Effects of altering CpG dinucleotide composition in small DNA viruses



Lisa Loew

Kellogg College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2020/2021

Abstract

Effects of altering CpG dinucleotide composition in small DNA viruses

A thesis submitted for the degree of *Doctor of Philosophy*

Lisa Loew, Kellogg College Michaelmas Term 2020

CpG dinucleotides are underrepresented in mammalian genomes and in many viruses infecting them. Previous studies have shown that increasing CpG frequencies in various RNA viruses and retroviruses reduces replicative fitness but the effects of this compositional modification are unknown in DNA viruses. To investigate this, the effects of increasing CpG frequencies were determined in the parvovirus minute virus of mice (MVM) at various stages of the viral life cycle and a system was set up for the murine polyoma virus (MPyV).

Part of the VP gene (24.4 % of the viral genome) of MVM was replaced by synonymously coded inserts with a CpG-high sequence and several other compositional mutations. Infection time courses and TCID50s were performed to test replicative fitness. Genome copy numbers and viral RNA levels were measured via RT-qPCR. Immunofluorescence, western blots and in vitro translation were used to visualise and measure protein production and distribution. Knockout cell lines were used to find antiviral mechanisms involved.

Compared to the wild type and scrambled virus, CpG-high viruses showed impaired growth curves *in vitro*. Immunofluorescence assays showed that the CpG-high viruses could establish infection and protein production in fewer cells. RT-qPCR revealed that viral mRNA levels were reduced in the CpG-high mutant, a phenotype that partially reverted in ZC3HAV1 (ZAP) knockout cells.

Increasing CpGs in MVM led to reduced mRNA levels, which impacted all aspects of the viral life cycle, leading to fewer productively infected cells and reduced viral progeny numbers though there might be further restriction later in the replication cycle. This is thought to be due at least in part to selective recognition of CpGs in RNA by ZAP.

Acknowledgements

I would like to thank everyone who has made this project possible. I would like to thank my supervisors Peter Simmonds and Paul Klenerman for welcoming me into their groups, their guidance and support with my project. I would also like to thank Jan Rehwinkel for their collaboration, discussions and feedback

Furthermore, I would like to thank David Pintel, Peter Tattersall, Hana Španielová and Jitka Forstová for sharing their knowledge and with providing plasmids, cell lines and antibodies, as well as my colleagues Niluka Goonawardane, Isabelle Dietrich, Jelke Fros, Andrea Magri, Kinda Alshaikhahmed and Dung Van Nguyen for their support and discussion in all work areas.

I am grateful to have this opportunity thanks to the funding received from the Nuffield Department of Medicine as well as the Wellcome Trust funding the overarching project.

Pre-face

A. Table of Content

Abstract ii

Acknowledgements iii

Pre-face iv

A. Table of Content iv

B. List of Figures viii

C. List of Tables ix

D. List of Abbreviations x

E. Statement of originality and contributions of others to this work xiii

Statement of originality xiii

Contributions of others xiii

1. Introduction 1

1.1 Repression of CpGs and TpAs/UpAs in vertebrate hosts and their viruses 2

1.2 Cellular recognition of viral nucleic acids 3

Toll-like receptors 3

RIG-I-like receptors 5

NOD-like receptors, AIM2 and inflammasomes 6

PKR 9

The cGAS-STING pathway 10

Other nucleic acid sensors 12

1.3 Cellular recognition of CpGs and TpAs/UpAs 14

ZAP 15

RNaseL 16

1.4 Consequences of viral infection and PRR activation 18

1.5 Small DNA viruses 19

1.6 Minute virus of mice (MVM) 20

Structure 21

Genome 21

Life cycle 22

1.7 Murine Polyoma Virus (MPyV) 25

Structure and genome 25
Life cycle 26
1.8 Project aims 28
2 Model selection and mutant virus generation 29
2.1 Dinucleotide repression in different DNA virus families 31
Murine mRNA 31
Anelloviridae 34
Circoviridae 35
Papillomaviridae 36
Parvoviridae 37
Polyomaviridae 38
Similarities and differences between families 39
2.2 Choosing a suitable small DNA model 40
Parvoviridae 41
Polyomaviridae 43
Other small DNA viruses 45
2.3 MVM mutant selection and generation 46
Region selection 46
Mutant selection – 1st generation 48
Mutant selection – 2nd generation 49
Construction of an R3 expression plasmid 51
2.4 MVM virus recovery and quantification 53
Virus recovery 53
Virus quantification - Infectivity 54
Virus quantification – Viral genome copies 58
DNA to infectivity ratios of the MVM mutants 58
2.5 MPyV mutant selection and generation 62
Region selection 62
Cloning strategy 64
2.6 MPyV virus recovery and quantification 65
Virus recovery 65
Virus quantification 65
2.7 Using MPyV as a second model virus. 67
3 Characterisation of MVM mutant phenotype 68
3.1 Replicative fitness of MVM mutants 69
1st generation mutants 69
2nd generation mutants 72

3.2 Viral protein expression 74
Visualisation and quantification of viral infection using Immunofluorescence 74
Western blot analysis 79
In vitro translation efficiency 80
3.3 Viral RNA expression 83
3.4 Selection pressure against CpG dinucleotides 85
Genetic stability of the MVM virus genome 85
Serial passage and miseq sequencing 86
4 Mechanisms of CpG-mediated attenuation of MVM replication 91
4.1 Screening for potential recognition molecules in murine knock out cells 92
Susceptibility to WT MVM 92
Attenuation of CpG-high MVM mutants 94
4.2 Involvement of TLR9 97
TLR9 mediated signalling in A9 cells 97
Effects of TLR9 inhibition on MVM infection 98
4.3 Involvement of ZAP and RNaseL 101
Susceptibility of A549 cells to MVM 101
Limitations of the A549 ZAP and RNAseL knockout system 103
RNA expression levels in ZAP and RNaseL knockout cells 104
Visualisation and quantification of viral infection using Immunofluorescence 107
5 Discussion 117
5.1 CpG and TpA/UpA suppression in small DNA viruses 118
5.2 Recognition of unmethylated CpGs by TLR9 119
5.3 Involvement of cytoplasmic RNA and DNA recognition receptors 121
5.4 Restriction of CpG-rich RNA by ZAP and RNaseL 122
ZAP 122
RNaseL 123
5.5 Attenuation of TpA/UpA-high viruses 124
5.6 Dose dependency of CpG attenuation 125
5.7 Effects of codon (de)optimisation and G+C content 127
5.8 Future work 130
MVM 130
MPyV 131
5.9 Conclusion 132

6 Methods 133
6.1 Mutant design and generation 134
In silico design of codon-optimised, GpC-, CpG- and UpA-modified viruses 134
Cloning and plasmid propagation 134
6.2 Cell culture and infectious virus work 137
Cell culture, cell lines and reagents 137
Virus recovery 137
Virus Titration 138
Virus infections 138
Serial passage 139
6.3 Molecular techniques 140
In vitro transcription and translation 140
RNA expression experiments 140
TLR9 stimulation 140
DNA and RNA extractions 141
PCR and Quantitative real-time PCR 141
Immunofluorescence 142
Immunoblotting 143
Miseq sequencing 144
7 Appendices 145
7.1 References 146
7.2 DNA sequences 187
MVM Inserts 187
pCAG_MVMp expression plasmid 194
In vitro translation plasmid 199
MPyV Inserts 202
7.3 Use of a small DNA virus model to investigate mechanisms of CpG dinucleotide-induced attenuation of virus replication - Loew et al., 2020, J Gen Virol 204

B. List of Figures

Figure 1.1 Genome organisation and transcriptional map of MVM 22
Figure 1.2 Genome organisation and transcriptional map of MPyV 26
Figure 2.1 CpG and TpA distribution in murine mRNA 33
Figure 2.2 CpG and TpA distribution in Anelloviridae 34
Figure 2.3 CpG and TpA distribution in Circoviridae 35
Figure 2.4 CpG and TpA distribution in Papillomaviridae 36
Figure 2.5 CpG and TpA distribution in Parvoviridae 37
Figure 2.6 CpG and TpA distribution in Polyomaviridae 38
Figure 2.7 CpG and TpA distribution in minute virus of mice 42
Figure 2.8 CpG and TpA distribution in murine polyomavirus 44
Figure 2.9 MVM genomic structure and the capsid gene region selected for mutagenesis 47
Figure 2.10 MVM genomic structure and R3 expression plasmid 52
Figure 2.11 TCID50 scoring in NB324K and A9 cells 56
Figure 2.12 DNA to infectivity ratios of MVM in A9 and NB324K cells 59
Figure 2.13 MPyV genomic structure and the regions selected for mutagenesis 63
Figure 3.1 MVM mutant growth kinetics 71
Figure 3.2 Immunofluorescence staining of viral NS and capsid proteins. 76
Figure 3.3 Quantification of immunofluorescence staining of viral NS and capsid proteins 78
Figure 3.4 MVM protein expression post infection 79
Figure 3.5 MVM VP in vitro translation 81
Figure 3.6 MVM mRNA levels post transfection 84
Figure 3.7 Read coverage of MVM next generation sequencing 87
Figure 4.1 MVM wild type replication in different cell lines 93
Figure 4.2 MVM mutant replication in different cell lines 96
Figure 4.3 IL-1β induction after TLR9 activation 98
Figure 4.4 Effects of TLR9 inhibition on MVM mutant phenotype 99
Figure 4.5 MVM mutant replication in A549 knockout cells 103
Figure 4.6 MVM R1 mRNA levels in parental A549 and ZAP and RNaseL knockout cells 106
Figure 4.7 Immunofluorescence staining of viral NS and capsid proteins in A549 cell lines 110
Figure 4.8 Quantitative analysis of immunofluorescence staining of viral NS and capsid proteins in A549 cell lines 112
Figure 4.9 Subcellular localisation of MVM proteins post-infection 114

C. List of Tables

Table 2.1 Dinucleotide composition of 1st generation MVM mutant regions 49
Table 2.2 Dinucleotide composition of 2nd generation MVM mutant regions 51
Table 2.3 Dinucleotide composition of MPyV mutant regions 64
Table 3.1 MVM sequence changes after serial passage 89

D. List of Abbreviations

2-5A	2′–5′ oligoadenylates
AIM2	Absent in melanoma 2
ALR	AIM2-like receptor
ASC	Apoptosis-associated speck-like protein containing a CARD
ATP	Adenosine triphosphate
cGAMP	Cyclic GMP-AMP
cGAS	Cyclic GMP-AMP synthase
CIITA	Class II, major histocompatibility complex, transactivator
CPE	Cytopathic effect
DAI	DNA-dependent activator of IFN-regulatory factors
DAMP	Damage-associated molecular pattern
DDX41	DEAD-box helicase 41
DHX	DEAH-box helicase
DNA-PK	DNA-dependent protein kinase
dsDNA/RNA	Double stranded RNA
E7	Echovirus 7
eIF	Eukaryotic translation initiation factor
ER	Endoplasmic reticulum
FADD	Fas-associated protein with death domain
HEK-hTLR9	Human kidney cell line HEK293 stably expressing human TLR9
HIV	Human immunodeficiency virus
IFI16	Interferon gamma inducible protein 16
IFN	Interferon

IFNAR	Interferon α/β receptor
IKK	IκB kinase
IL	Interleukin
IPAF	ICE-protease activating factor
IRF	Interferon regulatory factor
KHNYN	KH and NYN domain containing protein
k/o	Knockout
LGP2	Laboratory of genetics and physiology 2
MAVS	Mitochondrial antiviral signaling proteins
MDA5	Melanoma differentiation-associated protein 5
MEFs	Mouse embryonic fibroblasts
MHC	Major histocompatibility complex
MPyV	Murine polyoma virus
MVM	Minute virus of mice
MyD88	Myeloid differentiation primary response gene 88
NAIP	Neuronal apoptosis inhibitory protein
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NLR	NOD-like receptors
NLRC	NLR family CARD domain-containing protein
NLRP	NACHT, LRR and PYD domains-containing protein
NLS	Nuclear localisation sequences
NOD	Nucleotide-binding oligomerization domain
OAS3	Oligoanylate synthetase 3
ODN	Oligodeoxynucleotide
PACT	Protein activator of PKR
PAMP	Pathogen-associated molecular pattern

PARP13	Poly (ADP-ribose) polymerase-13
PBMC	Peripheral blood mononuclear cell
PKR	Protein kinase R
PPV	Porcine parvovirus
PRR	Pattern recognition receptor
p*x*	Passage *x*
PYHIN	Pyrin and HIN domain-containing
RIG-I	Retinoic acid-inducible gene I
shRNA	short hairpin RNA
siRNA	small interfering RNA
ssDNA/RNA	Single stranded DNA/RNA
STING	Stimulator of interferon genes
SV40	Simian virus 40
TBK1	TANK-binding kinase 1
$TCID_{50}$	50% tissue culture infective dose
TLR	Toll-like receptor
TNFα	Tumor necrosis factor α
TRIF	TIR-domain-containing adaptor inducing interferon-β
TRIM25	Tripartite motif-containing protein 25
ZAP	Zinc-finger antiviral protein
ZBP1	Z-DNA Binding Protein 1

E. Statement of originality and contributions of others to this work

Statement of originality

This thesis is submitted to the University of Oxford in partial fulfilment of the requirements of the degree of Doctor of Philosophy. This thesis contains no material that has been previously submitted for a degree or diploma at this University or any other institution. This research presents my own original work this research degree. Contributions of others towards this project are described below as well as in the relevant sections within the main text of the thesis. I confirm that this thesis contains no more than 50000 words as per University of Oxford guidelines.

The research in chapters 2, 3 and 4 has been published in *J Gen Virol* as Loew *et al.* (1). The methods described in Loew *et al.* (1) were copied from this thesis with minor modifications for publication. Chapter 1.1 was copied into the manuscript of the paper and used as a basis for the introduction in Loew *et al.* (1), leaving some overlap between both pieces of writing. The same applies to the results chapters 2-4, as an earlier draft these were used as the basis to write up the results found in Loew *et al.* (1). The discussion of this thesis and Loew *et al.* (1) were written up completely separate.

Contributions of others

Chapter 2, section 1:

Peter Simmonds ran the SSE analysis for the mouse mRNA and gave me the excel file containing the resulting data as my computer kept crashing under the processing burden when I tried it myself.

Chapter 3, section 2:

Niluka Goonawardane did the electronic image processing and quantification of the immunofluorescence images shown in this section.

Chapter 3, section 4:

Dung Nguyen processed the raw data obtained from the NGS run. Peter Simmonds made excel templates to easily analyse the data.

Chapter 4, section 3:

Niluka Goonawardane did the electronic image processing and quantification of the immunofluorescence images shown in this section.

1. Introduction

. … .

It's a dangerous business, Frodo, going out your door. You step onto the road, and if you don't keep your feet, there's no knowing where you might be swept off to.

- J.R.R. Tolkien, The Lord of the Rings

. … .

1.1 Repression of CpGs and TpAs/UpAs in vertebrate hosts and their viruses 2

1.2 Cellular recognition of viral nucleic acids 3

Toll-like receptors 3
RIG-I-like receptors 5
NOD-like receptors, AIM2 and inflammasomes 6
PKR 9
The cGAS-STING pathway 10
Other nucleic acid sensors 12

1.3 Cellular recognition of CpGs and TpAs/UpAs. 14

ZAP 15
RNaseL 16

1.4 Consequences of viral infection and PRR activation 18

1.5 Small DNA viruses 19

1.6 Minute virus of mice (MVM) 20

Structure 21
Genome 21
Life cycle 22

1.7 Murine Polyoma Virus (MPyV) 25

Structure and genome 25
Life cycle 26

1.8 Project aims 28

1.1 Repression of CpGs and TpAs/UpAs in vertebrate hosts and their viruses

Early investigations into the DNA composition of various organisms revealed differences in both the relative abundance of each nucleotide (A, C, G and T) and certain dinucleotide combinations (2-4). This was particularly obvious for CpG (the p indicating the phosphate bond between the two nucleotides) which were markedly reduced compared to their expected occurrences and their respective counterparts, GpC and ApT. This has primarily been attributed to DNA methylation processes, as methylated CpGs are more prone to C to T mutations than their non-methylated counterparts (5,6). In accordance with this, CpG suppression can be seen in organisms that highly methylate their genomes (e.g. mammals and other vertebrates), but is absent from those that show very low methylation levels (e.g. insects and other invertebrates) (5,7-10).

When virus genomes were sequenced, the dinucleotide bias observed in vertebrate genomes was reflected in various viruses infecting them. While almost all single stranded RNA (ssRNA) viruses, retroviruses and small DNA viruses (<20kb genome length) showed some degree of CpG suppression, double stranded RNA (dsRNA) viruses and large DNA viruses mostly did not (11-13). This is in contrast to the universal suppression of TpAs/UpAs which has been found to some extent in the majority of animals and viruses. This is thought to be mainly due to the antiviral, sequence specific RNase RNaseL which preferentially degrades UpA and UpU rich RNAs, thus affecting mRNA stability (14,15).

1.2 Cellular recognition of viral nucleic acids

Cells have evolved a sophisticated network of evolutionary conserved proteins sensing viral infections and activating signalling cascades to elicit antiviral responses. These pattern recognition receptors (PRRs) interact with various pathogen-associated molecular patterns (PAMPs), targeting structural differences between host and pathogens.

While PRRs can recognise a large range of different PAMPs associated with various pathogens, for this study, the focus has been on the recognition of viral nucleic acids, either in the form of viral RNA or DNA genomes or RNA intermediates, particularly dsRNA.

Toll-like receptors

Toll-like receptors (TLRs) are trans-membrane proteins mostly found in innate and adaptive immune cells (for an in-depth review see (16,17)). They signal through the adaptor molecules myeloid differentiation primary response 88 (MyD88) and/or TIR-domain-containing adaptor inducing interferon-β (TRIF). This leads to the downstream activation of pro-inflammatory cytokines through nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) and the production of type 1 interferons (IFNs) through interferon regulatory factor 3 (IRF3) and 7 (IRF7). An in depth review of the MyD88 and TRIF signalling cascades has been done by Kawasaki and Kawai (18).

Ten TLRs have been identified in humans (TLR1-TLR10), while 12 TLRs have been identified in mice (TLR1-9 and TLR11-13). TLR10 is a pseudogene in mice, but present in other mammals (19). In humans, TLR11 is a pseudogene, while TLR12 and TLR13 are completely absent.

TLR3, TLR7 and TLR8 are located in endosomes and lysosomes recognising RNA, while TLR9 recognises DNA in the same compartments. TLR7, TLR8 and TLR9 signalling is MyD88 dependent while TLR3 signals exclusively through TRIF.

TLR9 recognises short, ssDNA containing non-methylated CpGs and forms dimers upon activation (20-24). DNAse II digests dsDNA in lysosomes, generating ssDNA ligands for TLR9 (25). The distinction between self and non-self DNA is achieved by both CpG methylation as well as intracellular localisation (26). It has been shown that different types of oligodeoxynucleotides (ODNs) have different effects on TLR9 activation (25,27). Stimulation with CpG-A ODNs leads to type I interferon production in plasmacytoid dendritic cells (27). Stimulation with CpG-B ODNs induce the production of pro-inflammatory cytokines in dendritic cells and macrophages and activate B-cells (27). CpG-A requires DNAse II cleaving to activate TLR9, while CpG-B does not (25). TLR9 recognises both bacterial and viral DNA, playing a crucial role in controlling DNA viral infections (28-33).

TLR3 recognises dsRNA and stem loop structures in ssRNA. It is active against both RNA and DNA viruses (34,35). TLR7 and TLR8 detect single stranded RNA, with both being able to detect most long ssRNAs but showing composition specific differences in short RNA oligonucleotides (36-38). While recognising similar PAMPs, they lead to activation of different innate immune cells and induce different cytokines (39).

TLR1, TLR2 and TLR6 are found on the plasma membrane. They recognise bacterial lipoproteins and peptidoglycans as heterodimers (TLR1/TLR2 or TLR6/TLR2) and signalling through MyD88 (40). TLR10 form dimers with TLR2 as well, but has been found to function as a negative regulator of other TLRs and anti-inflammatory PRRs (19,41).

TLR4 and TLR5 are also located on the plasma membrane, recognising lipopolysaccharide (LPS) and flagellin, respectively (42). Beyond their detection of bacterial components, TLR2 and TLR4 has been shown to recognise viral proteins, especially glycoproteins (43-50).
TLR11 and TLR12 form heterodimers detecting profilin in parasite infections (51-53). TLR13 binds bacterial 23S RNA sequence specific (54,55), but has also been found to recognise respiratory syncytial virus (56).

RIG-I-like receptors

Unlike TLRs, Retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs) can be found in all cells containing a nucleus and have recently been reviewed in depth by Brisse and Ly (57) as well as by Rehwinkel and Gack (58). RLRs consist of three proteins: RIG-I, melanoma differentiation-associated gene 5 (MDA5), and laboratory of genetics and physiology 2 (LGP2), all of which bind to viral RNAs in the cytoplasm. RIG-I and MDA5 signal through mitochondrial antiviral signaling proteins (MAVS), an adaptor molecule found on the mitochondrial membrane feeding into the NF-κB, IRF3 and IRF7 activation cascade. LGP2 lacks the signalling domains of RIG-I and MDA5 required to activate MAVS, but it is crucial for RIG-I and MDA5 mediated RNA recognition (59-61). An in-depth review on MAVS signalling and regulation was recently published by Ren *et al.* (62).

RIG-I is constitutively expressed in an inactive form, undergoing activation upon viral infection through rapid dephosphorylation while MDA5 exists in an equilibrium between its active and inactive state (63,64). RIG-I and MDA5 recognise RNAs based on structural differences between viral and host RNA. MDA5 recognises long dsRNA, using these RNA species as a scaffold to oligomerise (64-68). RIG-I recognises 5’-diphosphorylate (5’-pp) or 5’-

triphosphorylated (5’-ppp), uncapped ssRNA and short dsRNA (65,69-72). Furthermore, RNA polymerase III recognises and transcribes cytosolic, AT-rich dsDNA into 5’-triphophate RNA which can then be recognised by RIG-I (73,74).

Both RIG-I and MDA5 oligomerise into filaments after RNA binding and are ubiquitinated (64,66,68,75-77). This leads to the activation and filamentous aggregation of MAVS on the mitochondrial membrane (75,78). Recently, RIG-I has been found to recognise viral RNAs within the nucleus, activating MAVS on perinuclear mitochondria (79).

RIG-I and MDA5 are able to detect a wide range of viruses, including positive and negative ssRNA viruses, dsRNA viruses and DNA viruses (80-84). An in-depth review of RLRs in DNA virus infections has been written by Zhao and Karijolich (84).

NOD-like receptors, AIM2 and inflammasomes

Nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) are a family of PRRs, which can recognise both PAMPs and damage-associated molecular patterns (DAMPs). NLRs can be divided into three phylogenetic subfamilies (85). The NOD subfamily contains NOD1-5 and class II, major histocompatibility complex, transactivator (CIITA). The NACHT, LRR and PYD domains-containing protein (NLRP, also called NALP) subfamily consists of NLRP1-14. The ICE-protease activating factor (IPAF) subfamily contains IPAF, also known as NLR family CARD domain-containing protein 4 (NLRC4), and Neuronal apoptosis inhibitory protein (NAIP). There are 22 human NLRs, but many more in mice due to gene duplication (85).

NOD1 and NOD2 are best characterised among the NLR family. They recognise bacterial peptidoglycan intermediates and degradation products (86-88). NOD2, but not NOD1, has also been shown to activate in response to Influenza A, vesicular stomatitis virus and

respiratory syncytial virus infection (89,90). NOD1 and NOD2 signal through several pathways leading to the activation of transcription factors NF-κB, AP1 and IRF3/IRF7. This has been reviewed in detail by Trindade and Chen (91).

Other members of the NLR family activate caspase-1 and IL-1β through forming a protein complex known as the inflammasome (92). The canonical inflammasome consists of the adaptor apoptosis-associated speck-like protein containing a CARD (ASC), pro-caspase-1 and an NLR, absent in melanoma 2 (AIM2) or Pyrin. Upon binding its ligands, the receptor oligomerises, activating caspase-1 (93). Caspase-1 cleaves the interleukin (IL)-1β precursor pro-IL-1β and the IL-18 precursor pro-IL-18, which accumulate in the cell as part of NF-κB activation (93,94). This cleavage produces biologically active IL-1β and IL-18 (93,94). Caspase-1 further activates Gasdermin D, leading to pyroptosis (95,96)(reviewed by Lieberman *et al.* (97)). Beyond this well-established composition and effector function of the inflammasome, non-canonical inflammasomes and further effector functions have been discovered. Rathinam and Fitzgerald (98) as well as Broz and Dixit (99) have written in-depth reviews on inflammasomes and their effector functions.

Different NLRs, AIM2 and pyrin form inflammasomes with a range of different ligand specificities. The best-characterised inflammasome is the NLRP3 inflammasome, where NLRP3 is proposed to be activated indirectly by a range of PAMPs and DAMPs. These include viral and bacterial nucleic acids and bacterial toxins, as well as cellular stress signals such as reactive oxygen species and metabolic crystals (100-107). This has been reviewed in depth by Wang and Hauenstein (108).

AIM2 is a member of the pyrin and HIN domain-containing (PYHIN) protein family. The PYHIN family, also known as AIM2-like receptors (ALRs), is genetically diverse and encompasses 4 human genes, but 14 mouse genes (109,110). Interferon Gamma Inducible Protein 16 (IFI16), another member of the PYHIN family is discussed in the next section. AIM2 is a cytosolic dsDNA sensor that activates both caspase-1 and caspase-8 through ASC (111-116). While AIM2 recognises dsDNA sequence-independent, it requires DNA with a minimum length of 70bp and assembles faster on longer DNA fragments (117,118). Upon DNA recognition, AIM2 forms filaments along the DNA, leading to AIM2 oligomerisation and ASC polymerisation (118). AIM2 has been shown to detect various viruses, including herpes- and papillomaviruses as well as genomic DNA leaking into the cyoplasm (111,119-125). An in-depth review of the AIM2 inflammasome has been written by Lugrin and Martinon (126).

Other well-studied inflammasomes feature NLRC4 and NLRP1. Both NLRC4 and NLRP1 are involved in the recognition of bacterial infections (127-130). NLRP1 has also been shown to recognise *Toxoplasma gondii (131-133).* An in depth review of the NLRC4 inflammasome has been done by Andrade and Zamboni (134) and Mitchell *et al.* have reviewed the NLRP1 inflammasome (135). Pyrin recognises changes in the activity of the cellular GTPase RhoA as a result of microbial infection (136). The pyrin inflammasome has been reviewed in detail by Schnappauf *et al.* (137).

The interactions with inflammasomes have been studied in several viruses, including human immunodeficiency virus (HIV), Influenza virus A and Epstein-Barr virus. An in-depth review has been written by Hanna Antushevich, exploring how viruses are hindered by inflammasomes or are hijacking them, as well as the consequences of the resulting inflammatory responses (138).

Another notable NLR is NLRC5. NLRC5 is activated in reaction to RNA virus infections, but does not directly bind to viral ligands (139,140). It is thought to amplify antiviral responses by acting as a transcriptional transactivator. Specifically, NLCR5 and RIG-I interact and NLCR5 leads to the upregulation of major histocompatibility complex (MHC) class I molecules on the cell surface (141-143).

PKR

Protein kinase R (PKR) was originally identified as an interferon-inducible dsRNA PRR, but has since been found to be involved in various pathways recognising and responding to cellular stress, most notably by facilitating translational shut-down.

PKR recognises dsRNA of at least 30bp length, with longer RNAs leading to better activation (144,145). Recently, short dsRNA with single-stranded ends has been found to activate PKR, including small nucleolar and nuclear RNAs (146-149). PKR further recognises some cellular mRNA UTRs and tRNAs without post-transcriptional modifications (150-153). Differently structured RNA may be activating or inhibiting PKR, though there is currently no unified code on which RNA species will have what effect (154-156). PKR activators and inhibitors have been reviewed in-depth by Bou-Nader *et al.* (157).

Beyond PKR recognition of dsRNA, various other cellular stresses lead to its activation. These include cytokines like tumor necrosis factor alpha (TNFα) and IL-1, TLR4 signalling and cellular stressors like arsenite and hydrogen peroxide (158-160).

Upon activation, PKR phosphorylates the α subunit of eukaryotic translation initiation factor 2 (eIF2α)(144). This phosphorylation leads to the inactivation of eIF2 and halts cap-dependent

translation initiation (161). This is part of the integrated stress response and contributes to the formation of stress granules (reviewed in depth by Pakos-Zebrucka *et al.* (162) and Eiermann *et al.* (163)). PKR is also involved in MDA5/MAVS signalling (164,165) and can induce apoptosis through the Protein activator of PKR (PACT)/Fas-associated protein with death domain (FADD)/caspase-8 pathway (166-171).

PKR is involved in the recognition of and antiviral response against various RNA and DNA viruses. Liu *et al.* (172) have written an in-depth review on how eIF2α affects viral infections, while Dzananovic *et al.* (173) have reviewed viral antagonism against PKR.

The cGAS-STING pathway

Stimulator of Interferon Genes (STING) has been established as a key adaptor protein that is activated by several cytoplasmic DNA sensors as well as endoplasmic reticulum (ER) stress and membrane fusions (174,175).

STING is constitutively expressed in the cell, located as a dimer on the ER membrane (176,177). Upon activation, STING translocates to the Golgi apparatus (178,179). Here, it signals through TANK-binding kinase 1 (TBK1) and IκB kinase (IKK) to activate IRF3 and NF-κB, respectively (180-185). This leads to the transcription and expression of type I interferons and pro-inflammatory cytokines. Beyond this, STING signalling has been found to facilitate autophagy and apoptosis (174,179,186).

The most prominent of the cytoplasmic sensors signalling through STING is cyclic GMP-AMP Synthase (cGAS). It binds dsDNA or RNA:DNA hybrids independent of sequence, with longer DNA fragments leading to cGAS oligomerisation, enhancing its enzymatic activity (187-191). Upon activation, cGAS converts ATP and GTP into cyclic GMP-AMP (cGAMP), which acts as a

second messenger to activate STING (192-195). cGAMP acts both intracellularly as well as in neighbouring cells after being transported through gap junctions (196-198). Besides cGAMP, STING is able to recognise other (bacterial) cyclic dinucleotides, such as cyclic diGMP or diAMP (182,199,200). An in-depth review of the cGAS-STING pathway has been written by Motwani *et al.* (201).

cGAS and STING are crucial in antiviral immunity, restricting RNA, DNA and retroviruses viruses, but are also antagonised by some of these viruses (181,202 ,203,204). This has been reviewed in depth by Ahn and Barber (205) and Phelan *et al.* (206).

Other cytoplasmic sensors activating STING include DNA-dependent Protein Kinase (DNA-PK), DEAD-Box Helicase 41 (DDX41) and IFI16.

DNA-PK is a helicase primarily involved in the DNA damage response, specifically in non-homologues end joining after DNA double-strand breaks (reviewed in-depth by Blackford and Jackson (207)). However, DNA-PK has been found to act as a cytosolic PRR for dsDNA, signalling both through STING and independent of it (208,209). DNA-PK impairs adenoviral replication, while being hijacked by adenovirus associated viruses and HIV (209,210). Furthermore, it has been shown that several viruses have proteins to antagonize DNA-PK (210-214). An in-depth review on the interactions between DNA viruses and the DNA damage response was written by Weitzman and Fradet-Turcotte (215).

DDX41 recognises cytosolic dsDNA and retroviral reverse transcripts, signalling through STING (216-218). It has also been found to recognise cyclic diGMP and cyclic diAMP and is required for their efficient recognition by STING (217,219). An in-depth review of DDX41 has been written by Jiang *et al.* (220).

IFI16 is found both in the nucleus and in the cytoplasm (221,222). IFI16 recognises dsDNA sequence independent in both cytosol and in the nucleus and signals through STING (117)(221). It enhances cGAS mediated cGAMP activity and is involved in recruiting TBK1 to the STING complex (223,224). Within the nucleus, it recognises viral DNA and recruits ASC (225-227). The complex then translocates to the cytoplasm, where it recruits and activates caspase-1, forming an IFI16 inflammasome (225,226).

Other nucleic acid sensors

Other notable nucleic acid sensors include DNA-dependent Activator of IFN-regulatory Factors (DAI) also known as Z-DNA Binding Protein 1 (ZBP1), DEAH-Box Helicase 9 (DHX9) and 36 (DHX36).

ZBP1 is mostly found in the cytoplasm, but accumulates in stress granules and processing bodies upon environmental stress (228,229). It was originally identified as a cytosolic dsDNA sensor signalling through TBK1 to activate IRF3 (230,231). However, more recently, it was found to recognise viral RNAs, specifically Z-form RNA (232-236). Furthermore, ZBP-1 was shown to activate the NLRP3 inflammasome as well as to induce cell death by signalling through RIPK1 and RIPK3 (234,237-239). Zheng and Kanneganti have written an in-depth review on recent ZBP1 discoveries (240).

DHX9 is a multifunctional protein that can unwind both DNA and RNA in various configurations, including triplex DNA and G-quadruplexes (241,242). It is involved in various aspects surrounding transcription, translation, DNA replication and genome stability. An in-depth review about the biological functions of DHX9 has been written by Lee and Pelletier (243). Depending on cell type, DHX9 has been shown to respond to different stimuli and to be part of several signalling pathways. It has been shown to recognise CpG DNA, specifically

CpG-B ODNs, as well as dsRNA and potentially other, DNA-independent mechanisms (244-246). DHX9 directly binds to and signals through MyD88 and has been suggested to mediate TLR signalling (244,247,248). It has also been found to signal through MAVS and interact with PKR (245,249). Several studies have found both antiviral activities of DHX9 as well as viral hijacking to promote viral proliferation (248,250-257).

Like DHX9, DHX36 binds to intracellular G-quadruplex DNA and RNA. It is involved in resolving these structures, leading to RNA degradation or translation depending on context (258-263). This has been reviewed in depth by Shen and Pelletier (264) as well as Schult and Paeschke (265). It can also to act as pattern recognition receptor in different cell lines. It has been shown to recognise CpG-A ODNs and dsRNA, signalling through TRIF, as well as interacting with PKR to stimulate RIG-I signalling (244,266-268).

1.3 Cellular recognition of CpGs and TpAs/UpAs.

While DNA methylation and the avoidance of TLR9 recognition could explain the CpG suppression in DNA and retroviruses, it fails to account for the widespread suppression in RNA viruses. The biological relevance of this suppression was highlighted when Burns *et al.* (269) mutated parts of the poliovirus (a positive ssRNA virus) capsid gene to evaluate the potential of different types of mutations for viral attenuation and potential vaccine candidates. While they set out to test the effects of using non-preferred codons on viral replication, they found that viral attenuation correlated with an increase in CpGs and UpAs, a side effect of the codon de-optimisation.

This attenuation can also be seen in other viruses, including effects on viral fitness and replication rates of Echovirus 7 (E7, also a positive ssRNA virus) in vitro (270,271) and on pathogenicity of Influenza A virus (a negative strand ssRNA virus) and the subsequent immune response towards the virus in vivo (272).

In parallel, mathematical models gave insight into the reverse correlation between CpG suppression and G+C content (i.e. the percentage of C and G nucleotides present in a sequence independent of order) in methylated genomes (6). These suggested that the conversion rate of C to T was the main driver of CpG suppression in genomic DNA, while DNA that was transcribed and translated showed further CpG suppression beyond the methylation pressure as well as a suppression of TpAs/UpAs (6,273,274).

However, identifying the cellular pathways involved in viral CpG recognition has not been as straightforward. Atkinson *et al.* (270) have found that the attenuating effect of CpG and UpA elevation in E7 was not IFN dependent nor were the dinucleotides recognised by the

cytoplasmic RNA sensors PKR, RIG-I or MDA5, but rather through an unknown pathway sensitive to the kinase inhibitors C16 and roscovitine.

During the course of this research project, zinc finger antiviral protein (ZAP) has been shown to specifically bind to CpG in HIV RNA sequences and attenuate the replication of CpG-high human immunodeficiency virus 1 (HIV-1) mutants (275). Similar studies with E7 showed that ZAP mediated the attenuation of both CpG-high and UpA-high mutants, but also found an involvement of oligoadenylate synthetase 3 (OAS3) and RNaseL (276).

ZAP

ZAP, also known as Poly (ADP-ribose) Polymerase-13 (PARP13) or ZC3HAV1 is an RNA binding protein involved in RNA degradation and translation repression (277-280). Both the PARP family (281) and ZAP (282) have been reviewed in depth in recent years.

ZAP is expressed as a long and short isoform (283-285). Both the long and short isoform been shown to be involved in antiviral responses (279,284,286-288). Recently, two further isoforms have been discovered (289).

ZAP signalling has been shown to be dependent on the presence of the proteins Tripartite motif-containing protein 25 (TRIM25) and KH and NYN domain containing protein (KHNYN) (290). While ZAP is constitutively expressed in cells, it gets upregulated upon infection through the IRF3 signalling cascade, but not through NF-κB (290-292).

ZAP has been shown to target specific cellular and viral mRNAs for degradation through the recruitment of exosome complexes leading to 3’-5’ decay of RNA (277,279). Furthermore, ZAP has been found to repress the translation of specific mRNAs by inhibiting translation-initiation

factor eIF4A in conditions of cellular stress and globally through initiating stress granule formation (280,293-295). It has also been shown to be able to amplify RIG-I signalling (284).

ZAP has been known to restrict a number of viruses from different viral families. These include the retroviruses murine leukaemia virus and HIV (279,294,296), RNA viruses such as Sindbis virus and influenza virus (284,286,297-299) and DNA viruses including hepatitis B virus, while promoting latency in herpesviruses (300,301). However, recently it has been shown that ZAP specifically binds to HIV-1 and E7 RNA with artificially increased CpG frequencies as well as a region naturally high in CpGs in Japanese encephalitis virus (275,276,298). In the E7 model, ZAP was shown to be involved in the UpA mediated restriction of the virus (276).

RNaseL

RNaseL is an endonuclease targeting viral and cellular single-stranded RNAs for degradation during infection and other cellular stresses. RNaseL is constitutively expressed in a latent form in the cell. It requires activation by 2'–5' oligoadenylates (2–5A), a second messenger produced by the OAS protein family upon detection of cytosolic double-stranded RNA (dsRNA) (302-306). The OAS/RNaseL pathway is one of the two main IFN effector pathways induced by dsRNA, alongside PKR activation. OAS3 is crucial for virus-induced RNaseL activation in humans (307,308) and an in depth review on the OAS family has recently been published by Schwartz and Conn (309).

RNaseL exhibits a range of antiviral functions, targeting both viral and cellular RNAs by cleaving RNA containing UpA or UpU motives (303-306,310). While it is thought to be most active in the cytoplasm, RNaseL has been associated with other cellular compartments,

including the mitochondria and the nucleus (311-314). It has been proposed to discriminate between cellular and viral RNAs by acting locally in areas with high dsRNA abundance (315). RNaseL can target both viral genomes (RNA viruses) and viral mRNAs (RNA and DNA viruses) (315-324). It can also cleave a wide variety of cellular RNAs, though a number of antiviral RNA are able to circumvent this down-regulation (325). It has a major down regulatory effect on translation by targeting a broad spectrum of cellular mRNAs interacting with translation termination release factor eRF3, as well as targeting ribosomal RNA and tRNA (317,326-332). RNaseL and 2-5A can be found in mitochondria where the activated RNaseL cleaves mitochondrial mRNA and 28S RNA, priming the cell for caspase 3 dependent apoptosis (311,316,328,333-338). As part of its nuclease activity, RNaseL produces short RNA species, which are often double-stranded and can be detected by RIG-I and MDA5 (339).

RNaseL has been shown to play an important role in restricting viral infections (324,340,341). Many viruses including Influenza A, Zika virus, HIV and polyomavirus simian virus 40 (SV40) antagonise the OAS/RNaseL pathway to overcome this restriction, either by inhibiting the pathway upstream or downstream of RNaseL or by producing 2-5A analogues which competitively inhibit RNaseL activation (318,342-344). In-depth review of viral inhibiton of the OAS/RNaseL pathway has been done by Drappier and Michiels (345).
OAS3 and RNaseL have recently been found to restrict viral replication in an UpA-high mutant of E7, as well as in CpG-high mutants, a novel specificity proposed for RNaseL (276).

1.4 Consequences of viral infection and PRR activation

Once activated, PRRs signal through both converging and cross-activating protein cascades to activate different transcription factors, including NFκB, AP-1, IRF3 and IRF7. These induce the production of pro-inflammatory cytokines, interferons and potentially apoptosis within the infected cell.

NFκB and AP-1 activation leads to the expression of a range of pro-inflammatory cytokines, including IL-1, IL-12, IL-18 and TNF-α. An in-depth review of these cytokines, as well as other members of the interleukin family had been written by Akdis *et al.* (346).

IRF3 and IRF7 activation leads to the production of IFNα and IFNβ. These IFNs will be secreted and bind to type I Interferon receptors (IFNAR) on the secreting cell as well as surrounding cells. Binding IFNα/β activates IFNAR which signals through the JAK/STAT pathway (an in-depth review on JAK/STAT signalling has been done by Villarino *et al.* (347)). This leads to the expression of interferon-stimulated genes and the induction an antiviral state within infected and neighbouring cells (ISGs). A large number of ISGs with various antiviral have been identified (reviewed in-depth by Schneider *et al.* (348) and Schoggins (349)). Their combined effects aim to eliminate viruses from infected cells (with or without eliminating the infected cells) as well as to prime the surrounding cells against incoming viruses.

1.5 Small DNA viruses

There are five families of small DNA viruses that infect vertebrate hosts: *Anelloviridae*, *Circoviridae*, *Papillomaviridae*, *Parvoviridae* and *Polyomaviridae*.

They all share non-enveloped, icosahedral virions between 15 and 55 nm in size. Their genomes are all less than 10kb long, but are either comprised of linear ssDNA (parvoviruses), circular ssDNA (anello- and circoviruses) or circular dsDNA (papilloma- and polyomaviridae). They encode between 1 and 7 genes with either unidirectional or bidirectional open reading frames.

They next sections explore in more detail the parvovirus Minute Virus of Mice (MVM) and murine Polyomavirus (MPyV). These two viruses are model viruses for their respective virus families and are the viruses used in this study.

In depth reviews of the other small DNA virus families have been written within Field's Virology (350) for circo- and papillomaviruses, while different aspects of anelloviruses have been reviewed by *Webb et al.* (351) and Kaczorowska and van der Hoek (352).

1.6 Minute virus of mice (MVM)

MVM was discovered in 1966 (353) as a contaminant in an adenovirus stock and was subsequently found to be a major contaminant in rat and mouse colonies around the world (354,355). It is spread readily via direct contact, but it is not airborne (355). While it is mostly apathogenic in adult mice, it can cause disease and death in foetal and newborn mice or hamsters (356-362). Tissue tropism and host range are determined by a few amino acid changes within the capsid proteins affecting host cell receptor binding as well as effective activation of the early promoter and NS2 expression (363,364).

Taxonomy

Parvoviridae are made of two subfamilies, *Densovirinae* infecting invertebrates and *Parvovirinae* infecting vertebrates. *Parvovirinae* encompass eight genera, *Amdoparvoviruses*, Aveparvoviruses, *Bocaparvoviruses*, *Copiparvoviruses*, *Dependoparvoviruses*, *Erythoparvoviruses, Protoparvoviruses and Tetraparvoviruses (365)*. *Dependoviruses* can only complete their viral life cycle in cells co-infected with a helper virus, with Adeno- or Herpesviruses being common helper viruses (reviewed in-depth by Meier *et al.* (366)). All other *Parvovrinae* are considered autonomous viruses. Viruses are grouped into these genera according to their genomic structure and biological properties while species level distinctions are based on antigenicity, DNA sequence relatedness and the hosts they infect. MVM, along with H-1 parvovirus, is part of the *Parvovirus* genus.

Structure

The MVM virion, like other parvoviruses, is a non-enveloped icosahedron comprised of 60 subunits of the capsid proteins with a diameter of 26 nm (367).

The capsid is made of three different capsid proteins, VP1-VP3 (368). VP1 and VP2 are derived from alternatively spliced RNA species. VP3 is derived from proteolytic cleavage of VP2 once the virion is assembled and contains DNA (369). The virion surface contains various features, including a pore at the five-fold symmetry of the capsid through which DNA is packaged into the virion (367,370). A single NS1 protein remains covalently attached to the 5'-end of the genome as the DNA gets packaged into the progeny virus (371,372). This NS1 molecule attached to the genome remains outside the virion, along with 20-30 bases of the DNA itself (371,372).

Genome

The genome of MVM is a single-stranded, negative-sense, linear DNA genome of approximately 5.1kb length (373,374). It has two genes, an NS gene encoding two non-structural proteins, NS1 and NS2, as well as a capsid gene encoding the capsid proteins, VP1 and VP2 (Fig. 1.1)(374-376). These genes are flanked by terminal palindromic hairpins, which are essential for genome replication (377,378).

The MVM genome contains two promoters, the early promoter p4 and the late promoter p38 (379). Promoter p4 is cell driven and gives rise to viral RNAs R1 and R2, the mRNAs required for translating NS1 and NS2, respectively (380). The p38 promoter is trans-activated by NS1 recruiting the cellular transcription machinery, leading to the transcription of R3 and the consequent translation of VP1 and VP2 from this RNA (381). MVM has a single polyadenylation site downstream of the capsid gene (382).

Viral RNAs undergo mRNA processing and splicing through the host cell machinery. All viral RNAs contain a small intron (map position 44-46)(383). This intron has two splice donor and acceptor sites, leading to three splice variants for each viral RNA, for a total of nine RNA species (383-385). For the R3 mRNA, if the upstream donor site is used for splicing, VP1 will be translated from the transcript. If the downstream donor is used, VP2 will be produced (383). To generate R2 mRNA, a large intron (map unit 10-39) is spliced out of the R1 RNA (386,387). The ratio of NS1 to NS2 produced varies across tissues and affects tissue tropism (388-390). Both NS1 and NS2 are essential for viral replication in mouse cell lines and animals (381,389,391).

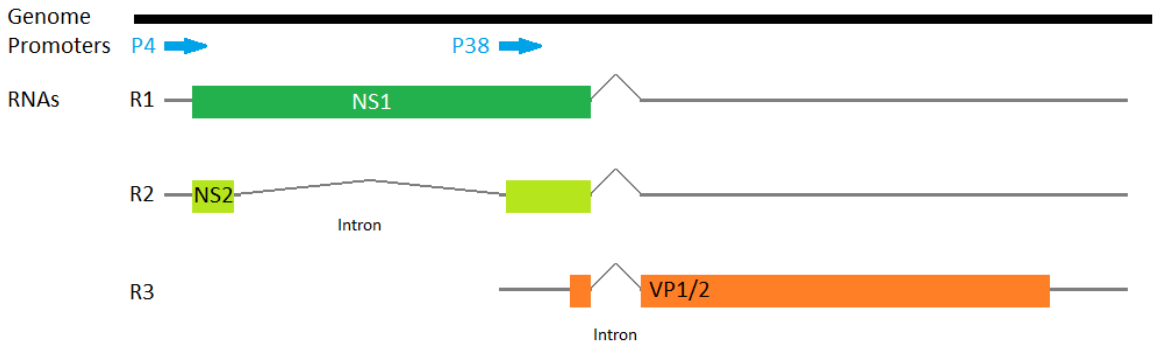


Figure 1.1 Genome organisation and transcriptional map of MVM
Diagram of the MVM genomic structure. Promoters (blue), transcripts (R1-R3) with ORFs (green and orange coloured boxes). Based on Berns & Parrish, Fig 57.10 (392).

Life cycle

Infectious particles bind to sialic acids on the cell surface of susceptible cells and are taken up into the cells by receptor-mediated endocytosis (393,394). Virions are taken by clathrin-mediated endocytosis leading them to endosomes (395). The virions escape or are released from the endosome in a perinuclear location, though it is currently unclear how exactly this occurs. The virions are actively transported to the nucleus along the cytoskeleton (reviewed

by Mäntylä *et al.* (396)). Nuclear entry is facilitated through nuclear localisation sequences (NLS) present in both the shared part of VP1 and VP2 as well as in the unique region of VP1 (397). It is currently unknown how the genome is released from the capsid.

Within the nucleus, the genome is replicated in discreet parvovirus-associated replication bodies (398). MVM uses the cellular DNA replication machinery once the infected cells to go enters S-phase (380,398,399). While MVM can halt the cell cycle once replication has been initiated, it is unable to induce cell cycle progression to S phase (400-405). Due to this restriction, MVM mainly infects dividing cells, such as foetal tissues, the gut lining and cancer cells (357,406,407).

Genome replication occurs via a rolling hairpin replication where MVM uses its own genome as a template to produce multimeric intermediates and subsequently cleaves these to obtain monomeric genomes (408,409)(reviewed in detail by Berns & Parrish (392)).

In short, the palindromic sequence at the (left-hand) 3'-end of the genome is used as a primer to synthesise the complementary strand of DNA. The newly synthesised strand is covalently joined to (right hand) 5'-end of the viral genome, forming a covalently closed double-stranded replication intermediate. This duplex is nicked by NS1 18 bases upstream of the original 5'-end of the genome, covalently binding the NS1 protein to the DNA in the process (371,372,409). This nick allows the cellular DNA repair machinery to elongate the newly synthesised DNA strand along the original right-hand palindromic sequence. The NS1 protein attached at the 5'end of the original strand acts as a natural stop to this step in DNA elongation. At this point, the DNA molecule undergoes a conformational change, forming back into the original palindromic configurations. This leads to a template change as the newly synthesised strand is extended back along itself and along the original virus genome until it is

stopped again by NS1 at the end of the template. This leaves the DNA in a double-stranded replicative form where the original 3’ palindrome is present in a double-stranded form in the middle of this DNA molecule. From this intermediate, the DNA can be further elongated into a tetrameric form or cleaved into monomeric forms (371).

To resolve multimeric replicative forms, NS1 cleaves a site within the double-stranded 3’ palindrome, attaching itself to the DNA in the process (410,411). The DNA then goes through several further conformational changes to resolve the replicative forms and generate progeny genomes containing the palindromes in the right orientation, rather than as flipped intermediates occurring during DNA elongation (409,411).

NS proteins are expressed early in infection and phosphorylated after translation (412,413). NS1 dimerises before translocating into the nucleus to facilitate replication (412). It also trans-activates the late promoter, leading to the transcription of the late R3 mRNA and the translation of the capsid proteins (381).

Capsid proteins are transported into the nucleus via NLS present in both VP1 and VP2 (397). In the nucleus, the capsids assemble into empty virions, before DNA is packaged into them (370,373,414).. While V2 is sufficient to assemble DNA-containing virions, VP1 is required for the infection of new cells (363,415). Viral release occurs via cell lysis and active egress (416,417).

1.7 Murine Polyoma Virus (MPyV)

MPyV was discovered in 1953 as a contaminant in a mouse leukemia virus stock causing various tumours alongside the expected leukemia in mice (418). Natural transmission occurs preferentially via intranasal entry and shedding occurs through saliva, urine and faeces (419,420). The infection is persistent and while there are some strains that can cause cell transformation and tumour growth, this is usually only seen in newborn, injured or immunocomprised mice (421-424), which has been reviewed in depth by Swanson *et al*. (420).

Structure and genome

MPyV has a double-stranded, circular DNA genome of 5.3 kb length that is packaged into an icosahedral, non-enveloped capsid of 45 nm diameter.

The virion is composed of the three viral capsid proteins VP1, VP2 and VP3. The virion forms a T=7 symmetrical icosahedron (425). It is made of 72 capsomeres with each containing VP1 molecules and one VP2 or VP3 molecule, which is not exposed on the surface (425). The DNA inside is packaged while wrapped onto cellular histones H2A, H2B, H3 and H4, forming nucleosome structures similar to cellular chromatin (426).

The genome is organised into a regulatory region called non-coding control region (NCCR), as well as the early region encoding the non-structural genes and the late region encoding the capsid genes (Fig. 1.2)(427). The NCCR contains the origin of replication and the promoter sequences for the two transcription units, which are read in opposite directions (428). Each coding region gives rise to several protein species, which are generated through alternative splicing (429,430).

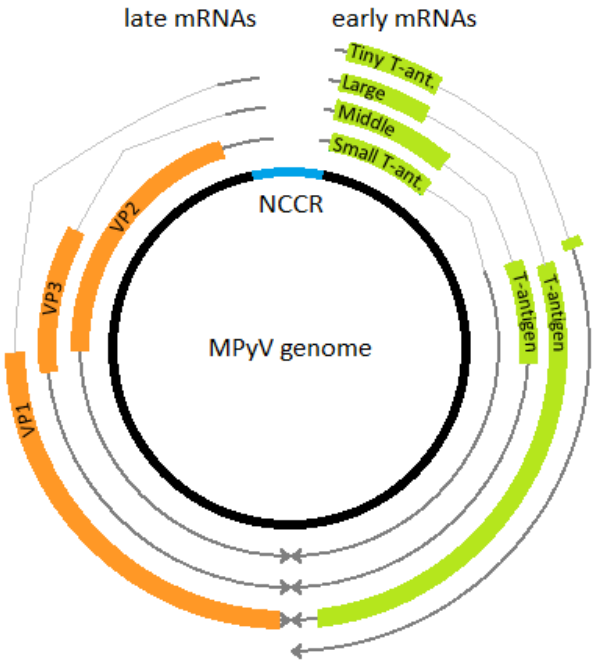


Figure 1.2 Genome organisation and transcriptional map of MPyV
Diagram of the MPyV genomic structure. Non-coding control region (NCCR, blue), transcripts (grey) with ORFs (green and orange coloured boxes). Based on Fig.1 by Carmichael, 2016 (427).

Life cycle

Infectious particles bind to the ganglioside GT1b or GD1a and enter the cells via endocytosis (431,432). The virus is transported via the endosome to the ER where the capsid begins to change and disassemble (433-435). Hydrophobic domains within the minor capsid protein VP2 are associated with virus exit from the ER while both major and minor capsid proteins interact with importin β to facilitate efficient translocation of the viral genome into the nucleus (436,437).

Transcription, translation and replication mechanics are reviewed in more detail by DeCaprio *et al*. (438). Briefly, once inside the nucleus, the early mRNA is transcribed by RNA polymerase II, processed through the cellular splicing machinery, and eventually gives rise to the non-structural proteins. The genome is replicated using the Large T antigen as a helicase unwinding the viral DNA and recruiting the cellular replication machinery.

In senescent cells, the virus is able to overcome the cell cycle arrest and to initiate S phase, allowing it to replicate in a broader range of cell types. As new genomes accumulate, the transcription switches from early to late mRNA and proteins. The capsid proteins translocate into the nucleus where progeny virus is assembled. It is currently not well understood how newly made virions are released.

1.8 Project aims

Several families of small DNA viruses show a striking underrepresentation of CpG dinucleotides in their genomes and protein coding sequences. Previous research suggests that these viruses are under selection pressure to keep CpG dinucleotide occurrence below their expected frequencies.

The main hypotheses for this project were:

1. Altering CpG dinucleotide frequencies might affect viral fitness
    a. Altering TpA/UpA dinucleotide frequencies and other compositional changes might also affect viral fitness
    b. Different parts of the viral life cycle might be affected
2. Specific cellular factors might be involved in CpG dinucleotide recognition.
    a. These may be similar or different to those found in RNA viruses and retroviruses

To investigate these questions, suitable small DNA model viruses were chosen to create mutant viruses with altered CpG frequencies and other compositional changes. They were tested for their viral fitness and how the mutations were affecting viral replication and other parts of the viral life cycle.

After characterising these mutants, knock out cell lines were used to investigate possible recognition pathways involved.

2 Model selection and mutant virus generation

In order to study the effects of dinucleotide composition in small DNA viruses, CpG and TpA/UpA distribution was determined in small DNA viral families infecting vertebrates. This expands on the analysis from Karlin *et al.* (11) and complements the analysis of RNA viruses from Simmonds *et al.* (273). Next, the different virus families were searched for potential model viruses allowing genetic manipulation, as well as *in vitro* and *in vivo* work.

Two viruses, the parvovirus minute virus of mice and murine polyomavirus, were selected as suitable models for further studies. The cloning strategy for generating mutants and the subsequent virus recovery are described for both viruses.

2.1 Dinucleotide repression in different DNA virus families 31

Murine mRNA 31
Anelloviridae 34
Circoviridae 35
Papillomaviridae 36
Parvoviridae 37
Polyomaviridae 38
Similarities and differences between families 39

2.2 Choosing a suitable small DNA model 40

Parvoviridae 41
Polyomaviridae 43
Other small DNA viruses 45

2.3 MVM mutant selection and generation 46

Region selection 46
Mutant selection – 1st generation 48
Mutant selection – 2nd generation 49
Construction of an R3 expression plasmid 51

2.4 MVM virus recovery and quantification 53

Virus recovery 53
Virus quantification - Infectivity 54
Virus quantification – Viral genome copies 58
DNA to infectivity ratios of the MVM mutants 58

2.5 MPyV mutant selection and generation 62
Region selection 62
Cloning strategy 64
2.6 MPyV virus recovery and quantification 65
Virus recovery 65
Virus quantification 65
2.7 Using MPyV as a second model virus. 67

2.1 Dinucleotide repression in different DNA virus families

In 1994, Karlin *et al.* (11) analysed viral genomes for their dinucleotide composition and found that many viruses of different groups and families suppressed CpGs. There has been much progress in viral sequencing since their study was published and there are now much larger viral data sets available. Simmonds *et al.* (273) analysed RNA viruses more depth in recent years, however, no update was made for DNA viruses. Hence, as a first step to study the effects of dinucleotide composition in small DNA viruses, CpG and TpA/UpA distribution was determined in small DNA viral families infecting vertebrates.

To analyse the CpG and TpA/UpA distribution of these viruses, genomic and protein coding sequences were obtained from RefSeq, filtering for vertebrate hosts. This included sequences from five virus families: *Anelloviridae*, *Circoviridae*, *Papillomaviridae*, *Parvoviridae* and *Polyomaviridae*.

Genetic sequences for both viral genomes and genes were analysed for their G+C content as well as their CpG and TpA/UpA frequencies using the composition scan within the SSE program v1.3 (439). These were then compared to those of murine mRNA (Peter Simmonds ran the composition scan for the murine mRNA and gave me the excel file containing the resulting data as my computer kept crashing under the processing burden when I tried it myself).

Murine mRNA

Sequences shorter than 500bp were excluded from the analysis as these would often produce outliers due to their small size. Murine mRNA shows a suppression in CpG dinucleotides that appears to be partially correlated to G+C content, with lower G+C content showing higher

levels of CpG suppression (Fig.2.1A). While the majority of genes form a cluster that spans between a G+C content of 0.3 to 0.7 with a corresponding CpG observed over expected ratio of 0.0 to 1.0, several smaller clusters and outliers can be observed. At the G+C rich end of the spectrum, the main cluster tethers out into G+C rich sequences with no or minimal CpG suppression. Furthermore, various outliers and smaller clusters can be observed, for example G+C rich CpG low sequences, G+C medium CpG rich sequences and CpG free sequences.

TpA dinucleotides show suppression in a similarly G+C content dependent fashion, although with a negative correlation (Fig. 2.1B). The main cluster ranges from an observed over expected ratio of 0.2 to 1.0 with the G+C content spanning from 0.3 to 0.7. Similar to the CpG ratios, several outliers can be observed, including TpA high sequences.

When CpG and TpA frequencies are plotted against each other, there appears to be no correlation between CpG and TpA suppression, though outliers can be found here as well. Interestingly, the CpG free sequences show varying frequencies of TpA. There is also a small number of CpG low, TpA high genes.

At this point it should be noted that we [the Simmonds research group] superficially looked at those outliers and smaller clusters to look for any proteins of interest, but they mostly appeared to be proteins with no known immune function. One protein family of interest observed to be CpG low was Interferon alpha. A finding that has previously been reported by Greenbaum *et al.* (440). Peter Simmonds did similar analyses using interferon stimulated genes (ISGs), but these showed similar distributions of CpG and TpA suppression as all mRNAs (data not published).

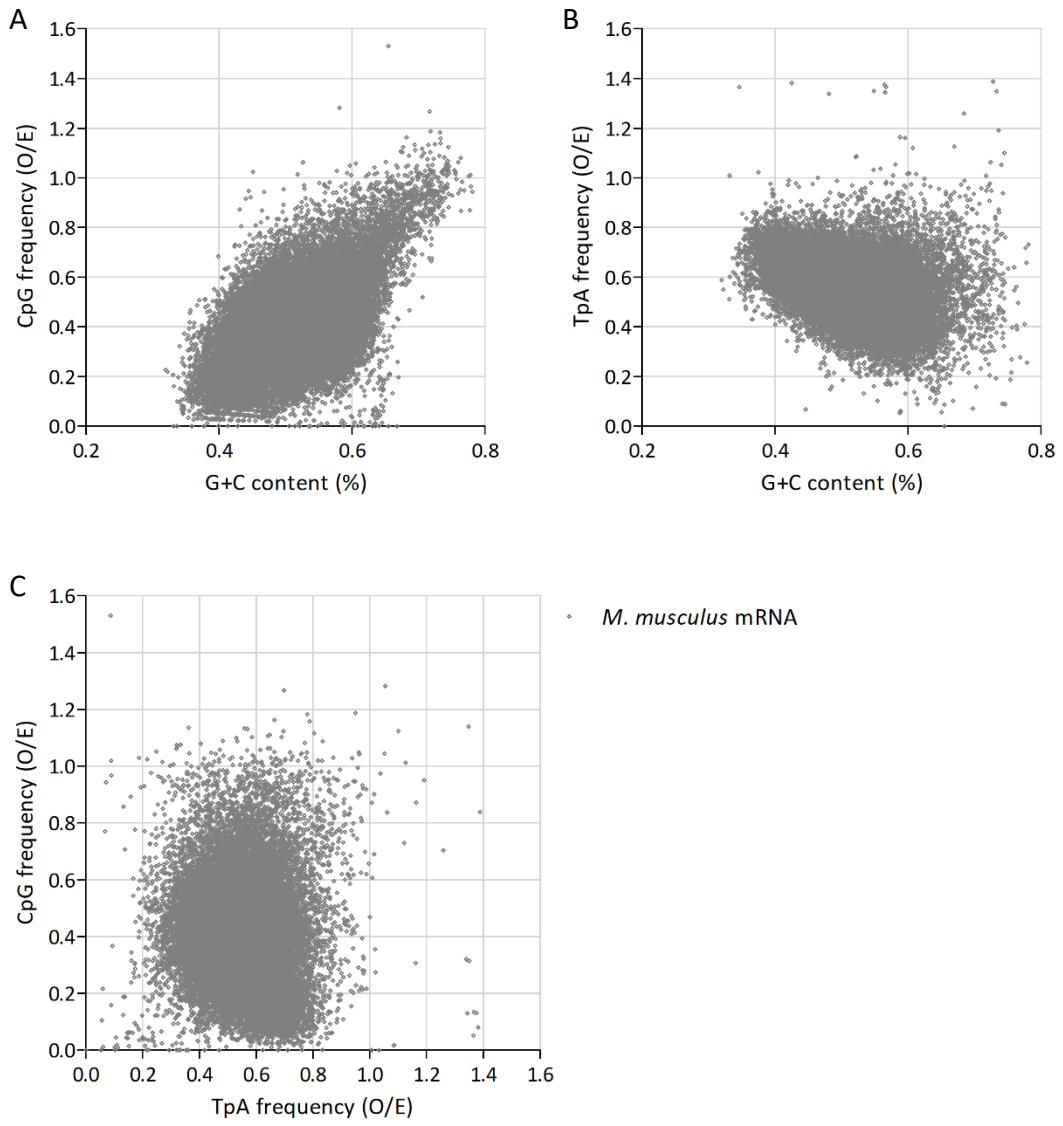


Figure 2.1 CpG and TpA distribution in murine mRNA
Each grey dot represents a separate gene. Observed (O) CpG and TpA frequencies were normalized to expected (E) values based on their G+C content. **A**. CpG frequencies relative to G+C content. **B.** TpA frequencies relative to G+C content. **C.** CpG frequencies relative to TpA frequencies.

Anelloviridae

Anelloviridae have a similar correlation between G+C content and CpG and TpA dinucleotide suppression as murine mRNA (Fig 2.2A+B), though both their CpG and TpA suppression appears to be less pronounced overall, especially at the genome level (Fig 2.2C).

