper, and by visual inspection for putative nuclear localization signals (NLSs). Such analysis revealed four putative NLSs (NLSa:$_{43}$GYW**RRKR** S**KK** A$_{53}$, NLSb:$_{74}$**RKK** IP**K** TPVGVVGPLWQGT**RKR** $_{95}$; NLSc:$_{111}$G**RK** I**KK** A**R** AP$_{120}$, NLSd: $_{141}$PP**RKRRR** VA$_{149}$), of which NLSb was identified a putative bipartite NLS (Fig. 1B). FrAdV1 and HAdV5 Pre-pVII share very little sequence identity (26.1%). Accordingly, the identified NLSs have no clear counterparts in HAdV5 pVII (Fig. 1C).

Since the structure of Pre-pVII proteins is yet to be determined, we employed AlphaFold2 to generate models

that may provide preliminary insights into their structure (Supplementary Fig. S1). Predictions suggest HAdV5 Pre-pVII to be disordered, with an unstructured N-terminus (comprised of a loose string of five alpha helices, and two beta strands separated by flexible loops) and a disordered C-terminus (Supplementary Fig. S1A-C). AlphaFold2 also predicts FrAdV1 Pre-pVII to be an intrinsically disordered protein with no structured regions (Supplementary Fig. S1D-E). Such disorder would potentially enable exposure of the putative NLS regions, making them accessible to the nuclear import receptors. Such a model would likely be flexible and dynamic and could adopt multiple conformations and be accessible to bind nuclear import receptors.

**FrAdV1 Pre-pVII protein strongly accumulates in the nucleolus in the absence of other viral proteins** . To investigate the subcellular localization of FrAdV1 Pre-pVII, we transiently expressed a Pre-pVII GFP-fusion protein, and compared its subcellular localization to similar fusions with HAdV2 Pre-pVII and PsSiAdV Pre-pVII, which are known to accumulate into the nucleoli and nucleoplasm, respectively[47]. As further controls, GFP alone and GFP fused to human cytomegalovirus (HCMV) DNA polymerase processivity factor UL44 were also expressed [48], while DsRed-fibrillarin was used as a marker for nucleolar localization (Fig. 2A). As expected, GFP alone localised with a diffuse pattern distributed between the nuclei and cytoplasm, whereas UL44 strongly accumulated in the cell nucleus, and both proteins were largely excluded from the nucleoli (Fig. 2B). As recently shown, Pre-pVII from PsSiAdV exhibited a localization reminiscent of UL44, while HAdV2 accumulated in the nucleoli and colocalised with fibrillarin, although a considerable fraction of the protein was detectable in the nucleoplasm (Fig. 2B). Intriguingly, FrAdV1 Pre-pVII strongly accumulated in the nucleoli, with limited detection in the nucleoplasm, displaying a different localization pattern to its HAdV2 and PsSiAdV orthologues (Fig. 2B).

**FrAdV1 Pre-pVII NLSs can bind to different cellular transporters.** Given their ability to interact with multiple IMPs, HAdVs pVIIs are believed to be transported into the nucleus by multiple pathways [10]. We therefore assessed the binding of each FrAdV-1 putative NLS to IMPα2 and IMPβ1 using a quantitative fluorescence polarization (FP) assay, commonly used in the field[23,25,34,35,47] (Figure 3 and Supplementary Table S3). Interestingly, NLSa interacted with IMPβ1 with high affinity, NLSd interacted with high affinity with IMPα2, while NLSb and NLSc bound with low affinity to both IMPα2 and IMPβ1 (Figure 3 and Supplementary Table S3). These results suggest that FrAdV1 Pre-pVII can interact with different IMPs via distinct NLSs, consistent with its ability to be transported into the nucleus via several nuclear import pathways.

**ΦρΑδ˝1 Πρε-π˝ΙΙ ΝΛΣς ςαν ςονφερ ΙΜΠα/β δεπενδεντ ανδ ινδεπενδεντ νυςλεαρ ιμπορτ το ηετερολογους προτεινς.** Since the FrAdV1 Pre-pVII strongly accumulated in the nucleoli of transfected cells and contains four putative NLSs (Fig. 1BC) which can mediate binding to several IMPs (Fig. 3), we analysed their ability to alter the subcellular localization of a reporter protein. To this end, we analysed the subcellular localization of fusion proteins between GFP and FrAdV1 Pre-VII putative NLSs when transiently expressed in mammalian cells (Fig. 4A), followed by quantitative analysis on nuclear and nucleolar accumulation levels (Fig. 4BC). We also expressed GFP alone and GFP fused to Simian Virus 40 (SV40) large tumour antigen (LTA) NLS, as negative and positive controls for nuclear accumulation respectively. As expected, GFP alone equally distributed between nucleus and cytoplasm (Fig. 4B) with a Fn/c of 1.2 (Fig. 4C); GFP-SV40-NLS strongly accumulated in the cell nucleus (Fig. 4B), with a Fn/c of 6.9 (Fig. 4C); and both proteins were partially excluded from the nucleoli with a Fno/Fn $< 1$ (Fig. 4D). Importantly, fusion to each FrAdV1 Pre-pVII-NLS affected GFP subcellular localization, although to different extents. The highest levels of nuclear accumulation were observed for GFP-FrAdV1 pVII-NLSd, with a Fn/c of 7.2 (Fig. 4C), and similarly to GFP alone, it was mainly excluded from nucleoli (Fno/n $< 1$, Fig. 4D). On the other hand, while nuclear accumulation mediated by either NLSa or NLSc was considerably lower as compared to NLSd, both sequences could significantly increase GFP targeting to the nucleolus (Fig. 4D). Finally, NLSb slightly increased GFP nuclear targeting (Fn/c of. c. 2, see Fig. 4C), but not its nucleolar accumulation (Fno/n $< 1$, see Fig. 4D). Our results suggest that FrAdV1 Pre-pVII contains four sequences which could contribute to its nuclear localization (NLSa-d), and two (NLSa and NLSc) which could contribute to its nucleolar targeting. Since we and others have extensively shown that the peptide inhibitor Bimax2 can abolish

IMPα/β dependent nuclear import [21,25,49], we decided to test its effect on the activity of FrAdV1 Pre-pVII NLSs. We therefore tested the effect of co-expression with mCherry-Bimax2 on the nuclear accumulation of the FrAdV1 Pre-pVII NLSs GFP fusion proteins (Supplementary Fig. S2). As expected, expression with Bimax2 completely abolished nuclear targeting of GFP-SV40 NLS (Supplementary Fig. S2), consistent with its ability to bind exclusively to IMPα/β via IMPα. Importantly, Bimax2 significantly reduced nuclear accumulation mediated by NLSd but not by NLSa, NLSb and NLSc (Supplementary Fig. S2B). These results are consistent with FrAdV1 Pre-pVII NLSd being a cNLS functionally interacting with IMPα.

