

UNIVERSITE D'AIX-MARSEILLE

ECOLE DOCTORALE

LABORATOIRE AFMB « Architecture et Fonction des
Macromolécules Biologiques »

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : ED 62 – SCIENCES DE LA VIE ET DE LA SANTE

Spécialité : Biochimie Structural

Ana Sofia FERREIRA RAMOS

Inhibitors of the mRNA capping machinery and
structural studies on *Macro* domains from
Alphaviruses

Soutenue le 08/07/2019 devant le jury :

Dr Laurence BRIANT	Université de Montpellier	Rapporteur
Pr Martijn van HEMERT	Leids Universitair Medisch Centrum	Rapporteur
Dr Hélène MALET	Université Grenoble Alpes	Examineur
Dr Bruno CANARD	AFMB, CNRS, Aix-Marseille Université	Examineur
Dr Gerlind SULZENBACHER	AFMB, CNRS, Aix-Marseille Université	Co- Directeur de thèse
Pr Bruno COUTARD	UVE, Aix-Marseille Université	Directeur de thèse

Numéro national de thèse/suffixe local : 2019AIXM0171/040ED62

THIS PAGE IS INTENTIONALLY LEFT BLANK

Acknowledgements

First and foremost, I want to express my sincere gratitude to Professor Doctor Bruno Coutard and Professor Doctor Bruno Canard who provided me an opportunity to join their team as a PhD student and the precious support to conduct this research and to write this manuscript. It was an honour to be part of the ANTIVIRALS consortium that provided opportunities for collaborations and that played an important role in the development of my scientific and non-scientific competences. I acknowledge the funding sources that made my PhD work possible. My PhD work was mainly founded by the Horizon 2020 Marie Skłodowska-Curie ETN “ANTIVIRALS”, grant agreement number 642434, and with beamtime from ESRF and SOLEIL synchrotrons.

The scientific guidance of Professor Doctor Bruno Coutard was the essence that inspired me during my labour days in the laboratory which resulted in the results presented in this manuscript. I want to specially thank to Professor Doctor Bruno Coutard for having fully supported me during my installation in Marseille and for the joy and enthusiasm he shared with me specially during the first half of my PhD. He has contributed immensely to all the opportunities I had during my PhD namely, conferences, workshops and, collaborations. I very much appreciated his enthusiasm, willingness and extraordinary ability to help me to manipulate crystals and to help me to prepare my first trips to the synchrotron. I will miss his inspirational discussions regarding my experiments and, I will always remember with joy all his gastronomic advices and suggestions.

Besides my supervisor Professor Doctor Bruno Coutard, I would like to express my gratitude to my co-supervisor Doctor Gerlind Sulzenbacher for her patience in teaching me X-ray crystallography and for her support in the more difficult times. I want to specially thank her for her helpful discussions and for being actively involved in supervise my work and the conception of this manuscript. During a huge part of my PhD, Doctor Gerlind Sulzenbacher was my source of inspiration for the X-ray crystallographic work, from data collection to structure refinement. I appreciated all her contributions of time and ideas to make my PhD experience productive and stimulating.

I would like specially to thank the bag group members that joined with me several synchrotron trips; Doctor Gerlind Sulzenbacher, Doctor Silvia Spinelli, Doctor Jennifer Roche, Doctor Brigitt Raux, Doctor Magali Saez-Ayala and Doctor Etienne Rebuffet. I was always integrated very quickly into each “team mission” in which I was involved. They were very good team players and supportive. I appreciated their expertise and with them it was possible to learn a lot and I got memories of joy, good team partnership, cooperativity, enthusiasm, friendship, and much more such as “sample changer jam” ...

I gratefully acknowledge to all our collaborators from Prestwick Chemical, Leiden University Medical Center and Cardiff for all the collaborative works we were all involved. I am very grateful for the opportunity I had to perform secondments at Prestwick Chemical and at Leiden University Medical Center and I take the opportunity to thank once again to both teams. I have really enjoyed the raining Strasbourg and the raining Leiden. Specially I would like to gratefully acknowledge Doctor Marie-Loiuse Jung, Doctor Jean Marie Contreras and Doctor Christophe Morice for receiving me and supervise my work at Prestwick Chemical and to Professor Doctor Martijn van Hemert and Doctor Ali Tas for receiving me and supervise my work at Leiden University Medical Center.

I would like also to express my sincere gratitude to Professor Doctor Martijn van Hemert for being my mentor and for all the follow-up, time, guidance and support that he has provided to me. Thank you for being always available to hear me!

A specially thank you to Doctor Irina Albuлесcu and my colleagues Natacha Ogando and Kristina Kovacikova for the food advices and the fun times we had in Leiden! I wish also to take this opportunity to acknowledge to my colleagues for the times we spend together in Marseille; Natacha Ogando, Birgit Zonsics, Kristina Kovacikova and Kristina Lanko. To Birgit Zonsics I would like to acknowledge for the opportunity to collaborate with her. It was a pleasure to welcome Birgit and to show to her the different techniques available in the lab.

I would like also to thank to Changqing Li for the few times we spend together and to Julie Lichiere for the initial technical support.

I thank to all my colleagues from the ANTIVIRALS network for all the nice meetings we had altogether. More importantly, I would like to specially present my gratitude with our ANTIVIRALS manager Clasien Omens for her excellence, expertise and efforts in organizing all our network meetings and industrial secondments. Her supportive, encouraging and patient attitude provided the necessary conditions for all of us to have a fruitful PhD time.

A very grateful acknowledge is here dedicated to my thesis committee for the follow up provided in the beginning of my 3rd year of thesis. I would also to express my sincere gratitude with Professor Doctor Pascal Marchot for the space in her office.

Lastly, I would like to thank to my family and friends for all their love and encouragement.

A very warm Thank You to everyone!

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

Resumé

Les alphavirus comme le virus Chikungunya (CHIKV) et le virus de l'encéphalite équine vénézuélienne (VEEV) sont des arbovirus (ré)-émergents préoccupants pour la santé publique. Ces virus transmis par des arthropodes ont un génome à ARN simple brin positif d'environ 11 à 12 kilobases comportant deux cadres ouverts de lecture (ORFs). L'ORF en 5' peut être directement traduit à partir de l'ARN génomique menant à la production des polyprotéines P123 et P1234 qui sont ensuite soumises à un clivage protéolytique pour générer les quatre protéines non structurales : nsP1, nsP2, nsP3 et nsP4. Ce manuscrit présente le travail réalisé sur nsP1 et de coiffage des ARNm d'alphavirus et la structure/fonction de nsP3.

Les alphavirus possèdent un mécanisme de coiffage de l'ARNm viral non classique catalysée par nsP1 et nsP2 qui conduit à la formation d'une structure cap-0 (m^7GpppN). La structure cap-0 est cruciale pour la réplication du virus car réduire la détection des ARN viraux étrangers, empêche la dégradation de l'ARN par exonucléase cellulaire et favorise la traduction de l'ARN viral en protéines. Par conséquent, le coiffage de l'ARNm de l'alphavirus est une cible antivirale attractive pour la conception d'inhibiteurs spécifiques.

L'enzyme nsP1 catalyse trois des quatre activités essentielles requises pour le coiffage d'ARNm viral : méthylation de GTP (MTase), guanylylation de nsP1 (GT), transfert sur le 5' l'ARNm (GTase). Un des objectifs de la thèse a consisté à développer un test de criblage permettant de détecter des inhibiteurs de nsP1, pour initier une approche antivirale. Pour cela, nous avons développé un immunotest permettant de cribler à haut débit (HT) la bibliothèque de Prestwick Chemical® contre l'activité GT de nsP1. 18 composés sont ressortis ce crible et trois séries de composés ont été sélectionnées pour une caractérisation plus poussée (Ferreira-Ramos et al., 2019). Ces composés inhibent peu une MTase cellulaire, ce qui suggère leur spécificité vis-à-vis de nsP1. Des analyses de relations structure/activité (SAR) ont également été initiées pour identifier les pharmacophores actifs. D'une manière générale, les résultats montrent que notre test basé sur l'enzyme HT constitue un moyen pratique pour sélectionner des composés spécifiques ciblant le coiffage de l'ARNm des alphavirus. Parallèlement au criblage, nous avons également testé de nouveaux analogues de la série MADTP.

L'organisation de nsP3 consiste en un *Macro* domaine au N-terminal, un domaine de liaison au zinc (ZBD) et une région hypervariable C-terminale (HVR). Le HVR joue un rôle clé dans la réplication virale en interagissant avec plusieurs partenaires parmi lesquelles des protéines à domaines SH3 ou G3BP. Le *Macro* domaine est essentiel aux

étapes précoces et tardives de la réplication par liaison à l'ADP-ribose (ADPr) et dé-ribosylation de protéines cellulaires.

Les études antérieures structure/fonction des *Macro* domaines permettent de classer les domaines *Macro* des alphavirus dans le groupe du *Macro* D2. Cependant, le mécanisme moléculaire de la dé-ribosylation par le *Macro* domaine des alphavirus reste à affiner. Afin de mieux comprendre ce mécanisme, nous avons lancé une étude structurale du *Macro* domaine du virus Getah (GETV). La séquence du *Macro* domaine du GETV montre une substitution particulière dans l'un des résidus conservés dans la boucle catalytique. Compte tenu de ces observations, nous avons produit et purifié le *Macro* domaine du GETV pour des études cristallographiques. Nous avons caractérisé plusieurs conformations adoptées par l'ADPr dans le site de fixation. Ces diverses conformations peuvent représenter plusieurs instantanés du mécanisme de l'ADP-ribosylhydrolase, mettant en évidence de nouveaux résidus à caractériser.

Mots Clés

Alphavirus, antiviraux, complexe de réplication, coiffage d'ARNm, nsP1, criblage à haut débit, nsP3, *Macro* domaines, ribosylation, ADP-ribose, cristallographie

Abstract

Alphaviruses such as Chikungunya virus (CHIKV) and Venezuelan equine encephalitis virus (VEEV) are important (re-)emerging arboviruses of public health concern. These arthropod-borne viruses carry a positive single strand RNA genome with approximately 11 to 12 kilobases and is divided into two open reading frames (ORFs). The 5' proximal ORF can be directly translated from the genomic RNA leading to the production of the P123 and P1234 polyproteins which are next subjected to proteolytic cleavage generating the four non-structural proteins: nsP1, nsP2, nsP3 and nsP4. The manuscript describes the work done on nsP1 and the alphavirus mRNA capping, and the structure/function of a domain of nsP3.

Alphaviruses own an unconventional viral mRNA capping mechanism catalysed by nsP1 and nsP2 which leads to the formation of a cap-0 structure (m^7GpppN -). The cap-0 structure is crucial for virus replication because it reduces the detection of foreign viral RNAs, prevents RNA degradation by cellular exonuclease, and promotes viral RNA translation into proteins. Therefore, the alphavirus mRNA capping is an attractive antiviral target for the design of specific inhibitors.

The nsP1 enzyme is known to catalyse three of the four crucial activities required for the viral mRNA capping, which are: N7-guanine methyltransferase (MTase), guanylylation (GT), guanylyltransferase (GTase) activities. The GT of VEEV nsP1 can be monitored by Western blot (WB) using an antibody recognizing the cap structure (Li *et al.*, 2015). Under the scope of this thesis a high throughput (HT) ELISA screening was developed by monitoring the GT reaction through the quantification of the m^7GMP -nsP1 adduct formation. The Prestwick Chemical library® was screened using this method and the IC_{50} was determined for 18 hit compounds. Three series of compounds were selected for further characterization (Ferreira-Ramos *et al.*, 2019). These compounds poorly inhibit a cellular MTase which suggests their specificity against VEEV nsP1. Analogue search and structural activity relationships (SAR) were also initiated to identify the active pharmacophore features. In general, the results show that our HT enzyme-based assay is a convenient way to select specific hit compounds targeting the viral mRNA capping of Alphaviruses. In parallel to the screening we also tested new analogues from MADTP series by WB.

The organization of nsP3 consists in a *Macro* domain at the N-terminal, a zinc binding domain (ZBD) and a C-terminal hypervariable region (HVR). The HVR plays a key role in the viral replication by interacting with several partners among which G3BP and SH3 containing proteins. The *Macro* domain is essential at both early and late steps of the

replication through ADP-ribose (ADPr) binding and de-ribosylation of cellular proteins (Abraham *et al.*, 2018).

Previous structural and sequence analysis together with functional characterization could classify alphavirus *Macro* domains into a group prototyped by *Macro* D2. However, the molecular mechanism of de-ribosylation by alphavirus *Macro* domain remains to be built up. In order to better understand this mechanism, we initiated a structure-based study of Getah virus (GETV) *Macro* domain. The sequence of GETV *Macro* domain shows a peculiar substitution in one of the conserved residues in the catalytic loop. Given these observations, we produced and purified the GETV *Macro* domain for crystallographic studies and characterized several conformations adopted by ADPr in the binding site. Together, these various conformations observed in the crystallographic structures may represent several snapshots of the ADP-ribosylhydrolase mechanism, highlighting new residues to be further characterised.

Keywords

Alphaviruses, antivirals, replication complex, mRNA capping, nsP1, guanylylation, approved drugs, HT screening, nsP3, *Macro* domains, ribosylation, ADP-ribose, deMARylation, crystallography, stress-granules, and G3BP

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

Index

ACKNOWLEDGEMENTS	III
RESUMÉ	VII
MOTS CLÉS	VIII
ABSTRACT	IX
KEYWORDS	X
INDEX	XIII
LIST OF FIGURES	XVII
LIST OF TABLES	XXXV
ABBREVIATIONS	XXXIX
I- INTRODUCTION	49
1. (RE-) EMERGING ARBOVIRUSES: ALPHAVIRUSES	51
1.1. <i>Transmission of arboviruses</i>	51
1.2. <i>Arboviruses from the Alphavirus genus</i>	54
1.3. <i>Impact and expansion of encephalitic alphaviruses in the society</i>	58
1.3.1 Epizootic circulation of VEEV	59
1.3.2 Enzootic circulation of VEEV	61
1.4. <i>Impact of the global expansion of CHIKV in the society</i>	66
1.4.1 Chikungunya West African (WA) lineage	69
1.4.2 Chikungunya East/Central/South Africa (ECSA) lineage	69
1.4.3 Chikungunya Indian Ocean sub-lineage (IOL)	71
1.4.4 Chikungunya Asian lineage	73
2. PREVENTION METHODS, CURRENT DIAGNOSTIC, TREATMENTS TO FIGHT ALPHAVIRUSES AND THE LACK OF ANTI-ALPHAVIRUS DRUGS.	76
2.1. <i>Diagnostic of Alphavirus infection</i>	79
2.2. <i>Treatments available and the lack of anti-alphavirus drugs</i>	81
3. WHAT ARE ALPHAVIRUSES?	89
4. THE ALPHAVIRUS RNA SYNTHESIS AND PROTEIN TRANSLATION	92
4.1. <i>Translation of structural proteins</i>	98
4.2. <i>Alphavirus assembly and budding</i>	104
4.3. <i>Translation of Non-structural proteins</i>	108
4.3.1 The alphavirus non-structural protein 1 (nsp1)	111
4.3.2 The alphavirus non-structural protein 2 (nsp2)	113
4.3.3 The alphavirus non-structural protein 3 (nsp3)	117
4.3.4 The alphavirus non-structural protein 4 (nsp4)	124

4.4.	<i>Assembly of the replication complex</i>	126
5.	RNA CAPPING PATHWAYS	131
5.1.	<i>Canonical cap (m⁷G) formation through unconventional pathways in viruses:</i>	136
5.1.1	Unconventional capping pathway through “Cap snatching”	136
5.1.2	Unconventional capping pathway with guanosine diphosphate (GDP)	138
5.1.3	Unconventional capping pathway with m ⁷ GMP	139
5.2.	<i>The discovery of non-canonical cap structures</i>	140
5.3.	<i>Unconventional RNA capping pathway in alphaviruses</i>	141
5.4.	<i>The connection of viral RNA capping and the innate immune response</i>	147
6.	INSIGHTS INTO <i>MACRO</i> DOMAINS: A KEY MODULE PARTICIPATING IN METABOLIC PATHWAYS REQUIRING ADP-RIBOSE (ADPR).	153
6.1.	<i>Structure and function of MacroD-type</i>	158
6.2.	<i>The connection of MacroD-type and the innate immune response</i>	161
7.	OUTLINE OF THIS THESIS	172
II-	RESULTS	176
8.	ANTIVIRALS TARGETING THE NON-STRUCTURAL PROTEIN 1 (NSP1) FROM ALPHAVIRUSES	178
8.1	<i>Approved drugs screening against the nsP1 capping enzyme of Venezuelan equine encephalitis virus using an immuno-based assay</i>	180
8.2	<i>Additional Results: Hit/series validation beyond the screened library against the VEEV nsP1 capping enzyme and pharmacophore evaluation of Prest-1291</i>	196
8.3	<i>Additional Results: Enzymatic screening to find more potent MADTP compounds targeting the nsP1 of VEEV</i>	206
8.4	<i>Discussion</i>	212
8.5	<i>Conclusions</i>	220
9.	UNDERSTANDING THE CATALYTIC MECHANISM BEHIND <i>ALPHAVIRUS MACRO</i> DOMAINS	224
9.1	<i>Crystallographic structures of Getah virus Macro domain reveal multiple conformation of ADP-ribose in the binding pocket</i>	226
9.2.	<i>Additional results: Evaluation of the effect of the viral Macro domains catalysis in the cell using cell-based assays</i>	266
9.3	<i>Discussion</i>	294
9.4	<i>Conclusion</i>	304
III-	GENERAL DISCUSSION	310
V-	GENERAL CONCLUSIONS	322
VI-	ANNEX	328
10.	ADDITIONAL COLLABORATION	330
10.1.	<i>Computer-aided design, synthesis and evaluation of novel compounds against Chikungunya nsP3</i>	332

THIS PAGE IS INTENTIONALLY LEFT BLANK

List of figures

Figure 1. Vertical transmission of mosquito-borne arboviruses. Mosquitoes such as *Aedes albopictus* or *Aedes aegypti* are endemic of inhabited densely vegetated rural areas and they had developed adaptability to colonize new habitats in urban and suburban areas. **(1)** Usually female mosquito needs to intake a blood meal to sustain their egg production. When feeding from an (arbovirus)-infected mammal, the arbovirus can be transmitted to the off-spring of this female mosquito through trans-egg transmission. **(2)** In urban and suburban areas, female mosquitoes utilize human-made water holding containers such as flowerpots, soda cans, buckets and tires that can collect and retain rainwater for a long period of time. During oviposition, females laid their eggs usually slightly above the water level. **(3)** Their eggs are resistant to drying out and when covered with water the larvae emerge from the egg and undergoes a process of development through four stages and feed primarily from organic material present in the water. **(4)** After completing the four stage the larvae enter the pupal stage. **(5)** The pupal stage last for a few days until the adult mosquito emerges and seeks a resting site in low vegetation. The female mosquitoes are then ready to start a new cycle and the male mosquitoes in some cases can transmit the arbovirus to an uninfected female during mating, through trans-ovarial transmission.

Figure 2. Transmission mode of mosquito-borne arboviruses. Most arboviruses normally circulate among non- human animals and can occasionally be transmitted to humans (Zoonotic pathogens). The transmission mode of alphaviruses usually requires alternating replication cycles between vertebrate and arthropod. The reservoir of these zoonotic pathogens are usually birds, rodents and non- human primates which are involved in the perpetuation of the enzootic cycle. Mosquito vectors such as species from the genus *Culex*, *Ochlerotatus* and *Psorophora* are involved in the transmission between vertebrate hosts. Usually the enzootic cycle of alphaviruses occurs in sylvatic habitats **(1)**. Enzootic and bridge- vectors might be involved in the indirect transmission to humans through the rural epizootic cycle **(1 → 2 → 3)**. The rural epizootic cycle **(3)** is associated with domestic animals and mosquitoes where the virus amplifies in the presence of intermediate hosts. Thus, **(3)** represents a large reservoir of viruses from where can result a severe spill over effect to dead-end hosts. Human infection and epidemics can arise when humans enter in the sylvatic enzootic habitats **(1 → 4)** or when the amplification of the virus occurs in domestic animals within a rural epizootic cycle (such as equines and pigs) followed by transmission to humans **(3 → 4)**. Adaptation of

arboviruses to the epizootic mosquito vectors and to the death- end hosts seems to be associated with several outbreaks. The movement of humans between rural and urban habitats are determinant for the initiation of urban epidemic cycles (5) which are dependent on the adaptation of the virus to the arthropod species present in the urban areas. Adapted from (Weaver & Barrett, 2004).

Figure 3. Phylogenetic tree constructed using the maximum-likelihood method in MEGA, version 6.0 with 1,000 bootstrap replicates based on multiple alignment of nucleotide sequences of the structural proteins. Bootstrap values greater than 50 are shown near the branch nodes and the scale bar indicates the number of substitutions per site. Adapted from Torii *et al.* (Torii *et al.*, 2018).

Figure 4. Locations of Venezuelan equine encephalitis virus outbreaks in the Americas (Weaver *et al.*, 2004). The location of all major VEE outbreaks in the Americas are highlighted in the map (regions shaded purple and labelled with text). The date (year) and respective VEE virus subtypes of the outbreak are shown. Symbols represent locations from which enzootic VEEV-complex virus have been isolated, with enzootic subtypes indicated in parentheses.

Figure 5. Origin, spread, and global distribution of CHIKV outbreaks. Country colours correspond to the decade of the first reported identification of the local transmission of CHIKV by either serological, molecular, or virological detection methods. (Image adapted from Nicole Lindsey in Powers 2018 (Powers, 2018)). The map shows the African origins of enzootic chikungunya virus strains and the patterns of emergence and spread of the Asian lineage and Indian Ocean lineage (IOL) of the virus during epidemics since the 1950s, based on phylogenetic studies. ECSA denotes East/ Central/ South Africa lineage. Adapted from centers for disease control and prevention (CDC) and (Weaver & Lecuit, 2015, Powers, 2018).

Figure 6. General schematic representation of an Alphavirus virion. Spherical particles with approximately 70 nm in diameter, containing an envelope made of the host lipid bilayer membrane and containing 80 spikes, each spike are trimer of E1:E2 heterodimer proteins. The central core contains an icosahedral capsid with T=4 icosahedral symmetry and is made of 240 monomers. The capsid with a diameter approximately between 30 to 40 nm is surrounded by the envelope structure and each heterodimer E2-E3 is in contact with the capsid. The green dot lines in form of triangle

highlights the icosahedral nucleocapsid symmetry $T=4$. Adapted from (Cheng *et al.*, 1995).

Figure 7. Three-dimensional (3D) reconstruction of two alphaviruses by cryo-EM analysis and image processing. (A) 3D surface structure of RRV virion (diameter ~70nm) viewed along an icosahedral threefold axis. The spikes trimers of E1-E2 heterodimers located at the threefold and quasi-threefold axes, have a flower-like head with three bilobal petals. The spikes are engaged in extensive lateral interactions close to the lipid bilayer via their skirts. These parts of the spikes are colored bluish. The lipid bilayer (yellow) is seen through openings in the spike-skirt protein layer at the twofold and fivefold symmetry axes. Adapted from (Cheng *et al.*, 1995) (B) Schematic representation (using RRV) of the interactions between the spikes (green) at the threefold (circled 3) or quasi-threefold (circled Q3) axes and the C molecules (yellow) of the capsomers in the underlying NC. Adapted from (Cheng *et al.*, 1995). (C) VEEV (TC-83) cryo-EM slice through the 3D density map 20 pixels from the origin Adapted from (Zhang *et al.*, 2011). (D) One asymmetric unit of VEEV containing four unique copies of E1 (magenta), E2 (cyan), E3 (orange) and CP (blue). The cryo-EM densities for the viral membrane (yellow) and genomic RNA (green) are also displayed at slightly lower isosurface threshold. Scale bar: 2 nm. (Zhang *et al.*, 2011) (E) Radially coloured 3D reconstruction of VEEV, showing the E1 basal triangle (green) and E2 central protrusion (blue) for each spike. Scale bar: 10 nm. (Zhang *et al.*, 2011)

Figure 8. General schematic description of Alphavirus RNA genome. The *Alphavirus* RNA genome is capped with a cap-0 structure at the 5'-end and polyadenylated at the 3'-end. Indicated are the two coding regions (ORF1 and ORF2), the three untranslated regions (UTRs) several cis-acting regulatory elements (CREs), for example CSE1, CSE2, CSE3, CSE4, UREs and RESs, the genomic and sub-genomic promoters, the location of the OPAL codon and the frame shift sequence. The regions coding for non-structural proteins (nsP1, nsP2, nsP3 and nsP4) and for structural proteins (Cp, E3, E2, 6K, TF and E1) are highlighted. For details see text.

Figure 9. Diagram representing the major steps during translation of non-structural and structural proteins. The non-structural polyproteins P123 and P1234 are directly translated from the viral genomic RNA previously released into the cytosol of the host cell. The two polyproteins are subject of proteolytic processing and drives the formation of the early replication complexes. The early replication complexes are known to synthesise the minus strand RNA and the complete maturation of the four non-

structural proteins (nsPs) leads to the switch to the synthesis of RNA of positive polarity (viral genomic and sub-genomic RNAs). The viral sub-genomic RNA is then used as a template for the translation of the structural polyproteins. A trans-frame region localized in the beginning of the region coding for the E1 protein is responsible for the translation of the two structural polyprotein Cp-E3-E2-TF and Cp-E3-E2-6K-E1. The two structural polyproteins are then proteolytic processed by the capsid and furin proteases and the signal peptidase. Thus, leading to the formation of the mature structural proteins, Cp, E3, E2, 6K, TF and E1.

Figure 10. Structural polyprotein membrane topology and processing. Pre-cleavage of structural polyprotein and conformation in the endoplasmic reticulum (ER) membrane. Capsid (CP) is first released as a soluble protein in the cytoplasm. The N-terminus of E3 contains a signal sequence that directs translocation across the membrane. Open arrows mark the furin cleavage site between E3 and E2, which occurs in the trans-Golgi network. The remaining envelope proteins are threaded through the membrane with insertion of transmembrane domains. Closed arrows mark signalase cleavage sites that are used in the ER. Post-cleavage conformations include palmitoylation sites (yellow). **(A)** The 6K polyprotein is the majority translation product. **(B)** The TF form of the polyprotein is the minority product resulting from frameshifting in the 6K gene and does not include E1. (Ramsey & Mukhopadhyay, 2017)

Figure 11. Model representing one hypothesis of the current understanding of TF and 6K. **(A)** TF is less abundant than 6K. It is palmitoylated and traffics from the ER to the plasma membrane (PM) where it is budded into virions. TF may oligomerize and function as an ion channel. **(B)** 6K is found concentrated at interior membranes where it likely interacts with the viral glycoproteins. The 6K channel properties resulting from oligomerization probably affects the ER. Additionally, 6K may interact with host proteins. (Ramsey & Mukhopadhyay, 2017)

Figure 12. Structural organization and function of *Alphavirus* glycoproteins. **(A)** The fusion proteins of an alphavirus are connected to the viral membrane and can consist of three components: E1, which contains domains I, II and III, and a fusion loop (star); E2, which contains domains A, B and C, and a ribbon-like connector region; and E3. The alphavirus envelope is peppered with trimers of E2-E1 pairs, although for simplicity only one E2-E1 pair is shown. E3 is a by-product of the cleavage of the E2 precursor protein p62 by the enzyme furin (not shown) and is released in some alphavirus species. **(B)**, On exposure to low pH, domain B and the connector region of

E2 move, exposing the E1 fusion loop. **(C)** E1 can then detach from E2 and insert into the host cell membrane to form an extended trimer. **(D)** Finally, to drive membrane fusion, E1 refolds into a hairpin-like structure through the movement of its domain III and the stem region, which lies adjacent to the membrane. For simplicity, only the final fused membrane is shown. (Kielian, 2010)

Figure 13. Schematic model of specialized budding sites of an alphavirus infected cell (gray). Alphaviruses assemble and bud from localized patches at the plasma membrane of the cell body, where host proteins are excluded, and Cp/NC serves as a scaffold to accumulate E2/E1 glycoproteins. Dramatic cytoskeletal remodelling occurs during infection, and short filopodia-like extensions (top) and long intercellular extensions (middle) are observed. Short extensions contain F-actin (yellow) only, while intercellular extensions contain F-actin and tubulin (blue) structures. Nascent virus particles bud along the short extension, which may facilitate dissemination from the infected cell. An intercellular extension projects from the infected cell and makes preferential contact with an uninfected cell (green). The contact site forms a stable, flattened tip on the target cell plasma membrane (right), producing a protected pocket where nascent particles bud and are subsequently internalized by the uninfected cell. Thus, intercellular extensions are close-ended membrane bridges that mediate cell-to-cell transmission. Host determinants involved in modulating the cytoskeleton, forming/stabilizing extensions, and promoting cell-to-cell virus transmission are currently unknown. (Brown *et al.*, 2018).

Figure 14. The major steps during the formation of the *Alphavirus* replication complex. The *ORF1* from the genomic RNA belonging to alphaviruses is directly translated into two non-structural polyproteins. The P1234 represents between 5 to 20 % of the translation and the P123 is the major product of translation representing more than 80 % of translation of the *ORF1*. The regulation of the translation of both polyproteins is achieved due to the presence of an OPAL codon localized at the end of the coding region for the non-structural protein 3 (nsP3). The polyproteins are further processed by proteolytic cleavage which is mediated by non-structural protein 2 (nsP2). The replication complex is assembled involving the viral non-structural polyproteins/proteins and host factors **(1)** The first cleavage occurs at the junction nsP3/nsP4, thus releasing the mature nsP4 which together with P123 forms the early replication complex. This early replication complex synthesizes mainly negative stranded RNA. **(2)** the maturation of the non-structural proteins proceeds and the junction between nsP1/nsP2 is cleaved and this cleavage is responsible for the switch in the syntheses from negative to positive polarity RNA. **(3)** Fully matured non-structural proteins (nsPs)

occurs after the last cleavage between the junction nsP2/nsP3 and forms the late replication complex that uses the negative stranded RNA as a template to synthesize the RNA of positive polarity which are the sub-genomic RNA and genomic RNA.

Figure 15. Domain organization of the matured non-structural protein 1 (nsP1).

The purple domain represents the domain with a Rossmann fold and that harbours the activities of methyltransferase (MTase) and guanylyltransferase (GTase). The His37 represents the histidine residue determined to bind covalently to the m⁷GMP moiety. The membrane-binding amphipathic helix are represented in the yellow rectangle and the palmitoylation sites are represented with text and yellow lines. Both membrane-binding amphipathic helix and palmitoylation site are membrane anchor sites. Amino acid numbers are indicated for VEEV nsP1 strain TC83. For more details see text. Adapted from (Rupp *et al.*, 2015)

Figure 16. The anti-host effect of nsP2 during alphaviruses infection.

The nsP2 is involved in host cell translation shut-off because inhibits the binding of the complex STAT1/STAT2/IRF9 complex to the interferon stimulated responsive element (ISRE) which is activated by the type I IFN through the JAK-STAT pathway. Therefore, nsP2 inhibits the expression of interferon stimulated genes (ISGs) such as *PARPs* including *PARP1* and *ZAP*, *IFIT* family, *TRIM* family, *OAS* which activates RNase L, *ATF4*, *XBP1* and *UPR* genes. Indirectly, nsP2 inhibits the activation of PKR kinase and this leads to protein translation inhibition. The activation of the caspase cascade is also affected, and this can compromise the maturation of pro-IL1 β into IL1 β . The dysregulation of *PARP* genes expression and alterations in the normal levels of host protein translation can induce the formation of stress granules. The stress granules are important structures where the inactive mTORC1 complex is stored when host cell is under stress. The mTORC1 is inactivated by AMPK and activation of AMPK requires PARP1. Indirectly, the decreased expression of PARP1 may lead to less inhibition of mTORC1. Another line to counteracts the host cell response by nsP2 is the induction of degradation mediated by proteasome of the subunit Rpb1 that belongs to the RNA polymerase II. For more details see text. Figure designed using <https://biorender.com/>.

Figure 17. Domain organization of the mature non-structural protein 2 (nsP2).

Organized in four major domains, nsP2 harbours an N-terminal domain, a helicase domain, a protease domain and an inactive MTase-like domain at the C-terminal portion of nsP2. The N-terminal domain is divided in D1 and D2 domains, the first with co-factor properties and the second likely involved in promoter selection. The helicase domain

also is composed of two domains, Rec 1-like 1A and Rec1-like 2A. the helicase domain also presents NTPase activity. The NLS1 and NLS2 are the sites of canonical nuclear localization sites. The amino acid range is indicated for VEEV strain TC83. For more details see text. Adapted from (Rupp *et al.*, 2015).

Figure 18. Domain organization of the non-structural protein 3 (nsP3). At the N-terminal is localized the *Macro* domain, after follows the Zinc-binding domain and the Hypervariable region. The last is localized at the C-terminal portion of nsP3 and is highly disordered and displays a multifunctional role during infection. The *Macro* domain is involved in removal of ADP-ribose from post-translationally modified proteins with mono- or poly- ADP-ribose. The Zinc-binding domain is known to coordinate a zinc ion, but its function is still unknown. Concerning the hypervariable region, a more complex role is in charge of this domain. In the diagram is represented 3 important sites of the hypervariable region. The YXXM motif which is required for activation of the PI3K through binding to the p85 subunit and that triggers the activation of the pathway that leads to activation of the mTORC1. The P(I/V)(P/A)PPR motif which is involved in binding to amphiphysin-2. Lastly, the FGDF motif (or FXR- binding motif) that is required to bind G3BP1/2 or FXR proteins respectively. The FGDF motif or FXR-binding motif is dependent on OW or NW viruses respectively. The interactions of nsP3 with p85, G3BP1/2 or FXR proteins are important for stress granules disassembly mediated by alphaviruses. The role of the *Macro* domain is still under investigation, but it is possible that is required for disassembly of stress granules. The amino acid range indicated with text is relative to the VEEV strain TC-83. Figure designed using <https://biorender.com/>.

Figure 19. The domain organization of the non-structural protein 4 (nsP4). The short N-terminal domain of the nsP4 is highly disordered and it is a putative scaffold for interaction with polyprotein P123 and the mature nsPs to form the replication complexes. The remaining (large) portion of nsP4 is the RNA-dependent RNA polymerase (RdRP) domain which is required for synthesis of the *Alphavirus* RNA and terminal divalent cation-dependent adenylyltransferase (TATase) activity. The GDD represented under a yellow line represents the approximated location of the Gly-Asp-Asp catalytic motif of the RdRp domain of nsP4. The amino acid range corresponds to the nsP4 from VEEV strain TC83. For more detail see text. Adapted from (Rupp *et al.*, 2015).

Figure 20. Assembly of the replication complex. The assembly of the replication complex occurs after the cleavage of nsP3/nsP4 junction. **(1)** The mature nsP4 together with P123 forms the early replication complex which is required for synthesizes of

negative stranded viral RNA that remains double stranded. **(2)** The early replication complex goes through a process of maturation involving the proteolytic cleavage of the remaining polyproteins. The cleavage of the junction nsP1/nsP2 is responsible for the switch in the syntheses from negative to positive polarity viral RNA. This early intermediate replication complex is still able to synthesise low levels of RNA of negative polarity and it starts to use the RNA of negative polarity as a template for the synthesis of the positive polarity RNA which are the sub-genomic and genomic RNAs. **(3)** Once fully matured, after total cleavage and maturation of non-structural proteins (nsPs), is then formed the late replication complex that only uses the negative stranded RNA as a template to synthesize the viral sub-genomic and genomic RNAs. See text for details.

Figure 21. The multifunctional role of nsP3 during Alphavirus replication. The late replication complex formed by the mature non-structural proteins have a key role in the synthesis of genomic and sub-genomic viral RNA. NsP3 is a multifunctional protein contained within the late replication complex that is assembled at the PM. It has been proposed that nsP3 contains a sequence motif, YXXM, that binds to the SH2 domain of the regulatory subunit of PI3K (the p85 subunit). Through interaction of this sequence motif with the SH2 domain of p85, the p110 subunit of PI3K is activated and converts the membrane bounded $PI(4,5)P_2$ to $PI(3,4,5)P_3$. The presence of $PI(3,4,5)P_3$ at the membrane recruits the PKD1 and AKT to the PM and PKD1 through phosphorylation activates AKT. The phosphorylation of AKT is also enhanced in the presence of mTOC2 complex. Interestingly, it has been reported that activation of AKT is correlated with translocation of the spherule containing the replication complex from the PM to cytopathic vacuoles I (CPV-I) that are found at the cytoplasm, in a process named as translocation of the spherule. In addition, the activated AKT triggers the inhibition of TSC1/2 complex which leads to activation of Rheb that converts GDP into GTP and leads to activation of the mTORC1 complex. The mTORC1 complex also induce phosphorylation of downstream targets to ensure efficient host cellular translation. The nsP3 also participates in binding and recruiting FXR proteins (not shown) or G3BP1/2 to the replication complex where occurs synthesis of viral RNA. It has been suggested that G3BP1/2 also participates in binding to the viral RNA near the spherule structures and the place where encapsidation of viral genomic RNA take place. For more details see text. Figure designed using <https://biorender.com/>.

Figure 22. Metal-independent triphosphatase activity that occurs during formation of the cap-0 structure in human cells. An enzyme with RNA triphosphatase activity (RTPase) produce a mRNA with a diphosphate 5'-end. Through a conserved cysteine

present in the conserved phosphate binding loop attacks the γ -phosphorus of triphosphate terminal RNA to form a covalent protein-cysteinyl-S-phosphate intermediate. The diphosphate RNA is later released and in a final step the inorganic phosphate is released after hydrolysis of the covalent bond between the phosphate and the enzyme. For more details see text. Adapted from (Ghosh & Lima, 2010).

Figure 23. Capping in human cells require guanylyltransferase activities to synthesise the cap-0 structure. An enzyme presenting RNA guanylyltransferase catalyzes the capping in a two-step reaction. First the enzyme binds GTP (colored blue) to catalyse the transfer of GMP to the active site lysine to form a covalent enzyme(lysyl-N)-GMP intermediate. Secondly, catalyse the transfer of the GMP to a diphosphate RNA end to form GpppRNA. For more details see text. Adapted from (Ghosh & Lima, 2010).

Figure 24. Methyltransferase activities during formation of the cap-0 structures during capping in human cells. The methylated cap structure (cap-0 structure) is produced co-transcriptionally and requires the enzymatic activity of the RNA guanine-N7 methyltransferase. The RNA huanine-N7 methyltransferase binds S-adenosylmethionine (SAM) (colored green) and GpppRNA (colored as in figure 23) and catalyze the transfer of the methyl group (colored green) to the guanine N7 position. For additional information see text. Adapted from (Ghosh & Lima, 2010).

Figure 25. Structures and synthesis of RNA caps. The cap-0 structure is then further methylated to originate the TMG cap, Cap-1 and cap-2. The TMG cap is formed through the transfer of two methyl groups (colored magenta) from S-adenosylmethionine (SAM). Concerning caps-1 and cap-2 structures, it is required methylation (colored magenta) of cap 0 at the ribose 2'-O-hydroxyl at the first and second nucleosides, respectively. For more detail see text. Adapted from (Ghosh & Lima, 2010).

Figure 26. Cap-snatching transcription mechanism of influenza polymerase. The PA-PB1-PB2 complex is localized in the nucleus of the infected cell. During transcription, the PB2 subunit binds the 5',7-methylguanosine cap of a host pre-mRNA molecule (red), which is subsequently cleaved 10–15 nucleotides downstream by the PA endonuclease. The resulting short capped RNA primer is used to initiate polymerization by the RNA-dependent RNA polymerase of the PB1 subunit using 5'- and 3'-bound vRNA (green) as template, resulting in capped, polyadenylated, chimeric mRNA molecules (red and blue) that are exported to the cytoplasm for translation into viral proteins. Adapted from (Boivin *et al.*, 2010).

Figure 27. Domain organization of VSV-L. The polymerase domain (RdRp) is in cyan; capping domain (Cap), green; connector domain (CD), yellow; methyl transferase (MT), orange; C-terminal domain (CTD), red. Amino acid residue numbers indicate functional domain boundaries. Flexible linkers 1 and 2 connect Cap to CD and CD to MT domain, respectively. Conserved regions within L proteins of non-segmented negative-strand (NNS) RNA viruses are labelled CR I – VI. Asterisks indicate the position of active site residues. For details see text. Adapted from (Liang *et al.*, 2015).

Figure 28. A proposed model of the polyribonucleotidyl transfer reaction catalysed by the unconventional mRNA capping enzyme L protein. For detail, see text. (Ogino *et al.*, 2010)

Figure 29. Current understanding of *Alphavirus* mRNA capping. The mRNA of alphaviruses is capped at the 5'-end with a type 0 cap by an unconventional mechanism. The Alphavirus capping enzyme is the non-structural protein 1 (nsP1) that presents the three enzymatic activities required for capping for the viral mRNA (genomic and sub-genomic). For the first step of the reaction is required the methyltransferase activity of the nsP1 enzyme. The first step consists in the methylation of a GTP molecule through the transfer of a methyl group from the donor S-adenosyl methionine (SAM) to the N7 position a molecule of GTP and generates S-adenosyl-L-homocysteine as a by-product of the reaction. A second step of the cap-0 formation is the attachment of a m⁷Gp (m⁷GMP) moiety to a diphosphate viral mRNA that is previously generated by nsP2 that presents 5' RNA triphosphatase (RTPase) activity. This step requires the guanylyltransferase activity (GTase activity) of nsP1 that involves a highly conserved histidine of nsP1 (histidine 37 in VEEV strain TC83). The GTase activity is a reaction that requires two steps, the first is the guanylylation (GT) of the nsP1 enzyme and the second is the transfer of the m⁷GMP moiety to the diphosphate end of the viral mRNA (transferase reaction). The conserved histidine attacks the α -phosphorus of the m⁷GTP and breaks the phospho-anhydride phosphoanhydride bond between the β and the α phospho-groups. During this step pyrophosphate is released and the nsP1-(histidyl-N)-m⁷GMP adduct is formed between the N ϵ (or possibly N δ) of the histidine and the α -phosphate of the m⁷GMP. For more details see text.

Figure 30. Activation of the innate immune response through sensing of the viral RNA. The RNA from alphaviruses protected with a type 0 cap structure or non-protected (5' triphosphate ends) can trigger the activation of the innate immune sensors such as RIG-I, MDA5, TLR3, TLR7 and TLR8. RIG-I and MDA5 sensors which are knowing to

activate the mitochondrial-anchored protein, the signalling adaptor IFN β promoter stimulator 1 (IPS1) which is also known as MAVS. At least two signalling cascades are activated by MAVS. One is the same cascade that is also triggered by TLR 3, 7 and 8 that leads to the phosphorylation of IRF3 and IRF7 that will dimerise and after translocation to the nucleus bind positive regulatory domains (PRD) II and III. Therefore, leading to the transcription of type I IFN and inflammatory cytokines such as pro-IL1 β . By another hand, MAVS also activates the TRAF6 signalling cascade. In order to activate the transforming growth factor β -activated kinase (TAK1) complex, an E1 ubiquitin activating enzyme (E1) is recruited to the TRAF6/TAK1/TAB2 complex together with a ubiquitin conjugating enzyme complex (E2) formed by Ubc13 and Uev1A. The TRAF6, an ubiquitin ligase, is ubiquitinated and activates TAK1 complex. Again, two possible pathways are triggered by TAK1 complex, one is the NF κ B pathway and the other is the MAP3Ks cascade. The I κ B α / β / γ complex is stabilized by NEMO in its ribosylated form. However, upon ubiquitination of NEMO, NEMO is degraded by proteasome mediated proteolysis and the I κ B α / β / γ complex is activated and leads to activation of NF κ B by directing for degradation the regulatory subunit (I κ B) of NF κ B and allowing the translocation to the nucleus of the p65-p50 of NF κ B. The p65-p50 of NF κ B usually binds to PRD I and allows the activation of antiviral genes. The activation of the MAP3Ks by the TAK1 complex also leads to downstream events of phosphorylation that will lead to the activation of several transcription factors belonging to the families of cJUN, cFOS and ATF2 and involved in the transcription of antiviral genes.