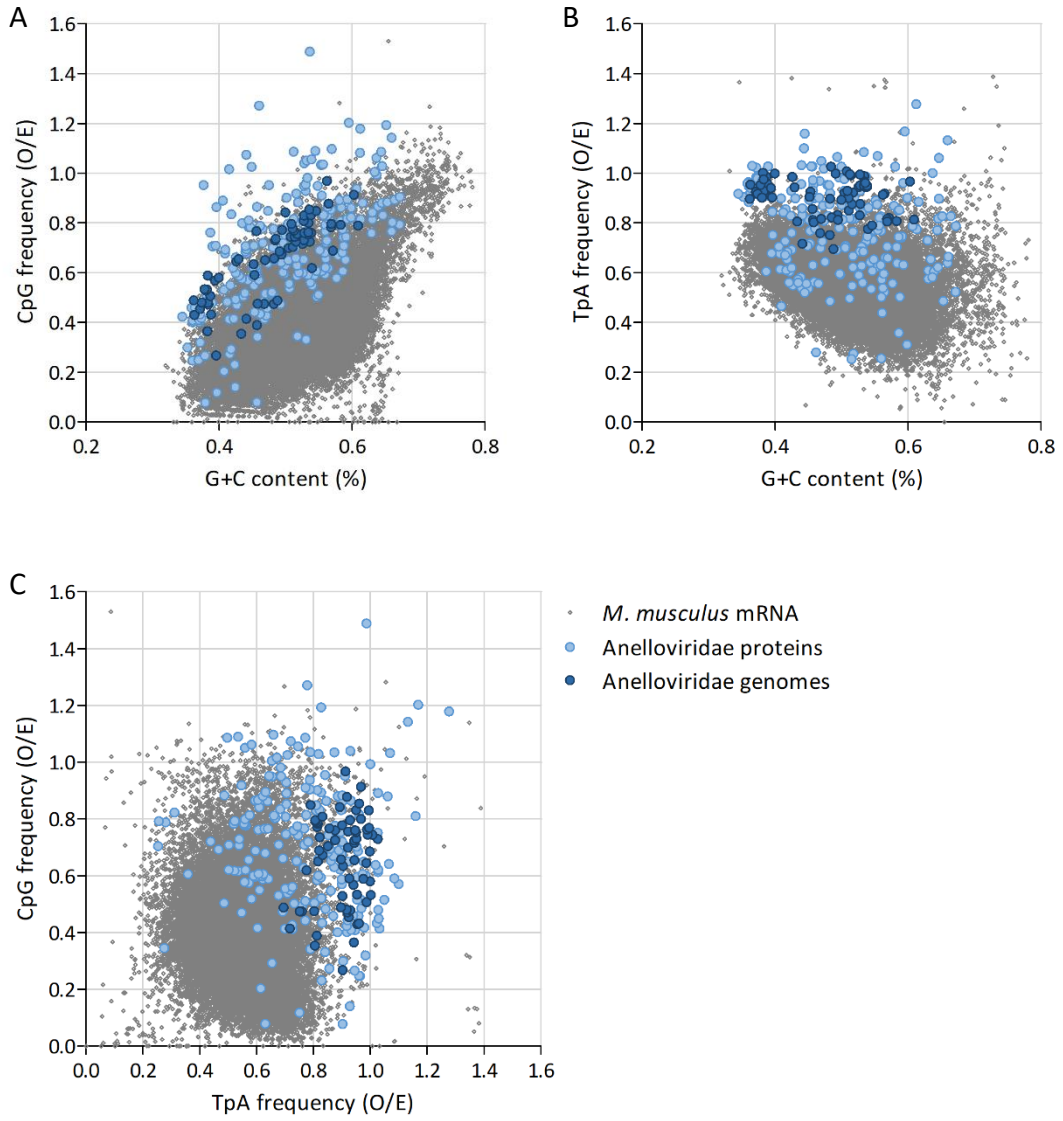


Figure 2.2 CpG and TpA distribution in Anelloviridae
CpG and TpA frequencies of viral genomic (dark blue circles) and protein sequences (light blue circles) were compared to those of murine mRNA (grey dots). Observed (O) CpG and UpA frequencies were normalized to expected (E) values based on their G+C content. **A**. CpG frequencies relative to G+C content. **B.** TpA frequencies relative to G+C content. **C.** CpG frequencies relative to TpA frequencies.

Circoviridae

Circoviridae genomes and proteins have a narrower range of G+C content (0.25-0.65) compared to murine mRNA and show very different CpG ratios compared to the murine host (Fig2.3A). While some viral proteins and genomes show low CpG frequencies, the majority of the reference genomes and corresponding viral proteins are not or only moderately CpG restricted. The TpA distribution of cirvoviruses is more similar to that of the human mRNA, though their TpA suppression does not reach the same level as some cellular RNAs (Fig2.3B). This difference in CpG and TpA expression is particularly obvious when plotted against each other, where the main circoviral cluster only partially overlaps with the murine mRNA cluster (Fig 2.3C).

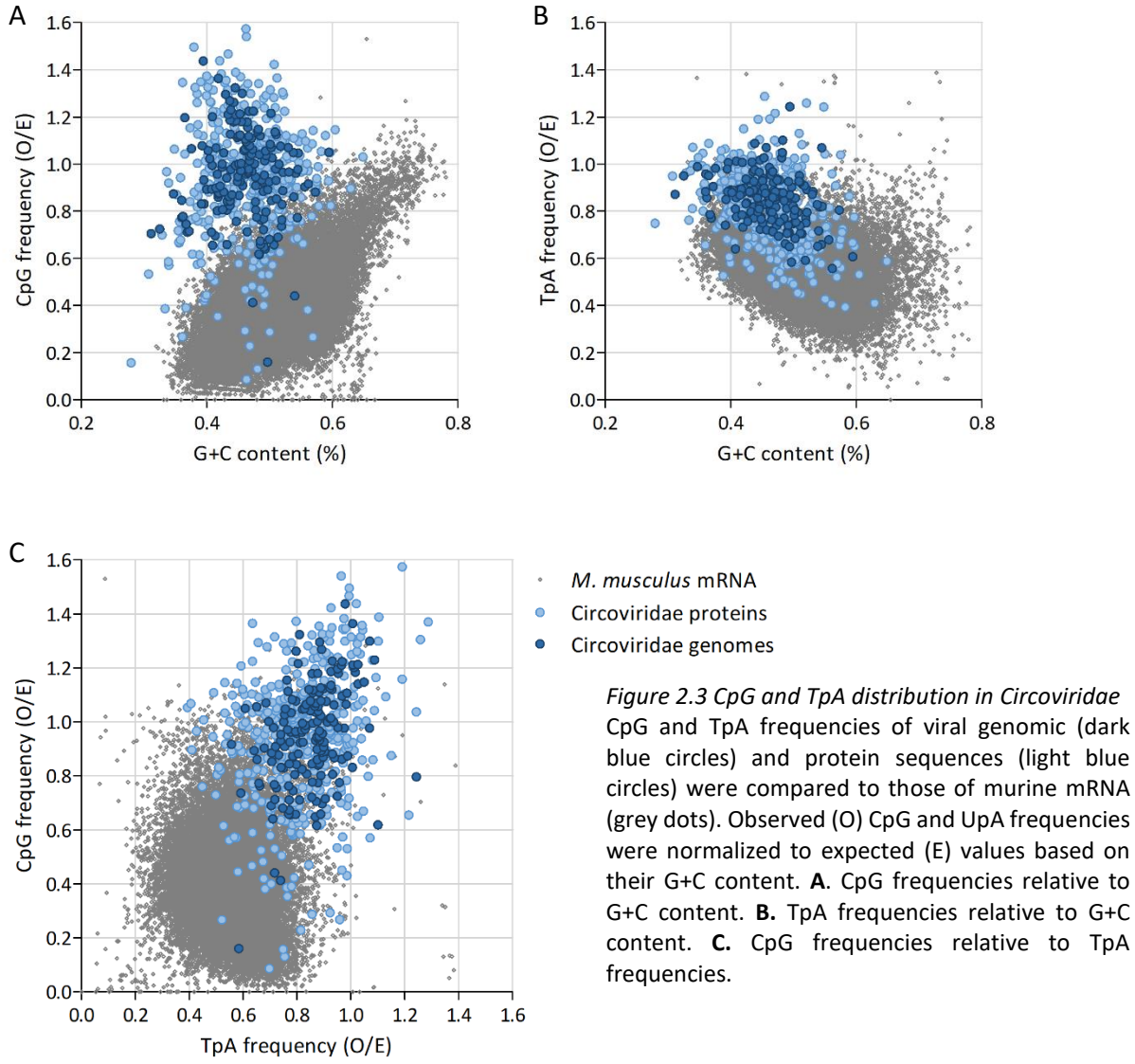


Figure 2.3 CpG and TpA distribution in Circoviridae
CpG and TpA frequencies of viral genomic (dark blue circles) and protein sequences (light blue circles) were compared to those of murine mRNA (grey dots). Observed (O) CpG and UpA frequencies were normalized to expected (E) values based on their G+C content. **A**. CpG frequencies relative to G+C content. **B.** TpA frequencies relative to G+C content. **C.** CpG frequencies relative to TpA frequencies.

Papillomaviridae

In contrast, *Papillomaviridae* mostly overlap with the murine mRNA clusters, both for CpG and TpA dinucleotides (Fig2.4A+B). Interestingly, their genomes cluster in a narrower range of G+C content (0.35-0.6), moderate to high CpG suppression and no to moderate TpA suppression (Fig 2.4C).

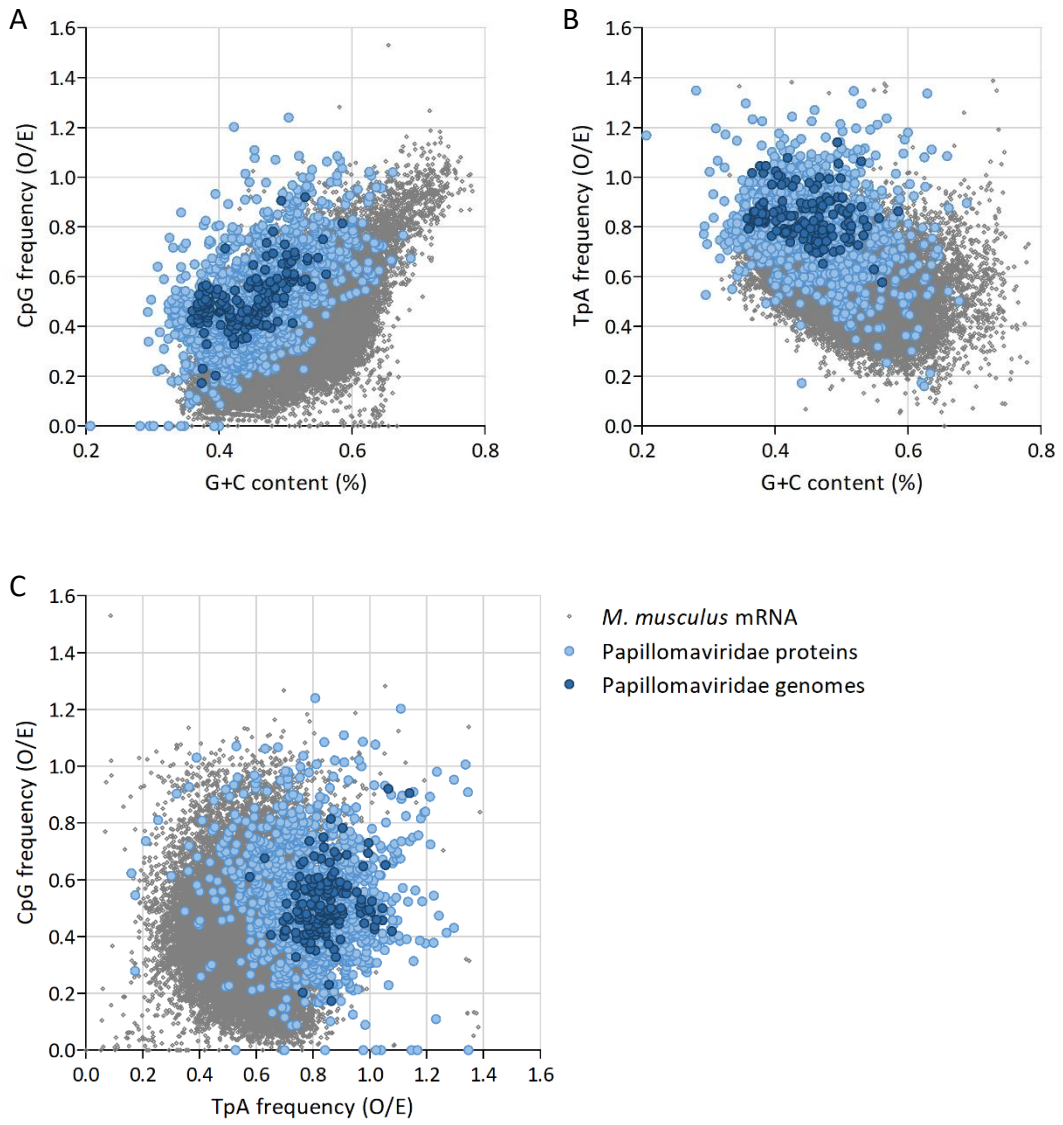


Figure 2.4 CpG and TpA distribution in Papillomaviridae
CpG and TpA frequencies of viral genomic (dark blue circles) and protein sequences (light blue circles) were compared to those of murine mRNA (grey dots). Observed (O) CpG and UpA frequencies were normalized to expected (E) values based on their G+C content. **A**. CpG frequencies relative to G+C content. **B.** TpA frequencies relative to G+C content. **C.** CpG frequencies relative to TpA frequencies.

Parvoviridae

Similar to *Papillomaviridae*, parvoviral genomes and proteins overlap with the human mRNA in terms of CpG and TpA suppression, though they cluster slightly differently. *Parvoviridae* have an overall narrower range of G+C levels (0.25-0.65) compared to murine mRNA, while their CpG levels are in a similar range (0-1.25) and likewise correlated to G+C content (Fig 2.5A). Furthermore, the TpA observed over expected ratios (0.4-1.25) are very similar to the murine mRNA levels at the respective G+C levels (Fig 2.5B). Overall, these viruses and their protein coding sequences show similar levels of CpG suppression as the murine host does, but with a narrower spectrum of TpA restriction, possibly due to a lack of G+C rich viruses (Fig 2.5C).

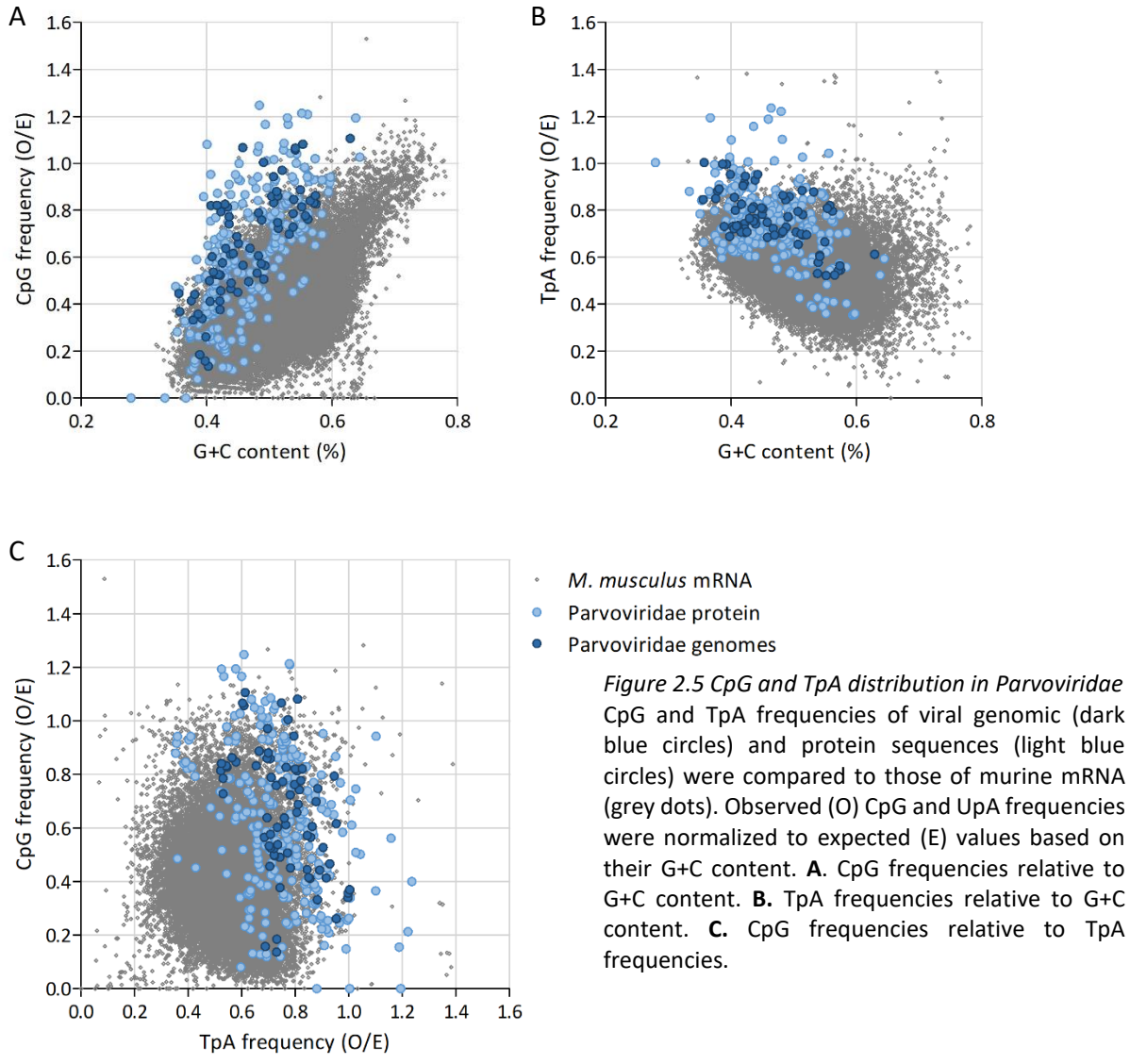


Figure 2.5 CpG and TpA distribution in Parvoviridae CpG and TpA frequencies of viral genomic (dark blue circles) and protein sequences (light blue circles) were compared to those of murine mRNA (grey dots). Observed (O) CpG and UpA frequencies were normalized to expected (E) values based on their G+C content. **A**. CpG frequencies relative to G+C content. **B.** TpA frequencies relative to G+C content. **C.** CpG frequencies relative to TpA frequencies.

Polyomaviridae

Polyomaviridae show a remarkable CpG suppression in both genomes and protein coding sequences and overall G+C low to medium genomes (Fig2.6A). Interestingly, they show no or minimal TpA restriction in their genomes, while their protein coding sequences show very similar TpA restriction compared to the murine mRNA (Fig 2.6B+C).

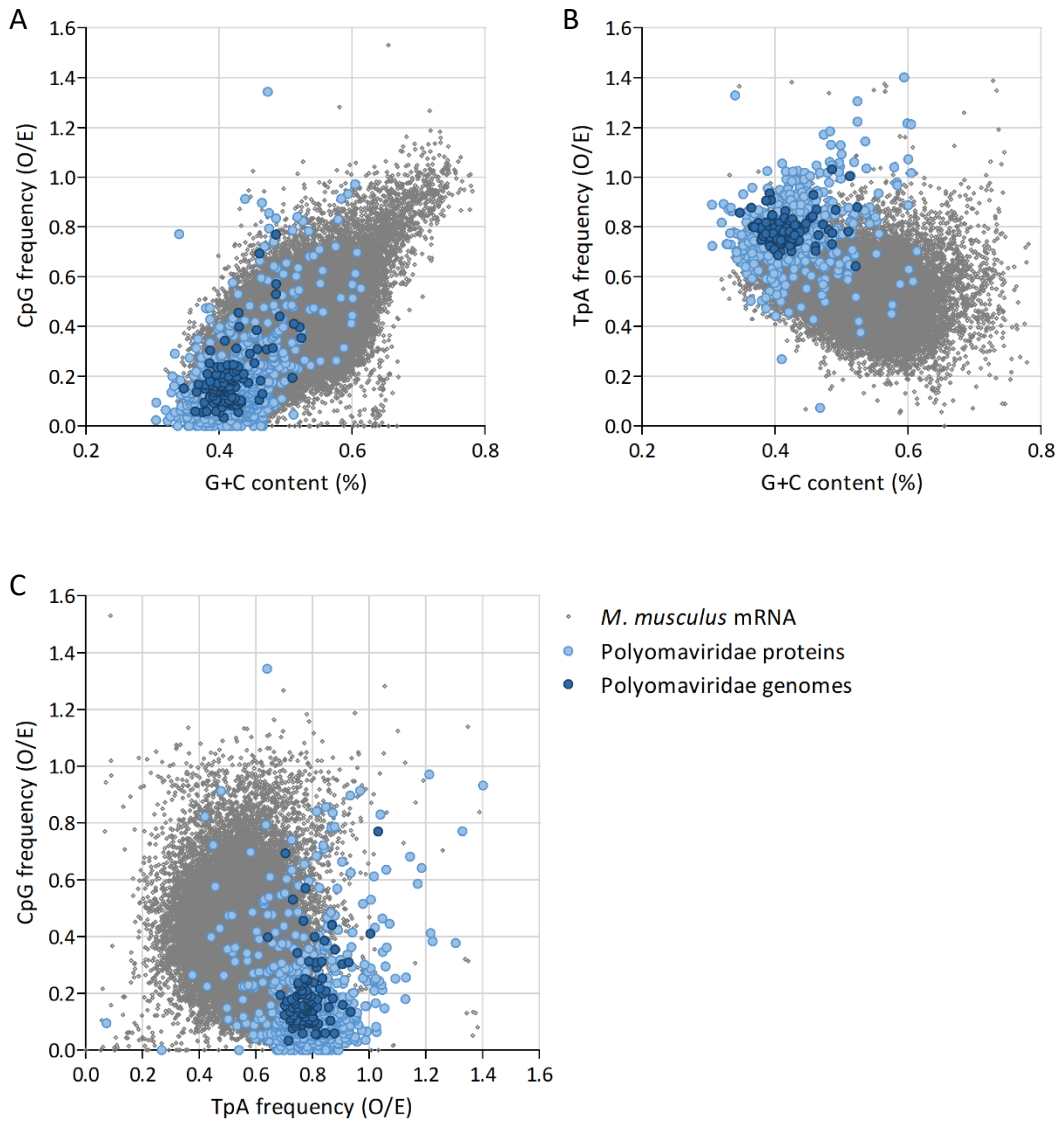


Figure 2.6 CpG and TpA distribution in Polyomaviridae
CpG and TpA frequencies of viral genomic (dark blue circles) and protein sequences (light blue circles) were compared to those of murine mRNA (grey dots). Observed (O) CpG and UpA frequencies were normalized to expected (E) values based on their G+C content. **A**. CpG frequencies relative to G+C content. **B.** TpA frequencies relative to G+C content. **C.** CpG frequencies relative to TpA frequencies.

Similarities and differences between families

While different viral families show distinct patterns of CpG and TpA restriction, there are some commonalities between them. With the exception of *Circoviridae*, these small DNA viruses appear to suppress CpG dinucleotides to a similar to mice (*Anelloviridae*, *Papillomaviridae* and *Parvoviridae*) or greater extent than mice (*Polyomaviridae*) and potentially other hosts. Interestingly, these viruses share an overall lower and less varied G+C content than murine RNA. And while their TpA suppression is mostly similar to the TpA suppression found at matching G+C levels in murine mRNA, their overall TpA suppression is less pronounced due to their lower G+C content.

It should be noted that viruses for this analysis were filtered using the host annotation in RefSeq. No attempt was made to verify whether the selected viruses were correctly classified or to further distinguish between different kinds of vertebrate hosts as this was beyond the scope of this study. Therefore, not much attention was paid to outliers or potential clusters within viral families, as there were too many unknown variables involved at this stage.

2.2 Choosing a suitable small DNA model

After analysing the CpG and TpA distribution of the five small DNA viral families infecting vertebrates, four families (*Anello*-, *Papilloma*-, *Parvo*- and *Polyomaviridae*) were brought forward as potential study models, while *Circoviridae* were excluded as they did not show sufficient CpG suppression.

In order to study the effects of dinucleotide composition, several criteria have to be fulfilled. First of all, a reverse genetics system has to be available to manipulate the genome of the chosen virus(es) and produce mutant viruses. On a secondary level, this includes finding genomic regions that are suitable for genetic manipulation, i.e. devoid of regulatory or structural elements, with no overlapping reading frames or nearby splice sites.

Next, there needs to be an established cell culture system that allows the characterisation of different aspects and stages of the viral life cycle. Ideally, the chosen virus(es) would be able to infect different cell lines, especially those were knockout cell lines are available. Furthermore, an animal model should be available for any chosen virus to be able to look at infection and immune response on a tissue and organism level. Lastly, any virus picked would ideally be a low risk pathogen to allow ease of use while facilitating a good immune response. In theory it is possible to use different viruses for different stages of this investigation. However, the long synthetic sequences required for making adequate mutants are costly and quickly add up when several mutants are of interest. Furthermore, for every new virus used, there is an optimisation process, to find suitable regions to mutate and to ensure the resulting mutant viruses are viable.

With these requirements and restrictions in mind, mouse viruses are great candidates to study compositional mutant viruses. Mouse cell lines are commonly used and mice with their

small size and easy maintenance are convenient for animal experiments. There are many strains of mice available and an ever-increasing number of knockout animals and cell lines along with them.

With this in mind follows an overview of viruses used and available for each of the four virus families and their suitability for this study.

Parvoviridae

Parvoviruses can be separated into autonomous viruses and dependoviruses, with the latter requiring helper viruses such as adenoviruses to complete their life cycle. Only autonomous parvoviruses were considered for this study to avoid two-virus.

Several autonomous parvoviruses have been extensively studied in the last decades, including those associated with human or animal disease (such as human parvovirus B19, canine parvovirus or porcine parvovirus). There are several known rodent parvoviruses, with the best studied ones being rat parvovirus H-1 and mouse parvovirus MVM.

While parvovirus H-1 has mainly been of interest as an oncolytic agent (reviewed by Bretscher and Marchini (441)), MVM has been used extensively as a model virus to gain insights into various aspects and stages of the (proto)parvoviral lifecycle (reviewed by Berns and Parrish (392)). Given these factors and the general preference for using a mouse virus, it was decided to investigate MVM as a small DNA virus model.

MVM has a well-established reverse genetics system (442). The infectious clone can be transfected into cells as the whole plasmid or after *BamHI* excision of the viral genome (442). Progeny virus will be produced without further intervention. The early promoter is cell driven and NS1 is able to initiate and complete replication from a double stranded template

(380,442). There are several well characterised, permissive cell lines available (443,444) and in vivo work is possible in mice (362).

Furthermore, MVM shows a marked suppression of CpG dinucleotides, both on a genomic level and within its ORFs (Fig 2.7A). Interestingly, the remaining CpGs are not spread equally, but seem to cluster towards the terminal regions and in areas with other restrictions, such as overlapping ORFs or splice sites while most CpGs are eliminated in regions with single ORFs (Fig. 2.7B).

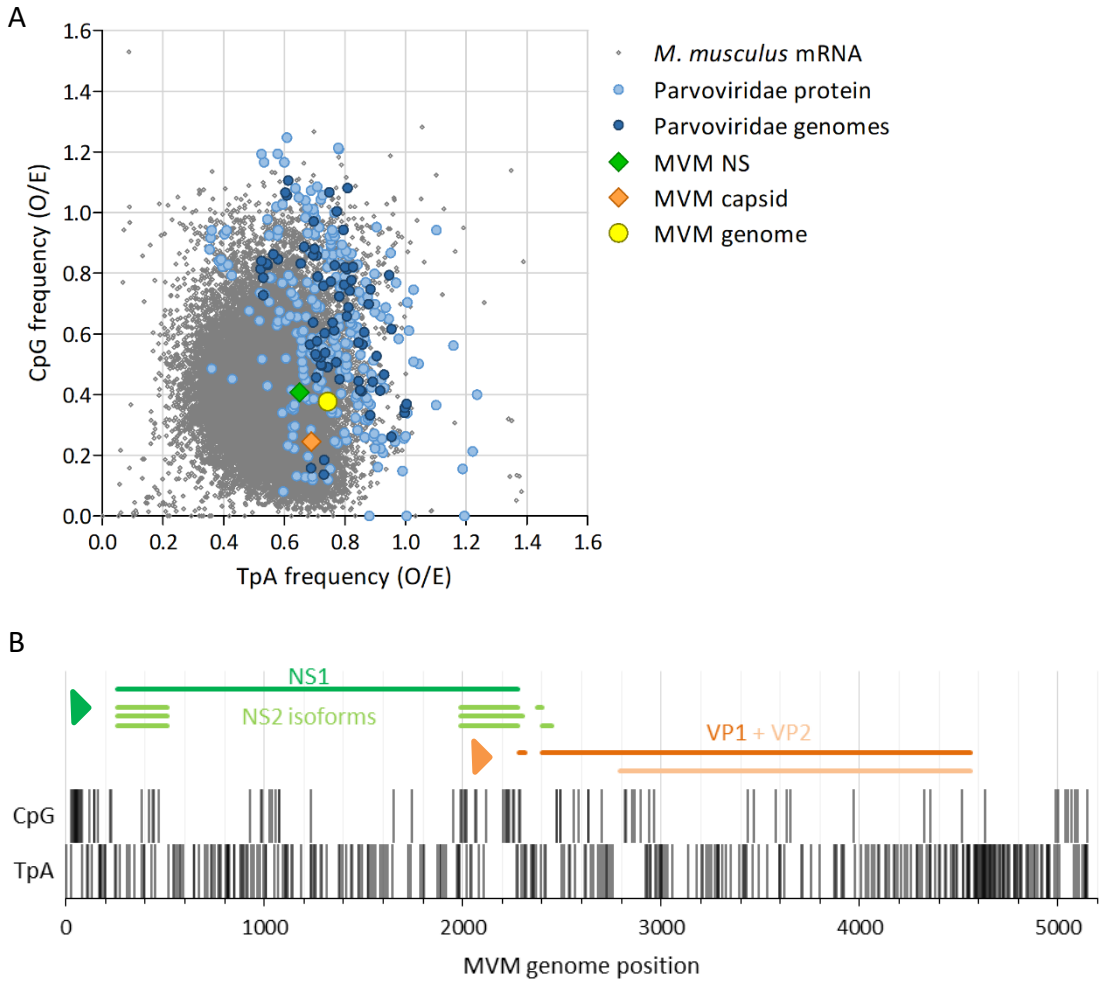


Figure 2.7 CpG and TpA distribution in minute virus of mice
A. CpG and TpA frequencies of the MVM genome (yellow circle) and NS (orange diamond) and capsid (green diamond) proteins were compared to parvoviral genomes (dark blue circles), parvoviral proteins (light blue circles) and human mRNA (grey dots). Observed (O) CpG and TpA frequencies were normalized to expected (E) values based on their G+C content.
B. Distribution of CpGs and TpAs along the MVM genome. NS (green) and capsid (VP, orange) ORFs are overlayed for orientation. Green and orange arrowheads indicate direction of translation.

Polyomaviridae

The first polyomavirus identified was MPyV and it has been well characterised since (Fig. 2.8). It has been one of the most studied polyomaviruses, along with SV40, which infects monkeys and humans.

MPyV has two reverse genetics systems, either using restriction enzymes for backbone excision and self-ligation or a newer system using a Cre-LoxP system in cell culture to recover infectious particles. Traditionally, MPyV infectious DNA was generated by excising the viral genome from the bacterial vector, followed by low concentration self-ligation using ligases such as T4. However, recently a Cre-LoxP system was established allowing the excision and ligation of the viral DNA inside of cells and thus eliminating the rate limiting step of T4 ligation. The only caveat of this method is a 31 bp loxP site which is left in the genome after successful excision (445).

MPyV can grow in 3T6 cells and primary MEFs, as well as several other cell lines, allowing various knockout studies as well as in vivo work in mice (420,446).

MPyV shows a marked suppression of CpG dinucleotide frequencies across its genome and within its ORFs.

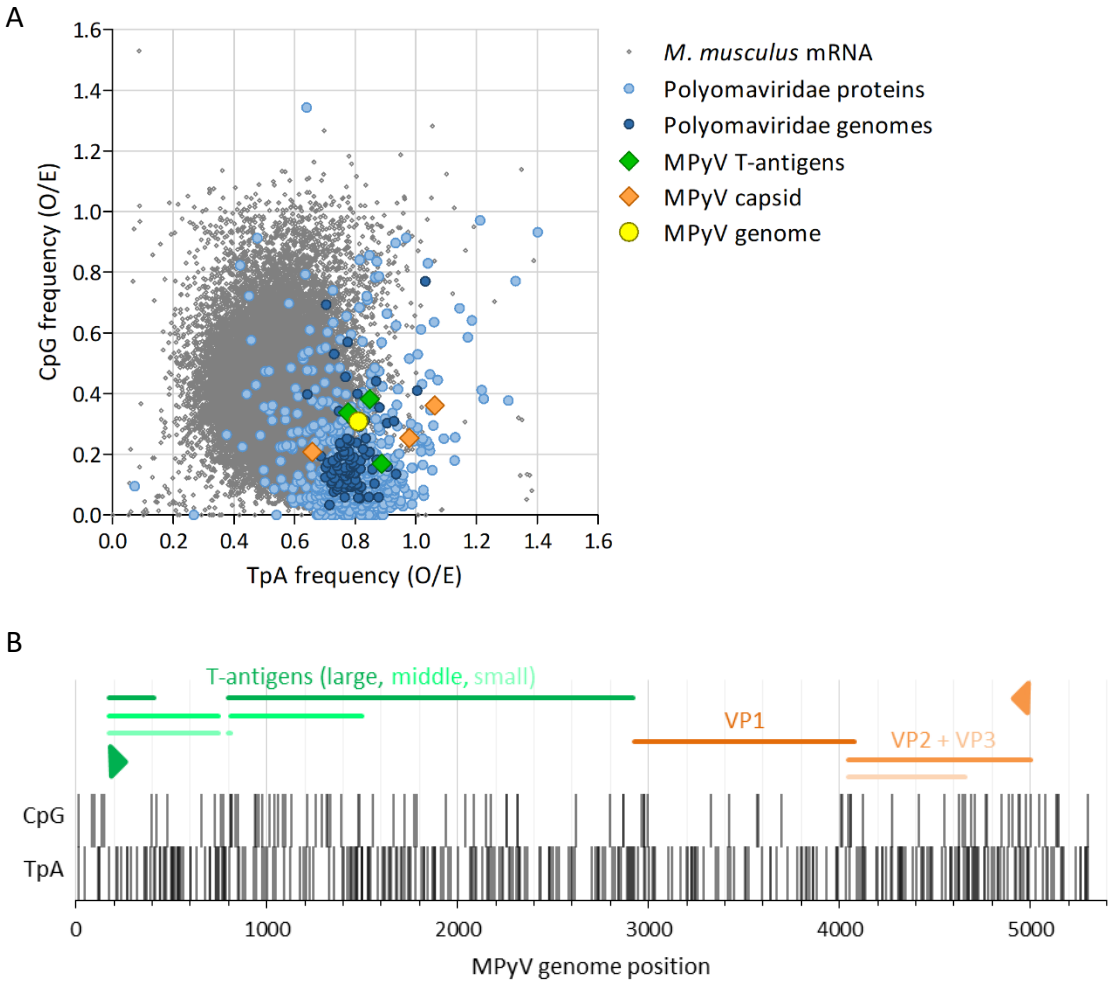


Figure 2.8 CpG and TpA distribution in murine polyomavirus
A. CpG and TpA frequencies of the MPyV genome (yellow circle) and NS (orange diamond) and capsid (green diamond) proteins were compared to polyomaviral genomes (dark blue circles), polyomaviral proteins (light blue circles) and human mRNA (grey dots). Observed (O) CpG and TpA frequencies were normalized to expected (E) values based on their G+C content. **B.** Distribution of CpGs and TpAs along the MPyV genome. NS (green) and capsid (VP, orange) ORFs are overlayed for orientation. Green and orange arrowheads indicate direction of translation.

Other small DNA viruses

For papillomaviruses, in vitro and cell culture experiments are usually conducted using human papillomaviruses and specialised cell culture systems are used to mimic natural infection of the skin (447 ,448). While a mouse papillomavirus was recently discovered, there is currently no cell culture system available (449-451). With these caveats, papillomaviruses appear to be quite a difficult group to study, as they would necessitate specialised cell culture as well as different viruses for cell culture and animal work.

For anelloviruses, there is currently no cell culture system available (as reviewed recently by Kaczorowska and van der Hoek (352)), rendering them not suitable for this study.

Due to these limitations, no candidates from these virus families were considered for further study given the scope of the project.

2.3 MVM mutant selection and generation

Region selection

In order to use MVM as a model, a suitable region for genetic manipulation was necessary. There are two commonly used strains of MVM, the prototype strain (MVMp) and immunosuppressive strain (MVMi). While both strains are very similar genetically, there are various mutations throughout the genome. In particular, non-synonymous mutations within capsid genes affect cell tropism (364,452). Due to a number of both synonymous and non-synonymous point mutations, mutant DNA sequences created for one strain cannot easily be transferred to the other strain. For this project, the prototype strain of MVM was used as this strain has been predominantly used for cell culture experiments.

To ensure that no overlapping open reading frames or structural elements were jeopardised, the available literature was consulted for the genetic map of the virus genome (Fig. 2.9A) and pairwise distance scans at synonymous sites were conducted using SSE (Fig. 2.9B)(439)). The pairwise distance scan was used to determine regions with high synonymous coding variability indicating less conserved codon usage for the given protein sequence and thus a likely lack of any conserved secondary structures and alternative ORFs. As seen in Fig. 2.9, the synonymous coding variability within the non-structural genes of MVM varies between 0.13 and 0.43. For the capsid genes, the synonymous coding variability spans from 0.16 to 0.70, with a sharp increase in variability around genome position 2900-3000.

The lower variability observed in the middle of the MVM genome is likely due to a combination of factors. NS1 and NS2 have overlapping ORF and there are several splice donor and acceptor sites in this region.

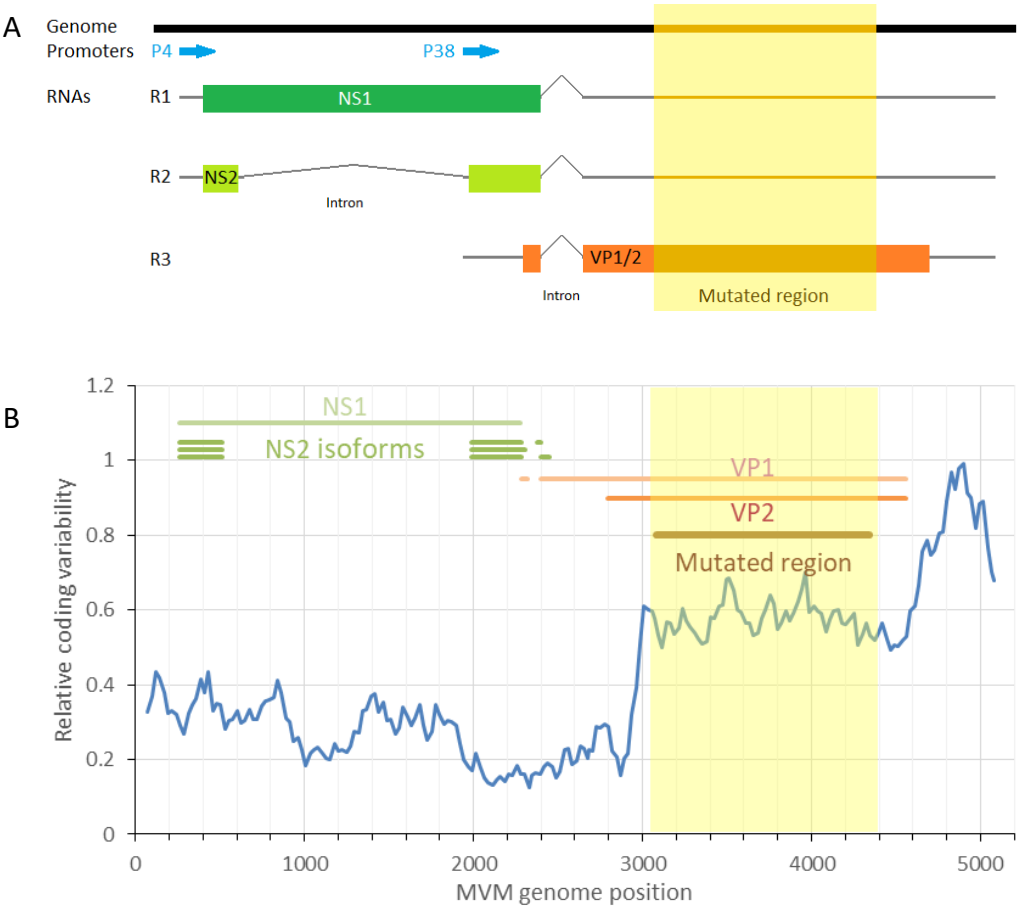


Figure 2.9 MVM genomic structure and the capsid gene region selected for mutagenesis
A. Diagram of the MVM genomic structure. Promoters (blue), transcripts (R1-R3) with ORFs (green and orange coloured boxes), mutated region highlighted in yellow. Based on Berns & Parrish, Fig 57.10 (392). **B.** SSE was used to calculate the synonymous coding variability observed between MVM and related parvoviruses. The yellow region indicates the region chosen for mutations.

A 1.3kb region (24.4 % of the genome) was chosen within the more highly variable part of the capsid genes. This region contained no known regulatory or structural features, encoding only a single open reading frame shared by VP1 and VP2 (Fig 2.9A. This region was replaced by various synthetic inserts using sticky end cloning.

In short, the mutated fragment spanned bases 3085-4342 of the MVM genome. Due to the location of suitable restriction sites (*Bpu10I* starting at base 2647 and *XbaI* starting at base 4342), the cloning fragment was several hundred base pairs longer than the mutated region. A *HpaI* restriction site was located relatively central within the chosen region (starting at base 3757) and allowed for smaller inserts to be made (approx. 0.7 kb and 0.6 kb, respectively).

Mutant selection – 1st generation

To confirm the bioinformatic inferences about the region being suitable, a scrambled mutant (CDLR) was made, where codons were randomly switched around within the mutated region in order to destroy any potential secondary or regulatory structures. This mutant served as a control to test the general suitability for mutating the chosen region and remained synonymously coding with the same codon usage while maintaining the CpG and TpA frequencies of the wild-type (WT) virus (Table 2.1).

Next, mutants with elevated CpG or TpA frequencies were made. For the CpG-high (CpG-H) mutant, CpG frequencies were maximised while maintaining native levels of TpA, increasing CpGs to 1.62 times the expected frequency given its G+C content (O/E ratio, table 2.1), a 13-fold increase over the wild-type sequence. Likewise, TpAs were maximised for the TpA-high (TpA-H) mutant to 1.23 times the expected frequencies, approximately doubling native levels while the number of CpGs remained constant. For the CpG-high mutant, half mutants were made (CpG-H 1st half and CpG-H 2nd half) as it was unclear how severe the CpG-dependent attenuation would be in this model. This was complemented by a codon-optimised mutant (CodOpt), which is CpG supressed as a by-product of mimicking cellular mRNA composition (O/E ratio of 0.61), but considerably higher than the wild-type ratio (O/E ratio of 0.13).

The sequence of the CodOpt mutant was generated by a proprietary algorithm by GeneArt, the company which synthesised all the synthetic inserts in this project. This was chosen as their algorithm is both optimised for expression as well as avoiding potential interfering elements such as cryptic splice sites.

The wild-type and these six mutants were used for the bulk of the experimental work described in this project. Further mutants described below were added to the project later

on, both to look into questions that had arisen during data collection and as further controls in anticipation of the peer review involved in publishing the material.

Table 2.1: Dinucleotide composition of 1st generation MVM mutant regions

Mutant	Fragment length	G+C content	CpG		GpC		TpA	
			Number	O/E ratio	Number	O/E ratio	Number	O/E ratio
WT	1257	42.32 %	7	0.126	58	1.04	59	0.585
CDLR	1257	42.32 %	7	0.126	58	1.04	59	0.585
CpG-H	1257	47.89 %	117	1.634	70	0.978	59	0.716
CpG-H 1st half	676	46.75 %	59	1.612	39	1.066	37	0.800
CpG-H 2nd half	582	49.14 %	58	1.660	31	0.887	22	0.606
TpA-H	1257	42.32 %	7	0.126	59	1.058	124	1.227
CodOpt	1257	61.10 %	70	0.611	108	0.943	23	0.524
Whole genome	5149	42.12 %	86	0.378	256	1.1251	312	0.7428

Table 2.1 Dinucleotide composition of 1st generation MVM mutant regions
O/E: observed over expected. Grey background used to highlight changes of interest.

Mutant selection – 2nd generation

Several further control mutants were made to rule out attenuating effects of other compositional modifications. Firstly, a GpC-high mutant was made with GpC frequencies elevated to match the 1.62 observed over expected ratio of the CpG-H mutant. This mutant was used as a control for increasing self-complimentary sequences by introducing CpGs (and TpAs) and the potentially resulting increase in forming secondary structures and increased RNA folding.

When the original CpG-H mutant was made, SSE, the software used to generate the mutants, was not able to keep G+C content consistent while increasing CpG frequencies, leading to an increase in G+C (from 42 % to 48 % for the CpG-H mutant). However, with a newer version of SSE, this became possible. This increase in G+C content could potentially affect viral fitness independent of CpG frequency. To differentiate better between effects due to changes in CpG frequency and G+C content, a CpG-high mutant with the same G+C content as wild-type was made (CpG-H 42% G+C). The mutant had 78 CpGs with a CpG O/E ratio of 1.4, 11x that of wild-type MVM. This was the limit for this fragment without changing the G+C content.

Interestingly, codon optimisation also lead to a major increase in G+C content (61% compared to a native 42%).

Ideally, a codon-optimised mutant with a fixed G+C content of 42% would have been generated, however that was technically not possible. Instead, a CpG-low codon-optimised mutant (CodOpt CpG-L) was designed. The online tool CooL (453) was used to construct the codon-optimised, CpG-low region. Using the codon adaptability index (CAI) as a proxy for codon optimisation, we tested different parameters for making the last mutant. The wild-type virus has a CAI of ~70% for the mutated region, while the original codon-optimised region designed by GeneArt has a CAI of 98%. With the CooL tool, it is possible to fix the number of CpGs and set a target percentage for the G+C level along with optimising the sequence. However, it was not possible to increase CAI without increasing G+C levels, while it was possible to reach a CAI of 94% while retaining the wild-type number of CpGs in the region (7 CpGs).

Table 2.2: Dinucleotide composition of 2nd generation MVM mutant regions

Mutant	Fragment length	G+C content	CpG		GpC		TpA	
			Number	O/E ratio	Number	O/E ratio	Number	O/E ratio
WT	1257	42.32 %	7	0.126	58	1.04	59	0.585
GpC-H	1257	42.32 %	7	0.125	91	1.626	59	0.585
CpG-H	1257	47.89 %	117	1.634	70	0.978	59	0.716
CpG-H 1st half	676	46.75 %	59	1.612	39	1.066	37	0.800
CpG-H 2nd half	582	49.14 %	58	1.660	31	0.887	22	0.606
CpG-H 42% G+C	1257	42.32 %	78	1.398	58	1.039	60	0.595
CodOpt	1257	61.10 %	70	0.611	108	0.943	23	0.524
CodOpt CpG-L	1257	57.04 %	7	0.069	92	0.908	25	0.444
Whole genome	5149	42.12 %	86	0.378	256	1.1251	312	0.7428

Table 2.2 Dinucleotide composition of 2nd generation MVM mutant regions
O/E: observed over expected. Selected 1st generation MVM mutants added for ease of comparison. Grey background used to highlight changes of interest.

Construction of an R3 expression plasmid

Due to the genomic structure of MVM, it is not possible to quantify the R3 RNA encoding the capsid proteins separately (Fig. 2.9). In order to measure R3 mRNA and VP protein expression independent of the non-structural proteins, a sub genomic expression vector was made. This plasmid contains a chicken actin gene promoter and enhancer cassette (pCAG), followed by the part of the genome that encompassed the (not yet spliced) R3 RNA (454,455). The R3 inserts were cloned into the pCAG backbone using high-fidelity PCR and sticky end cloning.

Two suitable restriction enzymes were chosen (*SacI* and *SwaI*) and restriction sides were added to the R3 insert using a nested PCR, before proceeding with a standard cloning protocol.

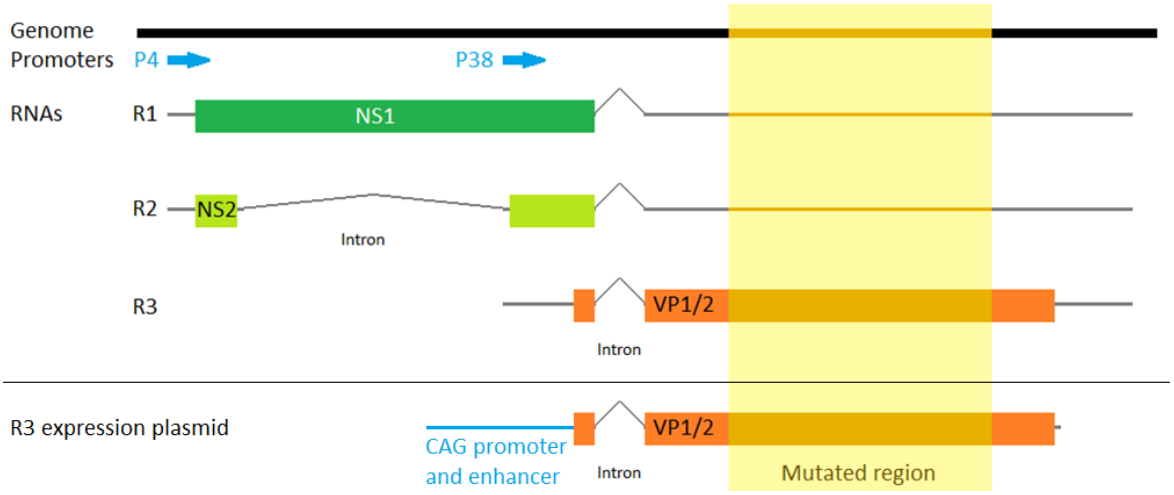


Figure 2.10 MVM genomic structure and R3 expression plasmid
Diagram of the MVM genomic structure. Promoters (blue), transcripts (viral R1-R3 plus R3 from expression plasmid) with ORFs (green and orange coloured boxes), mutated region highlighted in yellow. Upper part based on Berns & Parrish, Fig 57.10 (392).

2.4 MVM virus recovery and quantification

Virus recovery

To produce infectious particles of MVM, plasmids encoding the infectious clone (pMM984 (442)) and its respective mutants were transfected into NB324K cells (443) or A9 cells (444). After 24 hours, cells were transferred into appropriate tissue culture flasks and grown 3-4 days until confluent. The cells were harvested, resuspended in vTE (50 mM Tris-hydrochloride, 0.5 mM EDTA, pH 8.7 (368)) and freeze-thawed tree times to collect passage 0 (p0) virus (456). The presence of infectious particles was confirmed by either visible cytopathic effect (CPE) or viral mRNA transcription after infection.

Unexpectedly, no infectious particles were detected following the protocol described above. After transfection, mRNA was produced, though levels declined over time. Immunofluorescent staining for the capsid showed that protein was produced within 24 hours of transfection, but only in a fraction of the cells. However, this was partly due to poor transfection efficiencies, as GFP control transfections yielded similar numbers of fluorescent cells. Consequently, transfection was optimised by using the maximum amount of reagent and DNA without causing severe cell toxicity (400 ng DNA per 10^5 cells, with 3 µl Lipofectamine 2000 per µg DNA). NB324K cells were chosen over A9 cells for stock production due to a 10-fold greater transfection efficiency and higher viral yield (see below).

The infectious clone is prone to deleterious recombination in bacteria which is prohibitive for viral replication (457). To address this issue, recombination deficient SURE 2 Supercompetent Cells (Agilent, USA) were used for cloning and all new clones were screened for their intactness using stem loop PCRs as Sanger sequencing could not reliably sequence through the 3’ stem loop. With these adaptations, infectious particles were recovered (p0 virus). P1

virus stocks were grown by infecting NB324K cells with p0 virus at a low MOI and harvesting the cells at the onset of visible CPE to increase virus titre. Both NB324K and A9 cells were tested for p1 production, but A9 cells did not sufficiently support viral growth for the UpA and CpG high mutants to yield appropriate virus stock concentrations.

Virus quantification - Infectivity

The newly produced viruses were quantified using plaque assays and end-point dilution assays (50% Tissue culture Infective Dose, $TCID_{50}$).

In the literature, both A9 and NB324K cells have been used for MVM plaque assays (368,458). For WT virus, both cell lines are suitable, though NB324K cells usually have a ~10-fold higher plaque count compared to A9 cells (458). For the mutants however, the cell lines appeared to have different levels of susceptibility manifesting in their relative plaque count. While WT and TpA-H viruses showed the same count in NB324K cells, TpA-high virus produced approximately 10-fold fewer plaques in A9 cells compared to WT. No plaques could be observed for the CpG-high mutant in A9 cells. While plaques could be observed for the CpG-high mutant in the NB324K cells, the plaques produced by MVM in this cell line were rather small, making counting very time consuming.

It would have been possible to stain the cells with an anti-capsid antibody to count infectious foci rather than plaques, especially in the A9 cells where not all mutants produced plaques. However, given the high number of virus titrations anticipated (e.g. infectivity kinetics for the entire panel of mutant viruses including several time points), it was decided that $TCID_{50}$ assays would be more appropriate if possible.

The $TCID_{50}$ assay has been used with MVM before, but only with WT virus (459). While optimising the assay, it became apparent that the titrating the WT virus was more straightforward and robust than some of the mutants. Both cell density (before and after seeding) and the foetal bovine serum used affected the titre readout.

For the most reliable results, cells were seeded when the flasks were 80% confluent, cells were seeded at 3000 cells/well in a 96-well plate and mixed with the virus dilutions as soon as all plates were seeded. This low cell density was chosen to allow cell growth for as long as possible, allowing some of the slower viruses to undergo several rounds of replication before the wells were completely overgrown. While this made scoring more reliable, there was still some variation observed, where the readouts for the entire set would increase or decrease by up to one log when the same stocks were done on different occasions. In order to avoid introducing errors when new batches of virus stocks were created, WT and CpG-H virus stocks of known concentration were titrated along any new virus stocks and infectious titres were normalised to the WT virus.

Both A9 and NB324K cells were tested for their suitability to measure the infectivity of the various mutants. For this, the $TCID_{50}$ plates were scored at several time points to find the most appropriate time to score all viruses (Fig. 2.11). For best evaluation, cytopathic effects were scored as either partial (i.e. wells are visibly infected, but many/most cells are still alive and/or not affected yet) or complete (cells are dead or dying), with Fig. 2.11 counting partial CPE as positive. With the above protocol, reliable readouts were possible until day 13 for both cell lines, before deterioration of non-infected control cells was too profound to distinguish between viral and non-viral cytopathic effects and cell death. The different mutants showed different behaviours and titration endpoints in the two cell lines.

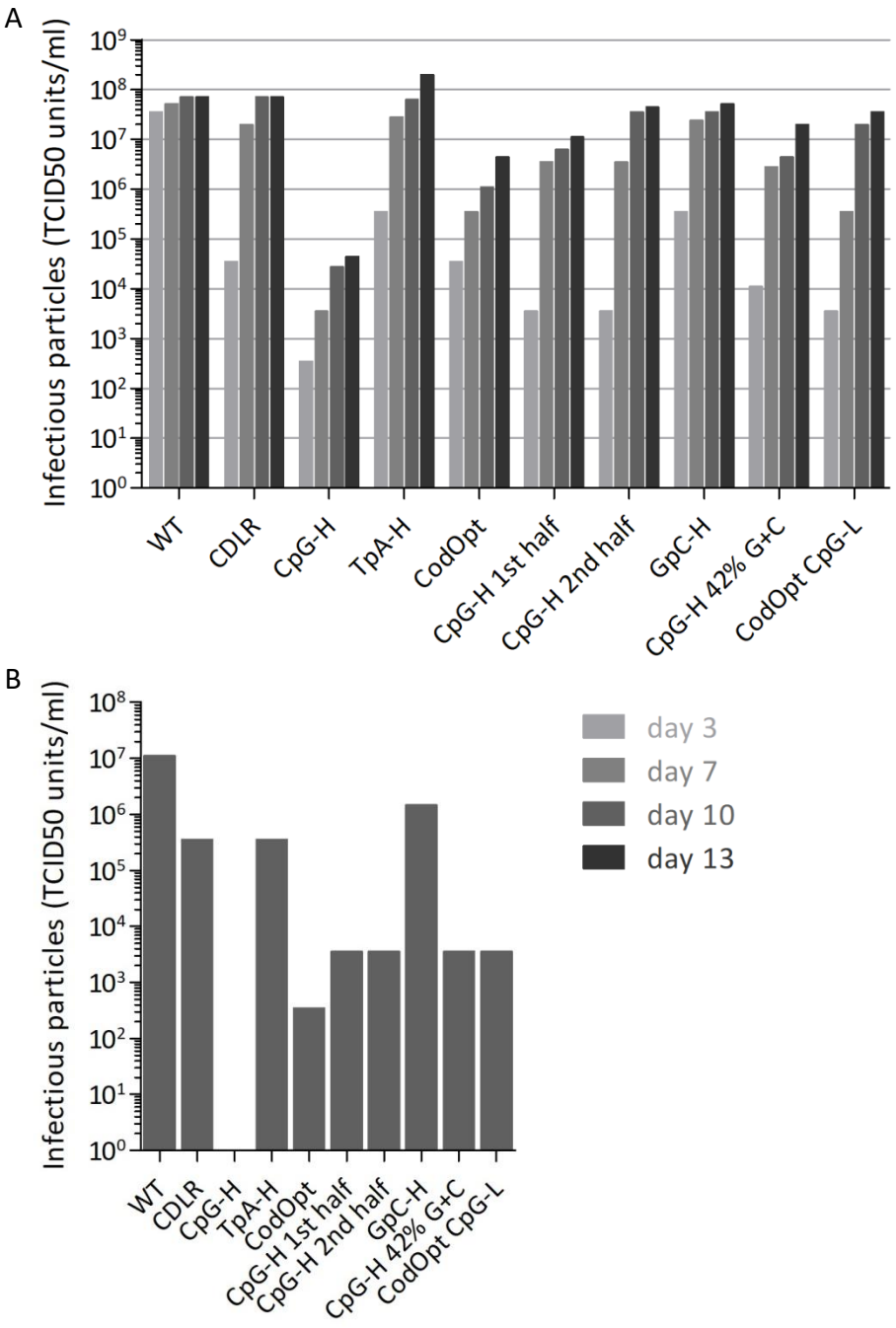


Figure 2.11 TCID50 scoring in NB324K and A9 cells
$TCID_{50}$ titres of MVM stocks. 8 wells per mutant and dilution. The same plate was read over the duration of the experiment. **A.** at 3-13 days post infection in NB324K cells and **B.** 10 days post infection in A9 cells

In NB324K cells, all viruses replicated to measurable titres (Fig 2.11A). Interestingly, all mutants scored lower than the WT on day four, with some of them catching up to WT levels by day ten while others remained lower throughout. While some viruses further increased in

titre between day 10 and 13, these were almost exclusively partial positives, while many partial positives from day 10 turned fully positive by day 13[1].

In contrast, in A9 cells (Fig. 2.11B), all mutants had lower infectivity titres throughout the scoring period, with several mutants being below or just above the detection threshold, making reliable quantification difficult. Interestingly, while the CpG-H mutant showed no CPE at all, there was still a visible difference between an (highly) infected and an uninfected well. In the infected well, cell density was lower, in a manner that seemed proportional to the amount of virus present. This could also be observed in some of the other low scoring mutants, but could not be used to reliably titre virus. Furthermore, there were no differences in scores between day ten and day 13, presumably because the cells were too overgrown at this point.

Ideally, the A9 cells would have been used for measuring infectivity, given that the cell line is a mouse cell line, rather than the human NB324K cells. However, given the inability to titre some of the mutants in A9 cells without using antibody staining, NB324K cells were chosen for infectivity titrations. The $TCID_{50}$ assay worked reliably with the different mutants in this cell line, yielding similar titres as the plaque assay. Day 10 was chosen for scoring, with a cut off for positive at "widespread CPE throughout the well" as these wells would progress to complete CPE by day 13. This was to avoid scoring them on day 13 in case virus-independent cytopathic effects or cell death would occur earlier for any reason.

[1] Please note that the data set shown in Fig. 2.11 was specifically created for this thesis towards the end of the project timeline to illustrate this progression for all 1st and 2nd generation mutants. The data shown in Fig. 2.11A are approximately one log lower on day 10 compared to the dataset established and used throughout the project (for both the DNA to infectivity ratios shown in the next section and all experimental work). As mentioned above, some variation was observed when titrating the same stocks in NB324K cells on different occasions. While it was important to account for these differences in the other sections, no adjustments were made for Fig. 2.11A, as the focus was on the relative changes over time.

Virus quantification – Viral genome copies

To complement the viral infectivity quantification, the genome copy numbers per ml of the virus stocks were determined using an amplicon in the NS region of MVM.

It was decided to use a region away from the mutated regions for three reasons. First, this would allow the use of a single primer pair for all mutants, avoiding the necessity for multiple standard curves. Second, it was unknown how genetically stable the generated mutants were when this assay was first performed. To avoid potential point mutations affecting primer binding, a region unlikely to be affected was chosen. Third, the chosen region had previously been established as suitable for qPCR, forgoing the need to establish suitable qPCR primers (460).

For viral genome quantification, 50 µl virus stock were treated with DNaseI prior to DNA purification. The samples were analysed via qPCR using the QuantiFast SYBR Green PCR Kit. Serially diluted BamHI –excised, gel-purified MVM DNA (as the exact length of pMM984 was not known) was used to make a standard curve for the samples.

Uninfected cell lysate prepared in the same way was used as a negative control. To account for DNA loss during DNA extraction, 50 µl of uninfected lysate was spiked with 10^9 copies of the standard and purified along the samples (without DNaseI treatment). The values obtained from the qPCR were then adjusted according to the loss rate.

All viruses had similar concentrations of viral DNA (1-9x10^{10} viral genomes / ml), indicating that differences in infectivity were due to viral attenuation rather than viral rescue.

DNA to infectivity ratios of the MVM mutants

After collecting data for both viral infectivity and genome copies, DNA to infectivity ratios for the different mutants were determined (Fig. 2.12).

In A9 cells, WT MVM had a lowest DNA to infectivity ratio of 7428, the lowest of all tested mutants. The CDLR and TpA-high mutants had DNA to infectivity ratios of 1.46 x 10^5 and 1.76 x 10^5 respectively, approximately 20 times higher than WT (Fig. 2.12A). In comparison in NB324K cells, DNA to infectivity ratios were similar for WT, CDLR and TpA-high with a ratio of 109, 57 and 66, respectively (Fig. 2.12B).

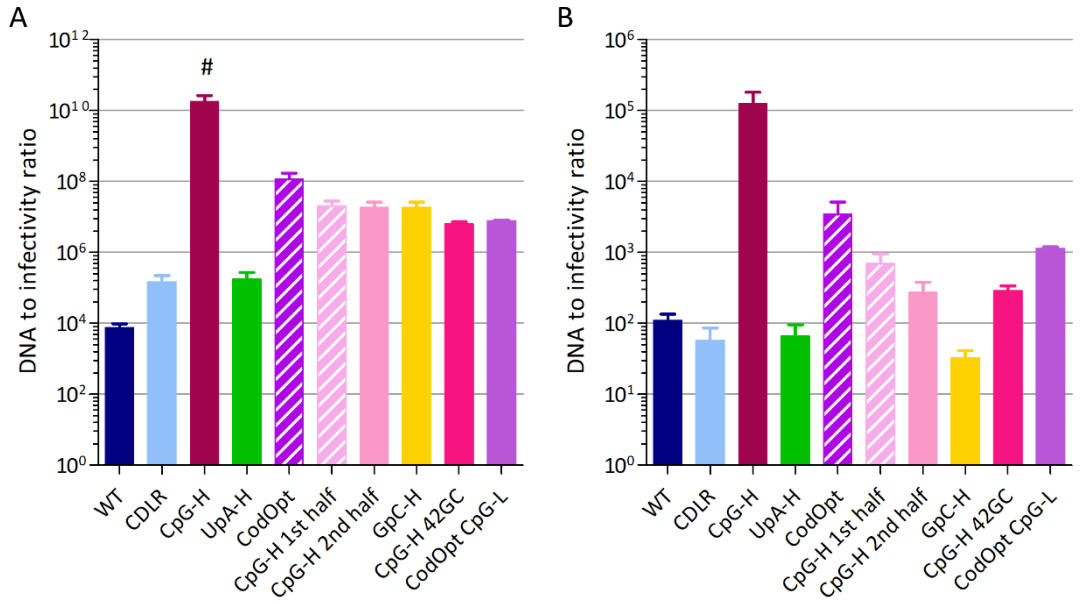


Figure 2.12 DNA to infectivity ratios of MVM in A9 and NB324K cells
DNA to infectivity ratios for the different MVM mutants. TCID50 titrations done once, with eight wells per mutant and dilution, qPCR done from three different DNA extractions per mutant. Error bars show SEM, # indicates that the $TCID_{50}$ score was below the measurable threshold, thus underestimating the DNA to infectivity ratio. **A.** A9 cells and **B.** NB324K cells

1st and 2nd generation CpG-high and codon-optimised mutants showed overall increased DNA to infectivity ratios in both cell lines, with bigger increases observed in A9 cells. In NB324K cells, the CpG-H mutant had a DNA to infectivity ratio of 1.24 x 10^5, a 3-log increase compared to WT virus, while the CpG-high half mutants and the CpG-high mutant with fixed G+C content had less than a 1-log increase in their ratios (CpG-H1: 690, CpG-H2: 271, CpG-H 42% G+C: 288). In A9 cells, the infectivity of the CpG-H mutant was below the detection limit of the $TCID_{50}$ assay used, giving it a DNA to infectivity ratio of at least 1.8 x 10^{10}, a 6-log increase

compared to WT MVM. The CpG-high half mutants and the CpG-high mutant with fixed G+C content showed an increase of approximately 3 logs compared to WT virus in A9 cells (CpG-H1: 2.03 x 10^7, CpG-H2: 1.87 x 10^7, CpG-H 42% G+C: 6.27 x 10^6).

For the codon-optimised mutants, the 1st generation codon-optimised mutant showed a DNA to infectivity ratio of 3446 in NB324K cells (1.5-log-increase compared to WT) while the CpG-low codon-optimised mutant had a DNA to infectivity ratio of 1136, a 1-log increase compared to WT virus in this cell line. In A9, the DNA to infectivity ratio for the 1st generation codon-optimised mutant was increased by 4 logs compared to WT (ratio of 1.17 x 10^8), while the CpG-low codon-optimised mutant displayed an increase of 3-logs compared to WT (DNA to infectivity ratio of 7.79 x 10^6).

