BIOINFORMATIC AND FUNCTIONAL CHARACTERIZATION OF VIRAL GADD34 HOMOLOGS

By

JEANNINE STROUP

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Microbiology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Stefan Rothenburg, Chair

Sean Collins

Barbara Shacklett

Committee in Charge

2022

Acknowledgements

I would like to express my heartfelt gratitude to all those whose support, guidance, advice, and encouragement were absolutely essential to my endurance, happiness, and success over the past three years. First and foremost I would like to give a special thanks to my mentor, Stefan Rothenburg, for believing in my abilities as a scientist even when I did not, and for his dedication to his role as a mentor, prioritizing my academic, professional, and personal growth despite my taking a different path than originally intended. I cannot adequately express how grateful I am to have been supported by such a kind, understanding mentor who treats his students and employees with utmost respect, listening and responding to our individual needs, frequently reminding us to take care of ourselves, and carefully fostering our confidence in ourselves as scientists by taking our ideas seriously.

I would also like to thank my mentor and friend, Greg Brennan, for fostering a fun work environment, providing a safe space for discussion, and teaching me perseverance and adaptability by never letting me off the hook easily. And of course, a huge thank you to all of my wonderful labmates alongside whom I have had the honor of working over the past three years: Julhasur Rahman, Chorong Park, Shefali Banerjee, Huibin Yu, Dewi Megawati, Ryan Bruneau, Ana Stoian, Chi Zhang, and Erin Markle. I thank you each for your patience as I pestered all of you with millions of questions as I found my footing in the lab; for your mentorship, guidance, and support as I navigated personal and professional challenges; and most of all, for your friendship, which made going to work something to look forward to and brought me great joy even in the most trying of times. I have been greatly inspired by each of you.

I would like to thank my committee members, Sean Collins and Barbara Shacklett, for their time and patience in my somewhat last-minute formation of a thesis committee and in reading a cumbersome, oversized master's thesis. You have both been incredibly kind and understanding throughout this process, offering support and guidance as I transition into a new phase of life.

I would also like to thank Renée Tsolis for her continued guidance as my graduate advisor throughout my time here at UC Davis, even throughout what I'm sure was an overwhelmingly busy transition to MGG chair.

A special thank you goes to Karryn Doyle, for always making herself available to answer my many, many questions and being a steadfast resource in navigating the everchanging landscape of graduate school and in particular the convoluted process of leaving with a master's degree. I am grateful to have had a program coordinator who I felt cared about me on a personal level and was always there offer to kindness and warmth in trying times.

I would like to thank the Deans Distinguished Graduate Fellowship (DDGF) for funding me and offering me an extra layer of freedom in choosing a lab and project.

Lastly, a huge thank you to my parents, siblings, and the many friends who have supported me throughout my time here in Davis, notably Julia, Taylor, Aashrita, Quimby, Danielle, Priscilla, Nitin, Luis, Chailah, Bianca, and Christina.

Abstract

Cells respond to a variety of stresses by phosphorylating the alpha subunit of eukaryotic initiation factor 2 (eIF2α) and thereby shutting off cap-dependent protein synthesis. During viral infection, eIF2α is phosphorylated by the double-stranded RNA-activated protein kinase R (PKR) and the resulting translational shut off inhibits viral replication. To overcome the antiviral effects of PKR, viruses have evolved a repertoire of PKR inhibitors, many of which play a role in determining host tropism. In humans, eIF2α is dephosphorylated by two phosphatase complexes, containing paralogous protein phosphatase 1 (PP1) cofactors GADD34 and CReP, which target PP1 phosphatase activity by directly binding eIF2α. Herpes simplex virus (HSV) 1, African swine fever virus (ASFV) encode homologs of GADD34/CReP, which also form eIF2α dephosphorylation complexes with PP1. Here, we use basic local alignment (BLAST) searches to identify additional GADD34 homologs in 39 viruses from 10 different viral families. These viral GADD34 homologs are diverse in amino acid sequences, but all contain predicted PP1 and eIF2α binding motifs. Phylogenetic analysis shows that viral GADD34 homologs separate into two clades – a GADD34-like clade and a CReP-like clade. Additionally, several GADD34 homologs from arthropod infecting viruses clade with cellular GADD34 homologs from their host species, providing support for the idea that these viral GADD34 homologs are derived from host gene acquisition events. We also show that viral GADD34 homologs from HSV1, ASFV, canarypox virus, a newly isolated puffinpox virus, a fruitbat infecting herpesvirus, and an amphibian infecting picornavirus show some ability to counteract PKR in mammalian cells. More specifically, these GADD34 homologs rescue replication of a highly attenuated vaccinia virus (VACV) in PKR-competent cells. Lastly, there appears to be both virus and host cell line specificity in the anti-PKR activity of viral GADD34 homologs.

Table of Contents

List of Figures

List of Tables

CHAPTER ONE: BIOINFORMATIC IDENTIFICATION AND CHARACTERIZATION OF VIRAL GADD34 HOMOLOGS

Jeannine Stroup[1], Huibin Yu[2], Stefan Rothenburg[1]

1. Department of Medical Microbiology and Immunology, School of Medicine, University of California, Davis, California, USA
2. Department of Immunology, School of Medicine, Yale University, Connecticut, USA

ABSTRACT

Cells respond to a variety of stresses by phosphorylating the alpha subunit of eukaryotic initiation factor 2 (eIF2) and thereby shutting down cap-dependent protein synthesis. During viral infection, eIF2α is phosphorylated by the double-stranded RNA-activated protein kinase R (PKR) and the resulting translational shut off inhibits viral replication. In humans, eIF2α is dephosphorylated by two phosphatase complexes, containing paralogous protein phosphatase 1 (PP1) cofactors GADD34 and CReP, which target PP1 phosphatase activity by directly binding eIF2α. Herpes simplex virus (HSV) 1 and African swine fever virus (ASFV) encode homologs of GADD34/CReP, which also form eIF2α dephosphorylation complexes with PP1. These viral GADD34 homologs can be thought of as indirect PKR antagonists. Here, we use basic local alignment (BLAST) searches to identify additional GADD34 homologs in 37 viruses from 10 different viral families. These viral GADD34 homologs are diverse in amino acid sequences, but all contain predicted PP1 and eIF2α binding motifs. Sequence analysis indicates that viral GADD34 homologs separate into two clades – a GADD34-like clade and a CReP-like clade. This provides support for the idea that these viral GADD34 homologs are derived from independent host gene acquisition events.

INTRODUCTION:

Translational repression is a hallmark of the eukaryotic integrated stress response (ISR) and is facilitated by the phosphorylation of the alpha subunit of eukaryotic initiation factor 2 (eIF2). In mammals, phosphorylation of eIF2α is facilitated by one of four different, stress-activated eIF2α kinases. Increased levels of phosphorylated eIF2α cause a shut off of cap-dependent translation and enact a transcriptional program of stress response genes[1]. During viral infection, eIF2α phosphorylation is mediated by protein kinase R (PKR). PKR is activated upon binding dsRNA, which is produced during the replication cycle of most viruses, and viral replication can be inhibited by the resulting shut off of cap-dependent translation[2].

Additionally, increased eIF2α phosphorylation levels induce translation of activating transcription factor-4 (ATF4) mRNA by enabling the 40S ribosomal subunit to bypass an inhibitory upstream open reading frame 5' to the ATF4 open reading frame. ATF4 induces several stress response genes, the products of which can help cells to recover after the stress stimulus is removed1. One such protein, known in humans as growth arrest and DNA damage-inducible protein 34 (GADD34, also known as Protein Phosphatase 1 Regulatory Subunit 15A (PR15A)), alleviates translational repression by facilitating dephosphorylation of eIF2α[3,4]. GADD34 complexes with the catalytic subunit of protein phosphatase 1 (PP1) to target it to eIF2α[5]. Human cells also encode a paralog of GADD34 known as constitutive repressor of eIF2α phosphorylation (CReP, also known as Protein Phosphatase 1 Regulatory Subunit 15B (PR15B)), which is constitutively expressed and is believed to maintain basal levels of eIF2α phosphorylation in unstressed cells[6]. Interestingly, the two paralogs have only 15% pairwise amino acid sequence identity, with most of the sequence identity being found within a 19 amino acid region in the C-terminus of each protein. This region was shown to be important for GADD34's interaction with and dephosphorylation of eIF2α in yeast and will hereafter be referred to as the eIF2α binding motif[7]. In addition, the C-termini of both proteins also contain two adjacent PP1 binding motifs, the RVxF and $\phi\phi$ motifs, located slightly N-terminal to the eIF2α binding motif. For both GADD34 and CReP, crystal

structure analysis of this small PP1 binding region bound to the catalytic subunit of PP1 has provided structural insight into the interactions facilitating the formation of PP1-GADD34 and PP1-CReP holoenzymes[8,9].

Prior to the identification of GADD34 in humans, studies of homologous proteins in mice and hamsters paved the way for understanding GADD34's multifaceted role in the mammalian stress response [10,11,12]. In addition, a mutant form of the rat homolog of GADD34 has been implicated in transformation and tumor progression in rats[13] . While CReP has been studied far less extensively than GADD34, studies in mice and mouse cell lines have revealed the importance of the mouse CReP homolog in embryonic development, obesity, and dampening of the integrated stress response [14, 15, 6]. While functional characterization of GADD34 and CReP has thus far been limited to human and rodent homologs, it is inferred that homologs of both proteins exist in all mammals as well as in other animal taxa.

Some viruses encode proteins that function similarly to GADD34 and CReP, scaffolding the interaction between PP1 and eIF2α. This targeted dephosphorylation of eIF2α counteracts PKR- and other eIF2a kinase-mediated translational repression and allows for translation of viral proteins. The most extensively studied of these GADD34-like viral proteins is the herpes simplexvirus 1 (HSV1) neurovirulence protein ICP34.5, more commonly known as γ34.5 [16,17]. The African swine fever virus (ASFV) protein DP71L is also known to reduce eIF2α phosphorylation in infection via the same mechanism [18,19]. For both HSV1 γ34.5 and ASFV DP71L, this PP1-eIF2α scaffolding function was mapped to the carboxyl terminus, which in both proteins is homologous to small regions in the carboxyl termini of GADD34 and CReP [20, 21, 7, 22]. PP1C and eIF2α binding motifs have been identified in both proteins based on sequence homology with corresponding motifs in GADD34, CReP, and other PP1- and eIF2α- binding proteins [7].

Previously, basic local alignment (BLAST) searches have identified additional protein sequences with predicted PP1 and eIF2α binding motifs in many other viruses, including canarypox virus (CNPV), a macropodid-infecting herpes virus (MaHV), herpes simplexvirus 2 (HSV2), amsacta moorei entomopoxvirus "L" (AmPV), wiseana iridescent virus (also known as invertebrate iridovirus 9 – IIV9), anopheles minimus irodovirus (AMIV), trichoplusia ni ascovirus 2c (TNAV2c), choristoneura occidentalis granulovirus (ClasGV-B), erinnys ello granulo virus (EreIGV), glossina pallidipes salivary gland hypertrophy virus (SGHV), invertebrate iridescent virus 22 (IIV-22), invertebrate iridescent virus 25 (IIV-25), and invertebrate iridescent virus 30 (IIV-30) [7]. For simplicity, this group of proteins will hereafter be referred to as viral GADD34 homologs. Here, we present the bioinformatic analysis of GADD34 homologs encoded by 37 different viruses from 10 different viral families.

MATERIALS AND METHODS

Bioinformatic exploration of animal GADD34 homologs:

To explore the presence of GADD34/CReP homologs among animal species, PR15A and PR15B amino acid sequences from many different animal species were used as BLAST queries against the GenBank database for each animal phylum listed in table 1, and then against each vertebrate class discussed. Queries consisted mainly of PR15A/PR15B sequences trimmed to the conserved, C-terminal 100-200 aa's. For our purposes, only hits with homology to the conserved PP1 and eIF2α binding regions were categorized as GADD34/CReP homologs. For context, the number of fully annotated genomes in GenBank per taxon evaluated was also noted. To explore homology among animal PR15A/PR15B from various taxa, multiple sequence alignments (MSAs) were generated using the MUSCLE alignment program in the EMBL-EBI suite.

Homology-based identification of viral GADD34 homologs:

To identify viral GADD34 homologs, both pBLASTs and psiBLASTs as well as nucleotide BLASTs were performed using full length PR15A and PR15B amino acid and nucleotide sequences as queries against each major viral family in the GenBank database. The same set of BLAST searches were performed using only the conserved C-terminal region containing the PP1 and eIF2α binding motifs. BLAST searches using queries limited to just the region spanning the PP1 and eIF2α binding motifs resulted in a large list of hits cluttered with proteins which did not appear to be true GADD34/CReP-like proteins and with large E values, so a larger region extending 50-100 aa's further in the N-terminal direction (100-200 aa's total) was used. After collecting the sequences of all viral protein hits with significant E values, a similar set of searches was performed using each viral protein as query. GADD34/CReP homologs from hosts of identified viruses were also used as queries. Additionally, genome annotations and ORFs for several viruses were manually inspected and literature searches were performed and to add context to the results presented here.

Multiple sequence alignments and phylogenies:

Multiple sequence alignments (MSAs) of viral GADD34 homologs were generated using the MUSCLE alignment program in the EMBL-EBI suite. MSA's shown were hand curated for more realistic alignments, insertions, and deletions. For the generation of the phylogeny in figure 3, amino acid sequences were trimmed to two aa's N-terminal to the RVxF site and 16 aa's C-terminal to the eIF2α binding motif, based on a MUSCLE MSA of all viral GADD34 homologs.

RESULTS:

Exploration of GADD34/CReP homologs throughout the animal kingdom

We used protein BLAST searches to explore the existence of GADD34/CReP homologs throughout the animal kingdom. We identified open reading frames (ORF) with homology to the

conserved C-terminal regions of GADD34 and CReP in the genomes of at least one species in the following phyla: *Porifera* (sponges), *Cnidaria*, *Mollusca*, *Annelida* (segmented worms), *Arthropoda*, *Echinodermata* as well as in the chordate subphyla *Urochordata* (tunicates) and *Vertebrata* (**Table 1).** Interestingly, despite an abundance of available genome data, proteins with homology to the conserved C-terminal region of GADD34/CReP were not identified in any *Platyhelminth* (flatworm) or *Nematoda* (roundworm) species. Proteins with homology to the conserved C-terminal region of GADD34/CReP were neither identified the third chordate subphylum: *Cephalochordata* (lancelets), nor in any *Bryozoa* species. However, it should be noted that only six and three genomes, respectively, are available in the GenBank nucleotide collection for *Cephalochordata* and *Bryozoa* species, making it possible that the absence of GADD34/CReP homologs in these phyla is merely an artifact of a dearth in available genome data. It is also possible that the species of a given taxa for which genome data is available do not encode GADD34/CReP homologs, but other species within the taxa for which genome data is not yet available *do* encode GADD34/CReP homologs, or that these homologs diverged enough to elude identification. Lastly, as extremely high sequence variation is seen among animal GADD34/CReP homologs, it is possible that GADD34/CReP homologs do exist within these phyla which are so divergent in sequence from other GADD34/CReP homologs that they are undetectable by the BLAST algorithm. This distribution of GADD34/CReP homologs among animal phyla matches that reported by Rojas et al. from similar BLAST search exploration[7].

***Table 1**: Identification of GADD34 and CReP homologs among animal phyla*

Group	Phylum	Approx. number of species	Number of genomes in GenBank	Common name(s)	Number of homologs per species
Non-bilateral	Porifera	5,000	4	Sponges	1
Non-bilateral	Cnidaria	11,000	75	Jellyfish, coral sea anemones	1
Lophotrochozoa	Platyhelminthes	25,000	50	Flatworms	0
Lophotrochozoa	Mollusca	112,000	91	Mollusks	1
Lophotrochozoa	Bryozoa	5,000	4		0
Lophotrochozoa	Annelida	17,000	20	Segmented worms	1
Ecdysozoa	Nematoda	1,000,000	152	Round worms	0
Ecdysozoa	Arthropoda	>1,100,000	1861		1
Deuterostomia	Echinodermata	7,000	27		1
Deuterostomia	Chordata	60,000	2168		
	Urochordata		17	Tunicates	1
	Cephalochordata		6	Lancelates	0
	Vertebrata		2145		2

The GADD34/CReP homologs we identified are extremely diverse in polypeptide length and amino acid sequence. Importantly, like GADD34 and CReP, these proteins possess the sequences of common PP1-binding motifs N-terminal to a 15-25AA stretch which displays homology to the eIF2α-binding motif in GADD34 and CreP[7]. Sequence identities of GADD34/CReP homologs from different phyla are mainly limited to a 40-50aa stretch encompassing these predicted PP1 and eIF2α binding motifs.

The presence of both a GADD34 and a CReP homolog is only seen in bony fish and higher vertebrates

We also examined whether other animal species possess both a GADD34 homolog and a CReP homolog or only a single GADD34/CReP homolog. A single GADD34/CReP homolog was found in all invertebrate phyla as well as in the chordate subphylum *Urochordata* (tunicates). The majority of these are labeled "uncharacterized protein" on GenBank and have not been assigned a name based on homology to

any known proteins. These proteins are so divergent in sequence from other PR15A and PR15B homologs that it is difficult to confidently classify them as PR15A homologs or PR15B homologs.

While still divergent in sequence, for GADD34/CReP homologs from vertebrate species the homology extends further in the N-terminal direction from the universally conserved 40-50aa stretch homologous to the PP1 and eIF2α binding region in GADD34 and CReP. Among vertebrate GADD34/CReP homologs, there is also an obvious distinction between proteins more similar in sequence to GADD34 and more similar in sequence to CReP. Therefore vertebrate GADD34/CReP homologs will hereafter be referred to separately as GADD34 homologs and CReP homologs. Interestingly, both a GADD34 homolog and a CReP homolog were not identified for all vertebrate species. While mammals, reptiles, amphibians, and ray-finned and lobe-finned bony fish possess both a GADD34 homolog and a CReP homolog, only one homolog per species was identified in cartilaginous and jawless fish. Many of the homologs identified in cartilaginous fish are labeled "PR15B-like" on GenBank and form longer alignments and have higher sequence identity with CReP homologs than with GADD34 homologs; Only one homolog was identified in jawless fish, and it has comparable sequence identity with GADD34 and CReP.

Exploration of GADD34/CReP homologs among viruses

We performed an extensive BLAST exploration of GADD34/CReP homologs (referred to here as GADD34 homologs for simplicity) among viruses. In total, we identified GADD34 homologs encoded by several herpes simplexviruses, poxviruses, iridoviruses, and baculoviruses as well as a single member of the following families: *Afsaviridae, Picornaviridae, Adintoviridae, Ascoviridae, Hytrosaviridae,* and *Caulimoviridae*. These viruses infect a diverse set of hosts encompassing many vertebrate and invertebrate species. This list includes well studied proteins HSV1 γ34.5 and ASFV DP71L, as well as the proteins identified by Rojas et al. in BLAST searches using the GADD34 eIF2α binding motif as query[7]. It also includes four additional proteins from arthropod-infecting viruses which have yet to be annotated based

on homology to known proteins. Excluding HSV1 γ34.5 and ASFV DP71L, and functional characterization performed in yeast by Rojas et al, on homologs from CNPV, MaHV, and AmPV, to our knowledge the remainder of these proteins are only putative proteins predicted from computational analysis of full genome sequences and have yet to be functionally characterized.

Viral GADD34 homologs display sequence homology to PP1 and eIF2α binding motifs in GADD34 and CReP

In general all viral GADD34 homologs identified are much shorter in polypeptide length than cellular GADD34/CReP homologs. The sequences of homologs from different viral families are strikingly different from one another (20-40% aa identity) as well as from cellular GADD34 homologs – including those of their hosts. For the most part, such high sequence divergence precludes any confident classification of these viral proteins as more GADD34-like or more CReP-like, but there are several key positions which for certain viral families show higher sequence conservation with either GADD34 or CReP.

GADD34 and CReP are a large, proline-rich proteins which are predicted to be largely unstructured[8,9]. GADD34 has been better characterized than CReP; it has an N-terminal ER localization helix and contains four central PEST motifs[7,8]. These are a series of four sequences rich in proline (P), glutamic acidic (E), serine (S), and threonine (T) residues. The role of PEST motifs is largely unknown, but they are common among proteins with a high turnover rate and are thus suspected of being important in protein turnover[23]. It has also been proposed that at least one of the GADD34 PEST motifs is important for increasing the specificity of the PP1-GADD34 holoenzyme for eIF2α[8]. None of the viral GADD34 homologs show homology to GADD34's N-terminal ER localization region or to its central PEST domains. Importantly, all viral GADD34 homologs identified do contain a 15-20aa region homologous to the GADD34/CReP eIF2α binding motif. Alignments between viral and cellular GADD34/CReP

homologs as well as among viral GADD34 homologs from different families are mostly limited to this predicted eIF2α binding region. Schematics annotated with the positions of the predicted eIF2α binding motif, as well as a few other regions of homology, for one or two representative GADD34 homologs from each viral family are shown in **figure 1**.

GADD34 possesses two adjacent C-terminal alpha helices (582WEQLARDRS590 and 610AARARA WARLRN621), which have been identified via NMR analysis. Crystal structure analysis of PP1 bound to a small fragment of GADD34 indicates that these two alpha helices do not interact directly with PP1[8]. It was therefore speculated that they instead function to recruit additional proteins, such as other PP1 substrates or additional PP1 regulatory subunits, to the PP1-GADD34 complex. The eIF2α binding motif, identified in the same year as the PP1-GADD34 crystal structure was published, does indeed span alpha helix 1, meaning that the sequence for alpha helix 1 is more or less conserved among viral GADD34 homologs[7]. While structural analysis has not been performed on any of these viral homologs, it is possible based on homology in this eIF2α binding region that they too form an alpha helix here. Several herpesvirus GADD34 homologs also display homology to GADD34 in the region encompassing alpha helix 2, and therefore might also form a second alpha helix here. CReP is predicted to form one amphipathic C-terminal alpha helix based on a commonly recognized sequence element. This sequence spans positions corresponding to those in alpha helix 2 of GADD34 and is a suspected binding site for G-actin, which has been shown to increase specificity of the PP1-CReP holoenzyme for eIF2α[9]. Interestingly, little to no homology to this region is found among the C-terminal sequences of viral GADD34 homologs and in fact, excluding herpesviruses, viral homologs diverge substantially in sequence from both GADD34 and CReP directly C-terminal to the eIF2α binding region. This far C-terminal region in viral homologs is generally conserved among proteins from the same viral family but differs greatly between families (**figure 2**).

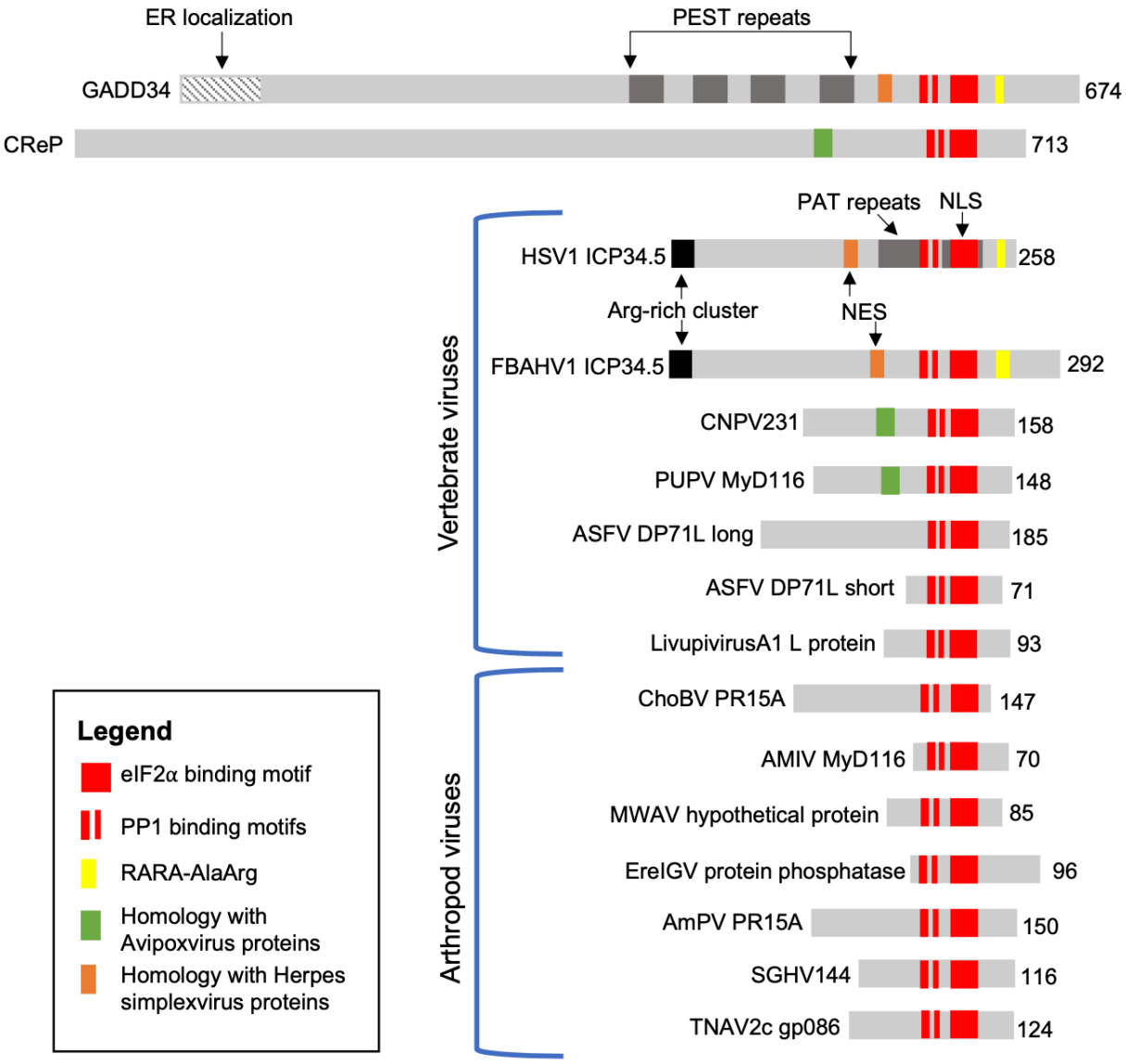

Figure 1: *Schematic representative of GADD34, CReP, and GADD34 homologs from 10 different viral families, to scale. Various sequence motifs are highlighted as described in the legend. Of note, all viral GADD34 homologs possess predicted PP1 binding motifs (red) and show homolog to GADD34 and CReP at the eIF2α binding motif (red). Virus abbreviations: HSV1 (herpes simplex virus 1); FBAHV1 (fruitbat alphaherpesvirus 1); CNPV (canarypox xvirus); PUPV (puffinpox virus); ASFV (African swine fever virus); ChoBV (Chionoecetes opilio bacilliform virus); AMIV (anopheles minimus iridovirus); MWAV (megastigmus wasp adintovirus); EreIGV (Erinnyis ello granulovirus); AmPV (amsacta moorei entomopox virus); SGHV (Glossina hytrosavirus); TNAV2c (Trichoplusia ni ascovirus 2c).*

N-terminal to the proposed eIF2α binding motif, two known PP1 binding motifs – the RVxF and ϕϕ motifs – can be found in all viral GADD34 homologs. The RVxF motif is considered the principal PP1

binding motif[24]. As can be seen in a multiple sequence alignment (MSA) of the conserved C-terminal region of GADD34, CReP, and all viral GADD34 homologs, the positions within viral proteins corresponding the RVxF sequence in GADD34 and CReP for the most part adhere to the RVxF consensus sequence: K/R V/I x F/W, with x being any residue other than F, I, M, Y, N, or P[24](**figure 2**). Several viral homologs stray from the consensus in the first position, having a threonine, histidine, or serine in place of arginine or lysine. In contrast, positions two and four are highly conserved among GADD34 homologs, with every protein possessing a phenylalanine in position 4 and all but two proteins (AMIV and ChoBV) possessing a valine in position 2; AMIV and ChoBV instead possess isoleucine here, which is still in agreement with the RVxF consensus. The second and fourth positions of the motif have been shown in several PP1-interacting proteins (PIPs), including GADD34 and CReP, to engage a hydrophobic pocket on the binding surface of the PP1 catalytic subunit[8,9]. The high sequence conservation among viral GADD34 homologs at these positions allows us to assert that these are bona-fide RVxF motifs.