**Ηιγη-ρεσολυτιον ςρψσταλ στρυςτυρε υνςοερς τηε ςλασσιςαλ βινδινγ ιντερφαςε οφ I-ΜΠα ΦρΑδ˝1 Πρε-π˝ΙΙ ΝΛΣδ.** Based on our observation that the FrAdV1 Pre-pVII NLSd bound to with high affinity to IMPα2, we undertook X-ray crystallography of this complex to gain a deeper understanding of the interactions between FrAdV1 Pre-pVII and host cellular transporters. The protein crystallization process was carried out using the hanging-drop vapor diffusion method, resulting in the formation of large rod-shaped crystals within a three-day incubation period (Fig. 5). These crystals diffracted to 2.2 Å at the Australian Synchrotron on the MX2 beamline, and the data were indexed in P212121, with unit cell parameters of a = 77.29, b = 89.29, and c = 95.80. The structure was solved using Phaser [41] through molecular replacement with the model derived from PDB entry 3UKX. The resolved structure identified one chain of FrAdV1 Pre-pVII NLSd (PP**RKRRR** VA-149) bound to IMPα2ΔIBB (residues 72-498) with a well-characterized lysine at the P2 site. Lys144 is the predominant binding determinant of the major binding site forming hydrogen bonds with Gly150, Thr155, and Asp192 at the P2 binding pocket of IMPα2, with a salt bridge formed between Asp192 of IMPα2 and Lys144 of the NLSd peptide. Arg145 of NLSd interact with the P3 binding site of IMPα2 at Asn188 and Asn288, via hydrogen bonding whereas Arg146 and Arg147 bound at P4 and P5 positions (Fig. 5) [39,40]. The full data collection and refinement statistics is given in Supplementary Table S4. Furthermore, when an excessive amount of peptide is employed, NLSd was found at IMP[?] minor binding site (Fig. 5B). Our results are therefore consistent with NLSd being a cNLS functionally interacting with IMPα major binding site.

**FrAdV1 Pre-pVII can accumulate in the nucleolus in a Ran independent process.** Since nuclear import of HAdV2 Pre-pVII and mature pVII has been shown to rely on IMPα/β, and IMPβ2, respectively[10], we analysed the effect of inhibitors of the IMPα/β and IMPβ2 nuclear import pathways on the subcellular localization of FrAdV1 Pre-pVII. Inhibitors tested include mcherry-Bimax2[30], which impairs IMPα/β dependent nuclear import, M9M-RFP [29], which inhibits IMPβ2 dependent nuclear import, and DsRed-RanQ69L [31], affecting Ran dependent nuclear import (Fig. 6A). GFP-UL44, which is imported into the nucleus by IMPα/β, was also expressed as a control [48]. As previously, in the absence of nuclear import inhibitors, both GFP-UL44 and FrAdV1 Pre-pVII strongly localised to the cell nucleus (Fig. 6B), with a Fn/c $> 20$ (Fig. 6C). Co-expression with either mCherry-Bimax2 (Fn/c 0.3) or DsRed-RanQ69L (Fn/c 1.1) strongly impaired nuclear accumulation of GFP-UL44, while no significant reduction was observed in the presence of RFP-M9M (Fig. 6C). Surprisingly, GFP-FrAdV1 Pre-pVII nuclear import was not significantly impaired by any inhibitor tested. The lack of response to either Bimax2 or M9M can be interpreted as the consequence of the ability of Pre-pVII to interact with multiple IMPs, while nuclear localization in the presence of RanQ69L could be the consequence of its ability to passively enter the nucleus due to its low molecular weight, followed by interaction with nuclear components.

**Molecular dissection of the role of FrAdV1 individual in nuclear and nucleolar targeting.** To verify these hypotheses**,** we dissected the contribution of each NLS identified here to FrAdV1 Pre-VII nuclear and nucleolar targeting. Therefore, we quantitatively analysed the subcellular localization of transiently expressed GFP fusion proteins with FrAdV1 Pre-pVII and several substitution derivatives thereof, whereby NLS basic residues were replaced either by A or T (Fig. 7A), using DsRed-Fibrillarin as a nucleolar marker. Importantly, individual inactivation of single NLSs did not significantly reduce nuclear accumulation (Fig. 7B and C), with barely no signal being detectable in the cytosol of cells expressing GFP-FrAdV1 Pre-pVII wt (Fn/c of 17.7), mNLSa (Fn/c of 14.7), mNLSb (Fn/c of 14.8), mNLSc (Fn/c of 17.9) and mNLSd (Fn/c of 15.69). Simultaneous inactivation of either NLSa and NLSc (mNLAac, Fn/c of 11), NLSa and NLSd, (mNLSad, Fn/c of 12) or NLSc and NLSd (mNLScd, Fn/c of 15) only mildly reduced nuclear localization

(Fig. 7C). These results confirm that, when any of FrAdV1 Pre-VII NLS is individually inactivated, the others can functionally compensate for its absence. However, simultaneous inactivation of NLSa, NLSc and NLSd was sufficient to significantly reduce nuclear accumulation (mNLSacd, Fn/c of 7.7). This data suggests that NLSb is functional in the context of full-length protein, but not sufficient to mediate optimal nuclear targeting. Accordingly, further inactivation of NLSb completely abolished nuclear accumulation (mNLSabcd, Fn/c 1.5). Therefore, FrAdV1 contains multiple functional NLSs responsible of active nuclear import, which are apparently functionally redundant.