For the GpC-high mutant, the DNA to infectivity ratio was 1.87 x 10^7 in A9 cells, an increase of 3-logs compared to WT. This was in contrast to NB324K cells, where the DNA to infectivity ratio was 33, 3-fold lower than the WT ratio.

These results indicate that there might be a difference in restriction between the two cell lines. It should be noted that there was some variation observed between assays, where the read outs of the entire set of viruses would increase or decrease while retaining the relative difference between different viruses. In NB324K cells, this variation spanned approximately 1.5 logs and numbers for new batches and mutants were normalised to the first set used for experiments (values ranging from 1 log lower to 0.5 logs higher of the original data set). In A9, fewer data points are available for this variation, as this cell line was only used in TCID50 assays to determine suitability and to produce Fig. 2.11B and Fig. 2.12B. There is thus a chance that the DNA to infectivity ratios are offset by a fixed ratio, especially in A9 cells where fewer data sets were acquired. Specifically, the 2-log difference between WT virus in NB324K and

A9 cells should be approached with caution. It has been shown before that the difference in infectivity of between A9 and NB324K cells is approximately one log (458) and in section 3.2 infecting A9 cells with 1000 viral genome copies/mL WT MVM resulted in approximately 40% infected cells.

2.5 MPyV mutant selection and generation

Region selection

The MPyV genome was first searched for suitable regions to introduce mutations, similar to the MVM parvovirus model. The MPyV BG strain was used for these experiments.

Viral early and late proteins do not share long stretches of DNA and can thus be manipulated and studied independently (Figure 2.13A). Similar to MVM, the genomic map of MPyV and the relative coding variability was used to determine suitable regions. The pairwise distance scan at synonymous sites were conducted using SSE(439) for the large T antigen and major capsid protein VP1. Due to their ORFs being encoded on opposite strands of the genome, these were carried out separately (F. 2.13B+C). For both genes relative coding variability was high (large T antigen: 0.71-0.92, VP1: 0.67-0.79) and two suitable regions were identified within the N-terminal regions of both the large T antigen (region T, bp 1920-2820) and VP1 (region V, bp 3893-3000) (Fig 2.13B+C). Both regions are approximately 0.9 kb long, but either can be shortened to approximately 0.6 kb as necessary (region Ts, bp 2236-2820, and Vs, bp 3611-3000).

To test the suitability of these regions, three scrambled mutants were made, CDLR T, CDLR V and CDLR TV (Table 2.3).

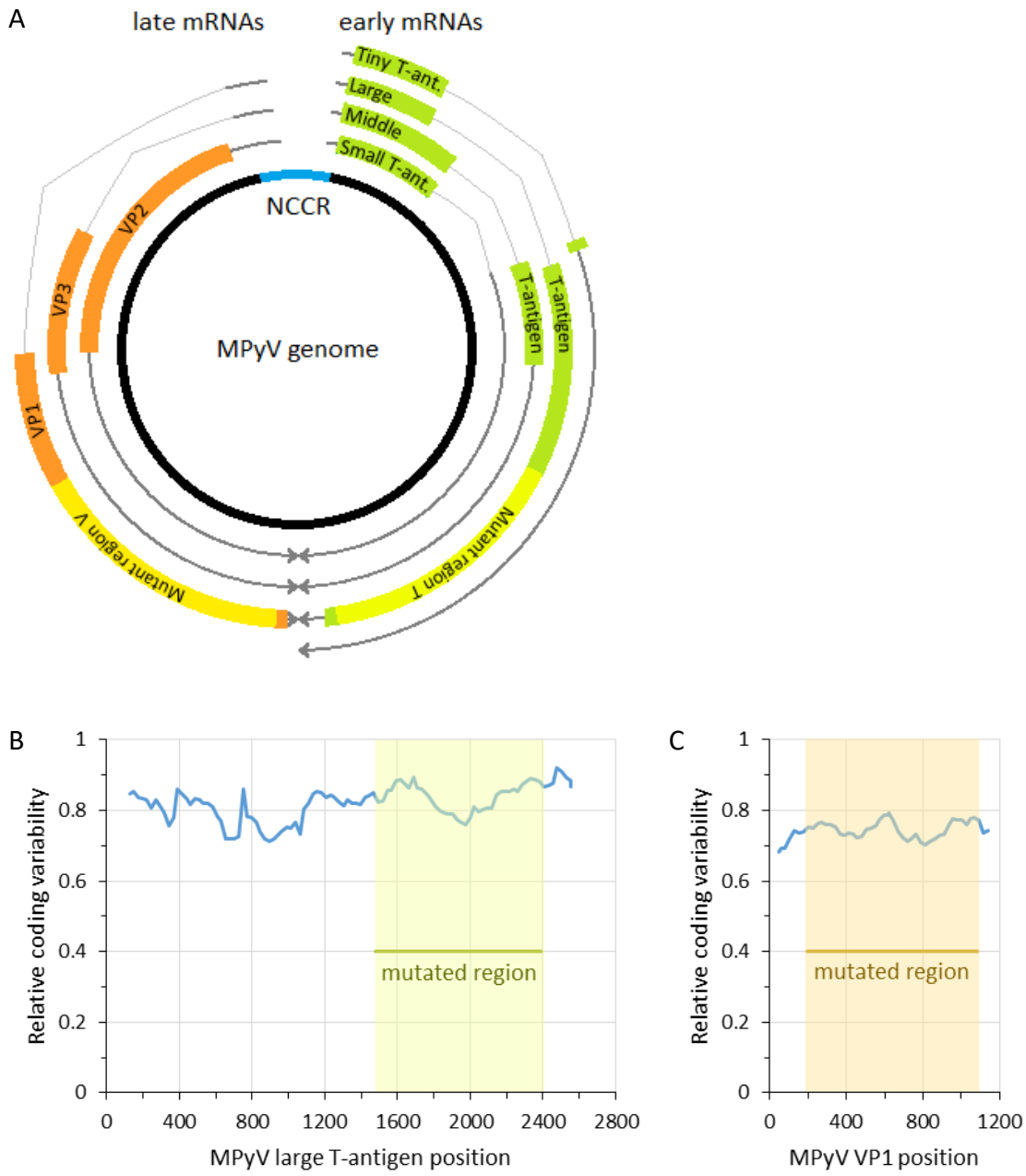


Figure 2.13 MPyV genomic structure and the regions selected for mutagenesis
A. Diagram of the MPyV genomic structure. Non-coding control region (NCCR, blue), transcripts (grey) with ORFs (green and orange coloured boxes), mutated region highlighted in yellow. Based on Fig.1 by Carmichael, 2016 (427). **B.+C.** SSE was used to calculate the synonymous coding variability observed between MPyV and related alphapolyomaviruses. The lime and orange region indicate the region chosen for mutations in the large T-antigen (B.) and VP1 (C.).

Table 2.3 Dinucleotide composition of MPyV mutant regions

Mutant	Fragment length	G+C content	CpG		TpA	
			Number	O/E ratio	Number	O/E ratio
WT (Region T)	900	44.33 %	11	0.249	47	0.679
CDLR (Region T)	900	44.33 %	11	0.249	47	0.679
WT (Region V)	888	47.76 %	6	0.119	37	0.647
CDLR (Region V)	888	47.76 %	6	0.119	37	0.647
WT (Regions T and V)	1788	46.00 %	17	0.169	84	0.667
CDLR (Regions T and V)	1788	46.00 %	17	0.169	84	0.667
Whole Genome	5380	47.16 %	91	0.308	300	0.810

Table 2.3 Dinucleotide composition of MPyV mutant regions
O/E: observed over expected

Cloning strategy

The cloning strategy described here can be used for either the pMJG vector (traditional reverse genetics system) or the pBS-loxPintron vector (Cre-LoxP reverse genetics system).

A synthetic DNA was ordered, ranging from bp 1920-4120, flanked by *NsiI* and *EcoRV* containing both insert regions. The entire synthetic sequence was cloned into the backbone to generate the TV mutant. In a second step the T region was excised from both wild-type and mutated TV clones using *KflI* (site 1 starting at bp 926 and site 2 starting at bp 2992) and ligated with the respective vectors to create both the T and V mutants. A PCR spanning vector and insert as well as Sanger sequencing were used to confirm the right orientation of the newly generated mutants.

It should be noted that inserts can be transferred between the pMJG and pBS-loxPintron clones, however the *KflI* restriction site is only suitable for cloning within the same backbone, as the bacterial backbone is inserted at different locations of the virus within the two plasmids, leading to different, incompatible fragments generated when cut with *KflI*.

2.6 MPyV virus recovery and quantification

Virus recovery

Jitka Forstová kindly gifted the lab necessary plasmids and cell lines for both the traditional and Cre-LoxP reverse genetics system. For the Cre-LoxP system, this included both parental 3T6 cells as well as stable, CRE expressing 3T6 cells, Called T2 cells. Given the ease of use, the Cre-LoxP system was used to recover virus.

A loxP-STOP-loxP-ZSgreen expression plasmid was used to determine both Cre expression and transfection efficiency of the 3T6 cell line. With this construct, only transfected cells with active Cre would express the green fluorescent ZSgreen protein. While there was low to medium intensity ZSgreen expression when transfected by itself, this could be enhanced by replacing 50% of the plasmid DNA with a Cre expression plasmid. This approach was then used to transfect WT and CDLR MPyV pBS-loxPintron into Cre expressing 3T6 cells when 50% confluent. The transfected cells were split the next day and cells were harvested after 3-5 days, freeze-thawed 3 times, filtered and aliquoted (p0). Fresh cells were then inoculated with the virus lysate to generate p1 stocks.

Virus quantification

Because 3T6 are fast growing cells, it was hard to distinguish between cytopathic effects and cells overgrowing. In the literature, plaque assays with primary cells such as MEFs or whole mouse embryo cells are recommended to use for quantifying infectivity in MPyV (446). However, the nature of the experiments used for this project (such as a time course with multiple mutants) generates large numbers of samples and requires large amounts of cells. For that purpose, having a suitable immortalised cell line for titrations would be more

suitable. Consequently, 3T6, NIH3T3 and immortalised MEFs were tested for a) their ability to show cytopathic effects or cell death upon MPyV infection and b) their suitability for a $TCID_{50}$ assay. 3T6 and NIH3T3 showed low to moderate cytopathic effects after infection, but there was no clear cut off, thus not providing a reliable readout for the $TCID_{50}$ assay. Unlike primary MEFs, the immortalised MEFs showed no cytopathic effects. However, given that they were immortalised using the SV40 large T antigen, it is very likely that this is interfering with an otherwise lytic infection.

Due to these circumstances, viral genome copies were measured using qPCR, using the same approach as for MVM (see section 2.4). MPyV p1 virus stocks had a concentration of $3\text{-}7x10^9$ copies/ml. Viral genome copies were similar for WT and the different CDLR mutants.

2.7 Using MPyV as a second model virus.

It was originally planned to replicate the findings from MVM in MPyV to evaluate whether the knowledge gained would be more broadly applicable as part of this project.

However, as the project progressed, it was decided that it would be more valuable to generate several further controls for the MVM system to ensure that the observed effects were not confounded by other factors. This was partly done in anticipation of what peer reviewers would be asking for, especially since creating and validating new mutants is a slow process.

While it was a strategic choice to postpone this work, it is nevertheless a potentially very valuable system since different potentially affected processes of the viral life cycle (transcription, translation and replication) can be addressed separately with the MPyV model.

3 Characterisation of MVM mutant phenotype

After rescuing MVM wild-type and mutant viruses, their replication kinetics were determined in two cell lines commonly used for MVM. As it became apparent that CpG-high and codon-optimised mutants were attenuated, different parts of the virus life cycle were analysed to narrow down which processes were impacted by the mutations. Immunofluorescence and western blot analysis were used to determine protein expression after infection and in a cell free *in vitro* translation system. Using reverse transcription real-time PCR, viral RNA levels were measured in the absence of DNA replication using both the infectious plasmid as well as a sub-genomic expression vector encompassing the R3 region of MVM. Lastly, genetic stability was determined in both virus stocks and after serial passage of the viruses.

3.1 Replicative fitness of MVM mutants 69
1st generation mutants 69
2nd generation mutants 72
3.2 Viral protein expression 74
Visualisation and quantification of viral infection using Immunofluorescence 74
Western blot analysis 79
In vitro translation efficiency 80
3.3 Viral RNA expression 83
3.4 Selection pressure against CpG dinucleotides 85
Genetic stability of the MVM virus genome 85
Serial passage and miseq sequencing 86

3.1 Replicative fitness of MVM mutants

After generating and recovering MVM and its mutants, growth kinetics of WT and mutant viruses were determined. For this, A9 or NB324K cells were infected with 1 $TCID_{50}$ unit/cell for one hour, washed three times with PBS and incubated with fresh complete DMEM. Supernatant was taken over a 48-hour time course and titrated in NB324K cells for infectious particles.

In both A9 and NB324K, the replication kinetics of WT and mutant viruses showed substantial differences (Fig 3.1A-C) with Figure 3.1D and 3.1E showing a cross section of the titres at 48 h for both 1st and 2nd generation mutants in both cell lines.

1st generation mutants

The WT virus showed a logarithmic growth after an initial lag phase to titres of 1.81 x 10^7 $TCID_{50}$ units/ml (A9) and 3.68 x 10^6 $TCID_{50}$ units/ml (NB324K) at 48 h post infection, an increase of 4.5-5 logs from the 1 h time point.

The CDLR control replicated similarly to the WT in both A9 and NB324K, reaching titres of 5.79 x 10^6 $TCID_{50}$ units/ml (A9) and 3.32 x 10^6 $TCID_{50}$ units/ml (NB324K) after 48 h. This finding indicates that scrambling the coding sequence did not disrupt any DNA/RNA structure, regulatory elements nor uncharacterised alternative ORFs. Thus, these can be ruled out as contributing factors for the attenuation and replication defects observed in other mutants as described below.

Modifying CpG frequencies profoundly affected replication of MVM. The titre of the CpG-high full mutant (containing 117 CpGs compared to 7 in the WT virus) increased only 5-fold over the 47 h window in the A9 cell line (to 339.4 $TCID_{50}$ units/ml), showing a 4 log attenuation of this mutant. In NB324K cells, the virus was also replication deficient, but could grow to 2.46 x

10^3 $TCID_{50}$ units/ml, a 1.5 log increase over the 1 h time point and a 3-log reduction compared to WT virus. The CpG-high half mutants showed a lesser degree of attenuation, consistent with their intermediary CpG dinucleotide frequencies (containing 58-59 CpGs). In A9 cells, the half mutants replicated to 2.39 x 10^4 $TCID_{50}$ units/ml (1^{st} half mutant) and 4.28 x 10^4 $TCID_{50}$ units/ml (2^{nd} half mutant), a 2.5 log reduction compared to WT. In NB324K cells, they grew to 8.06 x 10^4 $TCID_{50}$ units/ml (1^{st} half mutant) and 1.03 x 10^5 $TCID_{50}$ units/ml (2^{nd} half mutant), showing an attenuation by 2 logs compared to WT virus.

In contrast, the TpA-high mutant showed no signs of replication deficits in either cell line, growing to titres of 2.35 x 10^7 $TCID_{50}$ units/ml (A9) and 2.97 x 10^7 $TCID_{50}$ units/ml (NB324K) after 48 h.

Remarkably, the codon-optimised mutant was almost as attenuated in its replication as the CpG-high full mutant in A9 (2.14 x 10^3 $TCID_{50}$ units/ml, 4 log reduction compared to WT at 48 h post infection), despite the modified coding sequence VP region which would conventionally enhance translation. This attenuation was less severe in NB324K cells (3.04 x 10^5 $TCID_{50}$ units/ml, 1 log reduction compared to WT at 48 h post infection).

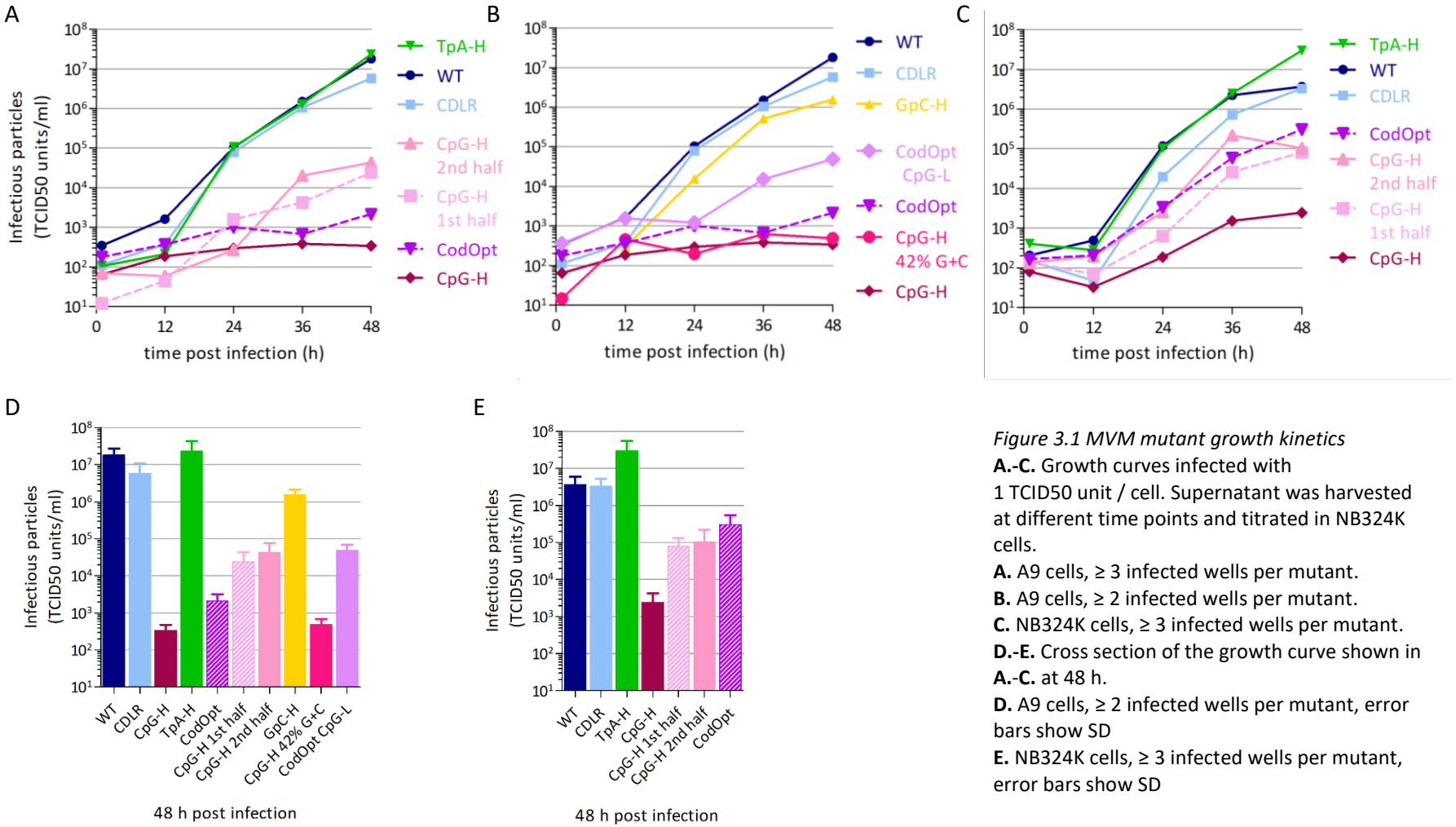


Figure 3.1 MVM mutant growth kinetics
A.-C. Growth curves infected with 1 TCID50 unit / cell. Supernatant was harvested at different time points and titrated in NB324K cells.
A. A9 cells, ≥ 3 infected wells per mutant.
B. A9 cells, ≥ 2 infected wells per mutant.
C. NB324K cells, ≥ 3 infected wells per mutant.
D.-E. Cross section of the growth curve shown in **A.-C.** at 48 h.
D. A9 cells, ≥ 2 infected wells per mutant, error bars show SD
E. NB324K cells, ≥ 3 infected wells per mutant, error bars show SD

2nd generation mutants

When the codon-optimised mutant was generated, two obvious changes were introduced. Compared to WT, more CpGs were present in the mutated region (from 7 CpGs in the WT to 70 CpGs in the codon-optimised mutant, raising the observed over expected ratio from 0.13 to 0.61) as well as an increase in G+C content (from 42 % to 61 %). Furthermore, the original CpG-high mutant had an increased G+C (from 42 % to 48 %) content due to software limitations at the time of creation.

To separate the effects of dinucleotide frequency alterations and G+C content changes, two new mutants were created. A CpG-high mutant with fixed (wild-type level) G+C content and a codon-optimised mutant was generated without an increase in CpG dinucleotide frequency (with a G+C content of 57 %). Ideally, a codon-optimised mutant with 42 % G+C content would have also been created, but that was not technically possible.

The CpG-high 42 % G+C mutant (containing 78 CpGs) replicated poorly (to 488 $TCID_{50}$ units/ml after 48 h in A9, not determined in NB324K), showing a 4-log attenuation in A9 cells and a similar level of attenuation as the original CpG-high mutant.

Surprisingly, the CpG-low codon-optimised mutant was also attenuated, reaching 4.88 x 10^4 $TCID_{50}$ units/ml at 48 h in A9 (2.5-log reduction compared to WT, not determined in NB324K).

The results from these two mutants suggests that while an increase in G+C content can have a detrimental effect on viral fitness, it does not appear to be the main driving force behind the severe attenuation the CpG-high mutants.

A third new mutant, a GpC-high mutant, was created to account for an increase in self-complimentary sequences and thus structural changes in the RNA of the CpG-high

mutants. The GpC-high mutant showed a slight impairment in replication, increasing to 1.54 x 10^6 $TCID_{50}$ units/ml at 48 h in A9 (a 1-log reduction compared to WT, not determined in NB324K). This suggests that the increase in self-complementary sequences may be contributing a small part of the attenuation seen in the CpG-high mutants, but does not account for the attenuation observed in these mutants.

Overall, these findings indicated that there is a dose dependent attenuation and replication deficiency when increasing CpG dinucleotide frequencies. Furthermore, increasing G+C content independently from CpG frequencies also resulted in attenuated viruses.

Going forward, this study focusses on the former, as the latter was discovered towards the end of the project timeline and taken over by the superseding PhD student, Jeremy Ratcliff.

3.2 Viral protein expression

Visualisation and quantification of viral infection using Immunofluorescence

After analysing the growth kinetics of wild-type and mutant MVM, immunofluorescence was used to visualise the infection of the various mutants and examine whether protein expression was influenced by dinucleotide frequency modifications.

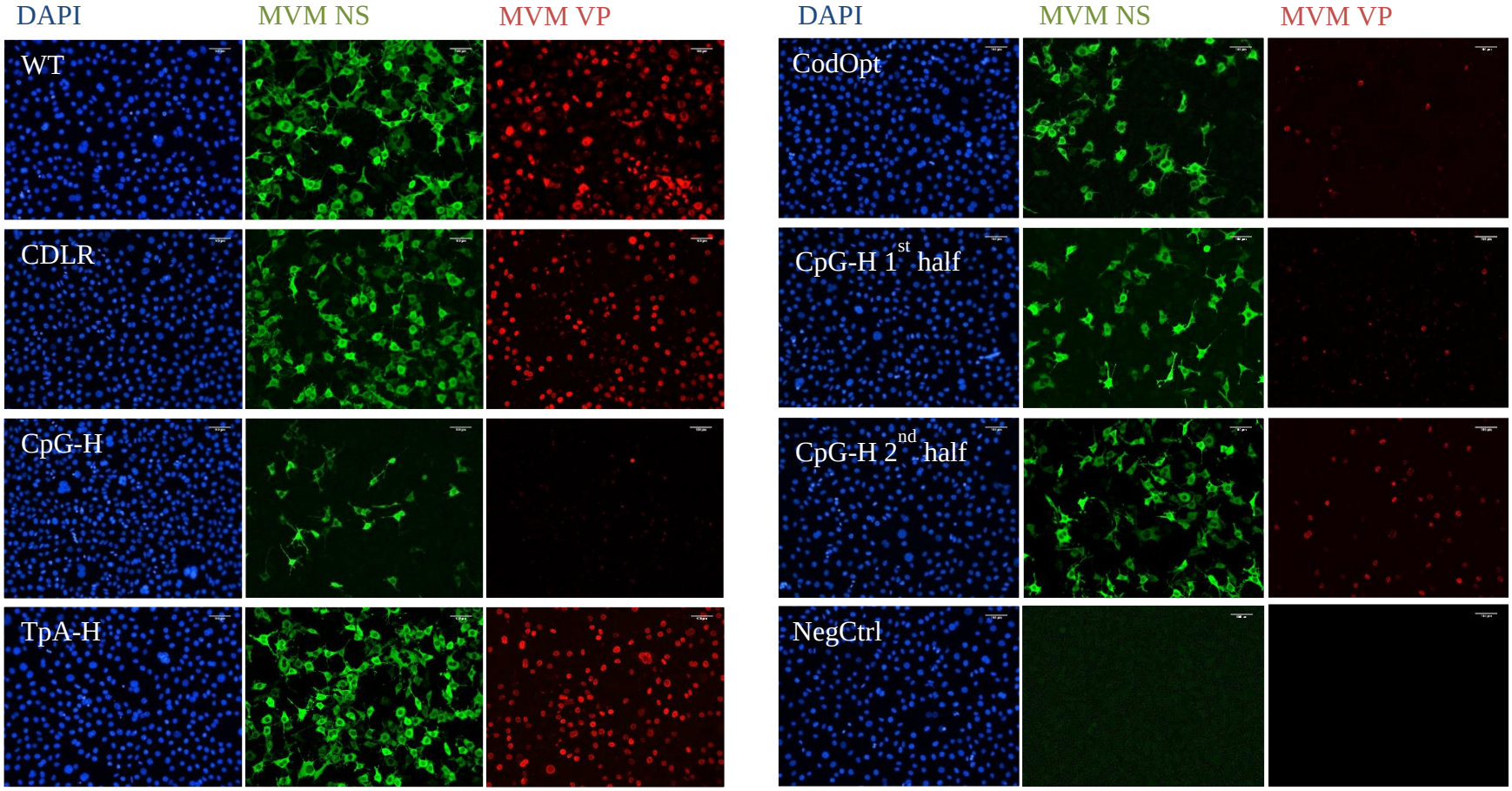
For this, cells were infected with either wild-type or mutant (1st generation) MVM, fixed with paraformaldehyde 18 hours post infection and stained with specific antibodies for viral NS and VP proteins (panel in Fig. 3.2A, close ups of merged images in Fig 3.2B, *image processing done by Niluka Goonawardane*). Cells were infected with equal copy numbers of the viral genomes to ensure that cells were exposed to equal amounts of virus particles.

Cells were infected with 1000 genome copies per cell, as this resulted in a robust infection and expression of NS and VP proteins for the wild-type virus in 30-40 % of cells, leaving room in either direction to measure differences in both infected cell numbers and relative fluorescence intensity caused by introducing mutations.

The CDLR mutant showed a slight reduction in the number of cells expressing the viral proteins as well as their intensity, while the TpA-high mutant showed similar levels to wild-type virus in both frequency and intensity of viral protein expression.

In contrast, the CpG-high full and half mutants as well as the codon-optimised mutant could establish measurable protein expression in far fewer cells and to lower levels. This effect was particularly pronounced in the capsid proteins, which contained the mutated region in their open reading frame.

A


DAPI
MVM NS
MVM VP
WT
CDLR
CpG-H
TpA-H
DAPI
MVM NS
MVM VP
CodOpt
CpG-H 1st half
CpG-H 2nd half
NegCtrl

B

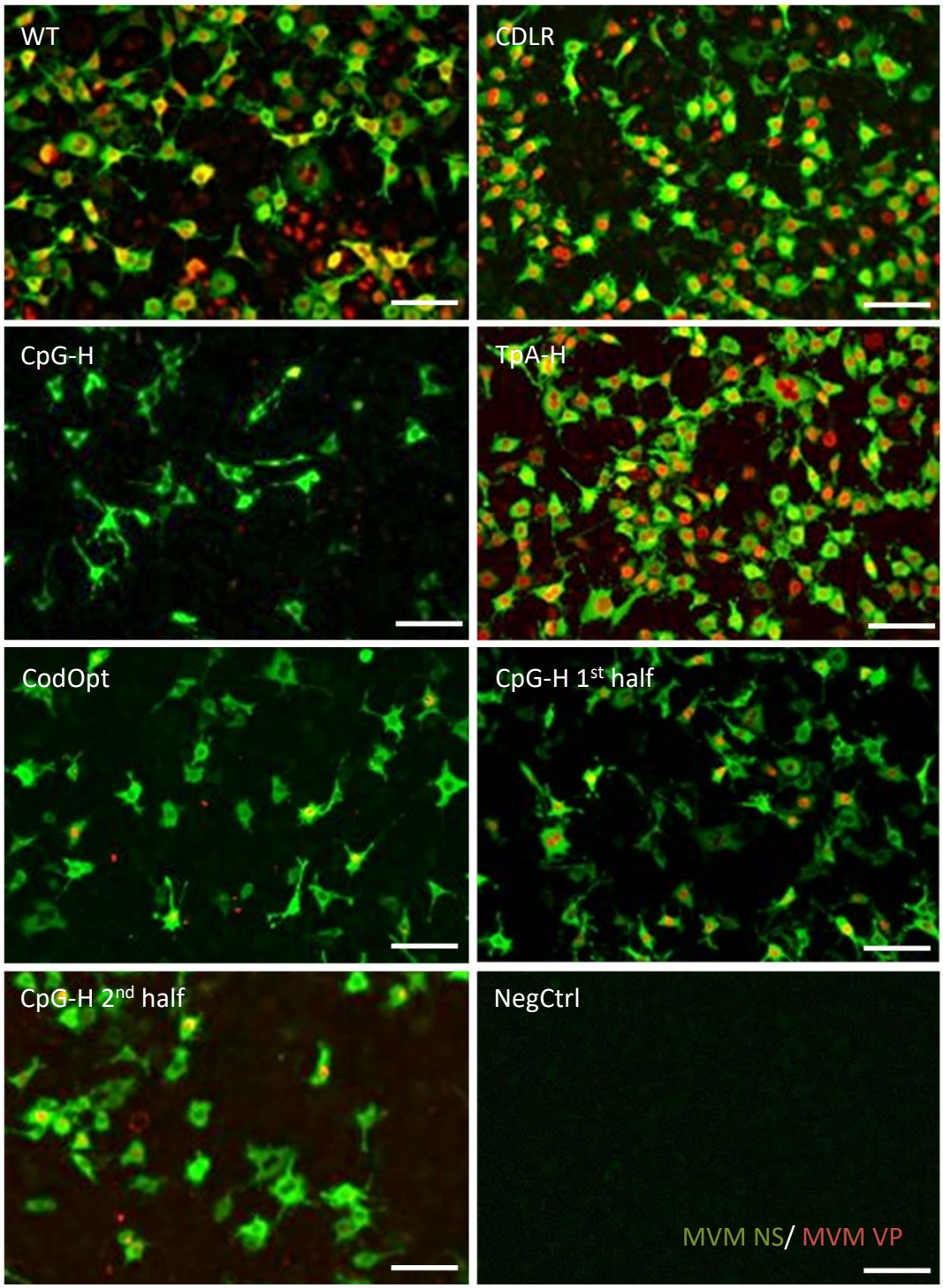


Figure 3.2 Immunofluorescence staining of viral NS and capsid proteins.
Immunofluorescence staining of DNA (blue), viral NS (green) and capsid (red) proteins at 18 h post infection with 1000 viral genome copies/cell. Uninfected cells were used as a negative control. Three wells were infected and stained per mutant, representative images shown. The scale bar measures 100 µm. *Image processing by Niluka Goonawardane.* **A.** Original panel **B.** Overlay of MVM NS and VP from A.

To quantify these differences, the number of fluorescent foci and their intensity was determined in representative areas of each sample (Fig 3.3, *image processing and quantification done by Niluka Goonawardane*). A minimum of 1806 cells were counted for each sample per triplicate.

When infected with MVM WT or the TpA-high mutant, approximately 40 % of cells expressed NS and and 37 % expressed VP proteins. When infected with CDLR, 33 % of cells were positive for NS and 22 % were positive for VP. While the TpA-high mutant expressed the viral proteins to the same intensity as WT virus, the intensity of both NS and VP was reduced by approximately 10% in CDLR infected cells.

In the CpG-H mutant, only 7 % of cells expressed NS (5-fold reduction) and 2 % expressed VP (23-fold reduction) while the reduction in fluorescence intensity was reduced by approximately 70 % for both NS and VP. In the CpG-high half mutants, approximately 20% of cells expressed NS and capsid proteins (2-fold reduction) when infected with the 2nd half mutant, while 12 % cells infected with the 1st half mutant expressed NS (3-fold reduction) and 3 % expressed VP (12-fold reduction). This variation between the two half mutants was also observed in the fluorescence intensity of the viral proteins. The relative fluorescence was reduced by 30 % for the 2nd half mutant and reduced by 60% for the 1st half mutant.

When infected with the codon optimised mutant, 13 % of cells expressed NS (3-fold reduction), while 5% expressed VP (7-fold reduction). The relative intensity of both NS and capsid proteins was reduced 40 % compared to cells infected with WT virus.

Only minor differences could be observed between wild-type, CDLR and TpA-high viruses in both frequency and intensity of cells expressing viral proteins. In contrast, this analysis

showed a profound reduction in viral protein expression both in frequency and intensity in CpG-high and codon-optimised mutants. Interestingly, while there was a greater reduction of cells expressing VP compared to those expressing the NS proteins, fluorescent intensity was similarly reduced for both NS and VP.

Due to time constraints, this experiment was not repeated when the 2nd generation mutants became available, but it would be interesting to study whether G+C content mediated attenuation shows similar or different effects in regards to protein expression.

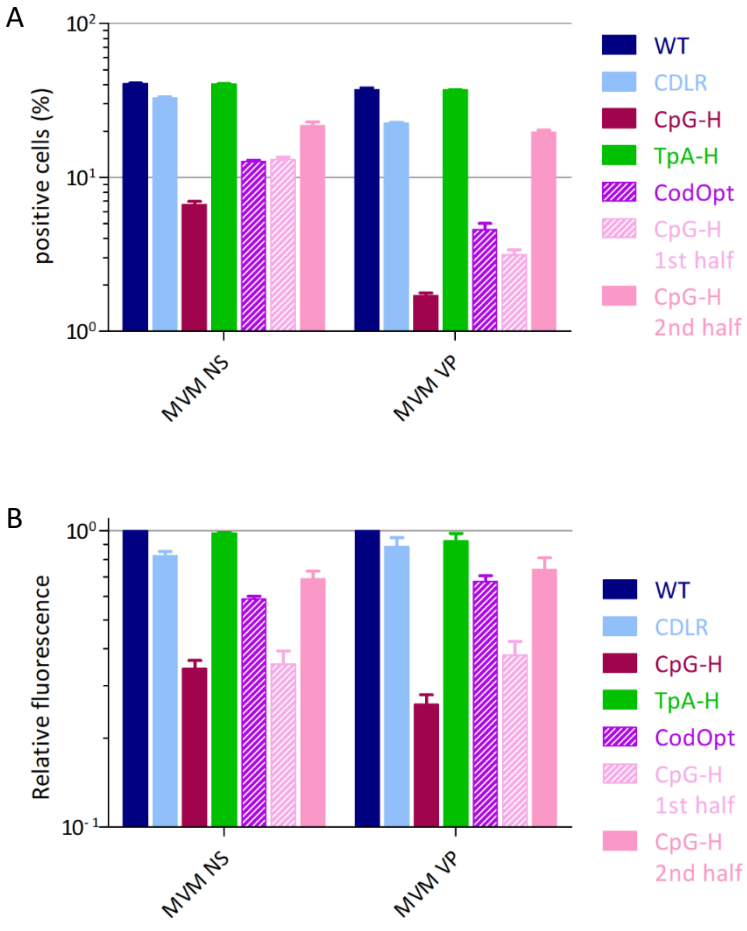


Figure 3.3 Quantification of immunofluorescence staining of viral NS and capsid proteins
At 18 h post infection with 1000 viral copies/cell. Uninfected cells were used as a negative control. Three fields of view from at least two different wells were analysed per mutant using Imaris and Fiji, error bars ± SEM. *Image processing and quantification done by Niluka Goonawardane.* **A.** Percentage of cells expressing NS or VP. **B.** Relative fluorescence intensity of NS or VP compared to WT MVM.

Western blot analysis

To separately verify the results from the immunofluorescence imaging with another method detecting viral proteins, western blotting was used to determine viral protein expression 18 h post infection, using 1000 viral genome copies per cell as before (Fig. 3.4). This assay further allowed distinguishing between the four viral proteins present in the cell (NS1, NS2, VP1 and VP2). This was not possible in the immunofluorescence imaging due to the antibodies targeting common regions of either the two NS proteins or the three VP proteins.

In accordance with the immunofluorescence staining, CDLR and TpA-high mutants showed no or only minor reductions in their protein expression compared to wild type, contrasting with much lower protein levels observed for the CpG-high and codon-optimised mutants.

Similar to the immunofluorescence assay, this part of the study was not repeated for the 2nd generation mutants due to their staggered implementation late in the project timeline.

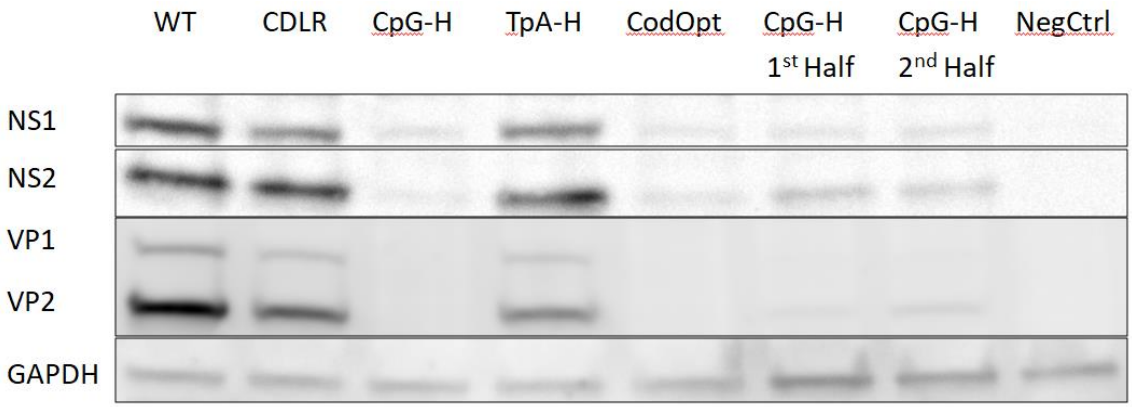


Figure 3.4 MVM protein expression post infection
Viral protein expression 18 h post infection with 1000 viral copies/cell. Uninfected cells were used as a negative control. Three wells were infected per mutant, representative image shown.

In vitro translation efficiency

To explore whether compositionally modifying the capsid region of MVM had any translational effects that might contribute to the observed differences in viral VP protein expression, translational rates were determined for wild-type and mutant viruses using a rabbit reticulocyte translation assay which allows extra-cellular *in vitro* translation of an RNA of choice.

For the experiment, the synthetic plasmid with the partial VP sequence used for the reverse engineering of the viral mutants (spanning positions 2647-4347 of the MVM genome) was used as it contained a conveniently placed T7 promoter with no splice sites upstream of the VP2 start codon, giving rise to a truncated VP2 with a size of roughly 60 kDa. Ideally, this experiment would have been carried out using radioactive labelling of the protein with ^{35}S-Methionine. However due to the hazard risk associated with radioactivity, it was simply not allowed to do experiments including radioactive substances. Instead, the VP antibody from the western blot above was tested for sensitivity and specificity for this particular assay. While there was some cross reactivity to the rabbit reticulocyte lysate at <25kDa, the antibody still bound to the truncated VP2 protein and was hence found suitable for the assay.

To assess translation efficiency, the synthetic plasmid was linearised using *ApaL*I and capped RNA transcripts were synthesized using the mMESSAGE mMACHINE T7 Transcription Kit and purified. RNA quality was assessed using the Bioanalyzer (Fig. 3.5A). Due to slight differences in RNA length observed on these gels, several repeats and technical duplicates were run and it was determined that the observed differences were likely due to problems with inconsistent denaturation, despite following the manufacturer's protocol. However, it should be noted that the RNA quality of the CDLR

and TpA-high constructs was consistently worse, despite trying different kits and approaches for DNA template preparation and working carefully to avoid contamination with RNases.

1 µg of each RNA transcript was then used with the rabbit reticulocyte translation assay and VP2 levels were determined after 1.5 hours using western blotting (Fig. 3.5B+C).

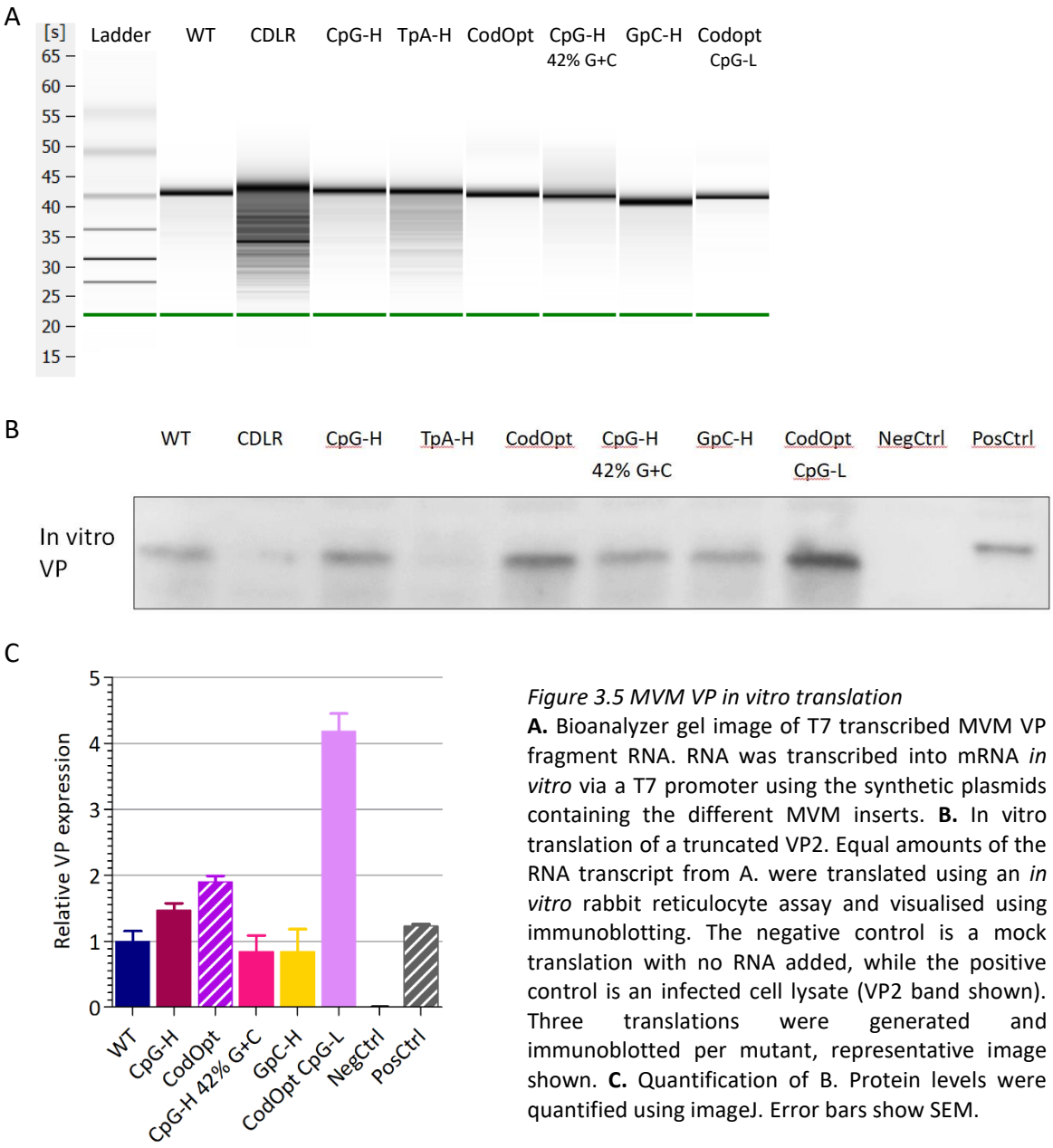


Figure 3.5 MVM VP in vitro translation
A. Bioanalyzer gel image of T7 transcribed MVM VP fragment RNA. RNA was transcribed into mRNA *in vitro* via a T7 promoter using the synthetic plasmids containing the different MVM inserts. **B.** In vitro translation of a truncated VP2. Equal amounts of the RNA transcript from A. were translated using an *in vitro* rabbit reticulocyte assay and visualised using immunoblotting. The negative control is a mock translation with no RNA added, while the positive control is an infected cell lysate (VP2 band shown). Three translations were generated and immunoblotted per mutant, representative image shown. **C.** Quantification of B. Protein levels were quantified using imageJ. Error bars show SEM.

In contrast to the infection setting, CpG-high and codon-optimised capsid sequences showed a similar translation efficiency as the wild-type sequence. This was also the case for the three 2nd generation mutants (GpC-high, CpG-high with 42 G+C content and CpG-low codon-optimised), which expressed similarly to wild type as well. In contrast, CDLR and TpA-high mutants barely translated. However, given their poorer RNA quality and robust expression in the infection context, it is very likely that this is a technical problem rather than an actual result, especially in a delicate system like this.

Given the similar expression levels in this extra-cellular translation assay, there was no evidence that the attenuated phenotype of the CpG-high and codon-optimised mutants was caused by differences in translation efficiency of the capsid protein.

3.3 Viral RNA expression

In the previous section it became apparent that CpG-high and codon-optimised viruses were restricted in both their NS and capsid protein expression while translation efficiency itself was not affected. Based on that observation, it was hypothesised that mRNA sequences with increased CpG content may be unstable, more readily subjected to degradation or otherwise targeted, offering an explanation why NS protein expression was affected in these mutants.

To investigate this question, viral mRNA levels were determined after transfection with plasmid DNA. To avoid compounding effects due to replication, cells were subjected to the replication inhibitor Aphidicolin before and during the transfection when the full-length viral clone was used. Due to the genomic organisation of MVM, it is possible to determine RNA levels of R1 or R2 individually or all viral RNAs, but not R3 individually, as R1 and R2 have unique splice sites, while the sequence encoding R3 completely overlaps with R1 and R2 (Fig. 3.6A). For this reason, R1 mRNA levels were determined using the full-length viral plasmid (pMM984), while R3 mRNA levels were separately determined using an expression plasmid made for this purpose. A commercial expression plasmid containing a chicken actin gene promoter and enhancer cassette was used to insert the part of the MVM genome corresponding to the pre-spliced R3 RNA (pCAG_MVM_R3).

Using the full-length plasmid, viral R1 mRNA was quantified using qPCR and gene specific primers after 8 h and 18 h (Fig. 3.6B). At the early 8 h time point, viral RNA levels were

still similar in all tested viruses. However, at 18 h post transfection, a reduction of R1 mRNA can be observed in the CpG-high mutant compared to the WT and CDLR mutant.

A similar trend was apparent in the R3 expression system, though the difference in RNA levels between wild-type virus and the CpG-high mutant was not as pronounced with a 3-fold reduction of CpG-high RNA in the R3 expression system versus a 10-fold RNA reduction in the full-length plasmid (Fig. 3.6C).

These findings suggest that there is indeed a restriction at the RNA level, independent of whether the mutated region is within the translated or non-translated region of the RNA.

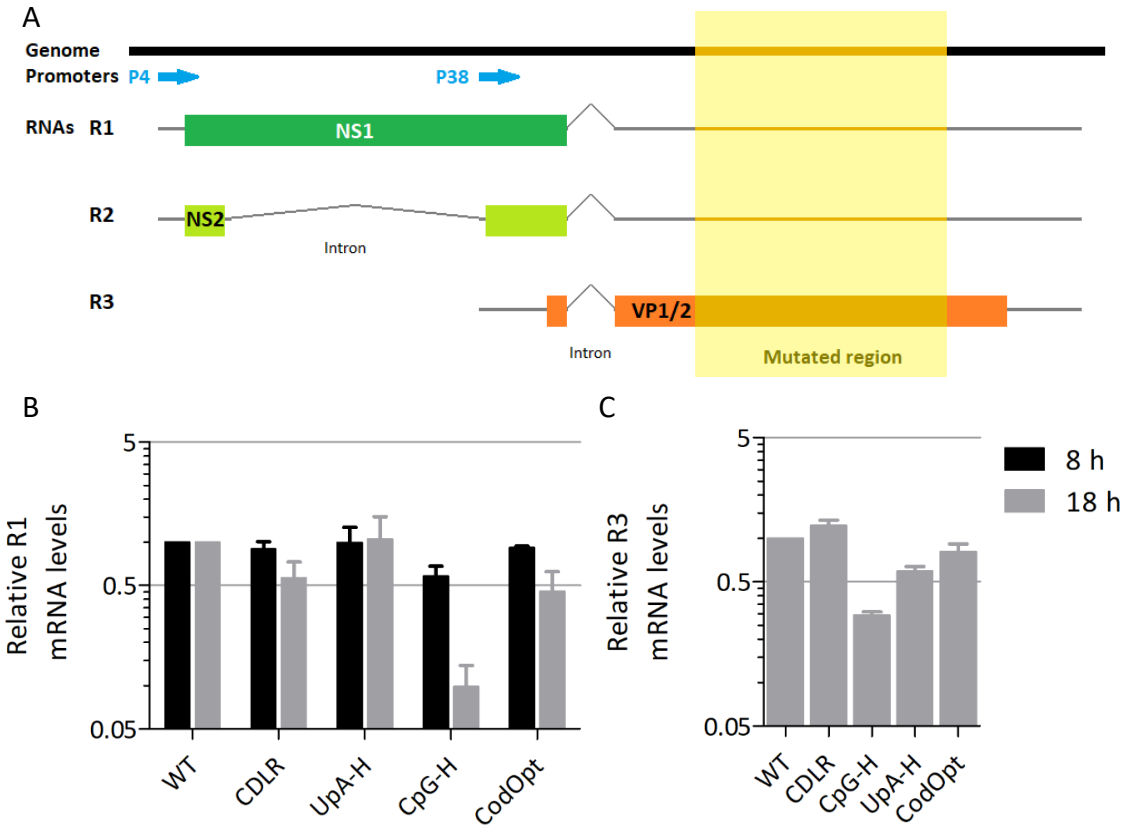


Figure 3.6 MVM mRNA levels post transfection
MVM R1 mRNA expression 8 h and 18 h post transfection normalised to cellular GAPDH and WT R1 or R3 levels. Three wells were transfected per mutant and condition, error bars show SEM.
A. Diagram of the MVM genomic structure. Promoters (blue), transcripts (R1-R3) with ORFs (green and orange coloured boxes), mutated region highlighted in yellow. Based on Berns & Parrish, Fig 57.10 (392). **B.** Plasmid pMM984, viral infectious clone, replication inhibited by Aphidicolin. **C.** Plasmid pCAG_MVMp_R3, expression vector containing the MVM R3 sub-clone.

3.4 Selection pressure against CpG dinucleotides

Genetic stability of the MVM virus genome

Given the severe repression of CpGs in the wild-type virus and the correlating severe attenuation of the CpG-high mutants, it was hypothesised that there is a selection pressure against CpGs within this virus. Consequently, it may be possible for the CpG-high viruses to (attempt to) revert to a lower number of CpGs in their genome as these reverting mutants would have a fitness advantage by removing CpGs from genomes. However, it was unclear over what period this might happen.

To investigate this, the virus was sequenced at different stages of passaging. First, the virus was rescued from the infectious plasmid using transfection and harvested in NB324K cells (virus passage 0). The virus was then propagated once more in NB324K cells (virus passage 1). These p1 virus stocks were used for the experiments throughout the project. In order to ensure no point mutations were present, the insert region of the plasmids and p1 virus DNA of wild-type MVM and all mutants were sequenced to determine the predominant clone present.

For the CpG-H virus (and wild-type virus, for comparison), an approximately 800 base fragment within each of the NS and VP region was amplified using PCR and ligated into a pJET vector. 20 clones of each region were picked for wild-type virus and the CpG-H mutant and analysed via Sanger sequencing. No reversion of CpGs were found in either sample. For both viruses, one clone each was found with a single mutation in the NS region (WT: base 609, C to T, CpG-H: base 558 G to T) while no mutations were found in the VP region. Furthermore, no deletions were found within these clones.

These results suggested that there was no reversion of CpGs at passage 1, as roughly 1300 CpGs (20x 66CpGs) were sequenced within the CpG-H VP fragment, but not a single reversion was seen. There was also no evidence of deletions occurring surrounding the increase in CpGs.

Serial passage and miseq sequencing

In a next step to analyse viral genomic stability, 1st generation viruses were serially passaged in A9 cells. For this, infectious supernatant was transferred to a new flask when full cytopathic effect was reached. The volume of transferred supernatant was adjusted for every mutant to allow full cytopathic effects within 3-4 days (WT: 1:50, TpA-H: 1:25, CpG-H1: 1:2, CpG-H2: 1:2). At passage 11, viral DNA was extracted from the supernatant. The viral DNA was processed into a genetic library and analysed using MiSeq (*data processing done by Dung Nguyen and Peter Simmonds*).

This protocol was attempted for the 1st generation CpG-high full mutant and the codon-optimised viruses, but unfortunately, these did not progress through cytopathic effects as the other mutants did, but rather appeared to arrest cell division and stay unchanged for several weeks before slowly progressing to show more cytopathic effects. It was attempted to passage these viruses over several months but it was not possible to raise the viral titres high enough to sequence them successfully.

Hence, for this experiment, the TpA-high and the CpG-high half mutants were used and compared to the wild-type virus. Passage 11 virus DNA was further compared to each plasmid DNA, to exclude false positives due to mutations inherent in the plasmids.

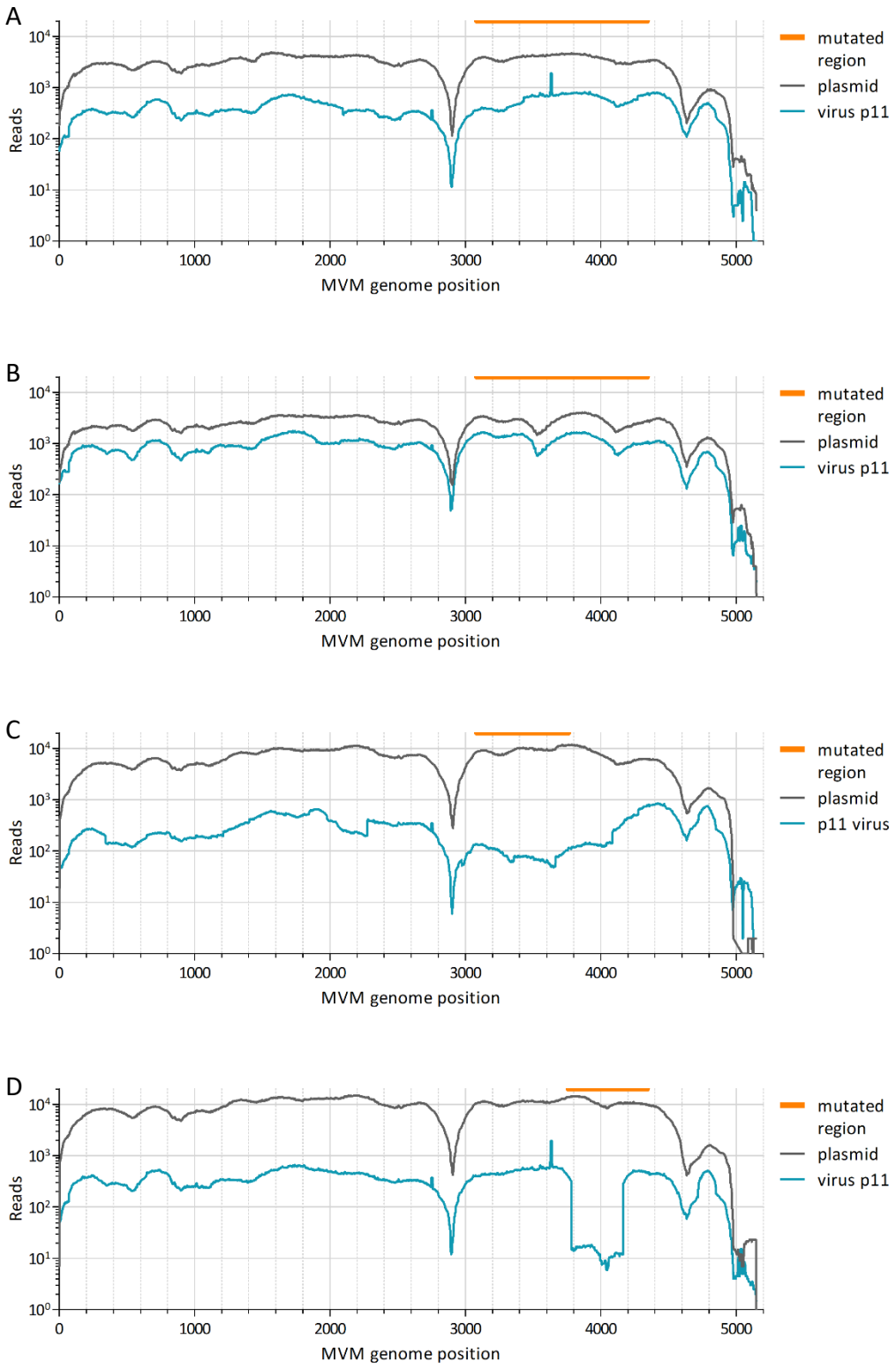


Figure 3.7 Read coverage of MVM next generation sequencing
MiSeq coverage for each base of the MVM genome of the infectious clone (plasmid pMM984) and of virus isolated after 10 passages (p11) in A9 cells. *Data processing done by Dung Nguyen and Peter Simmonds.* **A**. WT **B**. TpA-H **C**. CpG-H 1st half **D.** CpG-H 2nd half

Coverage was not uniform across the genome or between mutants (Fig. 3.7, *data processing done by Dung Nguyen and Peter Simmonds*). The 5'- and 3'-ends of the genome were consistently covered less, as was a G+C rich region around base 2900. Interestingly, there was dip in coverage in the CpG high area of the CpG-high 2nd half mutant and to a lesser extent in the 1st half mutant while the respective plasmid sequences showed consistent coverage over the CpG-high areas (Fig. 3.7C+D).

This phenomenon has been recognised before as an incomplete denaturation during DNA amplification steps (461). The protocol used for this experiment contained a PCR step to amplify DNA, the likely cause for the lower yield of reads in the CpG-high 2nd half mutant, however it is peculiar that is appears to only affect the CpG-high viruses, but not the corresponding plasmid sequences.

In parallel, the MiSeq reads for the CpG-high half mutants were checked for DNA recombination and intra-read gaps by Dung Nguyen, so see whether there was any evidence for large deletion mutations. No recombination events or intra-read gaps were found.

For future experiments, it would be recommended to use an alternative protocol or different kit to avoid these technical limitations. This would be particularly important if CpG sites were to be looked at specifically (such as for methylation studies).

Overall, all viruses were remarkably stable during cell culture passaging. All sequence changes that occurred with a frequency of 5% or more are summarised in Table 3.1. The wild-type and TpA-high mutant viruses showed single, partial changes in the capsid region, but neither of those had any effect on CpG or TpA frequencies (WT: position 3629, T to A mutation; TpA-H: position 3754, G to A mutation). The CpG-high 1st half mutant

acquired no reversions or other mutations within the mutated region, but two different minor populations of alterations were found within the non-mutated NS-region (position 1953, G to A mutation and position 1977, A to G mutation). In contrast, in the CpG-high 2nd half mutant, two CpG sites partially reverted in the mutated region (position 4072 and 4111, both G to A mutation). Four further mutations were found in the 2nd half mutant, a partial TpA reversion occurred within the NS region (position 1979, A to G mutation) as well as three mutations not related to CpG or TpA in the capsid (position 3177, G to A mutation; position 3629, T to A mutation; position 4048, C to T mutation). Consequently, the 2nd half mutant appeared to be more unstable than then 1st half mutant.

Table 3.1. MVM sequence changes after serial passage

Mutant	Position	Mutation	Mutation frequency
WT	3629	C**T**C → C**A**C	11.89 %
TpA-H	3754	G**G**G → G**A**G	18.41 %
CpG-H1	1953	G**G**C → G**A**C	6.76 %
	1977	T**A**T → T**G**T	8.31 %
CpG-H2	1979	T**A**A → T**G**A	28.42 %
	3177	G**G**C → G**A**C	92.51 %
	3629	C**T**C → C**A**C	15.90 %
	4048	G**C**T → G**T**T	8.33 %
	4072	C**G**G → C**A**G	9.52 %
	4111	C**G**A → C**A**A	57.69 %

Table 3.1 MVM sequence changes after serial passage
All sites with >5 % divergence from the plasmid sequence are shown.

Overall, these results indicate that the virus cannot easily counteract the negative effects conveyed by increasing CpG frequencies with escape mutants. This is in accordance with the attenuation of the CpG-high mutants observed in viral growth curves (Fig. 3.1) as well as the immunofluorescence imaging and quantification (Fig. 3.2-3.3). It appears the accumulation of many CpGs conveys attenuation. Thus, reverting any one CpG would likely yield a minimal increase in fitness and thus these reverted mutants would not easily outcompete the non-reverted viruses. This would be especially true if the CpG reversion came with a price in fitness attached to it (e.g. a non-synonymous mutation).

4 Mechanisms of CpG-mediated attenuation of MVM replication

After characterising CpG mediated attenuation at various steps of the viral life cycle, panels of knock out cell lines and a TLR9 inhibitor were used to investigate antiviral candidate proteins potentially involved in the restriction of CpGs in the small DNA virus context.

Within these cell lines, MVM mutants were tested for any changes in attenuation measuring viral output, as well as RNA and protein expression as appropriate.

Where available, MEF knockout cell lines were used, however, for ZAP and RNAseL, two proteins recently shown to recognise CpG-rich RNA, a human A549 cell system was used due to availability.

4.1 Screening for potential recognition molecules in murine knock out cells 92

Susceptibility to WT MVM 92

Attenuation of CpG-high MVM mutants 94

4.2 Involvement of TLR9 97

TLR9 mediated signalling in A9 cells 97

Effects of TLR9 inhibition on MVM infection 98

4.3 Involvement of ZAP and RNaseL 101

Susceptibility of A549 cells to MVM 101

Limitations of the A549 ZAP and RNAseL knockout system 103

RNA expression levels in ZAP and RNaseL knockout cells 104

Visualisation and quantification of viral infection using Immunofluorescence 107

4.1 Screening for potential recognition molecules in murine knock out cells

In chapter 3 it was observed that CpG-high and codon-optimised mutants of MVM were attenuated in their replication. However, the severity of the attenuation was cell type dependent, with a greater attenuation observed in murine A9 fibroblasts compared to the human kidney cell line NB324K (Fig. 3.1). In order to establish how cell type influences viral attenuation, a range of cell lines were screened for their susceptibility to wild-type MVM. In permissive cell lines, the relative attenuation of CpG-high and codon-optimised 1st generation mutants was measured.

Susceptibility to WT MVM

To investigate which cell lines were supportive for MVM infection, cells were infected with 1 $TCID_{50}$ unit/cell for 1 hour, washed three times with PBS and incubated for 72 hours. The supernatant was then harvested and titrated in NB324K cells. Both A9 and NB324K cells were used alongside a range of other cell lines that were available from the lab stocks as well as from collaborators (Fig. 4.1).

As shown previously, MVM replicated well in both A9 and NB324K cells, growing to titres of $5.3x10^7$ $TCID_{50}$ units/ml in A9 cells and $8.7x10^6$ $TCID_{50}$ units/ml in NB324K. Similarly, the hamster kidney cell line BHK was also susceptible to MVM replication, yielding $2.0x10^6$ $TCID_{50}$ units/ml at 72 hours post infection. While viral supernatant was not tested for residual virus from the infection at the 1 h time point to measure in this experiment, it averaged 348 $TCID_{50}$ units/ml (A9 cells) and 206 $TCID_{50}$ units/ml (NB324K cells) previously

with the same protocol (Fig. 3.1). This was consistent with other pilot experiments, where similar values were obtained independent of the cell line used.

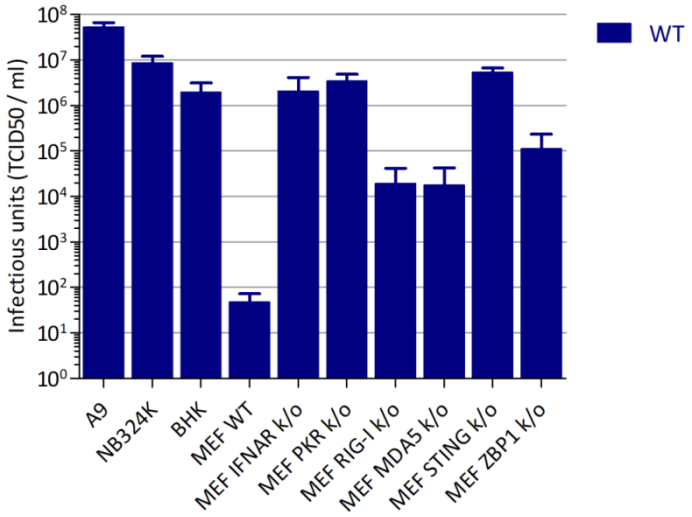


Figure 4.1 MVM wild type replication in different cell lines
Cell were infected with 1 $TCID_{50}$ unit/cell. Supernatant was harvested after 72 h and titrated in NB324K cells. ≥2 wells were infected for each cell line, error bars show SD.