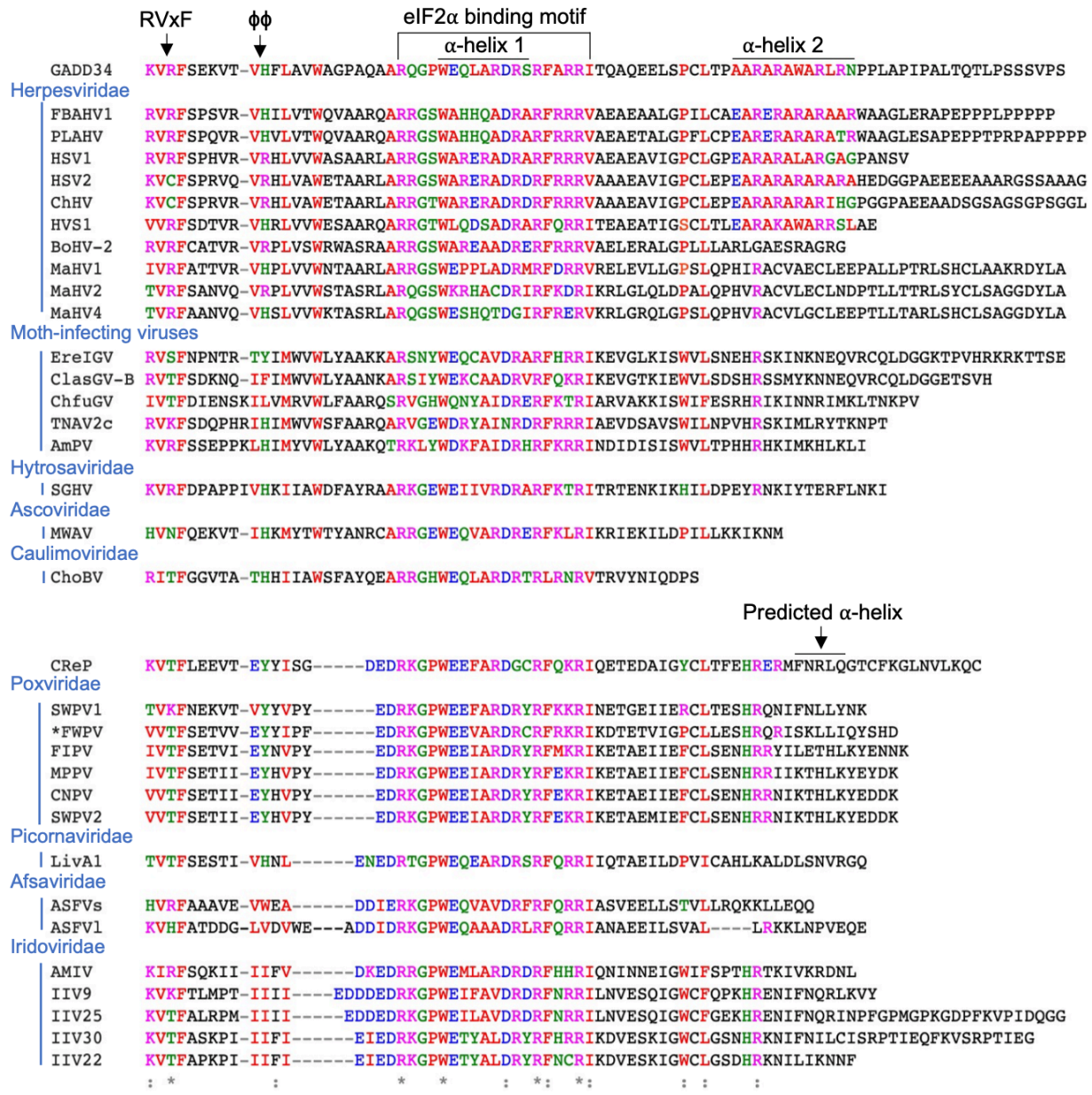

Figure 2: *Hand-curated, MUSCLE multiple sequence alignment (MSA) of GADD34, CReP, and viral GADD34 homologs with non-redundant aa sequences made using EMBL-EBI MSA program. RVxF, ϕϕ, and eIF2α binding motifs are annotated in GADD34, and corresponding residues are shown in color in CReP and viral GADD34 homologs. Secondary structural components are annotated in GADD34 and CReP, and regions are homology in viral proteins are shown in color. Additional conserved residues are also shown in color. Virus abbreviations: Note: the SWPV2 homolog is redundant in sequence with homologs from penguinpox virus 2, mudlarkpox virus, and cheloniid poxvirus 1 (sequences not shown here). Virus abbreviations: FBAHV1 (fruitbat alphaherpesvirus 1); PLAHV (pteropus lylei-associated alphaherpesvireus; HSV1 (herpes simplex virus 1); HSV2 (herpes simplex virus 2); ChHV (chimpanzee herpesvirus); HVS1 (squirrel monkey herpesvirus); BoHV-2 (Bovine herpesvirus 2); MaHV1,2,4 (Macropodid herpesvirus 1, 2,4); EreIGV (Erinnyis ello granulovirus); ClasGV-B (Clostera anastomosis granulovirus B); ChfuGV (Choristoneura fumiferana granulovirus); TNAV2c (Trichoplusia ni ascovirus 2c); AmPV (amsacta moorei entomopox virus); SGHV (Glossina hytrosavirus); MWAV (megastigmus wasp adintovirus); ChoBV (Chionoecetes opilio bacilliform virus); SWPV1,2 (shearwaterpox virus 1,2); FWPV*

(Fowlpox virus isolate 282E4; PUPV (puffinpox virus); FIPV (Finchpox virus); MPPV (Magpiepox virus); CNPV (canarypox xvirus); LivA1 (Livupivirus A1); ASFV (African swine fever virus); AMIV (anopheles minimus iridovirus); IIV (Invertebrate iridovirus)

The second PP1 binding motif, the ϕϕ motif, consists of two adjacent hydrophobic residues and is present in several families of PIPs. It is typically found several residues C-terminal to the RVxF site and has been shown via crystal structure analysis to interact with a deep hydrophobic pocket on the PP1 interaction surface[25,26]. The ϕϕ motif is particularly diverse in sequence even among PIPs within the same family, especially due to the conformational flexibility of the PP1 residue (Tyr78) with which it interacts. Thus far, the following ϕϕ sequences have been identified among three different families of PIPs for which the crystal structures of the PP1 binding region have been solved: II, VI, VF, VY, IF, IY, VH, YF, VR, IR, FH, FF, FY, IQ, FR, VS, VC, IN, HH, IH[25,26]. In GADD34 and CReP, the ϕϕ motif is found 5 residues C-terminal to the RVxF site, and has the following sequences, respectively: VH and EY. Crystal structure analysis has shown that both GADD34 and CReP engage with PP1 with both the RVxF and ϕϕ motifs[8,9]. Despite the CReP ϕϕ motif containing an acidic residue, the tyrosine in position 2 of the CReP motif was shown to occupy a similar position in the PP1-CReP crystal structure to the position 2 residue in the ϕϕ motifs in other PIPs[9]. The corresponding positions in viral GADD34 homologs are highly variable in sequence, with several herpesvirus homologs matching the GADD34 ϕϕ sequence (VH) and almost every poxvirus homolog matching the CReP ϕϕ sequence (EY) (**figure 2**). Homologs from the remaining viral families possess diverse sequences at these positions, some of which match the sequences of previously identified ϕϕ motifs from other PIP families, but many of which (TY, IL, LH, TH, VW, LV, II) may represent previously unidentified ϕϕ motif sequences.

It is typical for PIPs to have several PP1 binding motifs encompassing a large PP1 interacting interface, and substantial variation has been observed in the number, types, and sequences of PP1 binding motifs present in PIPs[24]. Some PIP families possess an Arg motif 7 or 8 residues C-terminal to the RVxF and ϕϕ motifs. CReP, but not GADD34, possesses an arginine in this position; this arginine was

shown via crystal structure analysis to engage a deep pocket on the PP1 binding surface and was thus labeled as a bona fide Arg motif[9]. An arginine is present in a corresponding position in all poxvirus and iridovirus homologs, as well as in the picornavirus homolog and ASFV DP71L. Interestingly, in these proteins this arginine also occupies the first position in the proposed eIF2α binding motif. Other known PP1 binding motifs, such as the SILK motif (G/S I L R/K), which is typically found N-terminal to the RVxF motif, and the myosin phosphatase N-terminal element or MyPhoNE (consensus: R x x Q V/I/L K/R x Y/W) have not been identified in GADD34 or CReP and also do not appear to be present in viral GADD34 homologs[24]. Lastly, there are a number of additional highly conserved positions near the recognized PP1 binding motifs in GADD34, CReP, and viral GADD34 homologs which may also contribute to PP1 binding in GADD34 homologs. Namely, a tryptophan 5 residues C-terminal to the $\phi\phi$ motif is highly conserved in GADD34, herpesvirus homologs, and homologs from all arthropod-infecting viruses excluding iridoviruses. We speculate that this tryptophan may partake in a pi-pi or hydrophobic interaction with a residue on the PP1 binding surface, thus contributing to PP1 binding stabilization.

***γ34.5 homologs are present in the genomes of some herpes simplexviruses but not others*:**

Our BLAST searches identified homologs of HSV1 γ34.5 in 9 additional herpesviruses, all of which are classified as simplexviruses except for a flying fox-infecting herpesviruses which has yet to be formally classified by the ICTV, but which clades with other simplexviruses in phylogenies (**Table 2**)[27]. Like γ34.5, these γ34.5 homologs are encoded by the RL1 gene and will hereafter be referred to as the formal protein name for γ34.5: ICP34.5. Herpes simplex virus 2 (HSV2) ICP34.5 has been studied for many years and is known to play an analogous role to HSV1 γ34.5 in conferring neurovirulence in HSV2 infection[28]. While infection with a HSV2 ICP34.5 null mutant virus results in increased levels of eIF2α phosphorylation, the role of HSV2 ICP34.5 as a scaffold in the PP1-eIF2α complex has not been directly shown as it has been for HSV1 γ34.5[29,28]. The remaining 8 herpesvirus ICP34.5 homologs are putative

proteins predicted via genome annotations and have yet to be functionally characterized. The viruses encoding these homologs include chimpanzee herpesvirus, squirrel monkey herpesvirus (Samiriine herpesvirus 1), bovine herpesvirus 2, three macropodid-infecting herpesviruses, and two and fruitbat-infecting herpesvirus[30, 27, 31, 32, 33]. Interestingly, the RL1 gene is not present in the genomes of all simplexviruses. Genome sequences from herpes B virus, simian agent 8, spider monkey alphaherpesvirus 1, and leporid herpesvirus 4 are lacking the RL1 gene which encodes ICP34.5[34, 35, 36, 37, 38]. Of note, ICP34.5 is present in all three simplexviruses known to infect *Hominidae* species (HSV1, HSV2, and CHeV) and absent from all simplexviruses known to infect old world monkeys. Interestingly, ICP34.5 is present in one new world monkey infecting simplexvirus (HVS1) but is absent in the other (spider monkey alphaherpesvirus 1), despite the two viruses belonging to the same clade according to phylogenetic analysis. Similarly, ICP34.5 is present in BoHV-2 but absent in leporid herpesvirus-4, despite the two viruses being relatively closely related based on genome sequence analysis[33]. The presence or absence of the RL1 gene in herpes simplexvirus genomes does not correlate with the presence or absence of US11: another well studied PKR inhibitor (**table 2**).

Table 2: *Presence of the RL1 gene encoding neurovirulence protein ICP34.5 in herpes simplexvirus genomes*
**virus has not yet been accepted as a separate species by ICTV*
***virus is technically unclassified on ICTV*

Virus	Host	ICP34.5?	Protein accession	Viral genome	US11?
Herpes simplex virus 1 (HSV1)	Human (*Homo sapiens*)	Y	ATI21962.1	Many	Y
Herpes simplex virus 2 (HSV2)	Human (*Homo sapiens*)	Y	QAU11051.1	Many	Y
Chimpanzee herpesvirus (ChHV)	Chimpanzee (*Pan troglodytes*)	Y	YP_009010985.1	PMID: 23508549	Y
Herpes B virus (HVB)	Macaques (*Macaca sp)*	N		PMID: 12743273	Y
Simian agent 8 (SA8)	Baboon (*Papio sp*)	N		PMID: 15629785	Y
Squirrel monkey herpesvirus 1 (HVS1)	Squirrel monkey (*Saimiri sp*)	Y	YP_003933841.1	PMID: 21130483	N
Spider monkey alphaherpesvirus 1 (HVA1)	Spider monkey (*Ateles sp*)	N		PMID: 28160144	N
Bovine herpesvirus 2 (BoHV-2)	Cattle (*Bos tauris*)	Y	QPO25207.1	PMID: 33315144	Y
Leporid herpesvirus 4 (LHV-4)	Domestic rabbits (*Oryctolagus cuniculus domesticus)*	N		PMID: 22921533	N
Macropodid herpesvirus 1 (MaHV-1)	Eastern grey kangaroo (*Macropus giganteus*)	Y	YP_009227215.1	PMID: 26800886	N
Macropodid herpesvirus 2 (MaHV-2)	Black forest wallaby (*Dorcop sis atrata*)	Y	QOD40248.1	Unpublished: MT900475.1	N
*Macropodid herpesvirus 4 (MaHV-4)	Eastern grey kangaroo (*Macropus giganteus*)	Y	QOD40177.1	Unpublished: MT900474.1	N
Fruit bat alphaherpesvirus 1 (FBAHV1)	Fruit bat (*Pteropus sp.*)	Y	YP_009042062.1	PMID: 24942567	N
**Pteropus lylei-associated alphaherpesvirus (PLAHV)	Lylei's flying fox (*Pteropus lylei*)	Y	BBM13171.1	PMID: 32669329	N

Presence of ICP34.5 homologs in simplexviruses

Multiple sequence alignments (MSA) of all 37 viral GADD34 homologs indicate that herpesvirus ICP34.5 homologs share between 22.9% (FBAHV1 with MaHV-2) and 82.9% (FBAHV1 and PLAHV) amino acid identity, with the most sequence conservation seen surrounding the eIF2α binding motif (**Figure 2, Table 3**). Generally, ICP34.5 homologs from macropodid infecting viruses and BoHV-2 are quite divergent in sequence compared to the other ICP34.5 homologs, whereas homologs from most of the primate-infecting viruses and both bat infecting viruses are more similar in sequence and even occupy the same clade in a neighbor joining tree constructed from a MUSCLE MSA of full length herpesvirus ICP34.5 amino acid sequences. Of the ICP34.5 homologs from macropodid infecting herpesviruses, the homologs from MaHV2 and MaHV4 are most similar to each other, sharing 82.9% amino acid identity, while MaHV2 and MaHV4 both share about 60% identity with MaHV1. The BoHV-2 homolog has less than 50% sequence identity with all other herpesvirus homologs. Of the homologs from primate infecting viruses, ICP345 from HSV2 and ChHV are most similar to each other, with 80.9% identity. HSV2 and ChHV homologs have 60% and 63.3% identity with HSV1 γ34.5, respectively (**table 3**). Of note, the ChHV homolog has an 80 aa C-terminal extension not present in any of the other homologs (**Figure 2**). The homologs from the two fruitbat infecting herpesviruses, FBAHV1 and PLAHV, share 81.8% identity with each other, and 46.1% and 44.3%, respectively, with HSV1 γ34.5, with which they share the highest sequence similarity of the primate infecting herpesviruses. Interestingly, the squirrel monkey herpesvirus (HVS1) ICP34.5 homolog is perhaps the most divergent in sequence of all the herpesvirus homologs, forming its own clade in the herpesvirus ICP34.5 phylogeny, and sharing no more than 45% identity with all other herpesvirus homologs, with the exception of the ChHV homolog, with which it shares only 63.3% identity (**table 3**).

Table 3: *Percent identity matrix based on MUSCLE sequence alignment of full-length aa sequences of all viral GADD34 homologs identified*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1: HVS1	100.00	14.48	20.43	20.00	17.36	15.07	15.91	20.00	17.95	17.95	20.19	17.60	27.14	22.68	22.45	29.79
2: ChoBV	14.48	100.00	25.97	17.86	17.36	18.63	16.24	20.48	16.82	15.89	17.24	13.64	32.73	18.75	18.52	19.27
3: Liv	20.43	25.97	100.00	27.27	28.41	32.94	32.95	31.76	30.68	29.55	23.86	22.99	26.56	23.81	20.45	23.91
4: SWPV1	20.00	17.86	27.27	100.00	39.00	44.33	42.00	45.65	44.00	43.00	30.00	31.31	31.43	31.18	30.21	19.19
5: *FWPV	17.36	17.36	28.41	39.00	100.00	47.90	42.14	50.00	48.39	48.39	33.02	30.40	34.29	27.37	28.28	18.18
6: PUPV	15.07	18.63	32.94	44.33	47.90	100.00	75.00	86.02	69.59	68.24	33.66	28.07	35.29	28.26	28.12	15.13
7: FIPV	15.91	16.24	32.95	42.00	42.14	75.00	100.00	81.25	66.46	65.19	31.43	26.89	34.29	28.42	26.00	15.15
8: MPPV	20.00	20.48	31.76	45.65	50.00	86.02	81.25	100.00	92.71	91.67	31.63	29.59	37.88	29.07	28.72	17.82
9: CNPV	17.95	16.82	30.68	44.00	48.39	69.59	66.46	92.71	100.00	98.10	30.77	27.73	35.71	27.37	28.28	13.49
10:SWPV2	17.95	15.89	29.55	43.00	48.39	68.24	65.19	91.67	98.10	100.00	30.77	27.73	35.71	27.37	28.28	13.49
11:IIV30	20.19	17.24	23.86	30.00	33.02	33.66	31.43	31.63	30.77	30.77	100.00	56.86	40.00	40.43	38.18	20.69
12:IIV22	17.60	13.64	22.99	31.31	30.40	28.07	26.89	29.59	27.73	27.73	56.86	100.00	41.43	41.94	44.21	18.26
13:AMIV	27.14	32.73	26.56	31.43	34.29	35.29	34.29	37.88	35.71	35.71	40.00	41.43	100.00	42.86	44.29	25.71
14:IIV9	22.68	18.75	23.81	31.18	27.37	28.26	28.42	29.07	27.37	27.37	40.43	41.94	42.86	100.00	63.04	20.83
15:IIV25	22.45	18.52	20.45	30.21	28.28	28.12	26.00	28.72	28.28	28.28	38.18	44.21	44.29	63.04	100.00	17.07
16:MaHV1	29.79	19.27	23.91	19.19	18.18	15.13	15.15	17.82	13.49	13.49	20.69	18.26	25.71	20.83	17.07	100.00
17:MaHV2	24.00	15.25	27.17	23.23	16.92	16.67	14.89	17.82	14.81	14.81	19.83	19.05	28.57	20.83	14.63	58.82
18:MaHV4	22.67	16.10	25.00	20.20	16.15	14.29	12.06	14.85	12.59	12.59	19.83	17.46	24.29	20.83	14.63	62.94
19:HSV1	29.47	15.56	26.09	17.17	21.58	17.81	15.70	17.65	13.46	13.46	20.75	18.55	24.64	19.79	19.00	33.80
20:HSV2	27.96	14.18	23.66	16.00	19.18	16.89	15.64	18.63	15.82	15.82	17.95	17.60	28.57	21.65	17.74	28.14
21:ChHV	29.03	15.00	24.73	16.00	20.00	16.22	13.48	16.67	14.56	14.56	20.51	18.40	28.57	21.65	21.77	27.65
22:BoHV-2	26.45	17.92	25.27	17.35	19.30	12.21	12.59	15.62	10.64	10.64	14.85	15.97	20.29	15.96	16.84	34.43
23:FBAHV1	29.44	18.06	25.81	18.00	18.18	12.84	12.50	13.73	10.76	10.76	18.42	17.60	22.86	17.35	14.88	26.35
24:PLAHV	28.17	16.78	25.00	17.17	19.01	14.97	14.86	14.85	11.46	11.46	18.97	17.60	24.29	19.39	14.63	26.04
25:MWAV	24.71	33.77	31.58	29.11	34.18	30.26	29.11	30.14	26.58	25.32	22.78	22.78	26.98	19.75	20.25	27.38
26:SGHV	24.59	24.76	21.98	25.00	24.79	29.25	24.79	26.67	29.36	29.36	25.51	22.86	34.29	23.71	21.88	19.30
27:ASFV_s	32.39	25.00	38.46	27.14	31.43	29.85	27.14	27.14	27.14	27.14	32.86	31.43	26.67	33.80	30.99	30.99
28:AmPV	18.37	11.28	20.43	28.00	23.20	23.68	20.17	24.24	21.01	21.01	26.21	24.19	35.71	25.00	22.68	19.01
29:EreIGV	35.62	30.36	27.27	31.82	31.88	32.84	31.43	31.88	30.43	30.43	28.75	33.33	32.31	28.99	24.44	22.11
30:ClasGV-B	22.94	18.48	20.43	26.00	25.71	23.53	23.58	26.04	23.81	23.81	20.35	23.76	30.00	23.71	22.94	20.17
31:ChfuGV	24.21	20.51	17.05	21.59	21.98	28.41	27.17	26.97	25.27	25.27	31.18	29.89	32.35	24.44	24.47	20.41
32:TNAV2c	26.32	20.62	18.18	23.66	23.36	20.00	20.18	21.84	19.44	19.44	24.21	25.49	31.43	26.80	23.66	21.24