On the other hand, almost any substitution introduced in FrAdV1 Pre-pVII NLSs significantly affected nucleolar accumulation, although to a different extent (Fig. 7B). Quantitative analysis revealed that GFP-FrAdV1 Pre-pVII nucleolar accumulation (Fno/n of c. 20) was minimally affected by substitutions within NLSa (mNLSa, Fno/n of 13.5) or NLSd (mNLSd, Fno/n of 10.8), while it was clearly affected by substitution of either NLSb (mNLSb, Fno/n of 3.7) and especially NLSc (mNLSc, Fno/n of 1.2). Overall, all derivatives containing substitutions within NLSc failed to accumulate in the nucleoli, confirming that NLSc ($_{111}$G**RK** I**KK** A**R** AP$_{120}$) has a major role in nucleolar accumulation, while NLSb (**RKK** IP**K** TPVGVVGPLWQGT**RKR** -95) also contributes to nucleolar targeting.

**ΦΑδ˝1 Πρε-π˝ΙΙ ις τρανσπορτεδ ιντο τηε νυςλευς βψ ΙΜΠα/β ανδ ΙΜΠβ, σπεςιφιςαλλψ ρεςογνισινγ ινδιιδυαλ ΝΛΣς.** To more precisely characterize the contribution of each NLS identified in this study to the functional interaction with specific IMPs, we analyzed the effect of overexpressing inhibitors of specific nuclear import pathways on the subcellular localization of the FrAdV Pre-pVII substitution derivatives. To this end, mammalian cells were transfected to express GFP-fusion proteins (Fig. 8A) in the presence and absence of mcherry-Bimax2, M9M-RFP or DsRed-RanQ69L, and microscopic images were captured by CLSM (Fig. 8B), followed by quantitative analysis of the levels of nuclear (Fn/c; Fig. 8C) and nucleolar (Fno/n; Fig 8D) accumulation. Despite the nuclear accumulation of FrAdV1 Pre-pVII not being affected by any inhibitor tested, nuclear import of all substitution derivatives lacking NLSc (mNLSc, mNLSac, mNLScd and mNLSabc) was significantly inhibited by RanQ69L (Fig. 8B and C, see Supplementary Figs S3-S5). This suggests that Pre-pVII is capable of passively diffusing into the nucleus, likely due to its small size, and accumulating in the nucleolus in the absence of active transport by binding to nucleolar components dependent on NLSc, which functions as a NoLS. In the absence of a NoLS, nuclear accumulation is dependent on Ran-dependent active nuclear import, and therefore inhibited by RanQ69L. Intriguingly, M9M did not inhibit nuclear import of any FrAdV1 derivative (Fig. 8BC, Supplementary Figs S6 and S7), despite the reduced accumulation of a GFP-GST-FUS fusion protein, which has been shown to be functionally depend on transportin (also called IMPβ2; Supplementary Fig. S8). These findings suggest that transportin does not play a major role in FrAdV1 Pre-pVII nuclear import. Finally, nuclear accumulation of most FrAdV1 Pre-pVII derivatives was not affected by Bimax2 co-expression, unless both NLSa and NLSc were simultaneously inactivated, such as in the case of FrAdV1 Pre-pVII mNLSac, and Pre-pVII mNLSacd (Fig. 8B and C, Supplementary Figs S9 and S10). Taken together, our data suggest that each signal identified in this study plays a specific and independent role in mediating FrAdV1 Pre-pVII subcellular trafficking. NLSa confers IMPβ1 dependent nuclear import, NLSb is an atypical sequence which contributes both to IMPα/β nuclear import and nucleolar targeting, and NLSc is the main NoLS, mediating nucleolar localization by interacting with intranuclear components, while NLSd is a cNLS conferring IMPα/β nuclear import (Fig. 9).

## Discussion

While several studies have identified NLSs in pVII from human mastadenovirus C, the presence, specific contribution, as well as their role in nuclear import of various other adenovirus members remained elusive. Here we have combined bioinformatics, microscopy imaging, biochemical and structural approaches to identify and characterize four putative NLSs in Pre-pVII from the ancient siadenovirus FrAdV1. Further, by co-expression of wild type and NLS substitution derivatives thereof in the presence and absence of several nuclear import pathways inhibitors (Fig. 2, 6, 7, 8) we have characterized the contribution of each NLS to the subcellular localization of FrAdV1 Pre-pVII, thus unveiling the contribution of different nuclear import pathways. This is the first study to characterize and dissect the functional contribution of

individual sequences to the subcellular localization of an AdV pVII protein and may provide a basis for future evolutionary comparisons regarding different AdV pVII proteins [8,10,12,50].

When expressed in the absence of other viral proteins, FrAdV1 Pre-pVII strongly accumulated in cell nucleoli (Fig. 2), in stark contrast to what was reported for Pre-pVII from both PsSiAdV which is known to accumulate in the nucleoplasm [47] and from HAdV2, which has been shown to localize in the nucleoplasm and in the nucleoli[8]. The reason and functional significance for such discrepancies remain elusive, but it is noteworthy that in the case of FrAdV-1, nucleolar accumulation is dependent on a highly basic sequence (NLSc:$_{111}$G**RK** I**KK** A**R** AP$_{120}$), which is the main determinant of nucleolar targeting, and a bipartite sequence located upstream (NLSb:$_{74}$**RKK** IP**K** TPVGVVGPLWQGT**RKR** $_{95}$; see Fig. 7), which also contributes to the process, although to a lesser extent. On the other hand, only one NoLS ($_{93}$M**RR** YA**K** MK**RRRRR** VA**RR** H**RRR** $_{112}$) has been identified in HAdV2 Pre-pVII [8], while none have been identified in PsSiAdV Pre-pVII [47]. Therefore, it is likely that the number and strength of NoLSs on AdV Pre-pVII proteins influences their degree of nucleolar targeting. Our findings strengthen the hypothesis that Pre-pVII from different AdVs are endowed with specific subcellular localization abilities, as exemplified by the ability of bovine adenovirus 3 Pre-pVII to localize to mitochondria during viral infection [51].