The other cell lines tested were a panel of immortalised (SV40 Large-T transformed) mouse embryonic fibroblasts (MEFs) which contained knockouts (k/o) for different genes involved in innate immune signalling pathways. The panel included a parental, wild-type cell line (MEF WT), as well as knockouts for Interferon α/β receptor (IFNAR), the RNA recognition receptors PKR, RIG-I and MDA5, as well as DNA recognition receptors STING and ZBP1. Immortalised cell lines were used in this experiment. This was both for their ease of use and because it has been established before that primary, wild-type MEFs are not susceptible to MVM infection (462). In contrast, cell transformation and cancer associated cell changes generally tend to make cells (more) susceptible to parvoviral infection (463). It was hoped that using these immortalised MEFs would facilitate MVM replication even in the parental cell line. Unfortunately, only 47 $TCID_{50}$ units/ml could be recovered at the 72 h time point in the parental MEFs (Fig 4.1). However, all knockout

mutants tested in the panel supported viral replication to different extents. In IFNAR, PKR and STING knockout cell lines, the virus could replicate to titres similar to those in A9 and NB324K cells (IFNAR k/o: $2.0x10^6$ $TCID_{50}$ units/ml, PKR k/o: $3.4x10^6$ $TCID_{50}$ units/ml, STING k/o: $5.3x10^6$ $TCID_{50}$ units/ml). In the RIG-1, MDA5 and ZBP1 knockout cells, lower replication rates were observed (RIG-1 k/o: $1.9x10^4$ $TCID_{50}$ units/ml, MDA5 k/o: $1.8x10^4$ $TCID_{50}$ units/ml, ZBP1 k/o: $1.1x10^5$ $TCID_{50}$ units/ml).

Attenuation of CpG-high MVM mutants

After establishing cell susceptibility to wild-type virus, the 1st generation CpG-high full and half mutants was well as the codon-optimised virus were tested on the different cell lines (Fig. 4.2A), and where appropriate, the degree of attenuation was determined and compared between the cell lines (Fig. 4.2B). As viral titres were very low or negative for the CpG-High full mutant and codon-optimised mutants in some of the cell lines, it seemed more appropriate to compare the level of attenuation of the CpG-High half mutants as these gave more robust titres.

In agreement with the cell line specific differences observed in the viral growth kinetics in chapter 3, the CpG-high full and half mutants as well as the codon-optimised mutant were attenuated to a greater degree in A9 cells compared to NB324K cells. This partial reversal of attenuation was also observed in BHK cells. In the MEF cells, all knockout cell lines were tested, however, for the three intermediary susceptible cell lines (RIG-I k/o, MDA5 k/o and ZBP1 k/o), most of the mutant viruses were very close to or below the detection limit. In these cell lines, it was difficult to determine whether any differences in attenuation (like a potential reduced attenuation in the RIG-I knockout cell line) were due to reduced attenuation in the mutants or due to reduced replication of the wild-type virus. However,

it can be noted that none of mutants completely reverted to the respective wild-type levels.

In the three fully susceptible MEF knockout cell lines (IFNAR k/o, PKR k/o and STING k/o), attenuation in IFNAR and STING knockout cells reached similar levels to A9 cells. In these cell lines, the CpG-high full mutant titres were below or just at the detection limit, likely underestimating the total degree of attenuation. In contrast, in MEF PKR knockout cells, the level of attenuation was similar as observed for NB324K and BHK cells.

These results show that wild-type MVM by itself is restricted by a range of cellular antiviral proteins, being susceptible to both DNA and RNA recognition receptors. While surprising, the results are in line with it having a very small genome and thus very little coding capacity to spare for antiviral antagonism.

The level of attenuation for the CpG-high and codon-optimised mutants varied between cell lines. Notably, neither disabling STING nor IFNα/β signalling affected the CpG mediated restriction of the MVM mutants. This excludes a number of DNA sensors that signal through STING as CpG detection mechanisms. Furthermore, the robust attenuation in IFNAR knockouts infers that CpG recognition occurs early within the infection and does not rely on an antiviral response through IFNα/β. In contrast, the reduced level of attenuation in PKR knockout cells suggest that PKR might be involved in CpG recognition of MVM, though further experiments would be required to study this in more detail.

It is unfortunate that the ability to estimate attenuation in any MEF knockout cell line is directly dependent on how well the knocked-out protein can restrict wild-type MVM. This obviously skews the achievable results, as it is impossible to find any molecule with this system that specifically recognises CpGs but fails to detect wild-type virus.

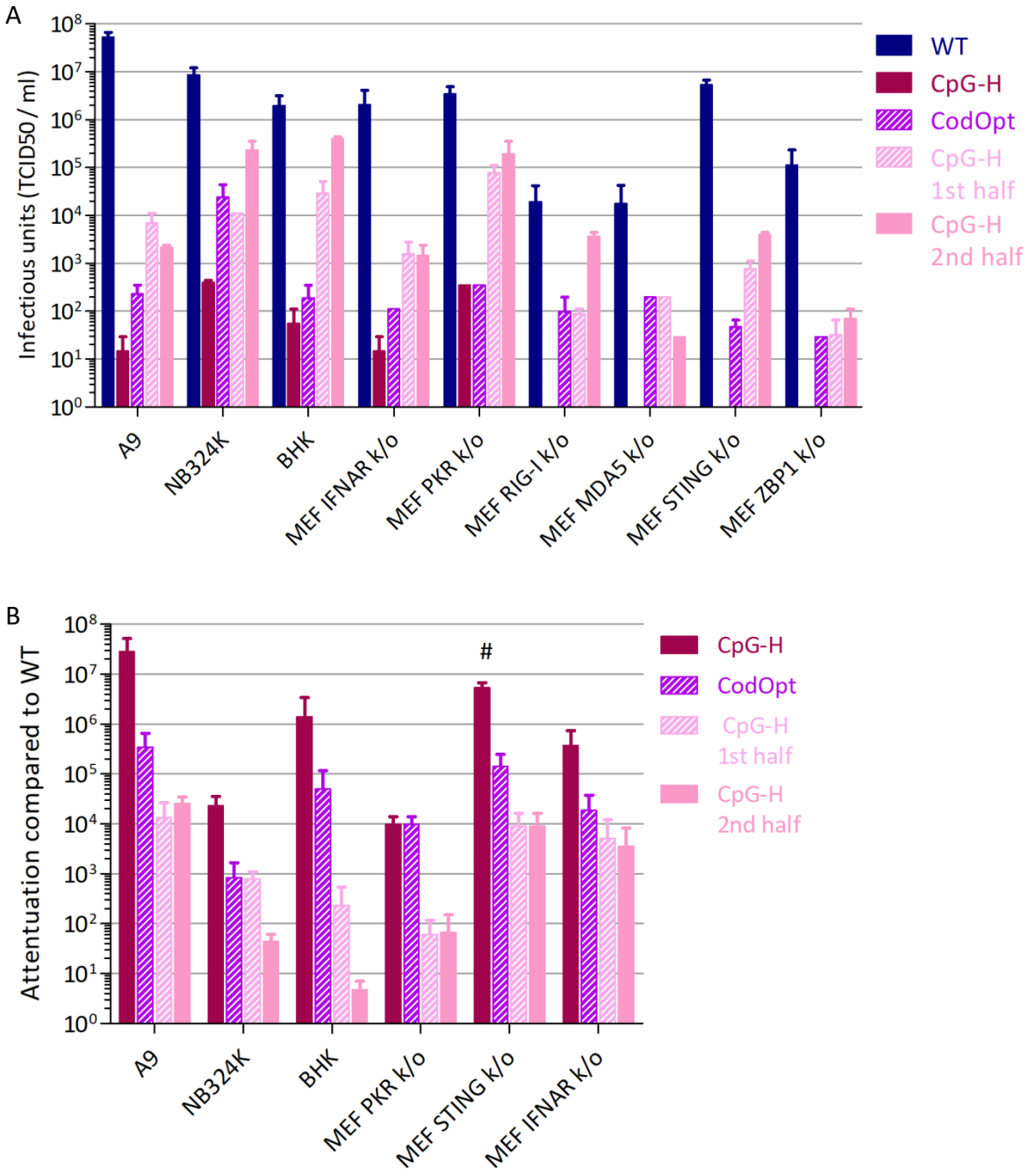


Figure 4.2 MVM mutant replication in different cell lines
Infectivity of MVM 72 h post infection (1 TCID unit / cell) in different cell lines. Two wells were infected for each mutant in each cell line, error bars show SD. **A.** Viral titres of MVM virus panel. **B.** Attenuation of CpG-high and codon-optimised mutants of compared to wild-type virus. # indicates samples where no virus could be recovered, thus underestimating attenuation.

4.2 Involvement of TLR9

TLR9 mediated signalling in A9 cells

It has previously been proposed that parvoviral genomes contain non-methylated CpGs which can be recognised by TLR9, activating antiviral cascades in the infected cell. It has been shown that human peripheral blood mononuclear cells (PBMCs) and HEK-hTLR9 (human kidney cell line HEK293 stably expressing human TLR9, which is otherwise absent from this cell line) can recognise unmethylated CpGs in MVM and thus mount an immune response through TLR9 (464). In contrast, murine bone marrow cells and plasmacytoid dendritic cells were unable to recognise MVM infection through TLR9 (465).

As A9 cells are fibroblasts, rather than immune cells, it was thought that TLR9 was unlikely to be the main restriction mechanism resulting in the severe attenuation observed in CpG-high viruses. To test TLR9 involvement, a dual approach was taken. A9 cells were tested for their ability to respond to extracellular DNA containing unmethylated CpGs (purified *E. coli* DNA, a TLR9 specific stimulant), establishing whether this cell line had a functional and responsive TLR9 pathway. In parallel. A9 cells were subjected to a TLR9 inhibitor during viral infection to see whether viral replication and mutant attenuation were altered when TLR9 was inhibited.

To investigate responsiveness to TLR9 stimulation, A9 cells and RAW264.7 cells (a monocyte/macrophage cell line known to have functional TLR9 signalling) were stimulated with *E. coli* DNA. IL-1β mRNA levels were determined with and without stimulation and normalised to GAPDH mRNA levels to test whether the TLR9 signalling cascade had been activated (466,467)(Fig. 4.3). In the same experiment, the effectiveness of ODN2088, a TLR9 specific antagonist (468), was tested.

In RAW264.7 cells, increased IL-1β mRNA levels were observed after TLR9 stimulation and this response could mostly be inhibited with a 2 h pre-treatment with ODN2088. In contrast, A9 cells showed no response to TLR9 stimulation.

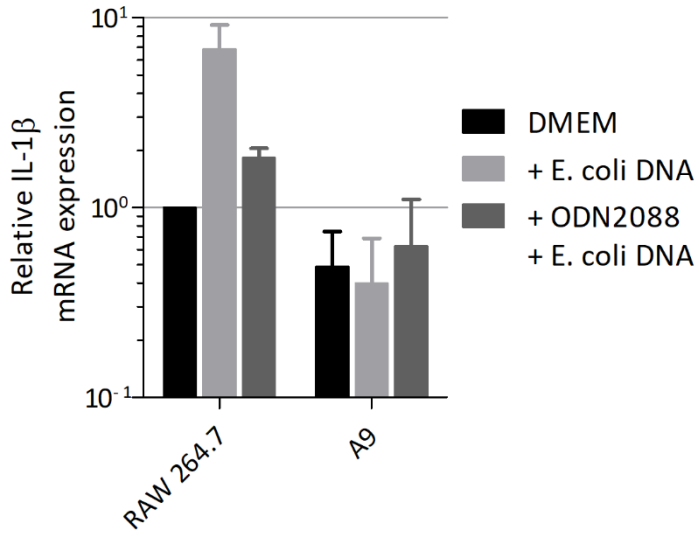


Figure 4.3 IL-1β induction after TLR9 activation
IL-1β induction after TLR9 activation by purified *E. coli* DNA in the presence or absence of TLR-9 antagonist ODN2088.Normalised to GAPDH using the ΔΔCT method. Three wells per condition and cell line, error bars show SEM.

Effects of TLR9 inhibition on MVM infection

While these results suggested that TLR9 signalling was not active in A9 cells, a functional assay was used to confirm these findings in the MVM infection model.

In order to test whether TLR9 expression had any effects on attenuation in A9 cells, cells were infected with wild-type virus, the TpA-high mutant or the CpG-high 1st half mutant at 1000 viral genome copies per cell (Fig 4.3B). This approach was chosen to subject cells to equal numbers of potential TLR9 ligands. As the CpG-high viruses had higher DNA to infectivity ratios than wild-type virus (i.e. more virus particles are necessary to achieve similar infection rates), one of the CpG-high half mutants was chosen to allow viral quantification by $TCID_{50}$. Replication of wild-type virus or the TpA-high mutant were minimally affected by TLR9 inhibition by ODN2088. There was also no relevant reversal of

the attenuation observed in the CpG-high 1st half mutant, confirming the non-involvement of TLR9 in the observed attenuation in this cell line.

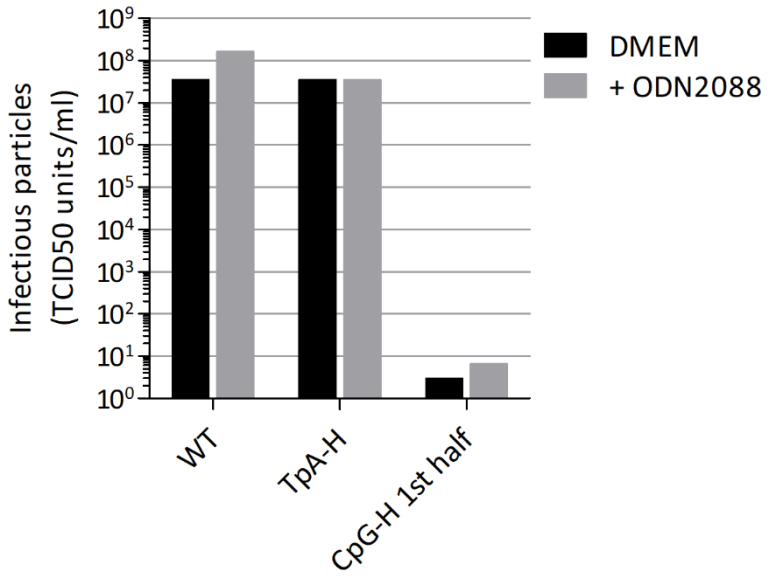


Figure 4.4 Effects of TLR9 inhibition on MVM mutant phenotype
Viral titres 72h post infection with 1000 viral genome copies/cell in the presence or absence of ODN2088. One well per mutant and condition, no error bars.

This experiment was only done once due to time restrictions before handing in my thesis. For the publication of this work, this specific experiment was repeated by Jeremy Ratcliff using qPCR to measure viral genome copies and more easily quantify the results. This was done in triplicate, depicted in Fig. 9b of Loew *et al.* (1), with a similar result (Appendix 7.3).

While it cannot be excluded that TLR9 might play a role in CpG recognition in the CpG-high or codon-optimised mutants created in this project, these experiments showed that it was not important for the attenuation observed in A9 cells.

It was decided that for this project, it would be of more interest to investigate factors responsible for the attenuation observed in A9 cells, rather than focus further onto TLR9 involvement.

However, it might be of interest to further test TLR9 involvement in CpG recognition in MVM or other small DNA viruses in the future. Ideally a MEF TLR9 knockout cell line would have been included in the panel in section 4.1, especially since it has been reported that MEF cells express and signal through TLR1-9 upon stimulation with the appropriate ligands (469). However, it was not possible to procure TLR9 knockout MEF cells, as most knockout studies on murine TLR9 have been done in animals rather than cell lines. However, given the robust inhibition of TLR9 activation by ODN2088, it would be possible to test TLR9 involvement in some of the permissive MEF knockout cell lines, to see if a (further) reversal of the CpG mediated attenuation could be observed.

4.3 Involvement of ZAP and RNaseL

ZAP has been known to restrict a wide variety of RNA, DNA and retroviruses (279,284-286,294,296-301). Recently it was discovered that ZAP specifically binds CpG-rich RNA, both in regions naturally elevated in CpGs within the Japanese encephalitis virus as well as artificially increased CpGs in HIV-1 and Echovirus 7 (275,276,298).

Since there are similarities in the replication strategy and life cycle of retroviruses and parvoviruses, this suggests a potential role for ZAP in the attenuation observed in the MVM mutants with increased CpG frequencies.

RNaseL has been found to cleave UpA or UpU sites in RNA, leading to faster degradation of UpA-rich RNAs. While increasing TpAs and thus UpA frequencies in MVM showed no to minimal attenuation, research of other members in the group on the RNA virus Echovirus 7 (E7) showed that knocking out RNaseL fully reversed UpA mediated attenuation while partially reversing CpG mediated attenuation.

As part of their research, Isabelle Dietrich and Valerie Odon created ZAP and RNaseL knockout cell lines in the human lung cancer cell line A549 (276). Knockout cell lines are both expensive and time consuming to generate and it was unclear whether either protein was involved in the observed attenuation of CpG-high and codon-optimised mutants of MVM. Hence, it was decided to test the suitability of these cell lines for MVM rather than E7 and whether any reversal of viral attenuation could be observed.

Susceptibility of A549 cells to MVM

First, it needed to be established whether A549 cells and the derived knockout cell lines would be able to support the full viral life cycle and how well the virus would be able to

propagate within these cell lines. This was followed by infecting the cell lines with the 1st generation CpG-high and codon-optimised mutants to examine the cell lines for any reversal in mutant attenuation (Fig. 4.5).

Cells were infected with 1 $TCID_{50}$ unit/cell for 1 hour, washed three times with PBS and incubated for 72 hours. The supernatant was then harvested and titrated in NB324K cells. The results were compared to the values of A9 cells as established in section 4.1 (Fig. 4.2A).

While there appeared to be some susceptibility to the wild-type virus in all A549 cell lines (A549: $2.1x10^4$ $TCID_{50}$ units/ml, A549 ZAP k/o: $7.3x10^4$ $TCID_{50}$ units/ml, A549 RNaseL k/o: $1.5x10^4$ $TCID_{50}$ units/ml), viral titres were markedly reduced compared to A9 cells (Fig 4.5). Interestingly, neither knockout had a big effect on susceptibility, unlike some of the MEF knockout cells.

For the mutants, similar obstacles as in section 4.1 were faced. Due to the lower wild-type replication, it was difficult to judge whether overall attenuation levels were reduced because of a lower restriction of the mutants or due to a higher restriction of the wild-type virus. However, no drastic reversion was seen in either cell line, a striking difference to the complete reversion of CpG attenuation of E7 observed in the ZAP knockout cell line (276).

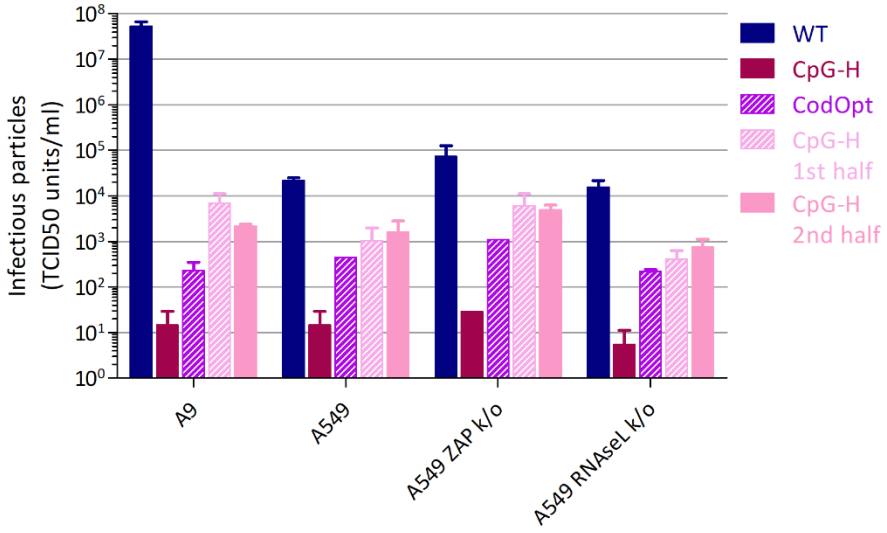


Figure 4.5 MVM mutant replication in A549 knockout cells
Infectivity of MVM 72 h post infection (1 $TCID_{50}$ unit/cell) in different cell lines. Two wells were infected per mutant and cell lines, error bars show SD.

For publication, Jeremy Ratcliff repeated this experiment after my departure trying an earlier time point to measure viral titre due to the limitations discussed in the next section. As can be seen in Fig. 11c in Loew *et al.* (1) (Appendix 7.3), the difference between WT and the CpG-H1 mutant as well as parental and ZAP knockout A549 cells was smaller at 48 h compared to the above results at 72 h.

Limitations of the A549 ZAP and RNAseL knockout system

It was apparent that there were some factors in place restricting the virus from completing its replication cycle or several rounds thereof, however it was unclear at what stage within the viral life cycle the restriction occurred. Since there were some virus particles released, the virus must have been able to enter the cell, though this could be very inefficient. It is also reliant on the cell machinery throughout its replication cycle and there could be some incompatibility at any stage. It might be that the viral NS proteins cannot interact as well with the human host proteins, though MVM replicates well in

human NB324K cells. Thus it is more likely that the restriction is due to a different/more potent antiviral response, especially since MVM propagation is restricted by various pattern recognition receptors as shown at the beginning of this chapter.

Given the pattern in Fig. 4.5, where wild-type MVM produced 1000-times fewer infectious progeny virus in any A549 cell line than the A9 cell line, while the CpG-high and codon-optimised mutants show much more similar titres between cell lines, the inhibiting factor might also be viral egress. In A9 cells, virus is released via active egress and cell lysis, however in A549 no CPE could be observed, even after infection with a very high multiplicity of infection. This might indicate that viral egress is impaired and might only occur through active egress or rare cell lysis, effectively trapping progeny virus in A549 cells and preventing subsequent rounds of infection.

With these considerations in mind, single round infections and transfections were used to test whether RNA and protein expression were sufficiently robust to compare between cell lines.

RNA expression levels in ZAP and RNaseL knockout cells

To investigate RNA expression, cells were either transfected with the plasmid pMM984 containing the infectious clone of the MVM genome using Aphidicolin as a replication inhibitor. In a second experiment cells were infected with 1000 viral genome copies per cell of MVM in either the presence or absence of Aphidicolin. This approach was used to examine the direct effects on RNA expression in the transfection setting, as well as in the infection context where other factors might affect viral RNA expression.

For the transfection, MVM R1 mRNA levels (Fig. 4.6B) were measured by qPCR at 8 and 18 hours post transfection, while only 18 h samples were analysed in the infection context

(Fig. 4.6C). This was due to the fact that RNA levels were just at or below the detection limit at 8 hours post infection. For these experiments, wild-type virus and the 1st generation CpG-high full mutant were compared, after a pilot experiment in A549 indicated that the other main 1st generation mutants (CDLR, TpA-high and codon-optimised) showed no or very little change in RNA expression levels (similar to their expression in A9 cells) and there would thus have been very little new insight to be gained from including these mutants.

In the transfection experiment, the A549 parental cell line showed a similar trend as the A9 cells, where RNA levels were similar between wild-type virus and the CpG-H mutant after 8 hours, but 4-fold lower RNA expression was observed in the CpG-high mutant after 18 h (Fig. 4.5B, Fig. 3.6C). Interestingly, in both ZAP and RNaseL knockout cell lines this difference in RNA expression was absent after 18 hours, indicating a potential involvement of both ZAP and RNaseL in CpG mediated restriction on the RNA level.

In the infection experiment, wild-type R1 mRNA levels were roughly 2 logs higher in each cell line when replication was not inhibited, which is in line with the virus progressing through its life cycle in these cell lines (Fig. 4.6C). RNA expression of the WT virus was increased approximately 3-fold in the ZAP knockout cell line compared to the parental cell line. In comparison, WT RNA levels were similar between the A549 parental and RNaseL knockout cell line. This effect was independent of replication. In contrast to the transfection experiment, all cell lines showed lower RNA levels in the CpG-H mutant post infection, though some differences could be observed (Fig4.6C).

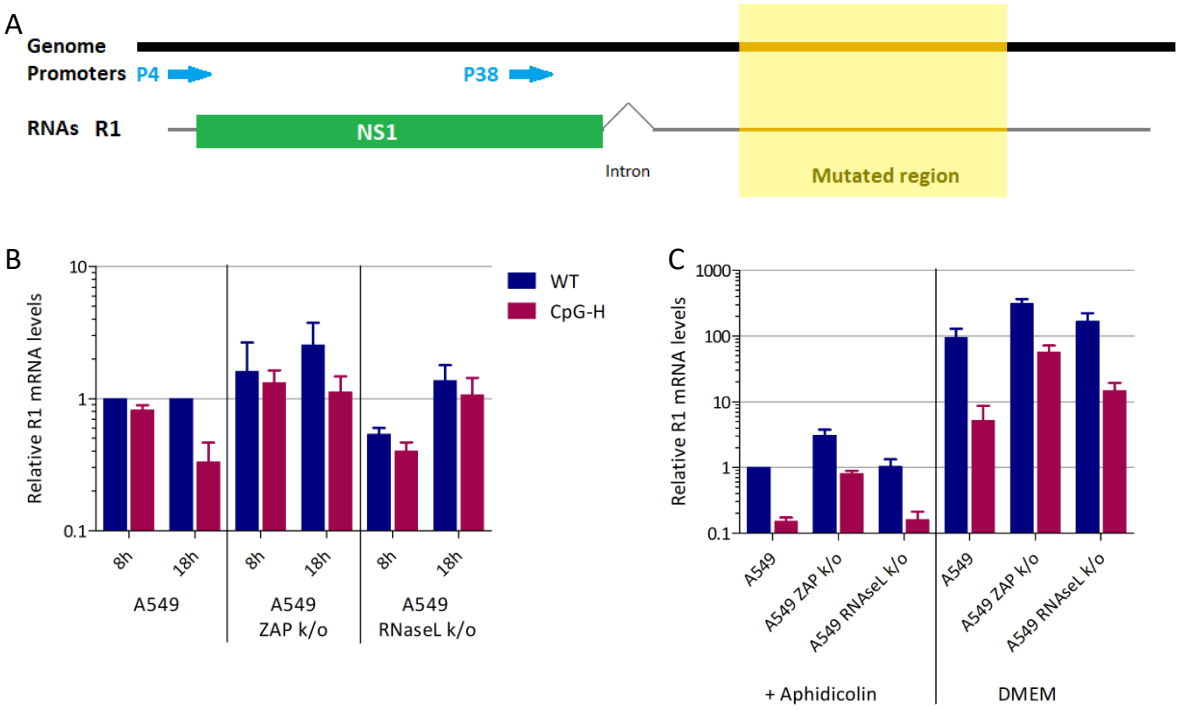


Figure 4.6 MVM R1 mRNA levels in parental A549 and ZAP and RNaseL knockout cells
A. Diagram of the MVM genomic structure. Promoters (blue), transcript R1 with ORF (green coloured box), mutated region highlighted in yellow. Based on Berns & Parrish, Fig 57.10 (392). **B.** MVM R1 mRNA expression 8 h and 18 h post transfection with pMM984 (viral infectious clone) in the presence of Aphidicolin, normalised to cellular GAPDH and MVM WT R1 mRNA in the parental A549 cell line. Three wells were transfected per condition, error bars show SEM. **C.** MVM R1 mRNA expression 18 h post infection with 1000 virus genome copies per cell in the presence or absence of Aphidicolin, normalised to cellular GAPDH and MVM WT R1 mRNA in the parental A549 cell line in the presence of Aphidicolin. Three wells were transfected per condition, error bars show SEM.

In the A549 parental cell line, the CpG-high mutant expressed 6-fold less RNA in the presence of Aphidicolin, while a 17- fold reduction was observed when replication was not inhibited. This indicates that further mechanisms might contribute to CpG mediated attenuation as MVM progresses through its replication cycle. In the RNaseL knockout cell lines, RNA expression of the CpG-high mutant was reduced 6-fold in the presence of Aphidicolin and 11-fold without Aphidicolin. Thus a similar restriction was observed in RNAseL knockout cells and in the parental A549 cell line in the presence of Aphidicolin, but the CpG-H virus appeared to be less restricted in the RNAseL knockout cells without Aphidicolin present compared to the parental cell line. In the ZAP knockout line, the CpG-

high mutant showed 3-fold less RNA expression in the presence of Aphidicolin and a 6-fold reduction in its absence. Due to the overall increased RNA expression in this cell line and the decreased CpG mediated attenuation, RNA levels of the CpG-high mutant in the ZAP knockout cell line almost recovered to the RNA levels observed wild-type virus in the A549 parental cell line.

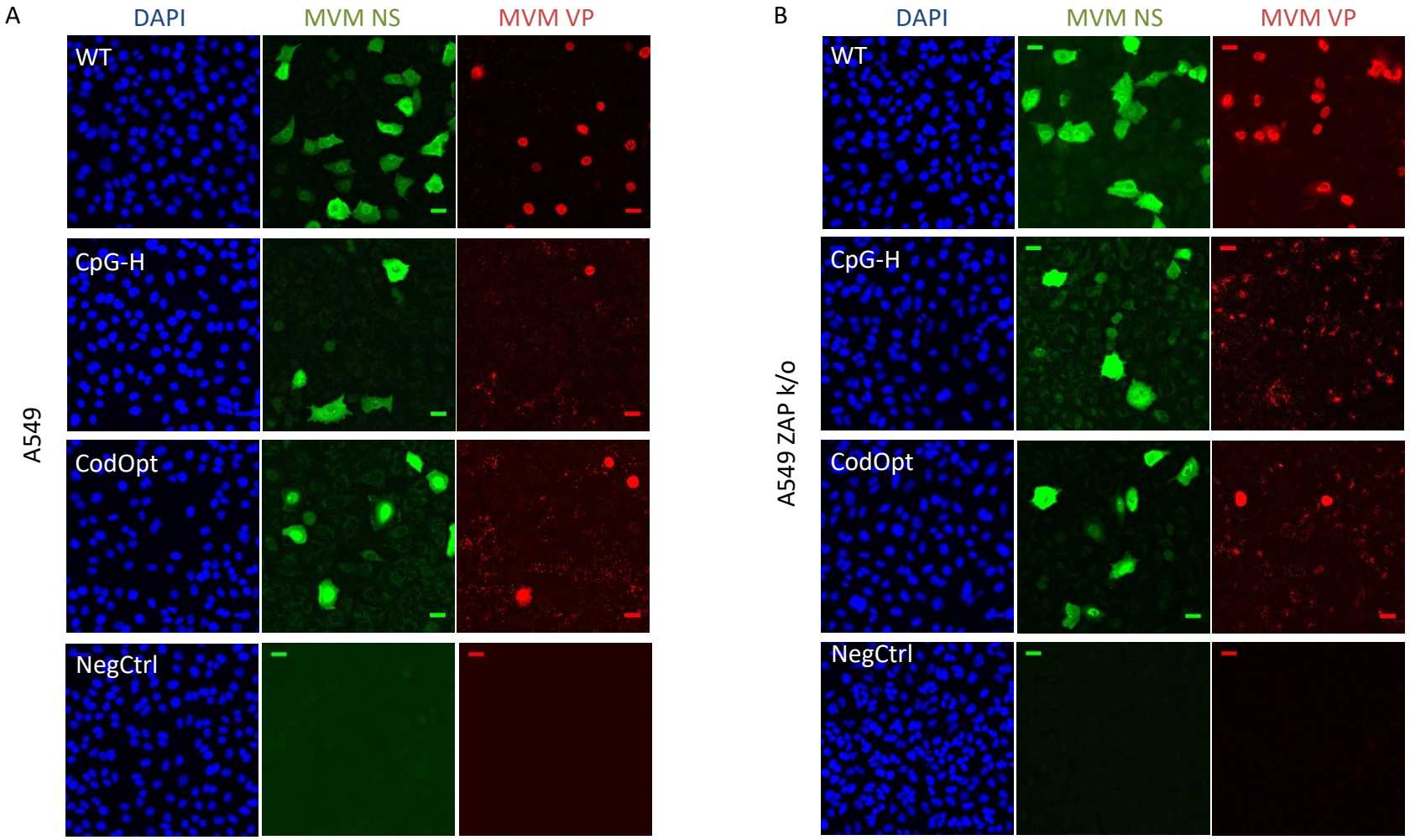
It appears that different restriction mechanisms are in action depending on the context surrounding CpG-rich viral RNA. These results suggest that both ZAP and RNaseL affect CpG rich RNA, but that other factors are involved in CpG mediated restriction in the infection context.

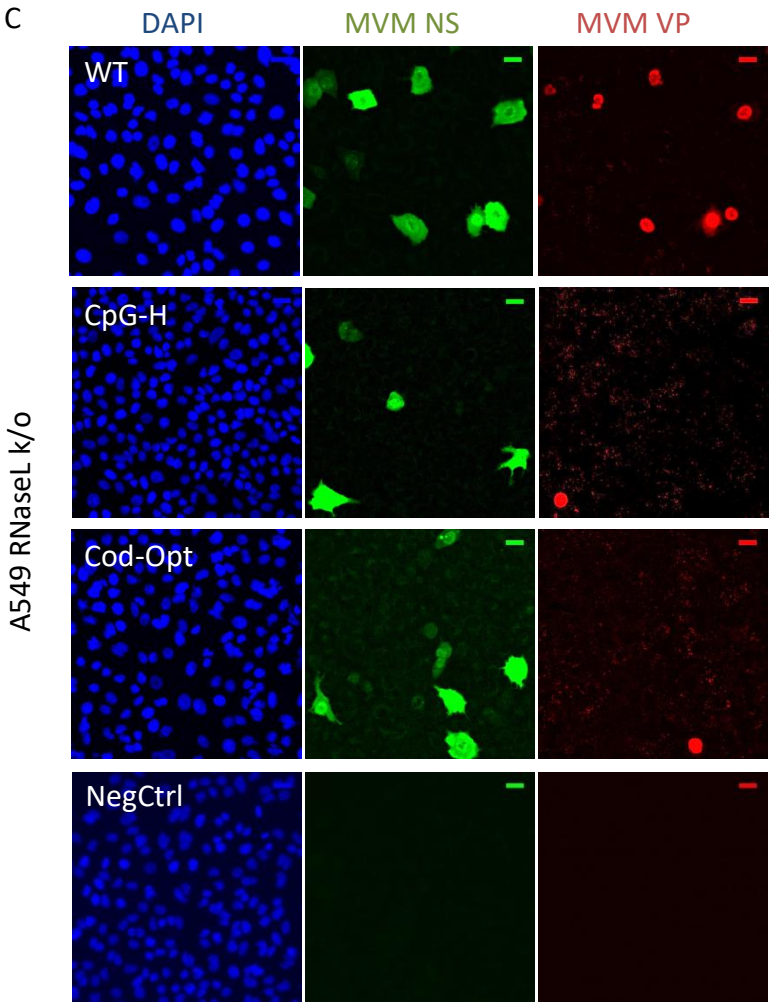
As these results were quite intriguing, it was decided to look at the infection via immunofluorescence assays to determine how these changes in RNA restriction would affect viral protein expression.

Visualisation and quantification of viral infection using Immunofluorescence

To examine protein expression in the A549 ZAP and RNaseL knockout cell lines, cells were infected with 1000 genome copies per cell of MVM for 1 hour, washed and stained for viral NS and VP proteins after 18 hours.

For this experiment, wild-type virus was compared to the 1st generation CpG-high and codon-optimised mutants, as it was unclear whether the attenuation observed in the codon-optimised mutant might also be affected by ZAP or RNaseL. A full representative panel for each cell line and virus can be found in Fig. 4.7A-C (*image processing done by Niluka Goonawardane)*, while merged images of all viruses in the parental A549 and ZAP knockout cells are shown in Fig. 4.7D-E (*image processing done by Niluka Goonawardane)*.


A
DAPI
MVM NS
MVM VP
WT
CpG-H
CodOpt
NegCtrl
A549
B
DAPI
MVM NS
MVM VP
WT
CpG-H
CodOpt
NegCtrl
A549 ZAP k/o


C
DAPI
MVM NS
MVM VP
WT
CpG-H
Cod-Opt
NegCtrl
A549 RNaseL k/o

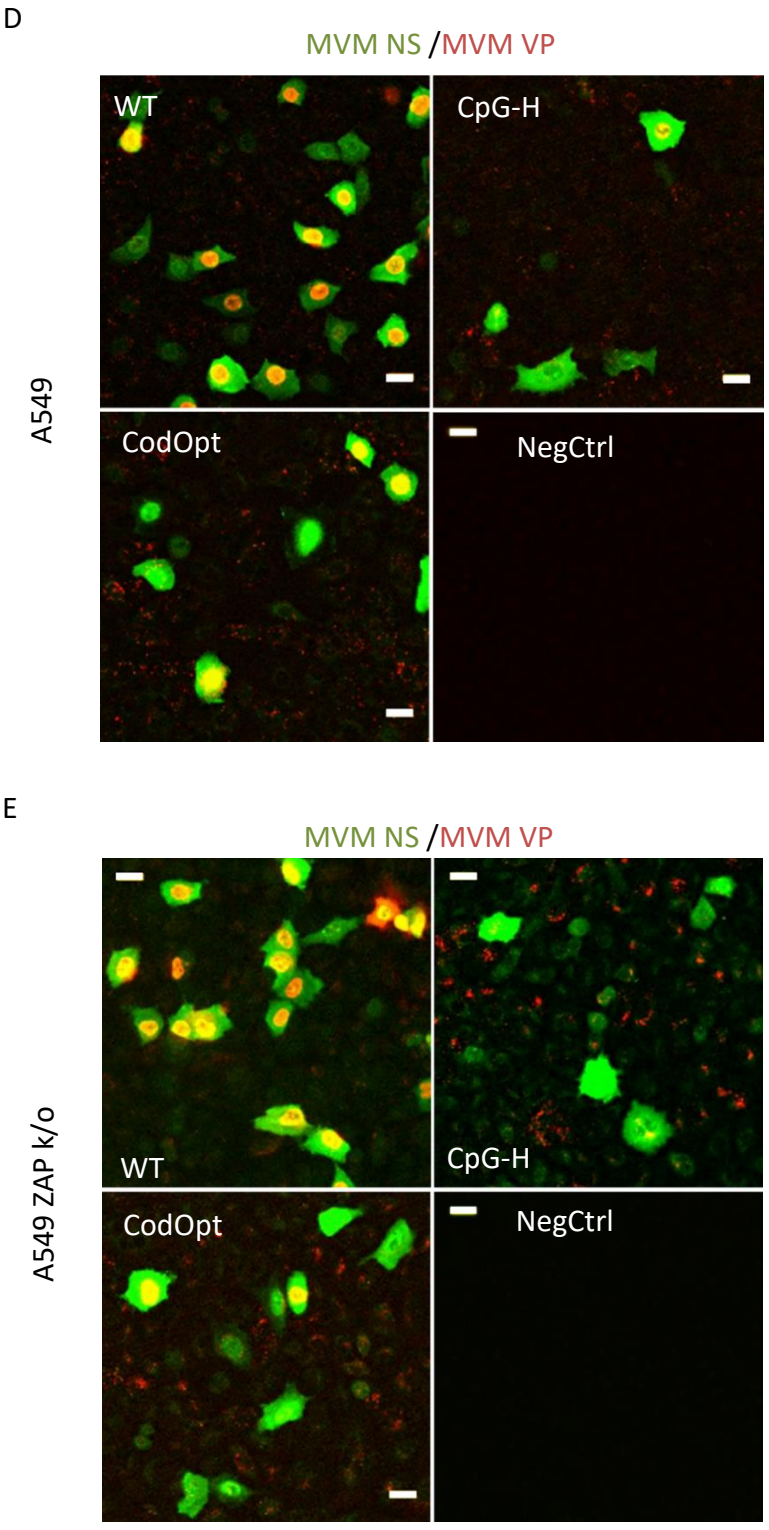


Figure 4.7 Immunofluorescence staining of viral NS and capsid proteins in A549 cell lines
Immunofluorescence staining of DNA (blue), viral NS (green) and capsid (red) proteins at 18 h post infection with 1000 viral genome copies/cell. Uninfected cells were used as a negative control. Three wells were infected per mutant/cell line, scale bar 50 µm, representative images shown. *Image processing and quantification done by Niluka Goonawardane* **A.-C.** Panel for each cell line. **A**. A549 cells **B**. A549 ZAP knockout cells. **C.** A549 RNaseL knockout cells. **D.-E**. Merged images. **D.** A549 cells **E.** A549 ZAP knockout cells.

The percentage of infected cells and relative fluorescence intensities were determined separately for the non-structural and capsid proteins and are shown in Fig. 4.8A+B (*image processing and quantification done by Niluka Goonawardane)*. A minimum of 1952 cells were analysed per triplicate for each sample.

In the parental A549 cell line infected with wild-type virus, 28 % of cells were positive for NS protein expression while 22 % of cells expressed the VP proteins, which is approximately a third lower than a comparable infection produced in A9 (Fig. 4.7A, Fig.3.3A). In contrast, in samples infected with CpG-high virus, there was a 3-fold reduction in the number of cells expressing viral NS or VP. In wells infected with the codon-optimised mutant, approximately 21% cells expressed NS and VP, a 25 % reduction compared to wild-type virus for the NS proteins, while similar amount of cells expressed VP when infected with wild-type or codon-optimised virus.

Similar effects were observed in terms of fluorescence intensity. A 4-fold reduction in fluorescence intensity of NS was observed for the CpG-high mutant compared to WT virus and a 9-fold reduction in intensity was observed for VP in the parental A549 cells. In the codon-optimised mutant, NS and VP expression was reduced by approximately 40% after infection compared to the wild-type virus.

These results showed a similar trend overall as the A9 cells, though the codon-optimised mutant showed a lower degree of attenuation in the parental A549 cells compared to the A9 cell line.

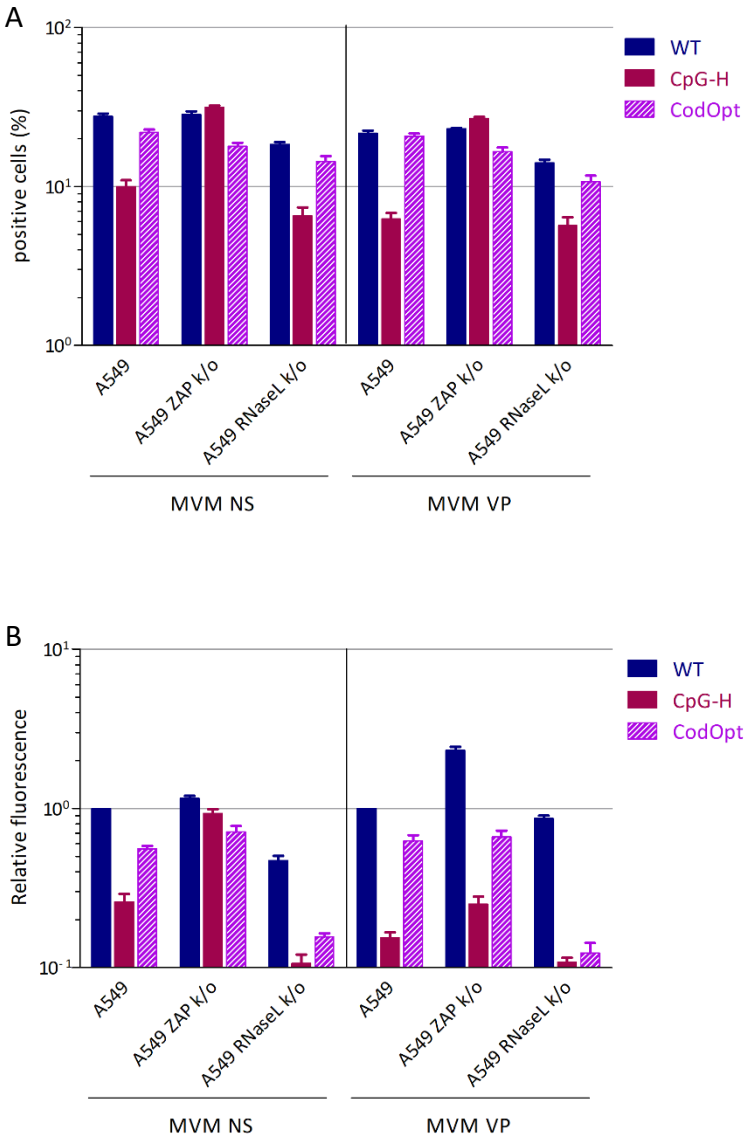


Figure 4.8 Quantitative analysis of immunofluorescence staining of viral NS and capsid proteins in A549 cell lines

Quantification of immunofluorescence staining of viral NS and capsid proteins using Imaris and Fiji. At 18 h post infection with 1000 viral genome copies/cell. Uninfected cells were used as a negative control. Three fields of view from at least two different wells were analysed per mutant and cell line, error bars show SEM. *Image processing and quantification done by Niluka Goonawardane*. **A.** Percentage of cells expressing MVM NS or VP. **B.** Relative fluorescence intensity of NS or VP compared to WT MVM.

In the ZAP knockout cells, 28 % of cells expressed the non-structural proteins after infection with the wild-type virus, while 23 % of cells expressed capsid proteins, similar values as observed in the parental A549 cells. The fluorescence intensity of NS proteins in cells infected with wild-type virus was similar in both parental A549 and ZAP knock cells, while capsid expression was doubled in ZAP knockout cells

In wells infected with the CpG-high virus, 32 % of cells expressed the non-structural proteins and 26 % of cells were positive for the capsid proteins, similar levels to wild-type virus. Within infected ZAP knockout cells, NS expression recovered to levels similar to cells infected with the wild-type virus. On the other hand, capsid protein expression remained 9-fold reduced compared to the wild-type expression in this cell line.

For the codon-optimised virus, there was little difference between infections in the two cell lines. In the ZAP knockout cell line, 18 % of cells expressed the non-structural proteins after infection, while 17 % of cells expressed the capsid proteins. Protein expression within infected ZAP knockout cells was similar to protein levels of the codon-optimised mutant in the parental A549 cell line.

Peculiarly, the subcellular distribution of the capsid proteins appeared different for the CpG-high mutant in the ZAP knockout cell line (Fig. 4.7A+B, close up images in Fig. 4.9). During parvoviral infection, the capsid proteins translocate into the nucleus to assemble into progeny virions. In infected A9 cells and parental and knockout A549 cells, capsid expression can be found within and outside the nucleus for the wild-type virus (Fig. 3.2, Fig. 4.7).

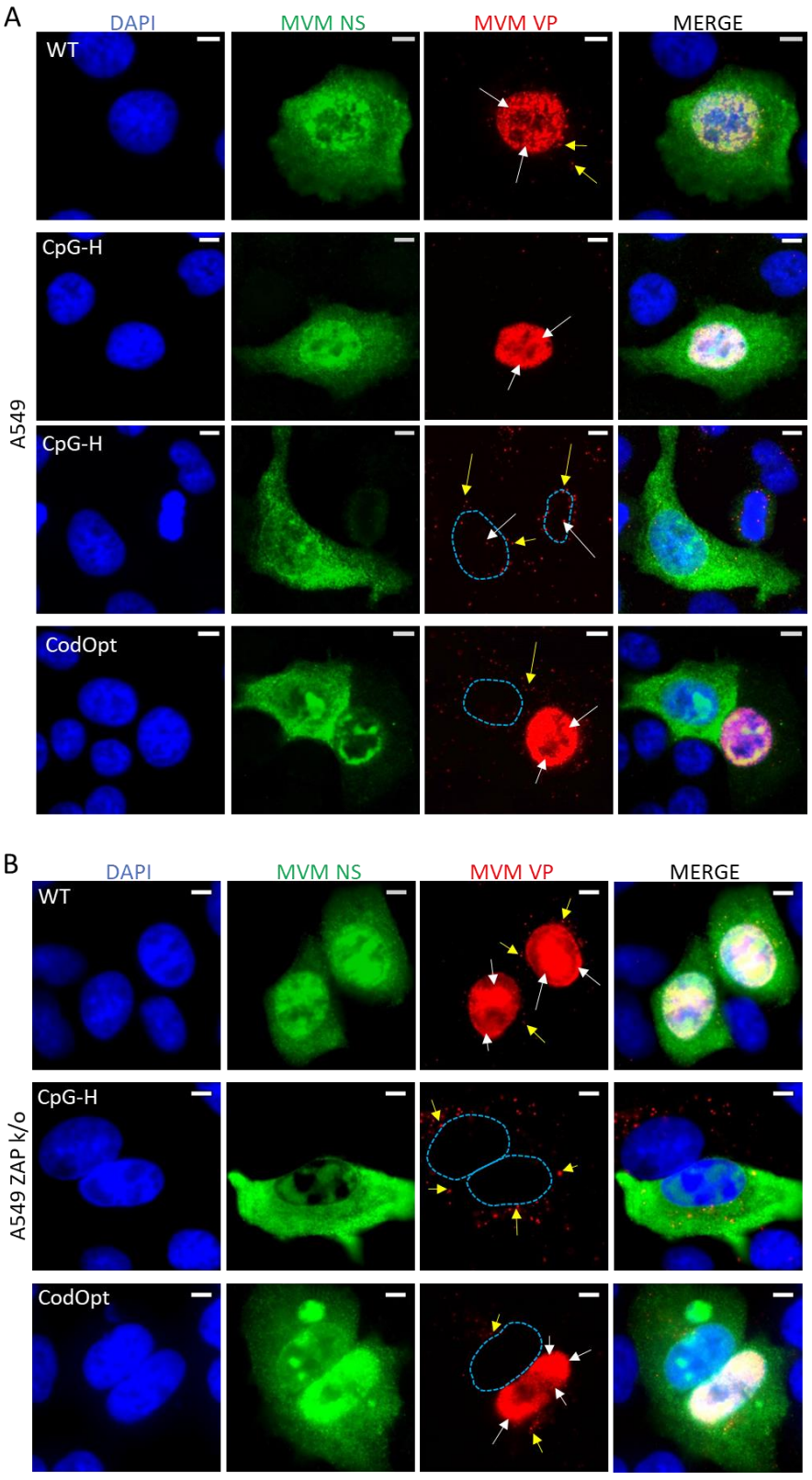


Figure 4.9 Subcellular localisation of MVM proteins post-infection
Immunofluorescence staining of DNA (blue), viral NS (green) and capsid (red) proteins at 18 h post infection with 1000 viral genome copies/cell. Light blue dotted lines represent nuclear outline, white arrows indicate nuclear VP expression and yellow arrows indicate VP expression outside the nucleus. Three wells were infected per mutant/cell line, scale bar 10 µm, representative images shown. *Image processing and quantification done by Niluka Goonawardane*. **A.** A549 cells. **B.** A549 ZAP knockout cells.

In A9 or parental A549 cells infected with CpG-high virus, fewer infected cells express the capsid proteins and nuclear translocation of them appeared to be less common. One possible explanation might be a slower progression through the viral life cycle and thus the infected cells might have been at different stages of infection when the cells were fixed and stained.

In ZAP knockout cells, none of the infected cells showed nuclear capsid expression. Instead, many infected cells showed capsid accumulation close to the nucleus (Fig. 4.7B, Fig. 4.9B). This capsid protein localisation could be observed both in cells expressing the non-structural proteins and within cells that did not. This raises the question whether the capsid proteins in these cells are from incoming virions that have been retained on their way to the nucleus or whether these represent newly expressed capsid in infected cells. Doing a time course of the infection would likely give a better understanding of this issue, however it was outside the scope of this project to investigate this in more detail.

In the RNaseL knockout cell line, infection with wild-type virus lead to expression of the non-structural proteins in 18 % of cells, while 14 % of cells expressed the viral capsid proteins. This was approximately one third less compared wild-type infection in the parental A549 cell line. The expression of the non-structural proteins was reduced by half, while the expression of the capsid proteins was similar to the amounts observed in the infection of the parental A549 cell line.

When infected with CpG-high virus, 7 % of cells expressed the non-structural proteins, while 6 % were positive for the capsid protein, which is similar to the 3-fold decrease of positive cells observed in the parental A549 cell line infected with CpG-high virus. There

was a 4-fold decrease in NS expression and an 8-fold decrease in VP expression in RNAseL knockout cells infected with CpG-high virus instead of wild-type virus. This was a similar restriction as observed between the two viruses infecting the parental A549 cell line.

When the RNAseL knockout cells were infected with codon-optimised virus, 14 % of cells expressed the non-structural proteins, while 11 % of cells were positive for the capsid proteins. This represented an approximately 20% decrease compared the fraction of positive cells after wild-type infection in this cell line. This was a similar ratio as observed for the two different viruses in the infection of the parental A549 cells. Interestingly, while similar fractions of cells were infected in the RNAseL knockout cell line, overall viral protein production was reduced relative to protein expression of the wild-type virus in these cells. The expression of the non-structural proteins was reduced 3-fold, while capsid protein expression was reduced approximately 6-fold.

From these results it appears that ZAP plays a role in restricting CpG-high MVM in cells, as ZAP knockout cells show a partial reversal of the phenotype, though other factors seem to be involved.

It is interesting that MVM is able to enter A549 cells, produce proteins and replicate, though not much progeny is produced. While there were clear differences in CpG mediated attenuation and restriction apparent between the parental and knockout cell lines, viral titres were very similar between them. This begs the question whether viral egress is the rate limiting step in these cell lines for the wild-type virus or whether other antiviral mechanisms prevent the virus from completing its replication cycle.

5 Discussion

Varying levels of CpG suppression have been observed in small DNA viruses. This chapter discusses what is known about CpG recognition within cells and how small DNA viruses and other viruses are subjected to these mechanisms, including TLR9, ZAP and RNaseL. This is compared and contrasted with the findings of this study using MVM mutants with altered dinucleotide frequencies. Furthermore, it is discussed how different compositional changes to MVM, specifically codon optimisation and altering G+C content affect the virus. To conclude, areas for further study for both MVM and MPyV are elaborated.

5.1 CpG and TpA/UpA suppression in small DNA viruses 118

5.2 Recognition of unmethylated CpGs by TLR9 119

5.3 Involvement of cytoplasmic RNA and DNA recognition receptors 121

5.4 Restriction of CpG-rich RNA by ZAP and RNaseL 122

ZAP 122

RNaseL 123

5.5 Attenuation of TpA/UpA-high viruses 124

5.6 Dose dependency of CpG attenuation 125

5.7 Effects of codon (de)optimisation and G+C content 127

5.8 Future work 130

MVM 130

MPyV 131

5.9 Conclusion 132

5.1 CpG and TpA/UpA suppression in small DNA viruses

When viral sequencing became available and more widespread, bioinformatics analyses showed that various virus groups and families differed in their dinucleotide composition. Vertebrate RNA and retroviruses as well as small DNA viruses such as parvoviruses, papillomaviruses and polyomaviruses showed a marked suppression of CpG dinucleotides and a noticeable suppression of TpA/UpA dinucleotides in their genomes (11,13,470). For this study, this analysis was repeated for small DNA viruses, as the number of complete virus genomes available had been expanded massively since the original study of Karlin *et al.* (11). This analysis showed that the degree of CpG and UpA repression varied across different small DNA viral families, leading to distinct distributions. Overall, anello-, parvo- and papillomaviruses showed somewhat similar distributions of CpG and TpA/Upa dinucleotides as murine mRNA in both their coding sequences and their genomes, leading to a general overlap with murine mRNA (Fig. 2.2, Fig. 2.4, Fig. 2.5). Polyomaviruses showed overwhelmingly very low CpG dinucleotide frequencies, though it can be argued that they might just be at the natural end of the CpG distribution seen in vertebrate host mRNA (Fig. 2.6). In contrast, CpG suppression was not very strong in circoviruses, with many of the viral genomes and genes showing no CpG suppression at all (Fig. 2.3).

These results suggest that different virus families might be affected differently by CpG recognition depending on their life cycle or ability to interfere with said recognition.

5.2 Recognition of unmethylated CpGs by TLR9

TLR9 has been well established in its ability to recognise unmethylated CpGs within both single and double stranded DNA. Over the years, different groups have studied how small DNA viruses interact with and are affected by TLR9 recognition.

While it is easy to imagine that viruses might gain an advantage by mimicking cellular mRNAs in their composition, the lack of CpG suppression in circoviruses seems peculiar. It has been shown that porcine circovirus 2 contains unmethylated CpG sites within its genome which can either be stimulating or inhibiting an antiviral response to the virus (471-473). Stimulatory sequences induced an interferon α (IFNα) response, while inhibitory DNA not only suppressed the CpG induced IFNα response, but facilitated an overall suppression of the IFNα response (471-473). Vincent *et al.* (471) showed that the inhibitory sequences in the porcine circovirus genome suppressed viral stimulatory sequences and TLR9 agonistic CpG-ODNs equally well, suggesting the involvement and suppression of TLR9 mediated DNA recognition.

Within the parvovirus family, several members have been studied in regards to TLR9 activation. Porcine parvovirus (PPV) has been shown to remain hypomethylated throughout its replication cycle and artificially methylating PPV genomes before transfecting them showed no enhancement in replication, instead they were slightly inhibited and progeny virus became non-methylated (474).

AAVs have been extensively studied for their use as gene therapy vectors. They have been found to activate both innate and adaptive immune response through TLR9

activation (475-480). Measures such as CpG depletion have been used to circumvent TLR9 recognition and to avoid immune activation (481).

Human PBMCs and HEK293 cells expressing functional TLR9 were able to recognise MVM and parvovirus H-1 via TLR9 (464). However in murine plasmacytoid dendritic cells, which express high level levels of TLR9, purified MVM genomic DNA could activate TLR9, while MVM virions did not elicit an immune response (465).

In this study, CpG-high mutants were greatly attenuated in the A9 cell line, however it was shown that A9 were not able to induce IL-1β after TLR9 stimulation and no reversal of the attenuated phenotype was observed in the presence of a TLR9 antagonist inhibiting TLR9 signalling (Figs. 4.3-4.4).

While it cannot be excluded that TLR9 would be involved in CpG recognition in cell lines where it is active, the observed restriction of CpG-high viruses was TLR9-independent in this context.

5.3 Involvement of cytoplasmic RNA and DNA recognition receptors

STING has been identified as a central adapter protein converging signals from a range of different cytoplasmic DNA recognition receptors (reviewed by Motwani *et al.* (201)). While there have been three PRRs shown to recognise CpG motifs in DNA, namely TLR9, DEAH-Box Helicase 9 (DHX9) and 36 (DHX36), none of them signal through STING. To investigate whether other PRRs might signal through STING, MEF knockout cells lacking STING or other proteins involved in antiviral responses were infected with MVM and tested for their susceptibility to both WT and mutant viruses. It has previously been shown that parental MEFs did not support viral growth, but knocking out PKR allowed viral replication and propagation (462). In this study, the same observation was made: parental MEFs were not susceptible to MVM, while virus grew to high titres in the PKR knockout cell line (Fig. 4.1). Additionally, MVM was able to grow in several other MEF knockout cell lines. It could grow to high titres in STING and IFNα/β receptor knockout cells, while growing to intermediate titres in RIG-I, MDA5 and ZBP1 knockout cells (Fig. 4.2).

There was no reversion in CpG-mediated attenuation in either STING or IFNα/β receptor knockout cells, potentially ruling out many DNA PRRs as well as other antiviral mechanisms converging in the IRF3 or IRF7 induced activation of the interferon pathway. Intriguingly, CpG-mediated attenuation was less pronounced in PKR knockout cells, one of the central RNA recognition receptors and associated with RNA virus recognition (482). It is unclear how the general susceptibility of MVM to PKR relates to the CpG specific restriction observed.

5.4 Restriction of CpG-rich RNA by ZAP and RNaseL

ZAP

Besides PKR, there might be other cytoplasmic RNA PRRs recognising parvoviral RNA in infected cells, leading to a CpG suppression in these viruses. ZAP has been shown to restrict CpG-rich RNAs in HIV-1, echovirus 7 (E7) and Japanese encephalitis virus (275,276,298). The ZAP-mediated CpG restriction of RNA is dependent on KH and NYN domain containing protein (KHNYN) and the targeted RNAs are likely sequestered or degraded as a consequence (290). Recently, the molecular mechanisms and precise binding sequence of ZAP have been discovered (483,484).

In this study, CpG-mediated attenuation could be partially reversed in a human knockout cell system (Figs. 4.6-4.8). In MVM, the CpG-high mutant showed a marked reduction of viral RNA expression levels in both A9 and A549 cells, even when viral replication was inhibited (Fig. 3.6, Fig. 4.6). This RNA restriction is absent in A549 ZAP knockout cells when transfected with plasmid DNA containing the infectious clone, while a reduced CpG-mediated RNA restriction occurs when these cells are infected with MVM.

Using immunocytochemistry, it was shown that CpG-mediated gene expression attenuation was similar in A9 and parental A549 cells post infection (Figs. 3.3-3.4, Figs. 4.7-4.8). In the A549 ZAP knockout cells, gene expression could be restored for the non-structural proteins. However, for the capsid proteins, gene expression was only marginally restored and there appeared to be some other restriction in plcae (Figs. 4.7-

4.8). This indicates that ZAP might restrict a very early point of the infection, while other factors might restrict MVM as it progresses through its replication cycle.

In contrast, CpG-mediated restriction in HIV-1 was completely reversed in ZAP knockout cells (275). This might be due to different factors being involved or the fact that the human system used in this study might have a different effect on MVM than a mouse system would. Further work is needed to investigate the role of ZAP in MVM infection and confirm these results in a mouse system.

RNaseL

RNaseL has been known to restrict a broad range of RNA and DNA viruses (315-324) and it has recently been shown to be involved in CpG- und UpA-mediated restriction of E7 (276).

In this study, RNaseL had some effect on reversing CpG-mediated restrictions in viral RNA expression in a transfection set up, but not when cells were infected with MVM (Figs. 4.6). There was no reversal of the CpG-mediated attenuation in the RNaseL knockout cells apparent during the immunocytochemistry, giving a further indication that RNaseL does not play an important role for MVM in the A549 system (Figs. 4.7-4.8).

5.5 Attenuation of TpA/UpA-high viruses

Previous studies on E7 and influenza A virus have shown that increasing UpA frequency in these viruses negatively affected viral fitness (270-272,485).

In MVM, increasing TpA frequency had little to no effect on viral fitness, despite mutating a comparable proportion of the viral genome (24.4 %) as in E7. The TpA-high MVM mutant replicated comparably to the wild-type (Fig. 3.1) whereas an UpA-high mutant of E7 showed 100-fold reduction in viral titres compared to wild-type virus in a similar assay. The TpA-high MVM mutant was furthermore comparable to wild-type virus in all other replication assays tested (Fig. 3.2-3.4) and did not show a reduction in its RNA expression (Fig. 3.6).

It is possible that different restriction mechanisms are involved in recognising and restricting CpG and TpA/UpA high sequences in parvoviruses. However, in E7 there has been evidence of a shared pathway involved in CpG and TpA/UpA restriction. Both CpG- and UpA-high E7 mutants are similarly restricted upon entry and by kinase inhibitor C16, but not through PKR, IFN or caspase activity C16 (270,485).

For this study, it was attempted to do similar experiments using C16 to explore the effects of kinase inhibition of MVM infection. However, C16 turned out to be highly cytotoxic for both A9 cells and NB324K cells and could thus not be used.

ZAP was found to be restricting both CpG and UpA high RNA in E7 (276). However, since there was no attenuation of the TpA-high MVM mutant in the fully permissive cell lines, the TpA-high mutant was not tested in the ZAP knockout system. It remains unclear how the TpA-high mutant remains unaffected by the antiviral proteins acting on viral RNAs.

5.6 Dose dependency of CpG attenuation

The CpG-mediated attenuation observed in this study was dose-dependent, where the CpG-high half mutants (58-59 CpGs) showed a 1000-fold reduction in replication compared to wild-type in A9 cells, but were still replicating a 1000-fold better than the CpG-high full mutant (117 CpGs) (Fig. 3.1). This less severe attenuation was also seen during the immunocytochemistry, were the half mutants showed an intermediate attenuation as well (Figs. 3.3-3.4). Interestingly, the CpG-high mutant with fixed G+C mutant (78 CpGs) was attenuated to a similar degree as the original CpG-high full mutant. This raises the question whether CpG recognition plateaus off at high CpG frequencies or whether other factors contribute to how effective the CpG-high viruses are restricted within infected cells.

This dose-dependent effect might also explain why other groups did not see an attenuation when they introduced lower numbers of CpGs into other parvoviruses (5-13 in H-1 and 9-29 in porcine parvovirus)(464,474). Interestingly, despite no measurable attenuation of the virus itself, they saw better immune responses to these viruses (464,474). Furhtermore, in those studies, kidney-derived cell lines (NB324K and HEK293) were used. These human kidney cell lines and the hamster kidney cell line BHK have shown less CpG-mediated restriction compared to fibroblast cell cultures, both in this project and in previous studies on E7 (276,485).

Interestingly, CpGs have been of concern in two very different areas of biomedical research.

CpG recognition is of interest in vaccine design, either as a TLR9-stimulating adjuvant in the form of CpG-ODNs (reviewed in depth by Bonam *et al.* (486)) or to produce attenuated live vaccine through genome recoding (reviewed in detail by Osterrieder and Kunec (487)). Introducing a large number of mutations as part of the attenuation process has a further advantage. It effectively prevents the occurrence of pathogenic reversion mutants, as for example seen in the live polio vaccine where only a few base changes dramatically change virulence (488).

In contrast, in gene therapy, CpGs are avoided. For one, this is to reduce possible immune responses to the administered DNA constructs (481,489). However, it has also been shown that CpGs within transfected or transduced DNA accelerated degradation and negatively affected retention of the DNA within the target cells (481,490-494).