Importantly, the detection of viral RNA leads to the production of type I IFN that through an autocrine and paracrine (and eventually endocrine) leads to the activation of the janus kinase-signal transducer and activator of transcription (JAK-STAT) signalling cascade to induce the expression of IFN-stimulated genes (ISGs). Among several ISGs are the interferon induced proteins with tetratricopeptide repeats (IFIT proteins) that recognise 5' triphosphate RNA and RNA bearing a cap-0 structure. However, alphaviruses have evolved in a way to escape to IFIT proteins because alphaviruses contain a special RNA secondary structure near the cap-0 structure and that blocks the recognition of the cap (or triphosphate ends) by IFIT proteins. Other ISG produced by the JAK-STAT signalling cascade is the protein kinase PKR that can be activated by viral RNA or through components of the JAK-STAT pathway and leads to phosphorylation of eIF2 α and inhibits host protein translation. Another one is the 2',5'oligo(A) oligonucleotides synthesized by oligoadenylyl synthases (OASs) that activates RNase L which is involved in degradation of viral RNA (capped and non-capped). See text for more details. Figure designed using <https://biorender.com/>.

Figure 31. Mono- and poly- ADP ribosylation and reversible post-translational modification. (A) Mono- or poly- (ADP-ribose) is covalently attached mainly to aspartic acid or glutamic acid of acceptor proteins and the ADPr units of the polymer are connected linearly or in a branched fashion (Leung, 2014). (B) The metabolism of poly-(ADP-ribose) requires PARPs (poly-[ADP-ribose] polymerases) for biosynthesis of poly-(ADP-ribose) from NAD⁺ while PARG (poly-[ADP-ribose] glycohydrolase) degrades polymer of poly-(ADP-ribose) to ADP-ribose. For more details see text. Adapted from (Tan *et al.*, 2012). Figure 30 B was designed using <https://biorender.com/>.

Figure 32. Model for genetic conflict involving PARP Macro domains. (A) Model for Macro PARP function. ADP-ribosylated host or viral proteins may be a signal for recruitment of PARP9, 14 or 15, which could facilitate additional recruitment of antiviral effectors, and amplify the initial ADPr signal. (B–D) Three models for how viruses may antagonize Macro PARP function. Viruses lacking their own Macro domains may use other proteins to directly antagonize Macro PARP proteins (B), driving recurrent positive selection in Macro PARP genes to escape antagonism. Macro domains encoded by viruses (e.g. corona- and togaviruses) may catalyze the removal of ADPr (C) or compete with Macro PARPs for binding to ADPr (D) in order to antagonize host ADPr-mediated signalling (Daugherty *et al.*, 2014).

Figure 33. Involvement of PARPs during the innate antiviral response. The RNA from alphaviruses protected with a type 0 cap structure or non-protected (5' triphosphate ends) can trigger the activation of the innate immune sensors such as RIG-I, MDA5, TLR3, TLR7 and TLR8. RIG-I and MDA5 sensors which are known to activate the mitochondrial-anchored protein, the signalling adaptor IFN β promoter stimulator 1 (IPS1) which is also known as MAVS. The signalling cascades activated were already described in figure 30 and are here presented again. The major new representation in this figure are the involvement of PARP proteins during the antiviral response. PARP-10 is known to be involved in ribosylation of NEMO which confers more stability to the I κ B α / β / γ complex in its inactive state. Most likely the effect of PARP10 can be counterbalanced by PARG. Therefore, the translocation of NF κ B to the nucleus can be modulated by PARP10. In addition, PARP1 seems to be involved in regulation of active NF κ B binding to the positive regulatory domains (PRD) I. Other transcription factors may be regulated in a similar fashion. For example, it has been described that PARP1 can form stable complexes with NF κ B, Oct-1, TEF-1, B-Myb, ATM and AP-1. Although PARP1 is highly present at the nucleus playing several pathways modulations, it has been described that PARP1 is also required for inactivation of mTORC1. To inactivate mTORC1, PARP1

activates AMPK that inhibits mTORC1. The inhibition of mTORC1 leads to its relocation inside stress granules under stress conditions. Importantly, alphaviruses developed original strategies to fight against the innate immune system. One of such is nsP2 that can inactivate the JAK-STAT signalling cascade as previously mentioned. However, another strategy involves the multifunctional role of nsP3. nsP3 has been associated with disassembly of stress granules and it has been associated with activation of mTORC1 complex as well. In a certain way it may be involved in reverse the post-translational modifications of mono- or poly- ADP ribosylation that are carried out by several PARP proteins. See text for more details. The figure was designed using <https://biorender.com/>.

Figure 34. Implication of autophagy in DNA damage repair. Endogenous (e.g., dysfunctional mitochondria, top right) or exogenous (e.g., radiations or genotoxic stimuli, bottom left) sources of ROS and RNS induce DNA damage, whose primary sensors are PARP1 and ATM. Once activated by DNA breaks, PARP1 catalyses poly-ADP ribosylation of itself, as well as of other nuclear proteins, thereby leading to a massive decrease of NAD⁺ and to a subsequent energetic stress. Upon DNA damage, ATM can activate p53-mediated transcription of autophagic genes. Alternatively, cytosolic pool of ATM could be directly activated by ROS through a still unidentified mechanism and it directly induces the activation of LKB1. The issue of whether cytosolic and nuclear pool of ATM are interconnected still waits to be demonstrated. Both PARP1 and ATM signalling pathways converge on AMPK, whose activation induces the autophagic machinery to remove the main source of DNA damage and contribute to its repair through a negative feedback loop. From (Filomeni *et al.*, 2015).

Figure 35. Stress granules formation and composition and the role of nsP3 in stress granules disassembly. Through its YXXM motif, nsP3 interacts with the SH2 domain from PI3K from the p85 subunit. The p85 subunit is the regulatory domain of the PI3K that activates the effector domain p110. Dissociation of the p110 from p85 leads to its activation and converts the membrane bounded PI(4,5)P₂ to PI(3,4,5)P₃. PKD1 and AKT are recruited to the PM because they harbour a pleckstrin homology domain (PH domain) that binds to PI(3,4,5)P₃. The formation of PKD1/AKT complex leads to activation of AKT through phosphorylation Thr308 and Ser473. The mTORC2 can may exercise some influence in the AKT activation. In addition, the activated AKT triggers the inhibition of TSC1/2 complex which leads to activation of Rheb that converts GDP into GTP and leads to activation of the mTORC1 complex. The mTORC1 complex also induce phosphorylation of downstream targets to ensure efficient host cellular

translation. Therefore, nsP3 may have a key role in inhibiting the translocation of inactive mTORC1 to the stress granules (Betz & Hall, 2013). Indeed, nsP3 have been associated with stress granules disassembly and to interact with G3BP1/2 or FXR proteins. Interestingly, under stress conditions G3BP1/2 is located inside stress granules in complex with caprin1, PKR, eIF4E and eIF4G1/2 and this complex ensure the inactive form of PKR. However, in certain conditions, an unknown host factor activates PKR and induces the formation of large stress granules. (Reineke *et al.*, 2015, Leung, 2014). The Figure was designed using <https://biorender.com/>.

Figure 36. Sequence alignment for GETV Macro domain against other viral Macro domains. The OW group from the genus *Alphavirus* is represented in the alignment by CHIKV, MAYV and SINV Macro domains, while NW group is represented by VEEV Macro domain. The human Coronavirus NL63 Macro domain as well as SARS Coronavirus are represented in the alignment. The alignment picture was obtained using ESPript server and residues highlighted with red boxes are strictly conserved, red residues represent similarity in a group, and blue frame represent similarity across groups. Secondary structure elements from the GETV Macro domain crystal structure obtained is represented above the alignment.

Figure 37. Effect of ADP-ribose concentrations on the thermostability of GETV Macro domain. Several concentrations of ADP-ribose at the μM range, presented in x axis, were tested on the thermostability of GETV Macro domain. For each concentration of ADP-ribose titrated onto the GETV Macro domain the melting temperature was calculated (T_m). The changes in T_m respective to the GETV Macro domain without ADP-ribose were determined and are presented in the graphic Y axis.

Figure 38. Cartoon representation of the structures for GETV Macro domain alone (a) and with homologues (b). (a) Overall structure of native GETV Macro domain with APD-ribose depicted as seen in the GETV Macro domain with ADP-ribose complex “pose1”. β -sheets are coloured in slate, α -helices in teal and loops in pink. Secondary structure elements and the N- and C-terminal of GETV Macro domain are labelled. ADP-ribose is presented in stick-mode, with carbon atoms coloured in grey, oxygens in red, nitrogen in blue and phosphorus atoms in orange. ADP-ribose was added for sake reference for the ADP-ribose binding pocket. (b) Overlap of native GETV Macro domain, same colour coding as in (a), APD-ribose depicted as seen in the GETV Macro domain with ADP-ribose complex “pose 1” and colour-coded as in (a) with the structures of CHIKV Macro domain (3GPO) in orange, VEEV Macro Domain (3GQ) in yellow, SARS

Macro domain (2FAV) in purple and the non-histone domain of the histone variant MacroH2A1.1 (1YD9) in grey. The N- and C-terminus and secondary structure elements are labelled in black for GETV *Macro* domain and in red for SARS *Macro* domain. For clarity, only the first β -sheets are labelled. ADP-ribose was added for sake reference for the ADP-ribose binding pocket.

Figure 39. The binding mode of ADP-ribose adopted a new conformation in the distal ribose. From diffraction data collected from Xtal-15, the initial electron density maps were calculated before incorporation of the ligand and the unbiased F_o-F_c difference electron density maps indicated the presence of ADP-ribose (data not shown). **(a)** After incorporation of the ligand ADP-ribose with the distal ribose in close conformation, the final weighted difference maps F_o-F_c (red and green) countered at 3 σ level and the $2F_o-F_c$ (blue) countered at 1 σ level suggested that distal ribose was not in closed conformation. **(b)** Then, the ligand was incorporated with the distal ribose in open conformation, and the final weighted difference maps F_o-F_c (red and green) countered at 3 σ level and the $2F_o-F_c$ (blue) countered at 1 σ level supported the hypothesis that the distal ribose was in an open conformation. The cartoon represents the GETV *Macro* domain from Xtal-15 (a) in yellow, for the model with ADP-ribose with the distal ribose in close conformation. (b) the grey colour represents the model of Xtal-15 with the ligand ADP-ribose with the distal ribose in open conformation.

Figure 40. Interaction network between ADP-ribose in “pose 1” with GETV *Macro*. The ADP-ribose binding to GETV *Macro* domain as observed in the complex of GETV *Macro* domain with ADP-ribose in “pose 1”. ADP-ribose and residues interacting with ADP-ribose are depicted in sticks. Atoms of ADP-ribose and carbon atoms of GETV *Macro* domain residues are colour-coded as in Fig. 38.

Figure 41. Comparison between GETV *Macro* domain in “pose 1” (a) and in “pose 2” (b). The interaction between GETV *Macro* domain with the adenine moiety, the proximal ribose and the phosphate groups are virtually identically for both “pose 1” (a) and “pose 2” (b). While the distal ribose in “pose 1” is tilted by approximately 90° with reference to the “conventional” pose of the distal ribose observed in the already described “pose 1”. Unbiased F_o-F_c difference electron density maps, calculated before incorporation of the ligands into the models and contoured at 3.0σ , are shown in green, and final weighted $2F_o-F_c$ maps and contoured at 1.0σ are shown in blue.

Figure 42. Conformational trajectory triggered by the binding of ADP-ribose to GETV *Macro* domain in the presence of aspartic or glutamic acid. (a) Three-dimension structure of GETV *Macro* domain in complex with ADP-ribose presenting the distal ribose in close conformation “Pose 1” obtained by soaking a crystal obtained by co-crystallization with 3 mM ADP-ribose in mother liquor supplemented with 15 mM aspartic acid (Xtal-91). (b) Three-dimension structure of GETV *Macro* domain in complex with ADP-ribose presenting the distal ribose in close conformation “Pose 2” obtained from diffraction data collected from a crystal obtained by co-crystallization with 3 mM ADP-ribose and 50 mM glutamic acid (Xtal-62). (c) A three-dimension structure was determined from diffracting data collected from a crystal obtained by co-crystallization of 3 mM ADP-ribose and 30 mM aspartic acid (Xtal-68) and in this three-dimension structure the distal ribose adopted a double conformation. In this panel we present the conformer A. The conformation of conformer B is the same as the one obtained for the three-dimensional structure containing a single conformation for ADP-ribose with the distal ribose in open conformation. (d) In this panel is illustrated the conformer B that is the same conformation adopted by the ADP-ribose from the three-dimension structure presenting the ADP-ribose with a single open conformation (Xtal-72). (f) The three-dimension structure representing the covalent bond established between 1”-C and the Cys34 SG. This three-dimension structure was obtained from diffracting data collected from a crystal obtained by co-crystallization with 3 mM ADP-ribose and 3 mM of aspartic acid (Xtal-44). ADP-ribose and residues interacting with ADP-ribose are depicted in sticks and colour-coded as in Fig. 38 and 40.

Figure 43. GETV *Macro* domain in complex with ADP-ribose with the distal ribose in open conformation. (a) the three-dimensional structure of the complex at resolution 1.85 Å where the open distal ribose was modelled in a single conformation. (b) the three-dimensional structure of the complex at resolution 1.6 Å where the distal ribose in open conformation was modelled in a double conformation. Unbiased F_o-F_c difference electron density maps, calculated before incorporation of the ligands into the models and contoured at 3.0σ , are shown in green, and final weighted $2F_o-F_c$ maps and contoured at 1.0σ are shown in blue.

Figure 44. GETV *Macro* domain in complex with ADP-ribose with the distal ribose in open conformation and stablishing a covalent bond with the lateral chain of cysteine 34. Unbiased F_o-F_c difference electron density maps, calculated before incorporation of the ligands into the models and contoured at 3.0σ , are shown in green, and final weighted $2F_o-F_c$ maps and contoured at 1.0σ are shown in blue.

Figure 45. Getah nsP3^{wt} induced cytoplasmic foci that do not contain markers from the genuine SGs induced by arsenite induced. In red is represented the staining with nsP3 (Panel A and E), G3BP1 (Panel B and F), eIF3 (Panel C and G) and PABP (Panel D and H), Green is the expression of GFP in transfected cells and blue is the nucleus stained with Hoechst.

Figure 46. Getah nsP3^{wt} (D) and mutants nsP3^{Asn24Ala} (E), nsP3^{Cys35Ala} (F) and nsP3^{Asp31Ala} (G) can recruit G3BP1 to cytoplasmic foci that resemble like SGs in a similar fashion as nsP3^{wt} CHIKV (C). The VEEVwt (B) was used a negative control for G3BP1 cytoplasmic foci, the empty vector pCAGGS was used to demonstrate that expression of GFP do not induce G3BP1 cytoplasmic foci (E) and untransfected cells (Mock control) treated with arsenite were used as positive control for bona fide G3BP1 cytoplasmic foci (A) which are characteristic of bona fide SGs. In red is represented the staining with G3BP1, Green is the expression of GFP in transfected cells and blue is the nucleus stained with Hoechst.

Figure 47. Getah nsP3^{wt} and mutants induced the formation of G3BP1-containing cytoplasmic foci that were resistant to cycloheximide (CHX) treatment. Panels on the top (A to E) are representing the control conditions without cycloheximide treatment. The panels on the bottom are representing the conditions in which cycloheximide was used. Arsenite induced stress granules were easily dispersed upon cycloheximide treatment (A and F). Getah nsP3^{wt} induced cytoplasmic foci that contain G3BP1 and that were resistant to cycloheximide treatment (B and G). For the nsP3 Getah mutants [nsP3^{Asn24Ala} (C and H), nsP3^{Cys35Ala} (D and I), nsP3^{Asp31Ala} (E and J)] it was observed the same effect observed for the Getah nsP3^{wt}. In red is represented the staining with G3BP1, Green is the expression of GFP in transfected cells and blue is the nucleus stained with Hoechst.

THIS PAGE IS INTENTIONALLY LEFT BLANK

List of tables

Table 1. Therapeutic candidates for CHIKV under development and primary characteristics of each candidate^a (Powers, 2018) and references therein.

Table 2. Selected compounds from the analogue search of the head of series 3 (Prest-1291) used for the evaluation of the GT activity carried by nsP1 VEEV. The analogue search was performed using the screened library from Prestwick Chemical and commercially available databases. n.d.: not determined

Table 3. Evaluation of the inhibition of the head of series 1 and 2 in both GT and MTase activity carried by VEEV nsP1. Determination of IC₅₀ for the GT activity was performed by western-blot and the determination of the % of inhibition of MTase activity at 50 μM was measured by DEAE filter binding assay (FBA). Sinefungin was included as a reference compound. n.i.: no inhibition, n.d.: not determined

Table 4. MADTP compounds used for the evaluation of the GT activity carried by nsP1 VEEV. The IC₅₀ was determined in duplicate, except when the standard error is not presented.

Table 5. Summary of the crystallization and soaking procedures used to produce the crystals of GETV *Macro* domain that led to the structures discussed in this work.

Table 6. Data collection and refinement statistics.

Table S1. List of sense and anti-sense primers designed to obtain the mutants of nsP3 for VEEV and GETV by (multi-) site- directed mutagenesis.

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

Abbreviations

[$\alpha^{32}\text{P}$]-GTP. guanosine triphosphate labeled on the α -phosphate group with ^{32}P	Protein kinase B (PKB). serine/threonine-specific protein kinase	AUG. translation initiation codon
μm. micrometre, unit of length from the international system of units (IS) (10^{-6} m)	Ala. alanine	AURAV. Aura virus
μM. micromolar ($10^{-6}\text{ mol}\cdot\text{L}^{-1}$)	ALC1. amplified in liver cancer 1	BEBV. Bebaru virus
^{32}P. radioactive isotope phosphorus-32	AMP. adenosine monophosphate	BFV. Barmah forest virus
4EBP1. eukaryotic initiation factor 4E-binding protein 1	AMPK. adenosine monophosphate (AMP)-activated protein kinase	B-Myb v-Myb, MYBL2. myeloblastosis viral oncogene homolog (avian)-like2
5'QC. 5'-end quality control mechanism	AP-1. activator protein 1	BSA. bovine serum albumin
6K. alphavirus structural protein 6K	AP-2. adaptor protein 2	BST-2. interferon-inducible host membrane protein, also known as thetherin
A. Adenosine monophosphate	Arg. arginine	C. cytosine monophosphate
Å. ångström (10^{-10} m)	Arp3. actin-related protein 3	C3. complement factor 3
a.a.. amino acids	Asn-X(6)-Gly-Gly-[Val/Leu/Ile]. specific signature motif from the catalytic loop 1 of MacroD-type Macro domains	CABV. Cabassou virus
A1''Pase. ADR-ribose-1''-phosphate phosphatase	Asp. aspartic acid	CARDs. caspase-recruitment domains
AAR. O-acetyl-ADP-ribose	ATF4. activating transcription factor 4	CCCH. Cys-Cys-Cys-His zinc finger domain
ADP. adenosine diphosphate	ATP. adenosine triphosphate	CDC. Centre Disease Control
ADPr ADP-ribose. adenosine diphosphate ribose	ATPase. adenosone-triphosphatase, adenosine triphosphatase activity	cdE2. carboxy-terminal domain E2
Ago. argonaute	AUD. alphavirus unique domain	CDS. coding sequences
AIF. apoptosis-inducing factor		CE. Chicken embryo cells
Akt		CEV. California encephalitis virus
		CHIKV. Chikungunya virus
		CHX. cycloheximide
		CK II. casein kinase II
		cm. centimetre (0.01 metres [m] or 10 millimeters [mm])

COX-2. regulate the cyclooxygenase-2	terminus of the E1 ectodomain	EEEV. Eastern equine encephalitis virus, Eastern equine encephalomyelitis virus, Eastern equine encephalomyelitis virus
Cp. capsid protein C	DII. distal β -barrel domain from the E1 ectodomain and contains the fusion loop	eIF2α. eukaryotic initiation factor 2 α subunit
CPVI. cytophatic vacuole I	DIII. membrane-proximal β -barrel domain from the E1 ectodomain	eIF4E. eukaryotic initiation factor 4E subunit
CPVII. type II cytopathic vacuole(s)	DMEM. Dulbecco's modified Eagle's medium	EILV. Eilat virus
CPVs. cytophatic vacuoles	DNA. Deoxyribonucleic acid	eIF3. eukaryotic initiation factor 3 subunit
CREs. cis-acting regulatory elements	dpCoA. desphospho-coenzyme A	eIF4B. eukaryotic initiation factor 4B subunit
cryo-EM. cryogenic Electron Microscopy	dsRNA. double stranded RNA	ELISA. enzyme-linked immunosorbent assay, enzyme-linked immunosorbent assay
CSE1. conserved sequence element 1	DXO. decapping and exoribonuclease protein, decapping exoribonuclease	EMEM. Eagle's Minimum Essential Medium
CSE2. conserved sequence element 2	<i>E. coli.</i> <i>Escherichia coli</i>	ER. endoplasmic reticulum
CSE3. conserved sequence element 3	E1. the structural protein named glycoprotein 1	eRC. viral early replication complex
CSE4. conserved sequence element 4	E2. the structural protein named glycoprotein 2	ESCRT. endosomal sorting complexes required for transport
CTD. C-terminal (CTD) domain of the RNAP	E3. the structural protein named glycoprotein 3	EVEV. Everglades virus
Cys. cysteine	ECSA. Chikungunya East/Central/South Africa Lineage	FAD. flavin adenine diphosphate
D1. first putative domain of nsP2	EDTA. Ethylenediaminetetraacetic acid	FBA. DEAE filter binding assay
D2. second putative domain of nsP2	EEE. Eastern equine encephalitis	FDA. US Food and Drug Administration
Da. Dalton (unified atomic mass unit, 1.66×10^{-27} Kg)	EEE complex. Eastern equine encephalitis antigenic complex	FGDF. Phe-Gly-Asp-Phe motif
deNADing. degradation of RNA through removal of the NAD ⁺ cap		FMV. Fort Morgan Virus
DENV. Dengue virus		FXR. fragile X syndrome proteins family
DENV-2. Dengue virus-serotype 2		
DEXH. Asp-Glu-X-His motif		
DI. β -barrel domain containing the N-		

g. gram (10 ⁻³ Kg)	loop 2 of MacroD-type	HS. heparan sulphate
G. guanosine monophosphate	Macro domains	Hsc70. heat shock protein 70
G3BP. Ras-GTPase activating protein (GAP) -(Scr Homology 3 [SH3] domain) binding protein	Gly-Gly-Gly-X(6-8)-Gln-Glu-Glu. specific catalytic motif from MacroD-type	HuR. Human antigen R protein
G3BP1 G3BP Stress Granule Assembly Factor 1. Ras-GTPase activating protein (GAP) -(Scr Homology 3 [SH3] domain) binding protein 1	GM-CSF. granulocyte macrophage-colony stimulating factor	HVR. hypervariable region
G3BP2 G3BP Stress Granule Assembly Factor 2. Ras-GTPase activating protein (GAP) -(Scr Homology 3 [SH3] domain) binding protein 2	GMP. guanosine monophosphate	IBV. Infectious Bronchitis Virus
GAP. Ras-GTPase activating protein	GpαpβpαNpNp-RNA. guanosine monophosphate cap linked to the 5'-end diphosphate mRNA	IC₅₀. half maximal inhibitory concentration
GDD. Gly-Asp-Asp motif present at the palm structure of the C-terminal portion of nsP4 (RdRp domain)	GT. guanylylation, guanylylation activity	ICAM-1 ICAM-1. <i>ICAM-1: Intercellular Adhesion Molecule 1 gene,</i> ICAM-1: Intercellular Adhesion Molecule 1 protein
GDP. guanosine diphosphate	GTase. guanylyl-transferase activity	ICTV. International Committee on Taxonomy of Viruses
GETV. Getah virus	GTP. guanosine triphosphate	IF. immunofluorescence
Gln. glutamine	HBSS. Hank's Balanced Salt Solution	IFA. immune fluorescence analysis, indirect immunofluorescent assay
Glu. glutamic acid	HEV. Hepatitis E Virus	iFCS. heat-inactivated fetal calf serum
Gly. glycine	His. histidine	IFIT. <i>interferon induced proteins with tetratricopeptide repeats,</i> interferon induced proteins with tetratricopeptide repeats
Gly-[Val/Ile/Ala]-[Tyr/Phe]-Gly. specific signature motif from the catalytic	His-Cys-X(5)-Arg-(Ser/Thr). phosphate binding loop or "P-loop" from the RTPase enzyme	IFIT1. Interferon-induced protein with tetratricopeptide repeats 1
	HIV. human immunodeficiency virus	IFN. interferon
	HJV. Highlands J Virus	IFN-γ. interferon- γ
	HMGB1. high mobility group box 1 protein	
	hnRNP. Heterogeneous nuclear ribonucleoprotein such as hnRNP K, A1, M, C and E1	

IgG. immunoglobulins G	INF β promoter	General Conference on
IgM. immunoglobulins M	stimulator 1	Weights and Measures
IKK. inhibitor of NF κ B (I κ B) kinase complex	IPTG. isopropyl β -D-1- thiogalactopyranoside	[CGPM], Versailles [France], 2018)
IL-1. interleukin-1	IRE1α. inositol requiring kinase enzyme 1 alpha	IRC. viral late replication complex
IL-1β IL-1β. <i>IL-1β: cytokine interleukin 1β gene, IL- 1β: cytokine interleukin 1β protein</i>	IRES. internal ribosome entry site elements	LRR. leucine-rich repeat motif
IL-2 IL-2. <i>IL-2: cytokine interleukin 2 gene, cytokine interleukine 2 protein</i>	IRFs. interferon regulatory factors	Lys. lysine Lys-X-Asp-Gly. sequence motif from the human GTase that contains the nucleophilic lysine that forms the GTase-GMP adduct
IL-6 IL-6. <i>IL-6: cytokine interleukin 6 gene, IL-6: cytokine interleukin 6 protein</i>	ISG. IFN-stimulated (antiviral) genes ISGs. <i>IFN-stimulated genes, IFN-stimulated genes</i>	m^{2,2,7}G cap TMG. 2,2,3- trimethylguanosine cap
IL-8 IL-8. <i>IL-8: cytokine interleukin 8 gene, IL-8: cytokine interleukine 8 protein</i>	IκB. inhibitor of NF- κ B J. Junction region JAK-STAT. <i>janus kinase (JAK)-signal transducer and activator of transcription (STAT), janus kinase (JAK)-signal transducer and activator of transcription (STAT), signalling pathway of janus kinase-signalling transducer and activator of transcription</i>	m⁶A_m. N6 methyl-adenosine with a methylation at the ribose-2'-O-hydroxyl m⁷G. canonical N ⁷ -methyl- guanosine m⁷GMP. N7 methyl guanosine monophosphate m⁷Gp. N7 methyl guanosine monophosphate m⁷Gppp. N7 methyl guanosine triphosphate m⁷GpppN-RNA. N7-methyl guanosine (m ⁷ G) linked to the first 5'-end nucleotide (N) of the mRNA through a triphosphate bridge (ppp).
Ile. isoleucine	Kb. kilobase pairs (10 ³ nucleotide residues)	
IMAC. Immobilized Metal Affinity Chromatography	kDa. kilodalton (10 ³ dalton)	
INFR. interferon receptor	Kg. Kilogram, base unit of mass in the International System of Units, defined by the Planck constant 6.62607015 $\times 10^{-34}$ m ² kg \cdot s ⁻¹ (26 th meeting of the	
iNOS iNOS. <i>iNOS: inducible nitric oxide synthase gene and iNOS: inducible nitric oxide synthase protein</i>		
IPS1 CARDIF, MAVS or VISA. the signalling adaptor		

m⁷Gp_αp_βp_αN_mN_m-RNA. canonical type 2 cap structure	mTORC1. mammalian target of rapamycin complex 1, mammalian target of rapamycin complex 1	NLRs. nucleotide-binding oligomerization domain (NOD)-like receptors
m⁷Gp_αp_βp_αN_m-RNA. canonical type 1 cap structure	MUCV. Mucambo virus	NLS. nuclear localization signals
m⁷Gp_αp_βp_αNpNp-RNA. canonical type 0 cap structure	MurNAc-pentapeptide. uridine diphosphate N-acetyl-muramyl-pentapeptide	nm. nanometer
MADV. Madariaga virus	MWAV. Mwinilunga virus	NO. nitric oxide
MARylation. mono-(ADP)-ribosylation	MYD88. myeloid differentiation primary response 88 adaptor protein	NOD. nucleotide-binding oligomerization domain
MAYV. Mayaro virus	N- nucleotidyl-transferase. N-nucleotidyl-transferase	NRf. nuclear respiratory factor 1
MBLs. mannose binding lectins	N⁶. nitrogen position 6 of adenine nucleoside	nsP1. <i>alphavirus</i> non-structural protein 1
MCP-1. macrophage chemoattractant protein-1	N⁷. Nitrogen at the position 7 of the guanosine and is in a sp ³ hybridized orbital. Its lone pair alone is not part of resonance or aromaticity of guanosine and is the most prone to alkylation.	nsP2. <i>alphavirus</i> non-structural protein 2
MDA-5. melanoma differentiation-associated gene 5	NAD⁺. oxidized form of nicotinamide adenine dinucleotide	nsP23. <i>alphavirus</i> intermediate non-structural protein 23
MDPV. Mosso das Pedras virus	NADH. reduced form of nicotinamide adenine dinucleotide	nsP3. <i>alphavirus</i> non-structural protein 3
Met. methionine	NC. nucleocapsid	nsP4. <i>alphavirus</i> non-structural protein 4
MHV. Mouse Hepatitis Virus	NDUV. Ndumu virus	nt. nucleotides
MIDV. Middelburg virus	NF-κB. nuclear factor-κB	NTF2. Nuclear Transport Factor 2-like domain
MIF. macrophage migrating inhibitory factor	NHC. β-D-N ⁴ -hydroxycytidine	NTPase. nucleoside triphosphatase
mL. milliliter		Nudix. nucleoside diphosphate linked to another moiety-X
MMPs. matrix metalloproteinases		NW. New World
MoA. mode of action		OAADPr. O-acetyl-ADP-ribose
mRNAs. messenger RNAs		OB-fold. oligonucleotide/oligosaccharide-binding fold is a five/six-stranded closed β-barrel
MTase. methyltransferase activity		
mTOR. mammalian target of rapamycin		

Oct-1. Octamer-binding transcription factor-1	PARYlation. poly-(ADP)-ribosylation	Phe. phenylalanine, phenylalanine
ONNV. O'nyong-nyong virus	PB1. polymerase base protein 1 from Influenza virus	PI(3,4,5)P₃. phosphatidylinositol-3,4,5-trisphosphate
ONOO[•]. peroxyneutron	PB2. polymerase base protein 2 from Influenza virus	PI(4,5)P₂. phosphatidylinositol-4,5-biphosphate
ORFs. open reading frames	PBS. phosphate buffered saline	PI3K. phosphoinositide-3-kinase protein
OW. Old World alphaviruses	PCBP1	PI3K-Akt-mTOR pathway. also known as the "pro-survival" pathway, <i>see</i> PI3K, Akt and mTOR
p110. PI3K catalytic subunit	hnRNP E1. Poly(rC)-binding protein 1, alternative name: Heterogeneous nuclear ribonucleoprotein E1 (hnRNP E1)	PIP5K1-α. lipid kinase phosphatidylinositol-4-phosphate 5-kinase1- α
P123. non-structural polyprotein 123	PCR. polymerase chain reaction	PIXV. Pixunia virus
P1234. non-structural protein 1234	PDK1. phosphatide-dependent kinase 1	PKC. protein kinase C
p62. autophagic membrane adaptor protein, same as pE2 (<i>see</i> pE2)	pE2. also known as p62, is a transmembrane protein intermediate product of E3 connected to E2	PKR. antiviral protein kinase R, protein kinase R
p65. p65 subunit (65 kDa) that dimerise with p50 subunit (50 kDa) to form the heterodimer NF- κ B	PERK1. proline-rich, extensin-like receptor kinase-1	PM. plasma membrane
p7. membrane-associated ion channel protein from hepatitis C virus (HCV)	PFA. paraformaldehyde	PMSF. phenylmethylsulfonyl fluoride
p85. PI3K regulatory subunit	PFU. plaque forming units	poly(A). poly- adenosine-monophosphate
PA. polymerase acidic protein from Influenza virus	pH. logarithmic scale for measure the acidity or alkalinity of a solution, $\text{pH}=\log_{10}[\text{hydrogen ion (H}^{\text{+}}\text{)}]$	poly-(ADP-ribose). poly- adenosine diphosphate ribose
PABP. poly(A)-binding protein	PH. pleckstrin-homology domain	PP_i. pyrophosphate
PAB. poly(A)-binding protein		PRNT. plaque reduction neutralization test
PAMPs. pathogen-associated molecular patterns		PRNTase. RNA GDP-polyribonucleotidyl-transferase
PAR. poly-(ADP-ribose)		Pro. proline, proline
PARG. poly-ADP-ribose glycohydrolase		PRR. pattern recognition receptors
PARPs. poly-(ADP-ribose) polymerases		

PRRs. pattern recognition receptors	RVFV. Rift Valley fever virus	SUD-M. middle of the SARS-unique domain
Rac1. Rac family small GTPase 1	S. Svedbergs (alternative abbreviation Sv) is a non-metric unit for sedimentation coefficient (deffined as 10^{-3} seconds)	T. triangulation number
RdRp. RNA-dependent RNA polymerase	S6. ribosomal protein S6	TALV. Tai Forest virus
RGG. Arg-Gly-Gly motif of the C-terminal portion of G3BP	SAGV. Sagiya virus	TARG1 OARD1 or C6orf130. the terminal ADPr protein glycohydrolase 1
RIG. retinoic acid-inducible gene	SAH AdoHcy. S-adenosyl-L-homocysteine	TATase. adenylyltransferase activity
RIG-I. retinoic acid-inducible gene I	SAM AdoMet. S-adenosyl-L-methionine	TB. Terrific Broth
RLRs. retinoic acid-inducible gene (RIG)-like receptors	SAV. Salmonid Alphavirus	TEF-1. elongation factor 1 alpha
RNA. ribonucleic acid	Ser. serine	TEM. transmission electron microscopy
RNAP II. RNA polymerase II	SESV. Southern elephant seal virus	TF. transframe form of the alphavirus structural protein 6K
RNase L. Ribonuclease L	SF. Semliki forest	T_H2 CD4⁺. T helper cells
RNV. Rio Negro virus	SF complex. Semliki forest antigenic complex	Thr. Threonine
ROS. reactive oxygen species	SF1. helicase signature motifs of superfamily1	TIA-1. T-cell-restricted intracellular antigen 1
Rpb1. DNA-directed RNA polymerase II subunit Rpb1	SFV. Semliki forest virus, Semliki forest Virus	TIAR. TIA-1-related protein
RRM. RNA Recognition Motif domain	SG. stress granule	TIR. Toll-IL-1 receptor
RRV. Ross River Virus	SGs. stress granules	TLRs. Toll-like receptors
RSEs. repeat sequence elements	SH2. Src Homology 2 domain	T_m. melting temperature
RT. reverse transcriptase	SH3. Src Homology 3 domain	TNF-α. tumor necrosis factor- α
RT-LAMP. Reverse Transcription loop-mediated isothermal amplification	SINV. Sindbis virus	TONV. Tonate virus
RTPase. RNA triphosphatase	snoRNAs. small nucleolar RNAs	Treg. CD4 ⁺ CD25 ⁺ /Foxp3 ⁺ regulatory T cells
RT-PCR. reverse transcriptase - polymerase chain reaction	snRNAs. small nuclear RNAs, small nuclear RNAs	TRIF. TIR-domain-containing adapter-inducing interferon- β
	SPDV. Salmon pancreatic disease virus	tRNA. transfer RNA
	ssRNA. single stranded RNA	tRNA^{Trp}. tRNA is an RNA that binds L-thryptophan

which is then transferred to ribosomes during translation	(U₆A) . hepta-nucleotide slippery sequence	VSV . Vesicular Stomatitis Virus
TROV . Trocara virus	Val . valine	WA . Chikungunya West Africa Lineage
Trp . tryptophan	VEE . Venezuelan equine encephalitis	WB . Western blot
TSA . thermal shift assay	VEE complex . Venezuelan equine encephalitis antigenic complex	WEE . Western equine encephalitis
Tyr . tyrosine	VEEV . Venezuelan equine encephalitis virus	WEE complex . Western equine encephalitis antigenic complex
U . uracil monophosphate	Vero cells . lineage of cells used in cell cultures	WEEV . Western equine encephalitis virus
UAA . ochre stop codon	VLP . virus-like particle	WHAV . Whataroa virus
UAG . amber stop codons	VPg . known as viral protein genome-linked. The VPg protein acts as a primer for initiation of RNA synthesis after its post-translational modification of uridylylation.	WNV . West Nile virus
UDP-GlcNAc . uridine diphosphate N-acetylglucosamine	Vpu . viral protein U from human immunodeficiency virus (HIV)	WT . wild type
UDP-glucose . uridine diphosphate glucose	vRC . viral replication complexes	XBP1 . X-box binding protein 1
UGA . opal stop codon		YBX1 . Y-box-binding protein 1
UNAV . Una virus		YFV . Yellow fever Virus
UPR . unfolded protein response, unfolded protein response		ZAP . zinc-finger antiviral protein, zinc-finger antiviral protein
URE . U-rich elements		ZBD . zinc-binding domain
US . United States		ZIKV . Zika virus
UTRs . untranslated regions		
UUUUUUA		

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

I-INTRODUCTION

THIS PAGE IS INTENTIONALLY LEFT BLANK

1. (Re-) emerging arboviruses: Alphaviruses

Arboviruses, whose acronym comes from arthropod-borne viruses, are emerging and re-emerging viruses. According to Iranpour *et al.* (Iranpour *et al.*, 2016) arboviruses account for 534 members listed in the International Catalogue of Arboviruses, of which 134 are human pathogens. In 1927 the first arbovirus was discovered, the Yellow fever (YFV) virus, and initially arboviruses were classified as Arbovirus Group A or Group B. Later, the viruses possessing a viral envelope acquired the jargon name “togavirus” that derives from Latin *toga*, a Roman “mantle” or “cloak” which refers to the envelope. In 1975, “togavirus” were officially nominated as an approved family named *Togaviridae* and comprising only two genera, *Alphavirus* (the former Group A arboviruses) and *Flavivirus* (the former Group B arboviruses). During several years, the classification was subject to deep modifications, as an example in 1984, *Flavivirus* were removed from the *Togaviridae* family and reached the level of family (Westaway *et al.*, 1985). Nowadays according to ICTV (International Committee on Taxonomy of Viruses) in 2018, the *Togaviridae* family is composed of the genera *Alphavirus* and *Rubivirus*. The *Flavivirus* is now a genus from the *Flaviviridae* family together with Hepacivirus, Pegivirus and Pestivirus genera.

Human pathogenic arboviruses belong mainly to four distinct genera of RNA viruses, namely *Alphavirus* (e.g., Venezuelan equine encephalitis Virus [VEEV], Chikungunya virus [CHIKV] and Eastern equine encephalomyelitis virus [EEEV]), *Flavivirus* (e.g., Zika virus [ZIKV], Yellow fever virus [YFV], Dengue virus [DENV], and West Nile virus [WNV]), *Orthobunyavirus* (e.g., California encephalitis virus [CEV]) and *Phlebovirus* (e.g., Rift Valley fever virus [RVFV]) (Lequime *et al.*, 2016). The focus of this manuscript will be mainly within the genera *Alphavirus*.

1.1. Transmission of arboviruses

Despite the global distribution of arboviruses, the majority circulates in tropical areas with suitable climate condition for year-round transmission. Transmission occurs through horizontal (non-parental transmission) or vertical transmission (hereditary transmission). Arboviruses are primarily transmitted between hematophagous arthropod vectors and vertebrates (horizontal transmission) during the enzootic cycle in sylvatic habitats (figure 1). The enzootic cycle (also known as sylvatic or jungle cycle) is attributed to the transmission that occurs in sylvatic habitats which circulates between wild animals and the arthropod vector. Spill over from the enzootic cycle can occur when humans or

domestic animals enter inside the sylvatic habitats. Usually, spill over from sylvatic transmission cycles to humans is incidental and humans are dead-end hosts (Weaver, 2013, Weaver & Barrett, 2004).

During this type of horizontal transmission, the blood feeding vectors become infected after a meal on a viremic vertebrate. The virus replicate after a development period within the vector and new vertebrate host can be infected during a subsequent blood meal. Horizontal transmission is probably an inefficient maintenance mechanism when the climacteric conditions are adverse. Thus, vertical transmission is an alternative mechanism for the persistence of the virus in an arthropod vector subpopulation. In mosquitoes, the vertical transmission occurs from the female to her off-spring by trans-ovarial transmission or trans-egg transmission. The trans-ovarial transmission occurs if the virus infects the germinal tissues and trans-egg transmission if the virus infects the eggs during the oviposition (Fig. 1). Combination of both horizontal and vertical transmission of the virus within the mosquito subpopulation might be important for long term arboviral maintenance (Weaver, 2013, Weaver & Barrett, 2004).

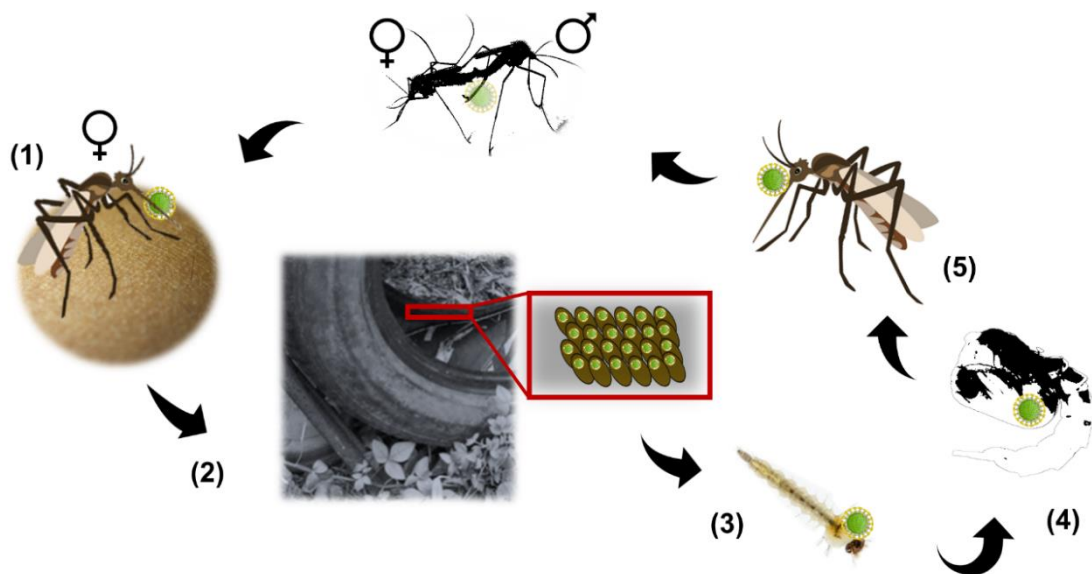


Figure 1. Vertical transmission of mosquito-borne arboviruses. Mosquitoes such as *Aedes albopictus* or *Aedes aegypti* are endemic of inhabited densely vegetated rural areas and they had developed adaptability to colonize new habitats in urban and suburban areas. **(1)** Usually female mosquito needs to intake a blood meal to sustain their egg production. When feeding from an (arbovirus)-infected mammal, the arbovirus can be transmitted to the off-spring of this female mosquito through trans-egg transmission. **(2)** In urban and suburban areas, female mosquitoes utilize human-made water holding containers such as flowerpots, soda cans, buckets and tires that can collect and retain rainwater for a long period of time. During oviposition, females laid their eggs usually slightly above the water level. **(3)** Their eggs are resistant to drying out and when covered with water the larvae emerge from the egg and undergoes a process of development through four stages and feed primarily from organic material present in the water. **(4)** After completing the four stage the larvae enter the pupal stage. **(5)** The pupal stage last for a few days until the adult mosquito emerges and seeks a resting site in low vegetation. The female mosquitoes are then ready to start a new cycle and the male mosquitoes in some cases can transmit the arbovirus to an uninfected female during mating, through trans-ovarial transmission.