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1: HVS1	24.00	22.67	29.47	27.96	29.03	26.45	29.44	28.17	24.71	24.59	32.39	18.37	35.62	22.94	24.21	26.32
2: ChoBV	15.25	16.10	15.56	14.18	15.00	17.92	18.06	16.78	33.77	24.76	25.00	11.28	30.36	18.48	20.51	20.62
3: Liv	27.17	25.00	26.09	23.66	24.73	25.27	25.81	25.00	31.58	21.98	38.46	20.43	27.27	20.43	17.05	18.18
4: SWPV1	23.23	20.20	17.17	16.00	16.00	17.35	18.00	17.17	29.11	25.00	27.14	28.00	31.82	26.00	21.59	23.66
5: *FWPV	16.92	16.15	21.58	19.18	20.00	19.30	18.18	19.01	34.18	24.79	31.43	23.20	31.88	25.71	21.98	23.36
6: PUPV	16.67	14.29	17.81	16.89	16.22	12.21	12.84	14.97	30.26	29.25	29.85	23.68	32.84	23.53	28.41	20.00
7: FIPV	14.89	12.06	15.70	15.64	13.48	12.59	12.50	14.86	29.11	24.79	27.14	20.17	31.43	23.58	27.17	20.18
8: MPPV	17.82	14.85	17.65	18.63	16.67	15.62	13.73	14.85	30.14	26.67	27.14	24.24	31.88	26.04	26.97	21.84
9: CNPV	14.81	12.59	13.46	15.82	14.56	10.64	10.76	11.46	26.58	29.36	27.14	21.01	30.43	23.81	25.27	19.44
10:SWPV2	14.81	12.59	13.46	15.82	14.56	10.64	10.76	11.46	25.32	29.36	27.14	21.01	30.43	23.81	25.27	19.44
11:IIV30	19.83	19.83	20.75	17.95	20.51	14.85	18.42	18.97	22.78	25.51	32.86	26.21	28.75	20.35	31.18	24.21
12:IIV22	19.05	17.46	18.55	17.60	18.40	15.97	17.60	17.60	22.78	22.86	31.43	24.19	33.33	23.76	29.89	25.49
13:AMIV	28.57	24.29	24.64	28.57	28.57	20.29	22.86	24.29	26.98	34.29	26.67	35.71	32.31	30.00	32.35	31.43
14:IIV9	20.83	20.83	19.79	21.65	21.65	15.96	17.35	19.39	19.75	23.71	33.80	25.00	28.99	23.71	24.44	26.80
15:IIV25	14.63	14.63	19.00	17.74	21.77	16.84	14.88	14.63	20.25	21.88	30.99	22.68	24.44	22.94	24.47	23.66
16:MaHV1	58.82	62.94	33.80	28.14	27.65	34.43	26.35	26.04	27.38	19.30	30.99	19.01	22.11	20.17	20.41	21.24
17:MaHV2	100.00	82.87	26.49	26.70	24.58	30.08	21.59	21.91	28.57	20.18	29.58	19.23	22.92	21.01	21.43	22.12
18:MaHV4	82.87	100.00	26.49	25.57	24.58	28.57	23.30	24.16	25.00	19.30	26.76	16.15	19.79	19.33	17.35	18.58
19:HSV1	26.49	26.49	100.00	51.17	52.07	31.82	42.36	42.98	23.81	25.00	26.76	14.69	31.58	19.82	21.65	25.00
20:HSV2	26.70	25.57	51.17	100.00	74.60	32.90	38.59	38.68	22.35	22.76	28.17	15.38	21.05	17.50	22.22	23.48
21:ChHV	24.58	24.58	52.07	74.60	100.00	31.61	36.96	36.68	22.35	24.59	26.76	14.69	23.16	16.67	23.23	23.68
22:BoHV-2	30.08	28.57	31.82	32.90	31.61	100.00	40.00	36.77	27.06	20.79	32.39	14.17	25.35	16.04	19.57	24.49
23:FBAHV1	21.59	23.30	42.36	38.59	36.96	40.00	100.00	81.44	29.41	23.33	28.17	15.44	25.00	18.49	20.20	30.70
24:PLAHV	21.91	24.16	42.98	38.68	36.68	36.77	81.44	100.00	28.24	23.53	25.35	16.22	24.21	18.49	20.20	32.46
25:MWAV	28.57	25.00	23.81	22.35	22.35	27.06	29.41	28.24	100.00	32.94	36.92	30.59	32.81	24.71	29.63	28.24
26:SGHV	20.18	19.30	25.00	22.76	24.59	20.79	23.33	23.53	32.94	100.00	26.76	26.79	33.78	20.00	29.90	28.07
27:ASFV_s	29.58	26.76	26.76	28.17	26.76	32.39	28.17	25.35	36.92	26.76	100.00	28.17	36.07	26.76	28.17	28.17
28:AmPV	19.23	16.15	14.69	15.38	14.69	14.17	15.44	16.22	30.59	26.79	28.17	100.00	45.83	27.78	32.63	34.86
29:EreIGV	22.92	19.79	31.58	21.05	23.16	25.35	25.00	24.21	32.81	33.78	36.07	45.83	100.00	51.19	36.36	43.24
30:ClasGV-B	21.01	19.33	19.82	17.50	16.67	16.04	18.49	18.49	24.71	20.00	26.76	27.78	51.19	100.00	28.28	29.41
31:ChfuGV	21.43	17.35	21.65	22.22	23.23	19.57	20.20	20.20	29.63	29.90	28.17	32.63	36.36	28.28	100.00	38.54
32:TNAV2c	22.12	18.58	25.00	23.48	23.68	24.49	30.70	32.46	28.24	28.07	28.17	34.86	43.24	29.41	38.54	100.00

Several functional motifs have been identified in HSV1 γ34.5. These include: an N-terminal arginine-rich cluster, which has been implicated in nucleolar localization; a series of proline-alanine-threonine (PAT) repeats sometimes referred to as the protein's backbone or central domain, which have been shown to determine subcellular localization of γ34.5; a nuclear export signal; a bipartite nuclear localization signal which, interestingly, spans the eIF2α binding motif; and the AlaArg motif, which lies just C-terminal to the eIF2α binding motif in γ34.5 and is believed to be important in the formation of the

γ34.5-containing eIF2α phosphatase complex[39, 40, 41, 42]. The AlaArg motif is also present in GADD34 and is preceded in both γ34.5 and GADD34 with the sequence: RARA. The RARA sequence has been prescribed a role in PP1 binding for GADD34 but not for γ34.5[42, 5]. For simplicity, these adjacent motifs will be hereafter referred to in combination as the "RARA-AlaArg" motif. Interestingly, the RARA-AlaArg motif in GADD34 is part of a 12 aa stretch which forms alpha helix 2 and may be part of a similar secondary structure in herpesvirus homologs[8]. The positions of aforementioned features are shown on the schematic of γ34.5 in **figure 2.** An N-terminal arginine-rich cluster can be identified in ICP34.5 from HSV2, ChHV, BoHV-2, HVS1, and both fruitbat herpesviruses, but is absent in all three macropodid-infecting virus proteins (**Table 4**). The sequence of the N-terminal arginine-rich cluster differs among ICP34.5 homologs in both the number and positions of arginine residues (**table 4**). γ34.5 from various HSV1 strains containing different numbers of arginine's in this motif have been shown to exhibit distinct subcellular localization pattens, and it is therefore likely that these ICP34.5 homologs from different herpesviruses also show variation in their subcellular localization patterns[40]. Of note, BoHV-2 ICP34.5 has a distinct N-terminal repetition of GR 20 times (**table 4**). Here we have classified this as an N-terminal arginine-rich cluster, but it is possible that this is just an artifact of computational prediction of this protein from genomic sequences. Interestingly, the HSV1 PAT repeats are not present ICP34.5 homologs from any other herpesviruses. As the HSV1 PAT repeats have been implicated as the predominant motif driving subcellular localization patterns of γ34.5, it is possible that subcellular localization of ICP34.5 from other herpesviruses is determined in part by other, yet unidentified sequence motifs[41]. The proposed HSV1 nuclear export signal is highly conserved in sequence in HSV2, ChHV, and the two fruitbat herpesviruses; in contrast, the sequences at the corresponding location in ICP34.5 from HVS1, BoHV-1, and MaHV1 are more divergent from that of HSV1, and almost no conservation is found in this location for MaHV2 and MaHV4, relative to HSV1 (**table 4**). Similar patterns in sequence variation are seen in the proposed bipartite nuclear localization signal, which contains the eIF2α binding motif (**table 4**). The Ala-

Arg motif, which is believed to be essential for formation of an eIF2α phosphatase complex for HSV1 γ34.5, is present in ICP34.5 homologs from HSV2, ChHV, HVS1, and both fruitbat infecting viruses, but is absent from ICP34.5 homologs from BoHV-2 and all three macropodid-infecting herpesviruses (**table 4**). Whether the absence of the Ala-Arg motif from these homologs affects their ability to dephosphorylate eIF2α and counteract PKR has yet to be explored. It should be noted however that HSV1 γ34.5 and MaHV-1 ICP34.5 displayed similar abilities to decrease eIF2α phosphorylation and counteract PKR toxicity in yeast assays[7].

Table 4: *Predicted functional motifs in herpes simplexvirus γ34.5 homologs*

	Motif/region	N term Arg-rich cluster	PAT repeats	NES	NLS	RARA AlaArg
	Prescribed function in HSV1 g34.5	Nucleolar localization	Subcellular localization	Nuclear export	Nuclear localization	Formation of eIF2a phosphatase complex
	Reference(s)	PMID: 12186925 PMID: 11788604	PMID:1178860 PMID:16099898	PMID:12186925	PMID:12186925	PMID:11264356
Sequences	**HSV1**	MARRRRRHRGPRRPRP	PAT	**LPPRLALRLR**	ASAARLARRGSWARERADRARFRRRVAEA	RARAL**AR**
	HSV2	MSRRGPRRRGPRRRPR	N/A	**LPPH**LALRLR	ETAARLARRGSWARERADRDRFRRRVAAA	RARAR**AR**
	ChHV	MSRRGPRRRPRHRPHR	N/A	**LPPH**LALRLR	ETAARLARRGTWARERADRDRFRRRVAAA	RARAR**AR**
	HVS1	MARKSRRRKGCVSRDR	N/A	**LPSW**LARRLL	ESAARQARRGTWLQDSADRARFQRRITEA	RAKA**WAR**
	BoHV-2	MGRGRGRGRGRGRGRG	N/A	**PPPRVALRL**A	RWASRAARRGSWAREAADRERFRRRVAEL	GAESRAG
	MaHV-1	MNTHHQAPSPFRFFIW	N/A	**LPPSL**THHLR	NTAARLARRGSWEPPLADRMRFDRRVREL	RACVAEC
	MaHV-2	MTTRSKPPSPSRFFIW	N/A	VPAAL**THRLQ**	STASRLARQGSWKRHACDRIRFKDRIKRL	RACVAEC
	MaHV-4	MNTRSKPPSPFRFFIW	N/A	VPAAL**THHLQ**	KTASRLARQGSWESHQTDGIRFRERVKRL	RACVAGC
	FBAHV1	MPRRGARRPRRRSVRP	N/A	**LPPRLARRLR**	QVAARQARRGSWAHHQADRARFRRRVAEA	RERAR**AR**
	PLAHV	MPRRGPRRPRRRSVRP	N/A	**LPPRL**VRRLR	QVAARQARRGSWAHHQADRARFRRRVAEA	RERAR**AR**
	GADD34	N/A	N/A	**APPRLPLRLQ**	AGPAQAARQGPWEQLARDRSRFARRITQA	RARA**WAR**
	CReP	N/A	N/A	N/A	YISGDEDRKGPWEEFARDGCRFQKRIQET	N/A

Interestingly, herpesvirus ICP34.5 homologs share an overall higher percent identity with GADD34 than with CReP, particularly within the conserved C-terminal region (**figure 2**). In addition to the RARA-AlaArg motif, GADD34 also displays homology to the proposed nuclear export signal in HSV1 γ34.5[39]. Neither of these motifs are present in CReP.

Figure 3: *Phylogeny generated from an MSA of only the conserved C-terminal regions of several animal GADD34/CReP homologs and all non-redundant viral GADD34 proteins. We are still working on generating a phylogeny with decent support. Will insert here once it's done, but for now the analysis presented here is all based on preliminary phylogenies.*

GADD34 homologs in poxviruses: the avipoxviruses and the green sea turtle poxvirus

In addition to previously identified GADD34 homologs in CNPV and AmPV, our BLAST searches identified GADD34 homologs in 9 additional poxvirus species (**table 5**). Eight of these newly identified homologs are encoded by other members of the Avipoxvirus genus. Avipoxviruses are important avian pathogens, infecting close to 300 species of wild and domestic birds worldwide and posing a substantial threat to several endangered species[43,44,45]. However, moves to better understand and molecularly characterize avipoxviruses have only recently been made. While the full genomes of fowlpoxvirus (FWPV) and canarypoxvirus (CNPV) were published in 2000 and 2004, respectively, full molecular characterization of poxviruses isolated from various bird species has only become standard within the last several years, with the advent widely available and accessible NGS methods [46,47]. Since 2014, whole genome sequences have been acquired for pigeonpox (FeP2), turkeypox (TKPV), two shearwaterpox viruses (SWPV1, SWPV2), two penguinpox viruses (PEPV, PEPV2), flamingpox virus (FGPV), magpiepox virus (MPPV), mudlarkpox virus (MLPV), and finchpox virus (FIPV)[48,43,49,44,50,51,52,53] (**table 5**). Phylogenies generated from concatenations of core genes from these avipoxvirus genomes show that the viruses cluster into three distinct clades: Clade A (FWPV-like): FWPV, PEPV, FeP2, and FGPV; Clade B (CNPV-like): CNPV, SWPV1, SPWV1, MPPV, MLPV, and PEPV2, and clade C (Psittacine)[48]. Clade C currently only contains turkeypox virus, which has a distinctly small genome and is missing several ORFs present in other avipoxviruses[43]. Finchpox virus was only recently isolated and characterized, and has yet to be included in this research, but shares highest genome sequence similarity with CNPV and is thus likely to cluster into clade B[53].

Table 5: *Identification of GADD34 homologs in avipoxviruses with fully sequenced genomes*

Virus	Host	MyD116?	Protein accession	Genome accesion & publication	Year sequenced
Fowlpox virus (FWPV)	Red junglefowl/chicken (*Gallus gallus*)	N	N/A	AF198100, PMID: 10729156	2000
Canarypox virus (CNPV)	Canary (*Serinus canaria*)	Y	NP_955254.1	NC_005309.1, PMID: 14671117	2004
Fowlpox virus (FWPV) FP9, attenuated strain	Red junglefowl/chicken (*Gallus gallus*)	N	N/A	AJ581527.1, PMID: 14769888	2010
Penguinpox virus (PEPV)	African penguin (*Spheniscus demersus*)	N	N/A	PMID: 24919868	2014
Pigeonpox virus (FeP2)	Feral pigeon (*Columba livia*)	N	N/A	KJ801920.1, PMID: 24919868	2014
Turkeypox virus (TKPV)	Turkey (*Meleagris gallopavo*)	N	N/A	NC_028238.1, PMID: 26282613	2015
Shearwaterpox virus-1 (SWPV1)	Flesh-footed shearwater (*Ardenna carneipes*)	Y	ARF02779.1	KX857216.1, PMID: 28407753	2017
Shearwaterpox virus (SWPV2)	Wedge-tailed shearwater (*Ardenna pacifica*)	Y	ARE67466.1	KX857215.1, PMID: 28407753	2017
Flamingopox virus (FGPV)	Lesser flamingo (*Phoenicoparrus minor*)	N	N/A	MF678796.1, PMID: 29207949	2017
*Fowlpox virus (FWPV) isolate 282E4	Red junglefowl/chicken (*Gallus gallus*)	Y	AYP74133.1	Unpublished: MG702259.1	2018
Magpiepox virus (MPPV)	Australian magpie (*Gymnorhina tibicen*)	Y	QGM48865.1	MK903864.1, PMID: 31726310	2020
Fowlpox virus (FWPV-S) standard vaccine strain	Red junglefowl/chicken (*Gallus gallus*)	N	N/A	MW142017.1, PMID: 33620554	2021
Penguinpox virus-2 (PEPV2)	Yellow-eyed penguin (*Megadyptes antipodes*)	Y	QRM15863.1	MW296038.1, PMID: 33525382	2021
Mudlarkpox virus (MLPV)	Magpie-lark (*Grallina cyanoleuca*)	Y	QRM15510.1	MT978051.1, PMID: 33385935	2021
Puffinpox virus (PUPV)	Puffin (*Fratercula spp*)	Y	N/A	In house	2021
Cheloniidpox virus 1 (ChPV)	Green sea turtle (*Chelonia mydas*)	Y	QRI42949.1	MT799800.1, PMID: 33572619	2021
Finchpox virus (FIPV)	House finch (*Haemorhous mexicanus*)	Y	UOX38668.1	OM869483.1, PMID: 35488713	2022

GADD34 homologs were identified in all clade B avipoxviruses, including finchpox virus, but not in any clade A or clade C avipoxviruses. Of note, a GADD34 homolog was also identified in the genome

of folwpox virus (FWPV) isolate 282E4. However, of the 10 FWPV genomes published in GenBank, both before and after the publication of the genome of this isolate, this is the only FWPV genome in which a GADD34 homolog was identified. The genome for isolate 282E4 is also about 10kb larger than any of the other FWPV genomes. Using ncbi's ORF finder tool, we manually scanned the ORFs in other FPWV genomes near the location of this GADD34 homolog in isolate 282E4 and failed to identify any ORFs containing even small stretches of sequences which are highly conserved among the other avipoxvirus GADD34 homologs. Unfortunately, this FWPV genome was a direct submission to GenBank and does not have an accompanying paper to provide more context. While it may be possible that some FWPV strains encode a GADD34 homolog while others do not, it is more likely that this particular isolate is actually a different species of avipoxvirus, which is possibly more closely related to other clade B avipoxviruses which also encode GADD34 homologs, like CNPV, MPPV, MLPV, SWPV1, and SWPV2. If this is the case, this may be the first clade B avipoxvirus isolated from a chicken, and may represent a novel avipoxvirus species. Further molecular analysis is needed to place this virus in an avipoxvirus phylogeny. For now, this virus will be referred to in subsequent sections simply as 282E4.

An additional avipoxvirus GADD34 homolog was identified in the genome of a virus isolated and characterized in-house, from pox-like lesions on an infected puffin. The viral genome was partially sequenced via nanopore sequencing and annotated using the CNPV genome as a reference. Based on the sequences of orthologs of the FWPV 140 gene, it clusters with other clade B avipoxviruses phylogenetic analysis, and appears to be very closely related to Finchpox virus (FIPV) (Rothenburg lab, unpublished data). Whether this puffinpox virus isolate is actually a strain of FIPV requires further comparison of the complete genomic sequences of the viruses. For now, we are provisionally calling this puffinpox virus (PUPV). The GADD34 homolog was identified via its homology with CNPV231 and will be included in the following sequence analysis.

Finally, the 9th additional poxvirus GADD34 homolog identified in our BLAST searches is encoded by cheloniidpox virus 1 (ChPV), a recently characterized virus isolated from an endangered green sea turtle species in Australia. This was the first poxvirus to be isolated from a turtle species, and one of only four poxviruses to be isolated from reptilian species thus far, with the other three being found in various species of crocodile[54]. Interestingly, the ChPV genome shows low sequence identity with the genomes of crododilepox viruses, and instead possess very high sequence identity (89.3%) to the SWPV2 genome. Phylogenies based on alignments of both full genome sequences and concatenations of core genes place ChPV *within* avipoxvirus clade B, along with SWPV2, CNPV, MPPV, and SWPV1[54]. Thus, this green sea turtle pox virus GADD34 homolog is hereafter grouped in with avipoxvirus homologs for further analysis. In contrast, the AmPV GADD34 homolog is strikingly different in sequence from the rest of the poxvirus homologs and is instead more similar in sequence to GADD34 homologs encoded by other insect-infecting viruses from other viral families. This interesting finding will be discussed in more detail in later sections. Besides the assays performed with CNPV231 in yeast, none of the avipoxvirus GADD34 homologs discussed here have been functionally characterized.

Sequence analysis of Avipoxvirus virus GADD34 homologs

Like the herpesvirus GADD34 homologs, there is much higher sequence identity among homologs within the avipoxvirus genus than exists between viral GADD34 homologs from different families, with homology between avipoxvirus GADD34 homologs extending beyond the PP1 and eIF2α binding motifs (**figure 2, figure 4**). Multiple sequence alignments of all viral GADD34 homologs show that the avipoxvirus proteins share between 16% and 100% amino acid sequence identity (**table 3**). Homologs from SPWV2, PEPV2, ChHV, and MLPV are 100% identical in amino acid sequence. These homologs are very similar in sequence to CNPV231, sharing 98.1% sequence identity. This makes sense, as these five viruses are very closely related, forming their own subclade within clade B, separate from SWPV1, and

having above 94% genome sequence identity [54,44]. The GADD34 homolog from SWPV1, which has a lower genome sequence identity with other clade B avipoxviruses, has only 48% and 49% identity with homologs from CNPV and SWPV2/ChHV/PEPV2/MLPV, respectively. The SWPV1 homolog is also much shorter in polypeptide length than most of the other homologs and appears to have an approximately 30-40 aa deletion in between the N-terminal and C-terminal segments of the protein, which align relatively well to the N-terminal and C-terminal segments of the other homologs (**figure 4**). The homolog from MPPV, which based on genomic sequence is very closely related to CNPV, SWPV2, ChHV, PEPV2, and MLPV and falls within the same subclade as these viruses, is slightly more divergent in sequence, having highest sequence identity with CNPV231 (91.2%). The MPPV homolog is similar in polypeptide length to the SWPV1 homolog but appears to be missing the first 40 aa's which are present, albeit not highly conserved, in the other avipoxvirus homologs (**figure 4**). Interestingly, the 282E4 homolog is quite divergent in sequence, only sharing between 44 and 52% identity with the rest of the homologs. The FIPV protein, which has 100% aa identity with the homolog from our PUPV, is longer in polypeptide length than the other homologs and appears to have a 22 aa insertion at position 28 (**figure 4**).

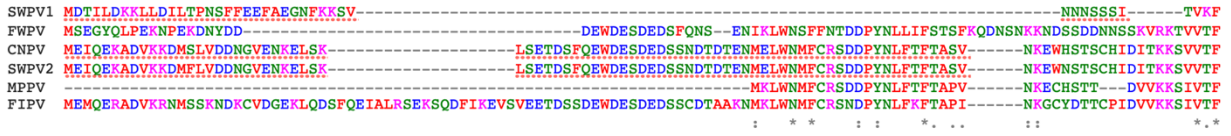

Figure 3: *Hand curated MUSCLE MSA of full length avipoxvirus GADD34 homologs generated using EMBL-EBI alignment program. Residues C-terminal to the RVxF motif not shown here. Note: the SWPV2 sequence is shown here as a representative of MLPV, PEPV2, and ChHV homologs, all of which are redundant in aa sequence with the SWPV2 homolog. PUPV homolog is 100% identical in sequence to FIPV sequence and is therefore not shown here.*

Interestingly, in addition to the eIF2α binding motif, avipoxvirus homologs share an additional 17 aa stretch of homology with CReP. For avipoxvirus homologs, this sequence occurs about 50 aa upstream of the PP1 binding motif, whereas for CReP this sequence occurs about 100 aa upstream of the PP1 binding motif (**figure 5**). This stretch falls within the apparent central deletion for the SWPV1 homolog and is thus

not present in this homolog. This region is weither found in GADD34, nor in the AmPV homolog. These results suggest avipoxvirus homologs may be more closely related to PR15B than PR15A.

```
                    CReP   LWNSFCNSDDPYNPLNF
                    CNPV   LWNMFCRSDDPYNLFTF
                    MPPV   LWNMFCRSDDPYNLFTF
                   *FWPV   LWNSFFNTDDPYNLLIF
SWPV2/MLPV/ChHV/PEPV2   LWNMFCRSDDPYNLFTF
                FIPV/PUPV   LWNMFCRSNDPYNLFKF
```

Figure 4: *sequences of a 17aa region of conservation between CReP and GADD34 homologs from avipoxviruses*

Identification of a GADD34 homolog in a picornavirus: the first of its kind?

Our BLAST searches identified a single GADD34 homolog in a picornavirus. This was the only GADD34 homolog identified in an RNA virus. The homolog is the leader protein from the predicted polyprotein of a recently identified picornavirus isolated from the fecal samples of healthy smooth newts (Lissotriton vulgaris) in Hungary[55]. The homolog is predicted to function similarly to GADD34 based on its homology to CNPV231 and ASFV DP71L and is the first of its kind to be discovered in a picornavirus. The virus, one of only two known amphibian-infecting picornaviruses, forms its own clade in a phylogeny made with several regions of polyprotein sequences from different picornaviruses and based on ICTV criteria has been proposed as a new genus named Livupivirus, after its amphibian host. The virus itself is being called livupivirusA1, and the homolog will hereafter be referred to as L protein. L protein is quite divergent in sequence compared to all other viral GADD34 homologs identified, sharing between 18% and 38% identity with other viral homologs (**table 3**).

When the livupivirus L protein was used as a query against picornaviruses and other RNA virus families, there were no hits with significant E values, and hits with E values lower than the threshold only

made small alignments (5-10aa's) which for the most part fell outside of the PP1 and eIF2α binding motifs. Ampivirus, the second known amphibian-infecting picornavirus which was also isolated from a smooth newt, appears to be somewhat distantly related to livupivirusA1 and does not encode a GADD34 homolog. L protein therefore appears to be the only GADD34 homolog identified in picornaviruses thus far.

Identification and sequence analysis of GADD34 homologs in arthropod infecting viruses from 7 different viral families

The viral GADD34 homologs discussed thus far have all been encoded by vertebrate-infecting viruses. The remaining 13 homologs identified are encoded by arthropod infecting viruses from a diverse set of viral families: Poxviridae, Iridovirae, Baculoviridae, Adintoviridae, Ascoviridae, Hytrosaviridae, and Caulimoviridae (**table 5**). All of these arthropod viruses infect insects (Hexapoda), except for the caulimovirus (ChoBV), which was purified from a snow crab (*Chionoecetes opilio*) [unpublished]. For Adintoviridae, Ascoviridae, and Hytrosaviridae, only a single virus encoding a GADD34 homolog was identified. However, it should be noted there are only 3, 8, and 2 fully sequences genomes available on GenBank, respectively, for each of these viral families. Overall, GADD34 homologs from arthropod infecting viruses are incredibly diverse in amino acid sequence, with even homologs from viruses within the same family having sequence identity as low as 30% (**table 3**).