In gene therapy, CpG-free sequences are advantageous over CpG-low sequences for long-term retention, yet in acute viral infections, the impact of small increases in CpG frequency appear less pronounced. These different facets of CpG-mediated restrictions raise some interesting questions. Do certain minimum requirements need to be met for cells to recognise CpG-high sequences and are those requirements different depending on context, e.g. an acute and persistent infection? Or alternatively, are all CpGs within DNA or RNA sequences affected, but it is less noticeable in acute viral infections (or even host cells) due to other mechanisms involved?

5.7 Effects of codon (de)optimisation and G+C content

Another potential mechanism for viral attenuation is the codon de-optimisation, especially since it is a side-effect of increasing CpGs or TpAs/UpAs frequencies. This has been proposed to lead to attenuation in a range of RNA viruses (495-498). Codon-de-optimisation can negatively affect translation and impact other part of the viral life cycles in turn.

In this study, the CpG-high MVM mutant was severely attenuated throughout its replication cycle (Figs. 3.1-3.4, Fig. 3.6), but the mutated VP protein translated similarly to wild-type in a cell free assay (Fig. 3.5).

Furthermore, a codon-optimised mutant of MVM was nearly as severely attenuated as the CpG-high mutant despite showing only a minor reduction in mRNA levels (Figs. 3.1-3.4, Fig. 3.6). Co-incidentally, when the region was codon-optimised, CpG frequencies were increased by tenfold due to the extreme suppression of CpGs in this region in the wild-type virus (Table 2.1). This lead to the codon-optimised mutant having a similar amount of CpGs as the CpG-high half mutants, though it is more spread out in the codon-optimised mutant due to the size of the insert region. For comparison, the observed over expected ratio of CpGs was approximately 1.6 for all CpG-high mutants within their mutated regions, while the wild-type ratio was close to 0.1 for the mutated region. The codon-optimised mutant had a CpG observed over expected ratio 0.6, thus still showing a CpG suppression though less than the wildtype virus.

This is consistent with the observation that codon optimisation also leads to attenuation in E7 due to increases in CpG and UpA dinucleotides (270,485). It is nevertheless surprising, as codon optimisation is widely used to enhance transgene expression in various DNA or viral vectors, including parvoviral adenovirus-associated virus AAV vectors (499-501). However, in viruses with very strong CpG suppression, codon optimisation might confer a disadvantage due to the increase in CpGs. This was observed in this project, where the viral gene expression of MVM was severely restricted, except in a cell free translation assay (Figs. 3.1-3.5). A similar effect has been observed in adenovirus, where most viral proteins are codon-optimised, yet optimising the non-optimised fiber protein led to a decrease in viral fitness (502).

In MVM, this restriction was not ZAP or RNaseL dependent (Figs. 4.7-4.8). The codon-optimised MVM mutant was restricted to a similar extent in both parental and ZAP or RNAseL knockout cells. Overall, the codon-optimised MVM mutant was less attenuated in these cell lines compared to a similarly designed codon-optimised E7 mutant, which showed marked attenuation in A549 cells previously (271). While it is unclear how these discrepancies arise, they might be connected to G+C content or differential recognition thereof.

When the codon-optimised mutant was created, the G+C content was raised from a native 42 % to 61 % alongside the codon optimisation. To explore whether changes in G+C content could affect viral fitness, a CpG-low codon-optimised mutant was created, as it was not technically possible to create a codon-optimised mutant with 42% G+C content. Remarkably, the CpG-low, codon-optimised mutant showed an intermediate

level of attenuation (Fig. 3.1), but translated well in a cell free translation assay (Fig. 3.5).

For MVM, G+C low (GCmin, 33 % G+C content) and G+C high (GCmax, 57 % G+C content) mutants were designed to see whether they would be attenuated without any other alterations to them. Unfortunately, this was not feasible in the time frame of this project and Jeremy Ratcliffe took over this work as part of his DPhil studies. Both the G+C low and G+C high MVM mutants were attenuated in A9 cells. The data for the G+C high mutant has been included in Loew et al. (1) along with the work presented in this thesis. While this concept has to be further explored to draw better conclusions from the data, it appears that both CpG frequency and G+C content influence viral fitness in MVM. This might also explain the observed differences in attenuation in A9 and A549 cells, as G+C content alterations might have different effects on RNA and DNA viruses.

5.8 Future work

MVM

There are two areas to further explore with this system, confirming the involvement of ZAP in the CpG-mediated restriction of MVM in a mouse system and exploring the effects of changing G+C content in more detail.

Ideally, ZAP knockout cells would be used to confirm the results obtained in the human A549 cell lines. While a collaborator was able to find ZAP knockout MEFs in China, their export restrictions did not allow for them to be shipped into the UK.

There are a few other options for available to knock out or knock down ZAP. Transient transfection with small interfering RNA (siRNA) is not suitable in A9 cells due to low transfection efficiency. However, short hairpin RNA (shRNA) could be transfected into cells as part of an expression plasmid also encoding neomycin or another antibiotic resistance allowing positive cell selection and prolonged shRNA expression. While it would take a couple of weeks to generate these cell lines, it would still considerably faster and far less labour intensive than CRISPR-CAS9 gene knockout, as single cell clonal lines need to be selected, grown up and verified.

As an alternative to or in conjunction with the above approach, ZAP binding assays could be performed with A9 cell lysate to determine whether ZAP (preferentially) binds CpG-rich RNA. This could be complemented with a pull-down essay to find binding partners for both viral RNA and DNA. While this approach is more labour intensive, it is open ended and could bring new leads towards other restriction mechanisms.

In regards to the G+C content mutant, there is currently very little information available both from this project and on a broader spectrum in virology.

It would thus be interesting to explore further which parts of the life cycle are affected, using a similar approach as for the 1st and 2nd generation MVM mutants.

MPyV

As the pilot work into MPyV has established it as a viable model system, it offers the opportunity to explore the different stages of CpG repression by altering early and late genes of the virus separately. Especially altering just the capsid region allows for the exploration of CpG mediated attenuation on the DNA level, as early translation and translation will not be affected by the mutations unlike in MVM. Polyomaviruses can cause persistent infections, using this system would further allow to study how changes in dinucleotide frequencies affect persistency.

This system would also present an excellent opportunity to study the effect of G+C content changes. Different parts of the viral life cycle can be targeted and examined more closely by altering either the early or the late genes.

5.9 Conclusion

In this study it was shown that increasing CpG dinucleotide frequencies or codon optimising leads to severe attenuation in MVM while a increasing TpA dinucleotides or scrambling part of the genome showed no or minimal attenuation compared to the wild-type.

In summary, the study provides compelling evidence for the action of cellular restriction of CpG modified mutants of MVM at the RNA level while TLR9-mediated recognition of non-methylated genomic DNA sequences was excluded in this context. This affected the transcription and translation of the non-structural proteins and all subsequent stages of the viral life cycle.

The study provides evidence for a commonality in restriction mechanisms and their effects on driving down CpG frequencies in many RNA viruses, retroviruses and small DNA viruses infecting vertebrates.

6 Methods

6.1 Mutant design and generation 134

In silico design of codon-optimised, GpC-, CpG- and UpA-modified viruses 134

Cloning and plasmid propagation 134

6.2 Cell culture and infectious virus work 137

Cell culture, cell lines and reagents 137

Virus recovery 137

Virus Titration 138

Virus infections 138

Serial passage 139

6.3 Molecular techniques 140

In vitro transcription and translation 140

RNA expression experiments 140

TLR9 stimulation 140

DNA and RNA extractions 141

PCR and Quantitative real-time PCR 141

Immunofluorescence 142

Immunoblotting 143

Miseq sequencing 144

6.1 Mutant design and generation

In silico design of codon-optimised, GpC-, CpG- and UpA-modified viruses

SSE (439) was used to determine suitable areas for mutation. Synonymous coding variability was calculated using the program SequenceDist within the package. A region with no suppression of synonymous sequences variability within the capsid ORF (MVM, MPyV) or large T-antigen (MPyV) was chosen as this was less likely to contain cryptic replication elements or undocumented alternative reading frames. Sequence Mutate in SSE was then used to produce CDLR, CpG-high, TpA-high and GpC-high sequences based on that region. For the codon-optimised construct, the GeneArt codon optimisation tool was used. For the codon-optimised CpG-low construct, the online program CooL was used (453). All constructs were synthesised by GeneArt into a commercial plasmid backbone.

Cloning and plasmid propagation

To generate the MVM mutants, the wild-type infectious clone pMM984 (kind gift from David Pintel, Missouri), as well as the insert containing plasmids, were cut with *Bpu*10I, *Hpa*I and *Xba*I as needed to generate sticky end fragments. These were gel purified (QIAquick Gel Extraction Kit, Qiagen, UK), ligated with T4 ligase (New England Biosciences UK, UK) and transformed into SURE 2 Supercompetent cells (Agilent, USA) according to the manufacturers' protocols. Bacterial colonies were grown in LB medium and plasmid preparations were purified using the HiSpeed Plasmid Midi Kit (Qiagen, Germany) and subsequently used to produce virus stocks.

For quantification purposes, the MVM genome was excised out of the pMM984 plasmid with *Bam*HI and gel purified.

For the MPyV mutants, the wild-type infectious clone pBS-loxPintron_ (kind gift from Jitka Forstová, Prague) and the synthetic inserts were cut with *NsiI*, *KflI* and *EcoRV,* as needed to generate sticky end fragments.

For the R3 subclones, the R3 RNA region of MVM was cloned into a pCAG-loxPSTOPloxP-ZsGreen vector (addgene #51269, kind gift from Lynn Dustin, Oxford) using *SwaI* and *SacI* restriction sites. The R3 region was amplified and restriction sites were added using two rounds of PCR with a Phusion High Fidelity DNA Polymerase (New England Biosciences, UK) according to the manufacturer's guidelines (primers:
FW1 5'-ATAATTTAAATCACCATTCACGACACCGAAAAGTAC-3',
RV1 5'-ATAGAGCTCTTAGTAAGTATTTCTAGCAACAGG-3';
FW2 5'-GATCGTAACGATAATTTAAATCACC-3',
RV2 5'-GATCGTAACGATAATTTAAATCACC-3'). PCR products and vector were processed as above, but grown in One Shot TOP10 Chemically Competent *E. coli* (ThermoFisher Scientific, UK).

For testing genetic stability using sanger sequencing, a region within the NS and VP genes was amplified via PCR with Phusion High Fidelity DNA Polymerase (New England Biosciences, UK) according to the manufacturer's guidelines (NS primers:
FW_MVMp_seq1 5'-CAAAAAAGAGCTGCAGGAGG -3',
RV_MVMp_NSseq 5'-TGGTTACTGTGGTTTCAACCG-3';

VP primers: FW_MVMp_ins_seq2 5’-ATAGATTCTTGGGTGACGGC-3’, RV_MVMp_VPseq 5’-GGCTTCAAAGTCATTGTGTGG-3’). PCR products were used with the pGET PCR Cloning Kit (ThermoFisher Scientific, UK) according to the manufacturer’s instructions.

6.2 Cell culture and infectious virus work

Cell culture, cell lines and reagents.

A9 cells(444) and NB324K cells(443) were a kind gift from Peter Tattersall, Yale. 3T6 and T2 cells were a kind gift from Jitka Forstová, Prague. These cells were grown in complete Dulbecco's modified Eagle's medium (DMEM/F-12, supplemented with 5 % foetal calf serum (FCS) and penicillin/streptomycin) and incubated at 37 °C and 5 % CO_2. A549 cells, A549 B8 ZAP k/o cells and A549 E10 RNaseL k/o cells (276), MEF cells (kind gift from Jan Rehwinkel, Oxford) and BHKs were grown as above but were supplemented with 10 % FCS.

Virus recovery

To produce MVM virus stocks, NB324K cells were transfected using Lipofectamine 2000 (ThermoFisher Scientific, UK) according to the manufacturer's recommendation.

10^6 NB324K cells were either transfected with 2 µg DNA (CpG high mutants, CodOpt), 1 µg DNA (CDLR, UpA-H, GpC-H, CodOpt CpG-L) or 500 ng DNA (WT) each. The next day, cells were split using trypsin and transferred into a T175 tissue culture flask. A non-transfected control was grown in parallel. After 3-4 days, cells were harvested by scraping the cells into the culture medium. Cells and medium were spun down and separated. The cell pellet was resuspended in 3 ml vTE (50 mM Tris-hydrochloride, 0.5 mM EDTA, pH 8.7 (368) and freeze-thawed three times. The cell lysate was spun down (1500 rpm, 20 min, 4 °C). The cleared lysate was sterile filtered, aliquoted, frozen at -80 °C and titrated.

For MPyV, a similar protocol was used. 10^6 T2 cells were transfected with 2 µg DNA (1 µg ploxPintron_ DNA and 1 µg pCAG Cre). This was done to ensure sufficient Cre expression for effective Cre-plox mediated excision of the bacterial backbone. The day after transfection, cells were transferred into a T175 tissue culture flasks and harvested after 3-5 days as above.

Virus Titration

$TCID_{50}$s were predominantly used to quantify infectious particles in MVM. 90 µl NB324K cell suspension containing 3000 cells were mixed with 10 µl viral suspension (serial 10-fold dilutions in complete DMEM) in a 96-well plate. After initial optimisation to determine the most robust scoring, the $TCID_{50}$ score was determined 10 days post infection using the Reed-Muench method(503).

Initially plaque assays were used to compare results between plaque assays and TCID50 assays. Cells were seeded at 4x10^4 cells/well (NB324K cells) or 6x10^4 cells/well (A9 cells) in 12-well plates. The next day, 200 µl viral suspension (serial 10-fold dilutions in PBS/2 % FCS) was added to each well and gently shaken intermittently. After an hour, 1 ml of a 1:1 mix of 2x MEM/4% FCS and 1.2% Avicell in PBS was added. After five days, the cells were fixed by adding 1.2 ml 8 % paraformaldehyde for 30 min and stained with 0.1% toluidine blue.

Virus infections

Cells were seeded over night before being infected. Plate sizes were chosen in accordance with individual experiment needs. Cell numbers were chosen according to experiment length, seeding at roughly 10 % confluence for experiments having total

incubation periods of 48 h or longer, while seeding at roughly 25 % confluence for experiments with a total incubation period of 24 h or less. Cells were infected with either equal $TCID_{50}$ units or equal viral genome copy numbers. Viral stocks were diluted in Opti-MEM. Cell supernatant was replaced by the viral suspension and incubated for 1 h at 37 °C, while shaking the plate occasionally. Afterwards, the cells were washed 3x with PBS and fresh DMEM with 5 or 10 % FCS was added as appropriate.

Serial passage

A9 cells were cultured in 25 cm^2 tissue culture flasks and were infected with virus stocks (0.5 $TCID_{50}$ units per cell). For the serial passage, viral supernatant was added to fresh A9 cells twice a week in a dilution that would allow complete CPE within 3-4 days (WT: 1:50, TpA-H: 1:25, CpG-H1: 1:2, CpG-H2: 1:2).

6.3 Molecular techniques

In vitro transcription and translation

The synthetic plasmids (pMKRQ_MVMp_synth) used the reverse engineering of the MVM mutants were linearised using *ApaL*I for 4h at 37°C and purified (HiSpeed Plasmid Midi Kit, Qiagen, UK). 1 µg of DNA was used as template to transcribe, cap and poly-adenylate messenger RNAs using the mMESSAGE mMACHINE T7 Transcription Kit (ThermoFisher Scientific, UK) according to the manufacturer's description. The newly made RNA was purified using the RNA Clean & Concentrator-25 kit (Zymo Research, USA). RNA integrity was determined via the Bioanalyser, using the RNA 6000 Pico Kit (Agilent, USA). 2 µg of RNA were used in the Rabbit Reticulocyte Lysate System, Nuclease Treated (Promega, UK) according to the manufacturer's instructions. Samples were then immnoblotted.

RNA expression experiments

Cells were seeded and used for transfections or infections the next day. Cells were either transfected with 500 ng DNA per 100,000 cells using Lipofectamine 2000 (ThermoFisher Scientific, UK) according to the manufacturer's guidelines. 5 µg/ml Aphidicolin (Santa Cruz Biotechnology, USA) were added 4 h post transfection.

TLR9 stimulation

For TLR9 stimulation, cells were stimulated with 1 µg/ml purified E. coli DNA (InvivoGen, UK) for 18 h, with or without 2 h pre-treatment with 5 µM ODN2088 (InvivoGen, UK).

For TLR9 inhibition, cells were pre-treated with 5 μM ODN2088 for 2 h, before being infected with 1000 virus genome copies per cell for 18 h.

DNA and RNA extractions

Viral stocks were pre-treated with DNAseI (Roche, Switzerland) prior to DNA extraction. 0.1 mg/ml DNase and 5 mM $MgCl_2$ were added to the samples and incubated for 25 °C for 25 min. Control (no-virus) lysate was spiked with 10^9 *Bam*HI-excised MVM copies to evaluate DNAse digestion and DNA recovery rate for MVM stocks. DNA was extracted and purified with either the DNA easy Blood and Tissue Kit (Qiagen, UK) or the Easy Viral DNA Kit (Zymo Research, USA) according to the manufacturers' guidelines. RNA was extracted and purified using the RNAeasy Mini Kit (Qiagen, UK) according to the manufacturer's guidelines. DNA and RNA concentrations were then measured using the Qubit (ThermoFisher Scientific, UK).

PCR and Quantitative real-time PCR

For PCRs, either GoTaq Flexi DNA Polymerase (Promega, UK) or Phusion High Fidelity DNA Polymerase (New England Biosciences, UK) were used according to the manufacturer's guidelines. Phusion High Fidelity DNA Polymerase was used whenever the PCR product was subsequently sequenced. For the MVM stem loop PCR, MVM_FW_seq6n (5'-CAAGATAGTTAGTTGGTTAATGTTAGATAG-3', binds bases 4703-4732 of the MVM genome) and PQE30 forward (5'- CCCGAA AAGTGCCACCTG-3', binds to pRB322 backbone approx. 100bp downstream of MVM) were used.

For the quantitative real-time PCR, the QuantiFast SYBR Green RT-PCR Kit (RNA) or QuantiFast SYBR Green PCR Kit (DNA) (Qiagen, UK) were used, loading 80 ng RNA or DNA per well. For viral DNA quantification, primers in the NS region of MVM (FW: 5'-AGTTTGCCATGCTATTTGC-3' and RV: 5'-AATCTGCTTGCTGCCTTT-3', from Zhan *et al.* (460)) or the Large T-antigen region of MPyV (FW: 5'-CGCACATACTGCTGGAAGAAGA-3', RW: 5'-TCTTGGTCGCTTTCTGGATACAG-3', from Kemball *et al.* (421)) with a standard curve derived from serially diluted *Bam*HI-excised MVMp (MVM) or undigested ploxPintron_ (MPyV).

For viral RNA quantification, primers targeting MVM R1 (FW: 5'- CCAACTCCTATAAATTTACTAG -3', RV: 5'-ACCAAGTGATTTCAGGCCTTAG-3', from Li *et al.*(377)) or R3 (FW: CACCATTCACGACACCGAAAAG, same RV as R1) were used. For IL-1β, FW: 5'-GCAACTGTTCCTGAACTCAACT-3'and RV: 5'-ATCTTTTGGGGTCCGTCAACT-3' were used (504). Samples were normalised to the species appropriate GAPDH (human FW: 5'-GAAATCCCATCACCATCTTCCAGG-3', RV: 5'-GAGCCCCAGCCTTCTCCATG-3', from Ahn *et al. (505)*; mouse FW: 5'- CGACTTCAACAGCAACTCCCACTCTTCC -3', FW: 5'- TGGGTGGTCCAGGGTTTCTTACTCCTT -3', from Banerjee *et al*.(506)) using the ΔΔCT method(507).

Immunofluorescence

Cells were seeded into cellview cell culture dishes (VWR, USA) or Falcon 96-well imaging microplates (Corning, USA). The next day, cells were infected with 1000 genome copies per cell. After 18 h incubation, the cells were submerged into 4 % paraformaldehyde for 10 min, neutralised with 0.1 M glycine for 5 min and permealised in 0.1 % Triton in PBS for 20 min. The cells were then blocked with 2 % FCS in PBS for 20 min washed 3x with

PBS and stained for 1 h with the primary antibodies (Mab3(508): 1:100, monoclonal mouse antibody; NS1/2 (386): 1:400, polyclonal rabbit antibody, both kind gifts from Peter Tattersall). Subsequently, samples were washed 3x with PBS and stained with the secondary antibodies (1:2000 anti-rabbit Alexa Fluor 488 and either 1:2000 anti-mouse Alexa Fluor 594 or 647; all from ThermoFisher Scientific, UK), stained with Hoechst 33342 (1:2000, ThermoFisher Scientific, UK) for 10 min, and washed 3x in PBS. Slides were analysed using a Leica DMIRE2 microscope and Q capture pro 7 software or an Olympus IX73 Fluorescence microscope and the accompanying software. Post-acquisition analysis was conducted using the Imaris (Bitplane) or Fiji (v1.49) software (509). The analyse particles module of Fiji (ImageJ) was used to substract background signal. NS and VP positive cells were counted in ImageJ and divided by the total cell determined with the Nucleus Counter plugin (MachBiophotonics ImageJ). (*Image processing and quantification done by Niluka Goonawardane*).

Immunoblotting

Cells were infected with 1000 viral genome copies per cell for 18h, washed with PBS, scraped of the cell culture dish and resuspended in Laemmli buffer (200 µl per $5x10^5$ cells). 30 µl in vitro translation products were diluted with 15 µl 4x Laemmli buffer (Bio-rad, UK). Samples were incubated for 5 min at 95°C and spun down. 10 µl lysate were loaded onto a 10-Well 4-15 % gradient mini protean TGX SDS-PAGE gel (Bio-rad, UK) and run for 4 h at 40 V. Gels were then transferred using the Trans-blot Turbo transfer system (Bio-rad, UK) according to the manufacturer's guidelines. Membranes were blocked in 5 % skimmed milk powder in 0.1 % PBS-Tween, incubated over night at 4 °C with primary antibodies (Tatt3(508): 1:10000; NS1/2(386): 1:10000, both polyclonal

rabbit antibodies and kind gifts from Peter Tattersall; GAPDH (ab70699, abcam, UK)). Membranes were washed 3x in 0.1 % PBS-Tween and incubated for 1 h at 25 °C with goat anti-rabbit IgG H&L HRP antibody (ab205718, abcam, UK). Membranes were washed 3x in 0.1 % PBS-Tween, substrate was added (Amersham ECL Prime Western Blotting Detection Agent, GE Lifesciences) and images were exposed using the ImageQuant LAS4000 (GE Healthcare, UK). Images were analysed using ImageJ. The background was digitally subtracted and bands were quantified by measuring the mean grey value over equal areas.

Miseq sequencing

Purified DNA underwent second strand synthesis using Sequenase Version 2.0 (ThermoFisher, UK) and random primer 9 (New England BioLabs, UK). Libraries were prepared from 1ng of the resulting double stranded DNA using the Nextera XT kit (Illumina, UK) and sequenced on the MiSeq platform with reagent kit v2 (500-cycles). After adapter removal and quality trimming using Trimmomatic (510), reads were aligned against the reference sequence (Genbank no. J02275) and the derived mutant sequences using Bowtie 2 (511) (*Data processing and analysis done by Dung Nguyen and Peter Simmonds*).

7 Appendices

7 Appendices 145

7.1 References 146

7.2 DNA sequences 187

MVM Inserts 187

pCAG_MVMp expression plasmid 194

In vitro translation plasmid 199

MPyV Inserts 202

7.3 Use of a small DNA virus model to investigate mechanisms of CpG dinucleotide-induced attenuation of virus replication - Loew et al., 2020, J Gen Virol 204

7.1 References

1. Loew, L., Goonawardane, N., Ratcliff, J., Nguyen, D. and Simmonds, P. (2020) Use of a small DNA virus model to investigate mechanisms of CpG dinucleotide-induced attenuation of virus replication.
2. Josse, J., Kaiser, A.D. and Kornberg, A. (1961) Enzymatic synthesis of deoxyribonucleic acid. VIII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. *J Biol Chem*, **236**, 864-875.
3. Swartz, M.N., Trautner, T.A. and Kornberg, A. (1962) Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J Biol Chem*, **237**, 1961-1967.
4. Russell, G.J., Walker, P.M., Elton, R.A. and Subak-Sharpe, J.H. (1976) Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J Mol Biol*, **108**, 1-23.
5. Bird, A.P. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res*, **8**, 1499-1504.
6. Simmen, M.W. (2008) Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics*, **92**, 33-40.
7. Jabbari, K. and Bernardi, G. (2004) Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene*, **333**, 143-149.
8. Hendrich, B. and Tweedie, S. (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet*, **19**, 269-277.
9. Tweedie, S., Charlton, J., Clark, V. and Bird, A. (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol*, **17**, 1469-1475.
10. Regev, A., Lamb, M.J. and Jablonka, E. (1998) The role of DNA methylation in invertebrates: Developmental regulation or genome defense? *Mol Biol Evol*, **15**, 880-891.
11. Karlin, S., Doerfler, W. and Cardon, L.R. (1994) Why Is Cpg Suppressed in the Genomes of Virtually All Small Eukaryotic Viruses but Not in Those of Large Eukaryotic Viruses. *J Virol*, **68**, 2889-2897.
12. Rima, B.K. and McFerran, N.V. (1997) Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *J Gen Virol*, **78 (Pt 11)**, 2859-2870.
13. Shackelton, L.A., Parrish, C.R. and Holmes, E.C. (2006) Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol*, **62**, 551-563.
14. Beutler, E., Gelbart, T., Han, J.H., Koziol, J.A. and Beutler, B. (1989) Evolution of the Genome and the Genetic-Code - Selection at the Dinucleotide Level by Methylation and Polyribonucleotide Cleavage. *P Natl Acad Sci USA*, **86**, 192-196.

15. Duan, J.B. and Antezana, M.A. (2003) Mammalian mutation pressure, synonymous codon choice, and mRNA degradation. *J Mol Evol*, **57**, 694-701.
16. Lester, S.N. and Li, K. (2014) Toll-like receptors in antiviral innate immunity. *Journal of molecular biology*, **426**, 1246-1264.
17. Thompson, M.R., Kaminski, J.J., Kurt-Jones, E.A. and Fitzgerald, K.A. (2011) Pattern recognition receptors and the innate immune response to viral infection. *Viruses*, **3**, 920-940.
18. Kawasaki, T. and Kawai, T. (2014) Toll-like receptor signaling pathways. *Front Immunol*, **5**, 461.
19. Jiang, S., Li, X., Hess, N.J., Guan, Y. and Tapping, R.I. (2016) TLR10 Is a Negative Regulator of Both MyD88-Dependent and -Independent TLR Signaling. *Journal of immunology (Baltimore, Md. : 1950)*, **196**, 3834-3841.
20. Krieg, A.M., Yi, A.K., Matson, S., Waldschmidt, T.J., Bishop, G.A., Teasdale, R., Koretzky, G.A. and Klinman, D.M. (1995) CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature*, **374**, 546-549.
21. Ohto, U., Shibata, T., Tanji, H., Ishida, H., Krayukhina, E., Uchiyama, S., Miyake, K. and Shimizu, T. (2015) Structural basis of CpG and inhibitory DNA recognition by Toll-like receptor 9. *Nature*, **520**, 702-705.
22. Kawane, K., Ohtani, M., Miwa, K., Kizawa, T., Kanbara, Y., Yoshioka, Y., Yoshikawa, H. and Nagata, S. (2006) Chronic polyarthritis caused by mammalian DNA that escapes from degradation in macrophages. *Nature*, **443**, 998-1002.
23. Pohar, J., Kužnik Krajnik, A., Jerala, R. and Benčina, M. (2015) Minimal Sequence Requirements for Oligodeoxyribonucleotides Activating Human TLR9. *The Journal of Immunology*, **194**, 3901.
24. Rutz, M., Metzger, J., Gellert, T., Luppa, P., Lipford, G.B., Wagner, H. and Bauer, S. (2004) Toll-like receptor 9 binds single-stranded CpG-DNA in a sequence- and pH-dependent manner. *Eur J Immunol*, **34**, 2541-2550.
25. Chan, M.P., Onji, M., Fukui, R., Kawane, K., Shibata, T., Saitoh, S., Ohto, U., Shimizu, T., Barber, G.N. and Miyake, K. (2015) DNase II-dependent DNA digestion is required for DNA sensing by TLR9. *Nat Commun*, **6**, 5853.
26. Barton, G.M., Kagan, J.C. and Medzhitov, R. (2006) Intracellular localization of Toll-like receptor 9 prevents recognition of self DNA but facilitates access to viral DNA. *Nat Immunol*, **7**, 49-56.
27. Vollmer, J., Weeratna, R., Payette, P., Jurk, M., Schetter, C., Laucht, M., Wader, T., Tluk, S., Liu, M., Davis, H.L. *et al.* (2004) Characterization of three CpG oligodeoxynucleotide classes with distinct immunostimulatory activities. *Eur J Immunol*, **34**, 251-262.
28. Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K. *et al.* (2000) A Toll-like receptor recognizes bacterial DNA. *Nature*, **408**, 740-745.

29. Tabeta, K., Georgel, P., Janssen, E., Du, X., Hoebe, K., Crozat, K., Mudd, S., Shamel, L., Sovath, S., Goode, J. *et al.* (2004) Toll-like receptors 9 and 3 as essential components of innate immune defense against mouse cytomegalovirus infection. *Proc Natl Acad Sci U S A*, **101**, 3516-3521.
30. Fiola, S., Gosselin, D., Takada, K. and Gosselin, J. (2010) TLR9 contributes to the recognition of EBV by primary monocytes and plasmacytoid dendritic cells. *J Immunol*, **185**, 3620-3631.
31. Lund, J., Sato, A., Akira, S., Medzhitov, R. and Iwasaki, A. (2003) Toll-like receptor 9-mediated recognition of Herpes simplex virus-2 by plasmacytoid dendritic cells. *J Exp Med*, **198**, 513-520.
32. Lima, G.K., Zolini, G.P., Mansur, D.S., Freire Lima, B.H., Wischhoff, U., Astigarraga, R.G., Dias, M.F., das Graças Almeida Silva, M., Béla, S.R., do Valle Antonelli, L.R. *et al.* (2010) Toll-like receptor (TLR) 2 and TLR9 expressed in trigeminal ganglia are critical to viral control during herpes simplex virus 1 infection. *The American journal of pathology*, **177**, 2433-2445.
33. Samuelsson, C., Hausmann, J., Lauterbach, H., Schmidt, M., Akira, S., Wagner, H., Chaplin, P., Suter, M., O'Keeffe, M. and Hochrein, H. (2008) Survival of lethal poxvirus infection in mice depends on TLR9, and therapeutic vaccination provides protection. *J Clin Invest*, **118**, 1776-1784.
34. Tatematsu, M., Seya, T. and Matsumoto, M. (2014) Beyond dsRNA: Toll-like receptor 3 signalling in RNA-induced immune responses. *Biochem J*, **458**, 195-201.
35. Tatematsu, M., Nishikawa, F., Seya, T. and Matsumoto, M. (2013) Toll-like receptor 3 recognizes incomplete stem structures in single-stranded viral RNA. *Nat Commun*, **4**, 1833.
36. Hornung, V., Barchet, W., Schlee, M. and Hartmann, G. (2008) RNA recognition via TLR7 and TLR8. *Handbook of experimental pharmacology*, 71-86.
37. Hornung, V., Guenthner-Biller, M., Bourquin, C., Ablasser, A., Schlee, M., Uematsu, S., Noronha, A., Manoharan, M., Akira, S., de Fougerolles, A. *et al.* (2005) Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nature medicine*, **11**, 263-270.
38. Diebold, S.S., Massacrier, C., Akira, S., Paturel, C., Morel, Y. and Reis e Sousa, C. (2006) Nucleic acid agonists for Toll-like receptor 7 are defined by the presence of uridine ribonucleotides. *Eur J Immunol*, **36**, 3256-3267.
39. Gorden, K.B., Gorski, K.S., Gibson, S.J., Kedl, R.M., Kieper, W.C., Qiu, X., Tomai, M.A., Alkan, S.S. and Vasilakos, J.P. (2005) Synthetic TLR agonists reveal functional differences between human TLR7 and TLR8. *J Immunol*, **174**, 1259-1268.
40. Oliveira-Nascimento, L., Massari, P. and Wetzler, L.M. (2012) The Role of TLR2 in Infection and Immunity. *Front Immunol*, **3**, 79-79.

41. Oosting, M., Cheng, S.-C., Bolscher, J.M., Vestering-Stenger, R., Plantinga, T.S., Verschueren, I.C., Arts, P., Garritsen, A., van Eenennaam, H., Sturm, P. *et al.* (2014) Human TLR10 is an anti-inflammatory pattern-recognition receptor. *P Natl Acad Sci USA*, **111**, E4478-E4484.
42. Hayashi, F., Smith, K.D., Ozinsky, A., Hawn, T.R., Yi, E.C., Goodlett, D.R., Eng, J.K., Akira, S., Underhill, D.M. and Aderem, A. (2001) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, **410**, 1099-1103.
43. Murawski, M.R., Bowen, G.N., Cerny, A.M., Anderson, L.J., Haynes, L.M., Tripp, R.A., Kurt-Jones, E.A. and Finberg, R.W. (2009) Respiratory syncytial virus activates innate immunity through Toll-like receptor 2. *J Virol*, **83**, 1492-1500.
44. Bieback, K., Lien, E., Klagge, I.M., Avota, E., Schneider-Schaulies, J., Duprex, W.P., Wagner, H., Kirschning, C.J., Ter Meulen, V. and Schneider-Schaulies, S. (2002) Hemagglutinin protein of wild-type measles virus activates toll-like receptor 2 signaling. *J Virol*, **76**, 8729-8736.
45. Ge, Y., Mansell, A., Ussher, J.E., Brooks, A.E., Manning, K., Wang, C.J. and Taylor, J.A. (2013) Rotavirus NSP4 Triggers Secretion of Proinflammatory Cytokines from Macrophages via Toll-Like Receptor 2. *J Virol*, **87**, 11160-11167.
46. Kurt-Jones, E.A., Popova, L., Kwinn, L., Haynes, L.M., Jones, L.P., Tripp, R.A., Walsh, E.E., Freeman, M.W., Golenbock, D.T., Anderson, L.J. *et al.* (2000) Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. *Nat Immunol*, **1**, 398-401.
47. Rassa, J.C., Meyers, J.L., Zhang, Y., Kudaravalli, R. and Ross, S.R. (2002) Murine retroviruses activate B cells via interaction with toll-like receptor 4. *Proc Natl Acad Sci U S A*, **99**, 2281-2286.
48. Jude, B.A., Pobezinskaya, Y., Bishop, J., Parke, S., Medzhitov, R.M., Chervonsky, A.V. and Golovkina, T.V. (2003) Subversion of the innate immune system by a retrovirus. *Nat Immunol*, **4**, 573-578.
49. Compton, T., Kurt-Jones, E.A., Boehme, K.W., Belko, J., Latz, E., Golenbock, D.T. and Finberg, R.W. (2003) Human cytomegalovirus activates inflammatory cytokine responses via CD14 and Toll-like receptor 2. *J Virol*, **77**, 4588-4596.
50. Boehme, K.W., Guerrero, M. and Compton, T. (2006) Human cytomegalovirus envelope glycoproteins B and H are necessary for TLR2 activation in permissive cells. *J Immunol*, **177**, 7094-7102.
51. Lauw, F.N., Caffrey, D.R. and Golenbock, D.T. (2005) Of mice and man: TLR11 (finally) finds profilin. *Trends in immunology*, **26**, 509-511.
52. Shukla, D., Chandel, H.S., Srivastava, S., Chauhan, P., Pandey, S.P., Patidar, A., Banerjee, R., Chattopadhyay, D. and Saha, B. (2018) TLR11 or TLR12 silencing reduces Leishmania major infection. *Cytokine*, **104**, 110-113.

53. Raetz, M., Kibardin, A., Sturge, C.R., Pifer, R., Li, H., Burstein, E., Ozato, K., Larin, S. and Yarovinsky, F. (2013) Cooperation of TLR12 and TLR11 in the IRF8-dependent IL-12 response to Toxoplasma gondii profilin. *J Immunol*, **191**, 4818-4827.
54. Li, X.D. and Chen, Z.J. (2012) Sequence specific detection of bacterial 23S ribosomal RNA by TLR13. *Elife*, **1**, e00102.
55. Oldenburg, M., Krüger, A., Ferstl, R., Kaufmann, A., Nees, G., Sigmund, A., Bathke, B., Lauterbach, H., Suter, M., Dreher, S. *et al.* (2012) TLR13 recognizes bacterial 23S rRNA devoid of erythromycin resistance-forming modification. *Science*, **337**, 1111-1115.
56. Shi, Z., Cai, Z., Sanchez, A., Zhang, T., Wen, S., Wang, J., Yang, J., Fu, S. and Zhang, D. (2011) A novel Toll-like receptor that recognizes vesicular stomatitis virus. *J Biol Chem*, **286**, 4517-4524.
57. Brisse, M. and Ly, H. (2019) Comparative Structure and Function Analysis of the RIG-I-Like Receptors: RIG-I and MDA5. *Front Immunol*, **10**, 1586.
58. Rehwinkel, J. and Gack, M.U. (2020) RIG-I-like receptors: their regulation and roles in RNA sensing. *Nature Reviews Immunology*, **20**, 537-551.
59. Venkataraman, T., Valdes, M., Elsby, R., Kakuta, S., Caceres, G., Saijo, S., Iwakura, Y. and Barber, G.N. (2007) Loss of DExD/H Box RNA Helicase LGP2 Manifests Disparate Antiviral Responses. *The Journal of Immunology*, **178**, 6444.
60. Satoh, T., Kato, H., Kumagai, Y., Yoneyama, M., Sato, S., Matsushita, K., Tsujimura, T., Fujita, T., Akira, S. and Takeuchi, O. (2010) LGP2 is a positive regulator of RIG-I– and MDA5-mediated antiviral responses. *Proceedings of the National Academy of Sciences*, **107**, 1512.
61. Bruns, Annie M., Leser, George P., Lamb, Robert A. and Horvath, Curt M. (2014) The Innate Immune Sensor LGP2 Activates Antiviral Signaling by Regulating MDA5-RNA Interaction and Filament Assembly. *Molecular Cell*, **55**, 771-781.
62. Ren, Z., Ding, T., Zuo, Z., Xu, Z., Deng, J. and Wei, Z. (2020) Regulation of MAVS Expression and Signaling Function in the Antiviral Innate Immune Response. *Front Immunol*, **11**, 1030-1030.
63. Kowalinski, E., Lunardi, T., McCarthy, A.A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D. and Cusack, S. (2011) Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell*, **147**, 423-435.
64. Berke, I.C. and Modis, Y. (2012) MDA5 cooperatively forms dimers and ATP-sensitive filaments upon binding double-stranded RNA. *Embo j*, **31**, 1714-1726.
65. Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T.S., Fujita, T. and Akira, S. (2008) Length-dependent recognition of double-stranded ribonucleic acids by

retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J Exp Med*, **205**, 1601-1610.
66. Peisley, A., Lin, C., Wu, B., Orme-Johnson, M., Liu, M., Walz, T. and Hur, S. (2011) Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition. *P Natl Acad Sci USA*, **108**, 21010-21015.
67. Peisley, A., Jo, M.H., Lin, C., Wu, B., Orme-Johnson, M., Walz, T., Hohng, S. and Hur, S. (2012) Kinetic mechanism for viral dsRNA length discrimination by MDA5 filaments. *P Natl Acad Sci USA*, **109**, E3340-E3349.
68. Wu, B., Peisley, A., Richards, C., Yao, H., Zeng, X., Lin, C., Chu, F., Walz, T. and Hur, S. (2013) Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. *Cell*, **152**, 276-289.
69. Hornung, V., Ellegast, J., Kim, S., Brzózka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M. *et al.* (2006) 5'-Triphosphate RNA is the ligand for RIG-I. *Science*, **314**, 994-997.
70. Pichlmair, A., Schulz, O., Tan, C.P., Näslund, T.I., Liljeström, P., Weber, F. and Reis e Sousa, C. (2006) RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science*, **314**, 997-1001.
71. Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A.J., Zillinger, T., Goldeck, M., Schuberth, C., Van der Veen, A.G., Fujimura, T., Rehwinkel, J. *et al.* (2014) Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. *Nature*, **514**, 372-375.
72. Baum, A., Sachidanandam, R. and García-Sastre, A. (2010) Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proc Natl Acad Sci U S A*, **107**, 16303-16308.
73. Ablasser, A., Bauernfeind, F., Hartmann, G., Latz, E., Fitzgerald, K.A. and Hornung, V. (2009) RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat Immunol*, **10**, 1065-1072.
74. Chiu, Y.H., Macmillan, J.B. and Chen, Z.J. (2009) RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell*, **138**, 576-591.
75. Jiang, X., Kinch, L.N., Brautigam, C.A., Chen, X., Du, F., Grishin, N.V. and Chen, Z.J. (2012) Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity*, **36**, 959-973.
76. Peisley, A., Wu, B., Yao, H., Walz, T. and Hur, S. (2013) RIG-I Forms Signaling-Competent Filaments in an ATP-Dependent, Ubiquitin-Independent Manner. *Molecular Cell*, **51**, 573-583.
77. Patel, J.R., Jain, A., Chou, Y.-y., Baum, A., Ha, T. and García-Sastre, A. (2013) ATPase-driven oligomerization of RIG-I on RNA allows optimal activation of type-I interferon. *EMBO reports*, **14**, 780-787.

78. Hou, F., Sun, L., Zheng, H., Skaug, B., Jiang, Q.-X. and Chen, Z.J. (2011) MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell*, **146**, 448-461.
79. Liu, G., Lu, Y., Thulasi Raman, S.N., Xu, F., Wu, Q., Li, Z., Brownlie, R., Liu, Q. and Zhou, Y. (2018) Nuclear-resident RIG-I senses viral replication inducing antiviral immunity. *Nat Commun*, **9**, 3199.
80. Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K.J. *et al.* (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, **441**, 101-105.
81. Triantafilou, K., Vakakis, E., Kar, S., Richer, E., Evans, G.L. and Triantafilou, M. (2012) Visualisation of direct interaction of MDA5 and the dsRNA replicative intermediate form of positive strand RNA viruses. *Journal of Cell Science*, **125**, 4761.
82. Schlee, M., Roth, A., Hornung, V., Hagmann, C.A., Wimmenauer, V., Barchet, W., Coch, C., Janke, M., Mihailovic, A., Wardle, G. *et al.* (2009) Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunity*, **31**, 25-34.
83. Kell, A.M. and Gale, M., Jr. (2015) RIG-I in RNA virus recognition. *Virology*, **479-480**, 110-121.
84. Zhao, Y. and Karijolich, J. (2019) Know Thyself: RIG-I-Like Receptor Sensing of DNA Virus Infection. *J Virol*, **93**, e01085-01019.
85. Schroder, K. and Tschopp, J. (2010) The Inflammasomes. *Cell*, **140**, 821-832.
86. Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., Ogura, Y., Kawasaki, A., Fukase, K., Kusumoto, S. *et al.* (2003) An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nature Immunology*, **4**, 702-707.
87. Girardin, S.E., Boneca, I.G., Carneiro, L.A.M., Antignac, A., Jéhanno, M., Viala, J., Tedin, K., Taha, M.K., Labigne, A., Zähringer, U. *et al.* (2003) Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science*, **300**, 1584-1587.
88. Girardin, S.E., Boneca, I.G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D.J. and Sansonetti, P.J. (2003) Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *Journal of Biological Chemistry*, **278**, 8869-8872.
89. Sabbah, A., Chang, T.H., Harnack, R., Frohlich, V., Tominaga, K., Dube, P.H., Xiang, Y. and Bose, S. (2009) Activation of innate immune antiviral responses by Nod2. *Nature immunology*, **10**, 1073-1080.
90. Le Bel, M. and Gosselin, J. (2015) Leukotriene B4 Enhances NOD2-Dependent Innate Response against Influenza Virus Infection. *PLOS ONE*, **10**, e0139856.
91. Trindade, B.C. and Chen, G.Y. (2020) NOD1 and NOD2 in inflammatory and infectious diseases. *Immunological Reviews*, **297**, 139-161.

92. Martinon, F., Burns, K. and Tschopp, J. (2002) The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. *Mol Cell*, **10**, 417-426.
93. Martinon, F., Burns, K. and Tschopp, J. (2002) The Inflammasome: A Molecular Platform Triggering Activation of Inflammatory Caspases and Processing of proIL-β. *Molecular Cell*, **10**, 417-426.
94. Broz, P., von Moltke, J., Jones, J.W., Vance, R.E. and Monack, D.M. (2010) Differential Requirement for Caspase-1 Autoproteolysis in Pathogen-Induced Cell Death and Cytokine Processing. *Cell Host & Microbe*, **8**, 471-483.
95. Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., Zhuang, Y., Cai, T., Wang, F. and Shao, F. (2015) Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature*, **526**, 660-665.
96. He, W.-t., Wan, H., Hu, L., Chen, P., Wang, X., Huang, Z., Yang, Z.-H., Zhong, C.-Q. and Han, J. (2015) Gasdermin D is an executor of pyroptosis and required for interleukin-1β secretion. *Cell Research*, **25**, 1285-1298.
97. Lieberman, J., Wu, H. and Kagan, J.C. (2019) Gasdermin D activity in inflammation and host defense. *Sci Immunol*, **4**.
98. Rathinam, V.A.K. and Fitzgerald, K.A. (2016) Inflammasome Complexes: Emerging Mechanisms and Effector Functions. *Cell*, **165**, 792-800.
99. Broz, P. and Dixit, V.M. (2016) Inflammasomes: mechanism of assembly, regulation and signalling. *Nature Reviews Immunology*, **16**, 407-420.
100. Duewell, P., Kono, H., Rayner, K.J., Sirois, C.M., Vladimer, G., Bauernfeind, F.G., Abela, G.S., Franchi, L., Nuñez, G., Schnurr, M. *et al.* (2010) NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. *Nature*, **464**, 1357-1361.
101. Kanneganti, T.-D., Özören, N., Body-Malapel, M., Amer, A., Park, J.-H., Franchi, L., Whitfield, J., Barchet, W., Colonna, M., Vandenabeele, P. *et al.* (2006) Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature*, **440**, 233-236.
102. Mariathasan, S., Weiss, D.S., Newton, K., McBride, J., O'Rourke, K., Roose-Girma, M., Lee, W.P., Weinrauch, Y., Monack, D.M. and Dixit, V.M. (2006) Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature*, **440**, 228-232.
103. Pétrilli, V., Papin, S., Dostert, C., Mayor, A., Martinon, F. and Tschopp, J. (2007) Activation of the NALP3 inflammasome is triggered by low intracellular potassium concentration. *Cell Death & Differentiation*, **14**, 1583-1589.
104. Sutterwala, F.S., Ogura, Y., Szczepanik, M., Lara-Tejero, M., Lichtenberger, G.S., Grant, E.P., Bertin, J., Coyle, A.J., Galán, J.E., Askenase, P.W. *et al.* (2006) Critical Role for NALP3/CIAS1/Cryopyrin in Innate and Adaptive Immunity through Its Regulation of Caspase-1. *Immunity*, **24**, 317-327.

105. Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A. and Tschopp, J. (2006) Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature*, **440**, 237-241.
106. Rawat, P., Teodorof-Diedrich, C. and Spector, S.A. (2019) Human immunodeficiency virus Type-1 single-stranded RNA activates the NLRP3 inflammasome and impairs autophagic clearance of damaged mitochondria in human microglia. *Glia*, **67**, 802-824.
107. Nieto-Torres, J.L., Verdiá-Báguena, C., Jimenez-Guardeño, J.M., Regla-Nava, J.A., Castaño-Rodriguez, C., Fernandez-Delgado, R., Torres, J., Aguilella, V.M. and Enjuanes, L. (2015) Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. *Virology*, **485**, 330-339.
108. Wang, L. and Hauenstein, A.V. (2020) The NLRP3 inflammasome: Mechanism of action, role in disease and therapies. *Molecular Aspects of Medicine*, 100889.
109. Brunette, R.L., Young, J.M., Whitley, D.G., Brodsky, I.E., Malik, H.S. and Stetson, D.B. (2012) Extensive evolutionary and functional diversity among mammalian AIM2-like receptors. *Journal of Experimental Medicine*, **209**, 1969-1983.
110. Cridland, J.A., Curley, E.Z., Wykes, M.N., Schroder, K., Sweet, M.J., Roberts, T.L., Ragan, M.A., Kassahn, K.S. and Stacey, K.J. (2012) The mammalian PYHIN gene family: Phylogeny, evolution and expression. *BMC Evolutionary Biology*, **12**, 140.
111. Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D.R., Latz, E. and Fitzgerald, K.A. (2009) AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature*, **458**, 514-518.
112. Fernandes-Alnemri, T., Yu, J.W., Datta, P., Wu, J. and Alnemri, E.S. (2009) AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature*, **458**, 509-513.
113. Bürckstümmer, T., Baumann, C., Blüml, S., Dixit, E., Dürnberger, G., Jahn, H., Planyavsky, M., Bilban, M., Colinge, J., Bennett, K.L. *et al.* (2009) An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat Immunol*, **10**, 266-272.
114. Roberts, T.L., Idris, A., Dunn, J.A., Kelly, G.M., Burnton, C.M., Hodgson, S., Hardy, L.L., Garceau, V., Sweet, M.J., Ross, I.L. *et al.* (2009) HIN-200 proteins regulate caspase activation in response to foreign cytoplasmic DNA. *Science*, **323**, 1057-1060.
115. Sagulenko, V., Thygesen, S.J., Sester, D.P., Idris, A., Cridland, J.A., Vajjhala, P.R., Roberts, T.L., Schroder, K., Vince, J.E., Hill, J.M. *et al.* (2013) AIM2 and NLRP3 inflammasomes activate both apoptotic and pyroptotic death pathways via ASC. *Cell Death Differ*, **20**, 1149-1160.
116. Pierini, R., Juruj, C., Perret, M., Jones, C.L., Mangeot, P., Weiss, D.S. and Henry, T. (2012) AIM2/ASC triggers caspase-8-dependent apoptosis in

Francisella-infected caspase-1-deficient macrophages. *Cell Death Differ*, **19**, 1709-1721.
117. Jin, T., Perry, A., Jiang, J., Smith, P., Curry, J.A., Unterholzner, L., Jiang, Z., Horvath, G., Rathinam, V.A., Johnstone, R.W. *et al.* (2012) Structures of the HIN domain:DNA complexes reveal ligand binding and activation mechanisms of the AIM2 inflammasome and IFI16 receptor. *Immunity*, **36**, 561-571.
118. Matyszewski, M., Morrone, S.R. and Sohn, J. (2018) Digital signaling network drives the assembly of the AIM2-ASC inflammasome. *Proc Natl Acad Sci U S A*, **115**, E1963-e1972.
119. Rathinam, V.A.K., Jiang, Z., Waggoner, S.N., Sharma, S., Cole, L.E., Waggoner, L., Vanaja, S.K., Monks, B.G., Ganesan, S., Latz, E. *et al.* (2010) The AIM2 inflammasome is essential for host defense against cytosolic bacteria and DNA viruses. *Nature Immunology*, **11**, 395-402.
120. Reinholz, M., Kawakami, Y., Salzer, S., Kreuter, A., Dombrowski, Y., Koglin, S., Kresse, S., Ruzicka, T. and Schauber, J. (2013) HPV16 activates the AIM2 inflammasome in keratinocytes. *Archives of Dermatological Research*, **305**, 723-732.
121. Torii, Y., Kawada, J.-i., Murata, T., Yoshiyama, H., Kimura, H. and Ito, Y. (2017) Epstein-Barr virus infection-induced inflammasome activation in human monocytes. *PLOS ONE*, **12**, e0175053.
122. Di Micco, A., Frera, G., Lugrin, J., Jamilloux, Y., Hsu, E.-T., Tardivel, A., De Gassart, A., Zaffalon, L., Bujisic, B., Siegert, S. *et al.* (2016) AIM2 inflammasome is activated by pharmacological disruption of nuclear envelope integrity. *Proceedings of the National Academy of Sciences*, **113**, E4671.
123. Lian, Q., Xu, J., Yan, S., Huang, M., Ding, H., Sun, X., Bi, A., Ding, J., Sun, B. and Geng, M. (2017) Chemotherapy-induced intestinal inflammatory responses are mediated by exosome secretion of double-strand DNA via AIM2 inflammasome activation. *Cell Research*, **27**, 784-800.
124. Strittmatter, G.E., Sand, J., Sauter, M., Seyffert, M., Steigerwald, R., Fraefel, C., Smola, S., French, L.E. and Beer, H.D. (2016) IFN-γ Primes Keratinocytes for HSV-1-Induced Inflammasome Activation. *J Invest Dermatol*, **136**, 610-620.
125. Hu, B., Jin, C., Li, H.-B., Tong, J., Ouyang, X., Cetinbas, N.M., Zhu, S., Strowig, T., Lam, F.C., Zhao, C. *et al.* (2016) The DNA-sensing AIM2 inflammasome controls radiation-induced cell death and tissue injury. *Science*, **354**, 765.
126. Lugrin, J. and Martinon, F. (2018) The AIM2 inflammasome: Sensor of pathogens and cellular perturbations. *Immunological Reviews*, **281**, 99-114.
127. Kofoed, E.M. and Vance, R.E. (2011) Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature*, **477**, 592-595.

128. Zhao, Y., Yang, J., Shi, J., Gong, Y.N., Lu, Q., Xu, H., Liu, L. and Shao, F. (2011) The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature*, **477**, 596-600.
129. Boyden, E.D. and Dietrich, W.F. (2006) Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nat Genet*, **38**, 240-244.
130. Neiman-Zenevich, J., Stuart, S., Abdel-Nour, M., Girardin, S.E. and Mogridge, J. (2017) Listeria monocytogenes and Shigella flexneri Activate the NLRP1B Inflammasome. *Infect Immun*, **85**.
131. Cirelli, K.M., Gorfu, G., Hassan, M.A., Printz, M., Crown, D., Leppla, S.H., Grigg, M.E., Saeij, J.P.J. and Moayeri, M. (2014) Inflammasome sensor NLRP1 controls rat macrophage susceptibility to Toxoplasma gondii. *PLoS pathogens*, **10**, e1003927-e1003927.
132. Gorfu, G., Cirelli, K.M., Melo, M.B., Mayer-Barber, K., Crown, D., Koller, B.H., Masters, S., Sher, A., Leppla, S.H., Moayeri, M. *et al.* (2014) Dual role for inflammasome sensors NLRP1 and NLRP3 in murine resistance to Toxoplasma gondii. *mBio*, **5**, e01117-01113.
133. Ewald, S.E., Chavarria-Smith, J. and Boothroyd, J.C. (2014) NLRP1 is an inflammasome sensor for Toxoplasma gondii. *Infect Immun*, **82**, 460-468.
134. Andrade, W.A. and Zamboni, D.S. (2020) NLRC4 biology in immunity and inflammation. *Journal of Leukocyte Biology*, **108**, 1117-1127.
135. Mitchell, P.S., Sandstrom, A. and Vance, R.E. (2019) The NLRP1 inflammasome: new mechanistic insights and unresolved mysteries. *Curr Opin Immunol*, **60**, 37-45.
136. Xu, H., Yang, J., Gao, W., Li, L., Li, P., Zhang, L., Gong, Y.-N., Peng, X., Xi, J.J., Chen, S. *et al.* (2014) Innate immune sensing of bacterial modifications of Rho GTPases by the Pyrin inflammasome. *Nature*, **513**, 237-241.
137. Schnappauf, O., Chae, J.J., Kastner, D.L. and Aksentijevich, I. (2019) The Pyrin Inflammasome in Health and Disease. *Front Immunol*, **10**.
138. Antushevich, H. (2020) Interplays between inflammasomes and viruses, bacteria (pathogenic and probiotic), yeasts and parasites. *Immunol Lett*, **228**, 1-14.
139. Kuenzel, S., Till, A., Winkler, M., Häsler, R., Lipinski, S., Jung, S., Grötzinger, J., Fickenscher, H., Schreiber, S. and Rosenstiel, P. (2010) The Nucleotide-Binding Oligomerization Domain-Like Receptor NLRC5 Is Involved in IFN-Dependent Antiviral Immune Responses. *The Journal of Immunology*, **184**, 1990.
140. Neerincx, A., Lautz, K., Menning, M., Kremmer, E., Zigrino, P., Hösel, M., Büning, H., Schwarzenbacher, R. and Kufer, T.A. (2010) A role for the human nucleotide-binding domain, leucine-rich repeat-containing family member NLRC5 in antiviral responses. *J Biol Chem*, **285**, 26223-26232.
141. Meissner, T.B., Li, A., Biswas, A., Lee, K.-H., Liu, Y.-J., Bayir, E., Iliopoulos, D., van den Elsen, P.J. and Kobayashi, K.S. (2010) NLR family member

NLRC5 is a transcriptional regulator of MHC class I genes. *Proceedings of the National Academy of Sciences*, **107**, 13794.
142. Ranjan, P., Singh, N., Kumar, A., Neerincx, A., Kremmer, E., Cao, W., Davis, W.G., Katz, J.M., Gangappa, S., Lin, R. *et al.* (2015) NLRC5 interacts with RIG-I to induce a robust antiviral response against influenza virus infection. *Eur J Immunol*, **45**, 758-772.
143. Guo, X., Liu, T., Shi, H., Wang, J., Ji, P., Wang, H., Hou, Y., Tan, R.X. and Li, E. (2015) Respiratory Syncytial Virus Infection Upregulates NLRC5 and Major Histocompatibility Complex Class I Expression through RIG-I Induction in Airway Epithelial Cells. *J Virol*, **89**, 7636-7645.
144. Dey, M., Cao, C., Dar, A.C., Tamura, T., Ozato, K., Sicheri, F. and Dever, T.E. (2005) Mechanistic Link between PKR Dimerization, Autophosphorylation, and eIF2α Substrate Recognition. *Cell*, **122**, 901-913.
145. Lemaire, P.A., Anderson, E., Lary, J. and Cole, J.L. (2008) Mechanism of PKR Activation by dsRNA. *J Mol Biol*, **381**, 351-360.
146. Nallagatla, S.R., Hwang, J., Toroney, R., Zheng, X., Cameron, C.E. and Bevilacqua, P.C. (2007) 5'-triphosphate-dependent activation of PKR by RNAs with short stem-loops. *Science*, **318**, 1455-1458.
147. Mayo, C.B. and Cole, J.L. (2017) Interaction of PKR with single-stranded RNA. *Sci Rep*, **7**, 3335.
148. Zheng, X. and Bevilacqua, P.C. (2004) Activation of the protein kinase PKR by short double-stranded RNAs with single-stranded tails. *Rna*, **10**, 1934-1945.
149. Youssef, O.A., Safran, S.A., Nakamura, T., Nix, D.A., Hotamisligil, G.S. and Bass, B.L. (2015) Potential role for snoRNAs in PKR activation during metabolic stress. *Proc Natl Acad Sci U S A*, **112**, 5023-5028.
150. Nallagatla, S.R., Jones, C.N., Ghosh, S.K., Sharma, S.D., Cameron, C.E., Spremulli, L.L. and Bevilacqua, P.C. (2013) Native tertiary structure and nucleoside modifications suppress tRNA's intrinsic ability to activate the innate immune sensor PKR. *PLoS One*, **8**, e57905.
151. Cohen-Chalamish, S., Hasson, A., Weinberg, D., Namer, L.S., Banai, Y., Osman, F. and Kaempfer, R. (2009) Dynamic refolding of IFN-gamma mRNA enables it to function as PKR activator and translation template. *Nat Chem Biol*, **5**, 896-903.
152. Ben-Asouli, Y., Banai, Y., Pel-Or, Y., Shir, A. and Kaempfer, R. (2002) Human interferon-gamma mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR. *Cell*, **108**, 221-232.
153. Davis, S. and Watson, J.C. (1996) In vitro activation of the interferon-induced, double-stranded RNA-dependent protein kinase PKR by RNA from the 3' untranslated regions of human alpha-tropomyosin. *Proc Natl Acad Sci U S A*, **93**, 508-513.
154. Bevilacqua, P.C., George, C.X., Samuel, C.E. and Cech, T.R. (1998) Binding of the protein kinase PKR to RNAs with secondary structure

defects: role of the tandem A-G mismatch and noncontiguous helixes. *Biochemistry*, **37**, 6303-6316.
155. Heinicke, L.A., Wong, C.J., Lary, J., Nallagatla, S.R., Diegelman-Parente, A., Zheng, X., Cole, J.L. and Bevilacqua, P.C. (2009) RNA dimerization promotes PKR dimerization and activation. *J Mol Biol*, **390**, 319-338.
156. Nallagatla, S.R., Toroney, R. and Bevilacqua, P.C. (2011) Regulation of innate immunity through RNA structure and the protein kinase PKR. *Curr Opin Struct Biol*, **21**, 119-127.
157. Bou-Nader, C., Gordon, J.M., Henderson, F.E. and Zhang, J. (2019) The search for a PKR code-differential regulation of protein kinase R activity by diverse RNA and protein regulators. *RNA (New York, N.Y.)*, **25**, 539-556.
158. Yeung, M.C., Liu, J. and Lau, A.S. (1996) An essential role for the interferon-inducible, double-stranded RNA-activated protein kinase PKR in the tumor necrosis factor-induced apoptosis in U937 cells. *Proc Natl Acad Sci U S A*, **93**, 12451-12455.
159. Goh, K.C., deVeer, M.J. and Williams, B.R. (2000) The protein kinase PKR is required for p38 MAPK activation and the innate immune response to bacterial endotoxin. *Embo j*, **19**, 4292-4297.
160. Ito, T., Yang, M. and May, W.S. (1999) RAX, a cellular activator for double-stranded RNA-dependent protein kinase during stress signaling. *J Biol Chem*, **274**, 15427-15432.
161. Samuel, C.E. (1979) Mechanism of interferon action: phosphorylation of protein synthesis initiation factor eIF-2 in interferon-treated human cells by a ribosome-associated kinase processing site specificity similar to hemin-regulated rabbit reticulocyte kinase. *Proc Natl Acad Sci U S A*, **76**, 600-604.
162. Pakos-Zebrucka, K., Koryga, I., Mnich, K., Ljujic, M., Samali, A. and Gorman, A.M. (2016) The integrated stress response. *EMBO Rep*, **17**, 1374-1395.
163. Eiermann, N., Haneke, K., Sun, Z., Stoecklin, G. and Ruggieri, A. (2020) Dance with the Devil: Stress Granules and Signaling in Antiviral Responses. *Viruses*, **12**.
164. Dalet, A., Gatti, E. and Pierre, P. (2015) Integration of PKR-dependent translation inhibition with innate immunity is required for a coordinated anti-viral response. *FEBS Lett*, **589**, 1539-1545.
165. Zhang, P. and Samuel, C.E. (2008) Induction of protein kinase PKR-dependent activation of interferon regulatory factor 3 by vaccinia virus occurs through adapter IPS-1 signaling. *J Biol Chem*, **283**, 34580-34587.
166. Gil, J. and Esteban, M. (2000) The interferon-induced protein kinase (PKR), triggers apoptosis through FADD-mediated activation of caspase 8 in a manner independent of Fas and TNF-alpha receptors. *Oncogene*, **19**, 3665-3674.

167. Thapa, R.J., Nogusa, S., Chen, P., Maki, J.L., Lerro, A., Andrake, M., Rall, G.F., Degterev, A. and Balachandran, S. (2013) Interferon-induced RIP1/RIP3-mediated necrosis requires PKR and is licensed by FADD and caspases. *Proc Natl Acad Sci U S A*, **110**, E3109-3118.
168. von Roretz, C. and Gallouzi, I.E. (2010) Protein kinase RNA/FADD/caspase-8 pathway mediates the proapoptotic activity of the RNA-binding protein human antigen R (HuR). *J Biol Chem*, **285**, 16806-16813.
169. Singh, M., Fowlkes, V., Handy, I., Patel, C.V. and Patel, R.C. (2009) Essential role of PACT-mediated PKR activation in tunicamycin-induced apoptosis. *J Mol Biol*, **385**, 457-468.
170. Marques, J.T., White, C.L., Peters, G.A., Williams, B.R. and Sen, G.C. (2008) The role of PACT in mediating gene induction, PKR activation, and apoptosis in response to diverse stimuli. *J Interferon Cytokine Res*, **28**, 469-476.
171. Peters, G.A., Li, S. and Sen, G.C. (2006) Phosphorylation of specific serine residues in the PKR activation domain of PACT is essential for its ability to mediate apoptosis. *J Biol Chem*, **281**, 35129-35136.
172. Liu, Y., Wang, M., Cheng, A., Yang, Q., Wu, Y., Jia, R., Liu, M., Zhu, D., Chen, S., Zhang, S. *et al.* (2020) The role of host eIF2α in viral infection. *Virol J*, **17**, 112.
173. Dzananovic, E., McKenna, S.A. and Patel, T.R. (2018) Viral proteins targeting host protein kinase R to evade an innate immune response: a mini review. *Biotechnology and Genetic Engineering Reviews*, **34**, 33-59.
174. Moretti, J., Roy, S., Bozec, D., Martinez, J., Chapman, J.R., Ueberheide, B., Lamming, D.W., Chen, Z.J., Horng, T., Yeretssian, G. *et al.* (2017) STING Senses Microbial Viability to Orchestrate Stress-Mediated Autophagy of the Endoplasmic Reticulum. *Cell*, **171**, 809-823.e813.
175. Holm, C.K., Jensen, S.B., Jakobsen, M.R., Cheshenko, N., Horan, K.A., Moeller, H.B., Gonzalez-Dosal, R., Rasmussen, S.B., Christensen, M.H., Yarovinsky, T.O. *et al.* (2012) Virus-cell fusion as a trigger of innate immunity dependent on the adaptor STING. *Nature Immunology*, **13**, 737-743.
176. Sun, W., Li, Y., Chen, L., Chen, H., You, F., Zhou, X., Zhou, Y., Zhai, Z., Chen, D. and Jiang, Z. (2009) ERIS, an endoplasmic reticulum IFN stimulator, activates innate immune signaling through dimerization. *Proc Natl Acad Sci U S A*, **106**, 8653-8658.
177. Srikanth, S., Woo, J.S., Wu, B., El-Sherbiny, Y.M., Leung, J., Chupradit, K., Rice, L., Seo, G.J., Calmettes, G., Ramakrishna, C. *et al.* (2019) The Ca(2+) sensor STIM1 regulates the type I interferon response by retaining the signaling adaptor STING at the endoplasmic reticulum. *Nat Immunol*, **20**, 152-162.
178. Mukai, K., Konno, H., Akiba, T., Uemura, T., Waguri, S., Kobayashi, T., Barber, G.N., Arai, H. and Taguchi, T. (2016) Activation of STING requires palmitoylation at the Golgi. *Nat Commun*, **7**, 11932.