Under appropriate ecological conditions, spill over of arboviruses can occur from enzootic sylvatic habitats. Both humans and domestic animals can be infected with arboviruses by enzootic- or bridge- arthropod vectors such as mosquitoes (Fig. 2). Usually, humans are dead-end or incidental hosts because they do not develop enough viremia to extend the transmission chain. However, some exceptions are distinguished, such as DENV, YFV and CHIKV that primarily infect people during outbreaks and then begin to use humans as amplification sources (Weaver, 2013). Domestic animals infected with arbovirus during spill over are more prone to establish a rural enzootic cycle. Upon amplification in the intermediate host, the rural enzootic cycle can evolve to an epizootic rural cycle (Fig. 2). The transmission of arboviruses to humans can occur preferentially from epizootic rural cycles more frequently than from sylvatic enzootic cycles. Such susceptibility is due to the close contact of humans with domestic animals than with wild animals. Through adaptation, arboviruses can also alter their host range to humans such as DENV and CHIKV. Once adapted, an urban epidemic cycle can be

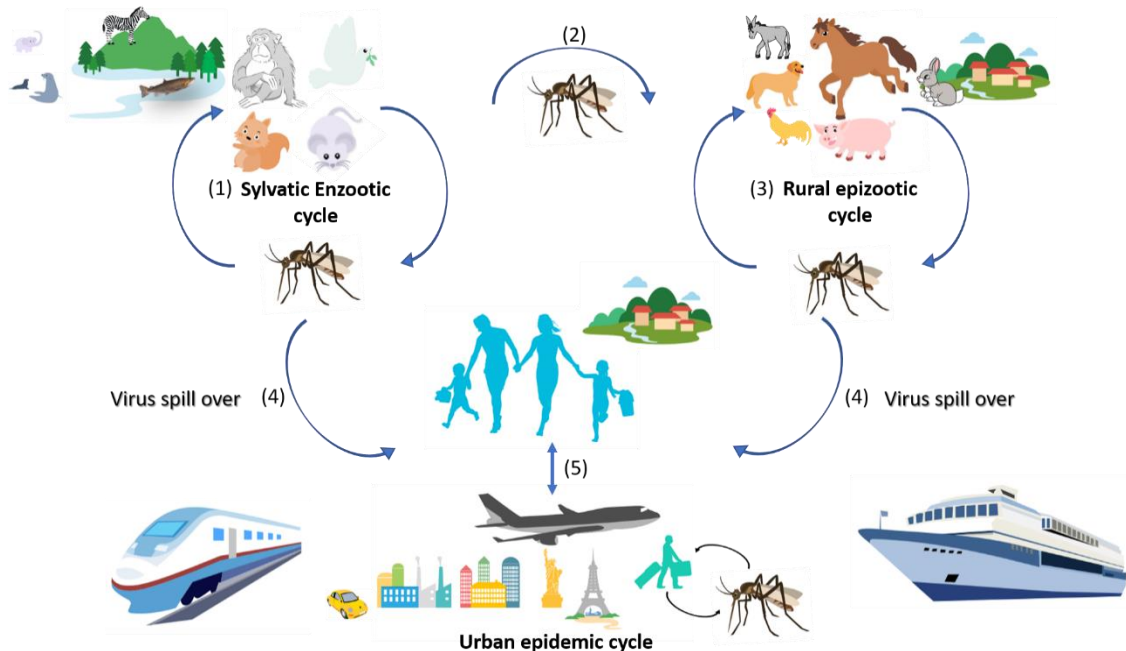


Figure 2. Transmission mode of mosquito-borne arboviruses. Most arboviruses normally circulate among non- human animals and can occasionally be transmitted to humans (Zoonotic pathogens). The transmission mode of alphaviruses usually requires alternating replication cycles between vertebrate and arthropod. The reservoir of these zoonotic pathogens are usually birds, rodents and non- human primates which are involved in the perpetuation of the enzootic cycle. Mosquito vectors such as species from the genus *Culex*, *Ochlerotatus* and *Psorophora* are involved in the transmission between vertebrate hosts. Usually the enzootic cycle of alphaviruses occurs in sylvatic habitats (1). Enzootic and bridge- vectors might be involved in the indirect transmission to humans through the rural epizootic cycle (1 → 2 → 3). The rural epizootic cycle (3) is associated with domestic animals and mosquitoes where the virus amplifies in the presence of intermediate hosts. Thus, (3) represents a large reservoir of viruses from where can result a severe spill over effect to dead-end hosts. Human infection and epidemics can arise when humans enter in the sylvatic enzootic habitats (1 → 4) or when the amplification of the virus occurs in domestic animals within a rural epizootic cycle (such as equines and pigs) followed by transmission to humans (3 → 4). Adaptation of arboviruses to the epizootic mosquito vectors and to the death- end hosts seems to be associated with several outbreaks. The movement of humans between rural and urban habitats are determinant for the initiation of urban epidemic cycles (5) which are dependent on the adaptation of the virus to the arthropod species present in the urban areas. Adapted from (Weaver & Barrett, 2004).

established if the necessary conditions are present in place (Fig. 2). Several factors have increased the risk for some arboviruses to trigger an urban epidemic cycle. Among which are: expansion of arthropod vector, increased urbanization, increase in global travel, trade of animals, global climate change and expansion of competent arthropod vectors to new territories. Thus, when all the conditions are present, the susceptibility for emergence and re-emergence of arboviruses increases.

Arbovirus members from the genera *Alphavirus* (family *Togaviridae*) are emerging and re-emerging. Some are known as important human pathogens because they can cause severe arthritis, nervous central diseases, coma or death. Thus, alphaviruses are nowadays a major concern of public health.

1.2. Arboviruses from the *Alphavirus* genus

The *Alphavirus* genus, according to ICTV in 2018, has 31 species registered, among which 11 are human pathogens. Alphaviruses can be divided in three groups, the larger group being specific to terrestrial species, and concerning the two other groups, one insect specific and the other specific to marine species (Nasar *et al.*, 2012). Terrestrial alphaviruses are widely distributed and mainly divided in two main groups which are the Old World (OW) and the New World (NW), Fig. 3 (Strauss & Strauss, 1994, Nasar *et al.*, 2012).

The definition of both worlds initially comes from a geographic perspective of Europeans who saw the Americas after the post-Columbian period (after 1492) or from the anthropologist's early point-of-view who saw Africa, Asia and Europe as the cradle of humanity (Merbs, 1992). The division of the world in Western and Eastern hemispheres appears primarily in 1494 with the creation of the Tordesillas meridian and later in 1529 with the Zaragoza antemeridian. The infectious diseases in both hemispheres evolved separately before 1492 with a documented history of infectious disease in the Eastern hemisphere. Whereas, in the Western hemisphere the history of infectious disease was unknown prior to 1492 (Merbs, 1992). It is possible that before 1492 some disease circulated between both hemispheres because most likely Christopher Columbus was not the first human being arriving to the Americas. Possibly, the migratory birds might have a role in the circulation of *Alphavirus* species between both hemispheres (Garmashova, Gorchakov, *et al.*, 2007, Strauss & Strauss, 1994, Weaver *et al.*, 1993, Forrester *et al.*, 2012). However, it is believed that the colonization of the Americas by Europeans highly contributed to the movement of disease between both hemispheres (Merbs, 1992). Furthermore, it is well documented that the impact of infectious disease

from the OW in the Americas (NW) after 1492 (Merbs, 1992, Strauss & Strauss, 1994). The position of Australia has been elusive but initially it was probably recognized as NW because Europeans discovered Australia in 1606 (after Americas). Moreover, Australia was at that time divided between the two hemispheres by the Zaragoza antemeridian. Nowadays, it is known that human occupation of northern Australia occurred around 65 000 years ago (Clarkson *et al.*, 2017) and this places Australia in the OW. In addition, the geographic division of both hemispheres is defined by the internationally recognized single meridian known as Greenwich meridian (established in 1884) and its antemeridian (180° meridian that pass-through open waters of the Pacific Ocean). The geographic definition of NW has changed to only Americas and concerning the OW it is still including Europe, Africa, Asia and now Australia.

The OW and NW division characterizes mainly the geographic location from where the species are endemic. However, the classification in to OW or NW is not only restricted to their geographic location but also to their symptoms of the disease. Therefore, NW alphaviruses are known to cause mainly high fever and encephalitis. Whereas the OW alphaviruses are mainly arthritogenic viruses causing high fever, rash, severe joint pain (arthralgia), and arthritis that can last weeks to months (Suhrbier *et al.*, 2012). Occasionally, the more severe OW alphaviruses are associated with encephalitis in animals and humans, such as Sindbis virus (SINV) (Griffin, 2005), Ross River Virus (RRV) (Harley *et al.*, 2001), Semliki forest virus (SFV) (Willems *et al.*, 1979), and CHIKV (Mehta *et al.*, 2018), with the first two also currently endemic in Australia (Gyawali *et al.*, 2017)). SINV and SFV usually do not cause, however, serious illness in humans.

Alphaviruses classified within the NW group are mainly organized in two antigenic complexes, excluding both Trocara virus (TROV) and Aura virus (AURAV) (Torii *et al.*, 2018, Nasar *et al.*, 2012), Fig. 3. The large antigenic complex from the NW group is known as the Venezuelan equine encephalitis (VEE) antigenic complex (VEE complex). The VEE complex includes strains from six subtypes of Venezuelan equine encephalitis virus (VEEV) (*i.e.*, VEE IC, VEE ID, VEE IAB, VEE IE and VEE 71D-1252), the Everglades virus (EVEV), Cabassou virus (CABV), Tonate virus (TONV), Mucambo virus (MUCV), Pixunia virus (PIXV), Rio Negro virus (RNV) and Mosso das Pedras virus (MDPV) Fig. 3 (Torii *et al.*, 2018). The second complex from the NW group is recognized as the Eastern equine encephalitis (EEE) antigenic complex (EEE complex) and comprises the species Madariaga virus (MADV), Eastern equine encephalomyelitis virus (EEEV) and Eastern equine encephalitis virus (EEEV).

The OW group is organized mainly in one complex named as Semliki forest (SF) antigenic complex (SF complex). The SF complex includes the following *Alphavirus*

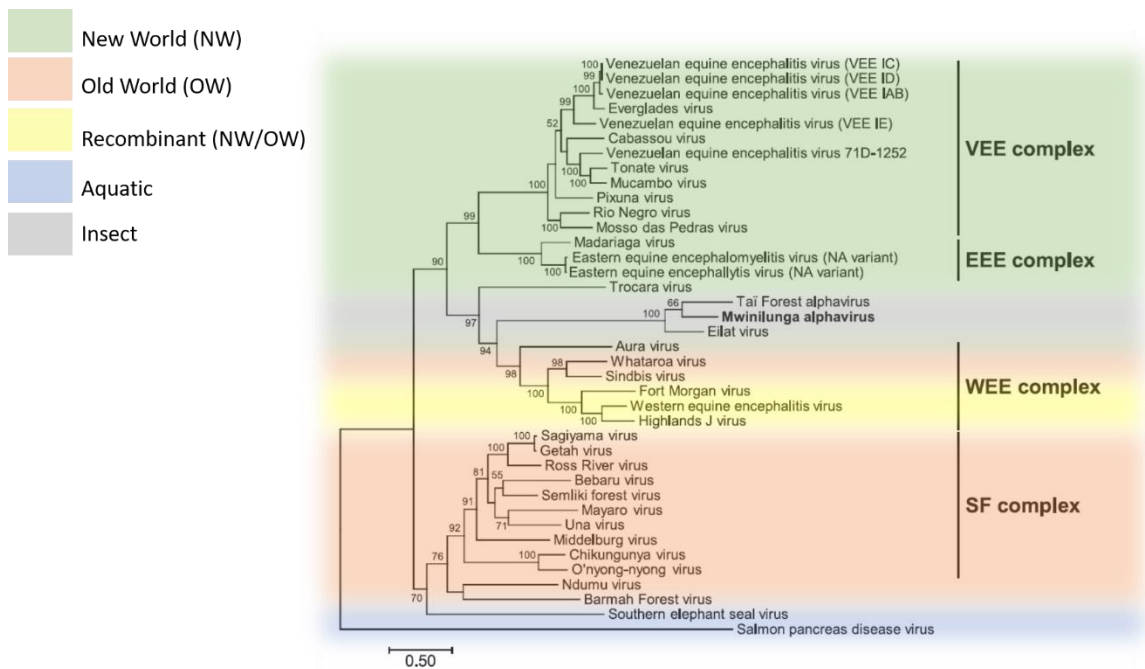


Figure 3. Phylogenetic tree constructed using the maximum-likelihood method in MEGA, version 6.0 with 1,000 bootstrap replicates based on multiple alignment of nucleotide sequences of the structural proteins. Bootstrap values greater than 50 are shown near the branch nodes and the scale bar indicates the number of substitutions per site. Adapted from Torii *et al.* (Torii *et al.*, 2018).

species; Sagiya virus (SAGV), Getah virus (GETV), Ross River virus (RRV), Bebaru virus (BEBV), Semliki forest virus (SFV), Mayaro virus (MAYV), Una virus (UNAV), Middelburg virus (MIDV), CHIKV and O’nyong-nyong virus (ONNV). In addition, Ndumu virus (NDUV), Barmah Forest virus (BFV), Whataroa virus (WHAV) and SINV are not included in the SF complex however are also included in the OW group, Fig. 3 (Nasar *et al.*, 2012, Torii *et al.*, 2018), Fig. 3.

The last known antigenic complex from the terrestrial alphaviruses is the Western equine encephalitis (WEE) antigenic complex (WEE complex). The WEE complex comprises NW viruses (*e.g.*, AURAV) and OW viruses (*e.g.*, WHATV and SINV) but also recombinant viruses that resulted from the recombination events among arboviruses (Torii *et al.*, 2018, Nasar *et al.*, 2012, Hahn *et al.*, 1988, Weaver *et al.*, 1997). The viruses known to result from the recombination between species from the NW and OW (probably from the EEE complex with SINV ancestors respectively) are Fort Morgan virus (FMV), Western equine encephalitis virus (WEEV) and Highlands J Virus (HJV), Fig. 3, (Nasar *et al.*, 2012, Torii *et al.*, 2018).

The recent discovery of the *Alphavirus* insect specific group introduced a new phylogenetic organization within the genera *Alphavirus* (Nasar *et al.*, 2012). Eilat virus (EILV) was the first discovered *Alphavirus* member with a host specificity in insects (Nasar *et al.*, 2012) followed by Tai Forest virus (TALV) (Hermanns *et al.*, 2017) and Mwinilunga virus (MWAV) (Torii *et al.*, 2018). The phylogenetic analysis previously

provided by Nasar *et al.* (Nasar *et al.*, 2012) showed that EILV isolated from *Anopheles coustani* mosquitoes from Negev desert in Israel (OW) was placed at the base of the WEE complex. Later, TALV was isolated from *Culex decens* mosquitoes collected in Ivory Coast (OW) and it was observed the same basal relationship to the WEE complex as observed by Nasar *et al.* with EILV (Nasar *et al.*, 2012, Hermanns *et al.*, 2017). In addition, TALV was characterized as sister of EILV. The recent discovery of MWAV (Torii *et al.*, 2018) isolated from *Culex quinquefasciatus* mosquitoes from Zambia in Southern Africa (OW) increased the knowledge on the diversity and evolution of alphaviruses. The genomic and phylogenetic analysis from Torii *et al.* introduced the MWAV in the same phylogenetic group as EILV and TALV. The insect antigenic complex shared the same ancestral virus with the WEE complex and was placed between TROV and AURAV as previously demonstrated, Fig. 3 (Torii *et al.*, 2018, Nasar *et al.*, 2012, Hermanns *et al.*, 2017).

Up to now, it is still elusive if insect specific viruses are present in the NW. To date, only three insect-specific alphaviruses were isolated and geographically belonging to the OW. Accordingly, to the new phylogenetic analysis proposed by Torii *et al.*, (Torii *et al.*, 2018), the insect specific *Alphavirus* group is positioned closer to the NW group, Fig. 2. In contrast to the basal position near the WEE complex proposed by Nasar *et al.* and Hermanns *et al.* (Hermanns *et al.*, 2017, Nasar *et al.*, 2012). The geographic origin of insect-specific alphavirus remains to be elucidated as geographic introductions of alphaviruses may have repeatedly occurred. Further studies are highly necessary to understand the evolution of alphaviruses between NW and OW and as well as the loss (or gain) of abilities to infect terrestrial animals. Much like in Forrester *et al.* reporting on aquatic alphaviruses for the first time (Forrester *et al.*, 2012), the works recently performed with insect-specific alphaviruses also suggests that the ancestor of alphavirus are probably of aquatic origin (Hermanns *et al.*, 2017, Nasar *et al.*, 2012, Torii *et al.*, 2018). The position of the Salmon pancreatic disease virus (SPDV) at the basal position of the trees defined by midpoint rooting reinforces this theory. As well as the phylogenetic placement of the Southern elephant seal virus (SESV) isolated from seal louse (*Lepidophthirus macrorhini*). The mode of transmission between aquatic alphaviruses is not well known and it has been suggested that aquatic alphaviruses diverged in a distant past from the mosquito-borne viruses. In an evolutionary perspective of view, Forrester *et al.* assumed that if the global movement of ancestral alphaviruses occurred before the age of frequent human transoceanic travel, it seems likely that zoonotic hosts like birds and arthropods were responsible for the alphavirus movement between hemispheres (Forrester *et al.*, 2012).

1.3. Impact and expansion of encephalitic alphaviruses in the society

Encephalitic alphaviruses are the most pathogenic from the *Alphavirus* genus as they can cause lethal encephalitis in the host. Nearly almost the encephalitic alphaviruses are from the NW and efficiently establish enzootic cycles in sylvatic habitats and between domestic animals such as equids in rural habitats (Ferro *et al.*, 2003, Salas *et al.*, 2001). The enzootic strains usually circulate in continuous cycles using rodents in forest and swamp habitats (Gonzalez-Salazar *et al.*, 2003). When humans enter in contact with sylvatic habitats spill over may occur directly to humans involving transmission mediated by mosquitoes. In contrast, the epizootic strains efficiently use equids as amplification host in rural habitats to intermediate the transmission to humans mediated by mosquitoes. Horses, donkeys and mules infected with epizootic strains usually develop serious illness and produce high viremia levels that are enough to infect competent mosquitoes (Wang *et al.*, 2001). Competent mosquitoes are usually from the genus *Ochlerotatus* (such as subgenus *Aedes*), *Psorophora* and *Culex*.

The clinical symptoms caused by human pathogenic viruses from the VEE complex are easy confused and mis-diagnosed with other arboviral disease such as dengue. It has been suggested that approximately 10 % of cases diagnosed as dengue in neotropical cities are probably cases of VEEV (Aguilar *et al.*, 2011, Forshey *et al.*, 2010, Manock *et al.*, 2009). Usually VEEV infection acute phase last for 3 to 5 days and cause fever, headache, chills, tremors, malaise, diarrhea, nausea, vomiting, arthralgia, prostration, retro-orbital pain. In few patients the VEE illness progress for neurological complications with symptoms such as convulsions, disorientation, drowsiness, mental depression and some cases death (Quiroz *et al.*, 2009, Weaver *et al.*, 1996, Weaver *et al.*, 2004). Neurological sequelae are more common in children and include recurrent seizures, motor impairment, psychomotor retardation and behavioural disorders (Bowen *et al.*, 1976). Rare symptoms such as haemorrhages were reported in Panama and Peru involving the enzootic subtype ID (Johnson *et al.*, 1968, Vilcarrromero *et al.*, 2009). Abortogenic effects were in the later 1980s associated with VEEV after a pregnant human exposed to the vaccine strain TC-83 VEEV lost her fetus (Casamassima *et al.*, 1987).

The human case-fatality rate of VEEV was estimated in 1974 as less than 1 % and severe encephalitis with neurological complications may affect about 4 % to 14 % of the cases, mainly children. During epizootics equine fatality rates were estimated between 19 % to 83 % (Johnson & Martin, 1974, Weaver *et al.*, 2004). The apparent misdiagnosed

cases with arboviral disease such as the almost 10 % of dengue cases may create bias in the determination of human fatality rate. The human fatality rate is probably higher than 1 % and varies with outbreaks. For example, VEE cases registered in Panama between 1961 and 2004 resulted in 5 % of human fatal cases (Quiroz *et al.*, 2009). A single case of VEE illness is a significant economic burden regarding treatment and long-term care (Armstrong & Andreadis, 2013).

Members of the VEE complex are extensively distributed through the Americas where outbreaks have been reported in Venezuela, Colombia, Peru, Ecuador, Bolivia, Costa Rica, Nicaragua, Honduras, El Salvador, Guatemala, Panama, Mexico and USA (Weaver & Reisen, 2010). The VEE complex consists of six distinct subtypes and a total of 13 variants. The subtype I consist of five variants (*i.e.*, IAB, IC, ID, IE and IF). The variant AB from subtype I resulted from the fusion of variant A and B that are difficult to distinguish. Only the variant *Everglades virus* is included in the subtype II. The subtype III includes four variants and each one represented by a single species, *i.e.* IIIA (*Mucambo virus*), IIIB (*Tonate virus*), IIIC (*Mucambo virus* strain 71D1252) and IIID (*Mucambo virus* strain V407660). The subtypes IV, V and VI are represented by the following variants respectively; *Pixuna virus*, *Cabassou virus* and *Rio Negro virus*.

1.3.1 Epizootic circulation of VEEV

Since 1920s VEE illness in equids was recognized as “peste loca” in northern South America, Central America (Mexico), USA (Texas). The VEEV epizootics were initially associated with strains from **subtype I variants AB** and **C** which have caused major outbreaks involving equines and humans. In 1935 the first registered outbreak involving equids was in Colombia (Magdalena River Valleys) and expanded to Venezuelan in 1936. In 1938 the virus was isolated for the first time from a post- mortem brain specimen of horse with encephalitis from an outbreak in Venezuelan (Beck & Wyckoff, 1938, Kubes & Rios, 1939). The **subtype I variant AB** caused several epizootics along the Pacific coast, occurring in Peru during the period 1942 to 1946 and reported in Trinidad in 1943 (Johnson & Martin, 1974, Lord, 1974). In 1952 in Espinal, Colombia, an epidemic registered with neurological complications and 0.7 % of human case fatality (Sanmartin-Barberi *et al.*, 1954). The human fatality rate in the 1962 outbreak in Colombia was the same as the 1952 outbreak and extended to Venezuela where it was registered a fatality rate of 0.7% and 4.1% of neurologic complications in humans. During 1967 to 1968 Colombia was again affected by an epizootic (Aguilar *et al.*, 2011), Fig. 4.

Immediately after, in 1969 the Ecuador was devastated by a large outbreak caused by **subtype I variant B** and resulted in 1% of human fatality rate and 20 000 equid deaths

(Gutierrez *et al.*, 1975). The outbreak started near Guatemala border between 1969-1970 and followed the path of Gulf coast, by expanding to in Ecuador, El Salvador, Mexico and eventually reached southern Texas (Hinman *et al.*, 1971, Sudia *et al.*, 1971, Aguilar *et al.*, 2011). In 1969, Texas reported an equine fatality rate of 71% and 110 human cases were reported (Aguilar *et al.*, 2011). Probably the origin of the outbreak in Texas was the use of formalin inactivated virus (subtype IAB) for vaccines that resulted



Figure 4. Locations of Venezuelan equine encephalitis virus outbreaks in the Americas (Weaver *et al.*, 2004). The location of all major VEE outbreaks in the Americas are highlighted in the map (regions shaded purple and labelled with text). The date (year) and respective VEE virus subtypes of the outbreak are shown. Symbols represent locations from which enzootic VEEV-complex virus have been isolated, with enzootic subtypes indicated in parentheses.

in incomplete inactivation of the virus (Kinney *et al.*, 1992a). The isolation of VEEV subtype I variant AB from humans that developed VEE illness upon vaccination with the

formalin inactivated virus was previously reported in 1954 (Sutton & Brooke, 1954). By the end of 1972 the outbreaks were controlled with the administration of the live attenuated TC-83 vaccine in equids that replaced the previous inactivated vaccine (Kinney *et al.*, 1992a, Aguilar *et al.*, 2011, Weaver *et al.*, 1999). A recent study reported the crypt circulation of the subtype I variant AB after 8 years of the 1973 outbreak in Venezuela (Medina *et al.*, 2015). An outbreak associated with equine epizootics caused by subtype I variant C followed a hiatus of 19 years of inactivity, resulting in re-emergence of VEEV in Venezuela in 1992 (Oberste, Fraire, *et al.*, 1998). More outbreaks were reported in 1993 involving Venezuela and Mexico (Oberste, Fraire, *et al.*, 1998), and in 1995 involving Venezuela and Colombia, Fig. 4.

Studies have suggested that the strain **subtype I variant C** that affected Venezuela from 1992 to 1993 resulted from the progenitor enzootic subtype I variant D (Kinney *et al.*, 1992b, Young & Johnson, 1969, Rico-Hesse *et al.*, 1988). Probably the equine virulence and amplification capacity of the subtype I variant C resulted antigenic shift due to amino acid substitutions that occurred in the E2 glycoprotein of the subtype I variant D (Powers *et al.*, 1997, Anishchenko *et al.*, 2006, Greene *et al.*, 2005, Wang *et al.*, 1999). During 1995 the subtype I variant C re-emerged again in the same area of the outbreak 1962 to 1964 in Venezuela and Colombia (Weaver *et al.*, 1996). The evolution of the epizootic subtype I variant AB and C resulted probably from circulation in cryptic transmission cycles, maintenance in animal populations involving mainly equines but rodents can also be involved, their existence as minor population within enzootic VEEV populations, emergence from an enzootic VEEV progenitor population, periodic emergency via mutations from enzootic strains. Lastly, and even for outbreaks after 1970 should be considered the incomplete inactivation of vaccine preparations that were used for vaccination and resulted in viremia and circulation of *wt* VEEV (Weaver *et al.*, 1999, Aguilar *et al.*, 2011, Johnson & Martin, 1974, Sutton & Brooke, 1954), Fig. 4.

1.3.2 Enzootic circulation of VEEV

The variants D, E and F from subtype I and subtypes II- VI are pathogenic to humans, clinically indistinguishable from epizootic strains and can be fatal, however these strains are not associated with major equine outbreaks or epidemics (Oberste, Weaver, *et al.*, 1998, Johnson *et al.*, 1968). The variants ID, IE and IF are known to present an enzootic circulation pattern and are not able to use equids as amplification hosts. The enzootic cycle of ID, IE, subtype II and IIIA usually involves a transmission cycle between rodents and insects, the subtype IIIB involves a transmission cycle between birds and insects (Barrera *et al.*, 2002, Monath *et al.*, 1980, Young *et al.*, 1969), Fig. 4.

For equids the **subtype I variant D** is avirulent, but human cases are known. The subtype I variant D circulates in northern South America namely in Colombia, Venezuela, Panama and Peru and now, since initially reported during 2005, in Bolivia (Aguilar *et al.*, 2009), Fig. 4. This serotype consists of **three lineages** which are; **(1) Colombia/Venezuela** identified in the 1970s and give rise in 1998 to the epizootic strain **(2) Panama/Peru** which is progenitor of **(3) Bolivia/Peru** (recent in Bolivia and Madre de Dios) (Aguilar *et al.*, 2009). The lineages are restricted geographically and may reflect limited dispersal potential of the rodents reservoir hosts and mosquito vectors (Weaver *et al.*, 2004).

In Colombia, the circulation of enzootic VEEV in urban areas, outside sylvatic habitats, has been suggested because vectors of the enzootic subtype I variant D were found in human habitations. Cases occurred prior to the identification in 1970s of the circulation of the subtype I variant D lineage Colombia/Venezuela. The subtype I variant D from Colombia belongs to the lineage of the ancestor that originated the epizootic subtype I variant AB and variant C, as well as the subtype I variant D lineage Panama/Peru (Wang *et al.*, 1999, Weaver *et al.*, 1999, Weaver *et al.*, 1992, Powers *et al.*, 1997). Probably the subtype I variant D lineage Colombia/Venezuela, caused epizootic viruses through mutations of the E2 envelope glycoprotein (Powers *et al.*, 1997, Weaver *et al.*, 1999, Wang *et al.*, 1999, Weaver *et al.*, 1992). Endemic Colombian VEE activity has been continuously documented. Human cases continued to be reported in Colombia as in 2003 in the middle Magdalena Valley (Ferro *et al.*, 2003). A most recent outbreak occurred in 2008 affecting 13% of the resident population in the village probably by local enzootic transmission (Aguilar *et al.*, 2011).

Next to Colombia, in the 1960s endemic VEE was reported in Mexico and in Panama. In Mexico probably was related with **subtype I variant E**, but in Panama circulation of subtype I variant D was detected in 1961. In 1967 a direct spill over from sylvatic habitats to humans was reported when US soldiers enter in the forest, but no fatal cases were reported. VEEV subtype I variant D was isolated from two of these soldiers and from sentinel hamsters suggesting the area as a focus of transmission (Young & Johnson, 1969, Franck & Johnson, 1970, Johnson *et al.*, 1968). No equine cases were reported during these outbreaks probably because subtype I variant D is avirulent and do not amplify efficiently in equines (Walton *et al.*, 1973, Srihongse *et al.*, 1967, Sanchez *et al.*, 1984). During the period 1960s to 1970s VEE illness in humans were continuously reported in Panama with symptoms usually characterized by fever, headaches, retro-orbital pain and tremors, in more severe cases neurological complications are involved,

mainly in children (Rossi, 1967, Quiroz *et al.*, 2009). The fatality rate reached 5%, higher than epizootic strains that amplify efficiently in equines (Johnson & Martin, 1974).

VEE is endemic through the amazon basin of Peru accounting on average with 10 to 15 cases per year (Forshey *et al.*, 2010). In 1969 one epizootic in Peru was associated with subtype I variant D (subtype III variant C was also cocirculating) and resulted in almost 600 equine deaths (Scherer & Chin, 1983, Scherer & Anderson, 1975, Scherer *et al.*, 1975, Aguilar *et al.*, 2011). In eastern Peru VEE cases were hardly detected due to the presence of Dengue and are continuously reported since 1995. The first direct evidence of human VEEV in Peru come in 1990s when ID was first identified in Pantoja and later in Iquitos (Watts *et al.*, 1998, Watts *et al.*, 1997). In Iquitos the subtype I variant D was isolated continuously from patients but affect mainly children and no neurological complications were reported (Forshey *et al.*, 2010, Aguilar *et al.*, 2004).

It has been suggested that during the period 1993 – 1995 in Peru (Iquitos) the ID Colombia/Venezuela lineage had been replaced by the Panama/Peru genotype (Watts *et al.*, 1998, Watts *et al.*, 1997, Aguilar *et al.*, 2004). The lineages ID Colombia/Venezuela and the ID Panama/Peru were cocirculating in eastern Peru (Iquitos) between 1993 and 2002 probably through an urban cycle. The Panama/Peru lineage represented most of the human isolates during this period. Several explanations for these high number were postulated as: high levels of circulation Panama/Peru lineage, production of higher titer human viremia, higher virulence and efficiency of transmission than the Columbia/Venezuela lineage. The Columbia/Venezuela sublineage continues circulating in Iquitos (eastern Peru) but may infect people at lower rate than the Panama/Peru lineage (Aguilar *et al.*, 2004). Recent outbreaks have been occurring since 2004 in Peru. Since 2005, Porto Maldonado and Madre de Dios are currently affected by the subtype I variant D (Morrison *et al.*, 2008). An outbreak in 2006 from Yurimaquast with the subtype I variant D Panama/Peru lineage originated a fatality rate of approximately 20% (Vilcarromero *et al.*, 2010, Vilcarromero *et al.*, 2009). During 2005-2006 in Iquitos urban transmission of the enzootic subtype I variant D was related with the illness of near 100 patients (Aguilar *et al.*, 2009).

In Ecuador is known that between 1975 to 1977 the lineage southwestern Colombia/Ecuador subtype I variant D was circulating in this area (Powers *et al.*, 1997).

A new lineage of the subtype I variant D was identified in Bolivia outbreak 2005 to 2007 (Aguilar *et al.*, 2009). Currently, the subtype I variant D that is circulating in Bolivia is closely related to the Peru lineage and distinct from the lineages found in Panama,

Colombia, Venezuela and Ecuador. The subtype I variant D lineage Bolivia/Peru can also induce neurological complications (Aguilar *et al.*, 2009).

The VEEV **subtype I variant E** is endemic in large coastal strips, both in Pacific and Gulf coasts of Mexico (Aguilar *et al.*, 2011). The Pacific coast lineage is also dispersed in the Pacific coast of Guatemala, Honduras, Nicaragua. The limited dispersal of the subtype I variant E is probably related with the limited mobility of rodents and insects (Adams *et al.*, 2012). **Three lineages** are known and have been maintained independently; **(1) Pacific coast of Mexico, (2) Central America, Gulf/Atlantic coasts of Mexico, and (3) western Panama** (Aguilar *et al.*, 2011). Epizootics have already been reported for the enzootic strain subtype I variant E. Variable virulence and neurovirulence have been observed for equines (Gonzalez-Salazar *et al.*, 2003). However, this variant does not amplify efficiently in equids (Walton *et al.*, 1973, Srihongse *et al.*, 1967, Sanchez *et al.*, 1984). In Mexico a survey from 1961 detected antibodies in humans suggesting that residents from tropical Atlantic lowlands of Southeastern Mexico were exposed to the virus probably even before 1961, (cited in Aguilar *et al.*, 2011). Later, it was suggested that VEE clinical cases occurred before and involving deaths and development of neurological sequels [De Mucha Macias (1963) (cited in Aguilar *et al.*, 2011)].

In 1963 VEEV strains were detected in mosquitoes and sentinel hamsters collected from a sylvatic habitat bordering a lagoon off the Gulf of Mexico coast at the village Sontecomapan in Veracruz (Scherer *et al.*, 1964). Accordingly, it was proposed that endemic VEEV strains were circulating in sylvatic habitats in southeast of Mexico and eventually could represent a health risk for inhabitants of Mexico and perhaps to United States due to the migrating birds. In 1965 a fatal human case was associated with VEEV reported in Veracruz region and approximately one year later an equine outbreak occurred with a fatality rate of approximately 30%. The lack of vaccination in Mexico suggested that an enzootic subtype IE was involved in this outbreak (cited in Aguilar *et al.*, 2011). In western panama 1968 subtype IE focus was identified and *Culex* as the enzootic vector covering the Canal zone of Panama in 1971 (Grayson & Galindo, 1968, Galindo & Grayson, 1971).

After a hiatus of 20 years, a new case of VEEV was reported in 1991 in Tabasco (Mexico) when the equine virulent subtype IE strain emerged in Mexico. Equine epizootics have been reported in the pacific coast of Mexico between 1993 – 1996 for the Pacific Mexican lineage (Oberste *et al.*, 1999, Aguilar *et al.*, 2011, Gonzalez-Salazar *et al.*, 2003) suggesting that the equine virulent strain emerged or was reintroduced recently in southern Mexico. The outbreak did not spread beyond southern Mexico because the

strain did not amplify efficiently in equids. In Venezuela and Colombia 1995, it was reported 3% of neurologic complications and 0.3% fatality rate in humans. (Oberste, Fraire, *et al.*, 1998, Centers for Disease & Prevention, 1995, Rivas *et al.*, 1997, Weaver *et al.*, 1996). In the Pacific coast the subtype IE acquired better adaptation to infect *Aedes taeniorhynchus* mosquitoes that have preference for large mammals and present increased dispersal. This adaptation probably occurred due to a mutation in E2 glycoprotein (Ser218Asn) (Brault *et al.*, 2004). In addition, recent studies have shown the susceptibility of *Culex (melanoconion) taeniopus* to the subtype I variant E (Deardorff & Weaver, 2010, Aguilar *et al.*, 2011). Evidences from a serological study conducted between 2000 -2001 in Chiapas supported that VEEV subtype I variant E is in continuing circulation. Probably rats and sentinel hamsters are reservoirs of the subtype I variant E in Mexico. The indications of a long-term enzootic and endemic VEEV circulation in this region highlights the risk for VEE illness in equines and humans (Estrada-Franco *et al.*, 2004, Deardorff & Weaver, 2010, Aguilar *et al.*, 2011).

Relatively to **subtype I variant F** (*Mosso das Pedras virus*) the virulence for equids is unknown but is circulating in sylvatic habitats in Brazil. Species from subtype II to VI were described to present an enzootic transmission pattern (Forrester *et al.*, 2017). The **subtype II** (*Everglades virus*) can cause VEE illness in humans (Calisher *et al.*, 1980). The subtype II is not virulent for equids and circulates with an enzootic pattern in sylvatic habitats from Southern Florida using wild rodents as the main reservoir (Chamberlain *et al.*, 1964, Chamberlain *et al.*, 1969). However, recently it was proposed the involvement of birds as a host involved in the maintenance of the virus (Forrester *et al.*, 2017, Bigler *et al.*, 1974). The **subtype III variant A** is not virulent for equine and regarding **subtype III variant B, variant C** and **variant D** the virulence for equine is unknown. However, Mucambo virus (BeAn 8) inoculated in horses caused fever, leukopenia and viremia (Shope *et al.*, 1964). The **subtype III variant C** was isolated in 1971 from mosquitoes in Peru (Iquitos) (Scherer & Anderson, 1975, Scherer & Chin, 1983). In 2002 the isolation of the virus from a sentinel hamster suggested that the subtype III variant C is still circulating in the Peruvian Amazon (Aguilar *et al.*, 2004). During the period 1995 to 2002 the subtype III variant C was circulating in enzootic habitats in Iquitos area, together with the new identified subtype III variant D isolated from mosquitoes and spiny rats (Turell *et al.*, 2006, Turell *et al.*, 2000, Yanoviak *et al.*, 2005). The subtype III variant C failed to be isolated from human's patients suggesting that is not transmitted to humans, or that the viremia is low to be detected or that the virus does not cause disease. New identified **variant D from subtype III** was isolated from human patients that presented fever, chills, and malaise during the period 1995 – 2002 (Aguilar *et al.*, 2004). Also, it is unknown the virulence for equines of the **subtype IV** (*Pixuna virus*) found in enzootic circulation in

Amazon region of Brazil and in Argentina (Shope *et al.*, 1964), **subtype V** (*Cabassou virus*) and **subtype VI** (*Rio Negro virus*) also found in Argentina circulating in enzootic habitats (Pisano *et al.*, 2014) (Weaver *et al.*, 2004).

1.4. Impact of the global expansion of CHIKV in the society

One of the characteristics of alphaviruses belonging to the OW is related to the clinical sign of arthritis in the patient. The OW alphavirus most pathogenic to humans and thus of most concern, is CHIKV, a member from the SF antigenic complex. The disease caused by CHIKV has a low fatality rate of approximately 0.1% (Renault *et al.*, 2008), but is very often leading to acute and chronic disability affecting the quality of life of the patient, together with a significant economic impact for society (Couturier *et al.*, 2012, Gerardin *et al.*, 2008, Schilte *et al.*, 2013). CHIKV has recently emerged as a global pathogen, with about 50% of the world's population at risk of infection (Powers, 2019, Powers & Waterman, 2017).

The virus is mainly transmitted to humans through the bite of an infected female mosquito from the genus *Aedes*, in general *Aedes aegypti* or *Aedes albopictus*. After an incubation period of usually 3 to 7 days (range, 1-12 days) the patient become symptomatic and manifest acute onset of symptoms but mainly fever (usually more than 39 °C) and polyarthralgia (Centers for Disease Control and Prevention [CDC], 2018). Symptoms such as headache, myalgia, arthritis, conjutivitis, nausea/vomiting, or maculopapular rash have also been registered to occur during the acute phase (CDC, 2018). The acute symptoms typically last from few days to a couple of weeks, approximately within 7-10 days (CDC, 2018). Some patients undergo to a chronic phase which can last up to months or even years and is characterized mainly by arthralgia, arthritis, joint pain and myalgia (Chow *et al.*, 2011, Petitdemange *et al.*, 2015). Although, the maternal-fetal transmission can occur intrapartum, leading to high rates of infant morbidity (Gerardin *et al.*, 2008, Gerardin *et al.*, 2014), studies have provided indications that transmission does not occur through breastfeeding because CHIKV have not been found in breast milk (CDC, 2018).

Recently, symptoms have been shown to expand to encephalitis (Nelson *et al.*, 2014) and can promote severe neurologic sequels in the patient (Silva & Dermody, 2017). Neurologic manifestation can include encephalitis, myelopathy, peripheral neuropathy and myopathy (Brizzi, 2017). Unusual clinical symptoms of this self-limiting disease were reported during the recent La Reunion outbreak, including hepatitis, autoimmune neurological pathologies (Guillain-Barré), cardiologic manifestations and deaths (Sang *et al.*, 2008, Rajapakse *et al.*, 2010).

Most likely the first epidemics were registered during 1779 in Batavia – Dutch East Indies (present day Central Jakarta, Indonesia) and in Cairo (Egypt). At that time the disease was designed as “Knuckle fever” and “Knee trouble” respectively (Carey, 1971), Fig. 5. However, the first well recorded CHIKV epidemic occurred from 1952 to 1953 in Tanzania (then Tanganyika) along the coastal plateaus of Mawia, Makonde and Rondo where human habitats were infested with *A. aegypti* and *Culex fatigans* (Lumsden, 1955, Weaver & Forrester, 2015, Gudo *et al.*, 2016). Symptoms were described as a “very sharp onset of crippling joint pains, severe fever, and eventually the conspicuous rash” (Ross, 1956, Weaver & Forrester, 2015) and initially was misdiagnosed with dengue virus (Carey, 1971).

The epidemic from Tanzania (1952 – 1953) is responsible for the origin of the name Chikungunya, Fig. 5. The name resulted from the “Makonde” world which means the disease “which bends up” the joints and describes the posture of the patient due to extreme joint pain (Ross, 1956). The name derives from the root verb *kungunyala* which means to dry up or become contorted. However, according to some Makonde health workers from the site of the first epidemic of chikungunya (northern Mozambican coastal town of Monçímboa da Praia) recognized that the correct name is chingwingwinda derived from gwingwindar (meaning to “bend up” and described the characteristic joint flexion) and the local remedy used was boiled cassava leaves rubbed into the affected joints (Gudo *et al.*, 2016).

Then, CHIKV extended to Uganda (Weinbren, 1958), and to many parts of sub-Saharan Africa, where indications were founded about CHIKV circulating in sylvatic/enzootic CHIKV cycle (Coffey *et al.*, 2014). In 1964 a serologic survey in Zimbabwe detected antibodies in non-human primates and evidences that vervet monkeys were competent amplification hosts through mosquito transmission (Paterson & McIntosh, 1964). CHIKV outbreaks extended to Asia during 1950s and 1960s (Weaver & Forrester, 2015), Fig. 5, and more recently disseminated to Europe and Americas with the presence *A. albopictus*.

Currently, autochthonous vector-borne transmission of CHIKV are already reported at least in 106 countries/territories. In sub-Saharan, 26 countries/territories are currently affected, and cases were already reported in Gulf countries such as Yemen and Saudi Arabia. India, China, and all most South East Asian countries and Pacific/Oceania countries have also reported the disease. In Europe cases were already reported in France, Italy and Spain. Recently, the virus has spread to the Americas and currently affects 46 countries/ territories (Leta *et al.*, 2018). It has been already reported that

CHIKV re-emerged in many countries with a gap of 7 to 20 years (Powers *et al.*, 2000), Fig. 5.

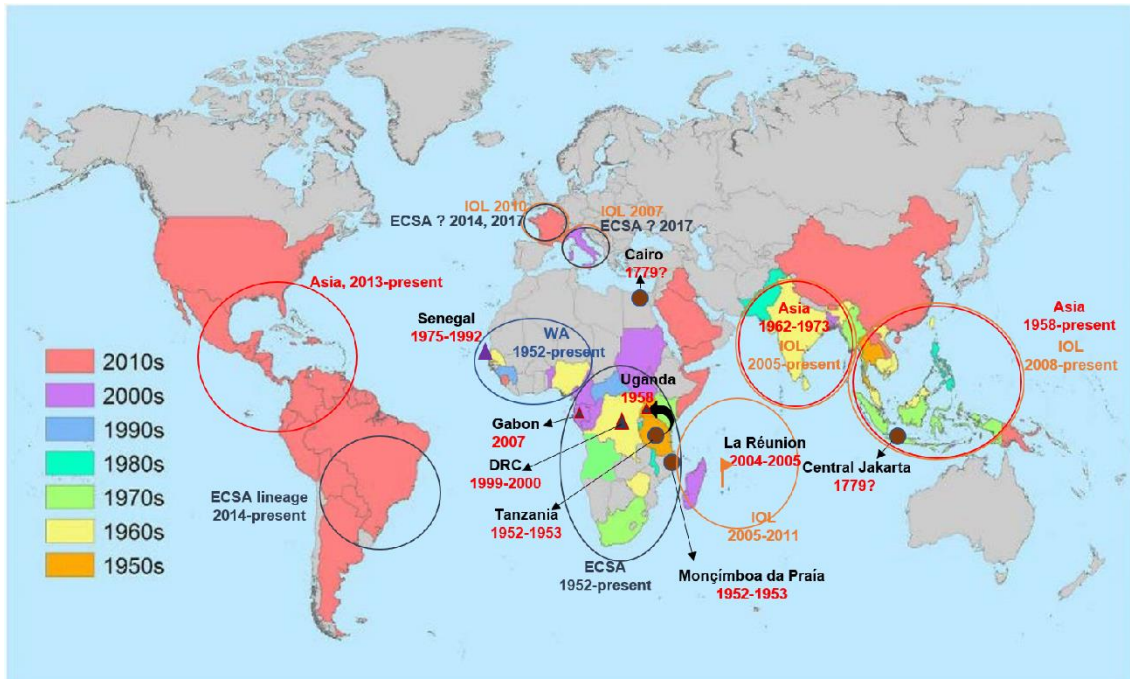


Figure 5. Origin, spread, and global distribution of CHIKV outbreaks. Country colours correspond to the decade of the first reported identification of the local transmission of CHIKV by either serological, molecular, or virological detection methods. (Image adapted from Nicole Lindsey in Powers 2018 (Powers, 2018)). The map shows the African origins of enzootic chikungunya virus strains and the patterns of emergence and spread of the Asian lineage and Indian Ocean lineage (IOL) of the virus during epidemics since the 1950s, based on phylogenetic studies. ECSA denotes East/ Central/ South Africa lineage. Adapted from centers for disease control and prevention (CDC) and (Weaver & Lecuit, 2015, Powers, 2018).