We propose a model for Frog Pre-pVII nuclear transport where the strong nucleolar targeting is due to the concerted action of the four NLSs identified here, each playing a specific role in protein subcellular localization (Fig. 9). The N-terminal NLSa appears to be a non-classical NLS, binding with high affinity to IMPβ1 directly (Fig. 3 and 4), similarly to HIV-1 tat and Rev NLSs [52]. NLSb is poorly active outside of its physiological context, both in terms of IMP binding (Fig. 3) and nuclear targeting activity (Fig. 4 and Supplementary Fig. S2). However, it is sufficient to confer IMPα/β1 dependent nuclear targeting to full-length FrAdV1 Pre-pVII once all other NLSs have been inactivated (Fig. 7 and Fig. 8). This suggest that NLSb might be an atypical NLS which binds IMPs thanks to its specific three-dimensional conformation, rather than in a liner fashion, such as reported for the NLS described in the intestinal fatty acid binding protein [53]. NLSc can partially target a heterologous protein to the nucleolus in an IMPα/β1 independent fashion, does not bind either IMPα2 nor IMPβ1 with high affinity (Fig. 3 and Supplementary Table S4), but substitution of its basic residues in the context of full-length Pre-pVII completely ablates nucleolar targeting (Fig. 7), suggesting that it is important for nucleolar localization, but by itself, not sufficient to localise proteins to the nucleolus. It therefore has similarities to regions within HAdV2 that act as an NoLS[8]. Finally, the C-terminal NLSd is a *bona fidae*monopartite cNLS mediating IMPα/β1 dependent nuclear transport, by directly interacting with IMPα major binding site (Figs. 3-5).

Therefore, our data are consistent with previous studies that reported the ability of Pre-pVII proteins to bind multiple IMPs and be important through multiple pathways [8,10,12,50]. Whilst HAdV5 Pre-pVII nuclear transport is dependent on IMPα/β1 [12], wild-type FrAdV1 Pre-pVII can accumulate in the nucleolus even in the presence of the highly efficient IMPα/β1 inhibitor Bimax2. Since simultaneous mutation of NLSc (conferring nucleolar accumulation) and NLSa (binding to IMPβ1 with high affinity) renders FrAdV1 Pre-pVII nuclear transport sensitive to Bimax2, FrAdV1 can be actively imported in the nucleus by IMPα/β1 (via NLSd) and IMPβ1 (via NLSa). The evidence that FrAdV1 Pre-pVII nuclear and nucleolar localization was not impacted by the transdominant negative RanQ69L, strongly supports the idea that it can enter the nucleus by passive diffusion and be therein retained after interaction with dsDNA and other nuclear components. Indeed, HAdV2 Pre-pVII has been shown to interact with cellular DNA and co-localize with human chromosomes [8], and mutation of the highly basic NLSc, which results in loss of nucleolar localization, is sufficient to render Frog Pre-pVII sensitive to RanQ69L, but not to Bimax2.

In conclusion, our findings revealed important differences between the import of FrAdV1 Pre-pVII in comparison to what has been reported for HAdV and PsSiAdV. In particular the protein accumulated in the nucleolus more strongly as compared to its orthologues, and its nuclear targeting is insensitive to inhibition of IMPα/β1. The findings of this study provide valuable insights into how different adenoviruses interact with the host cell's nuclear transport machinery. Understanding these interactions may have broader implications for viral replication and infection strategies across different adenovirus genera, and potential applications in

therapeutic approaches.

**Conflict of Interest**

All authors declare that they have no conflicts of interest.

**Funding information**

Subir Sarker is the recipient of an Australian Research Council Discovery Early Career Researcher Award (grant number DE200100367) funded by the Australian Government. Ajani Athukorala is supported by La Trobe University (LTU) under the LTU Graduate Research Scholarship and LTU Full Fee Research Scholarship. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. Camilla Donnelly is a THRIIVE postdoctoral fellow.

**Author Contributions**

The authors' contributions include conceptualisation by AA, BPM, JKF, GA, SS, experiments performed by AA, SN, BN, SP, VAD data acquisition and analysis by AA, SN, CMD, GA writing of initial draft, by GA, AA, SN review and editing by AA, CMD, SP, SN, VAD, BN, KJH, BPM, GA, JKF, SS, ED, supervision by GA and SS, and funding acquisition by SS. All authors have read and agreed to the published version of the manuscript.

**Supplementary Material**

The following Supplementary Material is available online.**Supplementary Table S1.** List of plasmids used in this study**. Supplementary Table S2.** List of peptides used in this study. **Supplementary Table S3.** Statistics for fluorescent polarisation analysis of NLS peptides-IMPs interaction.**Supplementary Table S4.** Refinement statistics of mIMPα2ΔIBB:FrAdV1 Pre-pVII NLSd complex.

**Supplementary Figure S1.** AlphaFold2 generated models of FrAdV1 and HAdV5 pVII. **Supplementary Figure S2.** FrAdV1 Pre-pVII contains IMPα/β dependent and independent NLSs. **Supplementary Figures S3-S10.** Full images and analysis from Figure 8.

**References**

1. Davison AJ, Wright KM, Harrach B. DNA sequence of frog adenovirus.*Journal of General Virology.* 2000;81:2431-2439.

2. Kovács ER, Benko M. Complete sequence of raptor adenovirus 1 confirms the characteristic genome organization of siadenoviruses.*Infection Genetics and Evolution.* 2011;11(5):1058-1065.

3. Benko M, Aoki K, Arnberg N, et al. ICTV Virus Taxonomy Profile: Adenoviridae 2022. *J Gen Virol.* 2022;103(3).

4. Nguyen TH, Ballmann MZ, Do HT, et al. Crystal structure of raptor adenovirus 1 fibre head and role of the beta-hairpin in siadenovirus fibre head domains. *Virology Journal.* 2016;13.

5. Singh AK, Berbís MA, Ballmann MZ, et al. Structure and Sialyllactose Binding of the Carboxy-Terminal Head Domain of the Fibre from a Siadenovirus, Turkey Adenovirus 3. *Plos One.* 2015;10(9).

6. Harrach B, Tarjan ZL, Benko M. Adenoviruses across the animal kingdom: a walk in the zoo. *FEBS Lett.* 2019;593(24):3660-3673.

7. Athukorala A, Phalen DN, Das A, Helbig KJ, Forwood JK, Sarker S. Genomic Characterisation of a Highly Divergent Siadenovirus (Psittacine Siadenovirus F) from the Critically Endangered Orange-Bellied Parrot (Neophema chrysogaster). *Viruses.* 2021;13(9).