Table 6: *GADD34 homologs in arthropod-infecting viruses*

Viral family	Virus	Host	Gene name	Protein name(s)	Accession	Viral genome
Adintoviridae	Megastigmatus wasp adintovirus (MWAV)	Douglas-fir seed chalcid (*Megastigmus spermotrophus*)	N/A	Hypothetical protein	DAC81332.1	PMID: 34646575
Iridoviridae	Anopheles minimus iridovirus (AMIV)	Mosquito (*Anopheles minimus*)	Locus tag: DH26_gp081	MyD116-like domain protein	YP_009021158.1	Direct submission: NC_023848.1
	Invertebrate iridescent virus 30 (IIV30)	corn earworm (*Helicoverpa zea*)	161R	Putative protein phosphatase 1 regulatory subunit	YP_009010455.1	PMID: 24394746
	Invertebrate iridescent virus 25 (IIV25)	Blackfly larva (*Simulium sp*)	163R	Putative phosphatase	YP_009010696.1	PMID: 24232916
	Invertebrate iridescent virus 22 (IIV22)	Blackfly larva (*Simulium sp*)	153R	MyD116-like domain protein	YP_008357451.1	PMID: 23804567
	Invertebrate iridescent virus 9 (IIV9)/ Wiseana iridescent virus	Porina (*Wiseana spp*)	Locus tag: WIV_gp103	Hypothetical protein	YP_004732886.1	PMID: 21632757
Ascoviridae	Trichoplusia ni ascovirus 2c (TNAV2c)	Cabbage Looper (*Trichoplusia ni*)	Locus tag: TNAV2c_gp086	Hypothetical protein	YP_803309.1	PMID: 16876847
Hytrosaviridae	Glossina hytrosavirus (SGHV)	Tsetse flies (*Glossina pallidipes*)	Locus tag: SGHV144	Hypothetical protein	YP_001687092.1	PMID: 18272583
Baculoviridae	Erinnyis ello granulovirus (EreIGV)	Cassava hornworm (*Erinnyis ello ello*)	pp-1	Protein phosphatase 1	YP_009091882.1	PMID: 25280947
	Clostera anastomosis granulovirus B (ClasGV-B)	Black-back prominent moth (*Clostera anastomosis*)	Locus tag: D1P97_gp038	Clas38	YP_009505980.1	PMID: 26168260
	Choristoneura fumiferana granulovirus (ChfuGV)	Eastern spruce budworm (*Choristoneura fumiferana*)	N/A	Hypothetical protein	YP_654457.1	Unpublished: NC_008168.1
Caulimoviridae	Chionoecetes opilio bacilliform virus (ChoBV)	Snow crab (*Chionoecetes opilio)*	Locus tag: SCV_094	Hypothetical protein	GAV93214.1	Unpublished: BDLS01000002.1
Poxviridae	Amsacta moorei entomopoxvirus (AmPV)	Red hairy caterpillar (*Amsacta moorei*)	Locus tag: AMV193	PR15A	NP_064975.1	PMID: 10936094

The baculovirus homologs are all encoded by betabaculoviruses, which are in general less well studied than alphabaculoviruses, and have thus far only been identified in Lepidoptera (butterflies and moths) hosts[56]. Of the 26 ICTV accepted betabaculovirus species with completely sequenced genomes,

only 3 species were identified as encoding GADD34 homologs. Two of these viruses, erinnyis ello granulovirus (EreIGV) and clostera anastomosis granulovirus B (ClasGV-B), are very closely related, forming their own subclade within the group b betabaculoviruses in a phylogenetic tree generated from concatenated core protein sequences[57]. The third virus, choristoneura fumiferana granulovirus (ChfuGV), appears also to be closely related to EreIGV and ClasGV-B. While ChfuGV has not been placed on a phylogeny based on full genome sequence or core protein concatenations, several ChfuGV share high aa identity with ORFs from Choristoneura occidentalis granulovirus (ChovGV), which is phylogenetically positioned most closely to EreIGV and ClasGV-B[58]. The baculovirus GADD34 homologs are in fact quite diverse in sequence from one another, with the EreIGV and ChfuGV homolog having higher sequence identity with homologs from Trichoplusia ni ascovirus 2c (TNIV2c) and AmEPV (Poxviridae) than with the ClasGV-B homolog.

The TNAV2c homolog is closest in sequence to the three baculovirus homologs and the AmPV homolog, having between 33 and 43% amino acid identity with these proteins (**table 3**). All five of these viruses infect moths (**table 5**). Interestingly, included in this clade as well are the cellular homologs from several moth species, including the GADD34/CReP homolog from the tobacco hornworm moth, which is very closely related to the EreIGV host: the cassava hornworm moth. Within this clade, the TNIV2c homolog forms a subclade with the GADD34/CReP homolog from its host species, the cabbage looper moth. Unfortunately there is no sequence data available for the exact host species of EreIGV, AmPV, ClasGV-B, or ChocGV, but it is possible that the homologs from these viruses would also cluster most closely with GADD34/CReP homologs from their natural host species.

A similar phenomenon can be seen from the homolog from the snow crab infecting caulimovirus (ChoBV). The ChoBV homolog is quite divergent in sequence from all other viral homologs, and instead forms its own clade with the cellular GADD34 homolog from a swimming crab (no genome data available for host species) (**figure 3**). The homologs from megastigmus wasp adintovirus (MWAV) and

glossina hytrosavirus (SGHV) are extremely divergent in sequence, having no higher than 33% identity with any other viral homologs (**table 3**). The SGHV homolog forms a subclade with GADD34/CReP homologs from its host species, tsetse flies, as well as other fly species. This subclade exists within a larger clade which contains GADD34/CReP homologs from several mosquito and wasp species as well as MWAV. (**figure 3**).

All five GADD34 homolog-encoding iridoviruses are members of the genus Chloriridovirus, which falls into the invertebrate-infecting Betairidovirinae family. Along with one additional virus – invertebrate iridovirus 3 (IIV-3) – these five viruses comprise the currently known members of genus Chloriridovirus. It is interesting that a GADD34 homolog is not found in IIV3, as it has been proposed via phylogentic analysis to be more closely related to the other Chlorirividoviruses than is AMIV[59]. However, this may not be all that surprising given the fact that the IIV GADD34 homologs do not share very high aa sequence identity among each other. IIV22 and IIV30 are the most similar of the iridovirus homologs, sharing 72.55% sequence identity. They share only a little over 40% sequence identity with homologs from AMIV, IIV9, and IIV25. IIV9 and IIV25 are also slightly more similar in sequence, with 63% amino acid identity (**table 3**). Interestingly, while the homologs from the rest of the arthropod infecting viruses clade with GADD34 homologs, the iridovirus homologs fall into the PR15B clade. In our phylogeny, all five IIV homologs clade together with livupivirus L protein and DP71L from several strains of ASFV (**figure 3**), suggesting that IIV homologs are more closely related to CReP than to GADD34. This supposition is further supported by the fact that the majority of hits from BLASTs using IIV homologs as queries are PR15B homologs, rather than PR15A.

DISCUSSION:

The presence of GADD34/CreP-like proteins throughout the animal kingdom suggests these are true homologs of an ancestral GADD34/CReP-like protein

GADD34/CReP homologs were identified in most animal phyla, including *porifera, cnidaria, mollusca, annelida, arthropoda, echinodermata, urochordata*, and *vertebrata*. While the apparent lack of GADD34/CReP homologs among *bryozoa* and *cephalochordata* species may be the product of a dearth of available genome data for these phyla, the true absence of GADD34/CReP homologs in nematodes is highly plausible, given the abundance of available genome data for both phyla and in particular the immense amount of research that has been performed on *C. elegans,* the prototypical nematode species widely used a an animal model for genetic and neuroscience studies[60,61]. The apparent pattern of existence of GADD34/CReP homologs among animal phyla does not coincide with the pattern of common anatomical features by which animal phyla are typically arranged in a phylogeny. However, it is likely that a GADD34/CReP-like protein existed in a common ancestor to all animals and has since diverged extensively with evolution and in some lineages has even been lost entirely.

The emergence of two paralogs in vertebrates: a gene duplication event

Among vertebrates, both GADD34 and CReP homologs were found in higher vertebrates (amphibians, reptiles, birds, and mammals) as well as both ray-finned and lobe-finned bony fish, while only a single GADD34/CReP homolog was found in jawless and cartilaginous fish. The correlation of this distribution pattern among vertebrate classes of a second GADD34/CReP paralog with proposed evolutionary models for vertebrates suggests that the second GADD34/CReP paralog arose from a gene duplication event which occurred somewhere in a common ancestor to both ray-finned fish and lobe-finned fish, from which "higher" vertebrates (tetrapods) are believed to have evolved[62,63,64]. Gene duplication is known to act as an important driving force of evolution by allowing for the acquisition of novel gene functions with the subsequent adaptation of duplicated genes[65,66]. Given that the GADD34/CReP homologs identified in cartilaginous fish appear to be more closely related to PR15B than to PR15A, we speculate that a CReP-like protein was present in an ancestral species common to all vertebrates, and the putative gene duplication event occurred in a common ancestor to ray-finned and lobe-finned fish, but not jawless fish.

Following this hypothesis, the duplicated gene evolved into a GADD34-like protein, homologs of which are seen today in lobe-finned fish as well as all lineages which are believed to have evolved from ray-finned fish (higher vertebrates).

Viral GADD34 homologs as horizontally acquired host genes

Given their homology to cellular PR15A and PR15B, it is likely that viral GADD34 homologs were originally acquired from cellular organisms via the horizontal transfer of host genes. Horizontal gene transfer (HGT) is widely accepted as a major driving force in the shaping of viral genomes and is believed to be the source in particular of many immunomodulatory viral genes involved in virulence[67,68]. Horizontally acquired genes provide a platform for adaptation to novel hosts as well as changes in virulence and viral pathogenicity[69]. Historically, viral genes have been identified as putative horizontally transferred genes solely based on bioinformatic analysis, but our lab has recently utilized a sophisticated experimental system to model the HGT process in real time, providing experimental evidence for one mechanism employed by viruses to acquire host genes[70].

A vast repertoire of genes encoded by viruses from several GADD34 homolog-encoding families are believed to be the products of HGT. Putative horizontally acquired genes have been identified in poxviruses, herpesviruses, iridoviruses, baculoviruses, ascoviruses, and hytrosaviruses, thus providing precedent for our proposition that GADD34 homologs identified in members of these viral families are also the products of HGT events[71,72,73,74,75] While HGT hasn't been directly surveyed in afsaviruses or ascoviruses, their putative close evolutionary relationship with other cytoplasmic dsDNA virus families like baculoviruses and poxviruses, as well as the existence of afsavirus and ascovirus homologs to several poxvirus and baculovirus proteins, respectively, suggest that afsavirus and ascovirus genomes are highly likely to have been shaped by HGT as well[76]. Adintoviruses are a newly defined family of linear dsDNA viruses named for their similarity to adenoviruses and their retrovirus-like integrase gene[77]. While HGT has not been surveyed in adintoviruses, one can imagine a scenario in which a host gene or gene

fragment located nearby to an adintovirus integration site in the host genome gets incorporated into adintovirus virions.

The extremely low sequence identity among GADD34 homologs from different viral families suggests these viral GADD34 homologs in different viral families are the products of independent acquisition events. For several viral GADD34 homologs, this conclusion is further supported by their phylogenetic placement nearer to several cellular GADD34 homologs than to GADD34 homologs from other viral families (**figure** 3). The substantially higher sequence identity seen among homologs encoded by viruses in the same family suggests that an acquisition event may have occurred in a common ancestor to the taxon containing multiple GADD34-encoding viruses, and that the acquired ORF has since evolved into multiple orthologs via speciation events. Lastly, viruses are speculated to have co-evolved alongside cellular organisms for an incredibly long time, possibly even from the very beginnings of life on earth[78]. Thus, the extreme divergence in aa sequence of some viral GADD34 homologs from cellular GADD34 homologs in their known hosts or any other extant cellular species may be explained by the hypothesis that some GADD34 homologs were acquired by ancestral viruses from ancestral cellular host species which are now extinct.

A GADD34 homolog was acquired by a common ancestor to herpes simplexviruses, but not other alphaherpesviruses

The herpesvirus GADD34 homolog, neurovirulence protein ICP34.5, is present in the genomes of 10 of 13 known simplexviruses and is not found in any other herpesvirus genera. The levels of sequence identity among simplexvirus ICP34.5 homologs correlate nicely with a proposed simplexvirus phylogeny. There are also regions of high sequence identity which extend beyond the universally conserved eIF2α binding motif among ICP34.5 homologs. Taken together, these data lend weight to the hypothesis that an ICP34.5-like protein was present in a common ancestor to simplexviruses, but not other alphaherpesvirsues, and was lost in some linages. Such an evolutionary pattern is not unprecedented, as viral genomes are believed to have been widely shaped through reductive evolution[79].

Further support for this hypothesis is found in the fact that all herpesvirus ICP34.5 homologs cluster together to form their own clade in our phylogeny. This clade does not include the cellular GADD34 homologs from the host species of any ICP34.5-encoding herpesviruses. In fact, this clade does not include any cellular GADD34 homologs at all, and the herpesvirus GADD34 homologs appear to be phylogenetically distantly related to any cellular GADD34 homologs used in our phylogeny. Instead, the top BLAST hits resulting when herpesvirus ICP34.5 homologs are used as queries include mainly PR15A from bats, rodents, wombats (marsupials), frogs, sharks, and skates. Further analysis is required to confidently predict whether the original, ancestral herpes simplexvirus GADD34 homolog which has given rise to ICP34.5 orthologs in modern herpes simplex viruses was in fact acquired from one of these or a closely related, possibly now extinct, species at some point in the evolutionary history of herpes simplexviruses.

Interestingly, while all ten herpesvirus ICP34.5 homologs were found in BLAST searches using GADD34 as bait, only the squirrel monkey herpesvirus ICP34.5 was identifiable in BLAST searches using CReP as bait. This indication that ICP34.5 may be more closely related to PR15A than to PR15B is further supported via sequence comparisons and phylogenetic analysis of PR15A and PR15B with herpesvirus ICP34.5 homologs. As mentioned, simplexvirus ICP34.5 homologs share several sequence motifs with PR15A, including the RARA-AlaArg motif, which is part of the sequence which forms alpha helix 2 in GADD34, as well as a putative nuclear export signal. In addition, ICP34.5 homologs clade with cellular PR15A homologs in our phylogeny. These data support the hypothesis that the ancestral ICP34.5 protein was derived from a cellular PR15A sequence, and not a PR15B sequence. Closer sequence analysis at the PP1 and eIF2α binding regions of these proteins revealed that herpesvirus ICP34.5 homologs look more like PR15A than PR15B within this conserved region, possessing the same $\phi\phi$ sequence as GADD34 but not CReP and lacking an arginine in the position corresponding to the Arg motif identified in CReP. Herpesvirus ICP34.5 homologs also possess a tryptophan 5 residues C-terminal to the $\phi\phi$ motif which is

highly conserved in cellular PR15A sequences but not in PR15B. Interestingly, there is a difference in spacing between the ϕϕ motif and eIF2α binding motif in PR15A and PR15B, with GADD34 (for example) having 12 aa's between these motifs and CReP (for example) having only 5 aa's between these motifs. Herpesvirus ICP34.5 homologs all possess the former spacing, matching that of PR15A.

A CReP homolog was acquired by a common ancestor to clade B, but not clade A or C, avipoxviruses

GADD34 homologs were identified in all known clade B avipoxviruses, including the newly identified finchpox virus, but not in any clade A or clade C avipoxviruses. This distribution pattern suggests that a GADD34 homolog was acquired in a common ancestor to clade B avipoxviruses, after the branching point between the three clades, and has since evolved via speciation into the nine different orthologs present in modern clade B avipoxviruses. Much like for the herpes simplexviruses, this hypothesis is further supported by the clustering of all avipoxvirus GADD34 homologs into their own clade, which excludes cellular homologs from avipoxvirus host species. Like the simplexvirus homologs, avipoxvirus homologs do not appear to be phylogenetically closely related to any cellular homologs used in our tree. As with the herpesvirus proteins, further analysis is required to identify the putative cellular origin of avipoxvirus GADD34 homologs.

Interestingly, avipoxviruses lack two well-studied PKR antagonists – K3 and E3 – which are present in other chordopoxvirus genera. It is possible that GADD34 homologs functionally replace these other two PKR antagonists in avipoxvirues. The restriction of GADD34 homologs to avipoxviruses could be explained by a number of factors, including particularities of the avian immunological response to viral infection compared to that of other chordopoxvirus hosts. The full host ranges of avipoxviruses have not been thoroughly explored, but the recent emergence of an avipox-like virus in green sea turtles illustrates a potential propensity of avipoxviruses to infect distantly related host species. Much as the host ranges for other chordopoxviruses have been shown to be determined in part by PKR antagonists K3 and

E3, the avipoxvirus host tropism landscape may be shaped, at least in part, by the activity of avipoxvirus GADD34 homologs. The ability of avipoxvirus GADD34 homologs to counteract PKRs from different host species should be explored in order to gain a better understanding of avipoxvirus virulence and potential host range.

As mentioned, avipoxvirus GADD34 homologs appear to be more closely related to cellular PR15B than PR15A, suggesting an evolutionary hypothesis which includes the acquisition of a cellular PR15B gene – and not PR15A – by a clade B avipoxvirus ancestor. Much like for the herpesvirus homologs, this hypothesis is further supported by closer sequence analysis of the PP1 and eIF2α binding regions, wherein almost all avipoxvirus GADD34 homologs possess the CReP $\phi\phi$ sequence, and all possess an arginine at the position corresponding to the CReP Arg motif. In addition to the aforementioned 17 aa stretch of homology to CReP N-terminal to the PP1 and eIF2α binding motifs, avipoxvirus homologs also show similar spacing between the $\phi\phi$ and eIF2α motifs as CReP, again indicating their closer relation to PR15B than to PR15A.

DP71L may represent the acquisition of a GADD34 homolog from ticks, a known vector for ASFV

DP71L from four different ASFV strains cluster phylogenetically with the livupivirus L protein and the iridovirus homologs. Together, these viral homologs form their own clade which excludes all cellular GADD34 homologs included in our tree. This clade falls within the larger PR15B clade, and in fact all of these viral GADD34 homologs appear to be more closely related to PR15B than to PR15A, displaying the PR15B-like spacing between the $\phi\phi$ and eIF2α motifs and possessing an arginine in the position corresponding to the CReP Arg motif, but lacking the tryptophan 5 residues C-terminal to the $\phi\phi$ motif which is conserved in GADD34, herpesvirus homologs, and homologs from all other arthropod-infecting viruses. Despite clustering together phylogenetically, these homologs display low sequence identity (27-38%) between families and likely represent independent gene acquisition events per family. Interestingly, DP71L from all four strains cluster closely with GADD34 homologs from several

tick species. Tick GADD34 homologs are also the top BLAST search results when DP71L is used as a query. Given that ticks are known vectors of ASFV and the apparent close phylogenetic relatedness of tick and ASFV GADD34 homologs, we hypothesis that a GADD34 homolog was transferred to the ASFV genome from a tick species[80].

Acquisitions of GADD34 homologs in other arthropod-infecting viruses likely occurred more recently

Excluding the iridovirus GADD34 homologs, when GADD34 homologs from arthropod viruses are used as queries in BLAST searches, the top hits are all GADD34 homologs from various arthropod species. Our phylogenetic analysis shows that arthropod virus GADD34 homologs appear to be very closely related to GADD34 homologs from their host species or closely related species. This is seen for all five moth-infecting viruses (TNAV2c, EPV, and the three baculoviruses), the snow crab infecting caulimovirus (ChoBV), and glossina hytrosavirus (SGHV). This close phylogenetic clustering of each of these viral proteins with their host counterparts suggests relatively recent horizontal acquisition of a GADD34 homolog, potentially even occurring in the current host-virus pair. This is in contrast to most vertebrate virus-encoded GADD34 homologs, which appear to have been acquired in an ancestral virus from an ancestral host species. The independent acquisitions of GADD34 homologs in TNAV2c, EPV, and the baculoviruses is not entirely unexpected, as it is known that the genomes of viruses which occupy the same ecological niche and are thus subjected to similar evolutionary pressures display convergent evolution, and in fact the parallel acquisition of other adaptive genes has been traced in entomopoxviruses and baculoviruses[81]. Again excluding the iridovirus homologs, GADD34 homologs from arthropod-infecting viruses appear to be more closely related to PR15A than to PR15B, being placed phylogenetically within the larger PR15A clade on our tree, displaying PR15A-like spacing between the $\phi\phi$ and eIF2α motifs, and possessing the tryptophan 5 residues C-terminal to the $\phi\phi$ motif which is conserved only in PR15A.

The livupivirus GADD34 homolog: an enigma in the HGT scheme

All of the aforementioned viral families fall into Baltimore class I, possessing large, dsDNA genomes. One can imagine how genetic material can easily be swapped between host and viral dsDNA genomes, especially for viruses for which genome replication occurs in the nucleus, like herpesviruses. For cytoplasmic viruses, like poxviruses, our lab has identified one mechanism in which HGT is facilitated by LINE-1 retrotransposons[70]. In addition to the shared nucleic acid type between these viral families and their hosts, the large genome sizes of these viral families can accommodate the acquisition of host genes more easily than viruses with smaller genomes. The livupivirus L protein, however, presents an enigma in our proposed HGT scheme. As mentioned, the L protein was the only identified protein encoded by an RNA virus with homology to cellular PR15A/PR15B. RNA virus proteins generally do not exhibit as much sequence identity to host proteins or anomalous nucleotide composition as many dsDNA viral proteins do, and thus are not commonly speculated to have been derived from HGT events. As the presence of this GADD34-like protein in a picornavirus genome seems to be anomalous, we cannot exclude the possibility that the livupivirus L protein is a product of convergent evolution of the picornavirus genome, as opposed to HGT. Picornavirus leader (L) proteins are in fact known to be quite divergent in sequence and to assume different functions for different picornaviruses[82]. One can imagine a scenario in which the relatively high mutation rate of picornavirus genomes and the distinct evolutionary pressures posed by livupivirus's amphibian host (this is one of only two picornaviruses thus far identified in amphibian hosts) resulted in the presence of a leader protein with a GADD34-like function in the livupivirus genome[83,84].

CHAPTER 2: FUNCTIONAL CHARACTERIZATION OF SELECT VIRAL GADD34 HOMOLOGS

Jeannine Stroup[1], Dewi Megawati[1], Greg Brennan[1], Stefan Rothenburg[1]

1. Department of Medical Microbiology and Immunology, School of Medicine, University of California, Davis, California, USA

ABSTRACT:

Cells respond to a variety of stresses by phosphorylating the alpha subunit of eukaryotic initiation factor 2 (eIF2) and thereby shutting down cap-dependent protein synthesis. During viral infection, eIF2α is phosphorylated by the double-stranded RNA-activated protein kinase R (PKR) and the resulting translational shut off inhibits viral replication. In humans, eIF2α is dephosphorylated by two phosphatase complexes, containing paralogous protein phosphatase 1 (PP1) cofactors GADD34 and CReP, which target PP1 phosphatase activity by directly binding eIF2α. Herpes simplex virus (HSV) 1 and African swine fever virus (ASFV) encode homologs of GADD34/CReP, which also form eIF2α dephosphorylation complexes with PP1. These viral GADD34 homologs can be thought of as indirect PKR antagonists. Our bioinformatic analysis identified additional GADD34 homologs in 37 viruses from 10 different viral families. Here we evaluate the anti-PKR pathway activity of a select panel of these putative PKR antagonists. We show that GADD34 homologs encoded by two herpes simplexviruses, two avian poxviruses, African swine fever virus, and an amphibian picornavirus (Livupivirus) counteract PKR in both a virus and cell line specific manner.

INTRODUCTION:

To overcome the antiviral effects of PKR, viruses have evolved a repertoire of PKR antagonists. Viral PKR antagonists can be grouped into six classes depending on mechanism of action: class 1: reduce free dsRNA levels, class 2: prevent PKR homodimerization, class 3: induce PKR degradation, class 4: interfere with eIF2α phosphorylation, class 5: induce P-eIF2α dephosphorylation, and class 6: inhibit P-

eIF2α-eIF2B interaction. Many viral PKR antagonists play a role in determining host tropism and are thus considered host range factors[85]. For instance, poxvirus PKR antagonists K3 and E3 – class 4 and class 1/2 type inhibitors, respectively – have been shown to be important for viral replication in cell lines from some host species but not others[86,87]. It is likely that additional viral proteins that antagonize PKR via other mechanisms, including viral GADD34 homologs (class 5 inhibitors), also display host specificity. It is important to delineate such host range functions, especially for understanding and predicting the virulence of viruses encoding such proteins in novel host species.

Given the presence of two common PP1 binding motifs and the conservation at the region corresponding to the GADD34 eIF2α binding motif, all 37 viral GADD34 homologs identified in chapter 1 represent potential class 5 PKR antagonists. GADD34 homologs from CNPV, MaHV, and AmPV have been shown to decrease eIF2α phosphorylation levels and counteract PKR toxicity in yeast[7]. However, the activity of these proteins and additional newly identified viral GADD34 homologs has not been explored in the context of viral infection nor in a mammalian system.

In order to characterize the ability of viral GADD34 homologs to act as class 5 PKR antagonists, we performed functional characterization in mammalian cell lines on a select panel of viral GADD34 homologs. This panel includes the previously well characterized proteins, HSV1 γ34.5 and ASFV DP71L, as well as CNPV231, which has thus far only been shown to counteract PKR activity in yeast[7]. In addition, we performed functional analysis here for the first time on GADD34 homologs from a recently isolated puffinpox virus (PUPV), a fruitbat infecting herpes simplexvirus (FABHV1), and one of two known amphibian infecting picornaviruses (Livupivirus). Each of these viruses has only recently been isolated from a single host species, and the full breadth of their host tropism has yet to be explored. It is therefore important to determine not only the ability of GADD34 homologs from these viruses to counteract the PKR response, but also effects on this activity caused by the different environments of cell lines from different host species. Here we show that HSV1 γ34.5, ASFV DP71L, CNPV231, PUPV MyD116,

FBAHV1 ICP34.5, and Livupivirus L protein rescue replication of a PKR antagonist-deficient, highly attenuated vaccinia virus (VACV) in both a virus and cell line specific manner, and that this rescue in replication correlates with a decrease in eIF2α phosphorylation for two cell lines tested.