In a very distant past, CHIKV has evolved independently from his sister ONNV (Weaver & Forrester, 2015, Khan *et al.*, 2002) and sequence analysis have indicated that the virus was originated in Africa. CHIKV likely diverged from a common ancestor that existed within the last 500 years ago and segregated in two clades (Volk *et al.*, 2010). One of the clades is the West Africa (WA) lineage and the other is the common ancestor of the two lineages East/Central/South Africa (ECSA) and Asian, Fig. 5 (Powers *et al.*, 2000, Schuffenecker *et al.*, 2006, Volk *et al.*, 2010, Powers & Logue, 2007). The recent analysis from Indian Ocean and Indian strains suggested that the IOL sub-lineage forms a monophyletic group descendent from the ECSA lineage (Volk *et al.*, 2010).

Concerning the ECSA and Asian lineages it has been proposed that they have split from a common ancestor during the last 150 years, between 1879 and 1956 and since then continued to circulate (Volk *et al.*, 2010). Probably the Asian lineage splitting into two clades, the Indian and Southeast Asian lineages, the first probably went extinct in 1973

and the second is possible to still circulating (Volk *et al.*, 2010, Arankalle *et al.*, 2007, Lahariya & Pradhan, 2006). Recently, the IOL lineage appears to have diverged from the ECSA lineage most likely within the last 70 years ago. Thus, several phylogenetic studies contributed to sort the division of chikungunya in to three lineages and one sub-lineage, (1) West Africa (WA) lineage, (2) East/Central/South Africa (ECSA) lineage, (2.1) Indian Ocean (IOL) sub-lineage and (3) Asian lineage, Fig. 5.

1.4.1 Chikungunya West African (WA) lineage

The WA lineage have been associated mainly with sylvatic/enzootic transmission and small focal outbreaks of human disease in countries located in western Africa (Powers & Logue, 2007). The WA genotype consists of isolates from Senegal and Nigeria. The identification of this enzootic lineage resulted from the sequence analysis of CHIKV isolates from Senegal during 1975 and 1992 (Diallo *et al.*, 1999, Powers *et al.*, 2000, Volk *et al.*, 2010). It has been suggested that CHIKV vector are strain specific because studies have shown that different populations of *Aedes* in Senegal have distinct susceptibilities to different CHIKV strains (Diallo *et al.*, 1999).

Studies performed in eastern Senegal reported a pattern of periodic major amplification due to the infection of sylvatic/enzootic vectors (Diallo *et al.*, 1999). The major sylvatic vectors of CHIKV were identified from a study in the Kedougou area, and in order of importance were *Aedes furcifer*, *Aedes luteocephalus* and *Aedes taylori* (Diallo *et al.*, 1999, Jupp & Kemp, 1996). The only domestic vector identified from this works was *A. Aegypti* but at the time the works suggested that *A. aegypti* has never been involved in an epidemic in that region. Serological evidences supported the involvement of non-human primates as amplifying hosts and suggested that other vertebrates could also serve as reservoir hosts. However, it is still unknown if the amplification of the virus occurs due to a stable enzootic circulation or if represents local extensions or reintroductions of CHIKV. The risk of development of enzootic cycles are exemplified by the introduction of non-human primates in new habitats, such as African green monkeys that were introduced in several Caribbean Islands and Latin America (Diallo *et al.*, 1999), Fig. 5.

1.4.2 Chikungunya East/Central/South Africa (ECSA) lineage

The ECSA lineage is original from Africa and circulates mainly through an enzootic cycle within sylvatic habitats. The name ECSA was attributed to this lineage based on its initial detection that occurred mainly in East, Central, and South Africa. Within the sylvatic habitats the transmission involves mainly wild non-human primates as vertebrate

reservoirs and primarily the mosquito vectors *A. furcifer* and *Aedes africanus* (Powers, 2015). Interestingly, it has been suggested that the main WA and ECSA lineages may overlap spatially in the enzootic cycle, at least on occasion (Volk *et al.*, 2010). The repeatedly spread of the ECSA lineage to new regions causes significant urban epidemics which are principally characterized by the absence of vertebrate reservoirs or sylvan transmission cycles outside Africa (Carey, 1971). Outside Africa the urban cycles usually occur associated with transmission through peridomestic mosquitoes *A. albopictus* or *A. aegypti*.

After a period of 39 years without reports of CHIKV cases in the Democratic Republic of Congo, two important outbreaks were then reported from 1999 and 2000 (Horwood & Buchy, 2015). During these two outbreaks caused by ECSA lineage of CHIKV resulted 50 000 cases (Pastorino *et al.*, 2004). Then, in 2007 an outbreak took place in Gabon (Horwood & Buchy, 2015) where the dengue virus- serotype 2 (DENV-2) was circulating. The epidemic from Gabon (2007) was named CHIKV/DENV outbreak and caused 20 000 cases in this year (Leroy *et al.*, 2009, Nkoghe *et al.*, 2012). During 2010 an outbreak of CHIKV/DENV was in circulation and almost 300 laboratory confirmed cases were reported (Nkoghe *et al.*, 2012).

Although, it was not the first episode of autochthonous transmission in Europe, the first episode in mainland France occurred in south-east during 2010 at the Var department in the city of Fréjus (Grandadam *et al.*, 2011). From this episode resulted two cases of autochthonous transmission that originated from one imported case from India (Rajasthan). Although, the strains from the three patients were related to strains from India within the ECSA lineage (Grandadam *et al.*, 2011), most likely the autochthonous transmission documented in France (2010) were caused by strains belonging to the IOL sub-lineage (Vazeille *et al.*, 2016, Grandadam *et al.*, 2011). Likely the same applies to the cases reported in Italy during 2007 and 2017 (Rezza *et al.*, 2007, Vazeille *et al.*, 2016, Lindh *et al.*, 2019).

A new introduction of a strain belonging to ECSA lineage occurred in October 2014 in France (Montpellier) and the sequencing results from two patients revealed that the strain from the ECSA lineage has harbouring an adaptative mutation in the glycoprotein E1 (Delisle *et al.*, 2015). The ECSA strain that was involved in the small outbreak in France (2014) related to *A. albopictus* transmission was imported from Cameroon. In the same year occurred the first introduction of an ECSA lineage into Americas, more precisely in Brazil, (Nunes *et al.*, 2015) probably due to adaptation of the virus to the *A. albopictus* vector which is abundant in temperate regions of America. This was considered as a risk

factor to the spread of CHIKV to US where *A. albopictus* is also abundant in the temperate regions (Kraemer *et al.*, 2015), Fig. 5.

The independent adaptation to *A. albopictus* of CHIKV belonging to the ECSA lineage has probably occurred in four different occasions (de Lamballerie, Leroy, *et al.*, 2008, Tsetsarkin *et al.*, 2007, Vazeille *et al.*, 2016). This include the adaptation to *A. albopictus* that occurred for the ECSA lineage during the outbreaks in Cameroon (2006) and in Gabon (2007) (Peyrefitte *et al.*, 2008, Vazeille *et al.*, 2016, de Lamballerie, Leroy, *et al.*, 2008), Fig. 5. However, the first well characterized cases of independent adaptation of CHIKV to *A. albopictus* occurred in outbreaks from La Réunion Island between 2004 to 2005 (Schuffenecker *et al.*, 2006) and from India during 2008 (Santhosh *et al.*, 2009) that were caused by the IOL sub-lineage (Vazeille *et al.*, 2016, Volk *et al.*, 2010).

The first emergence of ECSA lineage outside Africa occurred probably within the last 150 years in Asia (Volk *et al.*, 2010). Independent evolution of this strain gave rise to a distinct lineage, known as Asian lineage, and which continues in circulation in Asia and is responsible for numerous outbreaks (Powers *et al.*, 2000, Volk *et al.*, 2010). More recently, during the past decade, another descendent from the ECSA lineage emerged as the IOL sub-lineage.

1.4.3 Chikungunya Indian Ocean sub-lineage (IOL)

The IOL sub-lineage was already responsible for many outbreaks in Indian Ocean basin and Southeast Asia (Volk *et al.*, 2010). Studies from Dash *et al.* (Dash *et al.*, 2007) reported that sequence analysis of some of the strains among the samples collected during the outbreak in India from 2006 were clustered into the ECSA genotype. This evidence, from the outbreak in India that caused more than 1.3 million cases in 12 states of India, revealed that IOL genotype derives from the ECSA lineage (Dash *et al.*, 2007), Fig. 5.

A very interesting molecular signature of the Indian Ocean outbreak genomes was the detection of an adaptative mutation (E1-A226V) in some of the CHIKV isolates (Schuffenecker *et al.*, 2006). The selection of adaptative mutations promoted the increased viral fitness in *A. albopictus* without affect replication in *A. aegypti* (de Lamballerie, Leroy, *et al.*, 2008, Schuffenecker *et al.*, 2006, Hapuarachchi *et al.*, 2010, Kumar *et al.*, 2008, Sam *et al.*, 2009). Such mutations include the E1-A226V but also required lineage-specific epistatic interactions of two other mutations, such as E1-98A and E2-211T (Tsetsarkin *et al.*, 2016), where E1 and E2 represent the glycoproteins E1 and E2 respectively. The epistatic mutations have been observed within the lineages

ECSA and WA and within the sub-lineage IOL, but until now were never reported in the Asian lineage (Coffey *et al.*, 2013). Thus, the Asian lineage has a constrained genetic capacity to adapt to *A. albopictus* (Tsetsarkin, Chen, Leal, *et al.*, 2011). The occurrence of independent adaptative mutation in response to a similar requirement of transmission by *A. albopictus* that was already observed in the WA lineage, ECSA lineage and IOL sub-lineage may be suggestive of a rare phenomenon of "evolutionary convergence" (de Lamballerie, Leroy, *et al.*, 2008).

In 2004 an outbreak caused by CHIKV strains from the ECSA lineage emerged in the region of Kenya (East Africa), first in Lamu and then in Mombasa (Chretien *et al.*, 2007, Powers & Logue, 2007). The occurrence of adaptative mutations to *A. albopictus* in these epidemic stains originated the new IOL sub-lineage (Schuffenecker *et al.*, 2006, Volk *et al.*, 2010). In 2005 the IOL sub-lineage then spread to Indian Ocean islands and affected La Réunion island where approximately 300 000 cases were reported (Schilte *et al.*, 2013, Gerardin *et al.*, 2008).

It is known that in La Réunion CHIKV adapted to *A. albopictus* probably through an adaptative mutation that occurred in the E1 glycoprotein (E1-A226V) (de Lamballerie, Leroy, *et al.*, 2008, Tsetsarkin *et al.*, 2007, Vazeille *et al.*, 2007, Volk *et al.*, 2010). Then, the new epidemic IOL sub-lineage extended to South and Southeast Asia affecting millions of persons and continued to diverge and spread through Asia. The mechanism used by the IOL lineage to spread through South Asia and Southeast Asia was mainly regulated by the vector-adaptative mutation (Arias-Goeta *et al.*, 2013, Tsetsarkin, Chen, Leal, *et al.*, 2011, Tsetsarkin, Chen, Sherman, *et al.*, 2011, Tsetsarkin *et al.*, 2014, Tsetsarkin *et al.*, 2009, Tsetsarkin *et al.*, 2007, Vazeille *et al.*, 2007). Some IOL strain retained the alanine at the position 226 (E1-226A) and were predominantly found in vector populations of *A. aegypti*. In some regions *A. aegypti* are uncommon and *A. albopictus* are the most predominant populations of competent insect vectors. As a mechanism of adaptation to *A. albopictus*, some IOL strains acquired a mutation from alanine to valine at position 226 in the glycoprotein E1 (E1-A226V). Interestingly, the *A. albopictus* -adaptive mutations do not affect CHIKV fitness in *A. aegypti* (Tsetsarkin & Weaver, 2011, Tsetsarkin, Chen, Sherman, *et al.*, 2011, Tsetsarkin, Chen, Leal, *et al.*, 2011).

The IOL strains are still in circulation in Asia along islands in Oceania. Recent evidences of co-circulation with the Asian lineage have been supported by recent outbreaks in the same regions where IOL strains are still circulating, such as in French Polynesia (AbuBakar *et al.*, 2007, Kawashima *et al.*, 2014, Lanciotti & Valadere, 2014, Ledermann *et al.*, 2014, Roth *et al.*, 2014, Nhan *et al.*, 2014), Fig. 5. Since 2005, India, Indonesia,

Maldivas, Myanmar and Thailand have reported large numbers of cases. This outbreak expanded very quickly and during 2005 reached Comoros, La Réunion, other islands in the Indian Ocean (Sang *et al.*, 2008), India, and other parts of Southeast Asia (Thiberville *et al.*, 2013, Staples *et al.*, 2009, Powers & Logue, 2007, Schwartz & Albert, 2010, Arankalle *et al.*, 2007). As an idea of the dimensions of the epidemic, in Comoros more than 5 000 cases were reported during 2005. The number of cases from Kenya and Comoros in early 2005 were approximately 250 000 (Powers, 2015). The island more affected was La Réunion where the first cases were imported from Comoros and in 2006 an estimation of 244 000 cases were reported, corresponding to approximately 30% of the population (Paquet *et al.*, 2006) and accounting with 237 deaths (Mavalankar *et al.*, 2007). In India the virus re-emerged after a hiatus of 32 years and caused at least 1.3 million cases by 2006 in 12 states of the country (Mavalankar *et al.*, 2007, Schuffenecker *et al.*, 2006, Dash *et al.*, 2007), Fig. 5. In southern Thailand a large outbreak affected approximately 50 000 people between 2009 and 2010 (Rianthavorn *et al.*, 2010).

The epidemic in Asia involving the IOL strains infected travellers that imported CHIKV into northern Italy (Rezza *et al.*, 2007) and Southern France (Grandadam *et al.*, 2011) though transmission association with *A. albopictus*. Adaptability of CHIKV to *A. albopictus* was one of the main factors that contributed to autochthonous transmission of CHIKV (Severini *et al.*, 2018). Autochthonous transmission of these new epidemic stains occurred afterwards in Italy (region of Ravenna with less than 300 cases) during 2007 (Rezza *et al.*, 2007), Southeast Asia during 2008 (Ng *et al.*, 2009) and then in south-eastern of France during 2010 (Grandadam *et al.*, 2011). This epidemic of CHIKV caused by IOL sub-lineage affected naïve regions and it was probably one of the largest epidemics with more than 6 million estimated cases (Thiberville *et al.*, 2013, Staples *et al.*, 2009). In temperate regions from Europe and US a significant number of imported cases from India to new regions were registered (de Lamballerie, Leroy, *et al.*, 2008, Grandadam *et al.*, 2011, Lanciotti *et al.*, 2007), Fig. 5.

1.4.4 Chikungunya Asian lineage

In similar fashion to the other lineages, the name of the Asian lineage derived from the region where was initially identified the genotype. The presence of the periodomestic/anthropophilic mosquitoes *A. aegypti* and *A. albopictus* within the urban cycles in Asia explains the higher frequency of CHIKV outbreaks in Asia than in Africa. This is mainly because these mosquitoes are in more close association with humans than *A. furcifer*, *A. luteocephalus*, *A. taylori*, and *A. africanus*.

It has been suggested that the CHIKV Asian lineage diverged from ECSA lineage within the last 150 years (between 1879 and 1956) and is still in circulation (Volk *et al.*, 2010). Phylogenetic studies have suggested the Asian lineage evolved by splitting in two clades. One of the clades descendent from the Asian lineage is known as the Indian lineage and that was extinct in 1973. According to phylogenetic studies on isolates from Asian outbreaks occurred between 1958 and 1973, the Asian lineage is a monophyletic clad basal to the ECSA lineage (Volk *et al.*, 2010, Powers *et al.*, 2000, Dash *et al.*, 2007). The first CHIKV isolate from Asia was during 1958 when an outbreak affected Bangkok and Thailand (Hammon *et al.*, 1960). CHIKV associated outbreaks were identified in Cambodia and India between 1961 and 1963. Previously, CHIKV seroprevalence was detected in India during 1954 (Hammon *et al.*, 1960). In Sri Lanka seroprevalence suggested that CHIKV or a close related virus had circulated many years before 1967 (Hermon, 1967). Interestingly, in Southeast Asia, a suspected epidemic occurred before 1879, more precisely in 1779 in Batavia – Dutch East Indies (present day, Central Jakarta, Indonesia), however without evidences to prove that corresponds to the Asian lineage of CHIKV in circulation during 1971 (Carey, 1971).

The Southeast Asian lineage is the other descendent clade from the Asian lineage and is likely to continue in circulation since even (Volk *et al.*, 2010, Arankalle *et al.*, 2007, Lahariya & Pradhan, 2006). During outbreaks occurred in Asia since 2005 many travels were infected with Asian or IOL lineages that were co-circulating in Asia including the Indian Ocean Basin (Lanciotti *et al.*, 2007). In 2011 the first CHIKV case was detected in the Pacific Island in the region of New Caledonia (Cao-Lormeau & Musso, 2014) and in 2012 some islands from Oceania were affected by CHIKV (Cao-Lormeau & Musso, 2014, Roth *et al.*, 2014). For the first time both Asian and ESCA lineages were detected during 2012 in Philippines, Cambodia, Papua New Guinea, Malaysia, Bhutan and in the Middle East (Yemen) (Zayed *et al.*, 2012). In 2013 it were reported 992 cases in Singapore, 127 cases imported to Australia and a new Asian genotype emerged in Yap (belonging to the Federal States of Micronesia) which caused more than 900 cases (Powers, 2015). Most likely, the genotype found in Yap was the same from Philippines and China, indicating the movement between East Asia to western Pacific (Lanciotti & Valadere, 2014), Fig. 5. Since then, the Asian lineage has been responsible for many of the outbreaks in these regions and according to CDC have already affected at least 10 of 22 countries and territories (Nhan & Musso, 2015).

Infected travels are the main source responsible for continuing co-circulation and spread of strains from ECSA and Asian lineages and from the IOL sub-lineage within the Indian subcontinent, Southeast Asia, and Oceania (Tsetsarkin, Chen, Leal, *et al.*, 2011, Sam *et*

al., 2009, Arankalle *et al.*, 2007, Dupont-Rouzeyrol *et al.*, 2012, Horwood *et al.*, 2013, Mombouli *et al.*, 2013, Wangchuk *et al.*, 2013, Wu *et al.*, 2013). The global expansion of CHIKV was definitively more evident when during 2013 the first cases of local transmission of CHIKV occurred in Americas, and which was caused by an Asian lineage imported from Southeast Asia or Oceania (Leparc-Goffart *et al.*, 2014, Lanciotti & Valadere, 2014).

This Asian CHIKV strain was identified in the Caribbean island of French St. Martin (Cassadou *et al.*, 2014) with 300 suspected cases in 2013 (Fischer *et al.*, 2014). Probably this Asian strain was closely related to strains from East Asia and Yap (Fischer *et al.*, 2014, Powers, 2015). Until 2015, the virus reached all Central America countries, most of South America and recently North Americas. For example, in Florida 11 autochthonous cases were reported during the summer of 2014 (Weaver & Forrester, 2015). Since 2013, more than one million cases have been reported from Caribbean islands and Latin American countries, with Colombia, Brazil and Bolivia having the largest burden (Leta *et al.*, 2018). Probably, accordingly to CDC, more than two million suspected cases occurred by endemic transmission in almost 50 countries in the Americas, Fig. 5.

The local transmission of CHIKV Asian lineage in Americas was associated with *A. aegypti*. However, the global distribution of the arbovirus vectors *A. aegypti* and *A. albopictus* (Kraemer *et al.*, 2015) suggests that it is still possible that in the future it will occur the emergency of strains from the Asian lineage adapted to *A. albopictus* (Weaver & Forrester, 2015). Indeed, CHIKV strains from the Asian lineage are in circulation in territories with native *A. albopictus* for at least 50 years ago. Nevertheless, until date, the Asian lineage adaptation to *A. albopictus* was not yet observed and as a consequence it has been suggested that the Asian lineage has a constrained genetic capacity to adapt to *A. albopictus* (Tsetsarkin, Chen, Leal, *et al.*, 2011).

The aforementioned adaptation requires a mutation in the glycoprotein E1 of an alanine to a valine at the position 226 (E1-A226V) but also requires epistatic interactions of two other mutations, such as E1-T98A, E2-I211T and recently a third mutation, E2-L210Q, was also identified (Tsetsarkin *et al.*, 2016, Tsetsarkin, Chen, Sherman, *et al.*, 2011), where E1 and E2 represent the glycoproteins E1 and E2 respectively. As mentioned earlier, the epistatic mutations have been observed in the lineages ECSA and WA and in the sub-lineage IOL, but until now were never reported in the Asian lineage (Coffey *et al.*, 2013). Interestingly, to date, the E1-226V residue has never been detected in Asian lineage strains. Indeed, if this adaptative mutation (E1-A226V) occurs in the future, it is highly essential the occurrence of the epistatic mutations E1-T98A and E2-I211T at the

same time. These epistatic mutations are crucial for the E1-226V residue to exert its phenotype in *A. albopictus* (Chen, Wang, *et al.*, 2013, Tsetsarkin, Chen, Leal, *et al.*, 2011, Weaver & Forrester, 2015), Fig. 5.

2. Prevention methods, current diagnostic, treatments to fight alphaviruses and the lack of anti-alphavirus drugs.

Outbreaks are difficult to prevent due to the infection of bridge vectors or highly mobile hosts such as birds and mammals. The transmission of alphavirus infection through the live trade of asymptomatic infected animals is another difficult issue to control. The import of animals in densely urbanized areas may increase the risk of alphavirus emergency in Europe. For example, Belgium and Netherlands have a dense horse population with a land surface covered by suitable habitat for competent vector species such as *Culex pipiens* and *Aedes albopictus* and this may increase the risk of VEEV emergency in Europe (Durand *et al.*, 2013). Concerning CHIKV, it is already known that the virus has already expanded across the globe, from Africa to Asia, Europe and Americas through the *A. aegypti* vector or through adaptation to *A. albopictus* (CDC, 2018).

Thus, controls measures are usually based in vector elimination and limiting the human contact with *A. aegypti* and *A. albopictus* mosquitoes. Understanding the distribution of *Aedes* mosquitoes and the relation with climate conditions can help the prevention of human arbovirus infections. The preferred common habitats for oviposition usually are natural rain-filled artificial and natural containers, including discarded tins, used tires, tree holes and rock pools (Ding *et al.*, 2018, Morrison *et al.*, 2004). The mosquito expansion has a relationship with the intranational travel and trade routes (Gloria-Soria *et al.*, 2014).

Forested enzootic habitats and widespread in the neotropics are probably the main difficulties encountered to control endemic VEEV due to the continuous circulation of enzootic strains. However, deforestation and urbanization enhance contact of enzootic VEEV with humans and adaptation of enzootic viruses to urban circulation occurs probably through the adaptation of the virus to infect peri-domestic mosquitoes.

Regarding CHIKV, the difficult encountered to control CHIKV outbreaks are higher than in VEEV because Globalization of economic activities and cultures, together with year-round expansion of competent mosquito vectors, increases the risk of CHIKV become endemic in many parts of the world. For example, in tropical regions of Americas mainly in densely populated urban centres. Repetitive introductions and the establishment of

CHIKV into Americas it will probably influence the continuous spread of CHIKV outbreaks across the globe (Silva & Dermody, 2017).

The initiation of the human transmission cycle and potential spillback into the enzootic cycle can be efficiently controlled in new regions by the rapid identification of imported cases. The patients and family members should be accompanied and educated to minimize exposure to the insect vector. This measure when combined with vector control can efficiently control the spread of the virus as it happens in Italy 2007 (Rezza *et al.*, 2007). Usually mosquitoes in a radius of 50 to 100 meters are probably already infected and should be killed. The success of this intervention methods can be guaranteed if placed before secondary cases and further spread have occurred. However, strains with limited dispersal potential are more easily controlled such as the CHIKV Asian lineage that are transmitted by *A. aegypti* (Weaver & Forrester, 2015).

The controlling measures used in Italy to control *A. albopictus* included the use of fast-acting insecticides (synergised pyrethrins) for 3 days consecutively, applied with a truck-mounted atomiser in public spaces and a backpack mist blower in private spaces. Antilarval measures, using formulations of insect growth regulators and *Bacillus thuringiensis va israeliensis* were also implemented. House-to-house interventions were done to eliminate breeding places with the participation of the community. The measures were placed within a radius of 100 m to 300 m of the individual's residence of suspected cases of CHIKV (Rezza *et al.*, 2007).

However, these measures to control mosquitoes when applied in non-developed countries have been shown to be inefficient. Many efforts are generally focus in reducing or treating standing waters or containers for water storage, including backyard, nondegradable trash containers where eggs are laid and larvae develop (Weaver & Lecuit, 2015). Education of the population in problematic regions is also useful to limit the contact between humans and infected mosquitoes. People in affected regions are incentivised to wear protective clothing, sometimes impregnated with insecticides, or wearing repellents (Weaver & Lecuit, 2015). The use of insecticide impregnated curtains can also limit the entry of vector mosquitoes into homes. However, resistance of vector mosquitoes to insecticide can emerge as a limitation (Weaver & Lecuit, 2015, Lorono-Pino *et al.*, 2013).

To overcome the limitations currently encountered for vector mosquitoes' control, at least two measures are currently under the investigation. One of such is the possible release of transgenic *A. aegypti* engineered to carry a late-acting lethal genetic system (Phuc *et al.*, 2007). The other one is the use of Wolbachia bacteria, which, when introduced into

A. aegypti or *A. albopictus* mosquitoes, reduce their vector competence for CHIKV (Moreira *et al.*, 2009, Mousson *et al.*, 2010, van den Hurk *et al.*, 2012, Weaver & Lecuit, 2015).

The current methods used for prevention of alphavirus infection are not being efficient because highly pathogenic alphaviruses such as VEEV and CHIKV are still in circulation. The continuous circulation of enzootic VEEV within enzootic and urban areas has caused outbreaks with a high fatality rate for equines (Aguilar *et al.*, 2011, Aguilar *et al.*, 2009, Aguilar *et al.*, 2004). Despite the low fatality rate for humans [probably higher than 1 % (Quiroz *et al.*, 2009)], VEE illness can cause in humans neurological complications and is a significant economic burden regarding treatment and long-term care (Armstrong & Andreadis, 2013). While CHIKV has emerged as a global pathogen, to date with about 50% of the world's population at risk of infection (Powers & Waterman, 2017, Powers, 2019). Thus, efficient methods for prevention are highly required and vaccination would be a very efficient prophylactic method.

The prophylaxis of VEEV outbreaks usually is promoted by vaccination of equines in regions where the progenitors of epizootic strains are believed to circulate and where outbreaks have been documented (Aguilar *et al.*, 2011). Vaccination of equids during outbreaks can efficiently control the outbreaks if VEEV is detected on time (Weaver *et al.*, 2004). Two prophylactic vaccine are available for equids as prophylactic measures to control VEEV (Aguilar *et al.*, 2011). One is the live-attenuated TC-83 vaccine produced in Mexico, Venezuela and Colombia and is used for prophylaxis through Central and South America. In the past, this vaccine was also used for prophylaxis in humans, but its use was discouraged because between 20 to 40 % of the vaccinated people developed the disease with symptoms typical of natural VEEV infection and other adverse effects (Alevizatos *et al.*, 1967, Paessler *et al.*, 2006). The other vaccine is the inactivated multivalent alphavirus vaccine which is market in US and is used as prophylactic strategy for equines in US territory, however, it has been also used in some equines in the South America when their owners have access to the vaccine market in US (Weaver *et al.*, 2004). However, the inactivated multivalent alphavirus vaccine, in use to prevent VEEV in US, also presents some limitations and its use has been discouraged because (1) generates short lived immunity, (2) multiple boosters are required to induce and maintain protection, and (3) equines already vaccinated with the inactivated vaccine will not respond appropriately to the live vaccine (Weaver *et al.*, 2004). Vaccination with inactivated vaccines also has another limitation, for example, the possible occurrence of outbreaks caused by using incomplete inactivated vaccines. A very well-known case is

the 1969 Texas outbreak that was caused by VEEV subtype IAB that occurred due to the use of formalin poorly inactivated virus vaccines (Kinney *et al.*, 1992a).

Concerning CHIKV prophylaxis through vaccination, to date no licenced vaccine(s) are available to prevent CHIKV infections, despite of some CHIKV vaccine candidates have been evaluated in at least preliminary clinical trials (Chang *et al.*, 2014, Weaver & Lecuit, 2015, Weaver *et al.*, 2012, Powers, 2019). Until 2004, the CHIKV vaccine development that started in the 1960s, after a large outbreak in Thailand, had little progress (Powers, 2019). The global spread of CHIKV and the recognition that a high number of patients developed chronic symptoms, stressed the development of vaccine candidates for prophylaxis of CHIKV, since then, several new candidates have been developed (Powers, 2019). Some of the CHIKV candidate vaccines have presented promising results during clinical trials, this include; formalin inactivated vaccines (Harrison *et al.*, 1971, Eckels *et al.*, 1970, Deeba *et al.*, 2016), live attenuated vaccines (McClain *et al.*, 1998, Edelman *et al.*, 2000), and virus-like particle (VLP) vaccine (Akahata *et al.*, 2010, Silva & Dermody, 2017, Weaver *et al.*, 2012, Chang *et al.*, 2014). Other vaccine candidates based on a range of strategies and platforms, including chimeric viruses, DNA vaccines and subunit approaches, are also under development (Powers, 2018, 2019). From the eight vaccines already evaluated in preliminary clinical trials, only three have been evaluated in phase 2, in-human clinical trials, and presented a strong safety and immunogenicity profile and plans for phase 3 study are in preparation (Powers, 2019). It is highly promising that these works might lead to a protective, licenced product in the near future. It will be an important hallmark for the prophylaxis of CHIKV because it will allow the protection of all those at risk of infection (Powers, 2019). Although, the development of supportive therapies is also equally important to support patients affected by this severe disease.

2.1. Diagnostic of *Alphavirus* infection

The diagnostic of *Alphavirus* infection is made with both evaluation of clinical symptoms and routine laboratory diagnosis. For example, concerning CHIKV fever, individuals presenting acute onset of fever, usually with chills/rigors, lasting for 3-5 days, multiple joint pains/swelling of extremities that may continue for weeks to months are considered suspected cases (Wahid *et al.*, 2017). When patients from suspected cases also present an history of travel or residence in areas reporting outbreaks and are excluded other diseases known to cause fever with join pains (such as malaria and dengue), these cases are then classified as probable cases (Wahid *et al.*, 2017). The probable cases are then subject to confirmation through the use of multiple techniques such as; virus

isolation in cell culture or animal inoculation from acute phase serum, detection of the presence of viral RNA in acute phase serum by real-time reverse transcriptase (RT) polymerase chain reaction (PCR) (real-time RT-PCR), detection of virus-specific immunoglobulins M [IgM] antibodies in single serum sample in acute or convalescent stage, and detection of fourfold increase in virus-specific immunoglobulins G [IgG] titer in samples collected at least three weeks apart (Wahid *et al.*, 2017).

The isolation of the virus from serum samples, using cell culture or animal inoculation, was previously used but now is not a preferential technique because is time-consuming (Staples *et al.*, 2009, Adams *et al.*, 2012). Thus, detection of viral RNA by real-time RT-PCR and detection of specific immunoglobulins IgM/IgG by enzyme-linked immunosorbent assay (ELISA) or hemagglutination inhibition test (HAI) become preferred techniques (Deeba *et al.*, 2016, Adams *et al.*, 2012, Wahid *et al.*, 2017). To efficiently detect viral RNA from all CHIKV serotypes and distinguish them from dengue virus (DENV) it has been used the multiplex real-time RT-PCR assay as well (Wahid *et al.*, 2017, Cecilia *et al.*, 2015). The conventional PCR, the real-time PCR and the multiplex real-time RT-PCR have been developed targeting the envelope and the non-structural genes and can be used for early identification of infection before detection of antibodies (Deeba *et al.*, 2016, Adams *et al.*, 2012, Wahid *et al.*, 2017).

During acute phase of CHIKV infection the levels of viremia can be characterized by 10^5 – 10^{12} RNA copies per mL of blood and can be detected by real-time PCR (Das *et al.*, 2010). Also, within the first 7-10 days of infection, the viral loads of CHIKV can reach $1 \times 10^{6.8}$ plaque forming units (PFU) per mL that usually last for 4-6 days and eventually can persist up to 12 days (Lanciotti *et al.*, 2007, Laurent *et al.*, 2007). Thus, viral load can be detected by using real-time RT-PCR in samples collected during the first 7 days after illness (Edwards *et al.*, 2007, Staples *et al.*, 2009, Lanciotti *et al.*, 2007, Panning *et al.*, 2008) and the genotype of the virus can be identified by coupling DNA sequencing to the RT-PCR (Deeba *et al.*, 2016). A new technique named Reverse Transcription loop-mediated isothermal amplification (RT-LAMP) has also been used to detect CHIKV (Lakshmi *et al.*, 2008). The principle of this new technique is based in one step nucleic acid amplification and as an advantage is less time consuming and no need for specialized equipment's such as thermocycles (Staples *et al.*, 2009).

The IgM antibodies develop within 2-3 days after the onset of symptoms of CHIKV infection and may persist for several months (Sam & AbuBakar, 2006, Malvy *et al.*, 2009). During the convalescent phase the IgG antibodies are developed and may persist for several months to years (Cavrini *et al.*, 2009). The detection of antibodies (IgM/IgG) are routine used, however, some researchers have suggested that the sensitivity of

serologic techniques has limitations due to the possibility of cross reactivity with other arboviruses (Deeba *et al.*, 2016, Pialoux *et al.*, 2007). For the detection of antibodies usually are used immunoassays such as ELISA, HAI, indirect immunofluorescent assay (IFA) and plaque reduction neutralization test (PRNT) (Yergolkar *et al.*, 2006, Litzba *et al.*, 2008, Staples *et al.*, 2009). PRNTs are highly specific for alphaviruses and highly used for confirmation of serologic test results. However the requirement to used live viruses is a limitation of the PRNT technique for alphavirus that are biosafety level 3 agents (Staples *et al.*, 2009).

The methods used for the diagnostic of VEEV are highly similar to those applied to other alphaviruses such as CHIKV, for example. Initially, the cases are subject to the clinical evaluation of the symptoms and then probable cases are confirmed using routine laboratory methods of diagnosis. The real-time RT-PCR has been recognized as a reliable high-throughput technique for the detection and quantification of VEEV RNA in clinical and field samples and allows a rapid differentiation from potentially cocirculating EEEV and WEEV strains (Vina-Rodriguez *et al.*, 2016). This technique has been recognized as suitable for the surveillance of VEEV due to the capability to detect all known VEEV variants (Vina-Rodriguez *et al.*, 2016). The detection of antibodies by using immunoassays is also used. For example, PRNT is the most direct method for specific identification of neutralizing antibodies and recent works have shown that using a retroviral pseudotypes of VEEV, based on murine leukemia viruses, overcomes the safety limitation from the original PRNT assay that use the infectious VEEV (Kolokoltsov *et al.*, 2006). Interestingly, the modified PRNT assay is safe, sensitive, in general reproduce the neutralizing antibody titers obtained by conventional PRNT and is suitable for primary screen for detection of neutralizing antibodies against VEEV (Kolokoltsov *et al.*, 2006). The VEEV diagnosis through the use of monoclonal and polyclonal antibodies (Wang *et al.*, 2005) is also in use and other techniques are under development such the scFV fragments that can be used to fullsize IgGs for an ELISA assay or that might be applied for immune-PCR (Guo *et al.*, 2006) in order to increase the sensitivity of the detection (Kirsch *et al.*, 2008).

2.2. Treatments available and the lack of anti-alphavirus drugs

There are no specific therapeutic drugs to treat VEEV infected patients that result in severe meningoencephalitis and long-term sequelae and treatments have been limited to nonspecific antiviral agents, with suboptimal outcomes. Even the conventional use of human immune sera to treat human infections by passive immunity has never been used for CHIKV or VEEV infected patients. Very often, the passive immunity (alternatively

known as passive transfer) is particularly used for diseases caused by viruses that have no other known treatment and it has occurred for decades for infection with exotic viruses (Powers, 2018). Experiments have been performed to provide temporary protection against CHIKV by passive Immunity in mouse models for CHIKV infection. The administration of antibodies to an unimmunized mouse from an immune individual has been tested with efficiency in prevention and cure of CHIKV infection in mouse models, however until now is not being used in humans (Weaver & Lecuit, 2015, Couderc *et al.*, 2009, Silva & Dermody, 2017).

Since the early CHIKV outbreaks the treatments available for patients were only supportive care to treat the symptoms by using analgesics, anti-inflammatory drugs, and antipyretics. Thus, so far, to treat CHIKV infected patients are used non-salicylate analgesics and non-steroid anti-inflammatory drugs to control the symptoms including joint swelling (Pialoux *et al.*, 2007, Gould *et al.*, 2010a). In the 1980s a pilot study was created to evaluate the effect of a drug with antimalaria effect in CHIKV infected patients (Powers, 2018, Brighton, 1984).

The compound was cloroquine phosphate and its potential to treat CHIKV infected patients was noticed after a patient comment indicating that they felt less pain when taking cloroquine (Brighton, 1984, Powers, 2018). The pilot study was limited to 10 patients that received 250 mg per day of chloroquine phosphate for a study of 20 weeks and 7 patients reported that the treatment was effective (Powers, 2018). However, chloroquine failed to show efficiency to treat CHIKV infected patients during a study conducted on La Réunion Island (Gould *et al.*, 2010a, De Lamballerie, Boisson, *et al.*, 2008). Although, a different study conducted in India during the same outbreak reported that chloroquine treatment ameliorate the symptoms in patients (Powers, 2018, Chopra *et al.*, 2014). More independent evaluations of this drug are needed because the results observed might be strain dependent.

In contrast to cloroquine, another anti-malaria compound, Quinine presented more promising results *in vitro* assays and resistant mutants were obtained in the region of non-structural protein 1 (nsP1) from Chikungunya (Gould *et al.*, 2010a). However, no studies are available in humans. This suggests that non-structural protein 1 (nsP1) is probably one promising target for antiviral therapy.

Ribavirin belongs to the list of “US Food and Drug Administration” (FDA)-approved drug and it is used as inhaled antiviral agent for treatment of respiratory syncytial virus (RSV) infection and orally, in combination with alpha interferon (IFN- α), for treatment of chronic hepatitis C virus (HCV) infection. Interestingly, ribavirin presented a positive outcome in

patients with post-CHIKV arthritis (Ravichandran & Manian, 2008, Silva & Dermody, 2017, Powers, 2018), table 1. *In vitro* assays in which were tested the combination therapy of ribavirin with IFN- α 2b showed a synergistical antiviral effect against CHIKV infection (Briolant *et al.*, 2004). While, the combination of ribavirin with doxycycline induced a decrease in CHIKV replication *in vitro* and a decrease in viral loads and inflammation in mice models (Rothan *et al.*, 2015). Interestingly, pegylated interferon alpha appears to be an effective treatment against infection with VEEV and has profound effects on the host immune response to infection (lukaszewski and Brooks, 2000). In a similar fashion as ribavirin, a compound named VX-497 also showed more potent anti-VEEV effect when combined with interferon alpha, although the VX-497 seems to be a better inhibitor than the broad-spectrum Ribavirin, that exhibited a very limited effect against VEEV even if applied at concentrations higher than 500 μ M in infected cell cultures (Markland *et al.*, 2000).

Currently, a very potent anti-VEEV compound named as β -D-N⁴-hydroxycytidine (NHC) presented an half maximal inhibitory concentration (IC₅₀) inferior to 1 μ M and low levels of VEEV resistance (Urakova *et al.*, 2017). Interestingly, the NHC compound was recently tested against CHIKV using *in vitro* assays, and it has reported to behave as a pyrimidine ribonucleoside that selectively inhibits CHIKV replication in cell culture (Ehteshami *et al.*, 2017, Powers, 2018), table 1.

Another important discovered was the inhibition effect of CHIKV replication by favipiravir (T-705) though the use of *in vitro* assays (Delang *et al.*, 2014). This broad-spectrum antiviral was able to reduce the mortality rate and neurologic symptoms in mice experiments (Delang *et al.*, 2014) by probably targeting the RNA-dependent RNA polymerase (RdRp) of CHIKV. Interestingly, T-705 was previously reported to ameliorate the signs of the disease caused by WEEV and treatment with T-705 improved morbidity and mortality of WEEV-infected mice (Julander *et al.*, 2009).

Some quinazolinone compounds have been also identified with potential anti-VEEV effect (Schroeder *et al.*, 2014, Selvam *et al.*, 2007). For example, the quinazolinone CID15997213 compound was found to strongly inhibit the non-structural protein 2 (nsP2) activity and VEEV replication by *in vitro* studies and using small animal models (Chung *et al.*, 2014). Other example was the identification from *in vitro* works that the compound ML336 inhibited several VEEV strains at the nanomolar range and protected mice against infection (Schroeder *et al.*, 2014).

Some other compounds with potential anti alphavirus effect have been identified by screening existent libraries of compounds that inhibit the virus cytopathic effect (CPE) in

cell cultures (Wada *et al.*, 2017, Gigante *et al.*, 2014). Using this methodologies are included the identification of compounds harbouring a benzimidazole structure (Wada *et al.*, 2017) and the [1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones (Gigante *et al.*, 2014). These compounds were highly effective at low micromolar range against CHIKV and in some cases also targeted other alphaviruses. Interestingly, the [1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones were shown to inhibit the *in vitro* viral capping enzyme nsP1 from VEEV (Gigante *et al.*, 2014). Under antiviral pressure of a compound from the family [1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones was selected resistant CHIKV strains that carried a P34S substitution in the non-structural protein 1 (nsP1) (Gigante *et al.*, 2017). *In vitro* assays for the guanylation activity of VEEV nsP1 provided evidences that the amino acid at the position 34 was important for the inhibition of the guanylation activity exerted by compounds from this family (Gigante *et al.*, 2017). Thus, these works exhibited more evidences that suggested the nsP1 as an excellent target for the conception of new drugs against alphaviruses.

Currently several candidates have been identified by using high-throughput screenings of chemical libraries (Ashbrook *et al.*, 2016, Gigante *et al.*, 2014, Kaur *et al.*, 2013, Lucas-Hourani *et al.*, 2013) and by synthesis of designed drugs (Bassetto *et al.*, 2013, Das *et al.*, 2016). Several candidates with potential anti-CHIKV effect are presented in table 1 and it possible that some of them also present antiviral effect against VEEV, as it was the case of the NHC compound. To date researchers are focus in targeting the viral replication mainly through the inhibition of viral entry, protein synthesis, genome replication or enzymatic functions. Several compounds have been highlighted as potential lead compounds but further testing in animal models and humans are required for the development of those compounds (Abdelnabi *et al.*, 2015). Thus, antivirals candidates to treat alphavirus infections are urgently needed (Abdelnabi *et al.*, 2015).

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Table 1. Therapeutic candidates for CHIKV under development and primary characteristics of each candidate^a (Powers, 2018) and references therein.