8. Lee TWR, Blair GE, Matthews DA. Adenovirus core protein VII contains distinct sequences that mediate targeting to the nucleus and nucleolus, and colocalization with human chromosomes. *J Gen Virol.*2003;84(Pt

12):3423-3428.

9. Wickham TJ, Mathias P, Cheresh DA, Nemerow GR. Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell.* 1993;73(2):309-319.

10. Wodrich H, Cassany A, D'Angelo MA, Guan T, Nemerow G, Gerace L. Adenovirus core protein pVII is translocated into the nucleus by multiple import receptor pathways. *J Virol.*2006;80(19):9608-9618.

11. Greber UF, Suomalainen M. Adenovirus entry: Stability, uncoating, and nuclear import. *Mol Microbiol.* 2022;118(4):309-320.

12. Hindley CE, Lawrence FJ, Matthews DA. A role for transportin in the nuclear import of adenovirus core proteins and DNA. *Traffic.*2007;8(10):1313-1322.

13. Gorlich D, Kostka S, Kraft R, et al. Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. *Curr Biol.* 1995;5(4):383-392.

14. Timney BL, Raveh B, Mironska R, et al. Simple rules for passive diffusion through the nuclear pore complex. *The Journal of cell biology.* 2016;215(1):57-76.

15. Dingwall C, Robbins J, Dilworth SM, Roberts B, Richardson WD. The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. *The Journal of cell biology.*1988;107(3):841-849.

16. Kalderon D, Richardson WD, Markham AF, Smith AE. Sequence requirements for nuclear location of simian virus 40 large-T antigen.*Nature.* 1984;311(5981):33-38.

17. Alvisi G, Jans DA, Camozzi D, et al. Regulated transport into the nucleus of herpesviridae DNA replication core proteins. *Viruses.*2013;5(9):2210-2234.

18. Cingolani G, Bednenko J, Gillespie MT, Gerace L. Molecular basis for the recognition of a nonclassical nuclear localization signal by importin beta. *Mol Cell.* 2002;10(6):1345-1353.

19. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell.*1998;94(2):193-204.

20. Fontes MR, Teh T, Kobe B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *J Mol Biol.* 2000;297(5):1183-1194.

21. Kosugi S, Hasebe M, Entani T, Takayama S, Tomita M, Yanagawa H. Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling. *Chemistry & biology.*2008;15(9):940-949.

22. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.*2021;596(7873):583-589.

23. Hoad M, Cross EM, Donnelly CM, Sarker S, Roby JA, Forwood JK. Structural Characterization of Porcine Adeno-Associated Virus Capsid Protein with Nuclear Trafficking Protein Importin Alpha Reveals a Bipartite Nuclear Localization Signal. *Viruses.* 2023;15(2).

24. Teh T, Tiganis T, Kobe B. Crystallization of importin alpha, the nuclear-import receptor. *Acta crystallographica Section D, Biological crystallography.* 1999;55(Pt 2):561-563.

25. Alvisi G, Manaresi E, Cross EM, et al. Importin alpha/beta-dependent nuclear transport of human parvovirus B19 nonstructural protein 1 is essential for viral replication. *Antiviral Res.* 2023;213:105588.

26. Sinigalia E, Alvisi G, Mercorelli B, et al. Role of homodimerization of human cytomegalovirus DNA polymerase accessory protein UL44 in origin-dependent DNA replication in cells. *J Virol.*2008;82(24):12574-12579.

27. Dormann D, Madl T, Valori CF, et al. Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS.*Embo Journal.* 2012;31(22):4258-4275.

28. Gomez Corredor A, Archambault D. The bovine immunodeficiency virus rev protein: identification of a novel lentiviral bipartite nuclear localization signal harboring an atypical spacer sequence. *J Virol.* 2009;83(24):12842-12853.

29. Kino Y, Washizu C, Aquilanti E, et al. Intracellular localization and splicing regulation of FUS/TLS are variably affected by amyotrophic lateral sclerosis-linked mutations. *Nucleic Acids Res.*2011;39(7):2781-2798.

30. Tsujii A, Miyamoto Y, Moriyama T, et al. Retinoblastoma-binding Protein 4-regulated Classical Nuclear Transport Is Involved in Cellular Senescence. *J Biol Chem.* 2015;290(49):29375-29388.

31. Heilman DW, Teodoro JG, Green MR. Apoptin nucleocytoplasmic shuttling is required for cell type-specific localization, apoptosis, and recruitment of the anaphase-promoting complex/cyclosome to PML bodies. *Journal of Virology.* 2006;80(15):7535-7545.

32. Roman N, Christie M, Swarbrick CM, Kobe B, Forwood JK. Structural characterisation of the nuclear import receptor importin alpha in complex with the bipartite NLS of Prp20. *PLoS One.*2013;8(12):e82038.

33. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif.* 2005;41(1):207-234.

34. Cross EM, Akbari N, Ghassabian H, et al. A functional and structural comparative analysis of Large Tumor Antigens reveals evolution of different importin alpha-dependent nuclear localization signals.*Protein Sci.* 2023.

35. Cross EM, Marin O, Ariawan D, et al. Structural determinants of phosphorylation-dependent nuclear transport of HCMV DNA polymerase processivity factor UL44. *FEBS Lett.* 2023.

36. Aragao D, Aishima J, Cherukuvada H, et al. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J Synchrotron Radiat.* 2018;25(Pt 3):885-891.

37. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.*Acta crystallographica Section D, Biological crystallography.*2011;67(Pt 4):271-281.

38. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D.* 2011;67:282-292.

39. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta crystallographica Section D, Biological crystallography.* 2010;66(Pt 4):486-501.

40. Adams PD, Afonine PV, Bunkoczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D, Biological crystallography.* 2010;66(Pt 2):213-221.

41. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.*2007;40(Pt 4):658-674.

42. Messa L, Celegato M, Bertagnin C, et al. The Dimeric Form of HPV16 E6 Is Crucial to Drive YAP/TAZ Upregulation through the Targeting of hScrib. *Cancers.* 2021;13(16).