MATERIALS AND METHODS:

Cell lines:

HeLa (human, ATCC #CCL-2), HeLa PKR-knockout (kindly provided by Dr. Adam Geballe[88]), RK13 (rabbit, ATCC) RK13+E3+K3[89], OA1 (sheep, ATCC #CRL-6538), BSC40 (African green monkey, ATCC #CRL-2761), A549 (human, ATCC # CRM-CCL-185), A549 RNaseL KO (kindly provided by Dr. Bernard Moss[90]), A549 DKO (kindly provided by Dr. Bernard Moss[90]), Caco-2 (human, ATCC #HTB-37), were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 5% fetal bovine serum and 100 IU/ml penicillin/streptomycin (Gibco). RK13+E3L+K3L cell culture medium contained 500 μg/ml geneticin and 300 μg/ml zeocin (Life Technologies).

Plasmids:

All PKR and GADD34 homolog genes were cloned into the pSG5 expression vector (Stratagene) for transient transfection assays. Cloning of mouse PKR was described previously[91]. Full length and truncated human GADD34 and yeast-codon optimized genes for γ34.5, cnpv231, DP71L, and MaHV ICP34.5 were provided by Dr. Nels Elde[7]. 5′ 1X FLAG tag was introduced into yeast codon optimized γ34.5 for immunoprecipitation and immunoblot analysis. VACV codon optimized, 1X 5′ FLAG-tagged genes for γ34.5, FBAHV1 ICP34.5, CNPV231, MyD116, DP71L and L protein were cloned into both pSG5 and P837-GOI-mCherry-E3L for recombinant VACV generation using from synthesized DNA. All DNA sequences were validated by Sanger sequencing (UC Davis sequencing facility).

Luciferase based reporter assays:

Assays were performed as previously described[92] with the following modifications: 5×10^4 Hela PKR KO cells per well were seeded in 24-well plates 14 hours prior to the experiment. Cells were transfected with 50 ng of the mouse PKR expression plasmid, 200ng of each GADD34 homolog expression plasmid, and 50 ng of pGL3 firefly luciferase expression plasmid (Promega) using GenJet (Signagen) at a DNA to GenJet ratio of 1:2 following the manufacturer's protocol. Empty pSG5 vector was transfected to maintain a consistent DNA concentration between samples. 48 hours post-transfection cells were lysed with mammalian lysis buffer (GE Healthcare) and then luciferin (Promega) was added following the manufacturer's recommendations. Luciferase activity was measured using a GloMax luminometer (Promega). Experiments were conducted in triplicate.

Viruses and infection assays:

VACV Copenhagen strain VC2 was kindly provided by Dr. Bertram Jacobs. Generation of VC-R4 was described previously[93]. rVC-R4_GADD34, rVC-R4_γ34.5, rVC-R4_ICP34.5, rVC-R4_CNPV231, rVC-R4_MyD116, rVC-R4_DP71L, and rVC-R4_L were generated by the scarless integration of FLAG-tagged, VACV codon optimized, ORFs into the E3L locus of VC-R4 by the same method. The recombinant viruses were plaque-purified three times, the integrations were confirmed by Sanger sequencing (UC Davis sequencing facility), and expression of each GADD34 homolog was confirmed by immunoblot analysis.

Plaque assays were carried out in confluent six-well plates of the indicated cell lines, which were infected with 50 plaque forming units (pfu) of each indicated virus, as determined on RK13+E3+K3 cells. One hour post infection, the medium was replaced with DMEM containing 1% carboxymethylcellulose (CMC). After 48 hours, cells were stained with 0.1% crystal violet. Plates were imaged using an iBright Imaging System (Invitrogen).

Multiple-cycle virus replication assays performed in confluent six-well plates of the indicated cells, which were infected with each indicated virus at an MOI of 0.01. 48 hours post infection, cells and

supernatants were collected and subjected to three rounds of freezing at -80°C and thawing at 37°C. Lysates were sonicated for 15s, 50% amplitude (Qsonica Q500). Viruses were titered by 10-fold serial dilutions on RK13+E3+K3 cells. Infections were performed in triplicate (biological replicates) and viral titers were performed in duplicate (technical replicates).

Immunoprecipitation and immunoblot analyses:

To detect expression of GADD34 homologs and evaluate eIF2α phosphorylation levels, 6 well plates of confluent monolayers of indicated cells were infected with indicated viruses at an MOI of 3. 1 hour post infection, inocula were replaced with fresh cell culture medium. 6 hours post infection, cells and supernatants were collected and centrifuged at 800 RCF for 5 minutes. Cell pellets were lysed with 1% SDS in PBS (Gibco) and sonicated at 50% amplitude for 10 seconds twice. All protein lysates were separated on 10% SDS-polyacrylamide gels and transferred to PVDF membranes. Membranes were blocked with 5% (w/v) skim milk powder or bovine serum albumin (BSA) in TBST for 1 hour at room temperature (RT). Membranes were probed with antibodies against FLAG (1:1000 Sigma-Aldrich, F3165), total eIF2α (1:1000, Santa Cruz Biotech sc133132), phospho-eIF2α (1:1000 Cell Signaling Technology, CST 1297; 1:3000 Abcam E90 ab32157), and B-actin (1:5,000, Sigma-Aldrich, A1978) in 5% skim milk powder or bovine serum albumin (BSA) in TBST for either 1hour at RT or overnight at 4C. Membranes were then washed with TBST three times 5 mins each, and incubated for 1 hour at room temperature or overnight at 4C with donkey anti-rabbit or goat anti-mouse secondary antibodies conjugated to horseradish peroxidase (Invitrogen, A16110, 62–6520) at 1:10,000 in 5% (w/v) skim milk powder or BSA in TBST. The membranes were then washed five times for 5 min each and proteins were detected with Amersham ECL (GE Healthcare). Images were taken using the iBright Imaging System (Invitrogen).

For immunoprecipitation of GADD34 and γ34.5, 5ug expression plasmids encoding 5′ 1X FLAG-tagged GADD34 and γ34.5 (yeast codon optimized) were transiently transfected into 70-80% confluent monolayers of HeLa cells in 10cm dishes using GenJet (Signagen) at a DNA to GenJet ratio of 1:2

following the manufacturer's protocol. 48 hours post transfection, cells were washed twice with PBS (Gibco) and lysed with ice cold IP lysis buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% TRITON X-100, protease inhibitor cocktail tablet), and centrifuged at 14,000xg for 15min. The supernatant was then added to 40µl ANTI-FLAG M2 affinity gel (Sigma-Aldrich), prepared according to the manufacturer's protocol. The remainder of the immunoprecipitation was carried out according to the manufacturer's protocol. Binding step was performed overnight at 4C. Samples were eluted from resin by boiling at 95C for 5min and were diluted 1:2 prior to loading onto 12% SDS-PAGE gels. Immunoblot analysis was performed as described above with anti-FLAG and anti-totaleIF2α antibodies (specified above) at 1:5000 and 1:1000 dilutions, respectively. Samples bound to ANTI-FLAG M2 affinity gel were submitted to the UC Davis Proteomics Core for trypsin digest and MS analysis.

RESULTS:

GADD34, CReP, and Viral GADD34 homologs counteract PKR activity in a luciferase-based transfection assay

We first sought to determine whether our panel of viral GADD34 homologs can counteract PKR activity in mammalian cells using an established transfection-based luciferase reporter assay[94]. In this assay, PKR KO cells are co-transfected with plasmids encoding the luciferase enzyme and PKR. Luciferin is added a standard period of time and the amount of light produced via luciferase-catalyzed oxidation of luciferin is measured as a proxy for the level of translation occurring within the cells. Luciferase readout is low when PKR is transfected into cells, presumably because PKR-mediated translational repression blocks most luciferase from being expressed. Co-transfection of an effective PKR inhibitor can increase luciferase readout anywhere from 2-fold to 40-fold, as has been seen for VACV K3[94]. In the case of GADD34 homologs, we predict that GADD34 homolog-mediated dephosphorylation of eIF2α will counteract translational repression and increase luciferase readout in this assay. Previous testing of ASFV

DP71L activity in a similar system has shown a 2-fold increase in luciferase readout with co-transfection of DP71L compared to transfection of PKR alone[19]. ORFs encoding human GADD34 and CReP as well as GADD34 homologs from HSV1 (γ34.5), CNPV (CNPV231), ASFV (DP71L), FBAHV1 (ICP34.5), PUPV (MyD116), and Livupivirus (L) were cloned into a mammalian expression vector and transfected into HeLa PKR KO and RK13 PKR KO cells along with mouse PKR and the luciferase plasmid. GADD34, CReP, and viral GADD34 homologs showed differential abilities to counteract PKR-mediated translational repression. In HeLa PKR KO cells, both GADD34 and CReP showed increased luciferase readout 3- to 4-fold compared to transfection of PKR alone (**figure 1B).** Of the viral homologs, FBHAV1 ICP34.5 showed the highest increase in luciferase readout of about 5.5-fold compared to transfection of PKR alone (**figure 1B**). FBHAV1 is followed by PUPV MyD116, which increased luciferase readout 4.5-fold, and CNPV231 and ASFV DP71L, which increased luciferase readout around 3-fold, while HSV1 γ34.5 only increased readout slightly below 2-fold, and Livupivirus L proteins did not appear to have a big effect. In contrast, no anti-PKR activity was seen for GADD34, CReP, and FBAHV1 ICP34.5 in RK13 PKR KO cells (**figure 1B**). In RK13 PKR KO cells, activity of all homologs was generally lower than it was in HeLa PKR KO cells. PUPV MyD116, ASFV DP71L, and CNPV231 showed about 2- and 2.5-fold increase in luciferase activities when compared to transfection of PKR alone. These results confirm what has previously been shown for ASFV DP71L, show for the first time that CNPV231 can counteract PKR activity in a mammalian system, and represent the first functional characterization of PUPV MyD116, FBAHV1 ICP34.5, and Livupivirus L protein. While PUPV MyD116 displayed anti-PKR activity in both cell lines tested, FBAHV1 appeared to display cell line dependence in its ability to counteract PKR in this luciferase-based assay. This result is surprising given the assumption that FBAHV1 ICP34.5 functions similarly to HSV1 γ34.5, whose known binding partners – the PP1 catalytic subunit and eIF2α – are extremely highly conserved among host species.

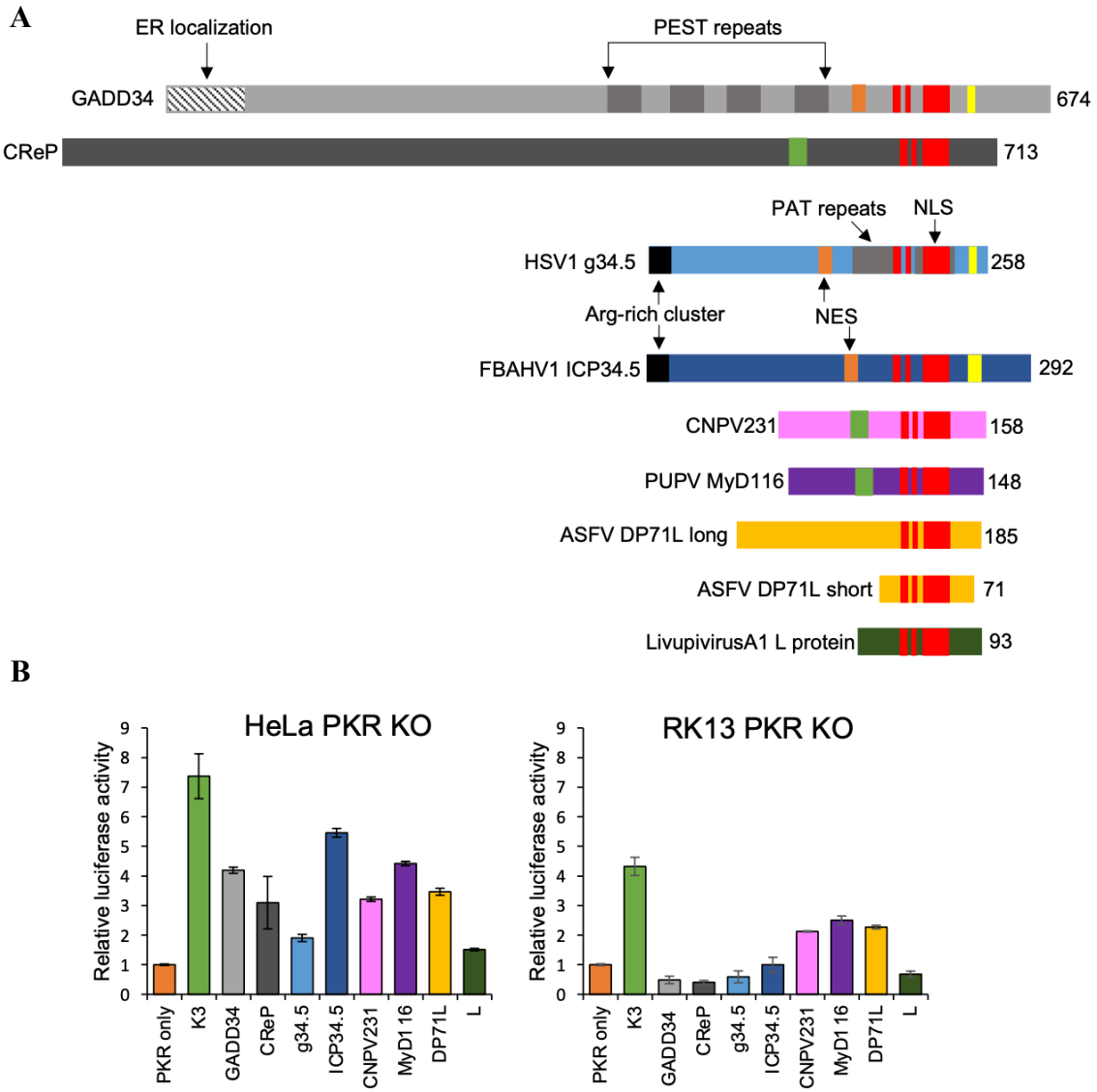

Figure 1: (A) *Scaled schematic representation of GADD34, CReP, and viral GADD34 homologs functionally characterized in this paper. Virus abbreviations: HSV1: herpes simplex virus 1; FBAHV1: fruitbat alphaherpesvirus 1; CNPV: canarypox virus; PUPV: puffinpox virus; ASFV: African swine fever virus (**B**) HeLa PKR KO cells and RK13 PKR KO cells were transfected with expression vectors encoding firefly luciferase (0.05 μg), mouse PKR (0.05 μg) and GADD34, CReP, or the indicated viral GADD34 homolog (0.2 μg). Luciferase activities were measured 72 or (HeLa) 48 (RK13) hours after transfection and normalized to PKR-only transfected cells to obtain relative luciferase activities. Error bars represent the standard deviations from three independent transfections.*

Recombinant vaccinia viruses (VACV) expressing GADD34 and viral GADD34 homologs form plaques in PKR competent cells

In order to determine the ability of GADD34 and viral GADD34 homologs to counteract PKR in the context of viral infection, FLAG-tagged and vaccinia virus (VACV) codon optimized GADD34 homologs from HSV1 (γ34.5), FBAHV1 (ICP34.5), CNPV (CNPV231), PUPV (MyD116), ASFV (DP71L), Livupivirus (L) and human GADD34 were inserted into the E3L locus of a VACV lacking PKR inhibitors E3L and K3L, where the ORFs are under control of the native E3L promoter[95]. This virus, referred to here as VC-R4, cannot replicate in PKR-expressing cell lines, and is used in our lab as an established system for determining the ability of viral PKR antagonists to rescue viral replication in PKR competent cells[96] . The resulting recombinant viruses are referred to as rVC-R4_GADD34, rVC-R4_γ34.5, rVC-R4_ICP34.5, rVC-R4_CNPV231, rVC-R4_MyD116, rVC-R4_DP71L, and rVC-R4_L. Expression of each GADD34 homolog in its respective recombinant virus was confirmed with an anti-FLAG immunoblot of lysates from the permissive RK13+E3+K3 cell line infected with each virus (**figure 2**). Excluding livupivirus L protein, expression of all homologs was also detected in the PKR-competent cell line BSC40.

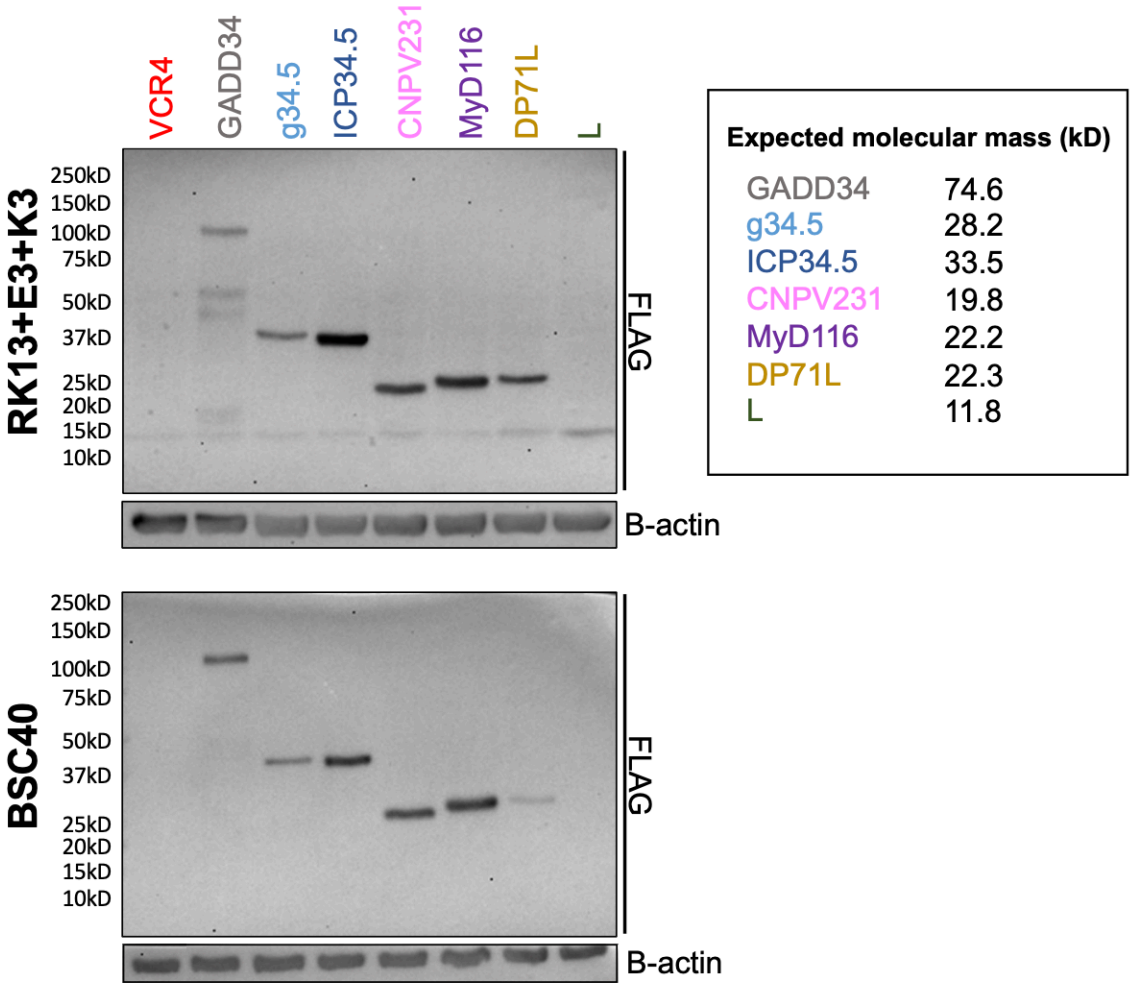

Figure 2: *Expression of GADD34 and viral GADD34 homologs in recombinant VC-R4. RK13+E3+K3 cells (permissive for infection with highly attenuated VACV, VC-R4) and BSC40 cells (PKR competent) were infected with recombinant VC-R4 (rVC-R4) encoding FLAG tagged versions of either human GADD34 or the indicated viral GADD34 homolog at an MOI of 3. Lysates were collected at 6 hours post infection (hpi), were run on 10% SDS-PAGE gels, and were analyzed by immunoblot analysis with an anti-FLAG antibody. B-actin was used as a loading control.*

In order to evaluate the ability of this panel of recombinant viruses to replicate in PKR-expressing cell lines from different host species, plaque assays were performed in BSC40 (African green monkey), RK13 (rabbit), OA1 (sheep), and HeLa (human) cells. As a control, RK13 cells expressing VACV PKR

antagonists K3 and E3, referred to hereafter as "RK13+E3+K3", were also infected. RK13+E3+K3 cells are permissive for VC-R4 replication, as the presence of K3 and E3 in the cell line overcomes VC-R4's lack of PKR inhibitors and allows the virus to replicate despite PKR activity; these cells have been used as a control in this experimental system in the past[94,91]. All cell lines were also infected with VC-R4 as a replication null control, as well as an E3 revertant virus as a replication positive control. The E3 revertant virus has the VACV E3 gene reinserted the E3L locus of VC-R4 and has been shown to replicate to high titers in several cell lines[96]. All viruses formed plaques in RK13+E3+K3 cells, including VC-R4, with E3 revertant virus forming slightly larger plaques than VC-R4 and rVC-R4 viruses (**figure** 3). This trend was seen across all cell lines and was more pronounced in RK13 and OA1 cells. Importantly, VC-R4 did not form plaques in any PKR competent cell lines. Interestingly, neither did rVC-R4_L. VC-R4_ICP34.5 formed plaques in RK13 cells, but not in BSC40 or OA1. This could be the result of either a complete inability of these viruses to replicate (and thereby form plaques) in these cell lines or of a decrease in plaquing efficiency of these viruses in these cell lines compared to in RK13+E3+K3 cells, which were used for determining the titer of the viral stock used for infection in these experiments. The remaining rVC-R4 viruses formed plaques in RK13, BSC40, and OA1 cells, indicating that these viruses, encoding GADD34, γ34.5, CNPV231, MyD116, and DP71L, can replicate at least to some extent in these PKR-expressing cells.

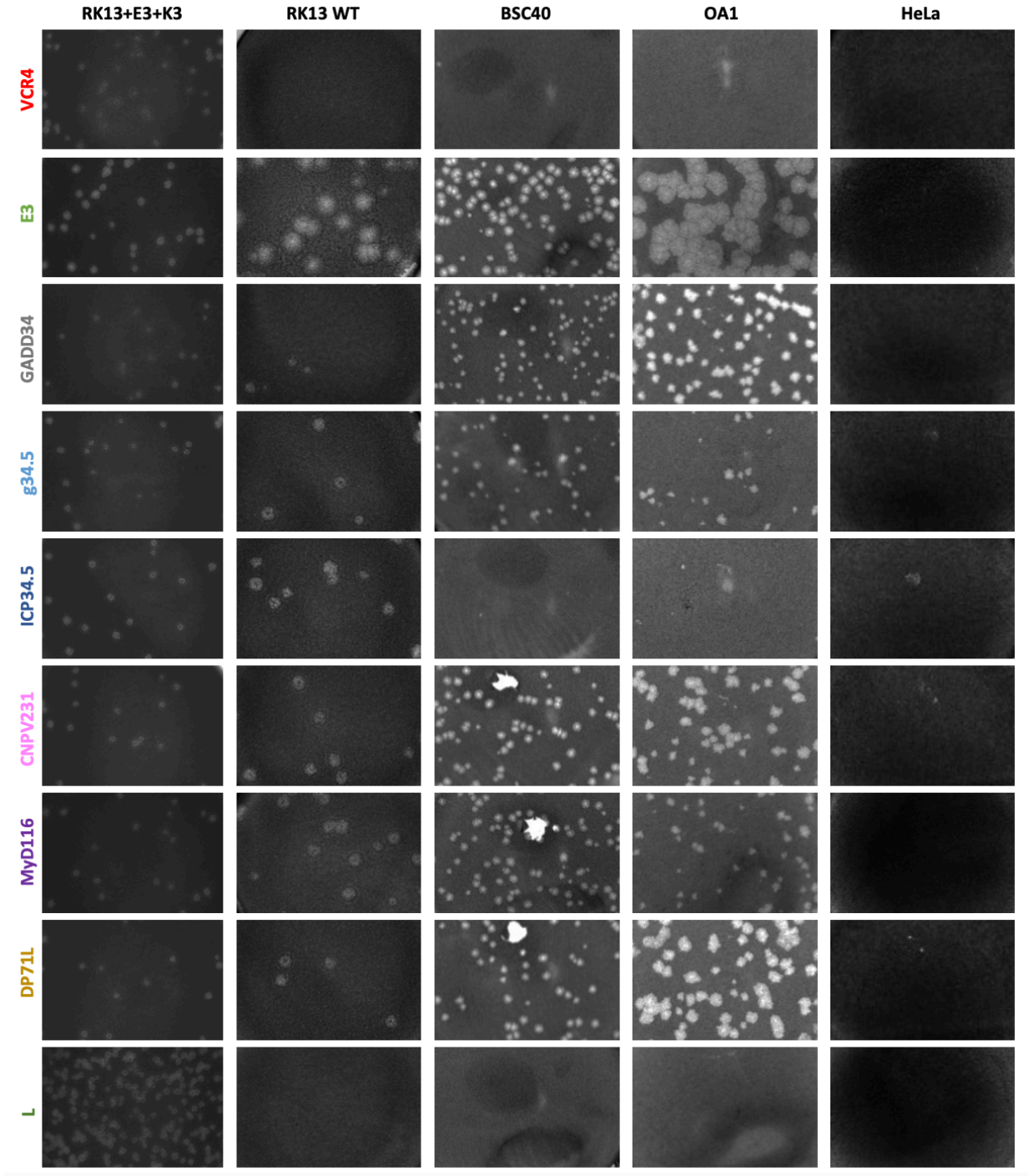

***Figure 3**: Plaque formation by recombinant VACV viruses expressing GADD34 and viral GADD34 homologs in PKR-competent cells. RK13+E3+K3, RK13 (WT), BSC40, OA1, and HeLa cells were infected with 50 pfu of VC-R4, E3 revertant, or a recombinant virus encoding the indicated GADD34 homolog for 1hour and then overlayed with DMEM + 1% carboxymethylcellulose (CMC) to promote plaque formation. Plates were stained with crystal violet two (RK13+E3+K3, BSC40, OA1), four (RK13WT, or seven (HeLa) days post infection (dpi). Results shown are representative of two or more independent infections.*

Recombinant Vaccinia viruses (VACV) expressing GADD34 and viral GADD34 homologs did not form plaques in three PKR-competent human cell lines

Interestingly, none of the viruses, including the E3 revertant virus, formed plaques in HeLa cells (**figure 3**). To further investigate this phenotype, we performed plaque assays in additional human cell lines: Caco-2 and A549. Despite the formation of plaques for the E3 revertant virus, none of the rVC-R4 viruses formed plaques in both Caco and A549 cells, indicating an inability of this entire panel of viral GADD34 homologs to counteract PKR in the context of infection of these human cell lines (**figure 4**). It has been shown that A549 cells have high OAS/RNaseL activity, which can preclude replication of viruses which are able to replicate in other PKR-competent cells.[90] To double check that the lack of plaque formation in A549 cells was not the result of high RNaseL activity in this cell line, we also infected A549 RNaseL KO cells. Again, only the E3 revertant virus formed plaques. As controls, we infected A549 RNaseL KO PKR KO (DKO) and HeLa PKR KO cells, to check that the viruses can replicate in these cell lines in the absence of PKR. All viruses, including VC-R4, formed plaques in both of these PKR KO cell lines, indicating that the inhibition of rVC-R4 replication in PKR-expressing human cell lines may somehow be related to the inability of encoded GADD34 homologs to counteract PKR activity in the context of the human cell environment. These results require further investigation.