Type of therapeutic	Compound(s)	Model system(s)	CHIKV strain(s) (genotype)	Functional concentration			Other measure(s) of inhibition	Selectivity index (CC ₅₀ /EC ₅₀)	In vivo trial or case study design		Reference
				IC ₅₀	EC ₅₀	CC ₅₀			Dose(s)	Route	
Known antiviral	Ribavirin	Vero cell culture	Ross C347 (ECSA)	58 μ M	83 μ M	NP		1 to 24	ND	ND	84
	Interferon alpha			>10,000 μ M	9.7 to 11.1 μ M			>900			
	Ribavirin and interferon alpha	Vero cell culture	181/25 (vaccine strain) (Asian)	NP	833 μ g/ml (ribavirin); 5,761 IU/ml (IFN- α)	NP		NP	ND	ND	85
	Ribavirin and doxycycline	Vero cell culture, ICR mice	Clinical isolate (ECSA)	NP	4.52 μ M	NP	92% inhibition of virus entry into cells	NP	5 to 50 mg/kg of body wt	i.p.	86
	Picolinic acid	Vero cell culture	DRDE-06 (ECSA)	NP	NP	NP	CHIKV RNA concn decreased by 2 log ₁₀ units	NP	ND	ND	140
	Nidocamide	BHK-21 cell culture	S27 (ECSA), 0611aTw (ECSA), 0810bTw (ECSA)	NP	0.85 to 0.95 μ M	>20 μ M	Decrease of viral titer up to 1.5 log ₁₀ units	11.17 to 23.53	ND	ND	95
	Nitazoxanide	BHK-21 cell culture	S27 (ECSA), 0611aTw (ECSA), 0810bTw (ECSA)	NP	1.96 to 4.95 μ M	25 μ M	Decrease of viral titer up to 2 log ₁₀ units	5.05 to 16.34	ND	ND	
	Curcumin	HeLa, BHK-21, Vero E6	06-049 (ECSA)	3.89 μ M	NP	11.6 μ M		NP	ND	ND	101
		293T	CHIKV E2/E1-pseudotyped lentivirus vector; CHIKV strain not specified	3.90 to 10.79 μ M	NP	<60 μ M		NP	ND	ND	102
	Berberine	BHK cell culture	LR2006 OPY1 (ECSA)	NP	1.8 μ M	>100 μ M	Decrease of virus production by 5 log ₁₀ units	>55.6	ND	ND	99
		Huh7.5 cell culture		NP	1.9 μ M	>100 μ M		>52.6			
	Berberine	Hek293 cell culture	LR 2006 OPY1 (ECSA), SGP11 (ECSA), CNR2023 (Asian)	NP	4.5 μ M	202.6 μ M		45			100
		HOS cell culture		NP	12.2 μ M	429.5 μ M		35			
		ATCC CRL-2522 cell culture		NP	35.3 μ M	NP		NP			
	C57BL/6 mice, 4 wk of age		ND	ND	ND		ND	10 ⁶ PFU virus	s.c.		
Coumarin A	Vero, C6/36 cell culture	CHIKV ACol (Asian)	NP	10.7 μ g/ml	3,150 μ g/ml		295.2	ND	ND	103	
Coumarin B			NP	0.5 μ g/ml	549 μ g/ml		1,021.0	ND	ND		
Voacangine			NP	304.3 μ g/ml	1,136 μ g/ml		3.7	ND	ND		
Lupeol acetate			NP	538.5 μ g/ml	4,015 μ g/ml		7.5	ND	ND		

(Continued on next page)

Table 1 Continuation.

Type of therapeutic	Compound(s)	Model system(s)	CHIKV strain(s) (genotype)	Functional concentration			Other measure(s) of inhibition	Selectivity index (CC ₅₀ /EC ₅₀)	In vivo trial or case study design		
				IC ₅₀	EC ₅₀	CC ₅₀			Dose(s)	Route	Reference
Known antimicrobial	Flavaglines	HEK293T17 cell culture	Thai isolate (ECSA)	NP	22.4 nM	92 to 138.5 nM		NP	ND	ND	87
	Suramin	BHK-21, U2O8, MRC-5 cell culture	S27 (ECSA)	NP	8.8 to 62.1 μM	~350 to 770 μM		19.3 to >39.1	ND	ND	88
	Suramin	C57BL/6 mice, 4 wk of age	0810bTw, 0611aTw, 0706aTw (ECSA)	ND	ND	ND		ND	0.25 to 2 mg suramin	i.p.	89
	Flavanoid (silymarin)	Vero, BHK-21 cell culture	My/065/08/FN295485 (ECSA)	16.9 μg/ml	50 to 100 μg/ml	305 to 425 μg/ml		25.1	10 ⁵ PFU CHIKV	s.c.	90
	Flavonoid (baicalin)	Virtual screening	NP	ND	ND	ND	Binding affinity of -9.8 kcal/mol	ND	ND	ND	92
	Flavonoid (naringenin)						Binding affinity of -8.4 kcal/mol				
	Flavonoid (quercetagenin)						Binding affinity of -8.6 kcal/mol				
	Flavonoid (baicalein)	Vero cell culture	FN295485 (ECSA)	6.997 μM	ND	356.3 μg/ml		188.4	ND	ND	91
	Flavonoid (fisetin)				29.5 μM	194.4 μg/ml		23.02			
	Flavonoid (quercetagenin)				43.52 μM	226.7 μg/ml		16.3			
	Flavonoid (green tea catechin)	HEK293T cell culture	NP	6.54 μg/ml	NP	NP	40% decrease in infection rate at 10 μg/ml	NP	ND	ND	96
	Cardiac glycoside (lanatoside C)	BHK cell culture	D1225Y08 (ECSA)	NP	NP	NP	IC ₇₀ 1 μM; EC ₃₀ 1 μM	NP	ND	ND	93
	Cardiac glycoside (digoxin)	U-2 OS cell culture HSF cell culture Vero cell culture ST2 cell culture C2C12 cell culture	SL15649 (ECSA), 181/25 (Asian)	NP	48.8 to 108.9 nM 43.9 nM 67.3 nM 16.2 μM 23.2 μM	NP		NP	ND	ND	94
	Vitamin C	Human case	Unknown	ND	ND	ND	Symptoms resolved	ND	100 g/day	i.v.	97
Chloroquine	Human clinical trial	Infected patients (ECSA)	ND	ND	ND		ND	250 mg/day	Oral	81	
Designer molecules	nsP2 protease inhibitors (dimethylbenzaldehyde derivative)	BHK-21 cell culture	LR2006 OPY1 (ECSA)	~50 to 100 μM	1.5 to >100 μM	>200 μM		>2.1 to >133.3	ND	ND	104
	Benzimidazole derivative (compound A)	Vero cell culture	SL10571, S27, (ECSA) BaH306 (Asian)	NP	0.54 to 0.98 μM	3.70 μM		NP	ND	ND	105
	Benzo-coumarin-arene conjugates	Vero cell culture	899 (ECSA)	NP	10.2 to >331 μM	13.8 to >284 μM		1.6 to 11.5	ND	ND	107

(Continued on next page)

Table 1 Continuation

Type of therapeutic	Compound(s)	Model system(s)	CHIKV strain(s) (genotype)	Functional concentration			Other measure(s) of inhibition	Selectivity index (CC ₅₀ /EC ₅₀)	In vivo trial or case study design		
				IC ₅₀	EC ₅₀	CC ₅₀			Dose(s)	Route	Reference
	LATA-PAP1-THAN peptide fusion protein	Vero cell culture, ICR mice, 5-6 wk of age	SGEHCHS277108 (ECSA)	NP	11.2 µg/ml	NP	89% plaque reduction <i>in vitro</i> , 100% survival in mice at 0.75 mg/kg	ND	0.5 to 1 mg/kg	i.p.	108
	[1,2,3]Triazolo[4,5-d]pyrimidin-7(6H)-ones	Vero cell culture	899, LR2006 OPY1 (ECSA)	NP	0.75 to >490 µM	82 to >872 µM		>200	ND	ND	106
	Benzimidazole/thiosemicarbazone hybrid (MBZM-N-IBT)	Vero cell culture	S27, DRCE06 (ECSA)	NP	38.68 to 58.93 µM	>800 µM	Reduction in infectious particle release	>21	ND	ND	109
Nucleic acids	Oligonucleotide (PMO) targeting AUG of ORFs	HeLa cell culture	SGEHCHD122508 (ECSA)	NP	NP	NP	>96% cell viability at 10 µM, 1- to 3-log ₁₀ reduction in CHIKV titers	NP			116
	Ribonucleoside analog (β-D-N ⁴ -hydroxycytidine)	BALB/c mice	LKIEHCH 6708 (ECSA)						5, 10, and 15 µg/g	i.p.	
	siRNA (against nsP3, E1)	Huh-7, BHK-21, Vero cell culture	CNR20235 (Asian), LR2006 OPY1 (ECSA)	NP	0.2 to 1.8 µM	7.7 µM		NP	ND	ND	141
	siRNA (against nsP1, E2)	Vero E6 cell culture	DRCE06 (ECSA)	NP	NP	NP	96.3 to 99.6% reduction of virus titer at 25 nmol siRNA	NP	ND	ND	112
	miRNA (against nsP1, nsP2, C)	Swiss albino, C57BL/6 mice, 3 to 4 wk of age	061573 (ECSA)	NP	NP	NP	>90% inhibition in cell culture 100% inhibition of viremia	NP	20 to 25 µg	i.v.	113
	shRNAs (against E1, nsP1, C)	Vero cell culture	DRDE-07 (ECSA)	ND	ND	ND	Decreases in RNA load and infectious virus titers	ND	ND	ND	114
		HeLa, BHK-21, RD cell culture	Various strains (ECSA and Asian)	NP	NP	NP	0.5- to 3.2-log ₁₀ reduction in CHIKV titers	NP			115
		C57BL/6 suckling mice					60 to 100% survival		10, 30, and 60 µg	i.p.	
Monoclonal antibodies	Human MAb C9	HEK293T cell culture	S27 (ECSA)	0.1 to 0.4 µg/ml	NP	NP	100% survival; NT ₅₀ 0.3 µg/ml	NP	0.5 mg MAb/mouse	i.p.	118
		C57BL/6 mice, 6 wk of age									121
	Mouse MAb CK47 (anti-E1)	293, Vero, B7, PAI cell culture	SL11131, clinical isolate (ECSA)	0.8 to 250 µg/ml	NP	NP	CHIKV titer reduction of ~2.5 log ₁₀ units	NP	NP	NP	121

(Continued on next page)

Table 1 Continuation.

Type of therapeutic	Compound(s)	Model system(s)	CHIKV strain(s) (genotype)	Functional concentration			Other measure(s) of inhibition	Selectivity index (CC ₅₀ /EC ₅₀)	In vivo trial or case study design		
				IC ₅₀	EC ₅₀	CC ₅₀			Dose(s)	Route	Reference
	Human MAbs, various	Vero 81 cell culture, <i>ifnar-1</i> ^{-/-} mice, 6 wk of age	SL15649 (ECSA)	NP	0.6 to 5,200 ng/ml	NP	50 to 100% survival in mice	NP	50 µg	i.p.	119
	Disease-modifying antirheumatic drug (CTLA4-Ig) with human MAb (4N12)	C57BL/6 mice, 4 wk of age	LR2006 (ECSA)	ND	ND	ND	Eliminated foot swelling, 100- to 10,000-fold decrease in viral load	ND	300 µg CTLA4-Ig and 300 µg anti-CHIKV MAb	i.p.	122
Host cell targets	Prostratin	BGM, Vero, HEL, ATCC CRL-2522 cell culture	CHIKV-899 (ECSA)	NP	0.2 to 8 µM	>100 µM	Decreased viral RNA levels	NP	ND	ND	123
	Multiple small-molecule inhibitors (i.e., TOFA, pimoside)	HEK293 cell culture C57BL/6J mice	C21 (not specified)	<0.01 to 3.36 µM	NP	NP	Decreased levels of viral replication Footpad swelling	NP	20 mg/kg (pimoside), 25 mg/kg (TOFA)	Oral (pimoside), i.p. (TOFA)	124
	SAT1 (spermidine-spermine acetyltransferase)	Huh7, BHK-21 cell culture	La Reunion 06-049 (ECSA)	ND	ND	ND	Enhanced transcription but limited virion production	ND			125

^aIC₅₀, 50% inhibitory concentration; CC₅₀, 50% cytotoxicity concentration; EC₅₀, 50% effective concentration; ND, not done; NP, not provided; HSF, human synovial fibroblasts; TOFA, 5-tetradecyloxy-2-furoic acid. Shown is information on candidates reported through April 2017.

3. What are alphaviruses?

Alphaviruses are enveloped spherical particles with dimensions of approximately 70 nm (Westaway *et al.*, 1985), a diameter of approximately 700 Å, a molecular mass of 5.2×10^6 Da and a density of $1.22 \text{ g}\cdot\text{cm}^{-3}$, (Cheng *et al.*, 1995, Paredes *et al.*, 1992). The three-dimensional reconstruction of alphaviruses virions was initially performed using cryogenic Electron Microscopy (cryo-EM) for SFV (Vogel *et al.*, 1986), SINV (Fuller, 1987, Paredes *et al.*, 1993) and RRV (Cheng *et al.*, 1995). The icosahedral structure of many alphaviruses has been determined at very high resolution using cryo-EM and crystallographic studies (Lescar *et al.*, 2001, Mukhopadhyay *et al.*, 2006, Kostyuchenko *et al.*, 2011).

The centre of alphaviruses consists of a nucleocapsid with icosahedral symmetry, Fig. 6. The diameter of the nucleocapsid ranges approximately between 30 to 40 nm (Strauss & Strauss, 1994, Westaway *et al.*, 1985) and accommodates the genomic RNA of the virus. The non-segmented single stranded RNA genome of positive polarity is complexed with multiple copies of single species of capsid protein of about 30 kDa (Strauss & Strauss, 1994). The nucleocapsid is formed by 240 copies of capsid protein C and is embedded by a lipid bilayer from host-cell membrane origin. From the lipid bilayer protrudes trimeric spikes that form icosahedral protein lattices. The three-dimensional reconstruction of SFV displayed a triangulation number (T) of 4 for the symmetry of the particle (Vogel *et al.*, 1986). The structural organization of alphavirions was then confirmed after the determination of the three-dimensional structure of virions from the alphavirus prototype SINV (Strauss & Strauss, 1994), Fig. 6 and 7.

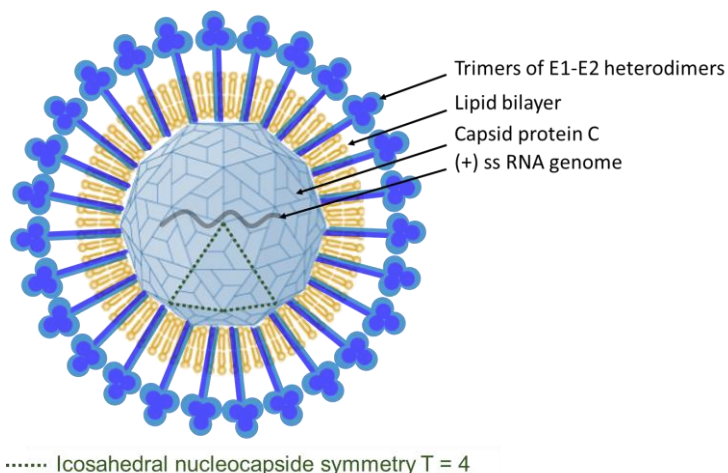


Figure 6. General schematic representation of an Alphavirus virion. Spherical particles with approximately 70 nm in diameter, containing an envelope made of the host lipid bilayer membrane and containing 80 spikes, each spike are trimer of E1:E2 heterodimer proteins. The central core contains an icosahedral capsid with T=4 icosahedral symmetry and is made of 240 monomers. The capsid with a diameter approximately between 30 to 40 nm is surrounded by the envelope structure and each heterodimer E2-E3 is in contact with the capsid. The green dot lines in form of triangle highlights the icosahedral nucleocapsid symmetry T=4. Adapted from (Cheng *et al.*, 1995).

The early reconstructions of SINV (Paredes *et al.*, 1993) and RRV (Cheng *et al.*, 1995) virions by cryo-EM elucidated the symmetry of both nucleocapsid and envelope of alphaviruses, Fig. 7. Both inner (nucleocapsid) and outer (envelope) shells of alphaviruses own an icosahedral lattice of triangulation number $T = 4$ with the presence of a five- and six- fold coordinated sub unities. This suggested the existence of 1:1 stoichiometric ratio between capsid protein and envelope glycoproteins (Strauss & Strauss, 1994), Fig. 7. The two glycoproteins (E1 and E2) that protrude from the membrane bilayer (outer shell) are in contact with the capsid protein C of the nucleocapsid (Paredes *et al.*, 1993). The glycoproteins E1 and E2 are transmembrane and the C-terminal endodomain of the E2 is in direct contact with the nucleocapsid. The structure of the outer shell icosahedron depends upon intramolecular disulphide bridges residing in the E1 glycoprotein (Anthony *et al.*, 1992, Fuller, 1987). The glycoproteins E1 and E2 are known to form 240 heterodimers that are arranged in sets of trimeric

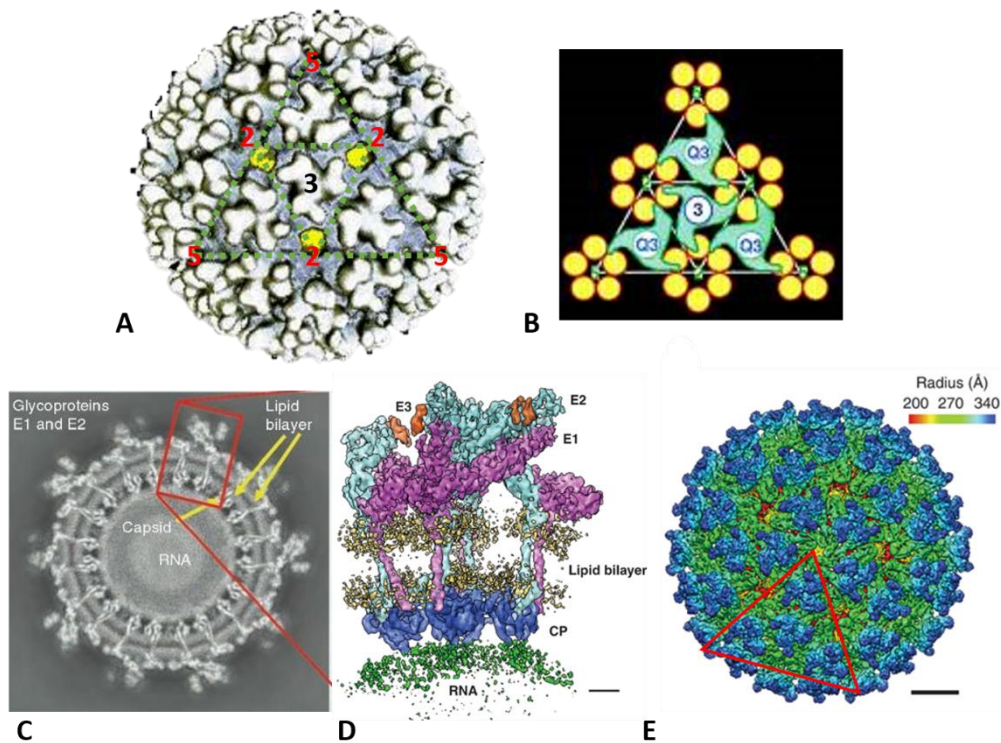


Figure 7. Three-dimensional (3D) reconstruction of two alphaviruses by cryo-EM analysis and image processing. (A) 3D surface structure of RRV virion (diameter ~70nm) viewed along an icosahedral threefold axis. The spikes trimers of E1-E2 heterodimers located at the threefold and quasi-threefold axes, have a flower-like head with three bilobal petals. The spikes are engaged in extensive lateral interactions close to the lipid bilayer via their skirts. These parts of the spikes are colored bluish. The lipid bilayer (yellow) is seen through openings in the spike-skirt protein layer at the twofold and fivefold symmetry axes. Adapted from (Cheng *et al.*, 1995) (B) Schematic representation (using RRV) of the interactions between the spikes (green) at the threefold (circled 3) or quasi-threefold (circled Q3) axes and the C molecules (yellow) of the capsomers in the underlying NC. Adapted from (Cheng *et al.*, 1995). (C) VEEV (TC-83) cryo-EM slice through the 3D density map 20 pixels from the origin Adapted from (Zhang *et al.*, 2011). (D) One asymmetric unit of VEEV containing four unique copies of E1 (magenta), E2 (cyan), E3 (orange) and CP (blue). The cryo-EM densities for the viral membrane (yellow) and genomic RNA (green) are also displayed at slightly lower isosurface threshold. Scale bar: 2 nm. (Zhang *et al.*, 2011) (E) Radially coloured 3D reconstruction of VEEV, showing the E1 basal triangle (green) and E2 central protrusion (blue) for each spike. Scale bar: 10 nm. (Zhang *et al.*, 2011)

structures. In total 80 distinct trimeric structures are present at the surface of the particle, Fig. 7. The glycoprotein E1 and E2 are formed after the cleavage of the its precursor polyproteins known as C-E3-E2-6K-E1 or C-E3-E2-TF. These proteolytic events also generate a small glycoprotein named as E3 and a small hydrophobic peptide that is produced as a linker between E1 and E2 which is named as 6K protein. Moreover, a transframe form of 6K protein (TF) is also produced. In some cases the E3 glycoprotein remains associated with the virion (*e.g.* SF) but not in others (*e.g.* SINV) (Strauss & Strauss, 1994). In contrast, the TF has been found to be associated with the virus in submolar quantities (7 to 30 molecules per virion)(Strauss & Strauss, 1994, Gaedigk-Nitschko *et al.*, 1990, Lusa *et al.*, 1991, Ramsey & Mukhopadhyay, 2017).

The alphavirus genome, a non-segmented single stranded RNA (non-segmented ssRNA) molecule with positive polarity of approximately 11.5 Kb in length, is capped at the 5' end with a type 0 cap structure and is polyadenylated at the 3' end with a poly(A)-tail (Westaway *et al.*, 1985). The characterization of the alphavirus genome, an important hallmark in the history of alphaviruses, was initiated during the 60's by using sedimentation sucrose gradients. The double stranded RNA (dsRNA) form of alphaviruses was reported for the first time by Sonnabend and colleagues in 1964 by using SFV (Sonnabend, 1964). Two years later, in 1966, Sreevalsan and Lockart (Sreevalsan & Lockart, 1966) described for the first time the identification of three species of viral RNA formed during the replication of WEEV. However, despite their attempts, they were not able to properly determine the infectivity of each form in Chicken embryo cells (CE). In the same year Friedman with his colleagues provided information about the replication pattern of SFV and they found three viral forms of RNA (Friedman *et al.*, 1966). They found that one of the forms sedimented at the fraction 42 Svedbergs (S) and a second form sedimented at 22 S. Their results suggested that the last fraction (22 S) was composed of two species of RNA, one that sediments at 26 S and other at 20 S (Friedman *et al.*, 1966). According to their work, the forms 42 S and 26 S were ribonuclease A sensitive while the 20 S form was resistant to the ribonuclease A treatment. Therefore, they have suggested that the 20 S RNA was a double stranded form and the 26 S was an intermediate form of single and double stranded RNA (Friedman *et al.*, 1966).

In 1967 Sonnabend and colleagues have worked also in the characterization of the RNA from SFV and presented slightly contradictory results from what has been published before. Using infected cells, three species of viral RNA were identified with the following sedimentation coefficients, 45 S, 26 S and 20 S. Similarly, to previous works, the forms 45 S and 26 S were ribonuclease A sensitive which was a characteristic of ssRNA.

Although the form 20 S was resistant to ribonuclease A treatment which served as indication of a possible dsRNA (Sonnabend *et al.*, 1967). For the first time, the infectivity of the three forms was properly accessed, they have shown that the 45 S RNA form was the most infectivity form and the other two forms presented low infectivity (Sonnabend *et al.*, 1967). They reported that 20 S RNA reached the maximum in the early infection while 26 S and 45 S are still increasing. As already suggested by others, the 45 S form is the form which is incorporated in the virus as it was similar to the form found in the RNA extracted from the virus (Sonnabend *et al.*, 1967). In the same study first-time evidences demonstrated that during early infection the 26 S form presents the highest rate of synthesis and is the major component in early infection. Only during the later infection, the form 45 S becomes a major component. Sonnabend and colleagues did not know the role of the form 26 S, but they have suggested that is an integral part of the virus replication and their results shown that 26 S is a precursor of 45 S. Moreover, it was proposed that 26 S RNA is probably the form in which viral RNA is released from the RNA polymerase despite other functions carry out by the 26 S RNA. For the first time it was suggested that the 26 S RNA could act as a messenger in the synthesis of viral proteins or as a template for viral RNA polymerase (Sonnabend *et al.*, 1967).

Studies were also performed in other alphaviruses such as SINV (Ben-Ishai *et al.*, 1968, Dobos & Faulkner, 1969) and was consistent with previous works of the era. The sedimentation coefficients of the three forms have a trend to be slightly different between different strains of alphavirus which is probably strongly related to the variation of the genome size (Dobos & Faulkner, 1969). In Fig. 6 is presented a graphical representation that summarizes the 3 species and their sedimentation coefficient zone. The work of Ben-Ishai *et al.* (Ben-Ishai *et al.*, 1968) also provided evidences that both 25 S and 20 S RNA species were found in association with the endoplasmic reticulum. Their work supported the hypothesis that viral RNA and coat synthesis was found to be localized in the cytoplasmic reticulum (Ben-Ishai *et al.*, 1968).

Since the breakthrough of the genome sequencing, the sedimentation sucrose gradient experiments are becoming less used. The sequencing of the genome of alphaviruses has provided more detailed information on the genome organization including the presence of regulatory elements in coding and non-coding regions.

4. The alphavirus RNA synthesis and protein translation

The *Alphavirus* genome consists of a ssRNA of positive polarity. Depending on the alphavirus strain, the genome can range from 11 to 12 Kb and, sedimentation coefficients from 40 S to 49 S RNA, respectively.

The genome of alphaviruses is also capped and polyadenylated at the 5'- and 3'-end respectively, which is important for the initiation of translation. The 5'-end of the genome is capped with a type-0 cap (m^7GpppN -RNA) (Ahola & Kaariainen, 1995, Dubin *et al.*, 1977, Hefti *et al.*, 1975, Strauss *et al.*, 1984, Wengler *et al.*, 1979), and the 3'-end of the genome is polyadenylated with a poly(A)-tail that range in length from approximately from 20 up to 250 nucleotides (nt) (Frey & Strauss, 1978, Ou *et al.*, 1981, Sawicki & Gomatos, 1976, Hyde *et al.*, 2015), Fig. 8. The size of the poly(A)-tail varies between species and inside species. However, it has been proposed that the minimal number of residues of about 11 – 12 nt in the poly(A)-tail are required for efficient production of the RNA of negative polarity (Hyde *et al.*, 2015, Hardy & Rice, 2005).

Within the genome of alphaviruses are present sequences and structural elements which are usually referred as *cis*-acting regulatory elements. In alphaviruses the *cis*-acting regulatory elements (CREs) are localized within both the untranslated regions (UTRs) and the translated regions (also known as coding sequences [CDS]), Fig. 8. The recognition of CREs by the alphavirus replication machinery is crucial for alphavirus RNA synthesis (Rupp *et al.*, 2015, Jose *et al.*, 2009, Pietila *et al.*, 2017). Two of the three

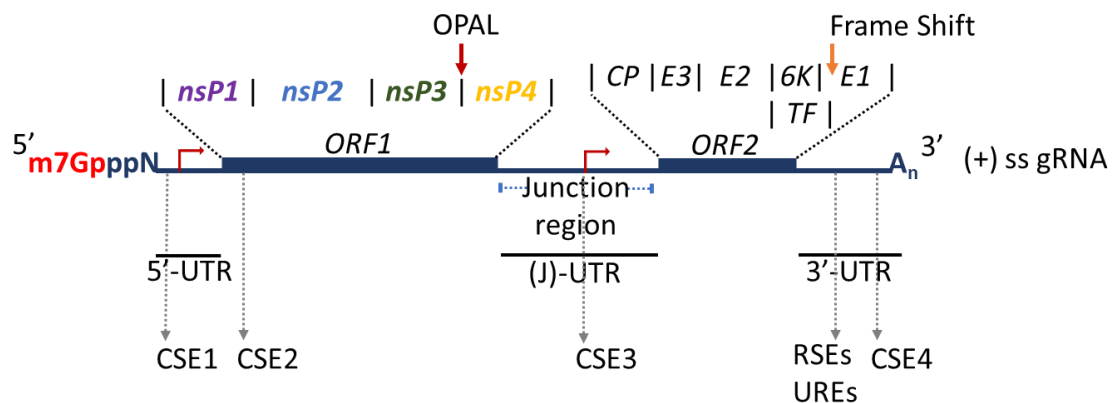


Figure 8. General schematic description of Alphavirus RNA genome. The Alphavirus RNA genome is capped with a cap-0 structure at the 5'-end and polyadenylated at the 3'-end. Indicated are the two coding regions (ORF1 and ORF2), the three untranslated regions (UTRs) several *cis*-acting regulatory elements (CREs), for example CSE1, CSE2, CSE3, CSE4, UREs and RESs, the genomic and sub-genomic promoters, the location of the OPAL codon and the frame shift sequence. The regions coding for non-structural proteins (nsP1, nsP2, nsP3 and nsP4) and for structural proteins (Cp, E3, E2, 6K, TF and E1) are highlighted. For details see text.

UTRs are flanking the 5'- and 3'- end of the alphavirus genome and are designated as 5'-UTR and 3'-UTR, respectively. The third UTR is localized between the two open reading frames (ORFs) of the alphavirus genome and named the junction region (J) (J-UTR) (Rupp *et al.*, 2015, Pietila *et al.*, 2017), Fig. 8.

The presence of UTRs flanking the 5'- and 3'- ends of RNAs is highly conserved among species. In alphaviruses the presence of 5'-UTR and 3'-UTR in the RNAs is not exception and the same is found for the J-UTR. Moreover, the three mentioned UTRs in the

alphavirus genome are conserved among the different species within the genus alphavirus (Hyde *et al.*, 2015, Rupp *et al.*, 2015), Fig. 8.

The UTRs vary greatly in size length, in sequence, and structure, both within a single alphavirus species and among strains of same species. Usually the 5'-UTRs between different alphaviruses range approximately from 27 nt, as observed in Salmonid Alphavirus (SAV), up to 85 nt, as noticed for SFV. Regarding the 3'-UTRs which are known to be longer than the 5'-UTRs, the length diverge approximately from 78 nt as in some strains from the VEE complex up to 723 nt as observed in CHIKV, (Hyde *et al.*, 2015).

The 5'-UTR of the alphavirus genome is usually characterized to host a CRE which is also referred as conserved sequence element 1 (CSE1), Fig. 8. The CSE1 corresponds approximately to the first 44 nt of alphavirus genome and it has been proposed to form a stem-loop structure (Niesters & Strauss, 1990b, Ou *et al.*, 1983). The stability of this stem-loop structure is important for polymerase access to the promotor and for regulation and initiation of RNA synthesis (Shirako & Strauss, 1998, Nickens & Hardy, 2008). In addition, recently it was suggested that circularization of the viral genome is necessary for the synthesis of viral RNA of negative polarity, and is in part mediated by CSE1 and involves the poly(A)-binding protein (PABP) (Frolov *et al.*, 2001, Kulasegaran-Shylini, Atasheva, *et al.*, 2009, Kulasegaran-Shylini, Thiviyathan, *et al.*, 2009). The CSE1 also plays undefined roles during infection in addition to being proposed to function as a promoter for the synthesis of viral single stranded RNA of positive polarity (Frolov *et al.*, 2001). The CSE1 has been suggested to camouflage the viral cap type-0 structure (the cellular cap is of type-1 structure) from being recognized by the host interferon-induced protein with tetratricopeptide repeats 1 (IFIT1) (Daffis *et al.*, 2010, Reynaud *et al.*, 2015). Thus, the stem-loop structure from CSE1 has been proposed as a determinant of pathogenicity because indirectly is involved in preventing the activation of the innate immunity (Daffis *et al.*, 2010, Reynaud *et al.*, 2015, Kinney *et al.*, 1993).

The 3'-UTR of the alphavirus genome also harbours a CRE, the conserved sequence element 4 (CSE4) of approximately 19 nt located upstream the poly(A) tract, Fig. 8. Most likely the CSE4 function as part of the core promoter for initiation of minus-strand synthesis. The mechanism involved in the synthesis of the viral RNA of negative polarity is not well understood, however, the CSE4 may provide an essential interaction between the 5'- and 3'- ends of the full-length viral genomic RNA (Frolov *et al.*, 2001, Kuhn *et al.*, 1990, Hardy, 2006). Indeed, it has been reported that 13 nt of the CSE4 and a poly(A) tail of at least 11 nt were found to be necessary for the synthesis of viral RNA of negative polarity (Hardy & Rice, 2005). Another important characteristic of the 3'-UTR is the

presence of other CREs which are the repeat sequence elements (RSEs) or U-rich elements (UREs), Fig. 8. The latter differ in their primary sequence and organization among several members of alphaviruses (Ou *et al.*, 1982, Chen, Wang, *et al.*, 2013). It has been suggested that both RSEs and UREs are required for repression of deadenylation of viral RNAs. A *trans*-acting factor is known to bind both CREs, the RSEs and UREs, and the human antigen R (HuR) protein is the *trans*-acting factor identified which modulates the stability and translational efficiency of mRNAs. Recently it has also found that HuR protein confers resistance to deadenylation in vitro and in tissue culture models of infection (Sokoloski *et al.*, 2010, Garneau *et al.*, 2008).

As already mentioned above, the CREs are present along the genome of alphavirus and are not restricted to untranslated regions only. The coding region of the genome is organized in two ORFs (Strauss *et al.*, 1984, Strauss & Strauss, 1994). The 5'-proximal ORF spans two thirds of the coding region and corresponds to the non-structural ORF that is directly translated from the genomic RNA (Strauss *et al.*, 1984, Strauss & Strauss, 1994). Within the non-structural ORF resides a CRE which is usually known as the conserved sequence element 2 (CSE2), Fig. 8. The CSE2 is located within the coding sequence of the alphavirus nsP1 and forms two stem-loops which may function as a transcriptional enhancer (Niesters & Strauss, 1990a). The CSE2 starts near the nucleotide 155 and may serve as part of the core promoter for the synthesis of viral RNA of negative polarity by using the genomic viral RNA (positive polarity) as a template. However, it has been reported that the CSE2 is also involved in the synthesis of the viral RNA of positive polarity. It is intriguing that the CSE2 is not present in the sub-genomic RNA (also of positive polarity) and thus suggests the reason why only viral full-length genomic RNAs can act as templates for the synthesis of the viral negative polarity RNA (Michel *et al.*, 2007, Frolov *et al.*, 2001). However, the CSE2 is dispensable for RNA replication (Gorchakov *et al.*, 2004, Frolov *et al.*, 2001).

Thus, in summary, the localization of the initiation site for the synthesis of viral RNA of negative polarity is dependent on the poly(A) tail and at least three 3' residues of the CSE4 (Hardy, 2006). The most important initiation site of the minus-strand synthesis is a cytidylate residue immediately before the poly(A) tail which is conserved among alphaviruses (Hardy, 2006, Adkins *et al.*, 1998). With regards to the synthesis of positive polarity RNA, both genomic and sub-genomic, it is always initiated with adenylate-uridylylate, apart from BFV which has an extra uridylylate residue at the 3' end of the RNA of negative polarity used as a template (Adkins *et al.*, 1998). The promoter for the genomic RNA synthesis is situated at the 3'-end of the RNA of negative polarity within the region corresponding to the CSE4 observed in the viral genomic RNA, thus

suggesting a complementary stem-loop structure to the viral RNA of negative polarity (Hardy, 2006, Adkins *et al.*, 1998).

Other CREs have been identified within the non-structural ORF, such as packing signals recognized by the capsid protein C (Cp), an early step during nucleocapsid (NC) formation (Weiss *et al.*, 1989). The NC assembly occurs during late infection, and only the viral genomic RNA is encapsidated except for AURAV which is not able to distinguish genomic from sub-genomic RNA, thus leading to the encapsidation of both viral genomic and sub-genomic RNAs (Rumenapf *et al.*, 1994). In SINV, VEEV, EEEV and WEEV packing signals were found within nsP1 coding region. They consist of four to six predicted stem-loops structures marked by a GGG conserved nucleotide sequence motif at the base of each loop (Kim *et al.*, 2011). In CHIKV, ONNV, SFV and RRV, though, three different sites were identified and localized within the non-structural protein 2 (nsP2) coding region (Frolova *et al.*, 1997).

Depending on the alphavirus strain, one or two non-structural polyproteins are translated from a single AUG initiation codon present in the non-structural ORF. The translation of two non-structural polyprotein, the P123 as a major product and the P1234 as a minor product, are partially regulated by the presence of the opal stop codon (UGA), Fig. 9, that has been reported in several alphaviruses isolates, such as isolates from SINV (Strauss *et al.*, 1984, 1983, Li & Rice, 1993), MIDV (Strauss *et al.*, 1983), VEEV (Kinney *et al.*, 1989), EEEV and WEEV (Weaver *et al.*, 1993), RRV (Strauss *et al.*, 1988), CHIKV (Morrison *et al.*, 2011) and ONNV (Lanciotti *et al.*, 1998). The termination for the translation of the P1234 is controlled by the presence of two amber stop codons (UAG) and one ochre stop codon (UAA) which are localized within the region transcribed into 26 S RNA, Fig. 9 (Strauss *et al.*, 1984).

Next to the stop codon of the non-structural ORF, the J-UTR separates the non-structural ORF from the initiation codon of the structural ORF (the 3'-proximal ORF). In 1989, Grakoui *et al.* have reported that in the J-UTR of SINV, an internal promoter exists for transcription of the sub-genomic mRNA (26 S RNA). Later, it was reported that this characteristic is common among alphaviruses (Grakoui *et al.*, 1989). The sub-genomic promoter has been identified as the conserved sequence element 3 (CSE3) which is a CRE of approximately 24 nt (Pushko *et al.*, 1997), Fig. 8. From the sub-genomic RNA transcript, the virus structural proteins are translated as a structural polyprotein of approximately 140 kDa, Cp-E3-E2-6K-E1 (Ramsey & Mukhopadhyay, 2017). The presence of a programmed ribosomal frameshifting near the end of the 6K gene leads

to the synthesis of the “trans-frame” structural polyprotein Cp-E3-E2-TF, Fig. 9 (Firth *et al.*, 2008).

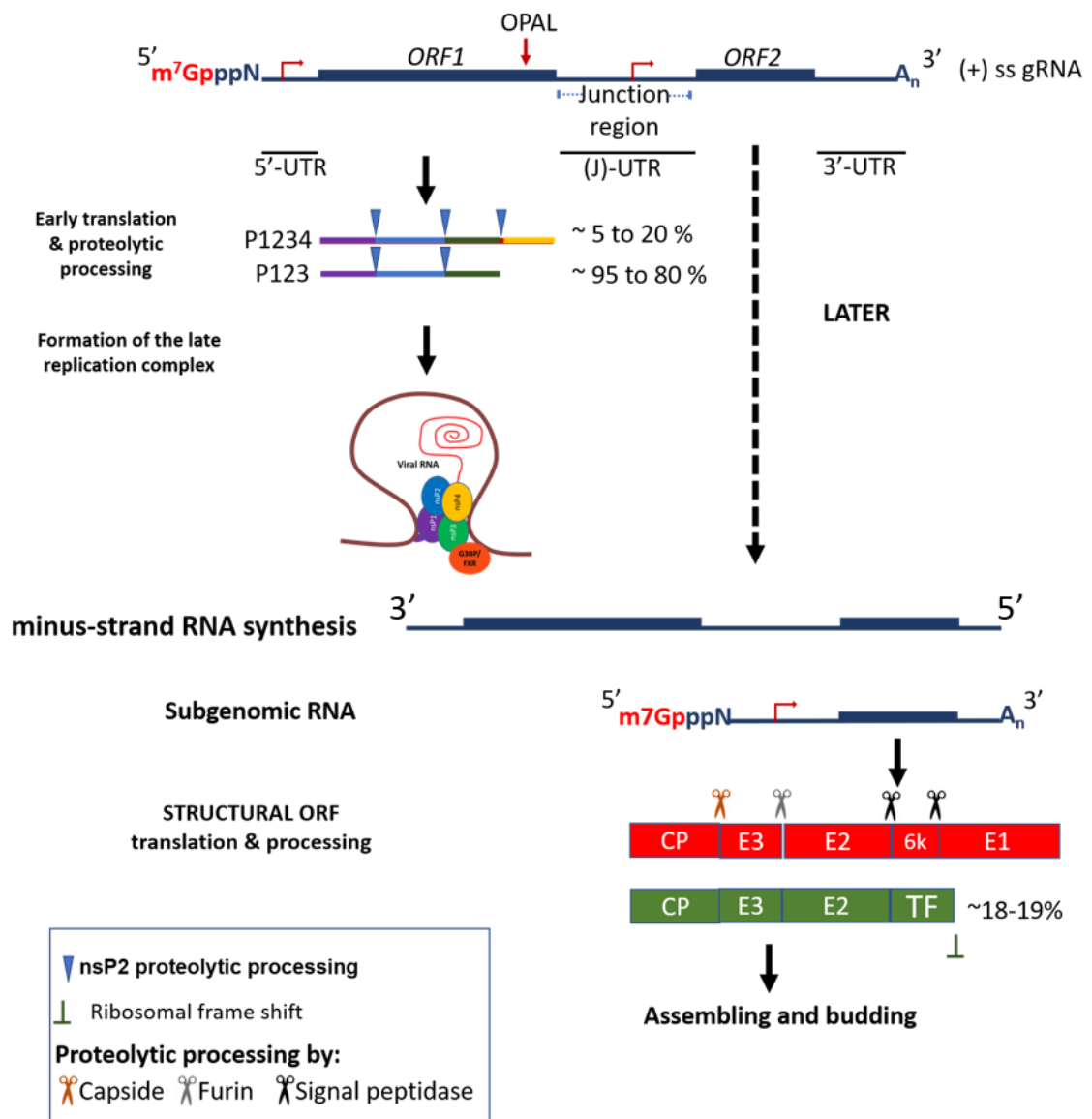


Figure 9. Diagram representing the major steps during translation of non-structural and structural proteins. The non-structural polyproteins P123 and P1234 are directly translated from the viral genomic RNA previously released into the cytosol of the host cell. The two polyproteins are subject of proteolytic processing and drives the formation of the early replication complexes. The early replication complexes are known to synthesise the minus strand RNA and the complete maturation of the four non-structural proteins (nsPs) leads to the switch to the synthesis of RNA of positive polarity (viral genomic and sub-genomic RNAs). The viral sub-genomic RNA is then used as a template for the translation of the structural polyproteins. A trans-frame region localized in the beginning of the region coding for the E1 protein is responsible for the translation of the two structural polyprotein Cp-E3-E2-TF and Cp-E3-E2-6K-E1. The two structural polyproteins are then proteolytic processed by the capsidase and furin proteases and the signal peptidase. Thus, leading to the formation of the mature structural proteins, Cp, E3, E2, 6K, TF and E1.

4.1. Translation of structural proteins

Enhancing sequences are present downstream the AUG initiation codon of the sub-genomic RNA, in SINV within the first 170 nt, with a predicted hairpin-like structure. The presence of a hairpin-like structure may present a barrier to the movement of the ribosomes during translation of mRNA (Frolov & Schlesinger, 1994). Thus, this pause may become the rate-limiting step in translation of the structural polyproteins.

During translation of Cp-E3-E2-6K-E1 or Cp-E3-E2-TF, the capsid protein (Cp) is translated first and folded in the cytoplasm. Then, the Cp is released from the nascent polyprotein chain by autoproteolysis (Garoff *et al.*, 1990). Cp is thus cleaved *in cis* via its chymotrypsin-like domain from the rest of the protein, Fig. 10 (Aliperti & Schlesinger, 1978, Hahn *et al.*, 1985, Hahn & Strauss, 1990, Melancon & Garoff, 1987). Therefore, Cp is a cytoplasmic protein of 35 kDa presenting a polybasic N-terminal domain that is predominantly unstructured and a structured serine protease domain at the C-terminal, (Brown *et al.*, 2018) and references therein. Still, the structured serine protease domain presents a chymotripsine-like structure which can be divided into two subdomains containing a hydrophobic cleft between them. Each subdomain holds six- or seven-stranded, antiparallel, β -barrel structure (Hahn & Strauss, 1990, Brown *et al.*, 2018).

A catalytic triad formed by His141, Asp163 and Ser215 (Hahn & Strauss, 1990, Melancon & Garoff, 1987) is important for the serine protease activity of the Cp (Choi *et al.*, 1991). During the autoproteolytic cleavage of Cp, the C-terminal Trp residue is left in the substrate-binding pocket to inhibit further proteolytic activity (Jose *et al.*, 2009). Once in the cytoplasm, the mature Cp can encapsidate the newly synthesized viral RNA with the help of the first highly basic 100 amino acids (a.a.) which supposedly bind the genomic RNA (Ramsey & Mukhopadhyay, 2017) (Coombs & Brown, 1989).