179. Gui, X., Yang, H., Li, T., Tan, X., Shi, P., Li, M., Du, F. and Chen, Z.J. (2019) Autophagy induction via STING trafficking is a primordial function of the cGAS pathway. *Nature*, **567**, 262-266.
180. Tanaka, Y. and Chen, Z.J. (2012) STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sci Signal*, **5**, ra20.
181. Ishikawa, H., Ma, Z. and Barber, G.N. (2009) STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature*, **461**, 788-792.
182. Jin, L., Hill, K.K., Filak, H., Mogan, J., Knowles, H., Zhang, B., Perraud, A.-L., Cambier, J.C. and Lenz, L.L. (2011) MPYS Is Required for IFN Response Factor 3 Activation and Type I IFN Production in the Response of Cultured Phagocytes to Bacterial Second Messengers Cyclic-di-AMP and Cyclic-di-GMP. *The Journal of Immunology*, **187**, 2595.
183. Abe, T. and Barber, G.N. (2014) Cytosolic-DNA-mediated, STING-dependent proinflammatory gene induction necessitates canonical NF-κB activation through TBK1. *J Virol*, **88**, 5328-5341.
184. Fang, R., Wang, C., Jiang, Q., Lv, M., Gao, P., Yu, X., Mu, P., Zhang, R., Bi, S., Feng, J.M. *et al.* (2017) NEMO-IKKβ Are Essential for IRF3 and NF-κB Activation in the cGAS-STING Pathway. *J Immunol*, **199**, 3222-3233.
185. Shang, G., Zhang, C., Chen, Z.J., Bai, X.C. and Zhang, X. (2019) Cryo-EM structures of STING reveal its mechanism of activation by cyclic GMP-AMP. *Nature*, **567**, 389-393.
186. Jin, L., Waterman, P.M., Jonscher, K.R., Short, C.M., Reisdorph, N.A. and Cambier, J.C. (2008) MPYS, a novel membrane tetraspanner, is associated with major histocompatibility complex class II and mediates transduction of apoptotic signals. *Mol Cell Biol*, **28**, 5014-5026.
187. Li, X., Shu, C., Yi, G., Chaton, C.T., Shelton, C.L., Diao, J., Zuo, X., Kao, C.C., Herr, A.B. and Li, P. (2013) Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity*, **39**, 1019-1031.
188. Zhang, X., Wu, J., Du, F., Xu, H., Sun, L., Chen, Z., Brautigam, C.A., Zhang, X. and Chen, Z.J. (2014) The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep*, **6**, 421-430.
189. Zhou, W., Whiteley, A.T., de Oliveira Mann, C.C., Morehouse, B.R., Nowak, R.P., Fischer, E.S., Gray, N.S., Mekalanos, J.J. and Kranzusch, P.J. (2018) Structure of the Human cGAS-DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell*, **174**, 300-311.e311.
190. Civril, F., Deimling, T., de Oliveira Mann, C.C., Ablasser, A., Moldt, M., Witte, G., Hornung, V. and Hopfner, K.P. (2013) Structural mechanism of cytosolic DNA sensing by cGAS. *Nature*, **498**, 332-337.
191. Mankan, A.K., Schmidt, T., Chauhan, D., Goldeck, M., Höning, K., Gaidt, M., Kubarenko, A.V., Andreeva, L., Hopfner, K.-P. and Hornung, V.

(2014) Cytosolic RNA:DNA hybrids activate the cGAS-STING axis. *The EMBO journal*, **33**, 2937-2946.
192. Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Röhl, I., Hopfner, K.P., Ludwig, J. and Hornung, V. (2013) cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature*, **498**, 380-384.
193. Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G. *et al.* (2013) Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell*, **153**, 1094-1107.
194. Zhang, X., Shi, H., Wu, J., Zhang, X., Sun, L., Chen, C. and Chen, Z.J. (2013) Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol Cell*, **51**, 226-235.
195. Diner, E.J., Burdette, D.L., Wilson, S.C., Monroe, K.M., Kellenberger, C.A., Hyodo, M., Hayakawa, Y., Hammond, M.C. and Vance, R.E. (2013) The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep*, **3**, 1355-1361.
196. Ablasser, A., Schmid-Burgk, J.L., Hemmerling, I., Horvath, G.L., Schmidt, T., Latz, E. and Hornung, V. (2013) Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP. *Nature*, **503**, 530-534.
197. Bridgeman, A., Maelfait, J., Davenne, T., Partridge, T., Peng, Y., Mayer, A., Dong, T., Kaever, V., Borrow, P. and Rehwinkel, J. (2015) Viruses transfer the antiviral second messenger cGAMP between cells. *Science*, **349**, 1228.
198. Gentili, M., Kowal, J., Tkach, M., Satoh, T., Lahaye, X., Conrad, C., Boyron, M., Lombard, B., Durand, S., Kroemer, G. *et al.* (2015) Transmission of innate immune signaling by packaging of cGAMP in viral particles. *Science*, **349**, 1232.
199. Shang, G., Zhu, D., Li, N., Zhang, J., Zhu, C., Lu, D., Liu, C., Yu, Q., Zhao, Y., Xu, S. *et al.* (2012) Crystal structures of STING protein reveal basis for recognition of cyclic di-GMP. *Nat Struct Mol Biol*, **19**, 725-727.
200. Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y. and Vance, R.E. (2011) STING is a direct innate immune sensor of cyclic di-GMP. *Nature*, **478**, 515-518.
201. Motwani, M., Pesiridis, S. and Fitzgerald, K.A. (2019) DNA sensing by the cGAS–STING pathway in health and disease. *Nature Reviews Genetics*, **20**, 657-674.
202. Gao, D., Wu, J., Wu, Y.-T., Du, F., Aroh, C., Yan, N., Sun, L. and Chen, Z.J. (2013) Cyclic GMP-AMP Synthase Is an Innate Immune Sensor of HIV and Other Retroviruses. *Science*, **341**, 903.
203. Franz, K.M., Neidermyer, W.J., Tan, Y.J., Whelan, S.P.J. and Kagan, J.C. (2018) STING-dependent translation inhibition restricts RNA virus replication. *Proc Natl Acad Sci U S A*, **115**, E2058-e2067.

204. Li, X.-D., Wu, J., Gao, D., Wang, H., Sun, L. and Chen, Z.J. (2013) Pivotal Roles of cGAS-cGAMP Signaling in Antiviral Defense and Immune Adjuvant Effects. *Science*, **341**, 1390.
205. Ahn, J. and Barber, G.N. (2019) STING signaling and host defense against microbial infection. *Experimental & Molecular Medicine*, **51**, 1-10.
206. Phelan, T., Little, M.A. and Brady, G. (2020) Targeting of the cGAS-STING system by DNA viruses. *Biochemical Pharmacology*, **174**, 113831.
207. Blackford, A.N. and Jackson, S.P. (2017) ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Molecular Cell*, **66**, 801-817.
208. Ferguson, B.J., Mansur, D.S., Peters, N.E., Ren, H. and Smith, G.L. (2012) DNA-PK is a DNA sensor for IRF-3-dependent innate immunity. *Elife*, **1**, e00047.
209. Burleigh, K., Maltbaek, J.H., Cambier, S., Green, R., Gale, M., James, R.C. and Stetson, D.B. (2020) Human DNA-PK activates a STING-independent DNA sensing pathway. *Science Immunology*, **5**, eaba4219.
210. Smith, J.A. and Daniel, R. (2006) Following the Path of the Virus: The Exploitation of Host DNA Repair Mechanisms by Retroviruses. *ACS Chemical Biology*, **1**, 217-226.
211. Scutts, S.R., Ember, S.W., Ren, H., Ye, C., Lovejoy, C.A., Mazzon, M., Veyer, D.L., Sumner, R.P. and Smith, G.L. (2018) DNA-PK Is Targeted by Multiple Vaccinia Virus Proteins to Inhibit DNA Sensing. *Cell Reports*, **25**, 1953-1965.e1954.
212. Peters, N.E., Ferguson, B.J., Mazzon, M., Fahy, A.S., Krysztofinska, E., Arribas-Bosacoma, R., Pearl, L.H., Ren, H. and Smith, G.L. (2013) A Mechanism for the Inhibition of DNA-PK-Mediated DNA Sensing by a Virus. *PLOS Pathogens*, **9**, e1003649.
213. Choi, Y.-K., Nash, K., Byrne, B.J., Muzyczka, N. and Song, S. (2010) The Effect of DNA-Dependent Protein Kinase on Adeno-Associated Virus Replication. *PLOS ONE*, **5**, e15073.
214. Schwartz, R.A., Carson, C.T., Schuberth, C. and Weitzman, M.D. (2009) Adeno-associated virus replication induces a DNA damage response coordinated by DNA-dependent protein kinase. *J Virol*, **83**, 6269-6278.
215. Weitzman, M.D. and Fradet-Turcotte, A. (2018) Virus DNA Replication and the Host DNA Damage Response. *Annual Review of Virology*, **5**, 141-164.
216. Stavrou, S., Blouch, K., Kotla, S., Bass, A. and Ross, Susan R. (2015) Nucleic Acid Recognition Orchestrates the Anti-Viral Response to Retroviruses. *Cell Host & Microbe*, **17**, 478-488.
217. Omura, H., Oikawa, D., Nakane, T., Kato, M., Ishii, R., Ishitani, R., Tokunaga, F. and Nureki, O. (2016) Structural and Functional Analysis of DDX41: a bispecific immune receptor for DNA and cyclic dinucleotide. *Scientific Reports*, **6**, 34756.

218. Zhang, Z., Yuan, B., Bao, M., Lu, N., Kim, T. and Liu, Y.J. (2011) The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat Immunol*, **12**, 959-965.
219. Parvatiyar, K., Zhang, Z., Teles, R.M., Ouyang, S., Jiang, Y., Iyer, S.S., Zaver, S.A., Schenk, M., Zeng, S., Zhong, W. *et al.* (2012) The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. *Nature Immunology*, **13**, 1155-1161.
220. Jiang, Y., Zhu, Y., Liu, Z.-J. and Ouyang, S. (2017) The emerging roles of the DDX41 protein in immunity and diseases. *Protein & Cell*, **8**, 83-89.
221. Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T., Latz, E., Xiao, T.S. *et al.* (2010) IFI16 is an innate immune sensor for intracellular DNA. *Nat Immunol*, **11**, 997-1004.
222. Li, T., Diner, B.A., Chen, J. and Cristea, I.M. (2012) Acetylation modulates cellular distribution and DNA sensing ability of interferon-inducible protein IFI16. *Proc Natl Acad Sci U S A*, **109**, 10558-10563.
223. Jønsson, K.L., Laustsen, A., Krapp, C., Skipper, K.A., Thavachelvam, K., Hotter, D., Egedal, J.H., Kjolby, M., Mohammadi, P., Prabakaran, T. *et al.* (2017) IFI16 is required for DNA sensing in human macrophages by promoting production and function of cGAMP. *Nature communications*, **8**, 14391-14391.
224. Almine, J.F., O'Hare, C.A., Dunphy, G., Haga, I.R., Naik, R.J., Atrih, A., Connolly, D.J., Taylor, J., Kelsall, I.R., Bowie, A.G. *et al.* (2017) IFI16 and cGAS cooperate in the activation of STING during DNA sensing in human keratinocytes. *Nat Commun*, **8**, 14392.
225. Kerur, N., Veettil, M.V., Sharma-Walia, N., Bottero, V., Sadagopan, S., Otageri, P. and Chandran, B. (2011) IFI16 acts as a nuclear pathogen sensor to induce the inflammasome in response to Kaposi Sarcoma-associated herpesvirus infection. *Cell Host Microbe*, **9**, 363-375.
226. Ansari, M.A., Dutta, S., Veettil, M.V., Dutta, D., Iqbal, J., Kumar, B., Roy, A., Chikoti, L., Singh, V.V. and Chandran, B. (2015) Herpesvirus Genome Recognition Induced Acetylation of Nuclear IFI16 Is Essential for Its Cytoplasmic Translocation, Inflammasome and IFN-β Responses. *PLoS Pathog*, **11**, e1005019.
227. Singh, V.V., Kerur, N., Bottero, V., Dutta, S., Chakraborty, S., Ansari, M.A., Paudel, N., Chikoti, L. and Chandran, B. (2013) Kaposi's sarcoma-associated herpesvirus latency in endothelial and B cells activates gamma interferon-inducible protein 16-mediated inflammasomes. *J Virol*, **87**, 4417-4431.
228. Pham, H.T., Park, M.-Y., Kim, K.K., Kim, Y.-G. and Ahn, J.-H. (2006) Intracellular localization of human ZBP1: Differential regulation by the Z-DNA binding domain, Zα, in splice variants. *Biochemical and Biophysical Research Communications*, **348**, 145-152.

229. Deigendesch, N., Koch-Nolte, F. and Rothenburg, S. (2006) ZBP1 subcellular localization and association with stress granules is controlled by its Z-DNA binding domains. *Nucleic Acids Research*, **34**, 5007-5020.
230. Takaoka, A., Wang, Z., Choi, M.K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., Honda, K. *et al.* (2007) DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature*, **448**, 501-505.
231. Wang, Z., Choi, M.K., Ban, T., Yanai, H., Negishi, H., Lu, Y., Tamura, T., Takaoka, A., Nishikura, K. and Taniguchi, T. (2008) Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules. *Proceedings of the National Academy of Sciences*, **105**, 5477.
232. Maelfait, J., Liverpool, L., Bridgeman, A., Ragan, K.B., Upton, J.W. and Rehwinkel, J. (2017) Sensing of viral and endogenous RNA by ZBP1/DAI induces necroptosis. *The EMBO Journal*, **36**, 2529-2543.
233. Sridharan, H., Ragan, K.B., Guo, H., Gilley, R.P., Landsteiner, V.J., Kaiser, W.J. and Upton, J.W. (2017) Murine cytomegalovirus IE3-dependent transcription is required for DAI/ZBP1-mediated necroptosis. *EMBO reports*, **18**, 1429-1441.
234. Thapa, R.J., Ingram, J.P., Ragan, K.B., Nogusa, S., Boyd, D.F., Benitez, A.A., Sridharan, H., Kosoff, R., Shubina, M., Landsteiner, V.J. *et al.* (2016) DAI Senses Influenza A Virus Genomic RNA and Activates RIPK3-Dependent Cell Death. *Cell Host Microbe*, **20**, 674-681.
235. Jiao, H., Wachsmuth, L., Kumari, S., Schwarzer, R., Lin, J., Eren, R.O., Fisher, A., Lane, R., Young, G.R., Kassiotis, G. *et al.* (2020) Z-nucleic-acid sensing triggers ZBP1-dependent necroptosis and inflammation. *Nature*, **580**, 391-395.
236. Zhang, T., Yin, C., Boyd, D.F., Quarato, G., Ingram, J.P., Shubina, M., Ragan, K.B., Ishizuka, T., Crawford, J.C., Tummers, B. *et al.* (2020) Influenza Virus Z-RNAs Induce ZBP1-Mediated Necroptosis. *Cell*, **180**, 1115-1129.e1113.
237. Rebsamen, M., Heinz, L.X., Meylan, E., Michallet, M.C., Schroder, K., Hofmann, K., Vazquez, J., Benedict, C.A. and Tschopp, J. (2009) DAI/ZBP1 recruits RIP1 and RIP3 through RIP homotypic interaction motifs to activate NF-kappaB. *EMBO Rep*, **10**, 916-922.
238. Upton, J.W., Kaiser, W.J. and Mocarski, E.S. (2012) DAI/ZBP1/DLM-1 complexes with RIP3 to mediate virus-induced programmed necrosis that is targeted by murine cytomegalovirus vIRA. *Cell Host Microbe*, **11**, 290-297.
239. Kuriakose, T., Man, S.M., Malireddi, R.K., Karki, R., Kesavardhana, S., Place, D.E., Neale, G., Vogel, P. and Kanneganti, T.D. (2016) ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci Immunol*, **1**.

240. Zheng, M. and Kanneganti, T.-D. (2020) The regulation of the ZBP1-NLRP3 inflammasome and its implications in pyroptosis, apoptosis, and necroptosis (PANoptosis). *Immunological Reviews*, **297**, 26-38.
241. Chakraborty, P. and Grosse, F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst)*, **10**, 654-665.
242. Jain, A., Bacolla, A., Chakraborty, P., Grosse, F. and Vasquez, K.M. (2010) Human DHX9 helicase unwinds triple-helical DNA structures. *Biochemistry*, **49**, 6992-6999.
243. Lee, T. and Pelletier, J. (2016) The biology of DHX9 and its potential as a therapeutic target. *Oncotarget*, **7**, 42716-42739.
244. Kim, T., Pazhoor, S., Bao, M., Zhang, Z., Hanabuchi, S., Facchinetti, V., Bover, L., Plumas, J., Chaperot, L., Qin, J. *et al.* (2010) Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. *Proc Natl Acad Sci U S A*, **107**, 15181-15186.
245. Zhang, Z., Yuan, B., Lu, N., Facchinetti, V. and Liu, Y.J. (2011) DHX9 pairs with IPS-1 to sense double-stranded RNA in myeloid dendritic cells. *J Immunol*, **187**, 4501-4508.
246. Ng, Yee C., Chung, W.-C., Kang, H.-R., Cho, H.-J., Park, E.-B., Kang, S.-J. and Song, Moon J. (2018) A DNA-sensing–independent role of a nuclear RNA helicase, DHX9, in stimulation of NF-κB–mediated innate immunity against DNA virus infection. *Nucleic Acids Research*, **46**, 9011-9026.
247. Jing, H., Zhou, Y., Fang, L., Ding, Z., Wang, D., Ke, W., Chen, H. and Xiao, S. (2017) DExD/H-Box Helicase 36 Signaling via Myeloid Differentiation Primary Response Gene 88 Contributes to NF-κB Activation to Type 2 Porcine Reproductive and Respiratory Syndrome Virus Infection. *Front Immunol*, **8**.
248. Dempsey, A., Keating, S.E., Carty, M. and Bowie, A.G. (2018) Poxviral protein E3-altered cytokine production reveals that DExD/H-box helicase 9 controls Toll-like receptor-stimulated immune responses. *The Journal of biological chemistry*, **293**, 14989-15001.
249. Sadler, A.J., Latchoumanin, O., Hawkes, D., Mak, J. and Williams, B.R.G. (2009) An Antiviral Response Directed by PKR Phosphorylation of the RNA Helicase A. *PLOS Pathogens*, **5**, e1000311.
250. Fujii, R., Okamoto, M., Aratani, S., Oishi, T., Ohshima, T., Taira, K., Baba, M., Fukamizu, A. and Nakajima, T. (2001) A Role of RNA Helicase A in cis-Acting Transactivation Response Element-mediated Transcriptional Regulation of Human Immunodeficiency Virus Type 1. *J Biol Chem*, **276**, 5445-5451.
251. Liao, H.-J., Kobayashi, R. and Mathews, M.B. (1998) Activities of adenovirus virus-associated RNAs: Purification and characterization of RNA binding proteins. *Proceedings of the National Academy of Sciences*, **95**, 8514.

252. Bolinger, C., Sharma, A., Singh, D., Yu, L. and Boris-Lawrie, K. (2010) RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic acids research*, **38**, 1686-1696.
253. He, Q.S., Tang, H., Zhang, J., Truong, K., Wong-Staal, F. and Zhou, D. (2008) Comparisons of RNAi approaches for validation of human RNA helicase A as an essential factor in hepatitis C virus replication. *Journal of Virological Methods*, **154**, 216-219.
254. Lenarcic, E.M., Ziehr, B.J. and Moorman, N.J. (2015) An unbiased proteomics approach to identify human cytomegalovirus RNA-associated proteins. *Virology*, **481**, 13-23.
255. Fu, W., Verma, D., Burton, A. and Swaminathan, S. (2019) Cellular RNA Helicase DHX9 Interacts with the Essential Epstein-Barr Virus (EBV) Protein SM and Restricts EBV Lytic Replication. *J Virol*, **93**.
256. Shen, B., Chen, Y., Hu, J., Qiao, M., Ren, J., Hu, J., Chen, J., Tang, N., Huang, A. and Hu, Y. (2020) Hepatitis B virus X protein modulates upregulation of DHX9 to promote viral DNA replication. *Cell Microbiol*, **22**, e13148.
257. Wang, Y., Chen, X., Xie, J., Zhou, S., Huang, Y., Li, Y.P., Li, X., Liu, C., He, J. and Zhang, P. (2019) RNA Helicase A Is an Important Host Factor Involved in Dengue Virus Replication. *J Virol*, **93**.
258. Tran, H., Schilling, M., Wirbelauer, C., Hess, D. and Nagamine, Y. (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol Cell*, **13**, 101-111.
259. Iwamoto, F., Stadler, M., Chalupníková, K., Oakeley, E. and Nagamine, Y. (2008) Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU. *Exp Cell Res*, **314**, 1378-1391.
260. Meier, M., Patel, T.R., Booy, E.P., Marushchak, O., Okun, N., Deo, S., Howard, R., McEleney, K., Harding, S.E., Stetefeld, J. *et al.* (2013) Binding of G-quadruplexes to the N-terminal recognition domain of the RNA helicase associated with AU-rich element (RHAU). *J Biol Chem*, **288**, 35014-35027.
261. Sauer, M., Juranek, S.A., Marks, J., De Magis, A., Kazemier, H.G., Hilbig, D., Benhalevy, D., Wang, X., Hafner, M. and Paeschke, K. (2019) DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat Commun*, **10**, 2421.
262. Creacy, S.D., Routh, E.D., Iwamoto, F., Nagamine, Y., Akman, S.A. and Vaughn, J.P. (2008) G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J Biol Chem*, **283**, 34626-34634.
263. Murat, P., Marsico, G., Herdy, B., Ghanbarian, A.T., Portella, G. and Balasubramanian, S. (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol*, **19**, 229.

264. Shen, L. and Pelletier, J. (2020) General and Target-Specific DExD/H RNA Helicases in Eukaryotic Translation Initiation. *Int J Mol Sci*, **21**, 4402.
265. Philipp, S. and Katrin, P. (2020) The DEAH helicase DHX36 and its role in G-quadruplex-dependent processes. *Biological Chemistry*, 000010151520200292.
266. Yoo, J.S., Takahasi, K., Ng, C.S., Ouda, R., Onomoto, K., Yoneyama, M., Lai, J.C., Lattmann, S., Nagamine, Y., Matsui, T. *et al.* (2014) DHX36 enhances RIG-I signaling by facilitating PKR-mediated antiviral stress granule formation. *PLoS Pathog*, **10**, e1004012.
267. Chalupníková, K., Lattmann, S., Selak, N., Iwamoto, F., Fujiki, Y. and Nagamine, Y. (2008) Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. *J Biol Chem*, **283**, 35186-35198.
268. Zhang, Z., Kim, T., Bao, M., Facchinetti, V., Jung, S.Y., Ghaffari, A.A., Qin, J., Cheng, G. and Liu, Y.J. (2011) DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity*, **34**, 866-878.
269. Burns, C.C., Campagnoli, R., Shaw, J., Vincent, A., Jorba, J. and Kew, O. (2009) Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J Virol*, **83**, 9957-9969.
270. Atkinson, N.J., Witteveldt, J., Evans, D.J. and Simmonds, P. (2014) The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res*, **42**, 4527-4545.
271. Tulloch, F., Atkinson, N.J., Evans, D.J., Ryan, M.D. and Simmonds, P. (2014) RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife*, **3**.
272. Gaunt, E., Wise, H.M., Zhang, H., Lee, L.N., Atkinson, N.J., Nicol, M.Q., Highton, A.J., Klenerman, P., Beard, P.M., Dutia, B.M. *et al.* (2016) Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *Elife*, **5**, e12735.
273. Simmonds, P., Xia, W., Baillie, J.K. and McKinnon, K. (2013) Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla--selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. *BMC Genomics*, **14**, 610.
274. Duret, L. and Galtier, N. (2000) The covariation between TpA deficiency, CpG deficiency, and G + C content of human isochores is due to a mathematical artifact. *Mol Biol Evol*, **17**, 1620-1625.
275. Takata, M.A., Goncalves-Carneiro, D., Zang, T.M., Soll, S.J., York, A., Blanco-Melo, D. and Bieniasz, P.D. (2017) CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*, **550**, 124-127.

276. Odon, V., Fros, J.J., Goonawardane, N., Dietrich, I., Ibrahim, A., Alshaikhahmed, K., Nguyen, D. and Simmonds, P. (2019) The role of ZAP and OAS3/RNAseL pathways in the attenuation of an RNA virus with elevated frequencies of CpG and UpA dinucleotides. *Nucleic Acids Res*.
277. Guo, X., Ma, J., Sun, J. and Gao, G. (2007) The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proc Natl Acad Sci U S A*, **104**, 151-156.
278. Atasheva, S., Frolova, E.I. and Frolov, I. (2014) Interferon-stimulated poly(ADP-Ribose) polymerases are potent inhibitors of cellular translation and virus replication. *J Virol*, **88**, 2116-2130.
279. Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y.T. and Gao, G. (2011) Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc Natl Acad Sci U S A*, **108**, 15834-15839.
280. Zhu, Y., Wang, X., Goff, S.P. and Gao, G. (2012) Translational repression precedes and is required for ZAP-mediated mRNA decay. *Embo j*, **31**, 4236-4246.
281. Brady, P.N., Goel, A. and Johnson, M.A. (2019) Poly(ADP-Ribose) Polymerases in Host-Pathogen Interactions, Inflammation, and Immunity. *Microbiology and Molecular Biology Reviews*, **83**, e00038-00018.
282. Todorova, T., Bock, F.J. and Chang, P. (2015) Poly(ADP-ribose) polymerase-13 and RNA regulation in immunity and cancer. *Trends Mol Med*, **21**, 373-384.
283. Vyas, S., Chesarone-Cataldo, M., Todorova, T., Huang, Y.H. and Chang, P. (2013) A systematic analysis of the PARP protein family identifies new functions critical for cell physiology. *Nat Commun*, **4**, 2240.
284. Hayakawa, S., Shiratori, S., Yamato, H., Kameyama, T., Kitatsuji, C., Kashigi, F., Goto, S., Kameoka, S., Fujikura, D., Yamada, T. *et al.* (2011) ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. *Nat Immunol*, **12**, 37-44.
285. Guo, X., Carroll, J.W., Macdonald, M.R., Goff, S.P. and Gao, G. (2004) The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *J Virol*, **78**, 12781-12787.
286. Bick, M.J., Carroll, J.W., Gao, G., Goff, S.P., Rice, C.M. and MacDonald, M.R. (2003) Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J Virol*, **77**, 11555-11562.
287. Kerns, J.A., Emerman, M. and Malik, H.S. (2008) Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet*, **4**, e21.
288. Charron, G., Li, M.M., MacDonald, M.R. and Hang, H.C. (2013) Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform. *Proc Natl Acad Sci U S A*, **110**, 11085-11090.

289. Li, M.M.H., Aguilar, E.G., Michailidis, E., Pabon, J., Park, P., Wu, X., de Jong, Y.P., Schneider, W.M., Molina, H., Rice, C.M. *et al.* (2019) Characterization of Novel Splice Variants of Zinc Finger Antiviral Protein (ZAP). *J Virol*, **93**.
290. Ficarelli, M., Wilson, H., Pedro Galao, R., Mazzon, M., Antzin-Anduetza, I., Marsh, M., Neil, S.J. and Swanson, C.M. (2019) KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *Elife*, **8**, e46767.
291. Zheng, X., Wang, X., Tu, F., Wang, Q., Fan, Z. and Gao, G. (2017) TRIM25 Is Required for the Antiviral Activity of Zinc Finger Antiviral Protein. *J Virol*, **91**.
292. Wang, N., Dong, Q., Li, J., Jangra, R.K., Fan, M., Brasier, A.R., Lemon, S.M., Pfeffer, L.M. and Li, K. (2010) Viral induction of the zinc finger antiviral protein is IRF3-dependent but NF-kappaB-independent. *J Biol Chem*, **285**, 6080-6090.
293. Leung, A.K., Vyas, S., Rood, J.E., Bhutkar, A., Sharp, P.A. and Chang, P. (2011) Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. *Mol Cell*, **42**, 489-499.
294. Lee, H., Komano, J., Saitoh, Y., Yamaoka, S., Kozaki, T., Misawa, T., Takahama, M., Satoh, T., Takeuchi, O., Yamamoto, N. *et al.* (2013) Zinc-finger antiviral protein mediates retinoic acid inducible gene I-like receptor-independent antiviral response to murine leukemia virus. *Proc Natl Acad Sci U S A*, **110**, 12379-12384.
295. Law, L.M.J., Razooky, B.S., Li, M.M.H., You, S., Jurado, A., Rice, C.M. and MacDonald, M.R. (2019) ZAP's stress granule localization is correlated with its antiviral activity and induced by virus replication. *PLoS Pathog*, **15**, e1007798.
296. Gao, G., Guo, X. and Goff, S.P. (2002) Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science*, **297**, 1703-1706.
297. Muller, S., Moller, P., Bick, M.J., Wurr, S., Becker, S., Gunther, S. and Kummerer, B.M. (2007) Inhibition of filovirus replication by the zinc finger antiviral protein. *J Virol*, **81**, 2391-2400.
298. Chiu, H.P., Chiu, H., Yang, C.F., Lee, Y.L., Chiu, F.L., Kuo, H.C., Lin, R.J. and Lin, Y.L. (2018) Inhibition of Japanese encephalitis virus infection by the host zinc-finger antiviral protein. *PLoS Pathog*, **14**, e1007166.
299. Li, M., Yan, K., Wei, L., Yang, J., Lu, C., Xiong, F., Zheng, C. and Xu, W. (2015) Zinc finger antiviral protein inhibits coxsackievirus B3 virus replication and protects against viral myocarditis. *Antiviral Res*, **123**, 50-61.
300. Mao, R., Nie, H., Cai, D., Zhang, J., Liu, H., Yan, R., Cuconati, A., Block, T.M., Guo, J.T. and Guo, H. (2013) Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. *PLoS Pathog*, **9**, e1003494.

301. Xuan, Y., Liu, L., Shen, S., Deng, H. and Gao, G. (2012) Zinc finger antiviral protein inhibits murine gammaherpesvirus 68 M2 expression and regulates viral latency in cultured cells. *J Virol*, **86**, 12431-12434.
302. Zhou, A., Hassel, B.A. and Silverman, R.H. (1993) Expression cloning of 2-5A-dependent RNAase: A uniquely regulated mediator of interferon action. *Cell*, **72**, 753-765.
303. Han, Y., Donovan, J., Rath, S., Whitney, G., Chitrakar, A. and Korennykh, A. (2014) Structure of human RNase L reveals the basis for regulated RNA decay in the IFN response. *Science*, **343**, 1244-1248.
304. Han, Y., Whitney, G., Donovan, J. and Korennykh, A. (2012) Innate immune messenger 2-5A tethers human RNase L into active high-order complexes. *Cell Rep*, **2**, 902-913.
305. Huang, H., Zeqiraj, E., Dong, B., Jha, B.K., Duffy, N.M., Orlicky, S., Thevakumaran, N., Talukdar, M., Pillon, M.C., Ceccarelli, D.F. *et al.* (2014) Dimeric structure of pseudokinase RNase L bound to 2-5A reveals a basis for interferon-induced antiviral activity. *Mol Cell*, **53**, 221-234.
306. Wreschner, D.H., McCauley, J.W., Skehel, J.J. and Kerr, I.M. (1981) Interferon action--sequence specificity of the ppp(A2'p)nA-dependent ribonuclease. *Nature*, **289**, 414-417.
307. Li, Y., Banerjee, S., Wang, Y., Goldstein, S.A., Dong, B., Gaughan, C., Silverman, R.H. and Weiss, S.R. (2016) Activation of RNase L is dependent on OAS3 expression during infection with diverse human viruses. *Proc Natl Acad Sci U S A*, **113**, 2241-2246.
308. Ibsen, M.S., Gad, H.H., Thavachelvam, K., Boesen, T., Despres, P. and Hartmann, R. (2014) The 2'-5'-oligoadenylate synthetase 3 enzyme potently synthesizes the 2'-5'-oligoadenylates required for RNase L activation. *J Virol*, **88**, 14222-14231.
309. Schwartz, S.L. and Conn, G.L. (2019) RNA regulation of the antiviral protein 2'-5'-oligoadenylate synthetase. *Wiley Interdiscip Rev RNA*, **10**, e1534.
310. Han, J.Q., Wroblewski, G., Xu, Z., Silverman, R.H. and Barton, D.J. (2004) Sensitivity of hepatitis C virus RNA to the antiviral enzyme ribonuclease L is determined by a subset of efficient cleavage sites. *J Interferon Cytokine Res*, **24**, 664-676.
311. Le Roy, F., Bisbal, C., Silhol, M., Martinand, C., Lebleu, B. and Salehzada, T. (2001) The 2-5A/RNase L/RNase L inhibitor (RLI) [correction of (RNI)] pathway regulates mitochondrial mRNAs stability in interferon alpha-treated H9 cells. *J Biol Chem*, **276**, 48473-48482.
312. Tnani, M., Aliau, S. and Bayard, B. (1998) Localization of a molecular form of interferon-regulated RNase L in the cytoskeleton. *J Interferon Cytokine Res*, **18**, 361-368.
313. Salehzada, T., Silhol, M., Lebleu, B. and Bisbal, C. (1991) Polyclonal antibodies against RNase L. Subcellular localization of this enzyme in mouse cells. *J Biol Chem*, **266**, 5808-5813.

314. Bayard, B.A. and Gabrion, J.B. (1993) 2',5'-Oligoadenylate-dependent RNAse located in nuclei: biochemical characterization and subcellular distribution of the nuclease in human and murine cells. *Biochem J*, **296 (Pt 1)**, 155-160.
315. Baglioni, C., De Benedetti, A. and Williams, G.J. (1984) Cleavage of nascent reovirus mRNA by localized activation of the 2'-5'-oligoadenylate-dependent endoribonuclease. *J Virol*, **52**, 865-871.
316. Silverman, R.H., Skehel, J.J., James, T.C., Wreschner, D.H. and Kerr, I.M. (1983) rRNA cleavage as an index of ppp(A2'p)nA activity in interferon-treated encephalomyocarditis virus-infected cells. *J Virol*, **46**, 1051-1055.
317. Li, X.L., Blackford, J.A. and Hassel, B.A. (1998) RNase L mediates the antiviral effect of interferon through a selective reduction in viral RNA during encephalomyocarditis virus infection. *J Virol*, **72**, 2752-2759.
318. Martinand, C., Montavon, C., Salehzada, T., Silhol, M., Lebleu, B. and Bisbal, C. (1999) RNase L inhibitor is induced during human immunodeficiency virus type 1 infection and down regulates the 2-5A/RNase L pathway in human T cells. *J Virol*, **73**, 290-296.
319. Rice, A.P., Roberts, W.K. and Kerr, I.M. (1984) 2-5A accumulates to high levels in interferon-treated, vaccinia virus-infected cells in the absence of any inhibition of virus replication. *J Virol*, **50**, 220-228.
320. Maitra, R.K., McMillan, N.A., Desai, S., McSwiggen, J., Hovanessian, A.G., Sen, G., Williams, B.R. and Silverman, R.H. (1994) HIV-1 TAR RNA has an intrinsic ability to activate interferon-inducible enzymes. *Virology*, **204**, 823-827.
321. Mordechai, E., Kon, N., Henderson, E.E. and Suhadolnik, R.J. (1995) Activation of the interferon-inducible enzymes, 2',5'-oligoadenylate synthetase and PKR by human T-cell leukemia virus type I Rex-response element. *Virology*, **206**, 913-922.
322. Sharp, T.V., Raine, D.A., Gewert, D.R., Joshi, B., Jagus, R. and Clemens, M.J. (1999) Activation of the interferon-inducible (2'-5') oligoadenylate synthetase by the Epstein-Barr virus RNA, EBER-1. *Virology*, **257**, 303-313.
323. Desai, S.Y., Patel, R.C., Sen, G.C., Malhotra, P., Ghadge, G.D. and Thimmapaya, B. (1995) Activation of interferon-inducible 2'-5' oligoadenylate synthetase by adenoviral VAI RNA. *J Biol Chem*, **270**, 3454-3461.
324. Park, I.H., Kwon, Y.C., Ryu, W.S. and Ahn, B.Y. (2014) Inhibition of hepatitis B virus replication by ligand-mediated activation of RNase L. *Antiviral Res*, **104**, 118-127.
325. Burke, J.M., Moon, S.L., Matheny, T. and Parker, R. (2019) RNase L Reprograms Translation by Widespread mRNA Turnover Escaped by Antiviral mRNAs. *Mol Cell*, **75**, 1203-1217.e1205.

326. Le Roy, F., Salehzada, T., Bisbal, C., Dougherty, J.P. and Peltz, S.W. (2005) A newly discovered function for RNase L in regulating translation termination. *Nat Struct Mol Biol*, **12**, 505-512.
327. Donovan, J., Rath, S., Kolet-Mandrikov, D. and Korennykh, A. (2017) Rapid RNase L-driven arrest of protein synthesis in the dsRNA response without degradation of translation machinery. *Rna*, **23**, 1660-1671.
328. Wreschner, D.H., James, T.C., Silverman, R.H. and Kerr, I.M. (1981) Ribosomal RNA cleavage, nuclease activation and 2-5A(ppp(A2'p)nA) in interferon-treated cells. *Nucleic Acids Res*, **9**, 1571-1581.
329. Clemens, M.J. and Williams, B.R.G. (1978) Inhibition of cell-free protein synthesis by pppA^{2′} p^{5′} A^{2′} p^{5′} A: a novel oligonucleotide synthesized by interferon-treated L cell extracts. *Cell*, **13**, 565-572.
330. Schmidt, A., Zilberstein, A., Shulman, L., Federman, P., Berissi, H. and Revel, M. (1978) Interferon action: Isolation of Nuclease F, A translation inhibitor activated by interferon-induced (2′–5′) oligo-isoadenylate. *FEBS Letters*, **95**, 257-264.
331. Nilsen, T.W., Maroney, P.A. and Baglioni, C. (1981) Double-stranded RNA causes synthesis of 2',5'-oligo(A) and degradation of messenger RNA in interferon-treated cells. *J Biol Chem*, **256**, 7806-7811.
332. Hovanessian, A.G., Wood, J., Meurs, E. and Montagnier, L. (1979) Increased nuclease activity in cells treated with pppA2'p5'A2'p5' A. *Proc Natl Acad Sci U S A*, **76**, 3261-3265.
333. Chandrasekaran, K., Mehrabian, Z., Li, X.L. and Hassel, B. (2004) RNase-L regulates the stability of mitochondrial DNA-encoded mRNAs in mouse embryo fibroblasts. *Biochem Biophys Res Commun*, **325**, 18-23.
334. Bisbal, C. and Silverman, R.H. (2007) Diverse functions of RNase L and implications in pathology. *Biochimie*, **89**, 789-798.
335. Castelli, J.C., Hassel, B.A., Wood, K.A., Li, X.L., Amemiya, K., Dalakas, M.C., Torrence, P.F. and Youle, R.J. (1997) A study of the interferon antiviral mechanism: apoptosis activation by the 2-5A system. *J Exp Med*, **186**, 967-972.
336. Castelli, J.C., Hassel, B.A., Maran, A., Paranjape, J., Hewitt, J.A., Li, X.L., Hsu, Y.T., Silverman, R.H. and Youle, R.J. (1998) The role of 2'-5' oligoadenylate-activated ribonuclease L in apoptosis. *Cell Death Differ*, **5**, 313-320.
337. Diaz-Guerra, M., Rivas, C. and Esteban, M. (1997) Activation of the IFN-inducible enzyme RNase L causes apoptosis of animal cells. *Virology*, **236**, 354-363.
338. Rusch, L., Zhou, A. and Silverman, R.H. (2000) Caspase-dependent apoptosis by 2',5'-oligoadenylate activation of RNase L is enhanced by IFN-beta. *J Interferon Cytokine Res*, **20**, 1091-1100.

339. Malathi, K., Dong, B., Gale, M., Jr. and Silverman, R.H. (2007) Small self-RNA generated by RNase L amplifies antiviral innate immunity. *Nature*, **448**, 816-819.
340. Flodstrom-Tullberg, M., Hultcrantz, M., Stotland, A., Maday, A., Tsai, D., Fine, C., Williams, B., Silverman, R. and Sarvetnick, N. (2005) RNase L and double-stranded RNA-dependent protein kinase exert complementary roles in islet cell defense during coxsackievirus infection. *J Immunol*, **174**, 1171-1177.
341. Samuel, M.A., Whitby, K., Keller, B.C., Marri, A., Barchet, W., Williams, B.R., Silverman, R.H., Gale, M., Jr. and Diamond, M.S. (2006) PKR and RNase L contribute to protection against lethal West Nile Virus infection by controlling early viral spread in the periphery and replication in neurons. *J Virol*, **80**, 7009-7019.
342. Min, J.Y. and Krug, R.M. (2006) The primary function of RNA binding by the influenza A virus NS1 protein in infected cells: Inhibiting the 2'-5' oligo (A) synthetase/RNase L pathway. *Proc Natl Acad Sci U S A*, **103**, 7100-7105.
343. Whelan, J.N., Li, Y., Silverman, R.H. and Weiss, S.R. (2019) Zika Virus Production Is Resistant to RNase L Antiviral Activity. *J Virol*, **93**.
344. Hersh, C.L., Brown, R.E., Roberts, W.K., Swyryd, E.A., Kerr, I.M. and Stark, G.R. (1984) Simian virus 40-infected, interferon-treated cells contain 2',5'-oligoadenylates which do not activate cleavage of RNA. *J Biol Chem*, **259**, 1731-1737.
345. Drappier, M. and Michiels, T. (2015) Inhibition of the OAS/RNase L pathway by viruses. *Curr Opin Virol*, **15**, 19-26.
346. Akdis, M., Aab, A., Altunbulakli, C., Azkur, K., Costa, R.A., Crameri, R., Duan, S., Eiwegger, T., Eljaszewicz, A., Ferstl, R. *et al.* (2016) Interleukins (from IL-1 to IL-38), interferons, transforming growth factor β, and TNF-α: Receptors, functions, and roles in diseases. *Journal of Allergy and Clinical Immunology*, **138**, 984-1010.
347. Villarino, A.V., Kanno, Y. and O'Shea, J.J. (2017) Mechanisms and consequences of Jak–STAT signaling in the immune system. *Nature Immunology*, **18**, 374-384.
348. Schneider, W.M., Chevillotte, M.D. and Rice, C.M. (2014) Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol*, **32**, 513-545.
349. Schoggins, J.W. (2019) Interferon-Stimulated Genes: What Do They All Do? *Annual Review of Virology*, **6**, 567-584.
350. Knipe, D.M. and Howley, P.M. (2013) *Fields Virology*. 6th ed. Wolters Kluwer Health, Philadelphia.
351. Webb, B., Rakibuzzaman, A.G.M. and Ramamoorthy, S. (2020) Torque teno viruses in health and disease. *Virus Research*, **285**, 198013.
352. Kaczorowska, J. and van der Hoek, L. (2020) Human anelloviruses: diverse, omnipresent and commensal members of the virome. *FEMS Microbiology Reviews*, **44**, 305-313.

353. Crawford, L.V. (1966) A Minute Virus of Mice. *Virology*, **29**, 605-&.
354. Kraft, V. and Meyer, B. (1990) Seromonitoring in Small Laboratory-Animal Colonies - a 5 Year Survey - 1984-1988. *Z Versuchstierkd*, **33**, 29-35.
355. Parker, J.C., Collins, M.J., Cross, S.S. and Rowe, W.P. (1970) Minute Virus of Mice .2. Prevalence, Epidemiology, and Occurrence as a Contaminant of Transplanted Tumors. *J Natl Cancer I*, **45**, 305-&.
356. Brownstein, D.G., Smith, A.L., Jacoby, R.O., Johnson, E.A., Hansen, G. and Tattersall, P. (1991) Pathogenesis of infection with a virulent allotropic variant of minute virus of mice and regulation by host genotype. *Lab Invest*, **65**, 357-364.
357. Brownstein, D.G., Smith, A.L., Johnson, E.A., Pintel, D.J., Naeger, L.K. and Tattersall, P. (1992) The pathogenesis of infection with minute virus of mice depends on expression of the small nonstructural protein NS2 and on the genotype of the allotropic determinants VP1 and VP2. *J Virol*, **66**, 3118-3124.
358. Kimsey, P.B., Engers, H.D., Hirt, B. and Jongeneel, C.V. (1986) Pathogenicity of fibroblast- and lymphocyte-specific variants of minute virus of mice. *J Virol*, **59**, 8-13.
359. Kilham, L. and Margolis, G. (1970) Pathogenicity of minute virus of mice (MVM) for rats, mice, and hamsters. *Proc Soc Exp Biol Med*, **133**, 1447-1452.
360. Kilham, L. and Margolis, G. (1971) Fetal infections of hamsters, rats, and mice induced with the minute virus of mice (MVM). *Teratology*, **4**, 43-61.
361. Segovia, J.C., Bueren, J.A. and Almendral, J.M. (1995) Myeloid depression follows infection of susceptible newborn mice with the parvovirus minute virus of mice (strain i). *J Virol*, **69**, 3229-3232.
362. Janus, L.M., Mahler, M., Kohl, W., Smoczek, A., Hedrich, H.J. and Bleich, A. (2008) Minute virus of mice: antibody response, viral shedding, and persistence of viral DNA in multiple strains of mice. *Comp Med*, **58**, 360-368.
363. Kontou, M., Govindasamy, L., Nam, H.J., Bryant, N., Llamas-Saiz, A.L., Foces-Foces, C., Hernando, E., Rubio, M.P., McKenna, R., Almendral, J.M. *et al.* (2005) Structural determinants of tissue tropism and in vivo pathogenicity for the parvovirus minute virus of mice. *J Virol*, **79**, 10931-10943.
364. Etingov, I., Itah, R., Mincberg, M., Keren-Naus, A., Nam, H.-J., Agbandje-McKenna, M. and Davis, C. (2008) An extension of the Minute Virus of Mice tissue tropism. *Virology*, **379**, 245-255.
365. Cotmore, S.F., Agbandje-McKenna, M., Canuti, M., Chiorini, J.A., Eis-Hubinger, A.-M., Hughes, J., Mietzsch, M., Modha, S., Ogliastro, M., Pénzes, J.J. *et al.* (2019) ICTV Virus Taxonomy Profile: Parvoviridae. *Journal of General Virology*, **100**, 367-368.

366. Meier, A.F., Fraefel, C. and Seyffert, M. (2020) The Interplay between Adeno-Associated Virus and its Helper Viruses. *Viruses*, **12**.
367. Agbandje-McKenna, M., Llamas-Saiz, A.L., Wang, F., Tattersall, P. and Rossmann, M.G. (1998) Functional implications of the structure of the murine parvovirus, minute virus of mice. *Structure*, **6**, 1369-1381.
368. Tattersall, P., Cawte, P.J., Shatkin, A.J. and Ward, D.C. (1976) Three structural polypeptides coded for by minite virus of mice, a parvovirus. *J Virol*, **20**, 273-289.
369. Mani, B., Baltzer, C., Valle, N., Almendral, J.M., Kempf, C. and Ros, C. (2006) Low pH-Dependent Endosomal Processing of the Incoming Parvovirus Minute Virus of Mice Virion Leads to Externalization of the VP1 N-Terminal Sequence (N-VP1), N-VP2 Cleavage, and Uncoating of the Full-Length Genome. *J Virol*, **80**, 1015.
370. Plevka, P., Hafenstein, S., Li, L., D'Abrgamo, A., Jr., Cotmore, S.F., Rossmann, M.G. and Tattersall, P. (2011) Structure of a packaging-defective mutant of minute virus of mice indicates that the genome is packaged via a pore at a 5-fold axis. *J Virol*, **85**, 4822-4827.
371. Cotmore, S.F. and Tattersall, P. (1988) The NS-1 polypeptide of minute virus of mice is covalently attached to the 5' termini of duplex replicative-form DNA and progeny single strands. *J Virol*, **62**, 851-860.
372. Cotmore, S.F. and Tattersall, P. (1989) A genome-linked copy of the NS-1 polypeptide is located on the outside of infectious parvovirus particles. *J Virol*, **63**, 3902.
373. Cotmore, S.F. and Tattersall, P. (2005) Encapsidation of minute virus of mice DNA: aspects of the translocation mechanism revealed by the structure of partially packaged genomes. *Virology*, **336**, 100-112.
374. Astell, C.R., Thomson, M., Merchlinsky, M. and Ward, D.C. (1983) The complete DNA sequence of minute virus of mice, an autonomous parvovirus. *Nucleic Acids Res*, **11**, 999-1018.
375. Pintel, D., Dadachanji, D., Astell, C.R. and Ward, D.C. (1983) The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. *Nucleic Acids Res*, **11**, 1019-1038.
376. Cotmore, S.F., Sturzenbecker, L.J. and Tattersall, P. (1983) The autonomous parvovirus MVM encodes two nonstructural proteins in addition to its capsid polypeptides. *Virology*, **129**, 333-343.
377. Li, L., Cotmore, S.F. and Tattersall, P. (2013) Parvoviral Left-End Hairpin Ears Are Essential during Infection for Establishing a Functional Intranuclear Transcription Template and for Efficient Progeny Genome Encapsidation. *J Virol*, **87**, 10501-10514.
378. Li, L., Cotmore, S.F. and Tattersall, P. (2012) Maintenance of the Flip Sequence Orientation of the Ears in the Parvoviral Left-End Hairpin Is a Nonessential Consequence of the Critical Asymmetry in the Hairpin Stem. *J Virol*, **86**, 12187-12197.

379. Clemens, K.E. and Pintel, D.J. (1988) The two transcription units of the autonomous parvovirus minute virus of mice are transcribed in a temporal order. *J Virol*, **62**, 1448-1451.
380. Deleu, L., Pujol, A., Faisst, S. and Rommelaere, J. (1999) Activation of promoter P4 of the autonomous parvovirus minute virus of mice at early S phase is required for productive infection. *J Virol*, **73**, 3877-3885.
381. Doerig, C., Hirt, B., Beard, P. and Antonietti, J.P. (1988) Minute virus of mice non-structural protein NS-1 is necessary and sufficient for trans-activation of the viral P39 promoter. *J Gen Virol*, **69 (Pt 10)**, 2563-2573.
382. Clemens, K.E. and Pintel, D. (1987) Minute virus of mice (MVM) mRNAs predominantly polyadenylate at a single site. *Virology*, **160**, 511-514.
383. Jongeneel, C.V., Sahli, R., McMaster, G.K. and Hirt, B. (1986) A precise map of splice junctions in the mRNAs of minute virus of mice, an autonomous parvovirus. *J Virol*, **59**, 564-573.
384. Haut, D.D. and Pintel, D.J. (1998) Intron definition is required for excision of the minute virus of mice small intron and definition of the upstream exon. *J Virol*, **72**, 1834-1843.
385. Haut, D.D. and Pintel, D.J. (1999) Inclusion of the NS2-specific exon in minute virus of mice mRNA is facilitated by an intronic splicing enhancer that affects definition of the downstream small intron. *Virology*, **258**, 84-94.
386. Cotmore, S.F. and Tattersall, P. (1986) Organization of nonstructural genes of the autonomous parvovirus minute virus of mice. *J Virol*, **58**, 724-732.
387. Cotmore, S.F. and Tattersall, P. (1990) Alternate splicing in a parvoviral nonstructural gene links a common amino-terminal sequence to downstream domains which confer radically different localization and turnover characteristics. *Virology*, **177**, 477-487.
388. Colomar, M.C., Hirt, B. and Beard, P. (1998) Two segments in the genome of the immunosuppressive minute virus of mice determine the host-cell specificity, control viral DNA replication and affect viral RNA metabolism. *J Gen Virol*, **79 (Pt 3)**, 581-586.
389. Choi, E.Y., Newman, A.E., Burger, L. and Pintel, D. (2005) Replication of minute virus of mice DNA is critically dependent on accumulated levels of NS2. *J Virol*, **79**, 12375-12381.
390. Schoborg, R.V. and Pintel, D.J. (1991) Accumulation of MVM gene products is differentially regulated by transcription initiation, RNA processing and protein stability. *Virology*, **181**, 22-34.
391. Naeger, L.K., Salomé, N. and Pintel, D.J. (1993) NS2 is required for efficient translation of viral mRNA in minute virus of mice-infected murine cells. *J Virol*, **67**, 1034-1043.

392. Berns, K.I. and Parrish, C.R. (2013) In Knipe, D. M. and Howley, P. M. (eds.), *Fields Virology* 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA, pp. 1768-1791.
393. Garcin, P.O. and Pante, N. (2015) The minute virus of mice exploits different endocytic pathways for cellular uptake. *Virology*, **482**, 157-166.
394. López-Bueno, A., Rubio, M.-P., Bryant, N., McKenna, R., Agbandje-McKenna, M. and Almendral, J.M. (2006) Host-Selected Amino Acid Changes at the Sialic Acid Binding Pocket of the Parvovirus Capsid Modulate Cell Binding Affinity and Determine Virulence. *J Virol*, **80**, 1563.
395. Linser, P., Bruning, H. and Armentrout, R.W. (1979) Uptake of minute virus of mice into cultured rodent cells. *J Virol*, **31**, 537-545.
396. Mäntylä, E., Kann, M. and Vihinen-Ranta, M. (2017) Protoparvovirus Knocking at the Nuclear Door. *Viruses*, **9**.
397. Lombardo, E., Ramírez, J.C., Garcia, J. and Almendral, J.M. (2002) Complementary Roles of Multiple Nuclear Targeting Signals in the Capsid Proteins of the Parvovirus Minute Virus of Mice during Assembly and Onset of Infection. *J Virol*, **76**, 7049.
398. Bashir, T., Rommelaere, J. and Cziepluch, C. (2001) In vivo accumulation of cyclin A and cellular replication factors in autonomous parvovirus minute virus of mice-associated replication bodies. *J Virol*, **75**, 4394-4398.
399. Wolter, S., Richards, R. and Armentrout, R.W. (1980) Cell cycle-dependent replication of the DNA of minute virus of mice, a parvovirus. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, **607**, 420-431.
400. Op De Beeck, A., Anouja, F., Mousset, S., Rommelaere, J. and Caillet-Fauquet, P. (1995) The nonstructural proteins of the autonomous parvovirus minute virus of mice interfere with the cell cycle, inducing accumulation in G2. *Cell Growth Differ*, **6**, 781-787.
401. Op De Beeck, A., Sobczak-Thepot, J., Sirma, H., Bourgain, F., Brechot, C. and Caillet-Fauquet, P. (2001) NS1- and minute virus of mice-induced cell cycle arrest: involvement of p53 and p21(cip1). *J Virol*, **75**, 11071-11078.
402. Adeyemi, R.O. and Pintel, D.J. (2014) Parvovirus-induced depletion of cyclin B1 prevents mitotic entry of infected cells. *PLoS Pathog*, **10**, e1003891.
403. Adeyemi, R.O. and Pintel, D.J. (2012) Replication of minute virus of mice in murine cells is facilitated by virally induced depletion of p21. *J Virol*, **86**, 8328-8332.
404. Ruiz, Z., Mihaylov, I.S., Cotmore, S.F. and Tattersall, P. (2011) Recruitment of DNA replication and damage response proteins to viral replication centers during infection with NS2 mutants of Minute Virus of Mice (MVM). *Virology*, **410**, 375-384.

405. Chen, A.Y. and Qiu, J. (2010) Parvovirus infection-induced cell death and cell cycle arrest. *Future Virol*, **5**, 731-743.
406. Itah, R., Tal, J. and Davis, C. (2004) Host cell specificity of minute virus of mice in the developing mouse embryo. *J Virol*, **78**, 9474-9486.
407. Marchini, A., Bonifati, S., Scott, E.M., Angelova, A.L. and Rommelaere, J. (2015) Oncolytic parvoviruses: from basic virology to clinical applications. *Virol J*, **12**, 6.
408. Tattersall, P. and Ward, D.C. (1976) Rolling hairpin model for replication of parvovirus and linear chromosomal DNA. *Nature*, **263**, 106-109.
409. Astell, C.R., Chow, M.B. and Ward, D.C. (1985) Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. *J Virol*, **54**, 171-177.
410. Costello, E., Sahli, R., Hirt, B. and Beard, P. (1995) The mismatched nucleotides in the 5'-terminal hairpin of minute virus of mice are required for efficient viral DNA replication. *J Virol*, **69**, 7489-7496.
411. Cotmore, S.F. and Tattersall, P. (2003) Resolution of Parvovirus Dimer Junctions Proceeds through a Novel Heterocruciform Intermediate. *J Virol*, **77**, 6245.
412. Cotmore, S.F. and Tattersall, P. (1986) The NS-1 polypeptide of the autonomous parvovirus MVM is a nuclear phosphoprotein. *Virus Res*, **4**, 243-250.
413. Corbau, R., Salomé, N., Rommelaere, J. and Nüesch, J.P.F. (1999) Phosphorylation of the Viral Nonstructural Protein NS1 during MVMp Infection of A9 Cells. *Virology*, **259**, 402-415.
414. Riolobos, L., Reguera, J., Mateu, M.G. and Almendral, J.M. (2006) Nuclear Transport of Trimeric Assembly Intermediates Exerts a Morphogenetic Control on the Icosahedral Parvovirus Capsid. *Journal of Molecular Biology*, **357**, 1026-1038.
415. Zádori, Z., Szelei, J., Lacoste, M.-C., Li, Y., Gariépy, S., Raymond, P., Allaire, M., Nabi, I.R. and Tijssen, P. (2001) A Viral Phospholipase A2 Is Required for Parvovirus Infectivity. *Developmental Cell*, **1**, 291-302.
416. Bar, S., Rommelaere, J. and Nuesch, J.P. (2013) Vesicular transport of progeny parvovirus particles through ER and Golgi regulates maturation and cytolysis. *PLoS Pathog*, **9**, e1003605.
417. Wolfisberg, R., Kempf, C. and Ros, C. (2016) Late Maturation Steps Preceding Selective Nuclear Export and Egress of Progeny Parvovirus. *J Virol*, **90**, 5462-5474.
418. Gross, L. (1953) A filterable agent, recovered from Ak leukemic extracts, causing salivary gland carcinomas in C3H mice. *Proc Soc Exp Biol Med*, **83**, 414-421.
419. Huang, G., Zeng, G., Huang, Y., Ramaswami, B. and Randhawa, P. (2016) Evaluation of the Gastrointestinal Tract as Potential Route of Primary Polyomavirus Infection in Mice. *PLOS ONE*, **11**, e0150786.

420. Swanson, P.A., 2nd, Lukacher, A.E. and Szomolanyi-Tsuda, E. (2009) Immunity to polyomavirus infection: the polyomavirus-mouse model. *Semin Cancer Biol*, **19**, 244-251.
421. Kemball, C.C., Lee, E.D., Vezys, V., Pearson, T.C., Larsen, C.P. and Lukacher, A.E. (2005) Late priming and variability of epitope-specific CD8+ T cell responses during a persistent virus infection. *J Immunol*, **174**, 7950-7960.
422. Atencio, I.A., Shadan, F.F., Zhou, X.J., Vaziri, N.D. and Villarreal, L.P. (1993) Adult mouse kidneys become permissive to acute polyomavirus infection and reactivate persistent infections in response to cellular damage and regeneration. *J Virol*, **67**, 1424-1432.
423. Szomolanyi-Tsuda, E., Dundon, P.L., Joris, I., Shultz, L.D., Woda, B.A. and Welsh, R.M. (1994) Acute, lethal, natural killer cell-resistant myeloproliferative disease induced by polyomavirus in severe combined immunodeficient mice. *American Journal of Pathology*, **144**, 359-371.
424. Rochford, R., Moreno, J.P., Peake, M.L. and Villarreal, L.P. (1992) Enhancer dependence of polyomavirus persistence in mouse kidneys. *J Virol*, **66**, 3287-3297.
425. Stehle, T. and Harrison, S.C. (1996) Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure*, **4**, 183-194.
426. Krüger, D.H. (1982) J. Tooze (Editor), Molecular Biology of Tumor Viruses (2nd Edition), Part 2: DNA Tumor Viruses. IX + 958 S., 116 Abb., 47 Tab., 20 Tafeln. Cold Spring Harbor 1980. Cold Spring Harbor Laboratory. $ 90.00. *Zeitschrift für allgemeine Mikrobiologie*, **22**, 79-79.
427. Carmichael, G.G. (2016) Gene Regulation and Quality Control in Murine Polyomavirus Infection. *Viruses*, **8**, 284.
428. Griffin, B.E. and Fried, M. (1975) Amplification of a specific region of the polyoma virus genome. *Nature*, **256**, 175-179.
429. Kamen, R., Favaloro, J. and Parker, J. (1980) Topography of the three late mRNA's of polyoma virus which encode the virion proteins. *J Virol*, **33**, 637-651.
430. Kamen, R., Favaloro, J., Parker, J., Treisman, R., Lania, L., Fried, M. and Mellor, A. (1980) Comparison of polyoma virus transcription in productively infected mouse cells and transformed rodent cell lines. *Cold Spring Harb Symp Quant Biol*, **44 Pt 1**, 63-75.
431. Tsai, B., Gilbert, J.M., Stehle, T., Lencer, W., Benjamin, T.L. and Rapoport, T.A. (2003) Gangliosides are receptors for murine polyoma virus and SV40. *The EMBO journal*, **22**, 4346-4355.
432. Gilbert, J. and Benjamin, T. (2004) Uptake pathway of polyomavirus via ganglioside GD1a. *J Virol*, **78**, 12259-12267.
433. Liebl, D., Difato, F., Hornikova, L., Mannova, P., Stokrova, J. and Forstova, J. (2006) Mouse polyomavirus enters early endosomes,

requires their acidic pH for productive infection, and meets transferrin cargo in Rab11-positive endosomes. *J Virol*, **80**, 4610-4622.
434. Qian, M., Cai, D., Verhey, K.J. and Tsai, B. (2009) A lipid receptor sorts polyomavirus from the endolysosome to the endoplasmic reticulum to cause infection. *PLoS Pathog*, **5**, e1000465.
435. Magnuson, B., Rainey, E.K., Benjamin, T., Baryshev, M., Mkrtchian, S. and Tsai, B. (2005) ERp29 triggers a conformational change in polyomavirus to stimulate membrane binding. *Mol Cell*, **20**, 289-300.
436. Huerfano, S., Ryabchenko, B., Spanielova, H. and Forstova, J. (2017) Hydrophobic domains of mouse polyomavirus minor capsid proteins promote membrane association and virus exit from the ER. *Febs j*, **284**, 883-902.
437. Soldatova, I., Prilepskaja, T., Abrahamyan, L., Forstova, J. and Huerfano, S. (2018) Interaction of the Mouse Polyomavirus Capsid Proteins with Importins Is Required for Efficient Import of Viral DNA into the Cell Nucleus. *Viruses*, **10**.
438. DeCaprio, J.A., Imperiale, M.J. and Major, E.O. (2013) In Knipe, D. M. and Howley, P. M. (eds.), *Fields Virology* 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA, pp. 1633-1661.
439. Simmonds, P. (2012) SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes*, **5**, 50.
440. Greenbaum, B.D., Rabadan, R. and Levine, A.J. (2009) Patterns of Oligonucleotide Sequences in Viral and Host Cell RNA Identify Mediators of the Host Innate Immune System. *PLOS ONE*, **4**, e5969.
441. Bretscher, C. and Marchini, A. (2019) H-1 Parvovirus as a Cancer-Killing Agent: Past, Present, and Future. *Viruses*, **11**, 562.
442. Merchlinsky, M.J., Tattersall, P.J., Leary, J.J., Cotmore, S.F., Gardiner, E.M. and Ward, D.C. (1983) Construction of an infectious molecular clone of the autonomous parvovirus minute virus of mice. *J Virol*, **47**, 227-232.
443. Shein, H.M. and Enders, J.F. (1962) Multiplication and cytopathogenicity of Simian vacuolating virus 40 in cultures of human tissues. *Proc Soc Exp Biol Med*, **109**, 495-500.
444. Littlefield, J.W. (1964) Three Degrees of Guanylic Acid--Inosinic Acid Pyrophosphorylase Deficiency in Mouse Fibroblasts. *Nature*, **203**, 1142-1144.
445. Hron, T., Spanielova, H., Suchanova, J. and Forstova, J. (2013) The Cre/loxP recombination system for production of infectious mouse polyomavirus. *Virus Res*, **176**, 128-136.
446. Hornikova, L., Zila, V., Spanielova, H. and Forstova, J. (2015) Mouse Polyomavirus: Propagation, Purification, Quantification, and Storage. *Curr Protoc Microbiol*, **38**, 14f.11.11-26.
447. Ryndock, E.J., Biryukov, J. and Meyers, C. (2015) In Keppler, D. and Lin, A. W. (eds.), *Cervical Cancer: Methods and Protocols*. Springer New York, New York, NY, pp. 39-52.