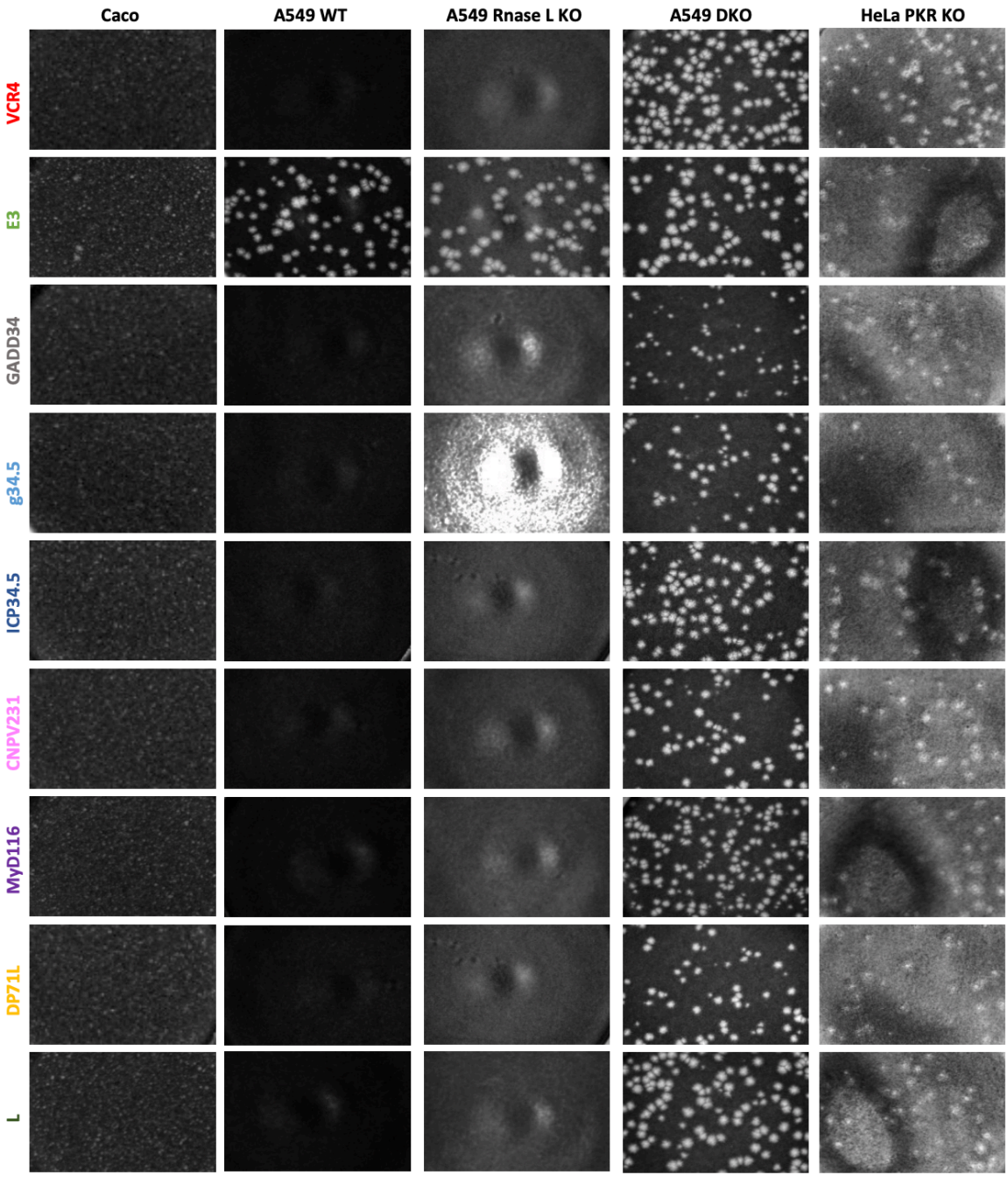

***Figure 4**: Plaque formation by recombinant VACV viruses expressing GADD34 and viral GADD34 homologs in human cell lines. Caco-2, A549 WT, A549 RNaseL KO, A549 DKO, and HeLa PKR KO cells were infected with 50 pfu of VC-R4, E3 revertant, or a recombinant virus encoding the indicated GADD34 homolog for 1hour and then overlayed with DMEM + 1% carboxymethylcellulose (CMC) to promote plaque formation. Plates were stained with crystal violet two dpi (A549 PKR KO, HeLa PKR KO), four dpi (Caco-2) or six dpi (A549WT, A549 RNaseL KO). Results shown are representative of two or more independent infections.*

Viral GADD34 homologs rescue replication of a highly attenuated VACV in a virus and cell line specific manner

In order to further compare the ability of these viral GADD34 homologs to support productive infection of PKR competent cells, we performed multiple cycle replication assays in BSC40, OA1, PK15, Caco-2, and HeLa cells, using RK13++ cells as a control. VC-R4 and the E3 revertant virus were again used as negative and positive replication controls, respectively. All rVC-R4 viruses showed comparable levels of replication in RK13++ cells, and rVC-R4 viruses replicating overall about 1 log lower than the E3 revertant virus and about half a log higher than VC-R4 (**figure 5**). In BSC40, OA1, and PK15 cells, rVC-R4 encoding GADD34, CNPV231, MyD116, and DP71L replicated to levels comparable to the E3 revertant virus, showing a complete rescue of replication by these homologs in these cells. Among these three cell lines, the replication for rVC-R4_γ34.5, rVC-R4_ICP34.5, and rVC-R4_L varied. In BSC40 cells, rVC-R4_γ34.5 replicated to levels comparable to the E3 revertant virus, whereas in OA1 and PK15 cells it replicated to slightly (0.5 to 1 log) lower levels. rVC-R4_ICP34.5 replicated to levels about 1 log lower than the E3 revertant virus in BSC40 cells, but 3 and 4 logs lower than the E3 revertant virus in OA1 and PK15 cells, respectively. While this is still several logs higher than VC-R4 titers in BSC40 and OA1 cells, indicating at least partial rescue of replication by ICP34.5 in these cell lines, it is only half a log higher than VC-R4 titers in PK15 cells, indicating that ICP34.5 only weakly rescued viral replication in PK15 cells. rVC-R4_L titers were lower than titers for the rest of the viruses across the board, showing no replication at all (compared to VC-R4) in OA1 and PK15 cells. rVC-R4_L did still show a substantial rescue of replication in BSC40 cells, reaching titers 4 log higher than VC-R4. Taken together, these results show that our panel of viral GADD34 homologs, to varying degrees, can rescue viral replication in PKR competent cells lines from some host species but not others. These results further solidify the apparent cell line dependency seen among GADD34 homologs in their anti-PKR activity as determined by the luciferase assay (**figure 1).** Despite displaying a partial rescue of viral replication, rVC-R4_ICP34.5 and

rVC-R4_L protein did not form plaques in BSC40 cells (**figure 3**). These results indicate that plaque formation might require surpassing a certain threshold of viral replication, or that the ability of these viruses to form plaques in these cell lines is not directly related to their ability to replicate within the cells. It should be noted that L protein did not appear to be expressed well at 6hpi in both BSC40 and RK13++ cells (**figure 2**). While the lysates used for immunoblot analysis represent different infection conditions from the lysates used to determine viral titer, it is possible that the low titers observed for rVC-R4_L in PKR-expressing cells can be explained by poor expression of the L protein. However, it should also be noted that it does not appear that high levels of expression as detected by Western blot are needed for successful rescue of replication by other GADD34 homologs (namely DP71L, which is only detected at very low levels 6hpi in BSC40 cells but rescues viral replication completely). It should also be noted that rVC-R4_ICP34.5, despite appearing to have strong expression at 6hpi in BSC40 cells, replicates to titers 1 log lower than viruses encoding most other homologs, further suggesting that viral replication rescue phenotypes are not directly correlated with the levels of expression of GADD34 homologs in viral infection.

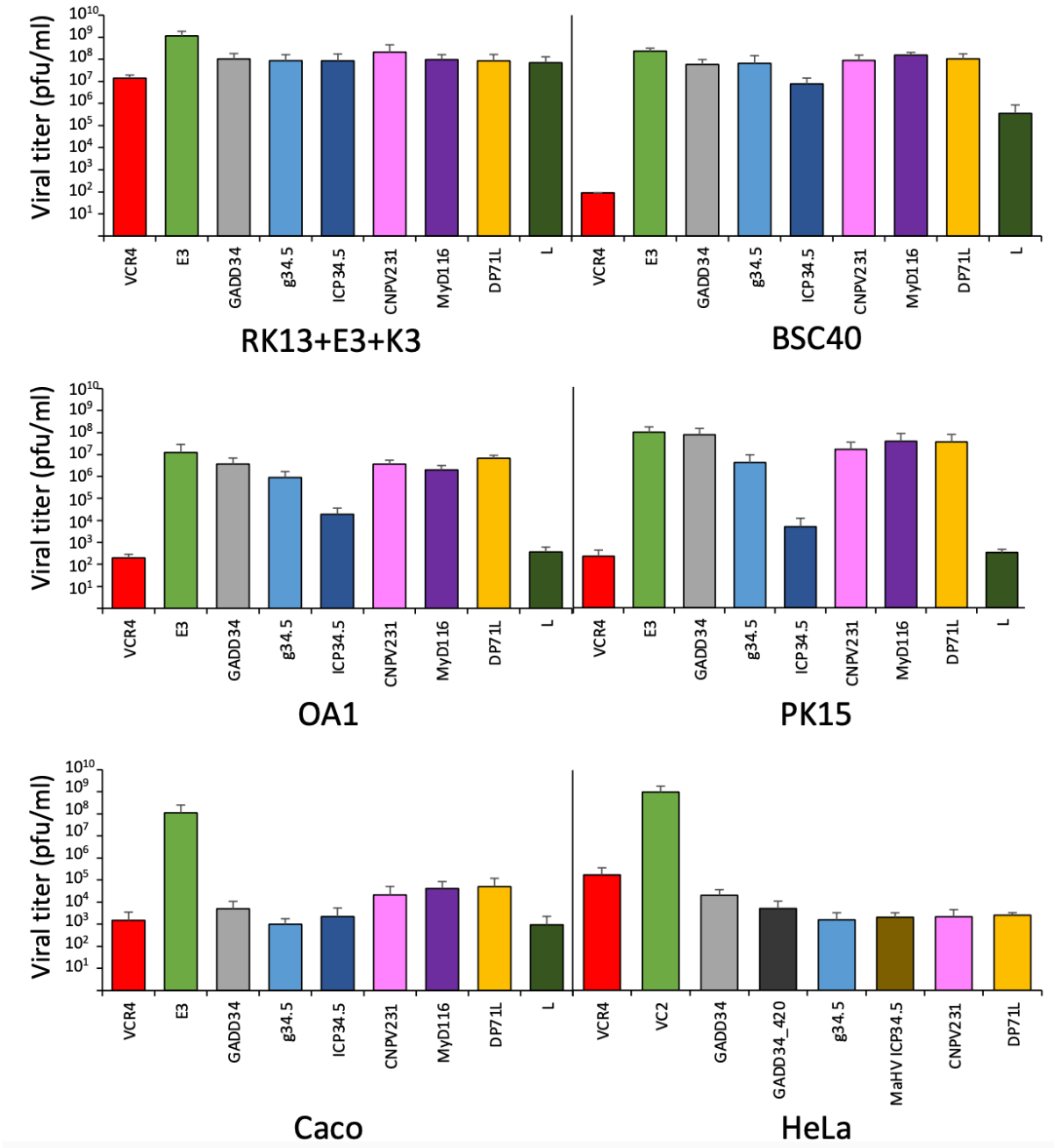

***Figure 5**: Replication of recombinant VC-R4 expressing GADD34 and viral GADD34 homologs in various PKR-competent cell lines. RK13++, BSC40, OA1, PK15, Caco-2, and HeLa cells were infected at an MOI of 0.01 with VC-R4, E3 revertant virus, VC2, or rVC-R4 viruses encoding the indicated GADD34 homolog. Lysates were collected 38hpi and tittered on RK13++ cells, in duplicate. Error bars represent the standard deviations from three independent infections (biological replicates).*

In correlation with the results of the plaque assays, replication was not rescued for most rVC-R4 viruses in both human cell lines tested, despite the E3 revertant virus reaching titers comparable to in other cells, indicating successful replication of this virus in these human cell lines (**figure 5**). In Caco-2 cells, titers for all rVC-R4 viruses were within 1 log of VC-R4 titers. HeLa cells were infected with a slightly different set of viruses, including rVC-R4 encoding a truncated version of GADD34 only containing the conserved C-terminal region and ICP34.5 from MaHV1, and lacking FBAHV1 ICP34.5, PUPV MyD116, and L protein. The viruses tested did not display any increase in titer relative to VC-R4 in HeLa cells. These results, taken together with the plaque assay results, indicate an overall inability of viral GADD34 homologs to counteract the PKR response in the human cell lines tested, and should be further explored.

Decreased eIF2α phosphorylation during infection with rVC-R4_GADD34 viruses

Presumably, the rescue of viral replication in PKR-expressing cells by viral GADD34 homologs seen in plaque and infectivity assays was the result of GADD34 homolog-mediated dephosphorylation of eIF2α. To investigate whether eIF2α was in fact being dephosphorylated in cells infected with rVC-R4 viruses, lysates from BSC40 and PK15 cells infected at a multiplicity of infection (MOI) of 3 were immunoblotted for phosphorylated- and total eIF2α. VC-R4 infection was used as a positive control, as it has been shown to induce high levels of eIF2α phosphorylation, and infection with the E3 revertant virus was used as a negative control, as it resulted in low levels of eIF2α phosphorylation[94]. In BSC40 cells, P-eIF2α levels were decreased in infection with all rVC-R4 viruses compared to VC-R4 (**figure 6**). P-eIF2α levels appear to be higher for rVC-R4_ICP34.5 and rVC-R4_L than for other recombinant viruses. These results correlate well with the plaque assay and infectivity results, altogether indicating that ICP34.5 and L protein did not counteract PKR as well as other viral GADD34 homologs in the context of VACV infection in BSC40 cells. In PK15 cells, P-eIF2α levels were decreased relative to VC-R4 for all rVC-R4

viruses expect rVC-R4_L, which appears to have similar levels of P-eIF2α as VC-R4. These results correlate well with the infectivity assay results in PK15, in which L protein failed to rescue viral replication.

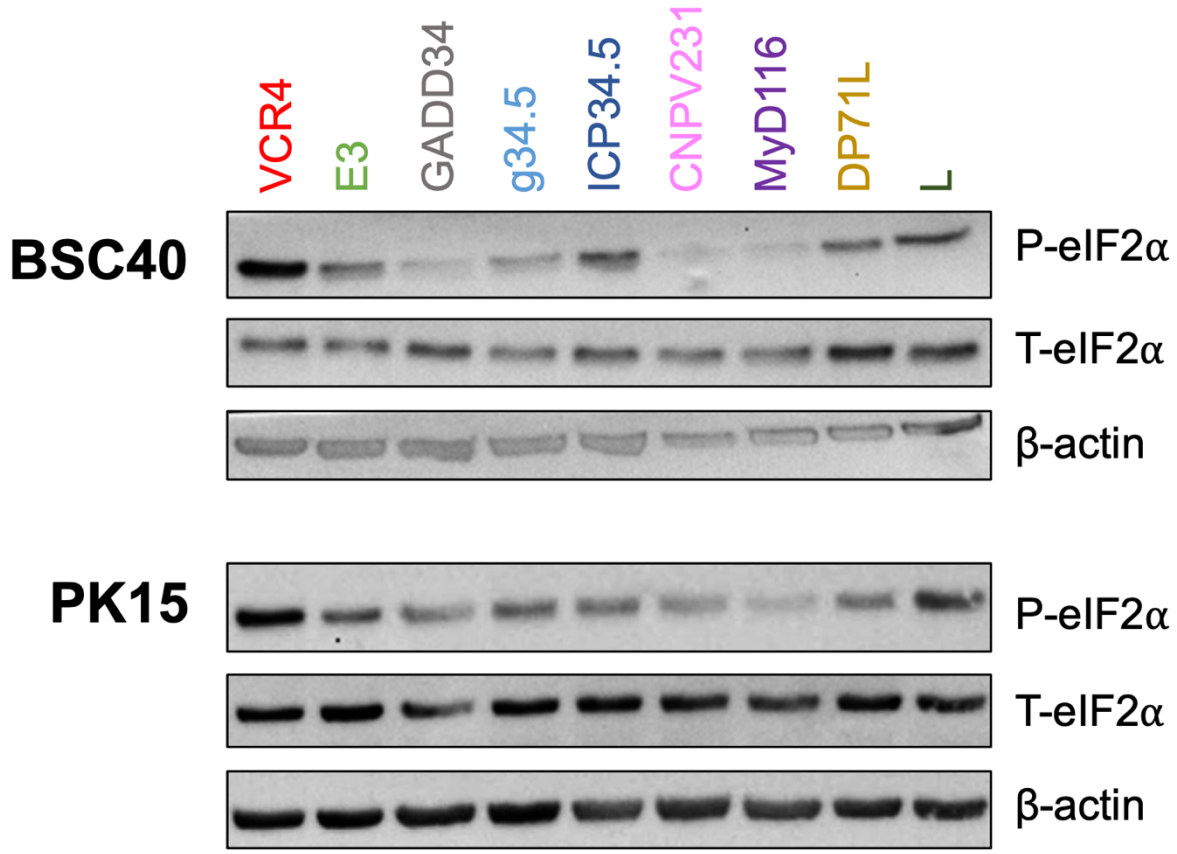

Figure 6: *eIF2α phosphorylation levels in cells infected with rVC-R4 expressing GADD34 and viral GADD34 homologs. BSC40 and PK15 cells were infected with rVC-R4 encoding GADD34 or the indicated viral GADD34 homolog at an MOI of 3. Lysates were collected at 6hpi, run on 10% SDS-PAGE gels, and analyzed by immunoblot analysis with an antibodies for phosphor- and total eIF2α. B-actin was used as a loading control.*

Establishing a system for exploring interaction partners of GADD34 and homologs

Results from the transfection-based luciferase reporter assay, plaque assays, and multiple cycle infectivity assays have indicated that cell line specificity may exist in the ability of viral GADD34 homologs to counteract PKR activity. It is surprising that such cell line specificity would exist for

GADD34 and homologs, given that both of the known interacting partners of GADD34, γ34.5, and DP71L – PP1C and eIF2α – are extremely highly conserved among host species. It is possible that GADD34 and viral GADD34 homologs interact with additional host proteins, which are less highly conserved in sequence among host species; differential interactions with these proteins in different host cell lines may underly the observed host species specificity among GADD34 homologs. In order to elucidate the cause of such host cell line specificity, we sought to identify additional GADD34 interacting partners and to compare the interactomes of viral GADD34 homologs. As a test run for establishing a system for evaluating interacting partners of GADD34 and homologs, FLAG-tagged GADD34 and γ34.5 were immunoprecipitated from HeLa cells transiently transfected with plasmids encoding each protein using anti-FLAG agarose beads, and eluents were sent for MS analysis. As a control for non-specific interacting partners, MS results for GADD34 and γ34.5 IP samples were compared to results for an IP from HeLa cells transiently transfected with FLAG-tagged VACV E3. It should be noted that E3 was not detected by Western blot in an IP sample run in parallel with the IP samples submitted for MS analysis. However, E3 was detected at extremely high intensities in two out of five replicate IP samples submitted for MS analysis.

GADD34 and γ34.5 were successfully detected at extremely high intensities in all five replicates of γ34.5 IP samples and 4 out of 5 replicates of GADD34 IP samples submitted (**figure 7**). Importantly, known GADD34 and γ34.5 interacting partners – eIF2α and the catalytic subunit of PP1 (PP1C) – were detected with significant log2-fold changes and p-values, relative to E3 samples, in γ34.5 IPs. Both PP1C and eIF2α were also detected in GADD34 IPs, but not at levels surpassing our assigned log2-fold threshold of 2 relative to E3 IPs, and with p-values greater than 0.05. This decreased detection of PP1C and eIF2α in GADD34 IPs compared to γ34.5 IPs may be the result of the fact that the IP buffer used was better optimized to co-IP of γ34.5 interacting partners than of GADD34 interacting partners, as determined by Coomassie staining of γ34.5 and GADD34 IPs run in SDS-PAGE (data not shown). The

detection of eIF2α in these IPs is somewhat unexpected, as the interactions with eIF2α are believed to be transient and thus difficult to capture via co-IP. The co-IP of eIF2α with both GADD34 and γ34.5 was confirmed via immunoblot analysis (**figure 8**). Several proteins, delineated in **table 1,** stood out as top hits for γ34.5, GADD34, and both GADD34 and γ34.5, having extremely high fold changes and extremely low p-values relative to E3. It is interesting that some proteins were highly enriched in the GADD34 pull-downs vs γ34.5, whereas the opposite was true for others. Whether these proteins are of any interest in understanding the mechanisms behind the observed virus and cell line specificity in GADD34 homolog activity remains to be explored.