After the release of Cp, the N-terminal signal sequence of the remaining polyprotein, which also contains a carbohydrate attachment site, is used for pE2 (also known as p62) chain translocation (Bonatti *et al.*, 1984, Garoff *et al.*, 1990, Garoff *et al.*, 1978) (Jose *et al.*, 2009). The transmembrane protein pE2 (or p62) consists in the intermediate product of E3 and E2 which remain connected (E3-E2). Similar to pE2, two other structural proteins are also membrane proteins which are directed produced by alternating signal

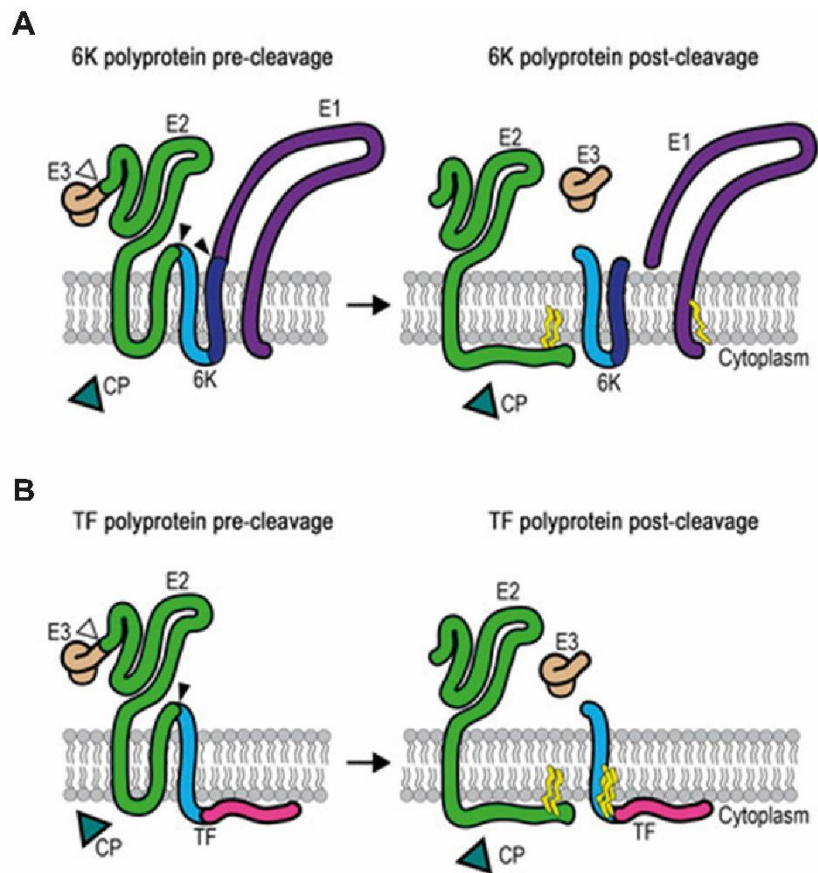


Figure 10. Structural polyprotein membrane topology and processing. Pre-cleavage of structural polyprotein and conformation in the endoplasmic reticulum (ER) membrane. Capsid (CP) is first released as a soluble protein in the cytoplasm. The N-terminus of E3 contains a signal sequence that directs translocation across the membrane. Open arrows mark the furin cleavage site between E3 and E2, which occurs in the trans-Golgi network. The remaining envelope proteins are threaded through the membrane with insertion of transmembrane domains. Closed arrows mark signalase cleavage sites that are used in the ER. Post-cleavage conformations include palmitoylation sites (yellow). **(A)** The 6K polyprotein is the majority translation product. **(B)** The TF form of the polyprotein is the minority product resulting from frameshifting in the 6K gene and does not include E1. (Ramsey & Mukhopadhyay, 2017)

and stop-transfer (anchor) sequences that function in translocation and cleavage of the virus precursor polyprotein (Liljestrom & Garoff, 1991). After pE2 translocation, the cellular protease furin cleaves E2/6K or E2/TF and later E1/6K, Fig. 10 (Dalbey *et al.*, 1997, Rice & Strauss, 1981, Liljestrom & Garoff, 1991, Strauss & Strauss, 1994). While in the endoplasmic reticulum (ER), pE2 together with cellular chaperones confer proper folding to E2 and E1 and later associates with E1 in heterodimers (Carleton & Brown, 1996, Molinari & Helenius, 1999, 2000, Mulvey & Brown, 1995, 1996, Parrott *et al.*, 2009). After, complexes named spikes are formed through the trimerization of those heterodimers. The spikes will be transported to the plasma membrane through the secretory system. Before the spikes reach the plasma membrane, furin cleaves E3 from E2 during its passage in the *trans*-Golgi and make the spikes fusion-competent (Lobigs *et al.*, 1990, Zhang *et al.*, 2003). (Ramsey & Mukhopadhyay, 2017). The cleavage occurs after the tetrabasic motif “Arg-His-Arg-Arg” and produces the mature E2:E1 heterodimer (Zhang *et al.*, 2003, de Curtis & Simons, 1988, Molloy *et al.*, 1999). However, cleavage

of E3 from the assembled spikes is not required for particle assembly/budding (Brown *et al.*, 2018) but is required to make the virus particles fusion competent (Ryman *et al.*, 2004), Fig. 10 and 11.

E3 is a small glycoprotein of 64 a.a. (approximately 7 kDa). It bears the signal sequence that, after Cp cleavage, directs the remaining polyprotein to ER membranes. There, in the ER, the remaining polyprotein cross the ER membrane several times, Fig. 10 (Ramsey & Mukhopadhyay, 2017). In addition, E3 probably serves as the insertion signal for E2 (Bonatti & Blobel, 1979, Bonatti *et al.*, 1984, Garoff *et al.*, 1990). The E3 is responsible for pE2 proper folding and mediates spike folding and maturation which is necessary for assembly of competent viral particles (Jose *et al.*, 2009). Participating in the pE2:E1 complex formation, the glycoprotein E3 is also involved in transport of viral structural components to the budding site. During the transport of the pE2:E1 complex to the budding site, E3 may display a stabilizing role to protect the fusion protein (E1) from the mildly acidic environment (Lobigs *et al.*, 1990, Lobigs & Garoff, 1990, Wahlberg *et al.*, 1989, Uchime *et al.*, 2013). Probably the interplay of a disulphide bond of E3 formed between two cysteines is crucial for E3 enzymatic or functional role in mediating spike transport to the plasma membrane (Parrott *et al.*, 2009). In the secretory pathway at low pH the cleaved E3 remains bound to E2 and is released at extracellular neutral pH (Sjoberg *et al.*, 2011). Thus, the E2:E1 heterodimer become sensitive to the pH of the endosome and prepares the mature virus for fusion and infection (Sjoberg *et al.*, 2011, Wahlberg *et al.*, 1989, Brown *et al.*, 2018).

The structural protein E2 is a transmembrane type I glycoprotein of approximately 40 to 50 kDa which functions as receptor binding protein and receptor mediated-endocytosis (Jose *et al.*, 2009). The glycoprotein E2 harbours two N-linked glycosylation sites and their elimination has been reported to increase replication owing to increase efficiency of binding to heparan sulphate (HS) on mammalian cells (Knight *et al.*, 2009).(Pletnev *et al.*, 2001). Indeed, the amino acid sequence of E2 is an important determinant of virulence (Tucker *et al.*, 1997). E2 has a highly-exposed leaf-like structure at the top of the spike followed by the narrower stem, which twists around the more tangentially disposed E1 molecule (Zhang *et al.*, 2002). The first 260 a.a. of E2 constitute the ectodomain, followed by approximately 100 a.a. that form the stem region and a 30 a.a. transmembrane helix. The carboxy-terminal domain E2 (cdE2) consists of 33 a.a. that interact with the NC core (Mukhopadhyay *et al.*, 2006, Pletnev *et al.*, 2001). Contacts are established between the leaf-like structure of E2 and the distal end of the E1 glycoprotein domain II, and between the stalk portion of E2 and domains I and III of E1 (Mukhopadhyay *et al.*, 2006). The receptor attachment site is near the residue 218 and

the carbohydrate associated with residue 216 of E2 are situated in the large, protruding, external, leaf-like surface, Fig. 12 (Smith *et al.*, 1995). This surface-accessible site is also the binding site for HS in a RRV mutant (Zhang *et al.*, 2005) as well as the Fab binding site for SINV- and RRV- neutralizing antibodies (Davis *et al.*, 1987, Meyer & Johnston, 1993, Smith *et al.*, 1995, Strauss *et al.*, 1991).

The presence of E1 is not required for the E2 transport to the plasma membrane, reviewed in (Brown *et al.*, 2018). Several signal sequences and post-translational modifications are responsible for regulation of glycoproteins translocation across the ER membrane. The E2 protein also containing the translocon signal (the start transfer sequence) for 6K (Ramsey & Mukhopadhyay, 2017) which is found in the C-terminus of E2 (Liljestrom & Garoff, 1991). One post-translational modification involved in the regulation of translocation is probably the palmitoylation of some or all the conserved cysteine residues of the cdE2 (Liljestrom & Garoff, 1991) (Gaedigk-Nitschko & Schlesinger, 1991, Ivanova & Schlesinger, 1993, Jose *et al.*, 2009). On the other hand, the start-transfer signal for E1 is found in the second hydrophobic stretch of E2 (Ramsey *et al.*, 2017, Liljestrom *et al.*, 1991, Loewy *et al.*, 1995).

The structural protein 6K is a transmembrane (type I) polypeptide which is known to be hydrophobic, cysteine-rich, acylated and can be found palmitoylated, Fig. 11. It has been suggested that this structural protein is formed by two membrane helices connected by a short cytoplasmic loop. The N-terminal of the 6K polypeptide is localized on the ER luminal side of the membrane (Liljestrom & Garoff, 1991) as well as the C-terminal of 6K due to the presence of the E1 start transfer sequence, Fig. 11 (Ramsey *et al.*, 2017, Liljestrom & Garoff, 1991, Hashimoto *et al.*, 1981, Melancon & Garoff, 1986). The 6K polypeptide was initially classified as a virus-derived ion channel, or viroporin (Fischer & Kruger, 2009) involved in ion modulation and that facilitates the virus budding (Ramsey & Mukhopadhyay, 2017, Liljestrom *et al.*, 1991, Gonzalez & Carrasco, 2003). Therefore, 6K can form cation-selective ion channels in planar lipid bilayers as it was reported for RRV and BFV with the following permeability affinity $\text{Na}^+ > \text{K}^+ > \text{Ca}^{2+}$ (Melton *et al.*, 2002, Madan *et al.*, 2005, Madan *et al.*, 2008). Viroporins are known as small proteins with hydrophobic stretches that oligomerize to form channels primarily comprised of alpha-helices (Martinez-Gil & Mingarro, 2015) and often associated with ion selectivity (Carrasco, 1995). It has been reviewed (Ramsey & Mukhopadhyay, 2017) that 6K may be similar to other viroporins such as viral protein U (Vpu) from the human immunodeficiency virus (HIV) or the protein p7 from hepatitis C virus (HCV) (Sanz & Carrasco, 2001, Gonzalez & Carrasco, 2001). The 6K protein can also participate in glycoprotein trafficking as it is thought to be associated with the pE2:E1 heterodimer

soon after synthesis and consequently transported to the plasma membrane, Fig. 11 (Lusa *et al.*, 1991). Induction caspase-dependent programmed cell death may be another processed requiring 6K (Madan *et al.*, 2008).

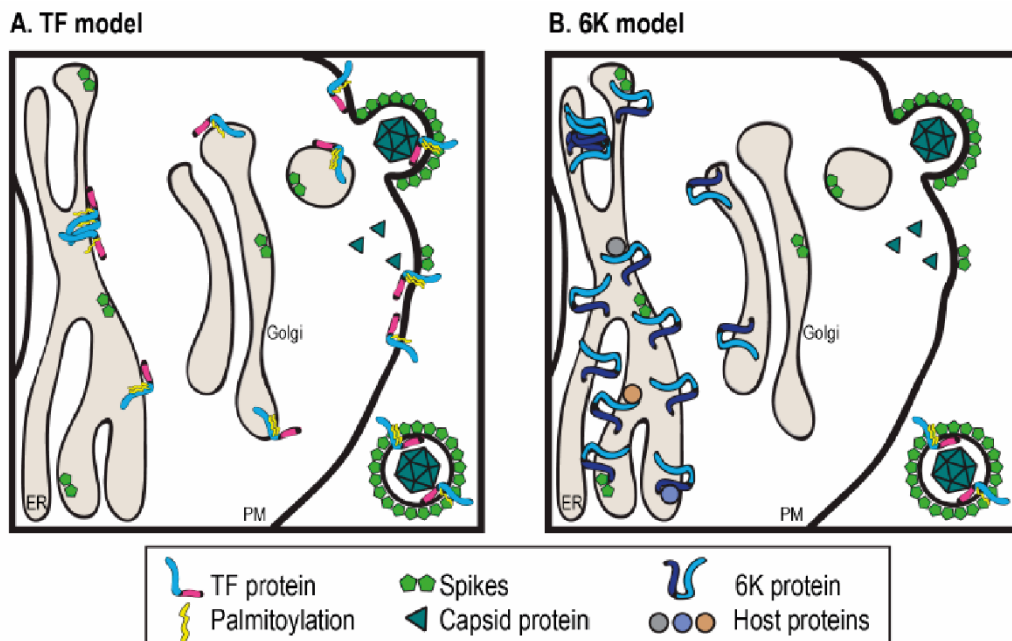


Figure 11. Model representing one hypothesis of the current understanding of TF and 6K. (A) TF is less abundant than 6K. It is palmitoylated and traffics from the ER to the plasma membrane (PM) where it is budded into virions. TF may oligomerize and function as an ion channel. **(B)** 6K is found concentrated at interior membranes where it likely interacts with the viral glycoproteins. The 6K channel properties resulting from oligomerization probably affects the ER. Additionally, 6K may interact with host proteins. (Ramsey & Mukhopadhyay, 2017)

Recently it was discovered that upstream the C-terminal end of 6K is localized a -1 open reading frame that provides the frameshifting required for the synthesis of the TF protein. The heptanucleotide slip site is localised approximately after roughly two-thirds of the 6K gene (Chung *et al.*, 2010, Kendra *et al.*, 2017). Proximal to the frameshift site complex arrangements of secondary structures (stimulatory elements) such as putative pseudoknot-like, RNA hairpin or intra-mRNA secondary structures can be observed upstream a poorly defined spacer (Chung *et al.*, 2010, Firth *et al.*, 2008, Firth *et al.*, 2011). For some alphaviruses, such as SINV, VEEV and BFV viruses; a stem-loop structure 3' adjacent to the hepta-nucleotide slippery sequence UUUUUUA (U₆A) enhances the -1 ribosomal frameshifting, resulting in the production of TF. Nevertheless, the structural element is not ubiquitous amongst the genus because SFV lacks an identifiable local secondary structure proximal to the U₆A motif (Chung *et al.*, 2010).

When frameshifting take place, the E3 and E2 proteins are produced as described above. However, rather than two transmembrane helices in 6K, the TF protein's unique C-terminus is translated into the cytoplasm and the E1 protein is not translated (Ramsey & Mukhopadhyay, 2017). On the contrary, in the absence of frameshift, the hydrophilic

C-terminal sequence of TF is acting as the signal peptide sequence for the E1 protein (Firth *et al.*, 2008). The TF protein preserves at the N-terminal approximately 71 to 83 % of the 6K N-terminal and contains an additional extension of approximately 8 to 50 a.a. at the C-terminal which is hydrophilic and depending on species it can be rich in basic residues (Sanz *et al.*, 2003, Chung *et al.*, 2010, Firth *et al.*, 2008). Thus, frameshifting may be necessary to provide TF with a hydrophilic C-term while maintaining a hydrophobic C-term in 6K to act as the signal peptide sequence for E1 (Greaves & Chamberlain, 2007).

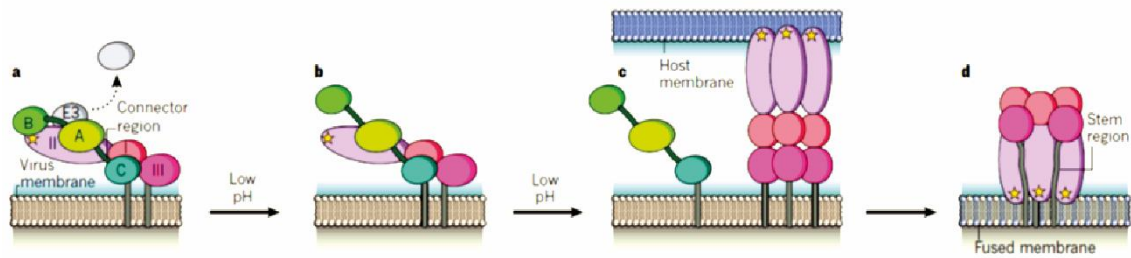


Figure 12. Structural organization and function of *Alphavirus* glycoproteins. (A) The fusion proteins of an alphavirus are connected to the viral membrane and can consist of three components: E1, which contains domains I, II and III, and a fusion loop (star); E2, which contains domains A, B and C, and a ribbon-like connector region; and E3. The alphavirus envelope is peppered with trimers of E2-E1 pairs, although for simplicity only one E2-E1 pair is shown. E3 is a by-product of the cleavage of the E2 precursor protein p62 by the enzyme furin (not shown) and is released in some alphavirus species. **(B)**, On exposure to low pH, domain B and the connector region of E2 move, exposing the E1 fusion loop. **(C)** E1 can then detach from E2 and insert into the host cell membrane to form an extended trimer. **(D)** Finally, to drive membrane fusion, E1 refolds into a hairpin-like structure through the movement of its domain III and the stem region, which lies adjacent to the membrane. For simplicity, only the final fused membrane is shown. (Kielian, 2010)

The TF protein (larger 6 kDa isoform), but not the 6K protein (the smaller 4.2 kDa isoform), appears to be much more heavily palmitoylated on N-terminal cysteines and is incorporated into the virion and perhaps confers stability to the latter, Fig. 11 (Firth *et al.*, 2008, Ramsey *et al.*, 2017, Snyder *et al.*, 2013, Gaedigk-Nitschko & Schlesinger, 1990). It has been suggested that the frameshifting (~10% to 18%) is apparently (~5%) higher than the molar ratio generally found in virions (~3% to 15%) and varies between species, Fig. 9. Thus, it is proposed that varying amounts of TF may be ‘diverted’ *en route* to the developing virion and/or that TF also plays other roles in infected cells, such as budding and membrane permeabilization, Fig. 11. In addition, production of E1 is predicted to be reduced by the level of frameshifting (*i.e.* ~10% to 18% for SFV and SINV), thus leaving an excess of Cp, E3 and E2. (Firth *et al.*, 2008). Low frameshifting rates could also reflect a need for maintaining a high amount of E1 translation for optimal spike assembly and low E2 misfolding levels. (Ramsey *et al.*, 2017)

Accordingly, to what has been reviewed, both 6K and TF are involved in the processing of the protein envelope membrane permeabilization, virus assembly, budding and virion structure, Fig. 11. Concerning the additional hydrophilic C-terminal segment in TF, a role

is probably involved in the viroporin activity and/or virion structure (Firth *et al.*, 2008, Chung *et al.*, 2010). Other possible functions for 6K and TF include the regulation of Cp-E2 binding, E2-E1 heterodimer formation, stability and trafficking, modulation of membrane curvature, Fig. 11 (Brown *et al.*, 2018).

The glycoprotein E1 (type I transmembrane protein) is a class II fusion protein that contains a hydrophobic fusion loop and refolds to drive virus fusion with the endosome membrane which requires low-pH, Fig. 12 (Jose *et al.*, 2009). The proper folding of E1 requires initial binding with pE2 as previously described. The complex pE2:E1 undergoes several intermediate states of folding that require chaperones and disulphide bond formation and exchange, Fig. 12 (Anthony *et al.*, 1992). Moreover, the complex is translocated and processed within the ER and undergoes post-translational modifications. High-mannose chains are added to all potential N-linked glycosylation sites, and the oligosaccharide chains are trimmed depending on the availability of the site (Sefton, 1977, Pletnev *et al.*, 2001). The addition of carbohydrate chains is usually required for proper folding and solubility of the protein and to prevent aggregation (Hsieh & Robbins, 1984, Knight *et al.*, 2009, Jose *et al.*, 2009). In addition, E1 can be also subject to palmitoylation as it has been shown that in SFV and SINV covalent attached palmitic acid was found near the membrane-spanning anchors (Schmidt, 1982, Schmidt *et al.*, 1979, Jose *et al.*, 2009).

This glycoprotein displays the N-terminus in the ER lumen and a short cytoplasmic C-terminal tail following its transmembrane anchor domain, Fig. 12 (Burge & Strauss, 1970, Rice *et al.*, 1982, Strauss & Strauss, 1994). The entire N-terminus ectodomain of E1 is found in the ER lumen, the protein immediately before it in the polyprotein would need to provide a start-transfer signal to translocate E1 into the ER lumen. The E1 ectodomain is composed of three β -barrel domains (DI to DIII), where DI contains the N-terminus and links membrane-proximal DIII to distal DII, and the top of DII contains the fusion loop, Fig. 12. Brown, 2018 #728} The E1 monomers were found at the base of each of the surface spikes and form a lattice on the virus surface. Residues from Pro383 to Trp409 comprise the E1 stem region, and the transmembrane helix of E1 enters the bilayer at the residue Trp409 and exits at residue Met433. The six carboxy-terminal residues of E1 extend past the inner lipid leaflet into the interior cavity of the virus (Jose *et al.*, 2009, Mukhopadhyay *et al.*, 2006).

4.2. Alphavirus assembly and budding

During virus assembly and budding are required binding of Cp with E2, E2:E1 heterodimer formation, pH protection of E1 by pE2 and spike lattice assembly. Also,

assembly and budding do not require pE2 cleavage (Zhang *et al.*, 2003) (Salminen *et al.*, 1992, Jain *et al.*, 1991, Lobigs & Garoff, 1990), nor packing of the genomic RNA (Frolov *et al.*, 1997, Suomalainen *et al.*, 1992, Pushko *et al.*, 1997). In rare cases infectious microvesicles containing viral RNA and envelope proteins can form in the absence of Cp (Ruiz-Guillen *et al.*, 2016, Jia *et al.*, 2017, Brown *et al.*, 2018). Interestingly, Cys palmitoylation in E2 endodomains is thought to orient the endodomain in a Cp-competent binding conformation (Ivanova & Schlesinger, 1993, Zhao *et al.*, 1994, Brown *et al.*, 2018). Several models have been suggested and some propose that the pre-assembled NCs bind E2 at the plasma membrane to induce spike lattice formation and particle budding (Suomalainen *et al.*, 1992). Other models propose that envelope proteins can interact with Cp prior to NC formation and induce particle budding (Snyder *et al.*, 2012, Brown *et al.*, 2018).

The mature Cp protein forms a lattice of pentamers and hexamers (called capsomers) within the icosahedral lattice. The first 100 a.a. which are highly basic provide to the Cp protein the ability to bind the newly synthesized viral RNA (Brown *et al.*, 2018, Ramsey & Mukhopadhyay, 2017) (Coombs & Brown, 1989).. Alphavirus RNA also have a specific packing signal utilized such that only viral RNA is encapsidated during NC formation (Weiss *et al.*, 1989). The packing signal is recognized by the CP, and this interaction is likely an early step in NC formation (Tellinghuisen & Kuhn, 2000, Linger *et al.*, 2004, Wengler & Wengler, 1984). The position of the packing signal in the viral RNA varies by species but is typically found in the 5' half of the genome such that only full-length RNA is packed. A single copy of RNA genome forms a complex with 240 copies of Cp (Jose *et al.*, 2009).

Thus, the capsomers encapsidate one single molecule of viral RNA and form the NC. A hydrophobic pocket is localized at the C-terminal domain of the Cp protein and binds a cdE2. Thus the surface spikes are linked to the internal NC core through the cdE2 (Lee *et al.*, 1996, Skoging *et al.*, 1996) which extends from the end of the E2 transmembrane helix into the NC shell, ending near to the hydrophobic pocket in the Cp protein where cdE2 residues are placed (Skoging *et al.*, 1996, Lee *et al.*, 1996, Wilkinson *et al.*, 2005). (Jose *et al.*, 2009).

The interaction of E1 with pE2 (Wahlberg *et al.*, 1989) protects E1 from prematurely fusing in the acidic environment of the trans-Golgi network (~pH 6.0). The prevention of the exposure of the E1 fusion loop is mediated by a conserved Tyr residue in the E3 portion from the pE2 that stabilizes the interaction of E2 with E1 (Voss *et al.*, 2010, Yap *et al.*, 2017, Snyder & Mukhopadhyay, 2012, Brown *et al.*, 2018).

Cp and E2 could interact at an early stage prior to NC formation (Snyder *et al.*, 2012). There are two hypotheses: The interaction may occur soon after translation and Cp must bind E2 in *cis*-dependent manner, and the complex is then co-transported to the plasma membrane via the secretory pathway (Jose *et al.*, 2012, Zheng & Kielian, 2013). Alternatively, the interaction may occur later within the type II cytopathic vacuoles (CPVII) that co-transport NCs and E1:E2 glycoproteins to the plasma membrane (Soonsawad *et al.*, 2010, Brown *et al.*, 2018).

The maturation of the E1:E2 heterodimers was already described and after being transported to the plasma membrane the heterodimers self-assemble into 80 trimeric spikes on the virus surface (Schmidt *et al.*, 1979, von Bonsdorff & Harrison, 1978). It is important to remember that uncleaved pE2 can also be efficiently transported and incorporated into virus particles. However, the cleavage of pE2 is required to generate infectious particles by making viral particles competent for entry and fusion activation in new cells (Salminen *et al.*, 1992, Heidner *et al.*, 1994).

Although, 6K and/or TF may support an accessory role for 6K/or TF during spike maturation, where they are not required for budding but enhance assembly, perhaps by stabilizing spikes (Loewy *et al.*, 1995, Ramsey *et al.*, 2017). The presence of TF in the *wt* virion and in the infected cell probably may function to exclude host proteins, corral spikes into the optimal budding conformation, or even interact with 6K (Ramsey & Mukhopadhyay, 2017). It has been also proposed that 6K (Rice *et al.*, 1987) acts as a spacer during spike assembly. Thus, 6K would aid in glycoproteins traffic and spike assembly and mediate the E2:Cp interaction necessary for proper budding (Gaedigk-Nitschko & Schlesinger, 1991, Ramsey & Mukhopadhyay, 2017).

Once the particles are formed, their release will be mediated through budding and alphaviruses use an endosomal sorting complexes required for transport (ESCRT) independent mechanism, and it is not known if other host proteins are required or if a viral protein aids in budding (Taylor *et al.*, 2007, Yondola & Carter, 2011). Moreover, budding is significantly enhanced by 6K, TF and cholesterol through an unknown mechanism (Lu & Kielian, 2000, Marquardt *et al.*, 1993, Vashishtha *et al.*, 1998, Ramsey & Mukhopadhyay, 2017, Liljestrom *et al.*, 1991, Lusa *et al.*, 1991, Loewy *et al.*, 1995, Kozlov *et al.*, 2010, Brown *et al.*, 2018). Physiological temperatures and neutral to mildly alkaline pH are the optimal budding conditions (Lu & Kielian, 2000, Lu *et al.*, 2001, Brown *et al.*, 2018).

The budding mechanism used by alphaviruses is still under discussion and to date two mechanisms have been proposed (Ramsey & Mukhopadhyay, 2017).

- i. The budding is driven by an association between viral glycoproteins followed by interactions with the capsid and RNA.
- ii. Pre-formed nucleocapsid cores drive interactions between glycoproteins that result in budding.

The host factors involved during assembly and budding has been poorly investigated. It has been suggested that E2:E1-containing CPVII vacuoles are trafficked to the cell surface along actin filaments by a mechanism involving the proteins: Rac family small GTPase 1 (Rac1), actin-related protein 3 (Arp3), and the lipid kinase phosphatidylinositol-4-phosphate 5-kinase1- α (PIP5K1- α). Thus, this mechanism could explain the trafficking of the alphavirus glycoproteins to localized sites of budding (Brown *et al.*, 2018). However, extensive protein interactions among the envelope proteins and the NC could be sufficient to drive both membrane curvature and scission (Kozlov *et al.*, 2010, Weissenhorn *et al.*, 2013, Snead *et al.*, 2017, Brown *et al.*, 2018).

During budding, alphaviruses induces a dramatic cytoskeletal remodelling at the budding site of the host cell, including the formation of two different types of filopodia-like extensions, the short- and long- extensions which are distinguished by their length, contacts and components, Fig. 13. Evidences from early works have suggested that nsP1 is present in filopodia and probably may induce the formation of filopodia through membrane association mediated by an amphipathic helix in nsP1 (Ahola *et al.*, 1999,

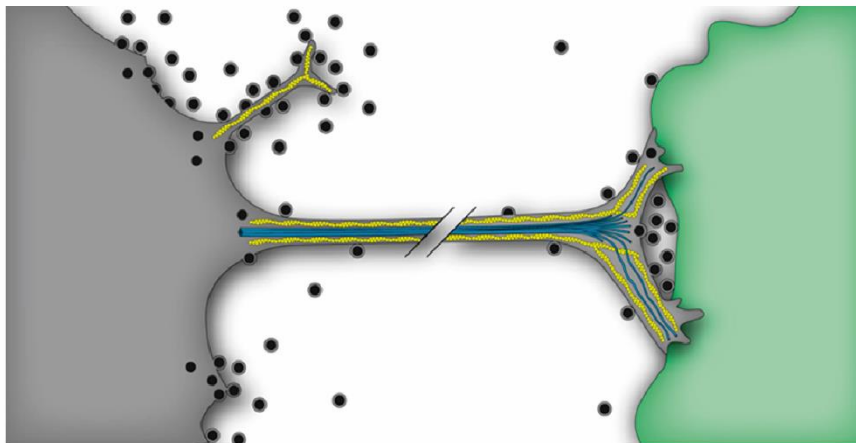


Figure 13. Schematic model of specialized budding sites of an alphavirus infected cell (gray). Alphaviruses assemble and bud from localized patches at the plasma membrane of the cell body, where host proteins are excluded, and Cp/NC serves as a scaffold to accumulate E2/E1 glycoproteins. Dramatic cytoskeletal remodelling occurs during infection, and short filopodia-like extensions (top) and long intercellular extensions (middle) are observed. Short extensions contain F-actin (yellow) only, while intercellular extensions contain F-actin and tubulin (blue) structures. Nascent virus particles bud along the short extension, which may facilitate dissemination from the infected cell. An intercellular extension projects from the infected cell and makes preferential contact with an uninfected cell (green). The contact site forms a stable, flattened tip on the target cell plasma membrane (right), producing a protected pocket where nascent particles bud and are subsequently internalized by the uninfected cell. Thus, intercellular extensions are close-ended membrane bridges that mediate cell-to-cell transmission. Host determinants involved in modulating the cytoskeleton, forming/stabilizing extensions, and promoting cell-to-cell virus transmission are currently unknown. (Brown *et al.*, 2018).

Spuul *et al.*, 2007) and by nsP1 palmitoylation (Spuul *et al.*, 2007, Ahola *et al.*, 2000, Karo-Astover *et al.*, 2010, Laakkonen *et al.*, 1996, Zusinaite *et al.*, 2007).

The filopodia are known as specialized sites of budding (Brown *et al.*, 2018). The short extensions are approximately 2-7 μm in length, it do not contain tubulin but mostly F-actin (Martinez *et al.*, 2014, Birdwell *et al.*, 1973), and may contain branching at distal ends, Fig. 13.

The long intercellular extensions are projected from infected cells to physically contact neighbouring cells (Martinez *et al.*, 2014, Martinez & Kielian, 2016) involving structures termed virological synapses, mediated by cell-cell contacts, such as tight junctions. These extensions are more than 10 μm in length, usually within 10-60 μm range and are tubulin and F-actin positive, specially at the root, but often display short branches that are tubulin and F-actin negative (Brown *et al.*, 2018). The long intercellular extensions are relatively large in diameter and usually detected once virus structural protein production is well underway (Martinez & Kielian, 2016) and they do not fuse with the target cell and do not mediate membrane or cytoplasm continuity between the two cells. The intercellular extensions of alphaviruses originate exclusively from infected cells and from stable physical contacts with neighbouring cells (Martinez & Kielian, 2016).

Host factors can also prevent alphavirus release such as tetherin (also known as BST-2) which is an interferon-inducible host membrane protein that prevents the release of enveloped viruses. The used mechanism involves direct tethering of budded particles to the plasma membrane (Neil, 2013, Sauter, 2014, Brown *et al.*, 2018).

Some viruses have developed strategies to overcome the broad antiviral activity of tetherin by encoding viral antagonists (Neil, 2013, Sauter, 2014). Alphaviruses are no exception and it has been suggested that nsP1 downregulates tetherin expression to promote VLP release (Jones *et al.*, 2013, Brown *et al.*, 2018).

4.3. Translation of Non-structural proteins

The non-structural proteins are immediately translated from the non-structural ORF after release of the alphavirus genomic RNA into the host cell, which occurs after fusion of the virus envelope with the endosome membrane. The translation of the non-structural proteins takes place into the cytosol where, depending on the alphavirus strain, one or two non-structural polyproteins are produced and further cleaved to produce the mature non-structural proteins nsP1, nsP2, nsP3 and nsP4 (Strauss & Strauss, 1994). In several alphavirus strains an opal codon (UGA) is found at the end of *nsP3* gene which regulates the expression of putative RNA polymerase nsP4 by a readthrough mechanism. Thus,

two non-structural polyproteins (P123 and P1234) are produced, however in some alphaviruses strains the opal codon is absent, replaced by a sense codon, and only P1234 is produced (Strauss & Strauss, 1994).

Beier and Grimm (2001) classified readthroughs in three types. Accordingly, the readthrough regions of alphavirus were classified as type II because immediately downstream the opal termination codon (UGA) is present either the codon CGG or CUA (Beier & Grimm, 2001, Firth *et al.*, 2011). The CUA codon immediately downstream of the UGA codon is sufficient for efficient translational readthrough (Li & Rice, 1993) and, in some alphaviruses such as SINV, the readthrough requires only a cytidine residue immediately downstream of the UGA codon (Li & Rice, 1993).

Despite the already reported role of the cytidine and the CGG or CUA codon immediately downstream of the UGA codon, an additional element was identified as important for the readthrough mechanism (Firth *et al.*, 2011). In VEEV and SINV a stem-loop of approximately 140 nt was identified downstream the UGA-CGG or UGA-CUA sequences respectively. The stem-loop presented a stimulatory effect on readthrough of 14-fold for VEEV and between 3 to 4-fold for SINV (Firth *et al.*, 2011). Thus, the findings of Firth and colleagues suggested that a highly conserved RNA structural element localized 3' upstream the UGA termination codon is also fundamental for the readthrough mechanism.

The occasional readthrough of the opal codon occurs with 5 to 20% efficiency (determined with *in vitro* translation reactions), which can be suppressed by several aminoacyl-tRNAs, and is responsible for the translation of P1234, Fig. 14 (Li & Rice, 1989, de Groot *et al.*, 1990, Firth *et al.*, 2011). Li and Rice, 1993, have suggested that in SINV the readthrough of the UGA codon may result from misreading by the cellular natural suppressor tRNA^{Trp} (anticodon U*CA) (Beier & Grimm, 2001, Li & Rice, 1993). However, a natural occurring Arg or Cys at the same position of the opal termination codon was already observed in other alphavirus studies. In some isolates of SFV (Takkinen, 1986, Strauss & Strauss, 1994), ONNV (Strauss *et al.*, 1988, Levinson *et al.*, 1990) and CHIKV (Jones *et al.*, 2017, Chen, Kam, *et al.*, 2013) the opal stop codon (UGA) was substituted by an Arg (codon CGA) (Tuittila *et al.*, 2000) and in SINV strain AR86 (Suthar *et al.*, 2005) was replaced by a cysteine. In these reported cases, the replacement of the opal stop codon by a sense codon confers a constitutive translation of the P1234. Early studies from the outbreak ONNV 1996- 1997 have suggested that ONNV circulates as a viral quasi-species containing both opal and Arg codons. However, the same study also reported that at the fifth passage in Vero cells the replacement of

the opal stop codon for an arginine was observed in the isolate ONN-1854 (Lanciotti *et al.*, 1998).

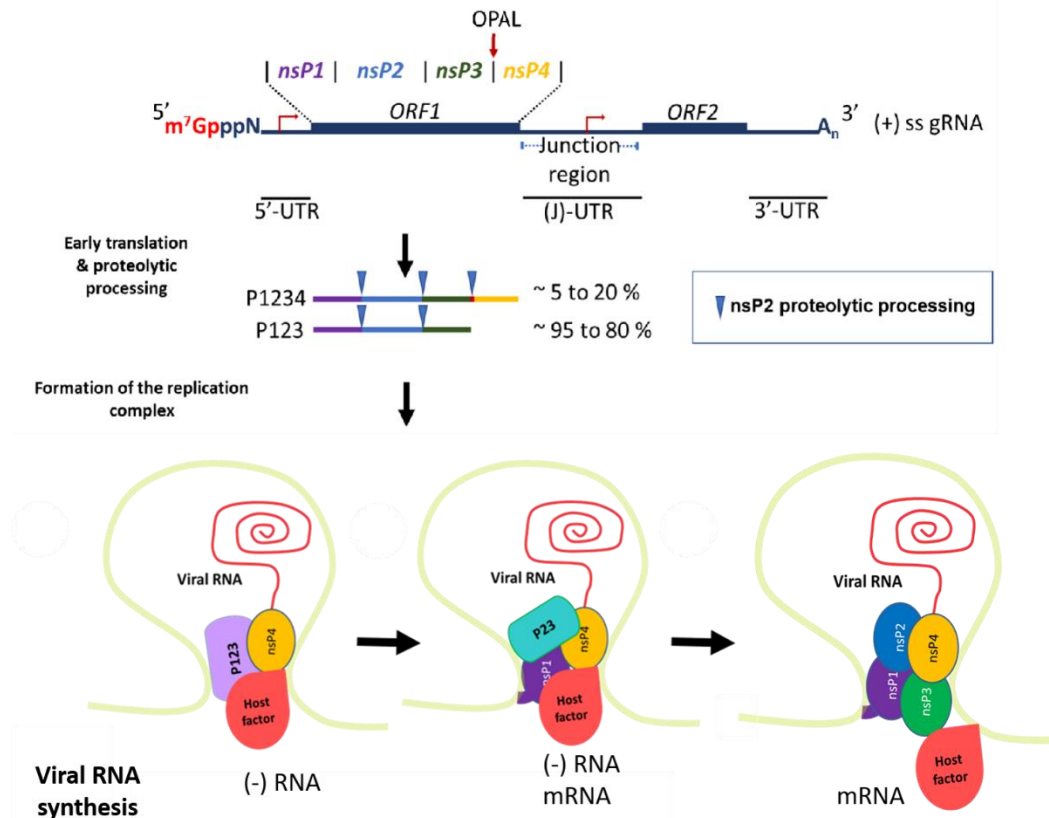


Figure 14. The major steps during the formation of the Alphavirus replication complex. The *ORF1* from the genomic RNA belonging to alphaviruses is directly translated into two non-structural polyproteins. The P1234 represents between 5 to 20% of the translation and the P123 is the major product of translation representing more than 80% of translation of the *ORF1*. The regulation of the translation of both polyproteins is achieved due to the presence of an OPAL codon localized at the end of the coding region for the non-structural protein 3 (nsP3). The polyproteins are further processed by proteolytic cleavage which is mediated by non-structural protein 2 (nsP2). The replication complex is assembled involving the viral non-structural polyproteins/proteins and host factors (1) The first cleavage occurs at the junction nsP3/nsP4, thus releasing the mature nsP4 which together with P123 forms the early replication complex. This early replication complex synthesizes mainly negative stranded RNA. (2) the maturation of the non-structural proteins proceeds and the junction between nsP1/nsP2 is cleaved and this cleavage is responsible for the switch in the syntheses from negative to positive polarity RNA. (3) Fully matured non-structural proteins (nsPs) occurs after the last cleavage between the junction nsP2/nsP3 and forms the late replication complex that uses the negative stranded RNA as a template to synthesize the RNA of positive polarity which are the sub-genomic RNA and genomic RNA.

Together with readthrough elements, the opal codon may play a regulatory role in alphavirus replication and infectivity (Li & Rice, 1993). It was suggested that in SINV strain AR86 the replacement of the opal stop codon for a sense codon (Cys) was a major determinant of neurovirulence (Suthar *et al.*, 2005). Studies performed with ONNV strain SG650, originally isolated from serum during the 1996 outbreak in Uganda, suggested that the natural occurring replacement of the opal stop codon (UGA) for an Arg (codon CGA) was associated with decrease in infectivity and most likely the opal stop codon may display a regulatory role in alphavirus replication (Myles *et al.*, 2006). However, a study using a Caribbean CHIKV strain from Sri Lanka (originally isolated from a patient

in St. Martin during the outbreak 2013, GenBank MG208125), that encodes both conserved opal termination codon and the Arg sense codon, concluded that the opal codon is important at the level of pathogenesis independently of effects on viral replication (Li & Rice, 1989, Jones *et al.*, 2017). These findings somehow are close to the observations of Myles *et al.* 2006, but contradictory to those of Tuittila *et al.* 2000, who reported that the opal codon is the main responsible for attenuation of virulence in SFV (Tuittila *et al.*, 2000).

Lastly, a survey of a CHIKV intra-outbreak analysis compared isolates with low passages *in vitro* from the epidemics in Reunion Island 2005-2006 with the prototype isolate in 1952 in Tanzania named S27 isolate (with multi passages *in vitro*). This work observed the opal stop codon (UGA) in the isolates from the Reunion Island 2005-2006 instead of the Arg (codon CGA). Therefore, it was proposed that is probably profitable for the viral quasi-species to maintain the opal codon *in vivo* (Schuffenecker *et al.*, 2006). According to Schuffenecker *et al.* 2006, changing to the Arg codon may confer a selective advantage *in vitro* as observed in previous works (Kim *et al.*, 2004, Lanciotti *et al.*, 1998). Their results suggest that probably both codons might be associated with fitness advantage but at different time in the transmission cycle due to alternate replication in both vertebrate and invertebrate hosts (Schuffenecker *et al.*, 2006). The evidences from the work of Schuffenecker 2006 and colleagues contributed to support the hypothesis already proposed by others (Myles *et al.*, 2006, Lanciotti *et al.*, 1998) that evolutionary pressure is probably the key for the presence either an opal termination or sense codon.

Once synthesised, the precursor polyproteins (P1234 and P123) are cleaved by a carboxyl-terminal protease domain of nsP2 (de Groot *et al.*, 1990). First the cleavage occurs in P1234 either in *cis* or *trans* at the junction of nsP3/4, followed by the nsP1/2 junction which occurs in *cis* only (Vasiljeva *et al.*, 2003). The last cleavage event occurs at the junction nsP2/3 producing fully mature nsPs, Fig. 14.

4.3.1 The alphavirus non-structural protein 1 (nsp1)

The alphavirus nsP1 protein is highly conserved among alphaviruses (Rozanov *et al.*, 1992). Its molecular weight is approximately 60 kDa and the nsP1 enzyme carries at the N-terminal signature sequence motifs of a Rossman fold-like methyltransferase (MTase), (Martin & McMillan, 2002, Rozanov *et al.*, 1992, Schluckebier *et al.*, 1995, Cross, 1983, Cross & Gomatos, 1981) and guanylyl-transferase (GTase) (Ahola & Karlin, 2015), which participate in the alphavirus mRNA capping, Fig. 15. The nsP1 MTase activity was reported for the first time from a work in which a nsP1 mutant of SINV was used which replicated in insect cells depleted of the SAM precursor (methionine) (Mi *et al.*, 1989).

Later, *in vitro* enzymatic assays were performed in order to monitor the MTase activity of SINV and SFV nsP1, and critical residues for the MTase activity were then identified (Laakkonen *et al.*, 1994, Mi & Stollar, 1991, Rozanov *et al.*, 1992). The guanylylation (GT) activity was later studied using lysates of SFV and SINV infected cells that were incubated with guanosine triphosphate labelled on the α -phosphate group with ^{32}P ($[\alpha^{32}\text{P}]\text{-GTP}$). Then, a covalent enzyme-guanylate complex was characterized by using SFV nsP1 expressed in recombinant baculovirus-infected cells (Ahola & Kaariainen, 1995). The critical residues involved in guanylyl-transferase-like activities were later identified together with some other residues crucial for the methyltransferase activity of nsP1 in SFV (Ahola *et al.*, 1997).