448. Bienkowska-Haba, M., Luszczek, W., Myers, J.E., Keiffer, T.R., DiGiuseppe, S., Polk, P., Bodily, J.M., Scott, R.S. and Sapp, M. (2018) A new cell culture model to genetically dissect the complete human papillomavirus life cycle. *PLOS Pathogens*, **14**, e1006846.
449. Cladel, N.M., Budgeon, L.R., Balogh, K.K., Cooper, T.K., Brendle, S.A., Christensen, N.D., Schell, T.D. and Hu, J. (2017) Mouse papillomavirus infection persists in mucosal tissues of an immunocompetent mouse strain and progresses to cancer. *Scientific Reports*, **7**, 16932.
450. Spurgeon, M.E., Uberoi, A., McGregor, S.M., Wei, T., Ward-Shaw, E. and Lambert, P.F. (2019) A Novel In Vivo Infection Model To Study Papillomavirus-Mediated Disease of the Female Reproductive Tract. *mBio*, **10**, e00180-00119.
451. Wei, T., Buehler, D., Ward-Shaw, E. and Lambert, P.F. (2020) An Infection-Based Murine Model for Papillomavirus-Associated Head and Neck Cancer. *mBio*, **11**, e00908-00920.
452. Ball-Goodrich, L.J. and Tattersall, P. (1992) Two amino acid substitutions within the capsid are coordinately required for acquisition of fibrotropism by the lymphotropic strain of minute virus of mice. *J Virol*, **66**, 3415-3423.
453. Chin, J.X., Chung, B.K.-S. and Lee, D.-Y. (2014) Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics*, **30**, 2210-2212.
454. Hitoshi, N., Ken-ichi, Y. and Jun-ichi, M. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene*, **108**, 193-199.
455. Jun-ichi, M., Satoshi, T., Kimi, A., Fumi, T., Akira, T., Kiyoshi, T. and Ken-ichi, Y. (1989) Expression vector system based on the chicken β-actin promoter directs efficient production of interleukin-5. *Gene*, **79**, 269-277.
456. Maroto, B., Ramirez, J.C. and Almendral, J.M. (2000) Phosphorylation status of the parvovirus minute virus of mice particle: mapping and biological relevance of the major phosphorylation sites. *J Virol*, **74**, 10892-10902.
457. Boissy, R. and Astell, C.R. (1985) An Escherichia coli recBCsbcBrecF host permits the deletion-resistant propagation of plasmid clones containing the 5'-terminal palindrome of minute virus of mice. *Gene*, **35**, 179-185.
458. Tattersall, P. and Bratton, J. (1983) Reciprocal productive and restrictive virus-cell interactions of immunosuppressive and prototype strains of minute virus of mice. *J Virol*, **46**, 944-955.
459. Heldt, C.L., Hernandez, R., Mudiganti, U., Gurgel, P.V., Brown, D.T. and Carbonell, R.G. (2006) A colorimetric assay for viral agents that produce cytopathic effects. *Journal of Virological Methods*, **135**, 56-65.
460. Zhan, D., Roy, M.R., Valera, C., Cardenas, J., Vennari, J.C., Chen, J.W. and Liu, S. (2002) Detection of minute virus of mice using real time

quantitative PCR in assessment of virus clearance during the purification of Mammalian cell substrate derived biotherapeutics. *Biologicals*, **30**, 259-270.
461. Veal, C.D., Freeman, P.J., Jacobs, K., Lancaster, O., Jamain, S., Leboyer, M., Albanes, D., Vaghela, R.R., Gut, I., Chanock, S.J. *et al.* (2012) A mechanistic basis for amplification differences between samples and between genome regions. *BMC Genomics*, **13**, 455.
462. Ventoso, I., Berlanga, J.J. and Almendral, J.M. (2010) Translation control by protein kinase R restricts minute virus of mice infection: role in parvovirus oncolysis. *J Virol*, **84**, 5043-5051.
463. Rommelaere, J. and Cornelis, J.J. (1991) Antineoplasic activity of parvoviruses. *Journal of Virological Methods*, **33**, 233-251.
464. Raykov, Z., Grekova, S.P., Hörlein, R., Leuchs, B., Giese, T., Giese, N.A., Rommelaere, J., Zawatzky, R. and Daeffler, L. (2013) TLR-9 contributes to the antiviral innate immune sensing of rodent parvoviruses MVMp and H-1PV by normal human immune cells. *PloS one*, **8**, e55086-e55086.
465. Mattei, L.M., Cotmore, S.F., Li, L., Tattersall, P. and Iwasaki, A. (2013) Toll-like receptor 9 in plasmacytoid dendritic cells fails to detect parvoviruses. *J Virol*, **87**, 3605-3608.
466. Cohen, P. (2014) The TLR and IL-1 signalling network at a glance. *Journal of Cell Science*, **127**, 2383.
467. Weber, A., Wasiliew, P. and Kracht, M. (2010) Interleukin-1 (IL-1) Pathway. *Science Signaling*, **3**, cm1.
468. Stunz, L.L., Lenert, P., Peckham, D., Yi, A.K., Haxhinasto, S., Chang, M., Krieg, A.M. and Ashman, R.F. (2002) Inhibitory oligonucleotides specifically block effects of stimulatory CpG oligonucleotides in B cells. *Eur J Immunol*, **32**, 1212-1222.
469. Kurt-Jones, E.A., Sandor, F., Ortiz, Y., Bowen, G.N., Counter, S.L., Wang, T.C. and Finberg, R.W. (2004) Use of murine embryonic fibroblasts to define Toll-like receptor activation and specificity. *J Endotoxin Res*, **10**, 419-424.
470. Hoelzer, K., Shackelton, L.A. and Parrish, C.R. (2008) Presence and role of cytosine methylation in DNA viruses of animals. *Nucleic Acids Res*, **36**, 2825-2837.
471. Vincent, I.E., Balmelli, C., Meehan, B., Allan, G., Summerfield, A. and McCullough, K.C. (2007) Silencing of natural interferon producing cell activation by porcine circovirus type 2 DNA. *Immunology*, **120**, 47-56.
472. Vincent, I.E., Carrasco, C.P., Guzylack-Piriou, L., Herrmann, B., McNeilly, F., Allan, G.M., Summerfield, A. and McCullough, K.C. (2005) Subset-dependent modulation of dendritic cell activity by circovirus type 2. *Immunology*, **115**, 388-398.
473. Wikström, F.H., Meehan, B.M., Berg, M., Timmusk, S., Elving, J., Fuxler, L., Magnusson, M., Allan, G.M., McNeilly, F. and Fossum, C. (2007) Structure-Dependent Modulation of Alpha Interferon Production by

Porcine Circovirus 2 Oligodeoxyribonucleotide and CpG DNAs in Porcine Peripheral Blood Mononuclear Cells. *J Virol*, **81**, 4919-4927.
474. Toth, R., Meszaros, I., Stefancsik, R., Bartha, D., Balint, A. and Zadori, Z. (2013) CpG distribution and methylation pattern in porcine parvovirus. *PLoS One*, **8**, e85986.
475. Zhu, J., Huang, X. and Yang, Y. (2009) The TLR9-MyD88 pathway is critical for adaptive immune responses to adeno-associated virus gene therapy vectors in mice. *J Clin Invest*, **119**, 2388-2398.
476. Martino, A.T., Suzuki, M., Markusic, D.M., Zolotukhin, I., Ryals, R.C., Moghimi, B., Ertl, H.C., Muruve, D.A., Lee, B. and Herzog, R.W. (2011) The genome of self-complementary adeno-associated viral vectors increases Toll-like receptor 9-dependent innate immune responses in the liver. *Blood*, **117**, 6459-6468.
477. Rogers, G.L., Suzuki, M., Zolotukhin, I., Markusic, D.M., Morel, L.M., Lee, B., Ertl, H.C. and Herzog, R.W. (2015) Unique Roles of TLR9- and MyD88-Dependent and -Independent Pathways in Adaptive Immune Responses to AAV-Mediated Gene Transfer. *J Innate Immun*, **7**, 302-314.
478. Ashley, S.N., Somanathan, S., Giles, A.R. and Wilson, J.M. (2019) TLR9 signaling mediates adaptive immunity following systemic AAV gene therapy. *Cell Immunol*, **346**, 103997.
479. Herzog, R.W., Cooper, M., Perrin, G.Q., Biswas, M., Martino, A.T., Morel, L., Terhorst, C. and Hoffman, B.E. (2019) Regulatory T cells and TLR9 activation shape antibody formation to a secreted transgene product in AAV muscle gene transfer. *Cell Immunol*, **342**, 103682.
480. Butterfield, J.S.S., Biswas, M., Shirley, J.L., Kumar, S.R.P., Sherman, A., Terhorst, C., Ling, C. and Herzog, R.W. (2019) TLR9-Activating CpG-B ODN but Not TLR7 Agonists Triggers Antibody Formation to Factor IX in Muscle Gene Transfer. *Hum Gene Ther Methods*, **30**, 81-92.
481. Faust, S.M., Bell, P., Cutler, B.J., Ashley, S.N., Zhu, Y., Rabinowitz, J.E. and Wilson, J.M. (2013) CpG-depleted adeno-associated virus vectors evade immune detection. *J Clin Invest*, **123**, 2994-3001.
482. Hull, C.M. and Bevilacqua, P.C. (2016) Discriminating Self and Non-Self by RNA: Roles for RNA Structure, Misfolding, and Modification in Regulating the Innate Immune Sensor PKR. *Acc Chem Res*, **49**, 1242-1249.
483. Meagher, J.L., Takata, M., Gonçalves-Carneiro, D., Keane, S.C., Rebendenne, A., Ong, H., Orr, V.K., MacDonald, M.R., Stuckey, J.A., Bieniasz, P.D. *et al.* (2019) Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proceedings of the National Academy of Sciences*, **116**, 24303-24309.
484. Luo, X., Wang, X., Gao, Y., Zhu, J., Liu, S., Gao, G. and Gao, P. (2020) Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein. *Cell Rep*, **30**, 46-52.e44.

485. Fros, J.J., Dietrich, I., Alshaikhahmed, K., Passchier, T.C., Evans, D.J. and Simmonds, P. (2017) CpG and UpA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. *Elife*, **6**.
486. Bonam, S.R., Partidos, C.D., Halmuthur, S.K.M. and Muller, S. (2017) An Overview of Novel Adjuvants Designed for Improving Vaccine Efficacy. *Trends in Pharmacological Sciences*, **38**, 771-793.
487. Osterrieder, N. and Kunec, D. (2018) Attenuation of Viruses by Large-Scale Recoding of their Genomes: the Selection Is Always Biased. *Curr Clin Microbiol Rep*, **5**, 66-72.
488. Famulare, M., Chang, S., Iber, J., Zhao, K., Adeniji, J.A., Bukbuk, D., Baba, M., Behrend, M., Burns, C.C. and Oberste, M.S. (2015) Sabin Vaccine Reversion in the Field: a Comprehensive Analysis of Sabin-Like Poliovirus Isolates in Nigeria. *J Virol*, **90**, 317-331.
489. Hyde, S.C., Pringle, I.A., Abdullah, S., Lawton, A.E., Davies, L.A., Varathalingam, A., Nunez-Alonso, G., Green, A.M., Bazzani, R.P., Sumner-Jones, S.G. *et al.* (2008) CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. *Nat Biotechnol*, **26**, 549-551.
490. Hodges, B.L., Taylor, K.M., Joseph, M.F., Bourgeois, S.A. and Scheule, R.K. (2004) Long-term transgene expression from plasmid DNA gene therapy vectors is negatively affected by CpG dinucleotides. *Mol Ther*, **10**, 269-278.
491. Davies, L.A., Hyde, S.C., Nunez-Alonso, G., Bazzani, R.P., Harding-Smith, R., Pringle, I.A., Lawton, A.E., Abdullah, S., Roberts, T.C., McCormick, D. *et al.* (2012) The use of CpG-free plasmids to mediate persistent gene expression following repeated aerosol delivery of pDNA/PEI complexes. *Biomaterials*, **33**, 5618-5627.
492. Chabot, S., Bellard, E., Reynes, J.P., Tiraby, G., Teissie, J. and Golzio, M. (2019) Electrotransfer of CpG free plasmids enhances gene expression in skin. *Bioelectrochemistry*, **130**, 107343.
493. Low, P.T., Lai, M.I., Ngai, S.C. and Abdullah, S. (2014) Transgene expression from CpG-reduced lentiviral gene delivery vectors in vitro. *Gene*, **533**, 451-455.
494. Suzuki, T., Goda, T. and Kamiya, H. (2018) Durable Transgene Expression Driven by CpG-Free and -Containing Promoters in Plasmid DNA with CpG-Free Backbone. *Biol Pharm Bull*, **41**, 1489-1493.
495. Le Nouen, C., Brock, L.G., Luongo, C., McCarty, T., Yang, L., Mehedi, M., Wimmer, E., Mueller, S., Collins, P.L., Buchholz, U.J. *et al.* (2014) Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proc.Natl.Acad.Sci.U.S.A.*, Aug 25. pii: 201411290. [Epub ahead of print].
496. Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E. and Mueller, S. (2008) Virus attenuation by genome-scale changes in codon pair bias. *Science.*, **320**, 1784-1787.

497. Mueller, S., Papamichail, D., Coleman, J.R., Skiena, S. and Wimmer, E. (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol*, **80**, 9687-9696.
498. Burns, C.C., Shaw, J., Campagnoli, R., Jorba, J., Vincent, A., Quay, J. and Kew, O. (2006) Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J Virol*, **80**, 3259-3272.
499. Ramezanpour, B., Haan, I., Osterhaus, A. and Claassen, E. (2016) Vector-based genetically modified vaccines: Exploiting Jenner's legacy. *Vaccine*, **34**, 6436-6448.
500. André, S., Seed, B., Eberle, J., Schraut, W., Bültmann, A. and Haas, J. (1998) Increased immune response elicited by DNA vaccination with a synthetic gp120 sequence with optimized codon usage. *J Virol*, **72**, 1497-1503.
501. Megati, S., Garcia-Hand, D., Cappello, S., Roopchand, V., Masood, A., Xu, R., Luckay, A., Chong, S.-Y., Rosati, M., Sackitey, S. *et al.* (2008) Modifying the HIV-1 env gp160 gene to improve pDNA vaccine-elicited cell-mediated immune responses. *Vaccine*, **26**, 5083-5094.
502. Villanueva, E., Martí-Solano, M. and Fillat, C. (2016) Codon optimization of the adenoviral fiber negatively impacts structural protein expression and viral fitness. *Scientific Reports*, **6**, 27546.
503. Ramakrishnan, M.A. (2016) Determination of 50% endpoint titer using a simple formula. *World J Virol*, **5**, 85-86.
504. Hamers, A.A.J., Uleman, S., van Tiel, C.M., Kruijswijk, D., van Stalborch, A.-M., Huveneers, S., de Vries, C.J.M. and van 't Veer, C. (2014) Limited role of nuclear receptor Nur77 in Escherichia coli-induced peritonitis. *Infect Immun*, **82**, 253-264.
505. Ahn, K., Huh, J.-W., Park, S.-J., Kim, D.-S., Ha, H.-S., Kim, Y.-J., Lee, J.-R., Chang, K.-T. and Kim, H.-S. (2008) Selection of internal reference genes for SYBR green qRT-PCR studies of rhesus monkey (Macaca mulatta) tissues. *BMC Mol Biol*, **9**, 78-78.
506. Banerjee, S., Xie, N., Cui, H., Tan, Z., Yang, S., Icyuz, M., Abraham, E. and Liu, G. (2013) MicroRNA let-7c regulates macrophage polarization. *J Immunol*, **190**, 6542-6549.
507. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402-408.
508. Cotmore, S.F., D'Abramo, A.M., Jr., Carbonell, L.F., Bratton, J. and Tattersall, P. (1997) The NS2 polypeptide of parvovirus MVM is required for capsid assembly in murine cells. *Virology*, **231**, 267-280.
509. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, **9**, 676-682.

510. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
511. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357.

7.2 DNA sequences

MVM Inserts

>CLDR

ATGGCCAAAGATGATGCACATGAGCAAATTTGGACACCATGGAGCTTGGTTGATGCAAATGC

GTGGGGAGTTTGGCTTCAACCAAGTGACTGGCAATACATTTGCAACACAATGAGTCAACTTAA

TTTGGTTTCACTTGATCAAGAAATATTCAATGTAGTGTTGAAAACTGTGACAGAGCAAGACTT

AGGAGGACAAGCTATAAAAATATACAACAATGACCTCACAGCATGTATGATGGTAGCAGTAG

ATTCTAACAACATATTGCCATACACTCCAGCAGCAAACTCAATGGAAACTCTTGGATTCTACCC

ATGGAAACCTACTATTGCTTCCCCATACAGATACTATTTTTGCGTTGACAGAGACCTATCAGTG

ACATATGAAAATCAAGAAGGCACTGTTGAACATAATGTCATGGGAACACCAAAAGGAATGAA

CTCACAATTTTTTACCATTGAGAACACACAGCAAATCACCTTGCTGAGAACTGGGGACGAATT

TGCCACAGGCACCTACTACTTTGACACAAATTCAGTTAAACTTACACACACTTGGCAAACAAAC

CGTCAACTAGGACAGCCTCCACTTCTTTCAACATTTCCTGAAGCAGACACTGATGCAGGTACAC

TAACTGCACAAGGAAGCAGACATGGGACAACTCAAATGGGAGTTAACTGGGTTAGTGAAGCC

ATAAGAACCAGACCAGCTCAAGTAGGCTTTTGTCAGCCACACAATGACTTTGAAGCAAGCAG

GGCAGGCCCATTTGCAGCACCTAAAGTTCCAGCTGACATTACACAAGGAGTGGACAAAGAAG

CCAATGGTAGTGTTAGATACAGCTATGGCAAGCAGCATGGTGAAAATTGGGCTTCACATGGA

CCTGCACCTGAACGTTACACATGGGATGAAACAAGTTTTGGTTCTGGCAGAGACACAAAAGA

CGGCTTCATTCAATCTGCACCACTAGTTGTGCCACCACCTCTAAATGGCATCCTTACCAATGCA

AATCCAATTGGGACAAAAAATGACATTCACTTTTCAAATGTTTTTAACAGTTATGGACCTCTAA

CTGCTTTTTCACACCCAAGTCCAGTGTACCCTCAAGGTCAAATATGGGATAAAGAACTAGATCT

TGAACACAAACCCAGACTTCACATTACTGCACCCTTTGTTTGTAAAAACAATGCACCTGGACAA

ATGTTGGTTAGGCTAGGACCCAACCTTACAGACCAATATGATCCCAACGGAGCAACACTTT

CpG-H

>CpG-H1

ATGGCAAAAGACGACGCACACGAGCAAATTTGGACACCGTGGAGCTTGGTTGACGCGAACG

CGTGGGGTGTTTGGCTACAGCCAAGCGACTGGCAATACATTTGCAACACTATGAGCCAGCTTA

ACTTGGTATCGCTTGATCAAGAAATATTCAATGTTGTACTAAAAACCGTTACCGAGCAAGACTT

AGGTGGTCAAGCAATAAAAATATACAACAATGACCTTACCGCGTGCATGATGGTTGCCGTTGA

CTCAAACAACATTTTGCCGTACACGCCCGCCGCGAACTCGATGGAAACTCTTGGTTTTTACCCG

TGGAAACCGACAATAGCGTCGCCATACCGTTACTATTTTTGCGTTGACCGTGATCTTTCCGTTA

CGTACGAAAATCAAGAAGGAACCGTTGAACATAACGTAATGGGAACGCCAAAAGGAATGAA

TTCACAATTTTTTACGATAGAGAACACGCAACAAATAACGTTGCTACGAACCGGTGACGAATT

TGCAACTGGTACGTACTACTTCGACACGAATTCGGTTAAACTAACGCACACGTGGCAAACAAA

CCGTCAACTTGGACAACCGCCGCTTCTTTCAACATTTCCCGAAGCCGACACGGACGCGGGTAC

ACTTACCGCACAAGGTAGCCGACACGGAACGACGCAAATGGGAGTT

>CpG-H2

AACTGGGTAAGTGAAGCAATACGAACACGACCCGCACAAGTCGGATTTTGTCAACCACACAA

CGACTTTGAAGCAAGCCGTGCCGGACCGTTTGCAGCGCCGAAAGTTCCTGCCGATATTACGCA

AGGTGTTGACAAAGAAGCAAATGGAAGTGTTCGATACAGTTACGGAAAACAACATGGTGAA

AATTGGGCGTCGCACGGACCGGCGCCCGAACGATACACGTGGGATGAAACGAGCTTCGGTTC

GGGTCGTGACACGAAAGATGGTTTTATTCAATCCGCGCCGCTCGTTGTTCCACCGCCGCTAAA

CGGAATTCTTACGAACGCGAACCCGATTGGAACCAAAAACGACATTCATTTTTCTAACGTTTTT

AACAGCTATGGTCCACTAACCGCGTTTTCGCACCCGAGTCCCGTATACCCACAAGGACAAATA

TGGGACAAAGAACTTGATCTTGAACACAAACCGCGACTTCACATAACTGCGCCATTCGTTTGT

AAAAACAACGCGCCCGGACAAATGTTGGTTCGATTAGGACCAAACCTAACCGACCAATACGA

TCCGAACGGTGCAACGCTTT

>TpA-H

ATGGCTAAAGATGATGCACATGAGCAAATATGGACACCATGGAGCTTAGTAGATGCTAATGC

TTGGGGAGTATGGCTACAGCCCAGTGACTGGCAATACATATGCAACACCATGAGCCAGCTTA

ACTTAGTATCACTAGATCAAGAAATATTCAATGTAGTCCTTAAAACAGTTACAGAGCAAGACT

TAGGAGGACAAGCTATAAAGATATACAATAATGACCTTACAGCCTGCATGATGGTAGCAGTA

GACTCTAACAACATATTACCATATACACCTGCAGCTAACTCTATGGAAACCCTAGGTTTCTACC

CCTGGAAGCCTACTATAGCATCACCATACAGGTACTATTTTTGCGTGGACAGAGATCTATCAG

TTACCTACGAGAATCAAGAGGGCACAGTAGAACATAATGTTATGGGCACCCCTAAAGGTATG

AATTCTCAATTTTTTACTATAGAGAATACACAACAAATTACATTACTTAGAACAGGGGACGAAT

TTGCCACAGGTACATACTACTTTGATACAAATTCAGTTAAACTTACACATACGTGGCAAACCAA

CCGTCAGCTAGGACAGCCTCCACTACTATCTACCTTTCCTGAAGCTGATACTGATGCAGGTACA

CTTACTGCCCAAGGTAGCAGGCATGGTACTACACAAATGGGGGTTAACTGGGTTAGTGAAGC

TATTAGAACCAGACCAGCCCAAGTAGGATTTTGTCAACCACACAATGACTTTGAAGCTAGCAG

AGCTGGACCATTTGCTGCCCCCAAAGTACCAGCAGATATTACTCAAGGAGTAGACAAAGAAG

CCAATGGTAGTGTCAGGTACAGTTATGGCAAACAGCATGGGGAAAATTGGGCTTCCCATGGA

CCAGCACCAGAGCGCTACACCTGGGATGAAACTAGCTTTGGTTCAGGTAGAGACACTAAGGA

TGGTTTTATACAATCAGCCCCACTAGTGGTACCACCACCACTAAATGGCATACTCACTAATGCA

AACCCTATAGGTACTAAAAATGACATACATTTTTCCAATGTATTTAATAGCTATGGGCCCCTAA

CAGCATTTTCACACCCCAGTCCAGTCTACCCTCAGGGACAAATATGGGACAAAGAACTAGATC

TTGAACATAAACCTAGACTACACATAACTGCTCCATTTGTATGTAAAAATAATGCCCCAGGGC

AAATGTTAGTTAGATTAGGACCTAACCTAACTGACCAATATGATCCTAACGGAGCCACACTGT

>CodOpt

ATGGCCAAGGACGACGCCCACGAGCAGATCTGGACACCTTGGAGCCTGGTGGACGCCAACG

CTTGGGGAGTGTGGCTGCAGCCTAGCGACTGGCAGTACATCTGCAACACCATGAGCCAGCTG

AATCTGGTGTCCCTGGACCAGGAAATCTTCAACGTGGTGCTGAAAACCGTGACCGAGCAGGA

CCTGGGCGGCCAGGCCATCAAGATCTACAACAACGACCTGACCGCCTGCATGATGGTGGCCG

TGGACAGCAACAACATCCTGCCCTACACCCCTGCCGCCAACAGCATGGAAACCCTGGGCTTCT

ACCCCTGGAAGCCCACAATCGCCAGCCCCTACCGGTACTACTTCTGCGTGGACAGGGACCTGT

CCGTGACCTACGAGAACCAGGAAGGCACCGTGGAACACAACGTGATGGGCACCCCCAAGGG

CATGAACAGCCAGTTCTTCACCATCGAGAACACCCAGCAGATCACCCTGCTGAGAACCGGCGA

CGAGTTCGCCACCGGCACCTACTACTTCGACACCAACAGCGTGAAGCTGACCCACACCTGGCA

GACCAACAGACAGCTGGGCCAGCCTCCTCTGCTGAGCACATTCCCTGAGGCCGATACCGACG

CCGGCACACTGACAGCACAGGGCAGCAGACACGGCACCACCCAGATGGGCGTTAACTGGGT

GTCCGAGGCCATCAGAACCAGACCTGCCCAAGTGGGCTTCTGCCAGCCCCACAACGATTTCGA

GGCCAGCAGAGCCGGCCCTTTCGCCGCTCCTAAAGTGCCTGCCGACATCACCCAGGGCGTGG

ACAAAGAGGCTAACGGCAGCGTGCGGTACAGCTACGGAAAGCAGCACGGCGAGAACTGGGC

CTCTCACGGACCTGCCCCTGAGAGATACACCTGGGACGAGACAAGCTTCGGCAGCGGCAGAG

ACACCAAGGACGGCTTCATCCAGAGCGCCCCACTGGTGGTGCCTCCACCTCTGAACGGCATCC

TGACCAACGCCAACCCCATCGGCACCAAGAACGACATCCACTTCAGCAACGTGTTCAACTCCT

ACGGCCCCCTGACAGCCTTCAGCCACCCCAGCCCAGTGTATCCACAGGGACAGATCTGGGAC

AAAGAACTGGACCTGGAACATAAGCCCAGACTGCACATCACCGCCCCCTTCGTGTGCAAGAA

CAACGCCCCTGGCCAGATGCTCGTGCGGCTGGGACCTAACCTGACCGACCAGTACGATCCCA

ACGGCGCCACCCTGT

>GpC-H

ATGGCAAAAGATGATGCTCATGAGCAAATTTGGACACCATGGAGCTTGGTTGATGCTAATGCT

TGGGGAGTTTGGCTGCAGCCAAGTGACTGGCAATACATTTGCAACACAATGAGCCAGCTTAA

CTTGGTATCACTTGATCAAGAAATATTCAATGTAGTGCTCAAAACTGTTACAGAACAAGACTTA

GGAGGGCAAGCTATAAAAATATACAACAATGACCTTACAGCTTGCATGATGGTTGCAGTAGA

TTCAAACAACATTTTGCCATACACTCCTGCAGCAAACTCAATGGAAACACTTGGATTTTACCCA

TGGAAGCCAACAATAGCATCACCATACAGGTATTATTTTTGCGTTGACAGAGATCTTTCAGTCA

CATACGAAAATCAAGAAGGCACAGTTGAGCATAATGTCATGGGTACACCAAAAGGCATGAAT

TCTCAATTTTTTACAATTGAAAACACACAGCAAATCACATTGCTGAGAACAGGTGATGAATTTG

CAACAGGCACTTACTACTTTGACACAAATTCAGTTAAGCTTACACACACTTGGCAAACAAACC

GCCAGCTTGGCCAGCCTCCACTGCTTTCAACATTTCCTGAAGCTGACACTGATGCAGGAACAC

TTACTGCACAAGGCAGCAGGCATGGCACAACACAAATGGGTGTTAACTGGGTTAGCGAAGCA

ATTAGAACAAGGCCTGCTCAAGTAGGCTTTTGCCAGCCACACAATGATTTTGAAGCAAGCAGA

GCTGGCCCATTTGCTGCTCCAAAAGTGCCAGCAGATATTACACAAGGAGTAGACAAAGAAGC

AAATGGCAGTGTTAGATACAGCTATGGCAAGCAGCATGGTGAAAATTGGGCTTCACATGGCC

CAGCACCAGAGCGATACACATGGGATGAAACAAGCTTTGGGTCAGGCAGAGACACAAAAGA

TGGTTTTATTCAATCAGCACCACTAGTTGTGCCACCACCACTAAATGGAATTCTTACAAATGCA

AACCCTATTGGAACTAAAAATGACATTCATTTTTCAAATGTTTTTAACAGCTATGGCCCACTAA

CTGCATTTTCACACCCAAGCCCTGTATATCCTCAAGGGCAAATATGGGACAAAGAACTAGATC

TTGAACATAAGCCTAGGCTGCACATAACTGCGCCATTTGTTTGCAAAAACAATGCACCTGGCC

AAATGTTGGTTAGATTAGGCCCAAATCTAACTGACCAATATGATCCAAACGGAGCAACACTTT

>CpG-H42

ATGGCAAAAGACGACGCACACGAACAAATTTGGACACCGTGGAGTTTGGTTGACGCAAACGC

ATGGGGCGTTTGGCTACAACCAAGCGATTGGCAATACATTTGCAACACAATGAGCCAACTTAA

TTTGGTATCGCTAGATCAAGAAATATTCAACGTTGTTCTTAAAACTGTTACAGAACAAGATTTA

GGCGGTCAAGCGATAAAAATATACAACAATGATCTTACAGCATGCATGATGGTTGCAGTTGAT

TCAAACAACATTTTGCCTTACACACCAGCTGCAAATTCAATGGAAACACTTGGTTTTTATCCAT

GGAAACCAACGATAGCATCGCCGTATCGTTATTATTTTTGCGTTGATCGCGATCTTTCTGTAAC

GTACGAAAATCAAGAAGGAACAGTTGAACATAATGTAATGGGAACACCAAAAGGAATGAATT

CGCAATTTTTTACTATAGAAAACACGCAACAAATAACATTGCTACGAACGGGCGACGAATTCG

CGACAGGTACATATTATTTTGACACAAATTCCGTTAAACTAACTCACACATGGCAAACAAATCG

TCAACTTGGACAACCGCCACTTCTATCAACATTTCCCGAAGCGGACACAGACGCAGGTACACT

TACCGCGCAAGGTAGTCGACATGGAACAACACAAATGGGCGTTAACTGGGTTAGCGAAGCAA

TACGAACACGACCTGCACAAGTTGGATTTTGTCAACCACATAACGATTTCGAAGCAAGTCGCG

CTGGACCATTCGCGGCGCCGAAAGTTCCTGCCGATATTACACAAGGAGTTGACAAAGAAGCA

AACGGAAGCGTTCGATACAGTTACGGAAAACAACACGGAGAAAATTGGGCATCGCATGGAC

CGGCACCAGAACGATACACATGGGACGAAACGAGTTTCGGTTCCGGTCGCGACACAAAAGAC

GGTTTTATTCAATCTGCACCACTTGTTGTTCCTCCACCTCTAAATGGAATTCTTACAAATGCAAA

TCCAATCGGAACAAAAAACGACATTCATTTTTCAAACGTTTTTAACAGTTATGGTCCTCTAACA

GCGTTTTCTCATCCGAGTCCTGTATATCCGCAAGGACAAATATGGGACAAAGAACTTGATCTT

GAACACAAACCACGACTTCACATAACTGCTCCTTTCGTTTGTAAAAACAACGCACCTGGACAA

ATGTTGGTTCGATTAGGACCAAATCTAACTGATCAATACGATCCAAACGGAGCAACTCTTT

>CodOpt CpG-L

ATGGCCAAGGATGATGCCCACGAGCAGATCTGGACCCCCTGGAGCCTGGTGGATGCCAATGC

CTGGGGGGTGTGGCTGCAGCCCTCTGACTGGCAGTACATCTGCAATACCATGAGCCAGCTGA

ATCTGGTGTCCCTGGATCAGGAGATCTTCAATGTGGTGCTGAAGACTGTGACAGAGCAGGAC

CTGGGAGGCCAGGCCATCAAGATCTACAACAATGACCTGACAGCCTGTATGATGGTGGCCGT

GGACTCCAACAACATCCTGCCTTACACACCAGCTGCCAACAGCATGGAGACCCTGGGCTTCTA

CCCATGGAAGCCCACCATTGCCTCCCCATACAGGTACTACTTCTGTGTGGACAGAGACCTCTCA

GTGACCTATGAGAACCAGGAGGGCACAGTGGAGCACAATGTGATGGGAACCCCAAAGGGCA

TGAATAGCCAGTTTTTCACCATCGAGAATACCCAGCAGATCACCCTGCTGAGAACCGGGGATG

AGTTTGCCACAGGCACCTACTACTTTGACACCAACTCTGTGAAGCTGACCCACACATGGCAGA

CCAACAGACAGCTGGGCCAGCCCCCCCTGCTGTCCACCTTCCCTGAGGCTGACACAGATGCCG

GCACCCTGACAGCCCAGGGCTCCAGACATGGCACCACCCAGATGGGAGTTAACTGGGTGTCA

GAGGCCATCAGAACCAGGCCTGCCCAGGTGGGATTCTGCCAGCCCCACAACGACTTTGAGGC

CTCCAGGGCAGGCCCCTTTGCTGCCCCTAAAGTGCCAGCTGACATCACCCAGGGAGTGGACA

AGGAGGCCAATGGCTCTGTGAGATACTCATATGGCAAGCAGCATGGCGAGAACTGGGCCAG

CCATGGACCTGCCCCAGAGAGATACACTTGGGATGAGACCAGCTTTGGCTCTGGCAGAGACA

CCAAAGATGGCTTCATCCAGTCTGCCCCCCTGGTGGTGCCCCCTCCACTGAATGGAATCCTGA

CCAATGCCAATCCAATTGGCACCAAGAATGACATTCACTTCAGCAATGTGTTCAACAGCTATG

GCCCCCTGACAGCCTTTAGCCACCCCTCCCCTGTGTATCCACAGGGGCAGATCTGGGACAAGG

AGCTGGACCTGGAGCACAAGCCAAGACTGCACATCACAGCCCCATTTGTGTGCAAGAATAAT

GCCCCTGGCCAGATGCTGGTGAGACTGGGCCCTAACCTGACAGACCAGTATGACCCCAATGG

GGCCACCCTGT

pCAG_MVMp expression plasmid

Insert region highlighted in grey

>pCAG_MVM_R3

GATTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGC

CCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAAC

GACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC

CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT

CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCC

CAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA

CCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCA

ATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGG

GCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGT

GCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGC

GGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGTTGCCTTCGCC

CCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCC

ACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGAC

GGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTAAAGGGCTCCGGGAGGGCCCTTTGTGCG

GGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGC

CCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCGT

GTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGCTGCGAGGGG

AACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGGCG

GTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGG

TGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCA

GGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGC

GCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTA

TGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGGCGGAGCCGAAATCT

GGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCGAAGCGGTGCGGCGCCGGCAGGA

AGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTCCATCTCCAG

CCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCG

GCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTT

CCTACAGATCCTTAATTAATAATACGACTCACTATAGGGGGTCGACCCGCCACCAATCATTTAA

ATCACCATTCACGACACCGAAAAGTACGCCTCTCAGCCAGAACTATGCACTAACTCCACTTGCA

TCGGATCTCGAGGACCTGGCTTTAGAGCCTTGGAGCACACCAAATACTCCTGTTGCGGGCACT

GCAGAAACCCAGAACACTGGGGAAGCTGGTTCCAAAGCCTGCCAAGATGGTCAACTGAGCCC

AACTTGGTCAGAGATCGAGGAGGATTTGAGAGCGTGCTTCGGTGCGGAACCGTTGAAGAAA

GACTTCAGCGAGCCGCTGAACTTGGACTAAGGTACGATGGCGCCTCCAGCTAAAAGAGCTAA

AAGAGGTAAGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGTTTTAC

AGGCCTGAAATCACTTGGTTTTAGGTTGGGTGCCTCCTGGCTACAAGTACCTGGGACCAGGG

AACAGCCTTGACCAAGGAGAACCAACCAATCCATCTGACGCCGCTGCCAAAGAGCACGACGA

GGCCTATGATCAATACATCAAATCTGGAAAAAATCCTTACCTGTACTTCTCTGCTGCTGATCAA

CGCTTTATTGACCAAACCAAGGACGCCAAAGACTGGGGAGGCAAGGTTGGTCACTACTTTTTT

AGAACCAAGCGCGCTTTTGCACCTAAGCTTGCTACTGACTCTGAACCTGGAACTTCTGGTGTA

AGCAGAGCTGGTAAACGCACTAGACCACCTGCTTACATTTTTATTAACCAAGCCAGAGCTAAA

AAAAAACTTACTTCTTCTGCTGCACAGCAAAGCAGTCAAACCATGAGTGATGGCACCAGCCAA

CCTGACAGCGGAAACGCTGTCCACTCAGCTGCAAGAGTTGAACGAGCAGCTGACGGCCCTGG

AGGCTCTGGGGGTGGGGGCTCTGGCGGGGGTGGGGTTGGTGTTTCTACTGGGTCTTATGAT

AATCAAACGCATTATAGATTCTTGGGTGACGGCTGGGTAGAAATTACTGCACTAGCAACTAGA

CTAGTACATTTAAACATGCCTAAATCAGAAAACTATTGCAGAATCAGAGTTCACAATACAACA

GACACATCAGTCAAAGGCAACATGGCAAAAGATGATGCTCATGAGCAAATTTGGACACCATG

GAGCTTGGTGGATGCTAATGCTTGGGGAGTTTGGCTCCAGCCAAGTGACTGGCAATACATTT

GCAACACCATGAGCCAGCTTAACTTGGTATCACTTGATCAAGAAATATTCAATGTAGTGCTGA

AAACTGTTACAGAGCAAGACTTAGGAGGTCAAGCTATAAAAATATACAACAATGACCTTACA

GCTTGCATGATGGTTGCAGTAGACTCAAACAACATTTTGCCATACACACCTGCAGCAAACTCA

ATGGAAACACTTGGTTTCTACCCCTGGAAACCAACCATAGCATCACCATACAGGTACTATTTTT

GCGTTGACAGAGATCTTTCAGTGACCTACGAAAATCAAGAAGGCACAGTTGAACATAATGTG

ATGGGAACACCAAAAGGAATGAATTCTCAATTTTTTACCATTGAGAACACACAACAAATCACA

TTGCTCAGAACAGGGGACGAATTTGCCACAGGTACTTACTACTTTGACACAAATTCAGTTAAA

CTCACACACACGTGGCAAACCAACCGTCAACTTGGACAGCCTCCACTGCTGTCAACCTTTCCTG

AAGCTGACACTGATGCAGGTACACTTACTGCTCAAGGGAGCAGACATGGAACAACACAAATG

GGGGTTAACTGGGTGAGTGAAGCAATCAGAACCAGACCTGCTCAAGTAGGATTTTGTCAACC

ACACAATGACTTTGA

AGCCAGCAGAGCTGGACCATTTGCTGCCCCAAAAGTTCCAGCAGATATTACTCAAGGAGTAG

ACAAAGAAGCCAATGGCAGTGTTAGATACAGTTATGGCAAACAGCATGGTGAAAATTGGGCT

TCACATGGACCAGCACCAGAGCGCTACACATGGGATGAAACAAGCTTTGGTTCAGGTAGAGA

CACCAAAGATGGTTTTATTCAATCAGCACCACTAGTTGTTCCACCACCACTAAATGGCATTCTT

ACAAATGCAAACCCTATTGGGACTAAAAATGACATTCATTTTTCAAATGTTTTTAACAGCTATG

GTCCACTAACTGCATTTTCACACCCAAGTCCTGTATACCCTCAAGGACAAATATGGGACAAAG

AACTAGATCTTGAACACAAACCTAGACTTCACATAACTGCTCCATTTGTTTGTAAAAACAATGC

ACCTGGACAAATGTTGGTTAGATTAGGACCAAACCTAACTGACCAATATGATCCAAACGGAGC

CACACTTTCTAGAATTGTTACATACGGTACATTTTTCTGGAAAGGAAAACTAACCATGAGAGC

AAAACTTAGAGCTAACACCACTTGGAACCCAGTGTACCAAGTAAGTGCTGAAGACAATGGCA

ACTCATACATGAGTGTAACTAAATGGTTACCAACTGCTACTGGAAACATGCAGTCTGTGCCGC

TTATAACAAGACCTGTTGCTAGAAATACTTACTAAGAGCTCGCTGATCAGCCTCGACTGTGCCT

TCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCA

CTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTC

TATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG

GCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCG

AGATCCACTAGTTCTAGCCTCGAGGCTAGAGCGGCCGCCACTGGCCGTCGTTTTACAACGTCG

TGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGC

TGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGG

CGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGC

GTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGC

CACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGT

GCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCG

CCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT

TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCC

GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAA

AATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGT

TTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCA

ATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT

GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAA

GATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGA

GAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC

GGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAA

TGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG

AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGA

TCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT

GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCC

TGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCG

GCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCC

TTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA

TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT

CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA

TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAAT

TTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTT

TTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTT

CTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG

GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAAT

ACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT

ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCG

GGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC

GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGC

TATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAG

GGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGT

CCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGG

AGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG

CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGA

GCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGG

AAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC

ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC

ACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGA

GCGGATAACAATTTCACACAGGAAACAGCTATGACCATGAGGCGCGCCG

In vitro translation plasmid

MVM region underlined, insert region highlighted in grey

>pMKRQ_MVMp_synth

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTT

TTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGG

TTGAGTGGCCGCTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGC

GTTTCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCG

ATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGC

GCGACGTAATACGACTCACTATAGGGCGAATTGAAGGAAGGCCGTCAAGGCCGCATCCTAAG

CTTGCTACTGACTCTGAACCTGGAACTTCTGGTGTAAGCAGAGCTGGTAAACGCACTAGACCA

CCTGCTTACATTTTTATTAACCAAGCCAGAGCTAAAAAAAAAACTTACTTCTTCTGCTGCACAGC

AAAGCAGTCAAACCATGAGTGATGGCACCAGCCAACCTGACAGCGGAAACGCTGTCCACTCA

GCTGCAAGAGTTGAACGAGCAGCTGACGGCCCTGGAGGCTCTGGGGGTGGGGGCTCTGGCG

GGGGTGGGGTTGGTGTTTCTACTGGGTCTTATGATAATCAAACGCATTATAGATTCTTGGGTG

ACGGCTGGGTAGAAATTACTGCACTAGCAACTAGACTAGTACATTTAAACATGCCTAAATCAG

AAAACTATTGCAGAATCAGAGTTCACAATACAACAGACACATCAGTCAAAGGCAACATGGCC

AAAGATGATGCACATGAGCAAATTTGGACACCATGGAGCTTGGTTGATGCAAATGCGTGGGG

AGTTTGGCTTCAACCAAGTGACTGGCAATACATTTGCAACACAATGAGTCAACTTAATTTGGTT

TCACTTGATCAAGAAATATTCAATGTAGTGTTGAAAACTGTGACAGAGCAAGACTTAGGAGG

ACAAGCTATAAAAATATACAACAATGACCTCACAGCATGTATGATGGTAGCAGTAGATTCTAA

CAACATATTGCCATACACTCCAGCAGCAAACTCAATGGAAACTCTTGGATTCTACCCATGGAA

ACCTACTATTGCTTCCCCATACAGATACTATTTTTGCGTTGACAGAGACCTATCAGTGACATAT

GAAAATCAAGAAGGCACTGTTGAACATAATGTCATGGGAACACCAAAAGGAATGAACTCACA

ATTTTTTACCATTGAGAACACACAGCAAATCACCTTGCTGAGAACTGGGGACGAATTTGCCAC

AGGCACCTACTACTTTGACACAAATTCAGTTAAACTTACACACACTTGGCAAACAAACCGTCAA

CTAGGACAGCCTCCACTTCTTTCAACATTTCCTGAAGCAGACACTGATGCAGGTACACTAACTG

CACAAGGAAGCAGACATGGGACAACTCAAATGGGAGTTAACTGGGTTAGTGAAGCCATAAG

AACCAGACCAGCTCAAGTAGGCTTTTGTCAGCCACACAATGACTTTGAAGCAAGCAGGGCAG

GCCCATTTGCAGCACCTAAAGTTCCAGCTGACATTACACAAGGAGTGGACAAAGAAGCCAAT

GGTAGTGTTAGATACAGCTATGGCAAGCAGCATGGTGAAAATTGGGCTTCACATGGACCTGC

ACCTGAACGTTACACATGGGATGAAACAAGTTTTGGTTCTGGCAGAGACACAAAAGACGGCT

TCATTCAATCTGCACCACTAGTTGTGCCACCACCTCTAAATGGCATCCTTACCAATGCAAATCC

AATTGGGACAAAAAATGACATTCACTTTTCAAATGTTTTTAACAGTTATGGACCTCTAACTGCT

TTTTCACACCCAAGTCCAGTGTACCCTCAAGGTCAAATATGGGATAAAGAACTAGATCTTGAA

CACAAACCCAGACTTCACATTACTGCACCCTTTGTTTGTAAAAACAATGCACCTGGACAAATGT

TGGTTAGGCTAGGACCCAACCTTACAGACCAATATGATCCCAACGGAGCAACACTTTCTAGAC

TGGGCCTCATGGGCCTTCCTTTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGC

ATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCTCACTGACTC

GCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGG

CCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGC

ATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG

GCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC

TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG

TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCG

CTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTG

GCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT

GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAA

GCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTA

GCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT

CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTG

GTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAAT

CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTAGAAAAATTCATCCAGCAGAC

GATAAAACGCAATACGCTGGCTATCCGGTGCCGCAATGCCATACAGCACCAGAAAACGATCC

GCCCATTCGCCGCCCAGTTCTTCCGCAATATCACGGGTGGCCAGCGCAATATCCTGATAACGA

TCCGCCACGCCCAGACGGCCGCAATCAATAAAGCCGCTAAAACGGCCATTTTCCACCATAATG

TTCGGCAGGCACGCATCACCATGGGTCACCACCAGATCTTCGCCATCCGGCATGCTCGCTTTC

AGACGCGCAAACAGCTCTGCCGGTGCCAGGCCCTGATGTTCTTCATCCAGATCATCCTGATCC

ACCAGGCCCGCTTCCATACGGGTACGCGCACGTTCAATACGATGTTTCGCCTGATGATCAAAC

GGACAGGTCGCCGGGTCCAGGGTATGCAGACGACGCATGGCATCCGCCATAATGCTCACTTT

TTCTGCCGGCGCCAGATGGCTAGACAGCAGATCCTGACCCGGCACTTCGCCCAGCAGCAGCC

AATCACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACACGGAACACCGGTGGTGGCC

AGCCAGCTCAGACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCTCAGATCGGTTTTC

ACAAACAGCACCGGACGACCCTGCGCGCTCAGACGAAACACCGCCGCATCAGAGCAGCCAAT

GGTCTGCTGCGCCCAATCATAGCCAAACAGACGTTCCACCCACGCTGCCGGGCTACCCGCATG

CAGGCCATCCTGTTCAATCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT

GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA

CATTTCCCCGAAAAGTGCCAC

MPyV Inserts

CDLR

>CDLR_Region_T

ATGCATAAAGTGCAGCAAAGAGGAGACTCGACTCCAAATACATTGGAAAAATCATAGAAAAC

ATGCAGAGAATGCAGACCTATTCCTAAACTGTAAGGCACAAAAGACTATATGTCAACAGGCC

GCTGACGGAGTGTTGGCCTCAAGGAGACTAAAATTAGTAGAATGCACACGTAGCCAGCTTTT

GAAAGAGAGGTTGCAACAGTCCCTACTCAGGCTAAAGGAACTTGGATCCTCAGATGCCCTACT

CTACCTAGCAGGTGTCGCTTGGTACCAGTGCCTTTTAGAGGATTTTCCTCAGACCCTATTTAAA

ATGCTTAAGCTGCTAACAGAAAATGTGCCAAAGCGTCGTAATATTCTCTTTAGAGGACCAGTT

AATTCTGGAAAAACAGGTCTGGCAGCCGCTCTTATTAGCCTGTTAGGTGGCAAATCTCTCAAC

ATAAATTGTCCTGCAGATAAACTTGCCTTTGAGCTTGGAGTAGCCCAGGATCAATTTGTGGTC

TGCTTTGAGGATGTGAAAGGCCAAATAGCTTTGAACAAACAGCTACAGCCAGGGATGGGAGT

GGCCAATCTTGACAACCTCAGGGACTACCTGGATGGGAGTGTAAAAGTCAATCTAGAAAAAA

AGCACAGCAACAAAAGGTCACAACTGTTTCCCCCATGTGTGTGTACAATGAATGAGTACCTTC

TGCCTCAAACAGTTTGGGCTCGTTTTCACATGGTATTAGATTTCACTTGCAAACCCCATTTGGC

GCAGTCCCTTGAAAAGTGTGAGTTTCTGCAGAGGGAAAGAATTATCCAAAGTGGAGATACCC

TGGCACTCTTACTCATATGGAATTTCACCTCAGATGTCTTTGATCCAGATATTCAGGGCCTTGT

GAAGGAGGTCCGTGACCAGTTTGCTAGT

>CDLR_Region_V

GAAAGCCTCACAGAAGGAGGGCAATATTATGGTTGGAGTAGAGGGATCAATTTAGCCACCTC

AGATACTGAGGACTCCCCCGAAAATAACACCCTCCCCACATGGAGCATGGCAAAGCTGCAGC

TTCCCATGCTGAATGAGGACCTCACCTGTGACACCCTCCAAATGTGGGAGGCTGTGTCAGTGA

AAACAGAGGTAGTGGGCTCAGGATCCCTCTTGGATGTACATGGCTTTAACAAACCAACAGAT

ACAGTGAACACAAAAGGAATTTCCACTCCAGTGGAGGGAAGCCAATATCATGTTTTTGCAGTT

GGCGGAGAGCCCCTGGATCTGCAGGGACTGGTCACAGATGCCAGGACAAAATACAAGGAGG

AAGGAGTTGTCACCATAAAAACAATCACAAAGAAAGACATGGTAAACAAGGACCAAGTGCTG

AATCCAATTAGTAAGGCAAAACTAGATAAGGACGGCATGTATCCAGTTGAAATCTGGCATCCA

GATCCTGCCAAGAATGAAAACACAAGGTATTTTGGCAACTACACTGGAGGCACAACAACTCC

ACCAGTCCTACAGTTCACAAACACCCTGACCACAGTGCTCCTAGATGAAAATGGGGTTGGCCC

TCTGTGCAAAGGGGAGGGGCTCTACCTATCCTGTGTAGATATAATGGGTTGGAGAGTAACAA

GAAACTATGATGTTCATCATTGGAGAGGGCTTCCCAGGTACTTTAAAATCACACTTAGAAAAA

GATGGGTCAAAAACCCCTATCCTATGGCATCGCTCATAAGCTCCCTCTTCAACAACATGCTCCC

CCAAGTCCAAGGCCAACCAATGGAAGGGGAGAATACCCAGGTGGAGGAGGTCAGAGTGTAT

GATGGAACTGAACCCGTCCCGGGGGACCC

7.3 Use of a small DNA virus model to investigate mechanisms of CpG dinucleotide-induced attenuation of virus replication - Loew et al., 2020, J Gen Virol

JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE
Loew et al., Journal of General Virology
DOI 10.1099/jgv.0.001477

MICROBIOLOGY SOCIETY
OPEN MICROBIOLOGY

Use of a small DNA virus model to investigate mechanisms of CpG dinucleotide-induced attenuation of virus replication

Lisa Loew†, Niluka Goonawardane, Jeremy Ratcliff, Dung Nguyen and Peter Simmonds*

Abstract

Suppression of the CpG dinucleotide is widespread in RNA viruses infecting vertebrates and plants, and in the genomes of retroviruses and small mammalian DNA viruses. The functional basis for CpG suppression in the latter was investigated through the construction of mutants of the parvovirus, minute virus of mice (MVM) with increased CpG or TpA dinucleotides in the VP gene. CpG-high mutants displayed extraordinary attenuation in A9 cells compared to wild-type MVM (>six logs), while TpA elevation showed no replication effect. Attenuation was independent of Toll-like receptor 9 and STING-mediated DNA recognition pathways and unrelated to effects on translation efficiency. While translation from codon-optimized VP RNA was enhanced in a cell-free assay, MVM containing this sequence was highly attenuated. Further mutational analysis indicated that this arose through its increased numbers of CpG dinucleotides (7→70) and separately from its increased G+C content (42.3→57.4 %), which independently attenuated replication. CpG-high viruses showed impaired NS mRNA expression by qPCR and reduced NS and particularly VP protein expression detected by immunofluorescence and replication in A549 cells, effects reversed in zinc antiviral protein (ZAP) knockout cells, even though nuclear relocalization of VP remained defective. The demonstrated functional basis for CpG suppression in MVM and potentially other small DNA viruses and the observed intolerance of CpGs in coding sequences, even after codon optimization, has implications for the use of small DNA virus vectors in gene therapy and immunization.

INTRODUCTION

Early investigations into the DNA composition of viruses and their hosts revealed differences in both the relative abundance of each nucleotide and certain dinucleotide combinations [1–3]. This was particularly obvious for CpG and TpA dinucleotides, which were markedly reduced compared to their expected occurrences and their respective counterparts, GpC and ApT. The avoidance of CpG in vertebrate genome DNA was primarily attributed to DNA methylation processes, as methylated CpGs are more prone to missense mutations than their non-methylated counterparts [4, 5]. Supporting this, CpG suppression in genomic DNA sequences is typically most apparent in organisms that highly methylate their genomes (e.g. mammals, other vertebrates), occurs at a reduced frequency in plants with their partial methylation, and is absent from organisms with very low methylation levels (e.g. insects and most other arthropods) [4, 6–9].

It has long been recognized that the dinucleotide bias observed in vertebrate genomes is reflected in RNA viruses infecting them. While almost all single-stranded RNA (ssRNA) viruses, retroviruses and small DNA viruses (<20 kb genome length) show CpG suppression, double-stranded RNA (dsRNA) viruses and large DNA viruses mostly do not [10–12]. This is in contrast to the suppression of UpAs that is present to some extent in the majority of eukaryotes and eukaryotic viruses. UpA composition plays a major in determining rates of turnover of mRNAs through the action of sequence-specific RNAses; there are additionally antiviral pathways, such as

Received 28 May 2020; Accepted 10 July 2020; Published 12 August 2020
Author affiliations: [1]Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford OX1 3SY, UK.
***Correspondence:** Peter Simmonds, peter.simmonds@ndm.ox.ac.uk
Keywords: parvovirus; CpG dinucleotide; TpA dinucleotide; zinc antiviral protein; codon optimization; minute virus of mice.
Abbreviations: AAV, adeno-associated virus; BHK, baby hamster kidney; CodOpt, codon-optimized construct; CodOpt CpG-L, CpGlow codonoptimized mutant; CPE, cytopathic effect; CpG-H, CpG-high; DMEM, Dulbecco's modified Eagle's medium; dsRNA, doublestranded RNA; FCS, foetal calf serum; IFNAR, interferon α receptor; k/o, knockout; MEF, mouse embryo fibroblasts; MVM, minute virus of mice; OAS, oligoanylate synthetase; O/E, observed over expected; PBS, phosphate-buffered saline; PKR, protein kinase R; PPV, porcine parvovirus; PRR, pattern recognition receptors; ssRNA, singlestranded RNA; STING, stimulator of interferon genes; TLR9, toll-like receptor 9; TpA-H, TpA-high; WT, wild-type; ZAP, zinc antiviral protein.
†Present address: Clinical Biomanufacturing Facility, University of Oxford, Old Road, Headington, Oxford OX3 7BN, UK.
Two supplementary tables and seven supplementary figures are available with the online version of this article.
001477 © 2020 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

RNAseL, which preferentially degrade UpA and UpU rich RNAs in infected cells [13, 14].

While DNA methylation could explain the CpG suppression in DNA viruses and retroviruses, it fails to account for its widespread avoidance in RNA viruses. Indeed, reduced frequencies of CpGs appears to be an adaptive response by the virus to avoid restriction pathways in the cell that recognize CpG-enriched sequences. Indeed, creation of poliovirus or echovirus 7 (E7) mutants with either increased CpG or UpA frequencies in their genomes showed profound attenuation of replication [15–17]; CpG modification reduced the pathogenicity of Influenza A virus and the subsequent immune response towards the virus in a mouse model [18]. Recently, zinc finger antiviral protein (ZAP) was shown to specifically bind to CpG in HIV RNA sequences and attenuate the replication of CpG-high HIV-1 mutants [19]. However, ZAP does not act directly to degrade RNA after binding, and its antiviral effect has recently been shown to depend on the presence of the cytoplasmic protein, KHNYN [20]. We have further shown that ZAP, along with oligoanylate synthetase (OAS) 3/RNAseL, mediates the attenuation of both CpG- and UpA-high mutants of E7 [21].

It is currently unknown whether ZAP is also the driving force behind the CpG suppression in small DNA viruses or whether other proteins and mechanisms are involved. To investigate the effects of CpG and TpA compositional modification in small DNA viruses, we developed an *in vitro* cell culture model based on minute virus of mice (MVM), where suppression of both dinucleotides is particularly evident. MVM is an autonomous parvovirus with a single-stranded, linear DNA genome that is 5.1 kb in length and packaged into an icosahedral, non-enveloped capsid with a diameter of 25 nm. The genome encodes two non-structural proteins, NS1 and NS2, driven by the early p4 promoter, and two capsid proteins, VP1 and VP2, whose expression is driven by the later promoter p38 (Fig. 1a) [22]. Capsid proteins are produced and transported into the nucleus, where they assemble into capsids and are packaged with genomic DNA [22]. The central region of the VP structural gene was compositionally modified and used to create mutants from full-length DNAs using an established reverse genetics system [23]. Viruses were assayed for their replication kinetics compared to wild-type (WT) virus; a range of pathway knockout (k/o) cell lines were used to investigate the mechanistic basis for the attenuation of viruses enriched for CpG.

METHODS

Cell culture, cell lines and reagents

A9 cells [24] and NB324K cells [25] were a kind gift from Peter Tattersall, Yale. Cells were grown in complete Dulbecco's modified Eagle's medium (DMEM) [F-12, supplemented with 5% foetal calf serum (FCS) and penicillin/streptomycin)] and incubated at 37 °C with 5% CO_2. A549 cells, A549 ZAP k/o cells [21], immortalised mouse embryo fibroblast (MEF) cells (a kind gift from Jan Rehwinkel, Oxford) and baby hamster kidney (BHK) cells were grown as above but were supplemented with 10% FCS.

In silico design of codon-optimized, GpC-, CpG- and TpA-modified viruses

SSE version 1.4 [26] (http://www.virus-evolution.org/Downloads/Software/) was used to determine suitable areas for mutation. Synonymous coding variability was calculated using the program SequenceDist within the package. A region with no suppression of synonymous sequences variability within the capsid ORF was chosen, as this was less likely to contain cryptic replication elements or undocumented alternative reading frames. Sequence Mutate in SSE was then used to produce the CDLR sequence and those with altered dinucleotide frequencies based on that region. For the codon-optimized construct (CodOpt), the GeneArt codon optimization tool was used, while Codon Optimization OnLine was used to for the CpG-low codon-optimized mutant (CodOpt CpG-L) [27]. Optimization was performed using codon usage tables for *Mus musculus*. All constructs were synthesized by GeneArt into a commercial plasmid backbone.

Cloning and plasmid propagation

To generate the mutants, the WT infectious clone pMM984 (a kind gift from David Pintel, Missouri), as well as the insert-containing plasmids, were cut with *Bpu*10I, *Hpa*I and *Xba*I as needed to generate sticky end fragments. These were gel-purified (QIAquick Gel Extraction Kit, Qiagen, UK), ligated with T4 ligase (New England Biosciences UK, UK) and transformed into SURE 2 Supercompetent cells (Agilent, USA) according to the manufacturers' protocols. Bacterial colonies were grown in Lysogeny broth and plasmid preparations were purified using the HiSpeed Plasmid Midi Kit (Qiagen, Germany) and subsequently used to produce virus stocks. For quantification purposes, the MVM genome was excised out of the pMM984 plasmid with *Bam*HI and gel-purified.

For the R3 subclones, the R3 RNA region of MVM was cloned into a pCAG-loxPSTOPloxP-ZsGreen vector (addgene #51269, a kind gift from Lynn Dustin, Oxford) using *SwaI* and *SacI* restriction sites. The R3 region was amplified and restriction sites were added using two rounds of PCR with a Phusion High Fidelity DNA Polymerase (New England Biosciences, UK) according to the manufacturer's guidelines (primers: FW1 5′-ATAATTTAAATCACCATTCACGAC ACCGAAAAGTAC-3′, RV1 5′-ATAGAGCTCTTAGTAA GTATTTCTAGCAACAGG-3′; FW2 5′-GATCGTAACGAT AATTTAAATCACC-3′, RV2 5′-GATCGTAACGATAATT TAAATCACC-3′). PCR products and vector were processed as above, but grown in One Shot TOP10 Chemically Competent *Escherichia coli* (Thermo Fisher Scientific, UK).

Transfection and virus recovery

To produce virus stocks, NB324K cells were transfected using Lipofectamine 2000 (Thermo Fisher Scientific, UK) according to the manufacturer's recommendations.

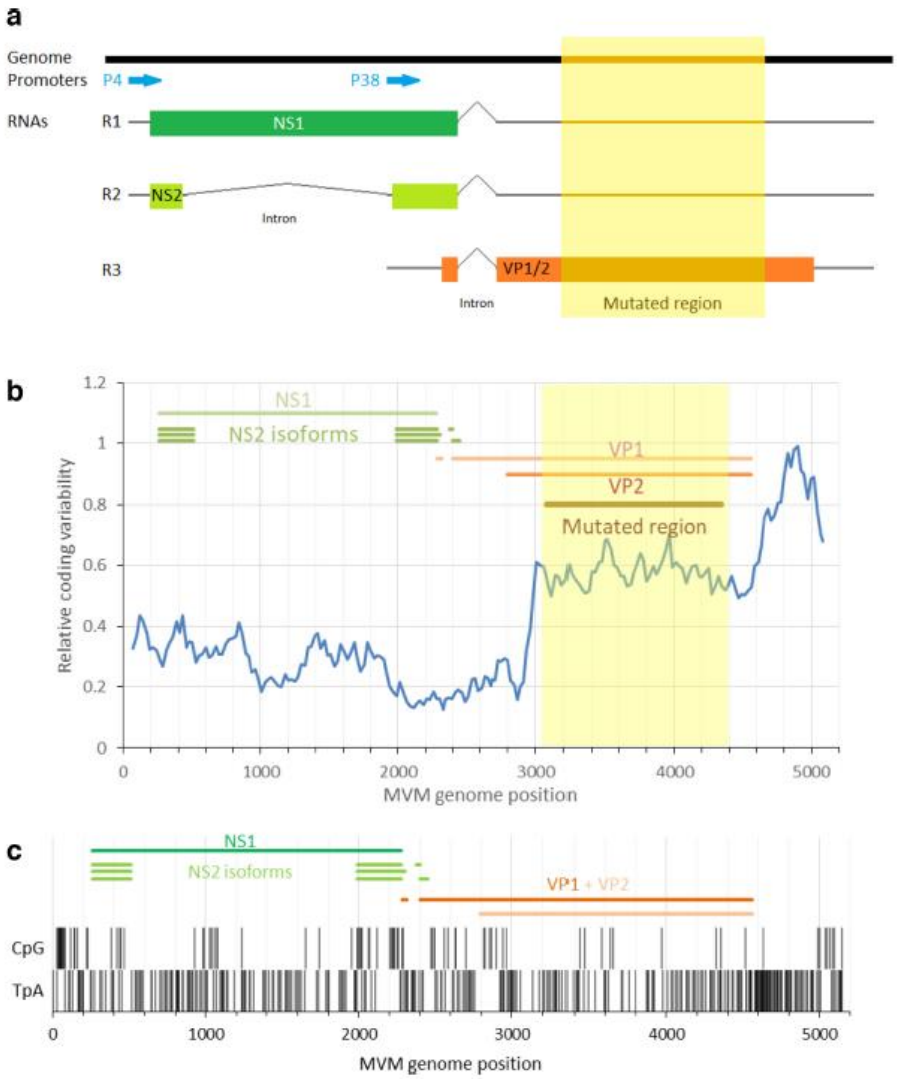


Fig. 1. MVM genomic structure and the capsid gene region selected for mutagenesis. (a) Diagram of the MVM genomic structure. Promoters (blue), transcripts (r1-r3) with ORFs (green and orange coloured boxes), mutated region highlighted in yellow. Based on, Fig. 57.10 of Berns and Parrish [65]. (b) SSE was used to calculate the synonymous coding variability observed between MVM and related parvoviruses. The orange region indicates the region chosen for mutations. (c) Distribution of CpGs and TpAs in the MVM genome.

One million NB324K cells were transfected with 0.5–2 µg DNA, depending on mutant recovery efficacy. The next day, the cells were split using trypsin and transferred into a T175 tissue culture flask. A non-transfected control was grown in parallel. After 3–5 days, at the onset of cytopathic effect (CPE), cells were harvested by scraping the cells into the culture medium. Cells and medium were spun down and separated. The cell pellet was resuspended in 3 ml vTE (50 mM Tris-hydrochloride, 0.5 mM EDTA, pH 8.7 [28]) and freeze-thawed three times. The cell lysate was spun down (1500 r.p.m., 20 min, 4 °C) and the cleared lysate was sterile-filtered, aliquoted, frozen at −80 °C and titrated.