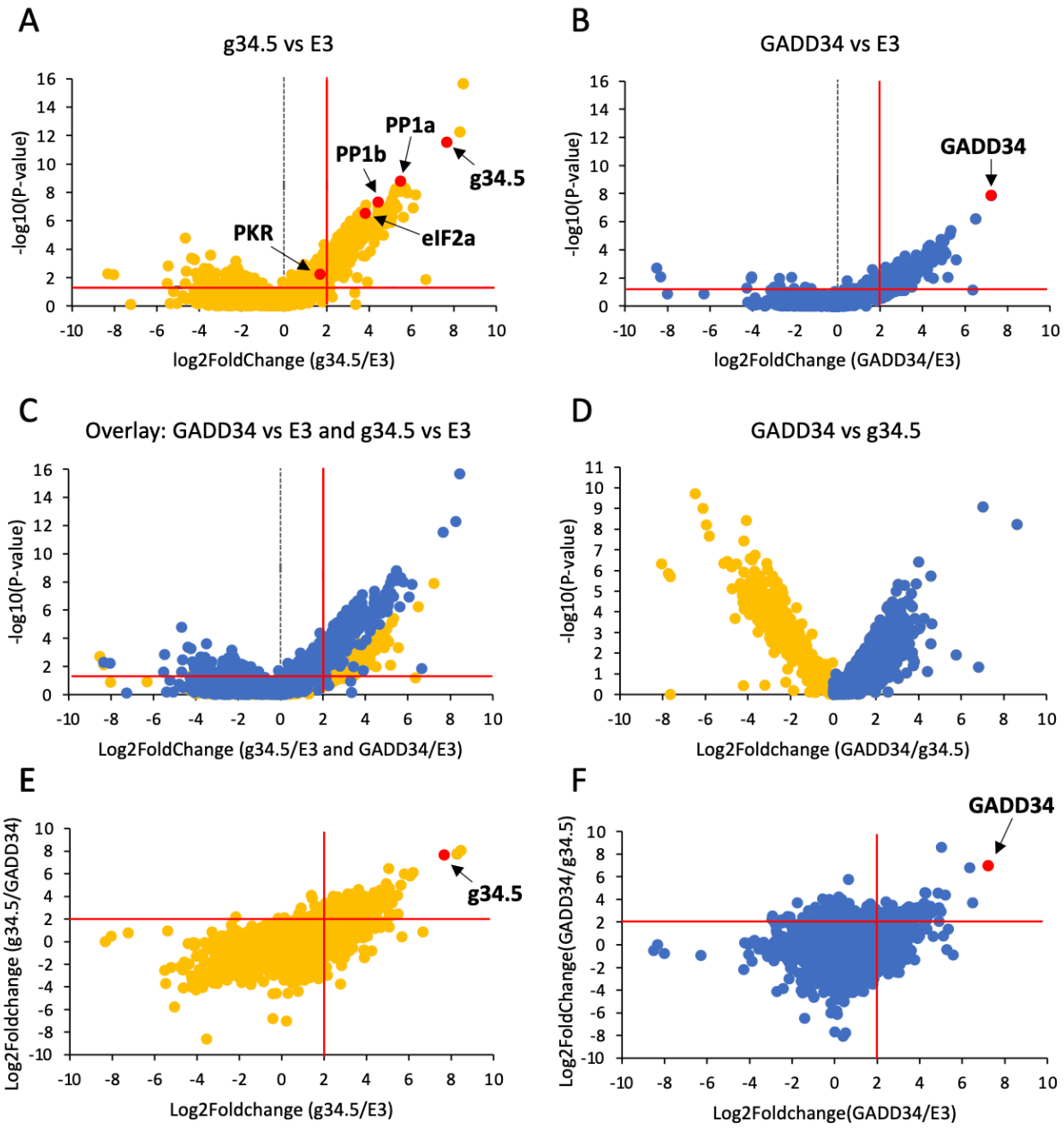

Figure 7: *Exploration of the GADD34 and γ34.5 interactomes. FLAG-tagged GADD34 and γ34.5 were transiently transfected into HeLa cells and immunoprecipitated using anti-FLAG agarose beads. IPs were sent for mass spec analysis, and results were analyzed using Simplifi software. Shown here are -log10(p-value) and log2foldchange, as determined via Excel analysis, for all analytes identified in IPs of (A) γ34.5 vs E3 (B) GADD34 vs E3. (C) overlay of graphs (A) and (B). (D) -log10(p-value) and log2foldchange for all analytes identified in IPs of GADD34 vs γ34.5. MS results for GADD34 and γ34.5 are further compared by graphing (E) log2foldchange(γ34.5/GADD34) vs log2foldchange(γ34.5/E3) and (F) log2foldchange(GADD34/γ34.5) vs log2foldchange(GADD34/E3).*

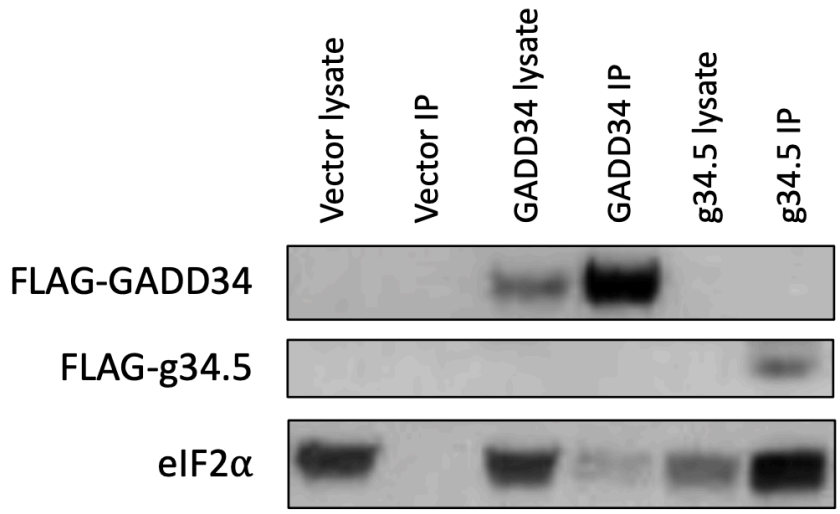

Figure 8: *Co-immunoprecipitation of eIF2α with GADD34 and γ34.5. FLAG-GADD34 and FLAG-γ34.5 were immunoprecipitated from transiently transfected HeLa cells. Total protein present in IP samples was separated on 10% SDS-PAGE and subjected to immunoblot analysis with anti-FLAG and anti-eIF2α antibodies. IP and immunoblot analysis were performed on HeLa cells transfected with empty vector (pSG5) as a control.*

Table 1: *MS hits with the ten highest fold changes for each of the following categories: GADD34/E3, GADD34/γ34.5, both GADD34/E3 and γ34.5/E3, γ34.5/E3, and γ34.5/GADD34*

			Average intensity			g34.5/E3		GADD34/E3		GADD34/g34.5	
	Accession	Description	E3	g34.5	GADD34	Log2 Foldchange	p-value	Log2 Foldchange	p-value	Log2 Foldchange	p-value
GADD34/E3	O75807	protein_phosphatase 1 regulatory subunit 15A	98	112	14466	0.23	7.3E-01	7.22	1.3E-08	8.60	5.9E-09
	Q8N490	Probable hydrolase PNKD	947	3085	41147	2.74	4.1E-05	6.49	6.0E-07	-4.55	1.3E+01
	P04632	Calpain small subunit 1	128	93	38184	-0.43	8.9E-01	6.35	6.8E-02	6.80	4.8E-02
	Q53H96	Pyrroline-5-carboxylate reductase 3	77	2486	4023	0.77	1.4E-01	5.19	8.6E-03	4.41	8.2E-02
	P07919	Cytochrome b-c1 complex subunit 6-like, mitochondrial	69	130	2893	4.23	3.3E-05	5.10	1.9E-04	-0.57	1.7E+00
	P48739	Phosphatidylinositol transfer protein beta isoform	1475	1363	36286	-3.56	1.4E-02	5.01	1.3E-04	-27.34	3.9E+02
	P00374	Dihydrofolate reductase	92	93	2960	2.16	1.2E-01	5.00	2.3E-05	-10.87	7.8E+00
	Q9UBX3	Mitochondrial dicarboxylate carrier	136	605	3875	2.83	1.2E-04	4.90	3.4E-05	-4.57	4.3E+00
	Q149N8	E3 ubiquitin-protein ligase SHPRH	68	236	2341	1.75	1.3E-01	4.89	4.8E-04	-7.38	9.7E+00
	P55084	Trifunctional enzyme subunit beta, mitochondrial	130	154	3641	-1.80	5.8E-01	4.86	1.8E-05	-19.01	2.4E+01
GADD34/g34.5	P48739	Phosphatidylinositol transfer protein beta isoform	1475	93	36286	-3.56	1.4E-02	5.01	1.3E-04	8.60	5.9E-09
	O75807	protein_phosphatase 1 regulatory subunit 15A	98	112	14466	0.23	7.3E-01	7.22	1.3E-08	7.00	8.7E-10
	P04632	Calpain small subunit 1	128	93	38184	-0.43	8.9E-01	6.35	6.8E-02	6.80	4.8E-02
	P63167	Dynein light chain 1, cytoplasmic;Dynein light chain 2, cytoplasmic	11906	240	39387	-5.07	6.6E-01	0.65	3.0E-01	5.76	1.3E-02
	A6NCE7	Microtubule-associated proteins 1A/1B light chain 3 beta 2	139	101	2642	-0.43	7.6E-01	4.23	2.4E-03	4.61	3.9E-04
	P55084	Trifunctional enzyme subunit beta, mitochondrial	130	154	3641	0.29	5.8E-01	4.23	2.4E-03	4.58	1.9E-06
	Q969M3	Protein YIPF5	161	131	4989	-0.25	8.8E-01	4.29	7.7E-03	4.58	3.7E-03
	Q53H96	Pyrroline-5-carboxylate reductase 3	77	130	4023	0.77	1.4E-01	5.19	8.6E-03	4.41	8.2E-02
	P23526	Adenosylhomocysteinase	2072	122	3053	-4.05	1.1E-02	0.19	5.6E-01	4.24	2.1E-04
	P50238	Cysteine-rich protein 1	2072	122	3053	-4.05	1.1E-02	-0.56	6.8E-01	4.07	5.5E-04

			Average intensity			g34.5/E3		GADD34/E3		GADD34/g34.5	
	Accession	Description	E3	g34.5	GADD34	Log2 Foldchange	p-value	Log2 Foldchange	p-value	Log2 Foldchange	p-value
Both GADD34 & g34.5/E3	Q99942	E3 ubiquitin-protein ligase RNF5	115	34813	5855	6.66	1.4E-02	5.56	4.9E-04	-0.86	2.3E-01
	P04004	Vitronectin	121	5893	4544	5.68	5.0E-09	5.26	7.8E-06	-0.42	2.4E-01
	O95478	Ribosome biogenesis protein NSA2 homolog	71	2285	1269	5.08	2.5E-07	3.75	4.5E-03	-1.30	2.7E-02
	Q03393	6-pyruvoyl tetrahydrobiopterin synthase	84	1837	844	4.65	8.0E-07	3.38	4.2E-03	-1.25	9.1E-03
	P07919	Cytochrome b-c1 complex subunit 6-like, mitochondrial	69	1363	2893	4.23	3.3E-05	5.10	1.9E-04	0.80	6.7E-01
	Q9UNE7	E3 ubiquitin-protein ligase CHIP	114	2072	1614	4.04	1.4E-04	3.60	3.2E-03	-0.49	4.7E-01
	Q96HS1	Serine/threonine-protein phosphatase PGAM5, mitochondrial	176	2486	7201	3.91	1.7E-06	5.33	4.2E-06	1.41	2.0E-01
	Q00587	Cdc42 effector protein 1	105	1548	2792	3.87	5.6E-06	4.62	9.1E-05	0.74	5.4E-01
	Q9NSI2	Protein FAM207A	87	1265	1084	3.85	7.1E-06	3.51	3.7E-03	-0.34	2.3E-01
	Q9NZJ7	Mitochondrial carrier homolog 1	66	638	1017	3.34	5.4E-05	3.90	1.2E-03	0.57	9.0E-01
g34.5/E3	Q07021	Complement component 1 Q subcomponent-binding protein, mitochondrial	187	64985	241	8.44	2.2E-16	0.38	3.9E-01	-8.06	5.0E-07
	Q5T1J5	Putative coiled-coil-helix-coiled-coil-helix domain-containing protein CHCHD2P9, mitochondrial	85	25106	117	8.27	5.5E-13	0.52	4.0E-01	-7.73	1.5E-06
	AWY10632.1	RL1_[Human alphaherpesvirus 1]	101	21257	104	7.66	3.0E-12	0.01	8.4E-01	-7.65	2.1E-06
	Q9NQT5	Exosome complex component RRP40	72	5710	84	6.19	1.5E-08	0.11	9.4E-01	-6.12	1.0E-09
	Q9BXS6	Nucleolar and spindle-associated protein 1	68	4997	73	6.07	1.2E-07	0.09	9.3E-01	-5.82	2.2E-08
	Q9BRT6	Protein LLP homolog	95	5415	88	5.79	1.1E-08	-0.19	8.4E-01	-5.97	6.5E-09
	O15479	Melanoma-associated antigen B2	68	3590	90	5.61	5.8E-07	0.44	5.9E-01	-4.99	3.7E-07
	Q13501	Sequestosome-1	123	5324	1022	5.48	8.3E-09	3.03	3.1E-03	-2.45	8.0E-04
	P62136	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	187	7441	435	5.47	1.6E-09	1.36	1.1E-01	-4.11	5.6E-05
	O75909	Cyclin-K	79	3056	226	5.32	1.2E-07	1.25	1.6E-01	-4.07	1.1E-05
g34.5/GADD34	Q07021	Complement component 1 Q subcomponent-binding protein, mitochondrial	187	64985	241	8.44	2.2E-16	0.38	3.9E-01	-8.06	5.0E-07
	Q5T1J5	Putative coiled-coil-helix-coiled-coil-helix domain-containing protein CHCHD2P9, mitochondrial	85	25106	117	8.27	5.5E-13	0.52	4.0E-01	-7.73	1.5E-06
	AWY10632.1	RL1_[Human alphaherpesvirus 1]	101	21257	104	7.66	3.0E-12	0.01	8.4E-01	-7.65	2.1E-06
	E9PRG8	Uncharacterized protein C11orf98	195	4888	55	5.06	4.6E-08	-1.43	2.9E-01	-6.48	1.9E-10
	Q9NQT5	Exosome complex component RRP40	72	5710	84	6.19	1.5E-08	0.11	9.4E-01	-6.12	1.0E-09
	Q9BRT6	Protein LLP homolog	95	5415	88	5.79	1.1E-08	-0.19	8.4E-01	-5.97	6.5E-09
	Q9BXS6	Nucleolar and spindle-associated protein 1	68	4997	73	6.07	1.2E-07	0.09	9.3E-01	-5.82	2.2E-08
	Q15633	RISC-loading complex subunit TARBP2	150	3797	113	4.94	7.5E-08	-0.19	5.9E-01	-5.15	4.6E-07
	O15479	Melanoma-associated antigen B2	68	3590	90	5.61	5.8E-07	0.44	5.9E-01	-4.99	3.7E-07
	O60353	Frizzled-6	88	2608	99	4.93	6.4E-07	0.15	8.6E-01	-4.78	6.8E-07

DISCUSSION:

Phosphorylation of eIF2α is an important regulatory mechanism by which eukaryotic cells respond to many different stressors, including viral infection. Activation of PKR and the ensuing translational repression are integral steps in the innate immune response to viral infection. Viruses have

evolved a a lot different mechanisms to circumvent PKR-mediated translational repression and successfully replicate in host cells. Such evasion of the host immune response often involves viral proteins referred to collectively as PKR antagonists. GADD34 homologs γ34.5 and DP71L are PKR antagonists, which counteract PKR activity indirectly, by decreasing the cellular pool of phosphorylated eIF2α. Our bioinformatic analysis suggests that multiple independent horizontal gene transfer events have distributed additional GADD34 homologs throughout diverse viral families. Here we sought to explore whether a subset of these putative viral GADD34 homologs function as true PKR antagonists in the context of viral infection.

We harnessed two well established experimental systems - a transfection-based luciferase reporter assay and a recombinant vaccinia virus (VACV) platform – to evaluate the anti-PKR pathway activity of GADD34, CReP, well studied viral GADD34 homologs γ34.5 and DP71L, avipoxvirus GADD34 homologs from Canarypox virus (CNPV) and a Puffinpox virus (PUPV) isolated in our lab, as well as newly identified putative homologs from fruitbat herpesvirus and a smooth-newt infecting picornavirus, livupivirusA1. Anti-PKR activity was seen in the luciferase assay for GADD34, CReP, FBAHV1 ICP34.5, CNPV231, PUPV MyD116, and DP71L, with each homolog increasing luciferase readout by at least 2-fold compared to transfection of PKR alone in at least one of the two cell lines tested. These results comprise the first evidence that the putative ORFs for PUPV MyD116 and FBAHV1 ICP34.5 are in fact functional proteins which can counteract PKR in a mammalian system. We provide further evidence that GADD34, γ34.5, ICP34.5 CNPV231, MyD116, and DP71L counteract PKR in the context of viral infection by showing their ability to rescue replication of a highly attenuated VACV lacking PKR antagonists K3 and E3. Whether the anti-PKR activity of these GADD34 homologs is essential for the replication of their parent viruses in their respective hosts cannot be extrapolated from this data alone. We have shown that these GADD34 homologs can functionally replace VACV PKR antagonists K3 and E3 and reverse the replication deficiency of VACV lacking both antagonists in cell lines from several

different host species. Lastly, we were able to show that eIF2α phosphorylation levels are decreased in infection with rVC-R4 encoding GADD34, γ34.5, ICP34.5 CNPV231, MyD116, and DP71L compared to infection with VC-R4 in both BSC40 and PK15 cells. Together, this data allows us to classify this panel of viral GADD34 homologs as bona fide, functional PKR pathway antagonists.

It is important to note that GADD34 was able to functionally replace K3 and E3 in VC-R4, rescuing replication completely in cell lines from several different host species. These results not only highlight why a horizontally acquired GADD34 homolog might have been maintained in a viral genome throughout exposure to evolutionary pressure but also exemplify the platform such a horizontal acquisition can provide for increases in virulence and host range. In the future, it would be interesting to use our rVC-R4_GADD34 virus to simulate the adaptive process that is believed to occur in viral genomes after the horizontal acquisition of host genes, by serially passaging the recombinant virus in PKR competent cells and evaluating subsequent adaptive changes in the architecture of both the GADD34 gene itself as well as other locations in the rVC-R4_GADD34 genome. Similar studies performed by our lab have provided valuable insight into the viral HGT process[69,70]. It is also important to note that both avipox virus GADD34 homologs fully rescued VC-R4 replication in several cell lines tested. These results provide evidence for our aforementioned hypothesis that GADD34 homologs functionally replace K3 and E3 in avipoxviruses, which lack both of these chordopoxvirus PKR antagonists. Furthermore, the ability of rVC-R4 encoding both avipox virus GADD34 homologs to replicate to high titers in cell lines from diverse host species – African green monkeys, sheep, and pigs – none of which are closely related to birds, highlights the need for a more thorough investigation of the seemingly broad host range of avipoxviruses and the molecular determinants of such.

The livupivirus L protein did not display any anti-PKR activity in the luciferase assay, nor was it able to rescue VC-R4 replication in most cell lines tested. This apparent lack of anti-PKR activity for L protein could simply be an artefact of poor expression, as was seen for L protein in infection of

RK13+E3+K3 and BSC40 cells with rVC-R4_L. However, the results from previous studies performed in our lab as well as the fact that DP71L rescues VC-R4 replication fully in BSC40 cells despite being detected at very low levels in western blot suggest that only a relatively low amount of protein, as detected by western blot, is sufficient to confer a phenotype in this assay. Furthermore, the partial rescue of VC-R4 by L protein seen in BSC40 cells shows that L protein is expressed at a high enough level in these cells to display some functionality, despite not being detectable by western blot. The low level of expression of L protein seen in our western blots could be related to the fact that in nature the L protein is translated as part of a larger polyprotein whereas in our experimental system it is translated as a single, distinct ORF. It is also possible that the experimental systems used here are not optimized for detecting anti-PKR activity of the livupivirus protein. Here we are evaluating L protein in the context of a poxvirus – a large dsDNA virus – which differs greatly in both replication cycle and gene expression program from picornaviruses, the family to which livupivirus belongs. We are also using cell lines from mammalian species, where livupivirus was isolated from an amphibian host. As one can imagine, there are a multitude of factors which could affect our readout of the anti-PKR activity of L protein in a system that is so foreign from the environment in which it is known to be expressed in nature.

Differences in anti-PKR activity can be seen among additional GADD34 homologs. In particular, HSV1 γ34.5 and FBAHV1 ICP34.5 display drastically different anti-PKR activity both in the luciferase assay and in their abilities to rescue VC-R4 replication, despite both being encoded by herpes simplexviruses. According to the infectivity data, γ34.5 rescues replication of VC-R4 to levels near the positive control in BSC40, OA1, and PK15 cells, displaying a broad cell line tropism, whereas the rVC-R4_ICP34.5 virus replicates to levels 1-3 logs lower than the rVC-R4_γ34.5 virus in these cell lines, displaying a restricted cell line tropism. We cannot necessarily extrapolate conclusions about the host range effects of these proteins from this data, as additional factors such as host tissue type and the immune response in the surrounding tissue or larger organ context must be taken into account to more

accurately predict host tropism. However, these results provide important initial evidence that GADD34 homologs may play a role in influencing viral host range; further investigations may allow us to categorize viral GADD34 homologs as host restriction factors. Furthermore, the differences in cell line tropism seen among our isogenic rVC-R4 viruses illustrate the potential of a single viral protein to incur broad differences in host range restriction between otherwise very closely related viruses.

There are several possible explanations for the differences in anti-PKR activity observed among GADD34 homologs. The extreme amino acid sequence divergence among our panel of GADD34 homologs most likely underlies differences in anti-PKR activity. Most obviously, differences in the eIF2α and PP1 binding motifs could dictate the eIF2α and PP1 binding capabilities of GADD34 homologs, which likely affects the level of eIF2α dephosphorylation each homolog can achieve. However, given the fact that γ34.5 and ICP34.5 only differ by 3 out of 19 positions in the eIF2α binding motif while γ34.5 and GADD34, which perform to comparable levels in the infectivity assays, differ at 10 of the 19 positions, differences in sequence in the PP1 and eIF2α binding regions likely do not entirely account for the observed differences in anti-PKR activity of these proteins.

The increased sequence divergence both N-terminal and C-terminal to the PP1 and eIF2α binding regions may affect the ability of GADD34 homologs to counteract PKR activity by determining subcellular localization of GADD34 homologs as well as their ability to interact with additional binding partners which may play currently unknown roles in eIF2α dephosphorylation complexes. Using the simplexvirus proteins γ34.5 and ICP34.5 as an example, sequence alignments show that in comparison to γ34.5, ICP34.5 (1) is missing PAT repeats entirely (2) differs in the number and positions of arginines in the N-terminal arginine-rich cluster and (3) differs in sequence in a putative bipartite nuclear localization signal (see chapter 1, **table 4**). Each of these motifs has been assigned a role in subcellular localization in γ34.5, and such differences in the ICP34.5 sequence may result in distinct subcellular localization of the two homologs. It is possible that ICP34.5 localizes to a nuclear or perinuclear region in which it comes

into contact with PP1 or eIF2α less frequently than does γ34.5. An additional possibility is that GADD34 homologs interact with additional proteins with which our panel of GADD34 homologs display differential binding affinities. Phosphatases are now understood to be targeted to specific substrates and subcellular locations via a diverse set of regulatory subunits, many of which may remain unidentified and/or unexplored[97,98]. Further, as PP1 regulation and targeting becomes more well understood, it has been proposed that most functional PP1 phosphatase complexes are often trimeric in nature[99]. It is possible that GADD34 and PP1 form a trimeric eIF2α phosphatase complex with an additional PP1 interacting protein (PIP). One such candidate is the protein phosphatase 1 regulatory subunit 1A (PR1A), otherwise known as Inhibitor 1 (I1), which has been shown to complex with GADD34, and is believed to form a heterotrimer with GADD34 and PP1. Pull-down assays have shown that PR1A binds somewhere in the central region of GADD34 (amino acids 180 to 483), where the four PEST repeats are found[98]. The viral GADD34 homologs used in our experiments display high sequence divergence at regions corresponding to the GADD34 PEST motifs and therefore likely display high variance in their abilities to bind PR1A, if they can bind at all. As PR1A is an inhibitor of PP1, we should see a negative correlation between PR1A binding ability and eIF2α dephosphorylation among GADD34 homologs. It is also possible that an additional interacting partner binds to the region directly C-terminal to the eIF2α binding region, which for GADD34 has been shown to have stable secondary structure – forming alpha helix 2 – and is thus a likely site for protein-protein interactions. Divergence in sequence is seen in this region among viral GADD34 homologs from different families, and once again could dictate relative binding abilities. Further investigation is needed to identify additional GADD34 interacting partners, map binding regions, and compare the binding abilities among GADD34 homologs. Such investigation can be performed using our established immunoprecipitation and mass spectrometry system.

There is also a possibility that GADD34 homologs interact preferentially with the different isoforms of the PP1 catalytic subunit (PP1C): alpha, beta, and gamma. While the PP1C alpha beta and

gamma isoforms are relatively conserved in sequence, some PIPs have been shown to distinguish between PP1C isoforms via their distinct C-terminal "tails"[100]. Differential subcellular localization of GADD34 homologs may also affect which PP1C isoforms they most frequently come into contact with, as the PP1C isoforms are believed to have distinct subcellular and subnuclear localization patterns[101]. However, whether differential interaction with the PP1C isoforms among GADD34 homologs would affect their anti-PKR activity is unclear. The PP1C isoforms share high sequence identity within their catalytic domains, and have similar catalytic capabilities and substrate specificity in vitro, suggesting that differential interaction of GADD34 homologs with the PP1C isoforms may not contribute to different levels of eIF2α dephosphorylation[97]. In our pilot IP experiment, PP1C isoforms alpha and beta were detected at comparable levels in both GADD34 and γ34.5 pull downs, whereas the gamma isoform was not detected at all in either. These preliminary results suggest there may be some selectivity among PP1C isoforms for GADD34 and γ34.5, and whether a similar or a different trend in selectivity is seen for other GADD34 homologs should be investigated with additional IP and mass spectrometry experiments.

Interestingly, the relative abilities of GADD34 homologs to counteract PKR in the luciferase assay do not correspond to their abilities to rescue replication of VC-R4. The most striking example of this is FBAHV1 ICP34.5. In HeLa PKR KO cells, ICP34.5 causes the highest increase in luciferase readout among the GADD34 homologs tested, but rVC-R4 encoding ICP34.5 replicates to titers between 1 and 3 logs lower than most other rVC-R4 viruses in BSC40, OA1, and PK15 cells. One possible explanation for this discrepancy in results between assays is the use of cell lines from different host species in the luciferase versus the infectivity assays; each cell line provides a completely different cellular context which may influence our measurement of anti-PKR activity for these proteins. This difference in cellular environment could underly the distinct pattern in anti-PKR activity in the luciferase assay seen among GADD34 homologs in HeLa cells versus RK13 cells. Unfortunately, we lack sufficient reliable viral infectivity data in HeLa and RK13 cell lines for this panel of rVC-R4 viruses to determine whether the apparently cell

line-dependent differential patterns in anti-PKR activity seen in the luciferase assay results correspond to patterns in viral titer in HeLa and RK13 cell lines.