Nsp1 can exhibit its activity either as a mature protein or in the form of the precursors P123 or P1234 (Salonen *et al.*, 2003). Interestingly, the order and mechanisms of both enzyme activities was found unconventional (see below the chapter 5 intitled “RNA capping pathways”): The MTase motif of nsP1 catalyses the transfer of a methyl group from SAM to the N⁷ position of a GTP molecule leading to the formation of m⁷Gppp. Then, the GTase motif binds the m⁷Gppp, forming a covalent bond with a highly conserved catalytic histidine (m⁷Gp-GTase), releasing PP_i (Ahola & Kaariainen, 1995, Ahola *et al.*, 1997, Lin *et al.*, 2012). Lastly, the m⁷Gp molecule is transferred to a 5'-diphosphate RNA to create m⁷GpppNp-RNA, reviewed in (Decroly *et al.*, 2011). The resulting cap structure is essential for viral mRNA translation and prevents the mRNA from being degraded by cellular 5' exonuclease (Abu Bakar & Ng, 2018).

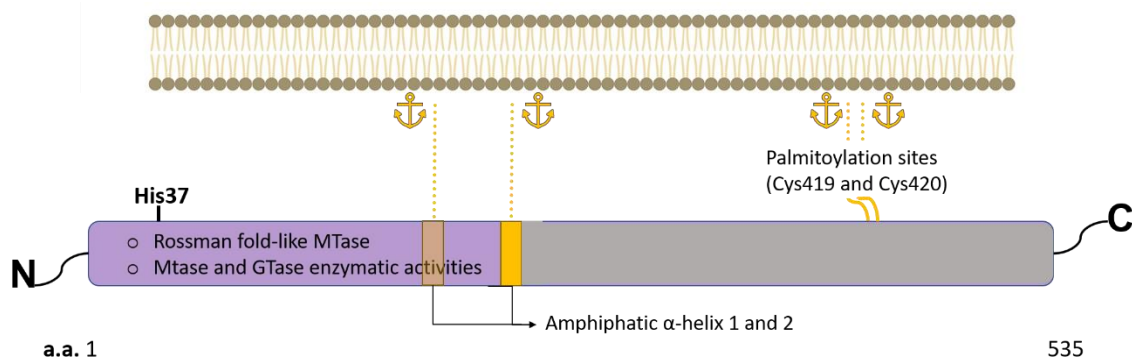


Figure 15. Domain organization of the matured non-structural protein 1 (nsP1). The purple domain represents the domain with a Rossman fold and that harbours the activities of methyltransferase (MTase) and guanylyltransferase (GTase). The His37 represents the histidine residue determined to bind covalently to the m⁷GMP moiety. The membrane-binding amphipathic helix are represented in the yellow rectangle and the palmitoylation sites are represented with text and yellow lines. Both membrane-binding amphipathic helix and palmitoylation site are membrane anchor sites. Amino acid numbers are indicated for VEEV nsP1 strain TC83. For more details see text. Adapted from (Rupp *et al.*, 2015)

Downstream the N-terminal domain are present features such as an amphipathic α -helix and cysteine palmitoylation sites required for association of nsP1 to the membranes. Both the amphipathic α -helix and cysteine palmitoylation sites are required to anchor

nsP1 and assemble the alphavirus replication complex to the host membrane, Fig. 15 (Ahola *et al.*, 1999, Ahola *et al.*, 2000, Laakkonen *et al.*, 1996, Lampio *et al.*, 2000, Peranen *et al.*, 1995, Spuul *et al.*, 2007). This occurs most likely through nsP1 interaction with the membrane's anionic phospholipids, reviewed in (Abu Bakar & Ng, 2018). Studies have reported that nsP1 alone is still targeted to the inner surface of the plasma membrane, however it is not enough for cytoplasmic vacuole formation (Peranen *et al.*, 1995). The association of nsP1 with lipids is necessary for its enzymatic activity depending on the alphavirus strain. For example, the enzymatic activity of SINV nsP1 is functional in the absence of lipids (Tomar *et al.*, 2011), unlike the nsP1 from SFV that was found to require the presence of those lipids (Ahola *et al.*, 1999). NsP1 is also post-translational modified by acylation but the latter remains to be understood functionally (Laakkonen *et al.*, 1996).

Concerning the palmitoylation, this post-translational modification of nsP1 is not essential for nsP1 enzymatic activity (Laakkonen *et al.*, 1994, Mi & Stollar, 1991). However, it has been reported that depalmitoylation mutants are associated to decreased pathogenesis in mice (Ahola *et al.*, 2000) which is likely associated to other alternative functions of nsP1 such as membrane and cytoskeletal rearrangements, development of cell filopodia, and cell-to-cell transmission of alphaviruses (Karo-Astover *et al.*, 2010, Laakkonen *et al.*, 1998). Jose *et al.* work confirmed by transmission electron microscopy (TEM) and immunofluorescence (IF) the localization of the non-structural proteins at the plasma membrane (PM) and internal vesicles. NsP1 was found to localize in filopodial extensions suggesting a role for nsP1 in transport of replication complexes to the PM and host actin modifications (Jose *et al.*, 2017).

NsP1 is important for minus-strand RNA synthesis (Hahn, Grakoui, *et al.*, 1989, Wang *et al.*, 1991) which depends on interactions of nsP1 within the replication complex. It has been reported that some mutations in nsP1 can negatively affect the minus-strand RNA synthesis without negatively impacting nsP1 enzymatic activity (Lulla *et al.*, 2008). In addition, studies have reported strong interactions between nsP1 and nsP4 as well as between nsP1 and nsP3. It is likely that nsP1 also establishes weak interactions with nsP2 (Salonen *et al.*, 2003, Zusinaite *et al.*, 2007). Interestingly, it has been suggested that nsP1 also regulates the proteinase activity of nsP2 and that nsP1/nsP2 cleavage is required before cleavage of nsP2/nsP3 (de Groot *et al.*, 1990).

4.3.2 The alphavirus non-structural protein 2 (nsp2)

NsP2 is a non-structural protein of approximately 90 kDa which is required for transcriptional and translational shutoff synthesis of the host macromolecules (Frolov *et*

al., 1999, Garmashova *et al.*, 2006, Gorchakov *et al.*, 2005, Kim *et al.*, 2004). nsP2 contributes to the inhibition of IFN-mediated antiviral responses and regulation of the translational machinery by viral factors, Fig. 16 (Gorchakov *et al.*, 2005, Breakwell *et al.*, 2007, Frolov *et al.*, 2009, Bhalla *et al.*, 2016). Indeed, it has been reported that nsP2 effectively inhibits the janus kinase-signalling transducer and activator of transcription (JAK-STAT) signalling pathway in CHIKV, Fig. 16, (Fros *et al.*, 2010) involved in upregulation of IFN-stimulated (antiviral) genes (ISG).

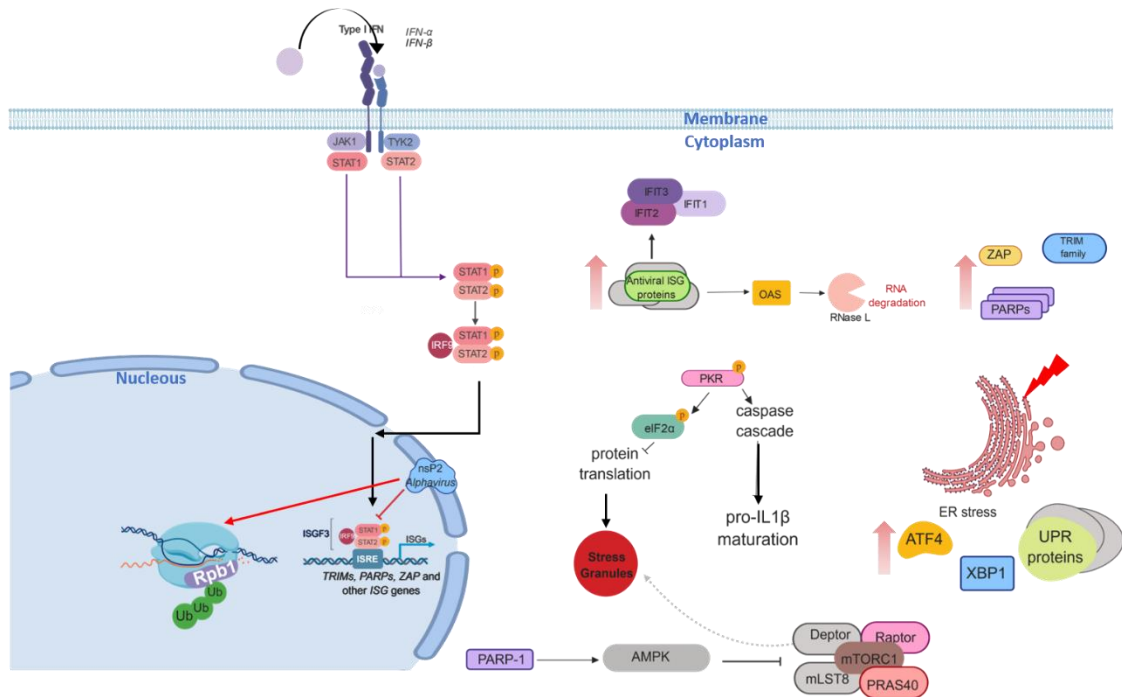


Figure 16. The anti-host effect of nsP2 during alphaviruses infection. The nsP2 is involved in host cell translation shut-off because inhibits the binding of the complex STAT1/STAT2/IRF9 complex to the interferon stimulated responsive element (ISRE) which is activated by the type I IFN through the JAK-STAT pathway. Therefore, nsP2 inhibits the expression of interferon stimulated genes (ISGs) such as *PARP1* and *ZAP*, *IFIT* family, *TRIM* family, *OAS* which activates RNase L, *ATF4*, *XBP1* and *UPR* genes. Indirectly, nsP2 inhibits the activation of PKR kinase and this leads to protein translation inhibition. The activation of the caspase cascade is also affected, and this can compromise the maturation of pro-IL1 β into IL1 β . The dysregulation of *PARP* genes expression and alterations in the normal levels of host protein translation can induce the formation of stress granules. The stress granules are important structures where the inactive mTORC1 complex is stored when host cell is under stress. The mTORC1 is inactivated by AMPK and activation of AMPK requires PARP1. Indirectly, the decreased expression of PARP1 may lead to less inhibition of mTORC1. Another line to counteracts the host cell response by nsP2 is the induction of degradation mediated by proteasome of the subunit Rpb1 that belongs to the RNA polymerase II. For more details see text. Figure designed using <https://biorender.com/>.

In SINV the host cell shutoff induced by nsP2 also affects the expression of IFNs and ISGs (Frolova *et al.*, 2002, Gorchakov *et al.*, 2005). Other advantage of alphaviruses in inducing host cell translation shutoff is the inhibition of the unfolded protein response (UPR) by preventing the expression of activating transcription factor 4 (ATF4), active X-box binding protein 1 (XBP1), and additional UPR target genes, Fig. 16 (Dryga *et al.*, 1997, Tamm *et al.*, 2008). The UPR is a response activated by stress of the ER that occurs due to the presence of unfolded or misfolded proteins in the ER and can also

occur due to the presence of large amounts of viral glycoproteins. The sensing of unfolded or misfolded proteins in ER leads to the phosphorylation of the eukaryotic initiation factor 2 α subunit (eIF2 α), and subsequent translation repression of general protein expression, although the expression of stress-related proteins such as ATF4 is upregulated (Fros *et al.*, 2015), reviewed in (Fros & Pijlman, 2016).

The host cell translational shutoff during alphavirus infections was proposed to be independent of protein kinase R (PKR) and eIF2 α phosphorylation, Fig. 16, and most likely the nsP2 may alter ribosomes by association with ribosomal protein S6 of vertebrates and mosquitoes (Montgomery *et al.*, 2006). Interestingly, it has been reported that in VEEV the host cell transcription shutoff is aided by the Cp (Garmashova, Atasheva, *et al.*, 2007) and that nsP2 is also involved in the packing of the viral genome into the infectious virions (Kim *et al.*, 2013). Concerning the nsP2 transcriptional host shutoff, it probably functions through degradation mediated by ubiquitination of the host DNA-directed RNA polymerase II subunit Rpb1 (Rpb1), a catalytic subunit of the RNA polymerase II polymerization complex, Fig. 16 (Akhrymuk *et al.*, 2012). Furthermore, a fraction of nsP2 (approximately 50 %, (Peranen *et al.*, 1990)) localizes to the nucleus and blocks cellular RNA export to the cytoplasm (Breakwell *et al.*, 2007, Rikonen *et al.*, 1994a). The translocation of nsP2 between the cytoplasm and the nucleus is in part due to the presence of two canonical nuclear localization signals (NLS) in the nsP2 sequence, Fig. 17.

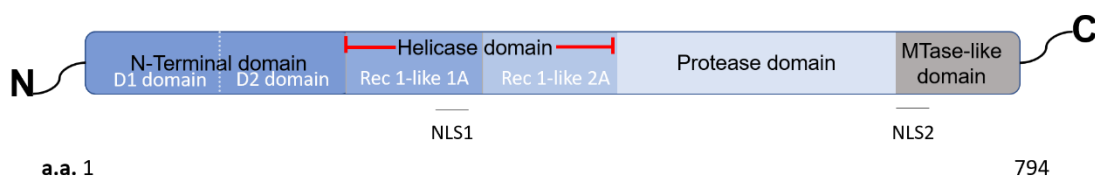


Figure 17. Domain organization of the mature non-structural protein 2 (nsP2). Organized in four major domains, nsP2 harbours an N-terminal domain, a helicase domain, a protease domain and an inactive MTase-like domain at the C-terminal portion of nsP2. The N-terminal domain is divided in D1 and D2 domains, the first with co-factor properties and the second likely involved in promoter selection. The helicase domain also is composed of two domains, Rec 1-like 1A and Rec1-like 2A. the helicase domain also presents NTPase activity. The NLS1 and NLS2 are the sites of canonical nuclear localization sites. The amino acid range is indicated for VEEV strain TC83. For more details see text. Adapted from (Rupp *et al.*, 2015).

This protein has been described to present three main important functions which are known as helicase, RNA triphosphatase (RTPase) and protease. However, its structural and functional organization is still under investigation. In the amino proximal region (also known as N-terminal domain) two putative domains were identified through mutational analysis, Fig. 17 (Atasheva *et al.*, 2007, Frolov *et al.*, 1999). The first putative domain (D1) was reported to display cofactor-like properties regarding the activity of the protease

domain. Concerning the second putative domain (D2) it was proposed to function in promoter selection (Rupp *et al.*, 2015).

The helicase domain is immediately next to the second putative domain (D2) and displays also nucleoside triphosphatase (NTPase) activity, Fig. 17. The helicase has seven signature motifs of superfamily1 (SF1) helicase (Gorbalenya *et al.*, 1989). The function of the helicase is dependent on the NTPase activity of the N-terminal domain as the latter provides energy to the RNA helicase activity (Karpe *et al.*, 2011, Rikkonen *et al.*, 1994b). The helicase function is supposedly to unwind RNA secondary structures formed during viral RNA replication (Gomatos *et al.*, 1980, Gomez de Cedron *et al.*, 1999, Gorbalenya *et al.*, 1988). Early experiments have identified mutations in the Walker A motif that negatively impact helicase activity (Rikkonen *et al.*, 1994b). Likely, helicase activity supposedly acts in coordination with the polymerase activity of nsP4 (Das *et al.*, 2014). The RTPase activity is associated with the N-terminal domain and is responsible for removal of the γ -phosphate from the 5' end of the nascent positive-sense RNAs to yield a diphosphate moiety at the 5' terminus. Thus, nsP1 is able to use the 5' diphosphate RNA as a substrate for the capping reaction (Vasiljeva *et al.*, 2000, Karpe *et al.*, 2011). The RTPase activity seems to be dependent on the same active site as the NTPase activity. Recently it has been also reported that the adenosine triphosphatase activity (ATPase) activity of nsP2 is enhanced in the presence of nsP1 (Kumar *et al.*, 2018).

Immediately upstream the C-terminal end of the helicase domain is found the protease domain which is crucial for the processing of the non-structural proteins, Fig. 17 (Hahn, Strauss, *et al.*, 1989, Hardy & Strauss, 1989). The protease domain does not function independently as the other nsP2 domains are also required for the correct processing of the non-structural proteins (Vasiljeva *et al.*, 2003). Thus, the protease domain is essential to virus replication (Lulla *et al.*, 2006). A crystallographic structure of the protease domain was already reported and shows similarity to known cathepsins but presenting a distinct cysteine protease fold (a novel α/β fold) (Russo *et al.*, 2006), recently proposed as a papain-like cysteine protease (Ramakrishnan *et al.*, 2017). The efficiency of this domain in the processing of the polyproteins is mainly dependent on the concentration of the polyprotein. For example, early in infection the concentration of the non-structural proteins is low which favour the *cis* cleavage, while later *trans* cleavage increases because the concentration of protease is high. The cleavage efficiency *cis/trans* is different for each of the three junctions. For example, the cleavage of the junction 2/3 is poor within the protein containing nsP1. Thus, there is a preference for the cleavage of the junctions nsP3/nsP4 (in *cis*) and nsP1/nsP2 (in *trans*) early in infection and the

cleavage of nsP2/nsP3 (in *trans*) later in infection (de Groot *et al.*, 1990, Shirako & Strauss, 1990), reviewed in (Rupp *et al.*, 2015, Jose *et al.*, 2009). In agreement with structural works, (Hu *et al.*, 2016, Russo *et al.*, 2006), in VEEV nsP2 the predicted active site for polyprotein processed is localized in a major surface groove which probably accommodates the substrate polyprotein to be cleaved. It has been suggested that the major enzyme groove may act as an enzyme mouth holding the protein to be processed, reviewed in (Abu Bakar & Ng, 2018).

Lastly, a SAM-dependent MTase-like domain presenting a classical MTase fold sharing structural similarity to FtsJ MTase, a 2'-O methyltransferase, was identified next to the end of the protease domain and proximal to the carboxy terminus of nsP2, Fig. 17 (Russo *et al.*, 2006). The MTase-like domain of nsP2 lacks the classical active site residues, the Lys-Asp-Lys-Glu tetrad, responsible for a 2'-O methyltransferase activity (Hu *et al.*, 2016). Thus, it has been suggested that the MTase-like domain probably does not have enzymatic activity. However, it has been proposed that the MTase-like domain may be involved in regulation of the minus-strand synthesis and is probably involved in the development of cytopathic effects (Mayuri *et al.*, 2008).

Despite the functions already mentioned for nsP2, it has been reported that nsP2 can act as transcription factor for sub-genome synthesis by binding to the sub-genomic promoter in the minus strand template (Hahn, Grakoui, *et al.*, 1989, Sawicki *et al.*, 1978, Suopanki *et al.*, 1998). Moreover, studies have reported the negative impact of mutations in nsP2 that alter the switching from the minus-strand to positive-sense RNA synthesis (De *et al.*, 1996), probably due to alterations in the nsP2 protease activity, or conformational modifications in the replication complex, or alteration of host effects including Ribonuclease L (RNase L) activation (De *et al.*, 1996, Sawicki *et al.*, 2006).

4.3.3 The alphavirus non-structural protein 3 (nsp3)

The precise role(s) of nsP3, a viral protein of approximately 60 kDa, during replication is still unknown. Several works have reported that nsP3 is a hub for multiple host protein interactions and it has been reviewed recently in (Gotte *et al.*, 2018). It has been proposed that nsP3 (in the context of polyprotein or in its mature state) is necessary for minus-strand or sub-genomic RNA synthesis (LaStarza, Lemm, *et al.*, 1994, Rupp *et al.*, 2011, Wang *et al.*, 1994)((Hahn, Strauss, *et al.*, 1989, Lemm *et al.*, 1994, Shirako & Strauss, 1994)). In infected mammalian cells at least two-types of nsP3-containing complex are probably formed. One it is likely associated with endosome-like vesicles and the plasma membrane. The other is possibly localized into the nuclear envelope which supports the hypothesis that nsP3 shuttles between the cytoplasm and the nucleus

(Cristea *et al.*, 2006). In SINV, both types have different composition and contain different ratios of the common components from the family Ras-GTPase activating protein (GAP) -(Scr Homology 3 [SH3] domain) binding protein (G3BP), such as G3BP1 and G3BP2, and the components Y-box-binding protein 1 (YBX1) and the heat shock protein 70 (Hsc70). In SINV infected mammalian cells, the membrane-associated complexes are important for viral RNA synthesis because high concentrations of dsRNAs were found associated with this type of complexes. Thus, it has been suggested that dsRNA synthesis takes place on the plasma membrane and later a partial fraction of dsRNA is transported through endocytic vesicles to the cytoplasm (Gorchakov, Garmashova, *et al.*, 2008). Other evidences, such as the acquired adaptative mutations occurred in SINV and VEEV nsP3s, suggested that nsP3 may interact with viral RNA in response to modifications of the 5' *cis*-acting elements in the virus genome (Fayzulin & Frolov, 2004, Michel *et al.*, 2007).

The N-terminal portion of nsP3 is conserved among alphaviruses and the full length nsP3 is organized into three distinct domains (Strauss & Strauss, 1994), Fig. 18, which are: (1) the Macro domain (also known as X-domain) at the N-terminal portion, (2) a central region named alphavirus unique domain (AUD), also known as zinc-binding domain

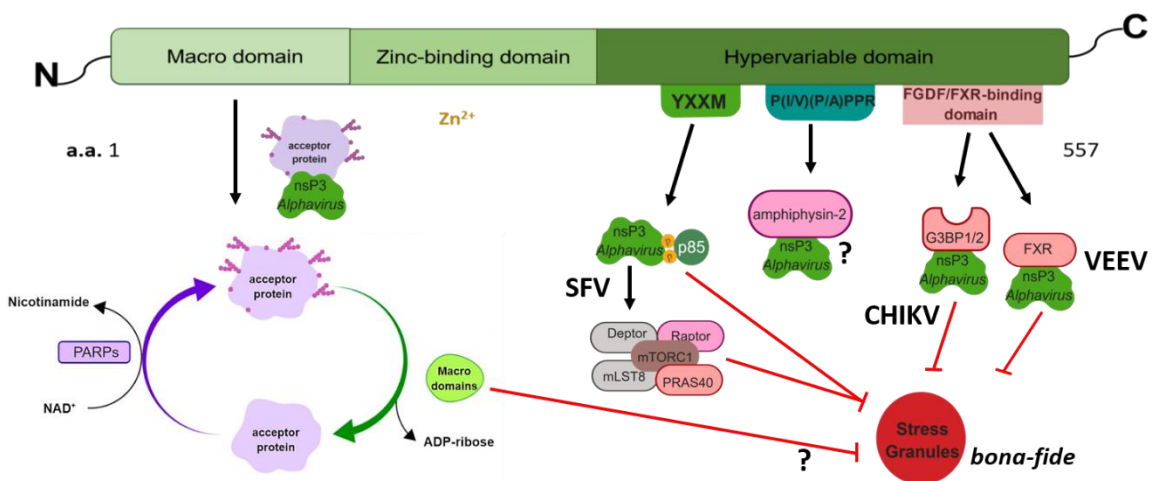


Figure 18. Domain organization of the non-structural protein 3 (nsP3). At the N-terminal is localized the Macro domain, after follows the Zinc-binding domain and the Hypervariable region. The last is localized at the C-terminal portion of nsP3 and is highly disordered and displays a multifunctional role during infection. The Macro domain is involved in removal of ADP-ribose from post-translationally modified proteins with mono- or poly- ADP-ribose. The Zinc-binding domain is known to coordinate a zinc ion, but its function is still unknown. Concerning the hypervariable region, a more complex role is in charge of this domain. In the diagram is represented 3 important sites of the hypervariable region. The YXXM motif which is required for activation of the PI3K through binding to the p85 subunit and that triggers the activation of the pathway that leads to activation of the mTORC1. The P(I/V)(P/A)PPR motif which is involved in binding to amphiphysin-2. Lastly, the FGDF motif (or FXR- binding motif) that is required to bind G3BP1/2 or FXR proteins respectively. The FGDF motif or FXR-binding motif is dependent on OW or NW viruses respectively. The interactions of nsP3 with p85, G3BP1/2 or FXR proteins are important for stress granules disassembly mediated by alphaviruses. The role of the Macro domain is still under investigation, but it is possible that is required for disassembly of stress granules. The amino acid range indicated with text is relative to the VEEV strain TC-83. Figure designed using <https://biorender.com/>.

(ZBD), and (3) a hypervariable region (HVR) at the C-terminal portion which contains several phosphorylation sites (Vihinen *et al.*, 2001).

Structural information is available for the folded N-terminal region of nsP3 which includes the protease domain of nsP2 fused with the N-terminal portion of nsP3 (Macro domain and AUD) (Shin *et al.*, 2012). The crystal structure of the nsP2/nsP3 fragment revealed the presence of a previously unknown zinc coordination site formed by four cysteines within the AUD (Shin *et al.*, 2012). It is likely that the AUD is important for minus-strand and sub-genomic RNA synthesis, polyprotein processing and neurovirulence (De *et al.*, 2003, Lastarza, Grakoui, *et al.*, 1994, Tuittila & Hinkkanen, 2003). The nsP3 Macro domain and the AUD forms a ring-like structure and an extensive charged interface with nsP2, encircling its MTase-like domain, (Shin *et al.*, 2012), reviewed in (Gotte *et al.*, 2018). Recently, it has been proposed that the AUD domain of nsP3 may be in or close to the opening of the replication complexes (the spherule). This model is supported by the fact that it has been suggested that this domain have RNA-binding activity. Thus, this function of AUD may support the idea that AUD domain controls the exit of the newly produced viral RNAs from the replication complex to the cytosol (Schulte *et al.*, 2016, Shin *et al.*, 2012). In addition, the residues located after the Macro domain participate in the positioning of the P23 cleavage site (Lulla *et al.*, 2012).

The Macro domain is conserved among alphaviruses, rubella virus, Hepatitis E Virus (HEV) and coronavirus (Koonin & Dolja, 1993). The crystallographic structures of Macro domains from VEEV and CHIKV are close to the homologous *Escherichia coli* domain (Malet *et al.*, 2009). Homologues of alphaviruses Macro domains are found across proteins from all kingdoms of life (Allen *et al.*, 2003, Park & Griffin, 2009, Pehrson & Fuji, 1998, Letunic *et al.*, 2006). However, it is not universally distributed in bacteria, archaea, and eukaryotes. It is frequently present on its own as a small open reading frame, but a more divergent version can be found attached to unusual histone variants, the *Macro* histones H2A (Pehrson & Fuji, 1998). The function of *Macro* domains may diverge among viral *Macro* domains and in SINV is important for virus replication in neurons and neurovirulence in mice (Park & Griffin, 2009). Alphavirus *Macro* domains, such as VEEV and CHIKV *Macro* domains, are known to bind nucleic acids and poly- adenosine diphosphate (ADP) ribose (poly-[ADP-ribose]) and were originally described to present adenosine diphosphoribose 1'-phosphate phosphatase activity (Malet *et al.*, 2009). In contrast, the severe acute respiratory syndrome coronavirus *Macro* domain presented low phosphatase activity but is still able to bind ADP-containing molecules (Egloff *et al.*, 2006). The adenosine diphosphoribose 1'-phosphate phosphatase activity may be related to induction of apoptosis in infected cells.

Alphavirus *Macro* domains are supposed to interact with one or more host proteins (Park & Griffin, 2009). It has been reported that *Macro* domains from alphaviruses are able to reverse ADP-ribosylation from post-translationally modified proteins (ADP-ribosylated proteins) through their hydrolytic activity, Fig. 18 (Li *et al.*, 2016, Abraham *et al.*, 2018, McPherson *et al.*, 2017, Ecker *et al.*, 2017). The binding of *Macro* domain to ADP-ribosylated proteins is likely necessary for alphavirus replication and its hydrolase activity promotes the amplification of replication complexes (Abraham *et al.*, 2018). Concerning alphaviruses, it has been reported that CHIKV *Macro* domain can act as a mono-ADP-ribosylhydrolase (McPherson *et al.*, 2017, Ecker *et al.*, 2017), whilst VEEV *Macro* domain can act either as mono-ADP-ribosylhydrolase or poly-ADP-ribosylhydrolase (Li *et al.*, 2016).

The hypervariable region has sequence features of natively unfolded proteins and is poorly conserved among alphaviruses (in length and sequence composition). However, conservation of some elements has been observed and some are preferentially selected indicating an advantage to their evolutionary retention (Aaskov *et al.*, 2011, Oberste *et al.*, 1996). Most likely the HVR is involved in pathogenesis/virulence and in viral transcription regulation (Tuittila *et al.*, 2000, Davis *et al.*, 1989, Galbraith *et al.*, 2006, Lastarza, Grakoui, *et al.*, 1994, Tuittila & Hinkkanen, 2003). Thus, the interactions between the alphavirus nsP3 HVR with virus-specific host factors may have a critical role in assembly of the viral replication complexes (vRC). Interestingly, recently it was reported a conserved proline-rich motif (Pro-[Ile/Val]-[Pro/Ala]-Pro-Pro-Arg) at the HVR of nsP3 which mimics the dynamin proline-rich motif and permit alphaviruses to recruit amphiphysin-2 to the viral replication complex through interaction with the Src Homology 3 (SH3) domain of amphiphysin-2, Fig. 18 (Tossavainen *et al.*, 2016). It is still unclear what is the role of amphiphysin-2 recruitment by alphaviruses (Tossavainen *et al.*, 2016), perhaps in supporting viral RNA replication (Neuvonen *et al.*, 2011). However, it is well known that amphiphysin-2 plays key roles in several cellular processes such as regulation of endocytosis and membrane recycling, cytoskeleton regulation, DNA repair, cell cycle progression and apoptosis, reviewed in (Prokic *et al.*, 2014). The participation of amphiphysins in endocytosis and membrane trafficking may suggest that they are important in the formation of the invagination of lysosomal membranes to form the viral replication complex, reviewed in (Lark *et al.*, 2017).

The HVR of nsP3s are rich in acidic residues as well as in serine, threonine and proline (Peranen *et al.*, 1988). The hyperphosphorylated form of nsP3 may play a role in RNA synthesis and the non-conserved C-terminus (the HVR) is heavily phosphorylated particularly on serines and threonines (Vihinen & Saarinen, 2000, Vihinen *et al.*, 2001).

It is likely that nsP3 had no autophosphorylation activity, thus requiring multiple cellular kinases. In SFV nsp3 it has been proposed that serine and threonine are phosphorylated in an approximate ratio of 2:1 respectively. Apparently, the sequence [Ser/Thr]-X-[Arg/Lys], where X is any a.a., (Kennelly & Krebs, 1991) seems to be conserved among several alphaviruses and it is probably involved in binding of protein kinase C (PKC) (Vihinen & Saarinen, 2000). Besides the PKC, other host phosphorylation enzyme has been suggested to interact with the HVR of nsP3. For example, casein kinase II (CK II) may be involved in the phosphorylation of other sites by recognizing the sequence [Ser/Thr]-X-X-[Asp/Glu], where X is any a.a. (Li *et al.*, 1990, Kennelly & Krebs, 1991).

Some works have reported that defective mutants in nsP3 phosphorylation exhibited a decrease rate of RNA synthesis and a strong reduction of pathogenicity in mice (Vihinen & Saarinen, 2000). Multiple phosphorylation of nsP3 with various phosphorylated states during infection, between different alphavirus species, mainly localized within the HVR, has been observed (Li *et al.*, 1990, Peranen *et al.*, 1988, Davis *et al.*, 1989, Lastarza, Grakoui, *et al.*, 1994, Vihinen *et al.*, 2001). The HVR plays an important role in the virus-host interaction and may be a determinant of pathogenesis through interactions with cell-type-specific factors, reviewed in (Rupp *et al.*, 2015). Among several host factors which are supposed to interact with nsP3, the cellular G3BP proteins were discovered to interact with the HVR of nsP3 (Rupp *et al.*, 2015, Cristea *et al.*, 2006, Cristea *et al.*, 2010, Frolova *et al.*, 2006, Scholte *et al.*, 2015, Panas *et al.*, 2014), as well as the fragile X syndrome (FXR) proteins family, Fig. 18 (Kim *et al.*, 2016, Foy *et al.*, 2013). The interaction of G3BP members has been characterized as specific of old world alphaviruses because until now the interaction has been observed in alphaviruses such as SINV, SFV and CHIKV (Cristea *et al.*, 2006, Frolova *et al.*, 2006, Scholte *et al.*, 2015, Panas *et al.*, 2012, Panas *et al.*, 2014). The HVR-G3BP interaction is determined by short repeating a.a. sequences located at the carboxy terminus of the HVR (Vognsen *et al.*, 2013) which are the Phe-Gly-Asp-Phe (FGDF) motifs (Schulte *et al.*, 2016, Panas *et al.*, 2014).

Through the interaction of the nsP3 HVR with G3BP, nsP3 interferes with the formation of the host cellular stress granules (SGs), which are involved in innate antiviral mechanisms, Fig. 18 (Panas *et al.*, 2012, Fros *et al.*, 2012). It is not clear what is the potential role for G3BPs in indirectly enhancing replication. It was suggested that from the association of nsP3 with G3BP1 and/or G3BP2 the translation of viral RNAs is reduced to allow the switch to genome amplification during early infection (Rupp *et al.*, 2015, Fros *et al.*, 2012, Scholte *et al.*, 2015).

Concerning alphaviruses belonging to the new world, some members such as EEEV can interact with all the members of FXR and G3BP protein families (Frolov *et al.*, 2017). However, FXR proteins were supposed to be specific to new world alphaviruses as it was previously reported for VEEV (Kim *et al.*, 2016, Foy *et al.*, 2013). The HVR-FXR interaction is likely important for VEEV infectivity, RNA replication initiation and rates of replication (Kim *et al.*, 2016). In a similar fashion to G3BP, the HVR interaction with FXR members exhibit a common potential to interact with RNAs and to self-assemble into higher order complexes like cellular stress granules (SGs) (Panas *et al.*, 2014). SGs are host intracellular organelles composed of a variety of cellular proteins in a dynamic phase transition into liquid droplets (a physical state associated with SGs) and G3BP appears to be critical for nucleation of SGs (White & Lloyd, 2012, Kedersha *et al.*, 2016, Calabretta & Richard, 2015, Kedersha *et al.*, 2013, Kato *et al.*, 2012).

The discovery of the potential interaction of nsP3 HVR from EEEV to interact with both G3BP and FXR families represents a significant evolutionary advantage which probably promotes efficient EEEV replication in a wide variety of cells and tissues, in which the host factor composition may be different (Frolov *et al.*, 2017). Interestingly, it is speculated that FXR-HVR interactions developed later in alphavirus evolution because G3BP-dependent CHIKV and SINV cannot utilize FXRs even if the binding sequences are inserted into their HVRS (Frolov *et al.*, 2017, Kim *et al.*, 2016, Foy *et al.*, 2013).

G3BPs have an important role in regulation of SGs formation (Kedersha *et al.*, 2016) (Tourriere *et al.*, 2003). G3BPs are organized into two globular domains, the Nuclear Transport Factor 2-like (NTF2) and the RNA Recognition Motif (RRM) domains, and structurally disordered regions with low amino acid complexity (Kedersha *et al.*, 2016, Irvine *et al.*, 2004). During SG formation the globular RNA-binding RRM domain and the disordered Arg-Gly-Gly (RGG) box RNA-binding motif are necessary for binding to RNA (Tourriere *et al.*, 2003).

The interaction of FXR or G3BP members with the nsP3 HVR occurs through their amino terminal domains, NTF2-like in G3BP and Agenet-like in FXR. In contrast, the downstream carboxy terminal domains are dispensable for HVR binding, however are essential for the protein function in viral replication complex formation and RNA replication, and it has been suggested that these domains are able to bind RNA through an unknown mechanism (Frolov *et al.*, 2017).

During alphavirus infection host-cell translation is inhibited and the SG are assembled because during viral replication the cytosolic sensors (also known as pattern recognition receptors [PRR]) such as retinoic acid-inducible gene I (RIG-I) and the melanoma

differentiation-associated gene 5 (MDA-5) can sense dsRNA replication intermediates. Thus, RIG-I and MDA-5 induce the activation of the antiviral protein kinase R (PKR) and the production of IFN type-I leading to the inhibition of translation and SG formation (Nikonov *et al.*, 2013, Akhrymuk *et al.*, 2016, McInerney *et al.*, 2005). In alphaviruses the formed cytoplasmic granules are clearly distinct from *bona fide* SGs as other components of SGs are lacking, such as: eukaryotic initiation factor 3 subunit (eIF3), eukaryotic initiation factor 4B subunit (eIF4B), small ribosomal subunit, poly(A)-binding protein (PAB or PABP), T-cell-restricted intracellular antigen 1 (TIA-1), TIA-1-related protein (TIAR), G3BP1 and G3BP2 (Scholte *et al.*, 2015). In addition, the cytoplasmic granules formed during alphavirus infection do not respond to chemical stimulation that either induce (arsenite) or disassemble (cycloheximide) *bona fide* SGs (Fros *et al.*, 2012). With the sequestration of G3BP into these viral nsP3-G3BP granules, the assemble of *bona fide* SG is effectively inhibited (Fros *et al.*, 2012, Schulte *et al.*, 2016). The mechanism that forms antiviral SGs is not fully understood but it is highly dependent on the RNA-binding proteins such as TIA1/R and G3BP1/2 and FXR members (Kim *et al.*, 2016, Anderson & Kedersha, 2008, Kanai *et al.*, 2004, Kedersha *et al.*, 1999, Tourriere *et al.*, 2003). Recently, it has been proposed that G3BP-mediated SG assembly is poly-(ADP-ribose) dependent (Isabelle *et al.*, 2012). Poly-(ADP-ribose) is probably involved in modulating the nuclear translocation of G3BP and it has been demonstrated that poly-(ADP-ribose) binds to the Gly-Arg-rich domain of G3BP in a non-covalent form (Isabelle *et al.*, 2012).

It has been proposed that host cell translation inhibition during alphavirus infection is mainly due to the activation of cellular stress response via phosphorylation of eIF2 α with the consequent transient formation of stress granules (SGs) containing the cellular TIA-1/R proteins. The viral replication occurs and the SGs are disassembled, synchronously with the switch from cellular to viral gene expression (McInerney *et al.*, 2005, Scholte *et al.*, 2015). Recent works have proposed that the recruitment of G3BP to nsP3 complexes may explain part of the mechanism by which SGs are disrupted in alphavirus infected cells (Fros *et al.*, 2012, Panas *et al.*, 2012). However, the replication complexes (with a spherule shape) are actively internalized into the body of the cell in a process that correlates with the activation of phosphoinositide-3-kinase (PI3K)-Akt-mammalian target of rapamycin (mTOR) pathway (Thaa *et al.*, 2015, Spuul *et al.*, 2010) which we may think that most likely occurs before the disassemble of SGs. Moreover, the 40S ribosome subunits were found in association with the Arg-Gly-Gly (RGG) motif of the C-terminal portion of G3BP, suggesting that during alphavirus infection the G3BP could direct the newly produced viral RNAs for translation after exit from the spherule (Kedersha *et al.*, 2016, Schulte *et al.*, 2016).

A variety of viruses have been reported to explore the PI3K-Akt-mTOR pathway, also known as the “pro-survival” pathway of rapamycin to prevent an apoptotic response. Experiments with nsP3 from SFV showed that nsP3 activates directly Akt only when associated with the plasma membrane through the hyperphosphorylated HVR which induces the internalization of the replication complex (Thaa *et al.*, 2015). In contrast, it has been reported that nsP3 from CHIKV only moderately activated Akt which was dependent on PI3K which did not stimulate the internalization of the replication complex for CHIKV (Thaa *et al.*, 2015, Spuul *et al.*, 2010). It has been suggested that the PI3K-AKT-mTOR pathway is not required for production of virions but is essential for internalization of the replication complex, reviewed in (Lark *et al.*, 2017).

4.3.4 The alphavirus non-structural protein 4 (nsP4)

The nsP4 is a protein of approximately 70 kDa and is involved in the alphavirus RNA synthesis. This non-structural protein is an RNA-dependent RNA polymerase (RdRp) because it contains the core RdRp domain and motifs at the C-terminal end. Thus, nsP4 participates in replicating the genomic RNA via a negative strand RNA intermediate and transcribing the 26 S sub-genomic RNA. The N-terminal of nsP4 is predicted to be disordered and the first 100 a.a. are conserved within alphaviruses only. This portion of the N-terminal seems to be required as a scaffold for interaction with polyprotein P123 and the mature nsPs to form replication complexes that are capable of synthesizing minus-strand from plus-strand templates (Rubach *et al.*, 2009, Rupp *et al.*, 2011, Tomar *et al.*, 2006). The interaction may occur with nsP1 or with (un)identified host proteins (Shirako *et al.*, 2000, Fata *et al.*, 2002). The C-terminal portion of nsP4 is predicted to present a typical RdRp structure with fingers, palm containing the Gly-Asp-Asp motif in the active site and thumb domains, Fig. 19 (O'Reilly & Kao, 1998, Rubach *et al.*, 2009, Tomar *et al.*, 2006).

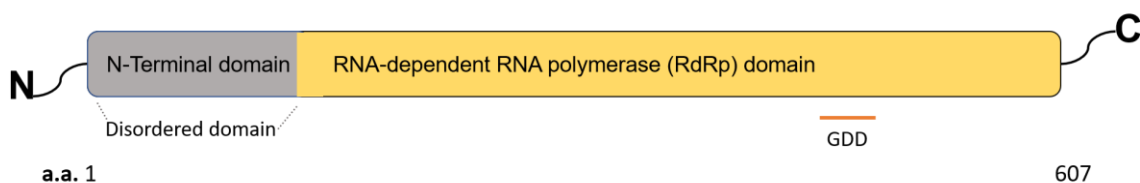


Figure 19. The domain organization of the non-structural protein 4 (nsP4). The short N-terminal domain of the nsP4 is highly disordered and it is a putative scaffold for interaction with polyprotein P123 and the mature nsPs to form the replication complexes. The remaining (large) portion of nsP4 is the RNA-dependent RNA polymerase (RdRp) domain which is required for synthesis of the *Alphavirus* RNA and terminal divalent cation-dependent adenylyltransferase (TATase) activity. The GDD represented under a yellow line represents the approximated location of the Gly-Asp-Asp catalytic motif of the RdRp domain of nsP4. The amino acid range corresponds to the nsP4 from VEEV strain TC83. For more detail see text. Adapted from (Rupp *et al.*, 2015).

The Gly-Asp-Asp (GDD) motif is required for the synthesis of the alphavirus RNA and terminal divalent cation-dependent adenylyltransferase (TATase) activity (Hahn, Grakoui, *et al.*, 1989) indicating a potential role in polyadenylation (Tomar *et al.*, 2006) at the 3' end of positive-sense RNAs (Tomar *et al.*, 2006). During infection the truncated alphavirus genomic RNA can be polyadenylated (Hill *et al.*, 1997, Raju *et al.*, 1999, Sawicki & Gorn, 1976). It has been suggested that the TATase activity is associated to the N-terminal of nsP4 and is independent of other viral factors (Tomar *et al.*, 2006). Currently there are two proposed mechanisms for polyadenylation in alphaviruses.

The first proposed mechanism, and currently less accepted, was related to the presence of a 5' poly(U) tract on the minus-strand RNA which was believed to be used as a template for the addition of a poly(A) tail at the 3' end of the RNA. More recently, a second mechanism was proposed after evidences indicating that the initiation of the minus-strand synthesis occurs immediately after the poly(A) tail (Rubach *et al.*, 2009, Tomar *et al.*, 2006, Raju *et al.*, 1999, Hill *et al.*, 1997). Thus, it was suggested that a template-independent mechanism is used for the addition of the poly(A) tail and a signal sequence is necessary to direct the polyadenylation. In SINV the polyadenylation signal is located within the 29 nucleotides located at the terminal of the 3'-UTR of the genome (Raju *et al.*, 1999). The template-independent polyadenylation mechanism is also supported by the TATase activity of nsP4 which has a role in addition, maintenance and repair of the poly(A) tail (Tomar *et al.*, 2006, Rubach *et al.*, 2009).

It has been reported that the full length recombinant nsP4 showed TATase activity and the *de novo* RNA synthetic activity was only observed after the addition of the other viral non-structural proteins supplied from mammalian cell membrane fractions (Rubach *et al.*, 2009). nsP4 is the unique non-structural protein presenting RdRp activity, but functional viral RNA synthesis requires the formation of the viral replicase complex, reviewed in (Rupp *et al.*, 2015).