Virus titration

$TCID_{50}$s were used to quantify virus concentrations. Ninety microlitres of NB324K cell suspension containing 3000 cells was mixed with 10 µl viral suspension (serial 10-fold dilutions in complete DMEM) in a 96-well plate. Ten days post-infection, the $TCID_{50}$ score was determined using the Reed–Muench method [29].

Virus infections

Cells were seeded overnight before being infected. Plate sizes were chosen in accordance with individual experiment needs. Cell numbers were chosen according to experiment length, seeding at roughly 10% confluence for experiments with total incubation periods of 48 h or longer, and seeding at roughly 25% confluence for experiments with a total incubation period of 24 h or less. Cells were infected with either equal $TCID_{50}$ units or equal copy numbers. Viral stocks were diluted in Opti-MEM. Cell supernatant was replaced by the viral suspension and incubated for 1 h at 37 °C, while shaking the plate occasionally. Afterwards, the cells were washed 3× with phosphate-buffered saline (PBS) and fresh DMEM with 5 or 10% FCS added as appropriate.

Serial passage

A9 cells were cultured in 25 cm^2 tissue culture flasks and were infected with virus stocks [multiplicity of infection (m.o.i.) of 0.5]. For the serial passage, viral supernatant was added to fresh A9 cells twice a week in a dilution that would allow complete CPE within 3–4 days (WT: 1 : 50; TpA-H: 1 : 25; CpG-H half-mutants: 1 : 2).

In vitro transcription and translation

Plasmids were linearized using *Apa*LI for 4 h at 37 °C and purified (QIAquickPCR Purification Kit, Qiagen, UK). One microgram of of DNA was used to transcribe capped and poly-adenylated messenger RNAs using the mMESSAGE mMACHINE T7 Transcription Kit (Thermo Fisher Scientific, UK) according to the manufacturer's description and purified using the RNA Clean and Concentrator-25 Kit (Zymo Research, USA). RNA integrity was determined via the Bioanalyser, using the RNA 6000 Pico Kit (Agilent, USA). Two micrograms of RNA were used in the Rabbit Reticulocyte Lysate System, Nuclease Treated (Promega, UK) according to the manufacturer's instructions. Samples were then immunoblotted.

Transcription experiments

Cells were seeded and used for transfections the next day. Cells were transfected with 500 ng DNA per 100 000 cells using Lipofectamine 2000 (Thermo Fisher Scientific, UK) according to the manufacturer's guidelines. Four hours post-transfection, 5 µg ml^{-1} Aphidicolin (Santa Cruz Biotechnology, USA) was added.

TLR9 stimulation and inhibition

For TLR9 stimulation, cells were stimulated with 1 µg ml^{-1} purified *E. coli* DNA (InvivoGen, UK) for 18 h, with or without 2 h pre-treatment with 5 µM ODN2088 (InvivoGen, UK). For TLR9 inhibition, cells were pretreated with 5 µM ODN2088 for 2 h, before being infected with 1000 virus genomes/cell as described above.

DNA and RNA extractions

Viral stocks were pretreated with DNAseI (Roche, Switzerland) prior to DNA extraction. Then 0.1 mg ml^{-1} DNase and 5 mM $MgCl_2$ was added to the samples, which were incubated at 25 °C for 25 min. Control (no-virus) lysate was spiked with 10^9 *Bam*HI-excised MVM copies to evaluate DNAse digestion and DNA recovery rate. DNA was extracted and purified with either the DNeasy Blood and Tissue Kit (Qiagen, UK) or the Quick-DNA Viral Kit (Zymo Research, USA) according to the manufacturers' guidelines. RNA was extracted and purified using the RNAeasy Mini Kit (Qiagen, UK) according to the manufacturer's guidelines. DNA and RNA concentration was then measured using the Qubit (Thermo Fisher Scientific, UK).

Quantitative real-time PCR

For the quantitative real-time PCR, the QuantiFast SYBR Green RT-PCR Kit (RNA) or QuantiFast SYBR Green PCR Kit (DNA) (Qiagen, UK) were used according to the manufacturer's instructions, loading 80 ng RNA or DNA per well. For viral DNA quantification, primers in the NS region were used (FW: 5′-AGTTTGCCATGCTATTTGC-3′ and RV: 5′-AATCTGCTTGCTGCCTTT-3′, from Zhan *et al.* [30]) with a standard curve derived from serially diluted *Bam*HI-excised MVMp. For viral RNA quantification, primers targeting MVM R1 (FW: 5′- CCAACTCCTATA AATTTACTAG –3′, RV: 5′-ACCAAGTGATTTCAGG CCTTAG-3′, from Li *et al.* [31]) or R3 (FW: CACCATTC ACGACACCGAAAAG, same RV as R1) were used. For IL-1β, FW: 5′-GCAACTGTTCCTGAACTCAACT-3′and RV: 5′-ATCTTTTGGGGTCCGTCAACT-3′ were used [32]. Samples were normalized to the species-appropriate GAPDH (human FW: 5′-GAAATCCCATCACCATCTTCCAGG-3′, RV: 5′-GAGCCCCAGCCTTCTCCATG-3′, from Ahn *et al.* [33]; mouse FW: 5′- CGACTTCAACAGCAACTCCC ACTCTTCC –3′, FW: 5′- TGGGTGGTCCAGGGTTTCTT ACTCCTT –3′, from Banerjee *et al.* [34]) using the ΔΔCT method [35].

Miseq sequencing

Purified DNA underwent second-strand synthesis using Sequenase Version 2.0 (Thermo Fisher, UK) and random primer 9 (New England BioLabs, UK). Libraries were prepared from 1 ng of the resulting double-stranded DNA using the Nextera XT kit (Illumina, UK) and sequenced on the MiSeq platform with reagent kit v2 (500 cycles). After adapter removal and quality trimming using Trimmomatic [36], reads were aligned against the reference sequence

(GenBank no. J02275) and the derived mutant sequences using Bowtie 2 [37].

Immunofluorescence (IF)

Cells were seeded into cellview cell culture dishes (VWR, USA) or Falcon 96-well imaging microplates (Corning, USA). The next day, cells were infected with 1000 genome copies per cell. After 18 h incubation, the cells were submerged into 4% paraformaldehyde for 10 min, neutralized with 0.1 M glycine for 5 min and permeabilized in 0.1% Triton in PBS for 20 min. The cells were then blocked with 2% FCS in PBS for 20 min, washed 3× with PBS and stained for 1 h with the primary antibodies (Mab3 [38]: 1 : 100, monoclonal mouse antibody; NS1/2 [39]: 1 : 400, polyclonal rabbit antibody, both kind gifts from Peter Tattersall). Subsequently, samples were washed 3× with PBS and stained with the secondary antibodies (1 : 2000 anti-rabbit Alexa Fluor 488 and either 1 : 2000 anti-mouse Alexa Fluor 594 or 647; all from Thermo Fisher Scientific, UK), stained with Hoechst 33342 (1 : 2000, Thermo Fisher Scientific, UK) for 10 min, and washed 3× in PBS. Slides were analysed using a Leica DMIRE2 microscope and Q capture pro 7 software or an Olympus IX73 fluorescence microscope and the accompanying software. Post-acquisition analysis was conducted using Imaris (Bitplane) or Fiji (v1.49) software [40].

Immunoblotting

Cells were infected with 1000 viral copies/cell for 18 h, washed with PBS, scraped off the cell culture dish and resuspended in Laemmli buffer. *In vitro* translation products were diluted with 4× Laemmli buffer (Bio-rad, UK). Samples were incubated for 5 min at 95 °C and spun down. Lysate was loaded onto a 10-well (10 µl) or 15-well (5 µl) 4–15% gradient mini protean TGX SDS-PAGE gel (Bio-rad, UK) and run for 4 h at 40 V. Gels were then transferred using the Trans-blot Turbo transfer system (Bio-rad, UK) according to the manufacturer's guidelines. Membranes were blocked in 5% skimmed milk powder in 0.1% PBS-T, incubated ON at 4 °C with primary antibodies (Tatt3 [38]: 1 : 10000; NS1/2 [39]: 1 : 10000, both polyclonal rabbit antibodies and kind gifts from Peter Tattersall; GAPDH (ab70699, abcam, UK)). Membranes were washed 3× in 0.1% PBS-T and incubated for 1 h at 25 °C with goat anti-rabbit IgG H and L HRP antibody (ab205718, abcam, UK). Membranes were washed 3× in 0.1% PBS-T, substrate was added (Amersham ECL Prime Western Blotting Detection Agent, GE Lifesciences) and images were exposed using the ImageQuant LAS4000 (GE Healthcare, UK). Images were analysed using ImageJ. The background was digitally subtracted and bands were quantified by measuring the mean grey value over equal areas.

RESULTS

Suppression of CpG and TpA dinucleotides in parvoviruses

Representative sequences of autonomous parvoviruses infecting mammals were divided into non-structural (NS) and VP1 (minor-capsid) coding regions and their mononucleotide and dinucleotide compositions were determined using the program Composition Scan in the SSE v1.3 package [26] and compared with those of mouse cellular mRNA genes (Fig. 2a, b). Both NS and VP1 genes of parvoviruses showed varying degrees of CpG suppression (Fig. 2a), with ratios of observed frequencies to expected frequencies based on G+C content ranging from <0.1 to 1.2 (*y*-axis). Suppression was greatest for parvovirus genomes with lower G+C content, a relationship recapitulated to an extent in mouse mRNA sequences. Frequencies of TpA dinucleotides in NS and VP1 genes showed a lower degree of suppression (Fig. 2b), again following a relationship observed in human mRNA with lower TpA frequencies in sequences with higher G+C content. The highlighted gene sequences of MVM show degrees of CpG and TpA suppression that are typical of other parvoviruses with low G+C content. CpG frequencies were 0.41 and 0.25

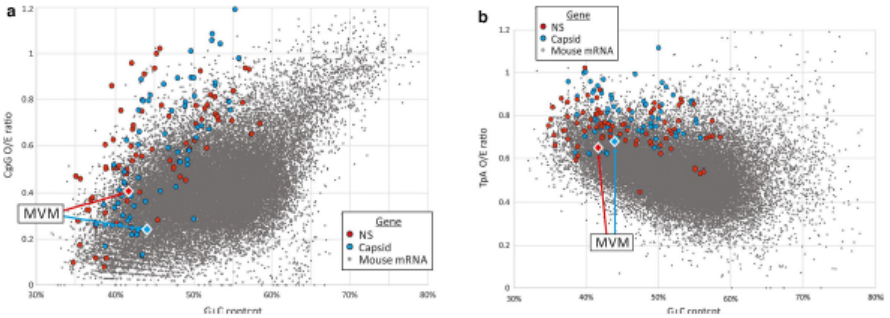


Fig. 2. CpG and TpA dinucleotide distribution in parvovirus genes. Comparison of frequencies of CpG and TpA dinucleotides in MVM NS- and capsid-encoding genes (red and blue filled diamonds) and representative other parvoviruses (circles) with those of mouse (*Mus musculus*) mRNA sequences (each grey point represents a separate gene). Observed (o) CpG and TpA frequencies were normalized to values expected (e) from their G+C content.

Table 1. Dinucleotide composition of MVM mutant regions

Mutant	Fragment length	G+C content	CpG		GpC		TpA	
			No.	O/E* ratio	No.	O/E ratio	No.	O/E ratio
WT	1257	42.32%	7	0.126	58	1.04	59	0.585
CDLR	1257	42.32%	7	0.126	58	1.04	59	0.585
CpG-H	1257	47.89%	**117**†	**1.634**	70	0.978	59	0.716
CpG-H first half	676	46.75%	**59**	**1.612**	39	1.066	37	0.8
CpG-H second half	582	49.14%	**58**	**1.66**	31	0.887	22	0.606
CpG-H 42% G+C	1257	42.32%	**78**	**1.398**	58	1.039	60	0.595
TpA-H	1257	42.32%	7	0.126	59	1.058	**124**	**1.227**
GpC-H	1257	42.32%	7	0.125	**91**	**1.626**	59	0.585
G+C-Max	1257	**57.44%**	7	0.068	88	0.858	59	1.061
CodOpt	1257	**61.10%**	**70**	**0.611**	108	0.943	23	0.524
CodOpt CpG-L	1257	**57.04%**	7	0.069	92	0.908	25	0.444
Whole genome	5149	42.12%	86	0.378	256	1.1251	312	0.7428

*O/E: observed over expected.
†Deliberately or secondarily modified frequencies are indicated in bold in a shaded cell.

for the NS and VP1 genes, respectively, while TpA was 0.65 and 0.69. A graphical representation of the positions of CpG and TpA dinucleotides is shown in Fig. 1c.

Effect of dinucleotide frequency modification on the replicative fitness of MVM

To study the mechanism or selection pressure underlying the suppression of CpG and TpA dinucleotides in MVM, the frequencies of CpG and TpA were modified in a region within the VP gene using synthetic DNA sequence inserts. This region was chosen as it contains a single open reading frame and has no known structural features or regulatory elements (Fig. 1a). This is supported by a high synonymous sequence variability, indicating conservation at the protein level but not the DNA or RNA level (Fig. 1b).

The native sequence in the chosen region (24.4% of the entire genome) was replaced by synthetically generated sequences containing a range of synonymous changes that left protein coding intact. The CDLR sequence was used as a mutagenesis control – codon order was scrambled to disrupt any underlying DNA or RNA structural elements or alternative reading frames in the sequence, while keeping coding, codon usage and dinucleotide frequencies identical to those of the WT sequence (Table 1). The CpG-high (CpG-H) mutant possessed maximized frequencies of CpG dinucleotides while retaining protein coding, achieving a ratio of 1.62× the expected frequency of CpGs based on this G+C content (O/E ratio; Table 1), 13× that of the native MVM sequence. TpA-high (TpA-H) sequences were similarly maximized to 1.23× the expected frequencies, approximately double that of the WT sequence.

To generate infectious virus, DNA clones of WT and mutated MVM genome sequences were transfected into NB324K cells and cell lysate was harvested at the onset of cytopathic effects (3–5 days post-transfection) and passaged for further replication in NB324K cells. Cell lysate from this passage was collected, aliquoted and titred, and used as a source of virus for phenotypic characterization. Viruses were titred in NB324K cells to estimate $TCID_{50}$s, while DNA concentrations in virus stocks were measured by qPCR using MVM-specific primers in the non-mutated NS region. Virus stocks of WT and mutant MVM virus stocks showed similar genome copy numbers, ranging between 10^{10} to 10^{11} DNA copies ml^{-1}.

Several MVM mutants showed substantial differences in replication kinetics from the WT virus in both in A9 and NB324K cell lines. These were apparent over the time course of a standard replication assay (Figs 3a and S1, available in the online version of this article) and in a cross-sectional analysis of virus titres in cultures in the A9 cell line at 48 h (Fig. 3b). The replication kinetics of the essential CDLR control was comparable to that of the WT virus in both cell

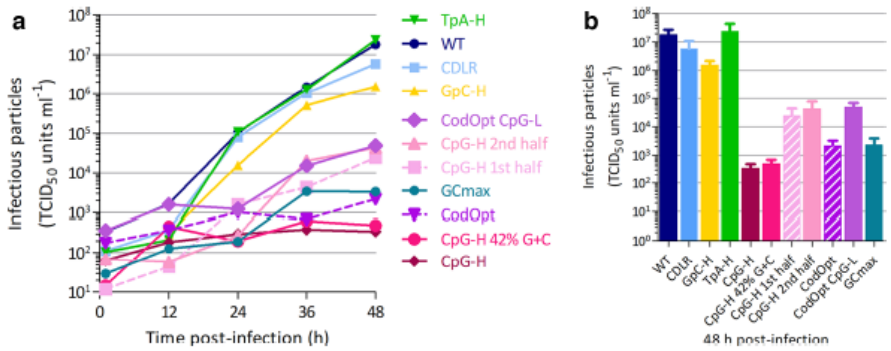


Fig. 3. Effects of V2 sequence modifications on MVM replication kinetics. (a) Multi-step growth curve in the A9 cell line infected with 1 $TCID_{50}$ unit/cell. Supernatant was harvested at different time points and titred in NB324K cells. Data points represent the mean of three biological replicates. (b) Cross-section of the growth curve at 48 h. Bar heights represent the mean of three biological replicates; error bar show SEMs.

lines, indicating that replication defects observed in other mutants (described below) did not originate from any DNA/RNA structure disruption caused by CDLR scrambling of the underlying coding sequence, or through the existence of uncharacterized alternative open reading frames.

However, modification of CpG frequencies led to a profound replication defect of MVM. The CpG-high mutant containing 117 CpG nucleotides instead of 7 barely replicated, with little net increase in supernatant infectivity titres over the duration of the assay. To investigate the dose-dependence of CpG addition on replication kinetics, a pair of mutants were generated in which the CpG modification was made either side of the *Hpa*I restriction site in the middle of the mutated region, with 59 in the 5′ half and 58 in the 3′ half (Table 1). Both CpG-high half-mutants were substantially attenuated but to a lesser extent than the full-region CpG-high mutant (containing 117 CpGs), with both showing an approximately three log reduction in replication abilities in A9 cells compared to WT. However, in contrast to the marked attenuation effects of increasing UpA sequences on the replication of E7 and influenza A virus [18, 21], the replication of the TpA MVM mutant matched that of the WT virus.

As controls we constructed a series of further mutants to investigate potential compounding effects of sequence manipulations on replication phenotypes. Firstly, increasing CpG frequencies potentially leads to a greater degree of DNA secondary structure in genomic DNA, or in RNA expressed in MVM structural gene mRNAs. To investigate this possibility separately from CpG frequency changes, we constructed a GpC-H mutant in which the similarly self-complementary GpC dinucleotides were increased in frequency (Table 1). However, the GpC-H mutant (with 91 GpC dinucleotides) replicated similarly to WT, providing no evidence that the attenuation of the CpG-H mutant originated from structural changes in the RNA. As a further control, we recognize that a further side-effect of increasing CpG content was to add further C and G residues to the sequence; in the VP2 mutated region this led to an increase in G+C content from 42.3–47.9%. To verify that the replication defect of the CpG-H mutant was not the result of the G+C content change, we constructed a new mutant in which the inserted C and G bases were compensated by removal elsewhere so that overall G+C content could be preserved. While retaining mononucleotide base composition limited the number of CpG dinucleotides that could be added (71 instead of 110), the resulting mutant (CpG-H 42% G+C) was highly attenuated (Fig. 3), indicating that perturbation in G+C content was unlikely to be the cause of attenuation in the CpG-H mutant.

Effect of codon optimization on MVM replication

Codon optimization is a frequently used mutational step to enhance expression of proteins from viral and vector-expressed mRNAs in biotechnology and vaccine applications. To investigate the effects of codon optimization on MVM replication, the insert region of VP2 was codon-optimized by a standard algorithm (see the Methods section) that is designed to maximize translation of the encoded protein (CodOpt). Remarkably, replication of the codon-optimized mutant was almost as attenuated as the CpG-high mutant, despite its VP region coding sequence being modified in a way that would conventionally enhance translation. Potentially, it was the addition of CpG dinucleotides to this genome region (7 to 70 and respective O/E ratios of 0.13 and 0.61; Table 1), rather than translation effects that led to its observed attenuation.

To investigate this potential compounding factor, we created a further mutant in which the MVM V2 coding region sequence was codon-optimized but the number of CpG dinucleotides was maintained at WT level (CodOpt CpG-L). This mutant,

however, also showed attenuation (approximately 2.5 logs lower than WT), although to a lesser extent than the four log reduction in the original CodOpt mutant. The other effect of codon optimization was to increase G+C content from 42.3 (WT) to 61.1 and 57.0% in the CodOpt and CodOpt CpG-L mutants, respectively. This may have alternatively influenced replication independently of codon usage and in a non-CpG-dependent manner. To investigate this, we constructed a further mutant in which G+C content was maximized (57.4%) while retaining native frequencies of CpG and TpA dinucleotides (GCmax; Table 1). This mutant showed comparable reduction in replication to that of the CodOpt mutant (approximately four logs; Fig. 3b), providing evidence that large-scale modification of genomic G+C composition does indeed have a substantial effect on replication separate from CpG representation.

Immunofluorescence assay and analysis

To compare the spatial and temporal distribution of gene expression from the NS and structural genes and how this was influenced by dinucleotide frequency modifications, infected cells were fixed with paraformaldehyde 18 h post-infection with WT and mutants of MVM and stained for the viral NS and VP proteins using specific antibodies (Figs 4 and S2). To ensure that cells were exposed to equal amounts of viral particles, cells were infected with equal copy numbers of WT and mutant virus stocks based on DNA/infectivity measurements (Fig. S3). Cells infected with WT and TpA-mutants of MVM showed similar frequencies and fluorescent intensities of NS and VP expression, while there was small reduction in the CDLR mutant. Contrastingly, the CpG-high and codon-optimized mutants showed greatly reduced frequencies and expression levels, particularly of the VP gene that contains the region modified by mutagenesis.

To quantify these reductions, representative areas of immunostaining from each mutant were analysed and the numbers of fluorescent foci and the intensity were determined in each section (Fig. 5a, b). This analysis showed only minor differences in frequencies and fluorescent intensity in expressing cells between WT, CDLR and TpA-high mutants. In contrast, it quantified a profound reduction in expressing cell frequencies and intensities in CpG-high and codon-optimized mutants. However, while the numbers of VP-expressing cells were lower than those expressing the NS protein, they both showed a comparable reduction in fluorescent intensity relative to WT in the cells that were infected. To separately verify these results using a different virus detection method, cells were infected for 18 h and viral protein content was measured using Western blot (Fig. 6a). Similarly to the immunofluorescence images, CDLR and TpA-H mutants showed only slight reductions in protein levels compared to WT, in contrast to the much greater reduction for the CpG-high and codon-optimized RNA.

RNA and protein expression levels

To investigate whether the observed differences in protein expression between WT virus and compositionally modified mutants was mediated at the translational or mRNA level, the translation rates of native and modified VP sequences were first determined in an extra-cellular *in vitro* translation assay. This assay determines translation efficiency independently of other potential virus-attenuating effects of altered composition mediated through mRNA stability or virus replication. A synthetic plasmid containing the partial VP sequence used for cloning (spanning positions 2647–4347 of the MVM genome) was linearized and capped RNA transcripts were synthesized. One microgram of each RNA was used in an *in vitro* rabbit reticulocyte translation assay and protein levels were assayed at 1.5 h using Western blot (Fig. 6b, c). In contrast to their marked effects on virus replication, the translation rates of CpG-H and CodOpt VP sequences were similar relative to that of the WT sequence. These observations provide no evidence that differences in the translation efficiency of the capsid gene influenced the attenuated phenotype of CpG-high mutants.

As an alternative mechanism of attenuation, mRNA sequences with high CpG frequencies may be unstable or more readily degraded in the cytoplasm and therefore disrupt MVM replication. To evaluate this possibility, mRNA levels of the full-length genomic transcripts were quantified in cells transfected with MVM DNA in the presence of a replication inhibitor (Aphidicolin) to avoid compounding the effects of virus replication differences on the measured outcomes. Because of the expression strategy of MVM, it was not possible to independently compare mRNA levels of NS- and VP1-encoding mRNAs, since the mRNA for NS contains the same modified sequences downstream from the NS gene as the VP1 gene (Fig. 1a). Viral RNA transcripts were quantified using primers targeting the NS1 region (R1) after 8 and 18 h (Fig. 7a). While the RNA levels at 8 h were initially similar, a reduction of R1 RNA levels relative to those of the WT and CDLR viruses was observed in the CpG-H mutant and to a lesser extent in the codon-optimized mutant at 18 h post-transfection. To further isolate the effects of virus replication on RNA expression, we constructed a sub-viral (R3) expression system in which WT and compositionally modified mutants of MVM cDNA sequences encoding the MVM VP proteins were cloned into a pCAG expression vector with transcription driven by the CAG promoter [41, 42]. This avoided possible compounding effects of MVM-related replication processes, such as expression of NS genes, on VP gene stability or turnover. Only the CpG-high mRNA showed a reduction in expression, 3-fold compared to the >10-fold reduction in the corresponding virus-expressed transcripts (Fig. 7b). The levels of the codon-optimized and UpA-high mRNAs were comparable to those of the WT and CDLR. These findings are therefore consistent with a role for RNA stability differences of CpG-modified sequences in their observed viral attenuation.

Stability of compositionally modified mutants

The attenuation of CpG-high mutants of MVM creates a selection pressure for the reversion of introduced CpG dinucleotides into the mutated VP region. To investigate the stability of the two CpG-high half-mutants along with

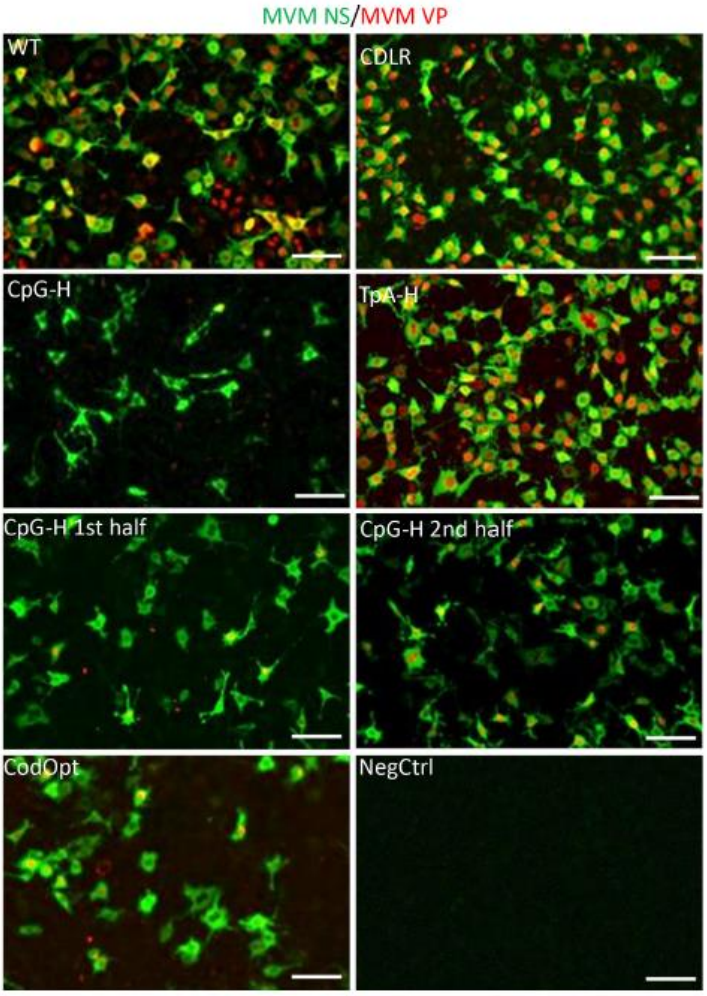


Fig. 4. Cellular distribution of MVM viral proteins in A9 cells. Immunofluorescence staining of viral NS (green) and VP (red) proteins at 18 h post-infection with 1000 viral copies/cell. Uninfected cells were used as a negative control. *n*=3, scale bar 100 µm, representative image shown.

the non-attenuated TpA-H and WT viruses as controls, viruses were serially passaged at high m.o.i. for 11 passages over a period of 5 weeks. Each passage was terminated on development of a full CPE. At the end of the passaging, viral genomic DNA was extracted and subjected to next-generation sequencing to investigate the occurrence of nucleotide (and inferred amino acid) substitutions in both mutated and non-mutated regions of the genome (Fig. S4). The CpG-H first and second half-mutants were remarkably stable on cell culture passage, with the CpG-H first half-mutant showing no reversions in the CpG-enriched mutated regions, and the occurrence of minor populations in the non-mutated NS region. The CpG-H second half-mutant showed two partial CpG reversions in the mutated region, and four partial or complete substitutions in the non-mutated first half-region. It also displayed a partial TpA reversion in the NS region. The TpA-H and WT viruses showed single, partial changes in the VP gene region that had no effect on CpG or TpA representation. In addition to demonstrating the high genetic stability of the MVM mutant on long-term *in vitro* passage,

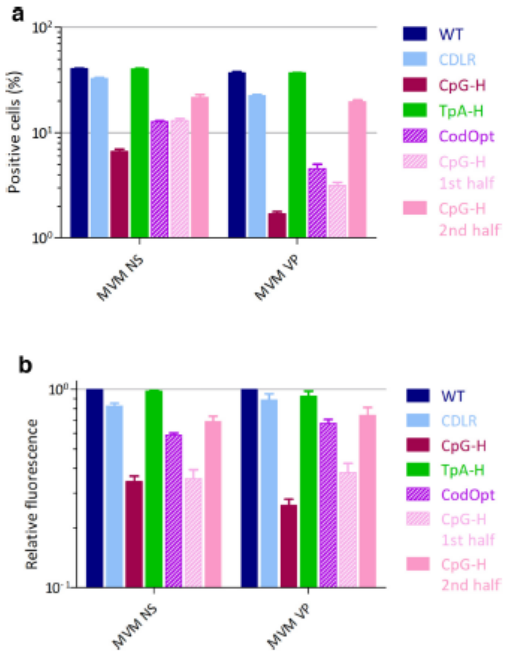


Fig. 5. Quantitative analysis of viral protein expression. Quantification of immunofluorescence staining of viral NS and VP proteins using Imaris and Fiji at 18 h post-infection with 1000 viral copies/cell. Uninfected cells were used as a negative control. $n=3$, +/−SEM. (a) Percentage of cells expressing NS or VP. (b) Relative fluorescence intensity of NS or VP compared to WT MVM.

nucleotide sequencing provided no evidence for any CpG-induced amino acid changes, deletions, insertions or other genome defects in either coding or replication regions that might underlie their substantial attenuation.

Mechanism of CpG-mediated attenuation of MVM replication

Differences in the degree of attenuation of CpG-high and codon-optimized mutants of MVM were apparent from their different replication kinetics in A9 and NB324K cell lines (Fig. 3 and S1). To investigate the extent to which attenuation was influenced by cell type, the susceptibility of a range of cell lines was evaluated through infection by WT virus at an m.o.i. of 1 and assay of supernatant infectivity at 72 h (Fig. 8a). NB324K, BHK and several immortalized mouse embryo fibroblast (MEF) cell lines with CRISPR-targeted knockout of different innate immune signalling pathway or effector genes were susceptible to MVM and could be evaluated for CpG-mediated attenuation (Fig. 8b; raw replication data are shown in Fig. S5).

The replication levels and degrees of attenuation of the CpG-H, CodOpt and both CpG-H half mutants were comparable in the NB324K and BHK cell lines. While the replication of MVM in unmodified MEF cells could barely be detected, knockout of protein kinase R (PKR), stimulator of interferon genes (STING) and interferon α receptor (IFNAR) reversed this restriction and enabled MVM replication to occur at comparable levels to that in A9 cells. However, in each k/o cell line, CpG-mediated attenuation was comparable to that observed in A9 cells, providing no evidence that these pathways participated in its attenuated phenotype compared to WT virus.

It has been previously proposed that non-methylated CpG sites in parvovirus genomic DNA activate a cellular antiviral response through Toll-like receptor 9 (TLR9) and it has been shown that MVM can be recognized by TLR9 in human PBMCs and HEK-hTLR9 cells, but not murine bone marrow cells or plasmacytoid dendritic cells [43, 44]. To investigate whether TLR9 mediated the attenuated phenotype of CpG-high mutants of MVM, we first determined whether the A9 cells in which CpG-mediated attenuation was most evident can actually signal through TLR9. IL-1β mRNA levels were measured before and after stimulation with *E. coli* DNA, a TLR-9-specific stimulant (Fig. 9a). In the same experiment, we also tested the efficacy of the TLR-9-specific antagonist ODN2088. This is a 15-mer oligodinucleotide (5′-TCC TGG CGG GGA AGT-3′) that inhibits binding of unmethylated CpG sequences to TLR9 [45].

As a positive control, RAW264.7 cells were responsive to TLR9 stimulation, with 2 h pre-treatment with ODN2088 inhibiting most of this response. In contrast, A9 cells showed no induction of IL-1 β mRNA. We additionally assayed the potential effects of TLR9 expression on attenuation in a functional assay. Replication of WT and the CpG-H first half-mutant was compared between A9 cells and those pretreated with ODN2088. Two hours' pretreatment of A9 cells with ODN2088 had no or minimal effects on the replication of WT or TpA-H MVM (Fig. 9b). Significantly, there was no relevant recovery of replication ability of the CpG-H first half-mutant, thereby providing no evidence that TLR9 controls its replication in the A9 cell line.

ZAP has previously been shown to participate in the attenuation of mutants of HIV-1 with elevated CpG frequencies. Similarities in the replication strategy of parvoviruses and retroviruses suggest a potential role for ZAP in mediating the restriction of CpG-modified mutants of MVM. We have previously constructed ZAP k/o A549 cell lines to investigate the attenuation pathways of E7 [21]. Replication of MVM in this human cell line was less efficient than in A9 cells, although both mRNA and protein expression and limited genomic replication occurred and could be quantified.

To quantify the effects of ZAP expression on MVM RNA levels, A549 and B8 (ZAP k/o) cells were transfected with DNA sequences of WT and CpG-H mutants of MVM in the presence of aphidicolin; R1 mRNA transcript levels were then quantified by qPCR at 8 and 18 h (Fig. 10). As

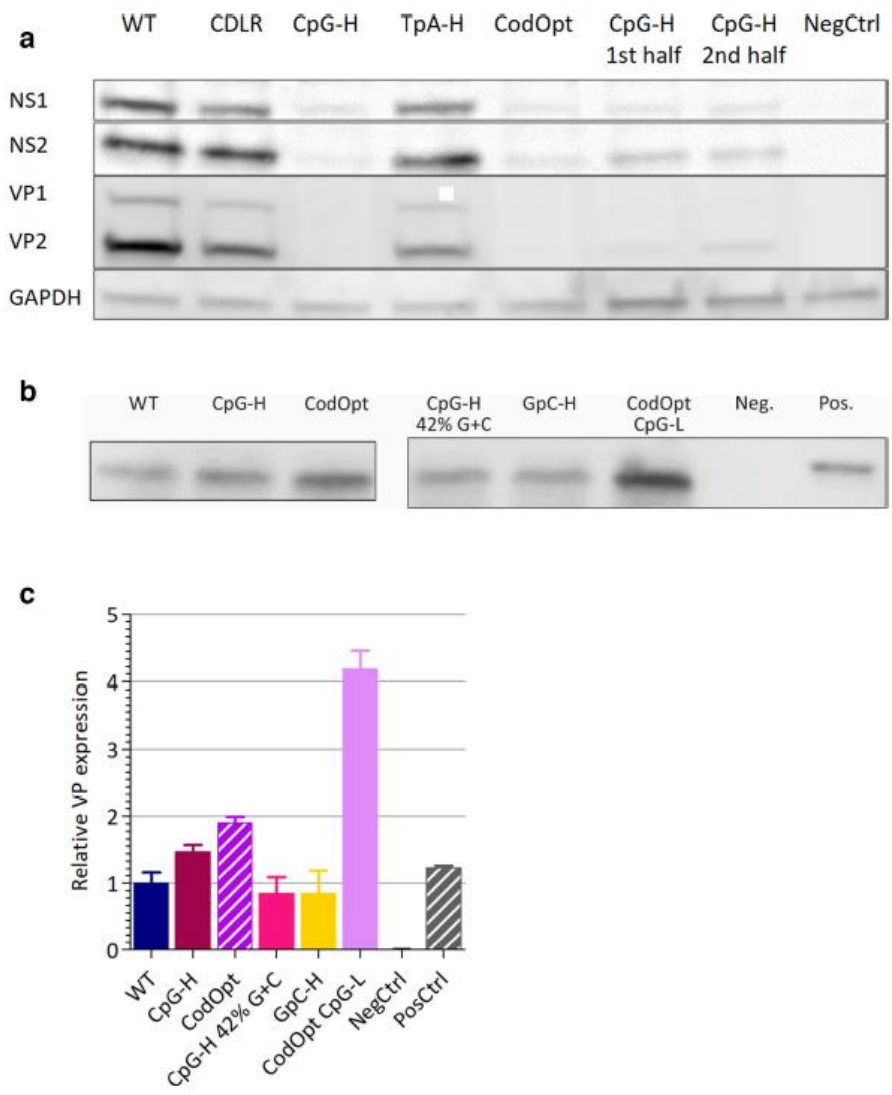


Fig. 6. MVM gene expression and translation. (a) Viral protein expression 18 h post-infection with 1000 viral copies/cell. Uninfected cells were used as a negative control. *n*=3, representative image shown. (b) *In vitro* translation of a truncated VP2. RNA was transcribed into mRNA *in vitro* via a T7 promoter using the synthetic plasmids containing the different MVM inserts. Equal amounts of the RNA were then translated using an *in vitro* rabbit reticulocyte and measured using immunoblotting. The negative control is a mock translation with no RNA added, while the positive control is an infected cell lysate (VP2 band shown). (c) Quantification of VP2 protein levels (from Fig. 6b) by ImageJ expressed as the ratio to that of the WT gene (*y*-axis). Bar heights shows the mean of three replicates; error bars show +/−SEM.

observed previously in A9 cells (Fig. 7), mRNA levels were comparable in A549s at 8 h, but then declined by approximately fourfold at 18 h. In contrast, there was no decline in RNA levels of the CpG-H mutant between 8 and 18 h in transfected B8 cells, indicating a potential role for ZAP in the observed reduced expression or instability of CpG-high RNA sequences in this context.

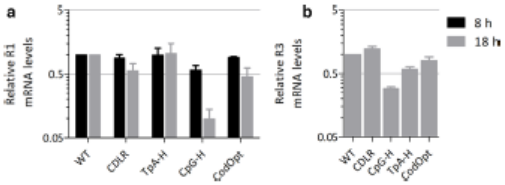


Fig. 7. MVM R1 mRNA levels post-transfection. MVM R1 mRNA expression 8 and 18 h post-transfection normalized to cellular GAPDH and WT R1 levels. *n*=3, +/−SEM. (a) Plasmid pMM984, viral infectious clone, replication inhibited by Aphidicolin. (b) Plasmid pCAG_MVMp_R3, expression vector containing MVM R3 subclone.

Further evidence for a role of ZAP in CpG-mediated attenuation was provided by quantitation of the virus replication of MVM and CpG-high mutants by immunocytochemistry and replication kinetics in the B8 ZAP k/o cell line. A549 and B8 cells were infected with WT, CpG-H or CodOpt virus and stained for both NS and VP proteins (representative sections are shown in Fig. 11a, full panel and close up in Figs S6 and S7). The frequencies and fluorescent intensities of stained capsid and NS proteins were separately quantified (Fig. 11b). The frequency of cells expressing NS or VP was reduced threefold for the CpG-H mutant in A549 cells but with a much lesser reduction in infection frequencies of the codon-optimized mutant. Similarly, the average intensity of cells infected with the CpG-H mutant was reduced 4-fold for the NS proteins and 9-fold for the VP proteins, while CodOpt was reduced 1.5–2-fold. In contrast, the proportion of B8 ZAP k/o cells expressing NS or VP was comparable to that of the WT virus, effectively reversing the restricted phenotype observed in A549 cells. The average fluorescence intensity of NS in B8 cells infected with the CpG-H mutants was also restored to WT levels (Fig. 11b). There was also restoration of VP expression, but remarkably there was no relocalization of VP to the nucleus of ZAP k/o cells, with protein remaining in punctate granules in the cytoplasm. These findings indicate that ZAP plays a role in restricting VP expression of CpG-high virus mutants, but other mechanisms also influenced by RNA composition potentially play roles in subsequent steps in MVM replication.

Infection of A549 cells led to a full infection cycle of MVM and produced detectable levels of infectious particles at 48 h post-infection, albeit at a lower level than in other cell lines (Figs. 8c, 11a). Replication of the CpG-H first half-mutant was impaired compared to WT in the A549 cell line, but not in the B8 ZAP k/o cell line (Fig. 11c).

DISCUSSION

This study was motivated by the observation that the suppression of CpG dinucleotide frequencies found in vertebrate RNA viruses and HIV-1 also occurs in small DNA viruses, such as parvoviruses, polyomaviruses and papillomaviruses [10, 12, 46]. In the current study, we document a remarkable

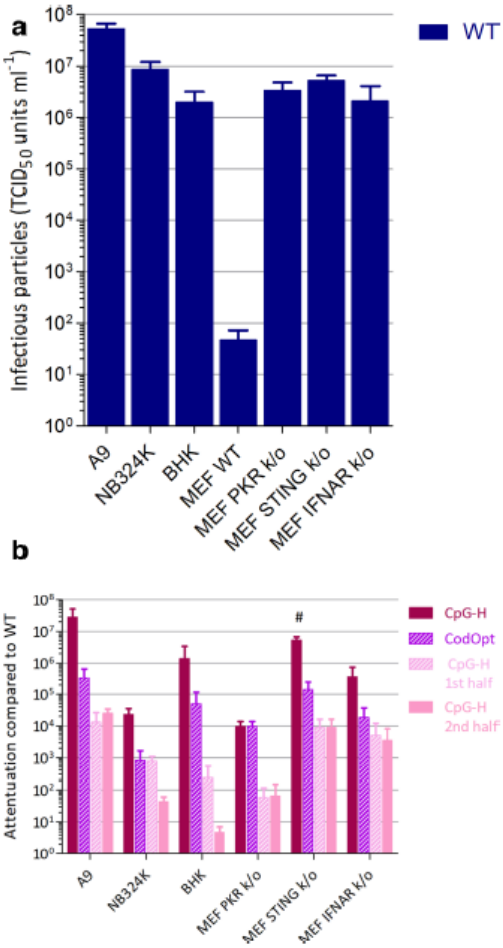


Fig. 8. Cell line-associated differences in MVM replication. Infectivity of MVM 72 h post-infection (1 TCID unit/cell) in different cell lines. *n*=2, +/−SD. (a) Viral titres of MVM WT virus. (b) Attenuation of CpG-high and codon-optimized mutants of compared to WT. # indicates samples where no virus could be recovered, thus underestimating attenuation.

attenuation of the replication ability of MVM constructs with artificially increased frequencies of CpG dinucleotides in the VP2 gene, with several log reductions in replication kinetics compared to WT virus and a detectable frequency of reversion of CpG sites on passaging CpG-H mutant viruses. The study used a variety of approaches to investigate the mechanism for this attenuation in the context of previous descriptions of potential effects of CpG addition on genome methylation, translation efficiency and mRNA stability.

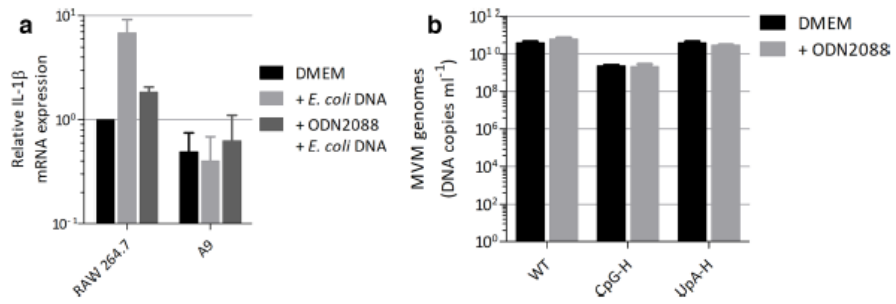


Fig. 9. Involvement of TLR9 in MVM mutant phenotype. (a) IL-1β induction after TLR9 induction by purified *E. coli* DNA in the presence or absence of TLR-9 antagonist ODN2088 . n=3, +/−SEM. b) Viral genome copies 72 h post-infection with 1000 viral genome copies/cell in the presence or absence of ODN2088. n=3.

It has previously been hypothesized that the suppression of CpG frequencies in parvoviruses and potentially other small DNA viruses results from methylation-induced hypermutation by cellular methyltransferases in the nuclear compartment. Such processes may reduce the visibility of DNA viruses to TLR9, which acts as a bacterial and viral pattern recognition receptor through its specific recognition of non-methylated CpG dinucleotides in DNA sequences. However, it has been since shown that the genome of porcine parvovirus (PPV) is hypo-methylated throughout its replication cycle, both in replicating DNA complexes and genomic DNA packaged into virions; genomes of adenoviruses, papillomaviruses and polyomaviruses are similarly methylation deficient [47–51]. Artificial methylation of PPV genomes prior to transfection showed a small inhibitory rather than enhancing effect on replication initiation and their infectious progeny became non-methylated.

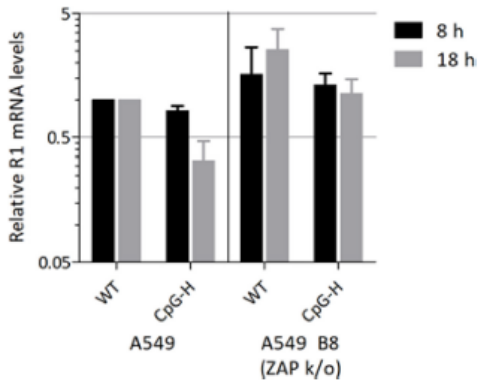


Fig. 10. MVM R1 mRNA levels in A549 andA549 B8 (ZAP k/o) cell lines post-transfection. MVM R1 mRNA expression 8 and 18 h post-transfection normalized to cellular GAPDH and MVM WT R1 levels in A549 cells. PlasmidpMM984, viral infectious clone, replication inhibited by Aphidicolin. n=3, +/−SEM.

MVM and parvovirus H-1 could be detected by TLR9 in human PBMCs and HEK-hTLR9 MVM, but adenovirus association virus 2 remained unrecognized by TLR9 expressed at high level on murine plasmacytoid dendritic cells [43, 44]. A further demonstration of a lack of involvement of TLR9 in CpG suppression was provided by our own observations that the A9 cell line in which CpG-H mutants of MVM were attenuated showed no induction of IL-1β after TLR9 stimulation and no reversal of attenuation when TLR9 signalling was inhibited with an TLR9 antagonist (Fig. 9). As previously proposed, it appears that the suppression of CpG dinucleotides in parvovirus genomes originates from a different selection or mutational pressure from that which drives vertebrate genomic CpG under-representation, although further attenuation through TLR9 involvement cannot be excluded in cell lines with a functional TLR9 pathway.

In the current study we have investigated other potential mechanisms underlying the observed suppression of CpG dinucleotide frequencies in MVM and other parvoviruses. In a range of RNA viruses, it has previously been proposed that CpG suppression reflects a process of avoiding unfavoured codons and codon pairs that contain CpG dinucleotides within codons or across codon boundaries [52–55]. Their presence may reduce translation rates from the mRNA template and restrict virus replication. However, the observation of high levels of attenuation of a mutant of MVM with a codon-optimized capsid gene is entirely inconsistent with the proposed translational basis for attenuation. While the CodOpt RNA template sequence translated more efficiently in a cell-free assay to the native WT sequence (Fig. 6b, c), the mutant virus constructed with this capsid sequence showed extreme attenuation (Figs. 4 and 6a), more compatible with effects of the concomitant increase in the number of CpG dinucleotides (WT: 7, CodOpt: 70; Table 1) and an increase in G+C content, which additionally influenced replication (Fig. 3). These findings

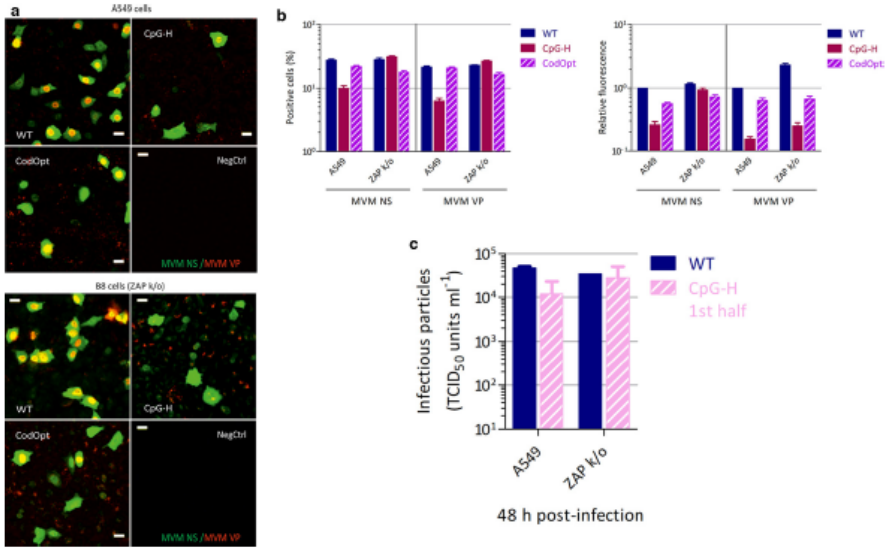


Fig. 11. Cellular distribution of MVM viral proteins in A549 and B8 (ZAP k/o) cells . (a) Immunofluorescence staining of viral NS (green) and VP (red) proteins at 18 h post-infection with 1000 viral copies/cell. Uninfected cells were used as a negative control. (b) Quantification of immunofluorescence staining of viral NS and VP proteins using Imaris and Fiji software separately for the frequency of cells expressing NS or VP (left) or relative fluorescence intensity of NS or VP compared to WT MVM (right). Bar heights represent the mean of three biological replicates; error bars show SEMs. (c) Replication of WT and CpG-H first half-mutant of MVM in A549 and ZAP k/o cells infected at an m.o.i. of 1. Cell supernatant was harvested at 48 h and infectivity was determined by titration on NB324K cells. Bar heights show the mean of two replicates; error bars show the SD.

are consistent with our previous analysis of mutants of E7 in which CpG and codon pair bias were independently manipulated with the outcome that attenuation originated from CpG and UpA frequency changes and the effects of codon pair changes were secondary to that [17]. They also concur with our previous observations that CpG (and UpA) dinucleotides added to a non-coding region of an E7 replicon construct produced similar attenuation to mutants where the same compositional modification was made to the coding region [56].

CpG-mediated attenuation of MVM was dose-dependent, with the half-mutants with only 58 or 59 CpGs added showing around 1000-fold enhanced replication compared to the full CpG-H mutant (Fig. 3a, b), although remaining 1000-fold less replication-competent than WT virus. This dose dependence potentially accounts for the previous failure to observe any substantial attenuation of parvovirus H-1 and porcine parvovirus with far fewer CpGs introduced into their genome (5–13 and 9–29, respectively). A contributory factor was the use of kidney-derived cell lines (NB324K and HEK293) that, along with BHK, show far less CpG-associated attenuation than fibroblast cell cultures [21, 56]. Construction of further mutants with a wider range of CpG compositions would be required to determine whether there was a minimum threshold concentration/spacing to achieve the CpG-mediated attenuation observed in the current study. Despite the severe attenuation, CpG-H mutants remained remarkably stable during *in vitro* passage (Fig. S3), indicating that the virus cannot readily escape the detrimental mutations while remaining viable.

STING has been identified as member of the principal pathway activated by a range of cytoplasmic pattern recognition receptors (PRRs) for viral DNA (reviewed in [57, 58]). Although undescribed to date, it is possible that there exist PRRs for viral genomic DNA with a sequence specificity for CpG that signal through STING. To investigate this, we investigated the susceptibility of STING and a range of other pathway k/os in MEF cells to MVM and a range of compositionally modified mutants. Parental MEFs did not support MVM replication, but MEFs with knockouts of PKR, STING and interferon α receptor (IFNAR) supported a high level replication, indicative of their potent roles in restricting parvovirus replication (Fig. 8a). The observed susceptibility of PKR k/o cells to MVM is consistent with previous observations [59]. While these three pathways substantially restricted MVM replication, there was no evidence for any dinucleotide compositional specificity in their restriction, with k/o cells showing equivalent attenuation of CpG-H mutant replication to that observed in A9 cells. Given the centrality of these three proteins and associated pathways in the innate cellular

response to virus infection, these observations effectively rule out a vast swathe of DNA PRRs and other activators of IRF3- or IRF7-induced interferon activated pathways in CpG-mediated restriction, including TLR9 (Fig. 8b).

If not through cytoplasmic PRRs for genomic DNA, CpG suppression in parvoviruses may alternatively originate from pathways recognizing CpG dinucleotides in viral mRNA sequences expressed during the replication cycle [18, 19, 56, 60]. Supporting this possibility are the data we obtained that document a role of ZAP in mediating the restriction of CpG-high mutants of MVM (Figs. 10 and 11), a pathway shown previously to be responsible for the attenuation of CpG-high mutants of HIV-1 [19] and in a different RNA virus model based on E7 [21]. Concerning the former, there are indeed several parallels in the strategy for gene expression of HIV-1 and parvoviruses, in nuclear RNA transcription and export of mRNAs into the cytoplasm, and therefore potentially in the steps in their replication targeted by ZAP. For HIV-1, ZAP has been shown to specifically target and bind to CpG sites in mRNA sequences of CpG-high mutants [19], needing KHNYN protein as a co-factor to exert its antiviral activity against CpG-rich RNAs [20]. The observation of higher rates of degradation of CpG-high MVM sequences in A549 cells (Fig. 10) and its reversion of ZAP k/o cells supports the involvement of this pathway in the attenuation of MVM. This is further indicated by restoration of NS and partial restoration of capsid gene expression in B8 cells from CpG-high MVM mutants on immunocytochemistry and of replication in A549 cells (Fig. 11).

However, a difference between the data collected for MVM in the current study and previous analyses of E7 and influenza A virus was the consistent absence of attenuation of the TpA-high mutant of MVM (Figs. 3–7). The increase in TpA dinucleotides achieved by mutagenesis and the proportion of the genome modified (24.4%) were comparable to those for E7, in which the UpA-high mutant showed an approximately 100-fold reduction in replication kinetics in an assay format comparable to that used for MVM (Fig. 3). The replication of the TpA-high mutant of MVM was similar to that for the WT in all other assay formats (Figs. 4–7), and did not show any evidence for the reduction in RNA stability observed for CpG-enriched sequences (Fig. 7). It is possible that this reflects the outcome of different restriction mechanisms responding to high CpG and TpA sequences and the inactivity of the latter pathways in parvovirus replication. However, our previous analyses provided consistent evidence that the pathway restricting CpG- and UpA-high mutants is shared. In recently completed work [21], we indeed demonstrated that ZAP is responsible for the attenuated phenotype of both CpG- and UpA-high mutants; ZAP furthermore shows a similar binding affinity to E7 viral RNA transcripts with inserted UpA-high or CpG-high regions. How the corresponding TpA-high mutant of MVM escapes these mechanisms operating at the RNA level on viral mRNAs is therefore currently unclear.

The marked attenuation of the codon-optimized mutant of MVM (Fig. 3) was unexpected and remarkable. Codon optimization is widely used to enhance the expression of transgenes in a variety of DNA and virus vectors, including other members of the *Parvoviridae*, such as adeno-associated viruses (AAVs) [61–63]. However, it appears that in MVM and perhaps in other small DNA viruses and retroviruses with suppressed CpG frequencies, codon optimization has a severely damaging effect on expression (Fig. 4). A similar effect has been observed in adenovirus, where most viral proteins are codon-optimized, but optimizing the non-optimized fibre protein led to a decrease in viral fitness [64].

In summary, this study provides compelling evidence for the action of cellular restriction of CpG-modified mutants of MVM operating at the RNA level, attenuating and affecting all downstream stages of the viral replication cycle. The study provides further evidence for a commonality in restriction mechanisms and their effect on driving down CpG frequencies in most RNA viruses and small DNA viruses infecting vertebrates, rather than through previously proposed TLR9 recognition.

Funding information
This work was supported by a Wellcome Trust Senor Investigator Award (WT103767MA).

Acknowledgements
The authors are grateful to Valerie Odon, Ahmad Ibrahim, Isabelle Dietrich, Jelke Fros and Paul Klenerman for valuable discussion and advice while the project was being performed. The authors would like to acknowledge David J. Pintel (Department of Molecular Microbiology and Immunology, University of Missouri School of Medicine, Columbia, Missouri, USA) for the generous provision of the MVM infectious clone, Peter Tattersall (School of Medicine, Yale University, New Haven, Connecticut, USA) for generous provision of the A9 and NB324K cell lines and immunocytochemistry reagents used for the MVM model and Jan Rehwinkel (Radcliffe Department of Medicine, University of Oxford, Oxford, UK) for generous provision of the MEF knockout cell lines and Lynn Dustin.

Conflicts of interest
The authors declare that there are no conflicts of interest

References

1. **Josse J, Kaiser AD, Kornberg A.** Enzymatic synthesis of deoxyribonucleic acid. VIII. frequencies of nearest neighbor base sequences in deoxyribonucleic acid. *J Biol Chem* 1961;236:864–875.
2. **Swartz MN, Trautner TA, Kornberg A.** Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J Biol Chem* 1962;237:1961–1967.
3. **Russell GJ, Walker PM, Elton RA, Subak-Sharpe JH.** Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J Mol Biol* 1976;108:1–20.
4. **Bird AP.** DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* 1980;8:1499–1504.
5. **Simmen MW.** Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics* 2008;92:33–40.
6. **Jabbari K, Bernardi G.** Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* 2004;333:143–149.
7. **Hendrich B, Tweedie S.** The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet* 2003;19:269–277.

8. Tweedie S, Charlton J, Clark V, Bird A. Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* 1997;17:1469–1475.
9. Regev A, Lamb MJ, Jablonka E. The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol Biol Evol* 1998;15:880–891.
10. Karlin S, Doerfler W, Cardon LR. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J Virol* 1994;68:2889–2897.
11. Rima BK, McFerran NV. Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *J Gen Virol* 1997;78 (Pt 11:2859–2870.
12. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol* 2006;62:551–563.
13. Beutler E, Gelbart T, Han JH, Koziol JA, Beutler B. Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. *Proc Natl Acad Sci USA* 1989;86:192–196.
14. Duan J, Antezana MA. Mammalian mutation pressure, synonymous codon choice, and mRNA degradation. *J Mol Evol* 2003;57:694–701.
15. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J *et al*. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J Virol* 2009;83:9957–9969.
16. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res* 2014;42:4527–4545.
17. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *eLife* 2014;3.
18. Gaunt E, Wise HM, Zhang H, Lee LN, Atkinson NJ *et al*. Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *Elife* 2016;5:e12735.
19. Takata MA, Gonçalves-Carneiro D, Zang TM, Soll SJ, York A *et al*. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* 2017;550:124–127.
20. Ficarelli M, Wilson H, Pedro Galão R, Mazzon M, Antzin-Anduetza I *et al*. KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *Elife* 2019;8:e46767.
21. Odon V, Fros JJ, Goonawardane N, Dietrich I, Ibrahim A *et al*. The role of ZAP and OAS3/RNAseL pathways in the attenuation of an RNA virus with elevated frequencies of CpG and uPA dinucleotides. *Nucleic Acids Res* 2019;47:8061–8083.
22. Cotmore SF, Tattersall P. Parvoviruses: small does not mean simple. *Annu Rev Virol* 2014;1:517–537.
23. Merchlinsky MJ, Tattersall PJ, Leary JJ, Cotmore SF, Gardiner EM *et al*. Construction of an infectious molecular clone of the autonomous parvovirus minute virus of mice. *J Virol* 1983;47:227–232.
24. Littlefield JW. Three degrees of guanylic acid-inosinic acid pyrophosphorylase deficiency in mouse fibroblasts. *Nature* 1964;203:1142–1144.
25. Shein HM, Enders JF. Multiplication and cytopathogenicity of Simian vacuolating virus 40 in cultures of human tissues. *Proc Soc Exp Biol Med* 1962;109:495–500.
26. Simmonds P. SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* 2012;5:50.
27. Chin JX, Chung BK-S, Lee D-Y. Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics* 2014;30:2210–2212.
28. Tattersall P, Cawte PJ, Shatkin AJ, Ward DC. Three structural polypeptides coded for by minite virus of mice, a parvovirus. *J Virol* 1976;20:273–289.
29. Ramakrishnan MA. Determination of 50% endpoint titer using a simple formula. *World J Virol* 2016;5:85–86.
30. Zhan D, Roy MR, Valera C, Cardenas J, Vennari JC *et al*. Detection of minute virus of mice using real time quantitative PCR in assessment of virus clearance during the purification of mammalian cell substrate derived biotherapeutics. *Biologicals* 2002;30:259–270.
31. Li L, Cotmore SF, Tattersall P. Parvoviral left-end hairpin ears are essential during infection for establishing a functional intranuclear transcription template and for efficient progeny genome encapsidation. *J Virol* 2013;87:10501–10514.
32. Hamers AAJ, Uleman S, van Tiel CM, Kruijswijk D, van Stalborch A-M *et al*. Limited role of nuclear receptor Nur77 in Escherichia coli-induced peritonitis. *Infect Immun* 2014;82:253–264.
33. Ahn K, Huh J-W, Park S-J, Kim D-S, Ha H-S *et al*. Selection of internal reference genes for SYBR green qRT-PCR studies of rhesus monkey (*Macaca mulatta*) tissues. *BMC Mol Biol* 2008;9:78.
34. Banerjee S, Xie N, Cui H, Tan Z, Yang S *et al*. MicroRNA let-7c regulates macrophage polarization. *J Immunol* 2013;190:6542–6549.
35. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;25:402–408.
36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–.
38. Cotmore SF, D'Abramo AM, Carbonell LF, Bratton J, Tattersall P. The NS2 polypeptide of parvovirus MVM is required for capsid assembly in murine cells. *Virology* 1997;231:267–280.
39. Cotmore SF, Tattersall P. Organization of nonstructural genes of the autonomous parvovirus minute virus of mice. *J Virol* 1986;58:724–732.
40. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M *et al*. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;9:676–682.
41. Niwa H, Yamamura K, Miyazaki J, Hitoshi N, Ken-ichi Y, Jun-ichi M. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 1991;108:193–199.
42. Miyazaki J, Takaki S, Araki K, Tashiro F, Tominaga A *et al*. Expression vector system based on the chicken beta-actin promoter directs efficient production of interleukin-5. *Gene* 1989;79:269–277.
43. Raykov Z, Grekova SP, Hörlein R, Leuchs B, Giese T *et al*. TLR-9 contributes to the antiviral innate immune sensing of rodent parvoviruses MVMp and H-1PV by normal human immune cells. *PLoS One* 2013;8::e55086-e.
44. Mattei LM, Cotmore SF, Li L, Tattersall P, Iwasaki A. Toll-like receptor 9 in plasmacytoid dendritic cells fails to detect parvoviruses. *J Virol* 2013;87:3605–3608.
45. Stunz LL, Lenert P, Peckham D, Yi A-K, Haxhinasto S *et al*. Inhibitory oligonucleotides specifically block effects of stimulatory CpG oligonucleotides in B cells. *Eur J Immunol* 2002;32:1212–1222.
46. Hoelzer K, Shackelton LA, Parrish CR. Presence and role of cytosine methylation in DNA viruses of animals. *Nucleic Acids Res* 2008;36:2825–2837.
47. Wienhues U, Doerfler W. Lack of evidence for methylation of parental and newly synthesized adenovirus type 2 DNA in productive infections. *J Virol* 1985;56:320–324.
48. Burnett TS, Sleeman JP. Uneven distribution of methylation sites within the human papillomavirus 1a genome: possible relevance to viral gene expression. *Nucleic Acids Res* 1984;12:8847–8860.
49. Danos O, Katinka M, Yaniv M. Molecular cloning, refined physical map and heterogeneity of methylation sites of papilloma virus type 1A DNA. *Eur J Biochem* 1980;109:457–461.
50. Gunthert U, Schweiger M, Stupp M, Doerfler W. DNA methylation in adenovirus, adenovirus-transformed cells, and host cells. *Proc Natl Acad Sci U S A* 1976;73:3923–3927.
51. Winocour E, Kaye AM, Stollar V. Synthesis and transmethylation of DNA in polyoma-infected cultures. *Virology* 1965;27:156–169.

52. Le Nouën C, Brock LG, Luongo C, McCarty T, Yang L *et al*. Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proc Natl Acad Sci U S A* 2014;111:pii: 201411290–.
53. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E *et al*. Virus attenuation by genome-scale changes in codon pair bias. *Science* 2008;320:1784–1787.
54. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol* 2006;80:9687–9696.
55. Burns CC, Shaw J, Campagnoli R, Jorba J, Vincent A *et al*. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J Virol* 2006;80:3259–3272.
56. Fros JJ, Dietrich I, Alshaikhahmed K, Passchier TC, Evans DJ *et al*. Cpg and uPA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. *elife* 2017;6:e29112.
57. Banete A, Seaver K, Bakshi D, Gee K, Basta S. On taking the sting out of immune activation. *J Leukoc Biol* 2018;103:1189–.
58. Luo W-W, Shu H-B. Delicate regulation of the cGAS-MITA-mediated innate immune response. *Cell Mol Immunol* 2018;15:666–675.
59. Ventoso I, Berlanga JJ, Almendral JM. Translation control by protein kinase R restricts minute virus of mice infection: role in parvovirus oncolysis. *J Virol* 2010;84:5043–5051.
60. Antzin-Anduetza I, Mahiet C, Granger LA, Odendall C, Swanson CM. Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology* 2017;14:49.
61. Ramezanpour B, Haan I, Osterhaus A, Claassen E. Vector-based genetically modified vaccines: exploiting Jenner's legacy. *Vaccine* 2016;34:6436–6448.
62. André S, Seed B, Eberle J, Schraut W, Bültmann A *et al*. Increased immune response elicited by DNA vaccination with a synthetic gp120 sequence with optimized codon usage. *J Virol* 1998;72:1497–1503.
63. Megati S, Garcia-Hand D, Cappello S, Roopchand V, Masood A *et al*. Modifying the HIV-1 env gp160 gene to improve pDNA vaccine-elicited cell-mediated immune responses. *Vaccine* 2008;26:5083–5094.
64. Villanueva E, Martí-Solano M, Fillat C. Codon optimization of the adenoviral fiber negatively impacts structural protein expression and viral fitness. *Sci Rep* 2016;6:27546.
65. Berns KI, Parrish CR. Parvoviridae. In: Knipe DM, Howley PM (editors). *Fields Virology*, 6th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2013. pp. 1768–1791.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.