Cell line specificity was also seen in the anti-PKR activity of GADD34 and CReP, both of which increased luciferase readout several-fold compared to PKR alone in HeLa PKR KO, but appeared to have no anti-PKR activity in RK13 PKR KO cells. Another drastic example of cell line specificity is seen in the ability of rVC-R4_ICP34.5 to form plaques in RK13 cells, but not in BSC40 or OA1, and yet again in the infectivity assays, where the livupivirus L protein provides partial rescue of VC-R4 replication in BSC40 cells, but no rescue at all in OA1 and PK15 cells. A similar pattern, although less pronounced, was seen for ICP34.5. Host specificity among PKR antagonists is not a novel concept. Vaccinia virus PKR antagonists K3 and E3, for example, are important for viral replication in cell lines from some host species but not others, making them host range factors. Both of these proteins interact directly with PKR – a host restriction factor under intense positive selection – and the PKR-K3 interface has been shown to be important for host species specificity[94]. However, the cell line dependency in the anti-PKR activity of GADD34 homologs is unexpected given the proposed mechanism of these proteins. It is unlikely that interactions of GADD34 homologs with their currently identified interacting partners – eIF2α and PP1 – are the source of such cell line specificity, because unlike PKR, eIF2α and PP1 are highly conserved among host species[102]. In fact, the PP1 sequences from humans, African green monkeys, pigs, and rabbits are 100% identical, and the eIF2α sequences from the aforementioned species are over 99% identical. Thus, we hypothesize that the cell line specific differences we observed in the anti-PKR activity of GADD34 homologs are likely the result of so far uncharacterized interactions between these viral GADD34 homologs and additional cellular proteins, such as the aforementioned PP1 inhibitor, PR1A, which differs in sequence slightly more than PP1 and eIF2α do among host species.

Another possibility is that GADD34 homologs, like VACV K3 and E3, interact directly with PKR, the well-established host restriction factor. GADD34 and PKR are likely to exist in close proximity within

the cell, as both proteins bind eIF2α. It is possible that some contact is made between GADD34 homologs and PKR in competing to bind eIF2α. It is also possible that PP1 actually dephosphorylates PKR in addition to eIF2α, thus inactivating PKR. PP1 has been shown to bind PKR in yeast two hybrid screens and pull down assays in mammalian cells[103]. GADD34 homologs may play an organized role in such dephosphorylation, which involves direct interaction with PKR. Through any of these mechanisms, the direct interaction of GADD34 homologs with PKR regions displaying high sequence divergence could explain the observed cell line dependence in our results. It should be noted that PKR was co-immunoprecipitated with both GADD34 and γ34.5 in our pilot MS experiments, albeit at levels lower than our threshold of significance. Without further experiments such as proximity labeling, we cannot determine whether this co-immunoprecipitation represents a direct interaction with or mere proximity of GADD34 homologs to PKR. In addition, we cannot rule out that some of these apparent difference in phenotype between cell lines may simply be the result of the fact that the viral replication cycle may follow different timelines in different cell lines. All infections were performed for the same amount of time, and it is possible that, for example, the phenotype for BSC40 cells might look more like the phenotype for OA1 and PK15 cells if lysates are collected at an earlier timepoint. Lastly, it is possible that the observed cell line specificity is not so much a product of the cells being from different host species as it is a product of them being of different tissue types. As the expression levels of PP1C isoforms have been shown to vary by tissue type, it is conceivable that tissue-specific differences between the PP1C isoform with which GADD34 homologs primarily come into contact with are affecting our measurements of anti-PKR activity [104]. However, as mentioned before, this explanation is predicated on the assumption that the different PP1C isoforms have varying catalytic capacities, which based on sequence analysis and in vitro assays seems unlikely.

Overall, we propose that differences in GADD34 activity within the same cell line are dictated by a combination of three factors: sequence divergence in the PP1 and eIF2α binding regions; subcellular

localization and resulting proximity to both PP1C and eIF2α; and differential interactions with additional, currently unidentified, interacting partners of GADD34 and homologs. Furthermore, we propose that one or more of these currently unidentified interacting partners serve as a host restriction factor, similarly to how the direct interaction of PKR with VACV K3 plays a role in determining VACV host tropism. These results further our understanding of the complex network of molecular interactions between host and viral proteins comprising the molecular arms race which controls viral virulence, host range, and transmission to novel host species.

REFERENCES

1. The integrated stress response. *EMBO Rep.* **17**, 1374–1395 (2016).
2. Wu, S. & Kaufman, R. J. A model for the double-stranded RNA (dsRNA)-dependent dimerization and activation of the dsRNA-activated protein kinase PKR. *J. Biol. Chem.* **272**, 1291–1296 (1997).
3. Lee, Y.-Y., Cevallos, R. C. & Jan, E. An Upstream Open Reading Frame Regulates Translation of GADD34 during Cellular Stresses That Induce eIF2α Phosphorylation *. *J. Biol. Chem.* **284**, 6661–6673 (2009).
4. Ma, Y. & Hendershot, L. M. Delineation of a Negative Feedback Regulatory Loop That Controls Protein Translation during Endoplasmic Reticulum Stress *. *J. Biol. Chem.* **278**, 34864–34873 (2003).
5. Brush, M. H., Weiser, D. C. & Shenolikar, S. Growth Arrest and DNA Damage-Inducible Protein GADD34 Targets Protein Phosphatase 1α to the Endoplasmic Reticulum and Promotes Dephosphorylation of the α Subunit of Eukaryotic Translation Initiation Factor 2. *Mol. Cell. Biol.* **23**, 1292–1303 (2003).
6. Jousse, C. *et al.* Inhibition of a constitutive translation initiation factor 2α phosphatase, CReP, promotes survival of stressed cells. *J. Cell Biol.* **163**, 767–775 (2003).
7. Rojas, M., Vasconcelos, G. & Dever, T. E. An eIF2α-binding motif in protein phosphatase 1 subunit GADD34 and its viral orthologs is required to promote dephosphorylation of eIF2α. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3466-3475 (2015).
8. Choy, M. S. *et al.* Structural and Functional Analysis of the GADD34:PP1 eIF2α Phosphatase. *Cell Rep.* **11**, 1885–1891 (2015).

9. Chen, R. *et al.* G-actin provides substrate-specificity to eukaryotic initiation factor 2α holophosphatases. *eLife* **4**, e04871 (2015).
10. Zhan, Q. *et al.* The gadd and MyD genes define a novel set of mammalian genes encoding acidic proteins that synergistically suppress cell growth. *Mol. Cell. Biol.* **14**, 2361–2371 (1994).
11. Shi, W. *et al.* GADD34–PP1c recruited by Smad7 dephosphorylates TGFβ type I receptor. *J. Cell Biol.* **164**, 291–300 (2004).
12. Novoa, I., Zeng, H., Harding, H. P. & Ron, D. Feedback Inhibition of the Unfolded Protein Response by GADD34-Mediated Dephosphorylation of eIF2α. *J. Cell Biol.* **153**, 1011–1022 (2001).
13. Hollander, M. C., Poola-Kella, S. & Fornace, A. J. Gadd34 functional domains involved in growth suppression and apoptosis. *Oncogene* **22**, 3827–3832 (2003).
14. Harding, H. P. *et al.* Ppp1r15 gene knockout reveals an essential role for translation initiation factor 2 alpha (eIF2α) dephosphorylation in mammalian development. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1832–1837 (2009).
15. Xu, X. *et al.* Preemptive Activation of the Integrated Stress Response Protects Mice From Diet-Induced Obesity and Insulin Resistance by Fibroblast Growth Factor 21 Induction. *Hepatol. Baltim. Md* **68**, 2167–2181 (2018).
16. He, B., Gross, M. & Roizman, B. The gamma134.5 protein of herpes simplex virus 1 has the structural and functional attributes of a protein phosphatase 1 regulatory subunit and is present in a high molecular weight complex with the enzyme in infected cells. *J. Biol. Chem.* **273**, 20737–20743 (1998).

17. Chou, J. & Roizman, B. The gamma 1(34.5) gene of herpes simplex virus 1 precludes neuroblastoma cells from triggering total shutoff of protein synthesis characteristic of programed cell death in neuronal cells. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3266–3270 (1992).
18. Rivera, J. *et al.* The MyD116 African swine fever virus homologue interacts with the catalytic subunit of protein phosphatase 1 and activates its phosphatase activity. *J. Virol.* **81**, 2923–2929 (2007).
19. Zhang, F., Moon, A., Childs, K., Goodbourn, S. & Dixon, L. K. The African Swine Fever Virus DP71L Protein Recruits the Protein Phosphatase 1 Catalytic Subunit To Dephosphorylate eIF2α and Inhibits CHOP Induction but Is Dispensable for These Activities during Virus Infection. *J. Virol.* **84**, 10681–10689 (2010).
20. Chou, J. & Roizman, B. Herpes simplex virus 1 gamma(1)34.5 gene function, which blocks the host response to infection, maps in the homologous domain of the genes expressed during growth arrest and DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5247–5251 (1994).
21. He, B., Chou, J., Liebermann, D. A., Hoffman, B. & Roizman, B. The carboxyl terminus of the murine MyD116 gene substitutes for the corresponding domain of the gamma(1)34.5 gene of herpes simplex virus to preclude the premature shutoff of total protein synthesis in infected human cells. *J. Virol.* **70**, 84–90 (1996).
22. Barber, C. *et al.* Identification of residues within the African swine fever virus DP71L protein required for dephosphorylation of translation initiation factor eIF2α and inhibiting activation of pro-apoptotic CHOP. *Virology* **504**, 107–113 (2017).
23. Rechsteiner, M. PEST sequences are signals for rapid intracellular proteolysis. *Semin. Cell Biol.* **1**, 433–440 (1990).

24. The extended PP1 toolkit: designed to create specificity | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0968000410000460?token=47EC0AAC54F53D81E0C47B189C274975FB0380F5A94BDA229C5F426BC8402B53F64DAA5261BC6FCB01DF75DF00181931&originRegion=us-east-1&originCreation=20220330200847 doi:10.1016/j.tibs.2010.03.002.
25. O'Connell, N. *et al.* The molecular basis for substrate specificity of the nuclear NIPP1:PP1 holoenzyme. *Struct. Lond. Engl. 1993* **20**, 1746–1756 (2012).
26. Choy, M. S. *et al.* Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code. *Proc. Natl. Acad. Sci.* **111**, 4097–4102 (2014).
27. Inagaki, T. *et al.* Characterization of a Novel Alphaherpesvirus Isolated from the Fruit Bat Pteropus lylei in Vietnam. *J. Virol.* **94**, e00673-20 (2020).
28. Davis, K. L., Korom, M. & Morrison, L. A. Herpes simplex virus 2 ICP34.5 confers neurovirulence by regulating the type I interferon response. *Virology* **468–470**, 330–339 (2014).
29. Km, W., Je, S. & La, M. Increased eIF2alpha phosphorylation attenuates replication of herpes simplex virus 2 vhs mutants in mouse embryonic fibroblasts and correlates with reduced accumulation of the PKR antagonist ICP34.5. *J. Virol.* **83**, (2009).
30. Sasaki, M. *et al.* Isolation and Characterization of a Novel Alphaherpesvirus in Fruit Bats. *J. Virol.* **88**, 9819–9829 (2014).
31. Tyler, S., Severini, A., Black, D., Walker, M. & Eberle, R. Structure and sequence of the saimiriine herpesvirus 1 genome. *Virology* **410**, 181–191 (2011).

32. Severini, A., Tyler, S. D., Peters, G. A., Black, D. & Eberle, R. Genome sequence of a chimpanzee herpesvirus and its relation to other primate alphaherpesviruses. *Arch. Virol.* **158**, 1825–1828 (2013).

33. Pfaff, F. *et al.* Full genome sequence of bovine alphaherpesvirus 2 (BoHV-2). *Arch. Virol.* **166**, 639–643 (2021).

34. Eberle, R. & Black, D. H. Sequence of the ateline alphaherpesvirus 1 (HVA1) genome. *Arch. Virol.* **162**, 1423–1425 (2017).

35. Tyler, S. D., Peters, G. A. & Severini, A. Complete genome sequence of cercopithecine herpesvirus 2 (SA8) and comparison with other simplexviruses. *Virology* **331**, 429–440 (2005).

36. Babra, B. *et al.* Analysis of the genome of leporid herpesvirus 4. *Virology* **433**, 183–191 (2012).

37. Perelygina, L. *et al.* Complete sequence and comparative analysis of the genome of herpes B virus (Cercopithecine herpesvirus 1) from a rhesus monkey. *J. Virol.* **77**, 6167–6177 (2003).

38. Tyler, S. D. & Severini, A. The complete genome sequence of herpesvirus papio 2 (Cercopithecine herpesvirus 16) shows evidence of recombination events among various progenitor herpesviruses. *J. Virol.* **80**, 1214–1221 (2006).

39. Cheng, G., Brett, M.-E. & He, B. Signals That Dictate Nuclear, Nucleolar, and Cytoplasmic Shuttling of the γ134.5 Protein of Herpes Simplex Virus Type 1. *J. Virol.* **76**, 9434–9445 (2002).

40. Mao, H. & Rosenthal, K. S. An N-terminal arginine-rich cluster and a proline-alanine-threonine repeat region determine the cellular localization of the herpes simplex virus type 1

ICP34.5 protein and its ligand, protein phosphatase 1. *J. Biol. Chem.* **277**, 11423–11431 (2002).

41. Jing, X. & He, B. Characterization of the triplet repeats in the central domain of the γ134·5 protein of herpes simplex virus 1. *J. Gen. Virol.* **86**, 2411–2419 (2005).
42. Cheng, G., Gross, M., Brett, M.-E. & He, B. AlaArg Motif in the Carboxyl Terminus of the γ134.5 Protein of Herpes Simplex Virus Type 1 Is Required for the Formation of a High-Molecular-Weight Complex That Dephosphorylates eIF-2α. *J. Virol.* **75**, 3666–3674 (2001).
43. Bányai, K. *et al.* Unique genomic organization of a novel Avipoxvirus detected in turkey (Meleagris gallopavo). *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **35**, 221–229 (2015).
44. Sarker, S., Athukorala, A., Bowden, T. R. & Boyle, D. B. Genomic Characterisation of a Novel Avipoxvirus Isolated from an Endangered Yellow-Eyed Penguin (Megadyptes antipodes). *Viruses* **13**, 194 (2021).
45. Offerman, K., Carulei, O., Gous, T. A., Douglass, N. & Williamson, A.-L. Phylogenetic and histological variation in avipoxviruses isolated in South Africa. *J. Gen. Virol.* **94**, 2338–2351 (2013).
46. Afonso, C. L. *et al.* The Genome of Fowlpox Virus. *J. Virol.* **74**, 3815–3831 (2000).
47. Tulman, E. R. *et al.* The Genome of Canarypox Virus. *J. Virol.* **78**, 353–366 (2004).
48. Offerman, K., Carulei, O., van der Walt, A. P., Douglass, N. & Williamson, A.-L. The complete genome sequences of poxviruses isolated from a penguin and a pigeon in South Africa and comparison to other sequenced avipoxviruses. *BMC Genomics* **15**, 463 (2014).
49. Sarker, S. *et al.* Genomic characterization of two novel pathogenic avipoxviruses isolated from pacific shearwaters (Ardenna spp.). *BMC Genomics* **18**, 298 (2017).

50. Carulei, O., Douglass, N. & Williamson, A.-L. Comparative analysis of avian poxvirus genomes, including a novel poxvirus from lesser flamingos (Phoenicopterus minor), highlights the lack of conservation of the central region. *BMC Genomics* **18**, 947 (2017).
51. Sarker, S. *et al.* Molecular characterisation of a novel pathogenic avipoxvirus from the Australian magpie (Gymnorhina tibicen). *Virology* **540**, 1–16 (2020).
52. Sarker, S., Athukorala, A. & Raidal, S. R. Molecular characterisation of a novel pathogenic avipoxvirus from an Australian passerine bird, mudlark (Grallina cyanoleuca). *Virology* **554**, 66–74 (2021).
53. McGraw, K., Penha, V. A. de S., Drake, D., Kraberger, S. & Varsani, A. Poxvirus infection in house finches (Haemorhous mexicanus): genome sequence analysis and patterns of infection in wild birds. *Transbound. Emerg. Dis.* (2022) doi:10.1111/tbed.14575.
54. Sarker, S., Hannon, C., Athukorala, A. & Bielefeldt-Ohmann, H. Emergence of a Novel Pathogenic Poxvirus Infection in the Endangered Green Sea Turtle (Chelonia mydas) Highlights a Key Threatening Process. *Viruses* **13**, 219 (2021).
55. Pankovics, P. *et al.* Genetic characterization of a second novel picornavirus from an amphibian host, smooth newt (Lissotriton vulgaris). *Arch. Virol.* **162**, 1043–1050 (2017).
56. Herniou, E. A., Olszewski, J. A., Cory, J. S. & O'Reilly, D. R. The genome sequence and evolution of baculoviruses. *Annu. Rev. Entomol.* **48**, 211–234 (2003).
57. Yin, F. *et al.* The Complete Genome of a New Betabaculovirus from Clostera anastomosis. *PloS One* **10**, e0132792 (2015).
58. Escasa, S. R., Lauzon, H. A. M., Mathur, A. C., Krell, P. J. & Arif, B. M. Sequence analysis of the Choristoneura occidentalis granulovirus genome. *J. Gen. Virol.* **87**, 1917–1933 (2006).

59. İnce, İ. A., Özcan, O., Ilter-Akulke, A. Z., Scully, E. D. & Özgen, A. Invertebrate Iridoviruses: A Glance over the Last Decade. *Viruses* **10**, E161 (2018).

60. Kaletta, T. & Hengartner, M. O. Finding function in novel targets: C. elegans as a model organism. *Nat. Rev. Drug Discov.* **5**, 387–399 (2006).

61. Lapierre, L. R. & Hansen, M. Lessons from C. elegans: signaling pathways for longevity. *Trends Endocrinol. Metab.* **23**, 637–644 (2012).

62. Clack, J. A. The Fin to Limb Transition: New Data, Interpretations, and Hypotheses from Paleontology and Developmental Biology. *Annu. Rev. Earth Planet. Sci.* **37**, 163–179 (2009).

63. Amemiya, C. T. *et al.* Analysis of the African coelacanth genome sheds light on tetrapod evolution. *Nature* **496**, 311–316 (2013).

64. The earliest known stem-tetrapod from the Lower Devonian of China | Nature Communications. https://www.nature.com/articles/ncomms2170.

65. Hurles, M. Gene Duplication: The Genomic Trade in Spare Parts. *PLOS Biol.* **2**, e206 (2004).

66. MAGADUM, S., BANERJEE, U., MURUGAN, P., GANGAPUR, D. & RAVIKESAVAN, R. Gene duplication as a major force in evolution. *J. Genet.* **92**, 155–161 (2013).

67. Moreira, D. & López-García, P. Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* **7**, 306–311 (2009).

68. Hughes, A. L. & Friedman, R. Poxvirus genome evolution by gene gain and loss. *Mol. Phylogenet. Evol.* **35**, 186–195 (2005).

69. Brennan, G., Kitzman, J. O., Rothenburg, S., Shendure, J. & Geballe, A. P. Adaptive Gene Amplification As an Intermediate Step in the Expansion of Virus Host Range. *PLoS Pathog.* **10**, (2014).

70. Rahman, M. J. *et al.* LINE-1 retrotransposons facilitate horizontal gene transfer into poxviruses. 2020.10.26.355610 Preprint at https://doi.org/10.1101/2020.10.26.355610 (2021).

71. Odom, M. R., Curtis Hendrickson, R. & Lefkowitz, E. J. Poxvirus protein evolution: Family wide assessment of possible horizontal gene transfer events. *Virus Res.* **144**, 233–249 (2009).

72. Hughes, A. L. Origin of ecdysosteroid UDP-glycosyltransferases of baculoviruses through horizontal gene transfer from Lepidoptera. *Coevolution* **1**, 1–7 (2013).

73. Schönrich, G., Abdelaziz, M. O. & Raftery, M. J. Herpesviral capture of immunomodulatory host genes. *Virus Genes* **53**, 762–773 (2017).

74. Hughes, A. L. & Friedman, R. Genome-Wide Survey for Genes Horizontally Transferred from Cellular Organisms to Baculoviruses. *Mol. Biol. Evol.* **20**, 979–987 (2003).

75. Kariithi, H. M. *et al.* Coevolution of hytrosaviruses and host immune responses. *BMC Microbiol.* **18**, 183 (2018).

76. Cheng, X.-W., Wan, X.-F., Xue, J. & Moore, R. C. Ascovirus and its evolution. *Virol. Sin.* **22**, 137 (2008).

77. Starrett, G. J. *et al.* Adintoviruses: A Proposed Animal-Tropic Family of Midsize Eukaryotic Linear dsDNA (MELD) Viruses. *Virus Evol.* (2020) doi:10.1093/ve/veaa055.

78. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of cells. *Biol. Direct* **1**, 29 (2006).

79. Shackelton, L. A. & Holmes, E. C. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol.* **12**, 458–465 (2004).

80. Gaudreault, N. N., Madden, D. W., Wilson, W. C., Trujillo, J. D. & Richt, J. A. African Swine Fever Virus: An Emerging DNA Arbovirus. *Front. Vet. Sci.* **7**, (2020).

81. Thézé, J., Takatsuka, J., Nakai, M., Arif, B. & Herniou, E. A. Gene Acquisition Convergence between Entomopoxviruses and Baculoviruses. *Viruses* **7**, 1960–1974 (2015).

82. Complete Genome Analysis of Three Novel Picornaviruses from Diverse Bat Species | Journal of Virology. https://journals.asm.org/doi/10.1128/JVI.02364-10.

83. Error Frequencies of Picornavirus RNA Polymerases: Evolutionary Implications for Virus Populations - Molecular Biology of Picornavirus - Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555817916.ch23.

84. Evolutionary and Structural Overview of Human Picornavirus Capsid Antibody Evasion - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6710328/.

85. Rothenburg, S. & Brennan, G. Species-Specific Host–Virus Interactions: Implications for Viral Host Range and Virulence. *Trends Microbiol.* **28**, 46–56 (2020).

86. Park, C., Peng, C., Brennan, G. & Rothenburg, S. Species-specific inhibition of antiviral protein kinase R by capripoxviruses and vaccinia virus. *Ann. N. Y. Acad. Sci.* **1438**, 18–29 (2019).

87. Langland, J. O. & Jacobs, B. L. The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range. *Virology* **299**, 133–141 (2002).

88. Carpentier, K. S., Esparo, N. M., Child, S. J. & Geballe, A. P. A Single Amino Acid Dictates Protein Kinase R Susceptibility to Unrelated Viral Antagonists. *PLOS Pathog.* **12**, e1005966 (2016).

89. Myxoma virus M156 is a specific inhibitor of rabbit PKR but contains a loss-of-function mutation in Australian virus isolates | PNAS. https://www.pnas.org/content/113/14/3855.
90. Liu, R. & Moss, B. Opposing Roles of Double-Stranded RNA Effector Pathways and Viral Defense Proteins Revealed with CRISPR-Cas9 Knockout Cell Lines and Vaccinia Virus Mutants. *J. Virol.* **90**, 7864–7879 (2016).
91. Peng, C., Haller, S. L., Rahman, M. M., McFadden, G. & Rothenburg, S. Myxoma virus M156 is a specific inhibitor of rabbit PKR but contains a loss-of-function mutation in Australian virus isolates. *Proc. Natl. Acad. Sci.* **113**, 3855–3860 (2016).
92. Rothenburg, S., Seo, E. J., Gibbs, J. S., Dever, T. E. & Dittmar, K. Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. *Nat. Struct. Mol. Biol.* **16**, 63–70 (2009).
93. Vipat, S., Brennan, G., Haller, S. L. & Rothenburg, S. *Rapid, seamless generation of recombinant poxviruses using host-range and visual selection*. http://biorxiv.org/lookup/doi/10.1101/467514 (2018) doi:10.1101/467514.
94. Park, C. *et al.* Orthopoxvirus K3 orthologs show virus- and host-specific inhibition of the antiviral protein kinase PKR. *bioRxiv* 2020.02.20.958645 (2020) doi:10.1101/2020.02.20.958645.
95. Vipat, S., Brennan, G., Park, C., Haller, S. L. & Rothenburg, S. Rapid, Seamless Generation of Recombinant Poxviruses using Host Range and Visual Selection. *J. Vis. Exp. JoVE* (2020) doi:10.3791/61049.
96. Park, C. *et al.* Orthopoxvirus K3 orthologs show virus- and host-specific inhibition of the antiviral protein kinase PKR. *PLoS Pathog.* **17**, e1009183 (2021).

97. Cohen, P. T. W. Protein phosphatase 1 – targeted in many directions. *J. Cell Sci.* **115**, 241–256 (2002).

98. Connor, J. H., Weiser, D. C., Li, S., Hallenbeck, J. M. & Shenolikar, S. Growth arrest and DNA damage-inducible protein GADD34 assembles a novel signaling complex containing protein phosphatase 1 and inhibitor 1. *Mol. Cell. Biol.* **21**, 6841–6850 (2001).

99. Virshup, D. M. & Shenolikar, S. From promiscuity to precision: protein phosphatases get a makeover. *Mol. Cell* **33**, 537–545 (2009).

100. Bertran, M. T. *et al.* ASPP proteins discriminate between PP1 catalytic subunits through their SH3 domain and the PP1 C-tail. *Nat. Commun.* **10**, 771 (2019).

101. Shenolikar, S. Protein serine/threonine phosphatases--new avenues for cell regulation. *Annu. Rev. Cell Biol.* **10**, 55–86 (1994).

102. Kimball, S. R. Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.* **31**, 25–29 (1999).

103. Tan, S.-L., Tareen, S. U., Melville, M. W., Blakely, C. M. & Katze, M. G. The direct binding of the catalytic subunit of protein phosphatase 1 to the PKR protein kinase is necessary but not sufficient for inactivation and disruption of enzyme dimer formation. *J. Biol. Chem.* **277**, 36109–36117 (2002).

104. Shima, H. *et al.* Identification of PP1 catalytic subunit isotypes PP1 gamma 1, PP1 delta and PP1 alpha in various rat tissues. *Biochem. Biophys. Res. Commun.* **192**, 1289–1296 (1993).