43. Di Antonio V, Palu G, Alvisi G. Live-Cell Analysis of Human Cytomegalovirus DNA Polymerase Holoenzyme Assembly by Resonance Energy Transfer Methods. *Microorganisms.* 2021;9(5).

44. Alvisi G, Jans D, Ripalti A. Human cytomegalovirus (HCMV) DNA polymerase processivity factor ppUL44 dimerizes in the cytosol before translocation to the nucleus. *Biochemistry.*2006;45(22):6866-6872.

45. Alvisi G, Paolini L, Contarini A, et al. Intersectin goes nuclear: secret life of an endocytic protein. *The Biochemical journal.*2018;475(8):1455-1472.

46. Smith KM, Di Antonio V, Bellucci L, et al. Contribution of the residue at position 4 within classical nuclear localization signals to modulating interaction with importins and nuclear targeting.*Biochim Biophys Acta.* 2018;1865(8):1114-1129.

47. Athukorala A, Donnelly CM, Pavan S, et al. Structural and functional characterization of siadenovirus core protein VII nuclear localization demonstrates the existence of multiple nuclear transport pathways.*J Gen Virol.* 2024;105(1).

48. Alvisi G, Jans D, Guo J, Pinna L, Ripalti A. A protein kinase CK2 site flanking the nuclear targeting signal enhances nuclear transport of human cytomegalovirus ppUL44. *Traffic.* 2005;6(11):1002-1013.

49. Alvisi G, Avanzi S, Musiani D, et al. Nuclear import of HSV-1 DNA polymerase processivity factor UL42 is mediated by a C-terminally located bipartite nuclear localization signal. *Biochemistry.*2008;47(52):13764-13777.

50. Kulanayake S, Tikoo SK. Adenovirus Core Proteins: Structure and Function. *Viruses.* 2021;13(3).

51. Anand SK, Gaba A, Singh J, Tikoo SK. Bovine adenovirus 3 core protein precursor pVII localizes to mitochondria, and modulates ATP synthesis, mitochondrial Ca<SUP>2+</SUP> and mitochondrial membrane potential. *Journal of General Virology.*2014;95:442-452.

52. Chook YM, Suel KE. Nuclear import by karyopherin-betas: recognition and inhibition. *Biochim Biophys Acta.* 2011;1813(9):1593-1606.

53. Suarez M, Canclini L, Esteves A. Identification of a non-classical three-dimensional nuclear localization signal in the intestinal fatty acid binding protein. *PLoS One.* 2020;15(11):e0242312.

54. Heilman DW, Teodoro JG, Green MR. Apoptin nucleocytoplasmic shuttling is required for cell type-specific localization, apoptosis, and recruitment of the anaphase-promoting complex/cyclosome to PML bodies. *J Virol.* 2006;80(15):7535-7545.

**Figures and Legends**

**Hosted file**

`image1.emf` available at https://authorea.com/users/757211/articles/729497-mechanistic-insights-into-an-ancient-adenovirus-precursor-protein-vii-shows-multiple-nuclear-import-receptor-pathways

**Figure 1. The evolution of pVII in siadenoviruses and identified NLSs in FrAdV1 pVII.** (**A** ) The phylogenetic tree exhibits the possible evolutionary relationship of siadenoviruses based on protein VII. The numbers on the left show bootstrap values as percentages, and the labels at branch tips refer to adenovirus species name, followed by GenBank accession number in parentheses. (**B** ) Frog adenovirus 1 (FrAdV1 pVII) (GenBank Accession no: NC_002501) [7] and human adenovirus (HAdV pVII) (GenBank Accession no: AC_000008) sequences were aligned using Geneious Prime (version 21.1.1, Biomatters, Ltd., Auckland, New Zealand) and previously described NLSs and NoLS are highlighted in dark and light blue shadings respectively. (**C** ) Simplified diagram displaying the identified putative NLS regions of FAdV pVII and previously described NLSs and NoLS from HAdV2[7,8,10].

**Hosted file**

`image2.emf` available at https://authorea.com/users/757211/articles/729497-mechanistic-insights-into-an-ancient-adenovirus-precursor-protein-vii-shows-multiple-nuclear-import-receptor-pathways

**Figure 2. FrAdV1 Pre-pVII proteins localizes to the cell nucleus, and accumulates in the nucleoli. (A)** Schematic representation of the GFP-fusion proteins expressed. Proteins are represented by *white horizontal bars* . NLSs are shown as *blue vertical bars* , along with their amino acidic sequence. The *single letter* amino acid code is used. Basic residues are in *boldface* . (B) The indicated GFP fusion proteins were transiently co-expressed in HEK293A with DsRed-fibrillarin by means of Lipofectamine 2000 transfection. 24h post transfection, cells were incubated with DRAQ5 to stain cell nuclei, fixed with paraformaldehyde and processed for CLSM analysis as described in the Materials and Methods section. Representative images of the 633 nm (nuclei), 488 nm (GFP), and 568 (fibrillarin) laser channels are shown, along with a merged image (merge), and a rgb profile plot across the indicated area.

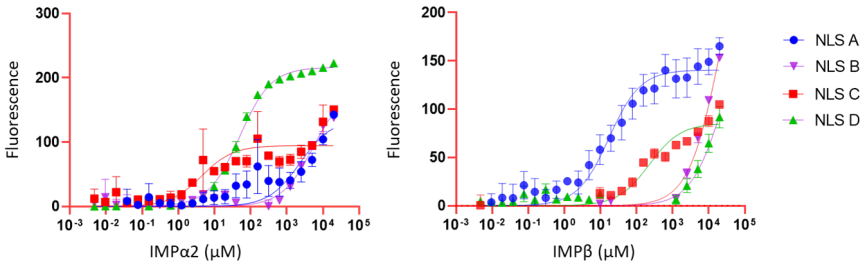


*Figure 3.* **Analysis of FrAdV1 Pre-pVII NLS peptides interaction with IMPs.** Fluorescence polarisation assay measuring the direct binding between FrAdV1 Pre-pVII NLSs and IMPα2 (left) and IMPβ (right). Data are shown as mean values ± SEM relative to three independent experiments.

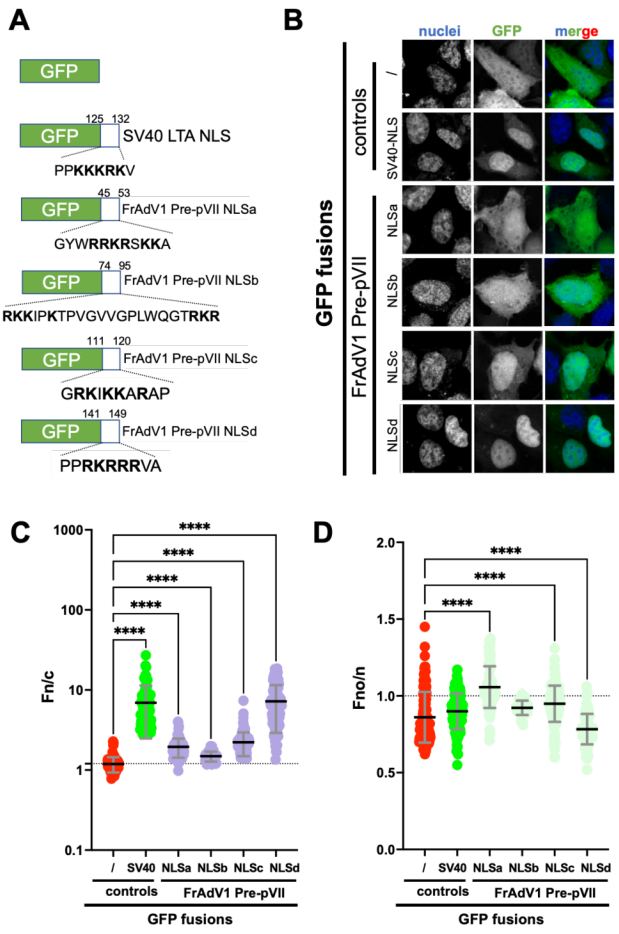


*Figure 4.* **FrAdV1 Pre-pVII contains multiple functional NLSs.** (A) Schematic representation of the GFP-fusion proteins expressed, with NLSs represented as white boxes, along with their amino acid sequence and position. (B) HEK 293-A cells were seeded on glass coverslips and transfected to express the indicated GFP fusion proteins. 24 h post transfection cells were incubated with DRAQ5 to stain cell nuclei, fixed and coverslips mounted on slide holders before being analysed by using a Nikon A1 CLSM equipped with a 60x oil immersion objective. Representative images of the 633 nm (nuclei), 488 nm (GFP), and 568 (fibrillarin) laser channels are shown, along with a merged image (merge). Micrographs such as those shown were used for image analysis by measuring nucleolar (Fno), nuclear (Fn), cytoplasmic (Fc) and background (Fb) fluorescence for single cells using FiJi. (C) The levels of nuclear accumulation (Fn/c) were calculated using the formula Fn/c = (Fn-Fb)/(Fc-Fb). (D) The levels of nucleolar accumulation (Fno/n) were calculated using the formula Fno/n = (Fno-Fb)/(Fn-Fb). Data shown are individual measurements (*circles* ), along with means (black horizontal bars) + standard deviation of the mean relative to > 45 cells from three independent experiments per each GFP fusion protein, along with the results of Welch and Brown-Forsythe ANOVA statistical analysis as compared to GFP alone (/). **** = p <0.0001.

### Hosted file

`image5.emf` available at https://authorea.com/users/757211/articles/729497-mechanistic-insights-into-an-ancient-adenovirus-precursor-protein-vii-shows-multiple-nuclear-import-receptor-pathways

*Figure 5.* **ˆρψσταλ στρυςτυρε οφ αδενοιρυς ΦΑρδˇ1 Πρε-πˇΙΙ ΝΛΣδ βουνδ το Ι-ΜΠα2ΔΙΒΒ.** Protein crystals of IMPα2ΔIBB in complex with FrAdV1 Pre-pVII NLSd FITC-labelled peptide were formed after 3 days of incubation at 23 $^0$C with 0.65M Na Citrate, pH 6.5. The structure solved at 2.2 Å resolution showed the FAdV1 pVII NLSd (shown in green color stick mode) bound to IMPα (shown in grey surface mode) at the major binding site. The P2 binding pocket (pink surface) is comprised of Gly150, Thr155 and Asp192.

**Hosted file**

`image6.emf` available at https://authorea.com/users/757211/articles/729497-mechanistic-insights-into-an-ancient-adenovirus-precursor-protein-vii-shows-multiple-nuclear-import-receptor-pathways

*Figure 6.* **ΦρΑδˇ1 Πρε-πˇΙΙ λοςαλιζε το τηε ςελλ νυςλευς ινδεπενδεντλψ οφ ΙΜΠα/β ανδ ΙΜΠβ2.** (A) Schematic representation of the GFP-fusions and of nuclear import inhibitors used in the assay. Proteins are represented by white horizontal bars. NLSs are shown as blue vertical bars, along with their amino acidic sequence and position. The single letter amino acid code is used. Basic residues are in boldface. (B) The indicated GFP fusion proteins were transiently expressed in HEK293A by means of Lipofectamine 2000 transfection, in the absence (*no add* ) or in the presence of plasmids mediating the expression of either mCherry-Bimax2 *(+ Bimax2* ), RFP-M9M *(+ M9M* ) or DsRed-RanQ69L *(+ RanQ69L* ). 24h post transfection, cells were incubated with DRAQ5 to stain cell nuclei, fixed with paraformaldehyde and processed for CLSM analysis as described in the Materials and Methods section. Representative images the 633 nm (*nuclei* ), the 488 nm (*GFP* ), and 561 nm (*inhibitor* ) channels are shown, along with a merged image (*merge* ). (C) Micrographs such as those shown in (B) were quantitatively analysed to calculate the levels of nuclear accumulation (Fn/c) relative to the indicated fusion proteins at the single cell level, as described in the Materials and Methods section. Data shown are individual measurements (*circles* ), along with means (black horizontal bars) + standard error of the mean (SEM), relative to pooled data from at least 27 cells from three independent experiments. The dashed horizontal line indicates a Fn/c of 1, which corresponds to equal distribution between nucleus and cytoplasm. Results from two-way ANOVA test for the nuclear accumulation of the indicated GFP-fusion proteins in the absence or in the presence of the indicated nuclear transport inhibitors are shown. ***: $p < 0.005$; ****: $p < 0.0001$.

**Hosted file**

`image7.emf` available at https://authorea.com/users/757211/articles/729497-mechanistic-insights-into-an-ancient-adenovirus-precursor-protein-vii-shows-multiple-nuclear-import-receptor-pathways

**Figure 7. Dissection of the individual contribution of each putative NLS to FrAdV1 nuclear and nucleolar targeting.** (A) Schematic representation of the GFP-FrAdV1 Pre-pVII fusion proteins expressed. Proteins are represented by a *white horizontal bar* . NLSs are shown as *blue vertical bars* , along with their amino acidic sequence. Individual amino acidic substitutions are indicated by arrows and their sequence is in *red* . The *single letter* amino acid code is used. Basic residues are in *boldface* . (B) The indicated GFP fusion proteins were transiently co-expressed with DsRed-fibrillarin in HEK293A by means of Lipofectamine 2000 transfection. 24h post transfection, cells were incubated with DRAQ5 to stain cell nuclei, fixed with paraformaldehyde and processed for CLSM analysis as described in the Materials and Methods section. Representative images of the 633 nm (nuclei), 488 nm (Pre-pVII), and 568 (fibrillarin) laser channels are shown, along with a merged image (merge), and a rgb profile plot across the indicated area. (C, D) Micrographs such as those shown in (B) were quantitatively analysed to calculate the levels of nuclear (C) or nucleolar (D) accumulation relative to the indicated fusion proteins at the single cell level, as described in the Materials and Methods section). Data shown are individual measurements (circles), along with are means (black horizontal bars) + standard error of the mean (SEM), relative to pooled data from at least 70 cells from three independent experiments. Results from Welch and Brown-Forsythe ANOVA test for accumulation of the indicated mutant Pre-pVII proteins as compared to the wild-type protein. ***: p <

0.001; ****: $p < 0.0001$.



**Hosted file**

`image8.emf` available at https://authorea.com/users/757211/articles/729497-mechanistic-insights-into-an-ancient-adenovirus-precursor-protein-vii-shows-multiple-nuclear-import-receptor-pathways

*Figure 8.* **Selective response of FrAdV1 Pre-pVII NLS substitution derivatives to nuclear transport inhibitors unveils the specific interaction of individual NLSs with cellular IMPs.** (A) Schematic representation of the GFP-FrAdV1 Pre-pVII fusion proteins expressed. Proteins are represented by a *white horizontal bar* . NLSs are shown as *blue vertical bars* , along with their amino acidic sequence. Individual amino acidic substitutions are indicated by arrows and their sequence is in *red* . The *single letter*amino acid code is used. Basic residues are in *boldface* . (B) The indicated GFP fusion proteins were transiently expressed in HEK293A by means of Lipofectamine 2000 transfection, in the absence (*no add* ) or in the presence of plasmids mediating the expression of either mCherry-Bimax2 *(+ Bimax2* ), RFP-M9M *(+ M9M* ) or DsRed-RanQ69L *(+ RanQ69L* ). 24h post transfection, cells were incubated with DRAQ5 to stain cell nuclei, fixed with paraformaldehyde and processed for CLSM analysis as described in the Materials and Methods section. Representative merged images of 633 nm (*nuclei, blue* ), 488 nm (*GFP, green* ), and 561 nm (*inhibitors, red)*laser lines are shown. Images of individual channels and additional histograms are shown in the Supplementary Figs S3-S10. (C, D) Micrographs such as those shown in (B) were quantitatively analysed to calculate the levels of nuclear (C) or nucleolar (D) accumulation relative to the indicated fusion proteins at the single cell level, as described in the Materials and Methods section. Data shown are individual measurements (circles), along with are means (black horizontal bars) + standard error of the mean (SEM), relative to pooled data from at least 20 cells from three independent experiments. Results from Welch and Brown-Forsythe ANOVA test for accumulation of the indicated GFP fusion protein expressed in the absence or in the presence of the indicated nuclear transport inhibitors. *:$p < 0.05$; **:$p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$; ns: non-significant.

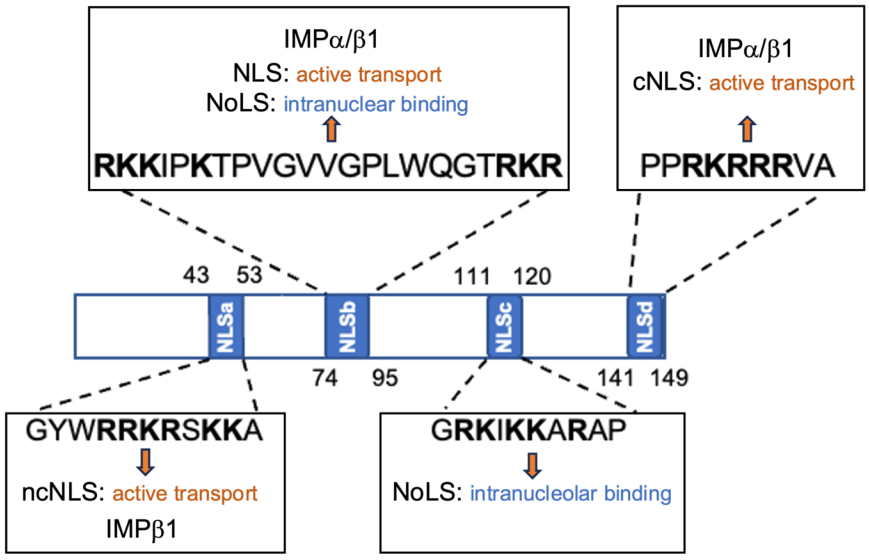


*Figure 9.* **A model for FrAdV1 Pre-pVII nuclear import.**Schematic representation of FrAdV1 Pre-pVII and its targeting sequences. The protein is represented by a *white horizontal bar* . NLSs are shown as *blue vertical bars* , along with their amino acidic sequence. The *single letter* amino acid code is used. Basic

residues are in *boldface* . NLS: nuclear localization signal; cNLS: classical NLS; ncNLS: non classical NLS; NoLS: nucleolar localization signal.