The formation of the replication complex involves incorporation of nsP4. Thus, nsP4 gains stability as it is a protein that is targeted for degradation by the N-end rule pathway (de Groot *et al.*, 1991). The N-terminal rule pathway relates the half-life of a protein to the identity of its N-terminal residue (Gonda *et al.*, 1989, Bachmair *et al.*, 1986). The presence of a tyrosine at the N-terminus confers a destabilizing effect and a short half-life (Varshavsky, 1992). It has been reported that a N-terminal tyrosine interacts with nsP1 for minus-strand synthesis (Shirako & Strauss, 1998) and different mutations in the predicted disordered N-terminal domain resulted in either minus-strand or positive-sense RNA defects, suggesting roles in each activity (Rupp *et al.*, 2011, Rupp *et al.*, 2015). Mutations of regions identified as binding specifically to either the genomic or

subgenomic promoters were found to abrogate the corresponding RNA synthetic activity (Li & Stollar, 2004, 2007, Li *et al.*, 2010).

The suggested mechanism for alphavirus 3'-end genome repair requires the addition of AU-rich sequences next to the poly(A) tails in genomes lacking the 3' CSE. Thus poly(A) tail is the prerequisite for the addition of the AU-rich sequences (Hill *et al.*, 1997, Raju *et al.*, 1999). Moreover, addition of such motifs to the newly synthesized minus strands results from polymerase stammers on AU-rich sequences and sliding, jumping, and stammering of the polymerase. Although it is a viral polymerase, nsP4 can also use other RNAs than viral RNA as a template to synthesize dsRNA as it synthesizes 5'-ppp dsRNA using host cell RNA as a template (Nikonov *et al.*, 2013). The host is likely able to restrict viral replication based on this unspecific activity of nsP4 as it may induce IFN- β production mediated by RIG-I-like receptors.

4.4. Assembly of the replication complex

The viral early replication complex (eRC) is formed during cleavage of the non-structural polyproteins after infection of the host cell, Fig. 20. The eRCs are formed before the cleavage of the nsP2/nsP3 junction and include interactions with host factors which occur with both P123 + nsP4 and nsP1 + nsP23 + nsP4. To form the eRC, the complexes of non-structural proteins, namely both P123 + nsP4 and nsP1 + nsP23 + nsP4, are assembled at the membrane forming interactions with host factors that favour the remodelling of the host membrane. Thus, small structures of single-membrane bulb-shaped invaginations of approximately 50 nm are assembled on the external surface of the plasma membrane. These structures which are also named "spherules" are connected to the cytoplasm through a neck-like opening structure with approximately 5 to 10 nm in diameter which allows exchanges of metabolites and export of viral nascent RNA (Froshauer *et al.*, 1988). Thus, the "spherules" are naturally associated with dsRNA and partially processed non-structural proteins are present on the "spherules" necks (Frolova *et al.*, 2010).

The spherule structures are mainly known as "factories" of viral RNA synthesis and the interior of the spherules are lacking the components from the translational machinery such as ribosomes. Ribosomes and viral proteins such as capsids are found side by side to the spherule opening, suggesting that the translation occurs near the replication complexes (Froshauer *et al.*, 1988).

The eRCs are short-lived complexes which are likely to synthesize preferentially viral RNA of negative polarity, Fig. 20 (Barton *et al.*, 1991, Shirako & Strauss, 1994). The

eRCs formed by P123 + nsP4 may synthesize as few as a single RNA molecule of negative polarity before being completely cleaved (Lemm & Rice, 1993a, Lemm *et al.*, 1994, Lemm *et al.*, 1998, Sawicki & Sawicki, 1980, 1994, Shirako & Strauss, 1994). The switch between the synthesis of negative polarity RNAs to positive polarity RNAs is evident after cleavage of P123 into nsP1 and nsP23 which forms the intermediate nsP1 + nsP23 + nsP4 replicase complex (Shirako & Strauss, 1994, Lemm *et al.*, 1994). The shutoff of viral non-structural protein expression and the rapid processing of non-structural polyproteins may induce inhibition of the synthesis of viral RNA of negative polarity (Gorchakov, Frolova, *et al.*, 2008). The nsP1 + nsP23 + nsP4 replication complex is capable of synthesizing both the genomic and sub-genomic RNA species by using the nascent RNAs of negative polarity as a template (Lemm *et al.*, 1994, Shirako & Strauss, 1994).

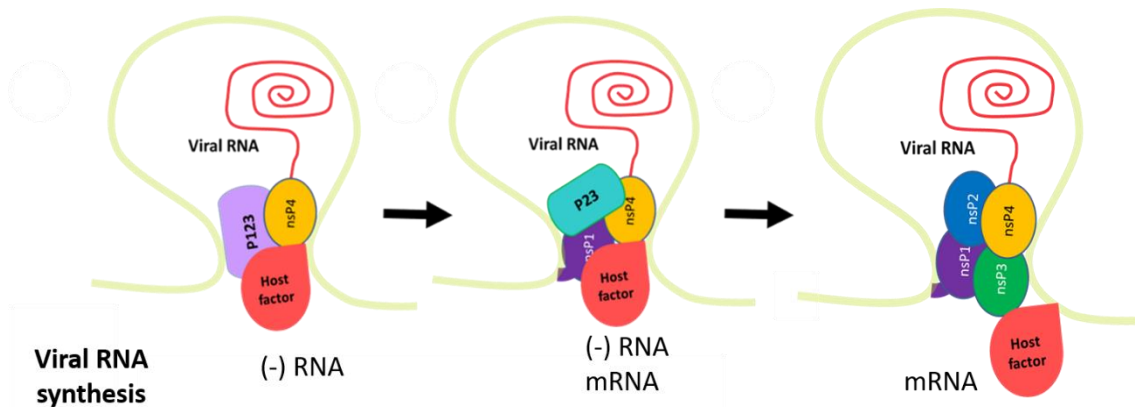


Figure 20. Assembly of the replication complex. The assembly of the replication complex occurs after the cleavage of nsP3/nsP4 junction. **(1)** The mature nsP4 together with P123 forms the early replication complex which is required for synthesizes of negative stranded viral RNA that remains double stranded. **(2)** The early replication complex goes through a process of maturation involving the proteolytic cleavage of the remaining polyproteins. The cleavage of the junction nsP1/nsP2 is responsible for the switch in the syntheses from negative to positive polarity viral RNA. This early intermediate replication complex is still able to synthesise low levels of RNA of negative polarity and it starts to use the RNA of negative polarity as a template for the synthesis of the positive polarity RNA which are the sub-genomic and genomic RNAs. **(3)** Once fully matured, after total cleavage and maturation of non-structural proteins (nsPs), is then formed the late replication complex that only uses the negative stranded RNA as a template to synthesize the viral sub-genomic and genomic RNAs. See text for details.

The switch of RNA template for synthesis of positive polarity RNAs, both genomic and sub-genomic RNAs, occurs after complete cleavage of the non-structural polyproteins and maturation of the non-structural proteins (Shirako & Strauss, 1994). The template for the positive polarity RNA synthesis is the nascent RNA of negative polarity which is initially produced. It has been proposed that the negative polarity RNA remains double stranded and results in different RNA species depending on which RNA is being synthesized, reviewed in (Rupp *et al.*, 2015). Therefore, the maturation of the non-structural proteins leads to the maturation of the replication complex into the late replication complex (IRC) (nsP1 + nsP2 + nsP3 + nsP4) which is required for efficient synthesis positive polarity viral RNA, both genomic and sub-genomic viral RNAs, Fig. 20

(Frolova *et al.*, 2006, Lemm & Rice, 1993b, Lemm *et al.*, 1994). Several host factors are probably important for the transition from eRCs to IRCs, one of such is RNase L which is known to recognize dsRNA and plays a crucial role for shutoff of the synthesis of RNA of negative polarity (Sawicki *et al.*, 2003).

Thus, the IRCs must regulate the synthesis of genomic and sub-genomic RNA from the same minus-strand template. This regulation probably occurs through nsP2 which may act as a transcription factor that associates with the sub-genomic promoter to recruit the RNA synthetic complex (Sawicki *et al.*, 1978, Suopanki *et al.*, 1998). In addition, nsP4 is also crucial for the regulation of the synthesis of both genomic and sub-genomic because nsP4 contains distinct site for binding of the two promoters (Li & Stollar, 2004, 2007, Li *et al.*, 2010). However, it has been suggested that the sub-genomic RNA is produced in excess relative to the viral genome (Lemm *et al.*, 1994, Shirako & Strauss, 1994).

Among the non-structural proteins, nsP1 and nsP3 are likely the main non-structural proteins with important roles for the structural architecture of the spherule. Regarding nsP1, it has been proposed that it is important to guide the assembly of the spherules at the PM due to their affinity for lipids specific for the cytoplasmic leaflet of the PM (Jose *et al.*, 2017). The observations from Jose *et al.* (Jose *et al.*, 2017) concerning the localization of nsP1 to the filopodial extensions emphasized the role of nsP1 in transport of replication complexes to the PM and host actin modifications. About nsP3, it has been reported that it plays a role in the induction of membrane curvature through the binding to cellular amphiphysin-1 and -2 (Neuvonen *et al.*, 2011). NsP3 is crucial for many other interventions on several host pathways. For example, the interaction of nsP3 with G3BP or FXR proteins promotes the bind of newly synthesized genomic RNA extruded from the eRCs present at the PM. Then, the complex of G3BP with newly synthesized genomic RNA in association with the viral replication complex will induce the formation and confer stability to the eRC (Kim *et al.*, 2016). Another role of G3BP during early infection was reported from CHIKV infection in which G3BPs were essential in the switch from translation to genome replication, possibly by removing the ribosomes from viral RNA (Scholte *et al.*, 2015).

In addition, it is well known that nsP3 participates in the activation of the PI3K-Akt-mTOR pathway, Fig. 21 (Mazzon *et al.*, 2018, Thaa *et al.*, 2015) which induces the internalisation of the spherules from the PM by endocytosis to give rise the cytopathic vacuoles 1 (CPV1) (Frolova *et al.*, 2010, Kujala *et al.*, 2001, Spuul *et al.*, 2010, Frolova *et al.*, 2006, Gorchakov, Garmashova, *et al.*, 2008). Spherules are probably internalised by endocytosis as cargoes through the endolysosomal pathway and the CPVs

(structures with approximately 600 to 200 nm in diameter) are usually positive for both endosomal and lysosomal markers (Grimley *et al.*, 1968, Froshauer *et al.*, 1988).

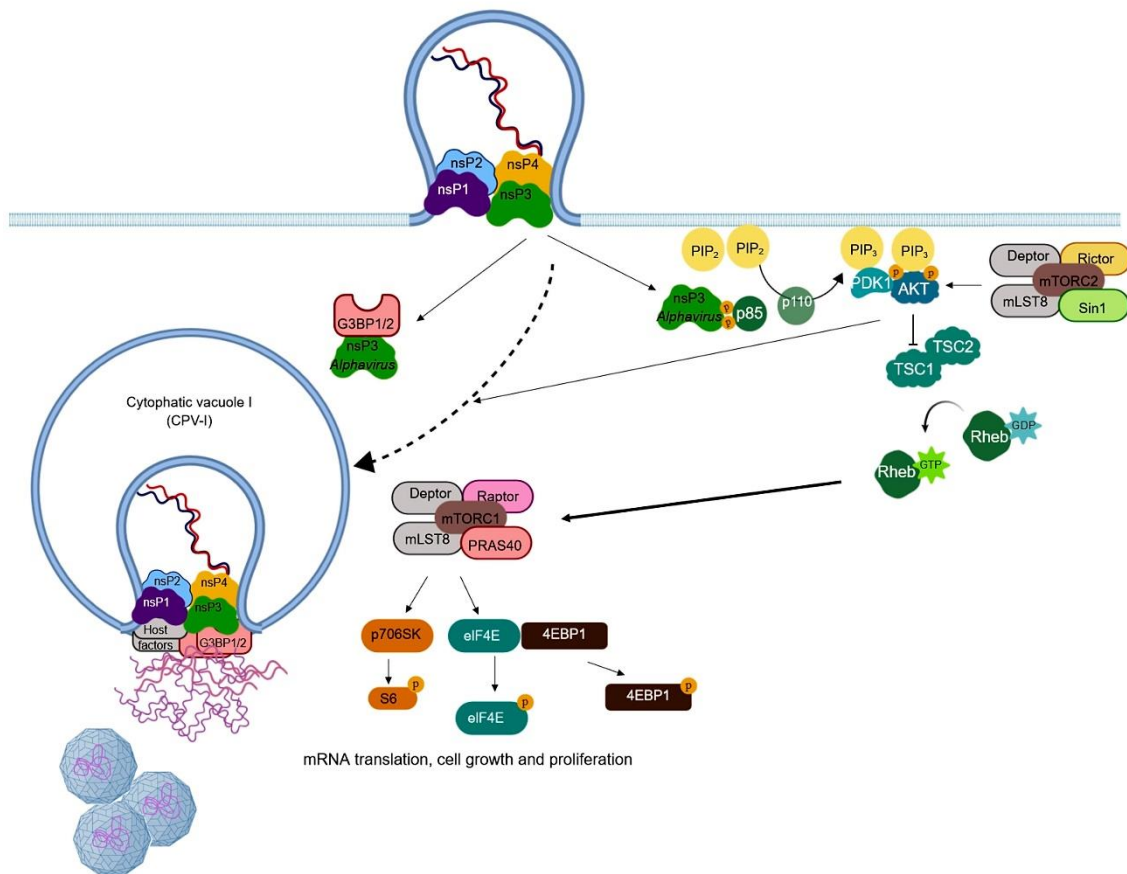


Figure 21. The multifunctional role of nsP3 during Alphavirus replication. The late replication complex formed by the mature non-structural proteins have a key role in the synthesis of genomic and sub-genomic viral RNA. NsP3 is a multifunctional protein contained within the late replication complex that is assembled at the PM. It has been proposed that nsP3 contains a sequence motif, YXXM, that binds to the SH2 domain of the regulatory subunit of PI3K (the p85 subunit). Through interaction of this sequence motif with the SH2 domain of p85, the p110 subunit of PI3K is activated and converts the membrane bound PI(4,5)P₂ to PI(3,4,5)P₃. The presence of PI(3,4,5)P₃ at the membrane recruits the PKD1 and AKT to the PM and PKD1 through phosphorylation activates AKT. The phosphorylation of AKT is also enhanced in the presence of mTORC2 complex. Interestingly, it has been reported that activation of AKT is correlated with translocation of the spherule containing the replication complex from the PM to cytopathic vacuoles I (CPV-I) that are found at the cytoplasm, in a process named as translocation of the spherule. In addition, the activated AKT triggers the inhibition of TSC1/2 complex which leads to activation of Rheb that converts GDP into GTP and leads to activation of the mTORC1 complex. The mTORC1 complex also induce phosphorylation of downstream targets to ensure efficient host cellular translation. The nsP3 also participates in binding and recruiting FXR proteins (not shown) or G3BP1/2 to the replication complex where occurs synthesis of viral RNA. It has been suggested that G3BP1/2 also participates in binding to the viral RNA near the spherule structures and the place where encapsidation of viral genomic RNA take place. For more details see text. Figure designed using <https://biorender.com/>.

However, the mechanism of formation and movement of higher-order membrane structures containing viral replication complexes remains unclear (Rupp *et al.*, 2015). The activation of the PI3K-Akt-mTOR pathway is mainly directed through the Tyr-X(2)-Met, where X is any a.a., motif (YXXM motif) present in the HVR of nsP3 of several alphaviruses including SFV, RRV, GETV, SAGV and MIDV. It was proposed that the

HVR of nsP3 works by mimicking activated growth factor receptors at the plasma membrane and results in the activation of the PI3K-Akt-mTOR pathway, Fig. 21 (Gotte *et al.*, 2018).

The inactive PI3K is a heterodimer composed of a catalytic subunit p110 and a regulatory subunit p85 that stabilizes and inhibits the p110 subunit. The regulatory subunit p85 contains two Src Homology 2 (SH2) domains, which interact with phosphotyrosine in Tyr-X(2)-Met motifs, where X is any a.a., on activated growth factor receptors (Songyang *et al.*, 1993). Thus, both Tyr and Met residues in the YXXM motif of the nsP3 HVR are the key players for the interaction with the SH2 domains of p85, Fig. 21 (Mazzon *et al.*, 2018, Thaa *et al.*, 2015). Consequently, the active subunit p110 dissociates from the heterodimer p85/p110 and is recruited to plasma membrane, where it converts the plasma lipid phosphatidylinositol-4,5-biphosphate [PI(4,5)P₂] to phosphatidylinositol-3,4,5-trisphosphate [PI(3,4,5)P₃]. Then, the phosphorylation and activation of Akt by the kinase phosphatide-dependent kinase 1 (PDK1) occurs because both are recruited into close proximity at the plasma membrane through their pleckstrin-homology (PH) domain that are known to bind PI(3,4,5)P₃, Fig. 21 (Gotte *et al.*, 2018). The activation of Akt by phosphorylation occurs on Thr-308 and Ser-473. After activation, Akt (Thr-308p and Ser-473p) phosphorylates a plethora of targets such as; mediators of proliferation, antiapoptotic mediators, and modulators of cytoskeleton dynamics. Indirectly, a downstream target of the active Akt is mTOR, a well-known central metabolic regulator in cells which sense nutrient availability, such as aminoacids, ATP and growth factors, through the large protein complex named as mammalian target of rapamycin complex 1 (mTORC1). Active mTOR also ensues efficient cellular translation by inducing phosphorylation of downstream targets such as eukaryotic initiation factor 4E-binding protein 1 (4EBP1) and the ribosomal protein S6, Fig. 21 (Mazzon *et al.*, 2018, Thaa *et al.*, 2015).

The cytophatic vacuoles (CPVs) induced by the activation of Akt, mediated by the interaction between nsP3 and p85, are known as the sites for RNA synthesis, Fig. 21 (Grimley *et al.*, 1968, Salonen *et al.*, 2005). Within these modified membrane structures, the RNA synthesis take place on the cytoplasmic side of CPVs, but partially sequestered within the spherules (Froshauer *et al.*, 1988, Kujala *et al.*, 2001). The formation of this spherules has been found to be dependent on non-structural proteins and as well on active RNA synthesis (Frolova *et al.*, 2010, Spuul *et al.*, 2011).

The alphavirus RNA synthesis must benefit from being synthesised in the spherule's structures because within spherules the viral dsRNA intermediates are protected from the host cell detection and disruption. The assembly of the replication complexes is

dependent on the stage of the polyprotein. Once formed, the membrane structures of the spherules and CPVs have been suggested to act as scaffolds that stabilize the complex once the polyprotein is cleaved. Indeed, these membrane structures effectively increase the concentration of replication factors at the site of synthesis (Salonen *et al.*, 2003). Despite several host factors being required for membrane remodeling and spherule formation, other host factors are probably required for viral RNA synthesis such as the Poly(rC)-binding protein 1 (PCBP1) and a set of other heterogeneous nuclear ribonucleoprotein (hnRNP) such as the hnRNP K, A1, M and C, as reviewed by (Rupp *et al.*, 2015).

After the synthesis of the genomic RNA of positive polarity, the nascent RNA may be extruded from the spherules followed by co-assembly with Cp. The possible involvement of G3BP or FXRs proteins in assist this process through their RNA binding motif remains to be addressed. Thus, the NC is likely to assemble near the cytopathic vacuole I (CPVI), Fig. 21 (Jose *et al.*, 2017).

5. RNA capping pathways

With the discovery of the canonical N⁷-methyl-guanosine (m⁷G) cap, during almost 40 years the 5'-end mRNA capping was believed to be a characteristic of viral and eukaryotic mRNAs (Wei *et al.*, 1975a, Mauer *et al.*, 2017, Daffis *et al.*, 2010, Kiledjian, 2018, Furuichi, Morgan, Muthukrishnan, *et al.*, 1975, Wei *et al.*, 1975b, Furuichi, 2015). The m⁷G cap was the first cap identified in eukaryotes and it is referred in the literature as the canonical cap (Furuichi, Morgan, Muthukrishnan, *et al.*, 1975, Wei *et al.*, 1975b, Wei *et al.*, 1975a).

In higher eukaryotes such as humans, the canonical m⁷G cap promotes stability of the mRNA and facilitate mRNA translation, as well as contributes to RNA splicing and RNA transport (Ramanathan, Robb, *et al.*, 2016). The stability of mRNAs is, in part, achieved through protection against exoribonucleases by the addition of the m⁷G cap to the 5'-end of mRNAs (Furuichi *et al.*, 1977, Wang & Kiledjian, 2001). The m⁷G cap is also used as an alternative signal to direct the translational machinery to the 5'-end of protein-encoding RNAs. The recognition of the cap by the translational machinery, formed by eukaryotic initiation factor 4E subunit (eIF4E) and other proteins, recruits the 40 S ribosome subunit (Gingras *et al.*, 1999). The additional methylations at the N⁶- and 2'-O-hydroxyl positions (m⁶A_m), when the first nucleotide following the m⁷G cap is an adenosine, confers a key layer of epitranscriptomic regulatory information to the 5'-ends of the m⁶A_m eukaryotic RNAs, for example, it may confer a key layer of protection to decapping (Mauer *et al.*, 2017). These modifications facilitate mRNA translation and

contribute to distinguishing host from foreign viral mRNAs (Daffis *et al.*, 2010, Devarkar *et al.*, 2016).

Furthermore, capping does not proceed to completion in eukaryotic cells. Incomplete caps lacking the N⁷-methyl moiety or lacking a cap altogether can also be generated. These incomplete caps are detected and cleared by a 5'-end quality control (5'QC) mechanism employing the decapping exoribonuclease (DXO) family of proteins (Chang *et al.*, 2012, Jiao *et al.*, 2013, Jiao *et al.*, 2010). In parallel to the capping quality control pathway, a mechanism of cytoplasmic recapping of 5'-end monophosphorylated RNA may occur in cellular mRNAs (Schoenberg & Maquat, 2009, Trotman & Schoenberg, 2018).

In humans (and eukaryotes in general) the canonical cap formation (m⁷G cap) occurs co-transcriptionally. The addition of the m⁷G cap to the 5'-end of a nascent RNA transcript generated by the RNA polymerase II (RNAP II) after transcriptional initiation within a narrow window of between +20 to +40 nascent oligonucleotides (Salditt-Georgieff *et al.*, 1976, Shatkin, 1976). The capping process is known to require three enzymatic activities which are the RTPase, GTase and MTase activities (Ghosh & Lima, 2010, Shuman, 1995). Eukaryotes can present two (*e.g.* mammals) or three (*e.g.* yeast) enzymes as the main players in the capping pathway to catalyse these three enzymatic activities. For example, in mammals the RTPase and GTPase activities are carried out by a single bifunctional capping enzyme (Yue *et al.*, 1997).

In humans, the first step to occur during the m⁷G capping of mRNAs is the hydrolysis of the 5'- triphosphate end of the nascent pre-mRNA by removal of the γ -phosphate. This step is carried out by an enzyme with RTPase activity to yield a mRNA with a diphosphate 5'-end. The reaction is catalysed by a two-step phosphoryl-transfer reaction. A conserved cysteine from the signature motif His-Cys-X(5)-Arg-(Ser/Thr) (the phosphate binding loop, or "P-loop") attacks the γ -phosphorus of triphosphate terminal RNA to form a covalent protein-cysteinyl-S-phosphate intermediate and then a diphosphate RNA product is released (Changela *et al.*, 2001, Shuman, 2002). Second, the covalent phosphor-enzyme intermediate is hydrolysed to release inorganic phosphate, Fig. 22 (Takagi *et al.*, 1997, Changela *et al.*, 2001, Shuman, 2002).



Figure 22. Metal-independent triphosphatase activity that occurs during formation of the cap-0 structure in human cells. An enzyme with RNA triphosphatase activity (RTPase) produce a mRNA with a diphosphate 5'-end. Through a conserved cysteine present in the conserved phosphate binding loop attacks the γ -phosphorus of triphosphate terminal RNA to form a covalent protein-cysteinylyl-S-phosphate intermediate. The diphosphate RNA is later released and in a final step the inorganic phosphate is released after hydrolysis of the covalent bond between the phosphate and the enzyme. For more details see text. Adapted from (Ghosh & Lima, 2010).

Then, the next step of the capping is the addition of a guanosine monophosphate (GMP) to the 5'-end diphosphate mRNA. The GMP is transferred from a GTP molecule through a reversible “ping-pong” mechanism which consists in two steps. The GTase, present in a bifunctional protein together with RTPase, reacts with GTP to make a covalent enzyme-GMP intermediate complex. The capping is coupled to the early stage of transcription because the elongating RNAP II binds directly their phosphorylated C-terminal (CTD) domain to the GTase enzyme. This induces an allosteric effect that mediates a two-fold increase in affinity of the GTase for GTP (Ho & Shuman, 1999). In an “open conformation” the enzyme was observed to bind GTP and the interaction of GTP with the oligonucleotide/oligosaccharide-binding fold (OB-fold) induces the “closed conformation” which is important for the formation of the enzyme-(lysyl-N)-GMP adduct, Fig. 23 (Fabrega *et al.*, 2003, Hakansson *et al.*, 1997, Ghosh & Lima, 2010).

The enzyme-(lysyl-N)-GMP adduct intermediate is formed by a phosphoamide linkage to the ϵ -amino group of the nucleophilic lysine residue in the signature Lys-X-Asp-Gly sequence with the α -phosphate from GTP. The sequence Lys-X-Asp-Gly is localized in a nucleotide-binding pocket which is part of the N- nucleotidyl-transferase (NTase) domain (Shuman & Schwer, 1995, Ghosh & Lima, 2010). Then, the enzyme shifts to an “open conformation” to release pyrophosphate. It has been suggested that the covalent nucleoside adduct likely undergoes a conformational change from *syn* to *anti* to properly position the nucleotide for attack by the incoming diphosphate 5'-end RNA substrate (Shuman & Lima, 2004, Ghosh & Lima, 2010). The GMP is then transferred from the GTase to the β -phosphate of the 5'-diphosphate end of the mRNA and thus forming a

5'-5' linked guanosine monophosphate cap ($Gp_{\alpha}p_{\beta}p_{\alpha}NpNp$ -RNA), Fig. 23 (Furuichi, Morgan, Muthukrishnan, *et al.*, 1975, Wei *et al.*, 1975b).

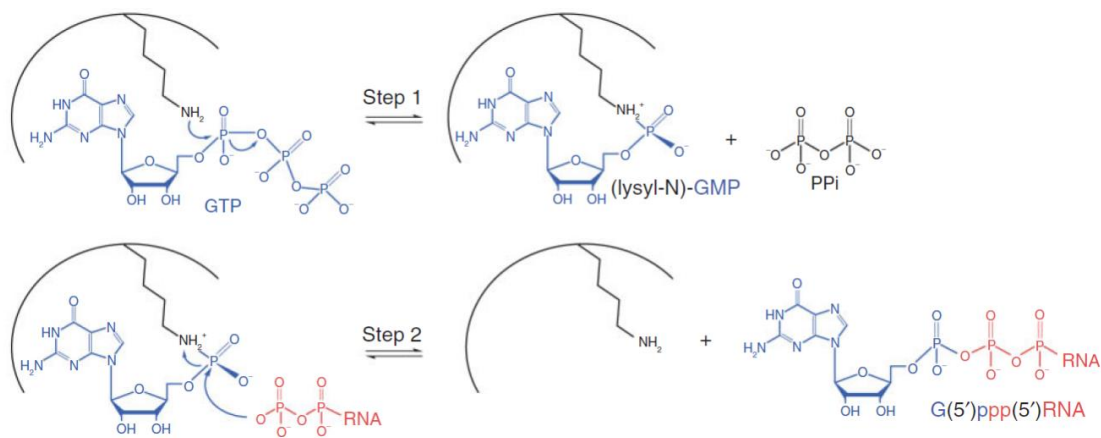


Figure 23. Capping in human cells require guanylyltransferase activities to synthesise the cap-0 structure. An enzyme presenting RNA guanylyltransferase catalyzes the capping in a two-step reaction. First the enzyme binds GTP (colored blue) to catalyse the transfer of GMP to the active site lysine to form a covalent enzyme(lysyl-N)-GMP intermediate. Secondly, catalyse the transfer of the GMP to a diphosphate RNA end to form GpppRNA. For more details see text. Adapted from (Ghosh & Lima, 2010).

Subsequently, the methylated cap structure (m^7G) is produced co-transcriptionally through the activity of an N^7 -MTase that binds to GTase-RNAP II complexes. The reaction, catalysed by the N^7 -MTase, uses SAM as a methyl donor to produce the cap structure $m^7Gp_{\alpha}p_{\beta}p_{\alpha}NpNp$ -RNA (type 0 cap structure) and releases S-adenosyl-L-homocysteine (SAH) as a by-product of the reaction, Fig. 24 (Furuichi, 1974, 2015, Desrosiers *et al.*, 1974, Nuss *et al.*, 1975).

The formation of the cap-0 does not terminate the capping reaction in eukaryotes: The m^7G capped mRNA may be further modified by several methylations, Fig. 25. The first ribose-2'-O-hydroxyl methylation in eukaryotic mRNAs is m^7G cap depend and occurs in the nucleus ($m^7Gp_{\alpha}p_{\beta}p_{\alpha}N_m$ -RNA, type 1 cap structure), while the second ribose-2'-O-hydroxyl methylation occurs in the cytoplasm ($m^7Gp_{\alpha}p_{\beta}p_{\alpha}N_mN_m$ -RNA, type 2 cap structure) and is catalysed by specific methyltransferases, Fig. 25 (Adams & Cory, 1975, Wei *et al.*, 1975a, Ghosh & Lima, 2010). Usually higher organisms have more extensively methylated caps (Furuichi, 2015). For example, a class of small nuclear RNAs can be hypermethylated in the cytoplasm on the m^7G to generate the trimethylated $m^{2,2,7}G$ cap (trimethylguanosine, TMG), which is important for the reimport of these RNAs into the nucleus, Fig. 25 (Hamm & Mattaj, 1990, Lamond, 1990).

Viruses have developed specific strategies to hide their genomic information and replication from the host cell. They can inactivate or utilize the host capping system to benefit the progress of their replication cycle (Furuichi, 2015, Decroly *et al.*, 2011).

For example, retroviruses make use of the capping enzymes from the host cell (Goff, 2007). Some viruses have really evolved a strategy to translate uncapped viral mRNAs through cap-independent mechanisms (Topisirovic *et al.*, 2011). Few viruses (*e.g.*, picornaviruses) use *cis*-acting internal ribosome entry site (IRES) elements, present in the 5'-UTR of the viral RNA (Cullen, 2009), for direct recruitment of ribosomal subunits (Cullen, 2009, Hellen & Sarnow, 2001, Topisirovic *et al.*, 2011). Other viruses (*e.g.*,

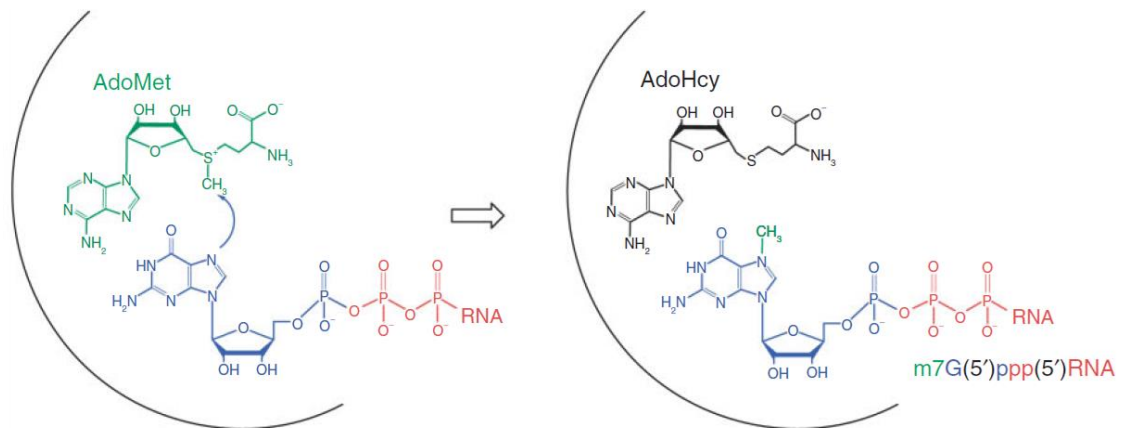


Figure 24. Methyltransferase activities during formation of the cap-0 structures during capping in human cells. The methylated cap structure (cap-0 structure) is produced co-transcriptionally and requires the enzymatic activity of the RNA guanine-N7 methyltransferase. The RNA guanine-N7 methyltransferase binds S-adenosylmethionine (SAM) (colored green) and GpppRNA (colored as in figure 23) and catalyzes the transfer of the methyl group (colored green) to the guanine N7 position. For additional information see text. Adapted from (Ghosh & Lima, 2010).

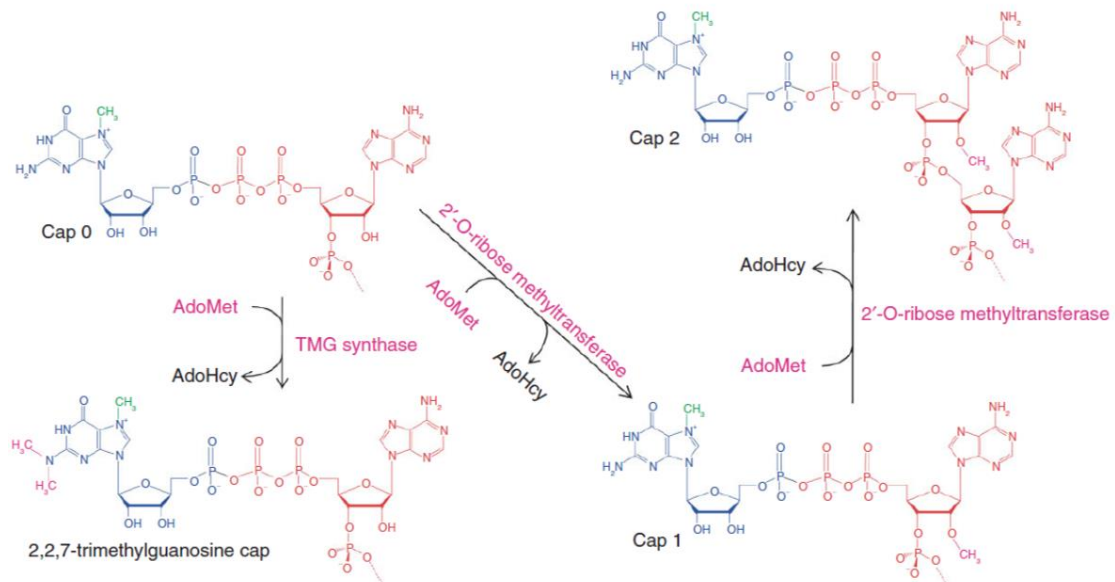


Figure 25. Structures and synthesis of RNA caps. The cap-0 structure is then further methylated to originate the TMG cap, Cap-1 and cap-2. The TMG cap is formed through the transfer of two methyl groups (colored magenta) from S-adenosylmethionine (SAM). Concerning caps-1 and cap-2 structures, it is required methylation (colored magenta) of cap 0 at the ribose 2'-O-hydroxyl at the first and second nucleosides, respectively. For more detail see text. Adapted from (Ghosh & Lima, 2010).

caliciviruses and potiviruses) have acquired the ability to translate their own RNAs through covalently coupling of a translational regulator protein (e.g., VPg (Rieder *et al.*, 2000)) to the first nucleotide of the 5'-end of their RNA, and interacting at the same time with transcription factors (e.g., eIF4E) to ensure efficient translation of viral mRNAs (Topisirovic *et al.*, 2011, Oruetxebarria *et al.*, 2001, Goodfellow *et al.*, 2005). Although the cap-independent mechanism of translation has been explored by a subset of viruses, a plethora of unconventional pathways for the canonical cap formation has been explored by viruses, as well. For a significant fraction of viruses, the cap structure is crucial for their life cycle and they had to adapt their own cap-dependent mechanisms in order to survive inside the host cell.

5.1. Canonical cap (m⁷G) formation through unconventional pathways in viruses:

Some viruses such as the ones from the genus *Flavivirus* and those from the order *Nidovirales* may form a group of conventional RNA-capping viruses (Decroly *et al.*, 2011). However, the capping of viral mRNAs with a canonical cap is known to deviate from the conventional RNA-capping pathway for several viruses. For example, viruses such as negative polarity ssRNA viruses (e.g., rhabdoviruses) and positive polarity ssRNA viruses (e.g., alphaviruses) can synthesize a viral cap similar to a cellular RNA cap by using unconventional capping mechanisms.

5.1.1 Unconventional capping pathway through “Cap snatching”

A fraction of RNA viruses lacks their own machinery to synthesise a cap structure and in this manner these viruses developed strategies to steal the cap from the host RNA and this process is named “cap snatching”. This capping mechanism is present among negative polarity ssRNA viruses from the families *Arenaviridae*, *Bunyaviridae* and *Orthomyxoviridae* which have a segmented RNA genome (Decroly *et al.*, 2011, Bouloy *et al.*, 1978, Caton & Robertson, 1980). The “cap snatching” mechanism usually requires three steps. A specific site in the viral RdRp (or possibly the N protein (Qi *et al.*, 2010)) mediates the first step in which the specific sites binds to the 5' cap-1 or cap-2 structure of a host mRNA. Then, the second step is the endo-nucleolytic cleavage of the cellular mRNA that take place several nucleotides downstream from the cap structure. Lastly, the third step is the use of the capped host-RNA fragment by the RdRp as a primer to synthesise the viral mRNA (Decroly *et al.*, 2011). The “decapped” cellular mRNAs are then subject to degradation leading to downregulation of cellular mRNAs (Decroly *et al.*, 2011).

A practical example of “cap snatching” is the mechanism used by Influenza virus (*Orthomyxoviridae* family) to promote the virus replication, Fig. 26. In influenza the RdRp is a complex formed between three proteins subunits which are the polymerase base protein 1 (PB1), polymerase base protein 2 (PB2) and polymerase acidic protein (PA) (Sugiyama *et al.*, 2009, Reich *et al.*, 2014). The assembly of the RdRp complex in the nucleus (Huet *et al.*, 2010) drives the binding of PB2 subunit to the cap structure of the host capped RNA. Then, the first 10 to 15 nucleotides of the capped RNA are excised through the endonuclease activity of the subunit PA. The excised capped RNA-fragment is then used as a primer to initiate the viral mRNA transcription, Fig. 26 (Dias *et al.*, 2009, Plotch *et al.*, 1979, Plotch *et al.*, 1981, Ruigrok *et al.*, 2010). Influenza seems to prefer, as snatching targets, the host capped non-coding RNAs (such as small nuclear RNAs [snRNAs] U1 and U2) rather than host capped mRNA or pre-mRNA (Gu *et al.*, 2015).

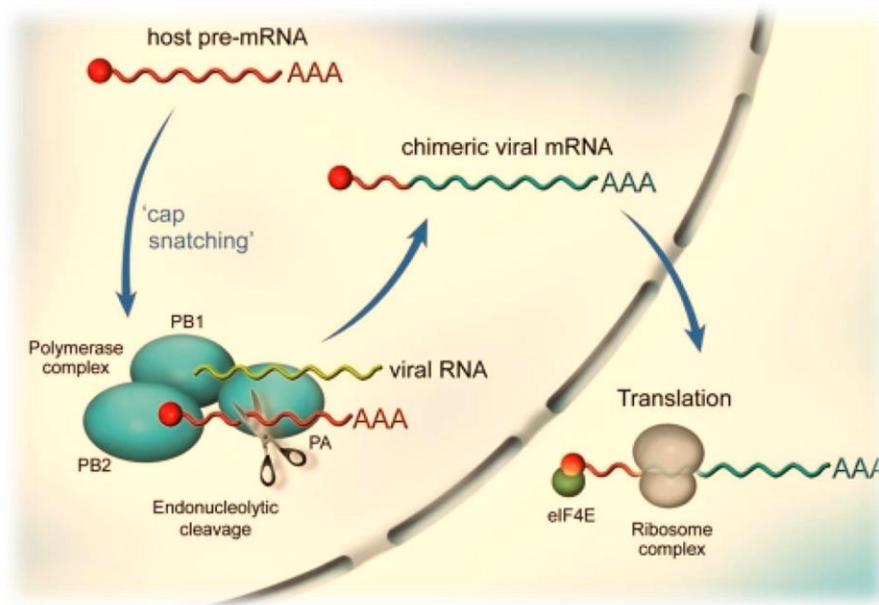


Figure 26. Cap-snatching transcription mechanism of influenza polymerase. The PA-PB1-PB2 complex is localized in the nucleus of the infected cell. During transcription, the PB2 subunit binds the 5',7-methylguanosine cap of a host pre-mRNA molecule (red), which is subsequently cleaved 10–15 nucleotides downstream by the PA endonuclease. The resulting short capped RNA primer is used to initiate polymerization by the RNA-dependent RNA polymerase of the PB1 subunit using 5'- and 3'-bound vRNA (green) as template, resulting in capped, polyadenylated, chimeric mRNA molecules (red and blue) that are exported to the cytoplasm for translation into viral proteins. Adapted from (Boivin *et al.*, 2010).

5.1.2 Unconventional capping pathway with guanosine diphosphate (GDP)

An unconventional capping pathway which is a unique characteristic of rhabdoviruses such as Vesicular stomatitis virus (VSV). The L protein from rhabdoviruses is known to present several activities such as RdRp, RNA GDP-polyribonucleotidyl-transferase (PRNTase), MTase, and NTPase activities, Fig. 27. The RdRp activity is contained within the RdRp domain localized at the N-terminal of L protein and includes the conserved regions I, II and III, Fig. 27. The L protein also harbours the activities of PRNTase, within the cap domain and includes the conserved region V of L protein. While the MTase activity is present in the MTase domain immediately upstream the CTD domain of L protein and including the conserved region VI, Fig. 27 (Bujnicki & Rychlewski, 2002, Ferron *et al.*, 2002, Liang *et al.*, 2015, Li *et al.*, 2005).

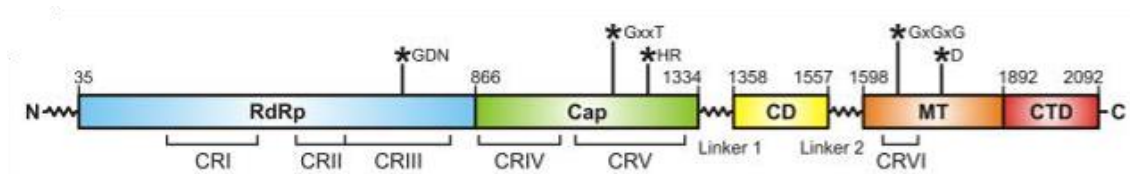


Figure 27. Domain organization of VSV-L. The polymerase domain (RdRp) is in cyan; capping domain (Cap), green; connector domain (CD), yellow; methyl transferase (MT), orange; C-terminal domain (CTD), red. Amino acid residue numbers indicate functional domain boundaries. Flexible linkers 1 and 2 connect Cap to CD and CD to MT domain, respectively. Conserved regions within L proteins of non-segmented negative-strand (NNS) RNA viruses are labelled CR I – VI. Asterisks indicate the position of active site residues. For details see text. Adapted from (Liang *et al.*, 2015).

The 5'- triphosphate end of a nascent RNA is used as a substrate by the L protein PRNTase, Fig. 28. There is formation of a covalent bond between the 5' α -phosphate of the viral RNA and a conserved histidine from the "His-Arg" (HR) motif, which is present in the conserved region V of the cap domain. The L protein is likely involved in the generation of GDP from GTP by an NTPase activity, which likely requires the conserved region VI and the CTD domain. The GDP is then transferred onto a PRNTase-bound pRNA intermediate during cap synthesis in VSV (Liang *et al.*, 2015, Ogino & Banerjee, 2007, Paesen *et al.*, 2015). Next, the enzyme transfers the 5'-monophosphate RNA to GDP (Ogino & Banerjee, 2007) and thus forming the $Gp_{\alpha}p_{\beta}p_{\alpha}Np$ -RNA structure through a 5'-5' triphosphate bridge, Fig. 28. Later, this structure is further converted into a canonical type 1 cap structure through ribose-2'-O-hydroxyl methylation and N⁷ guanine methylation (Decroly *et al.*, 2011). Thereby, the VSV MTase first methylate the structure $Gp_{\alpha}p_{\beta}p_{\alpha}Np$ -RNA at the ribose-2'-O-hydroxyl of the first nucleotide generating the $Gp_{\alpha}p_{\beta}p_{\alpha}N_m$ -RNA, followed by methylation at the guanine-N⁷ position ($m^7Gp_{\alpha}p_{\beta}p_{\alpha}N_m$ -RNA) (Rahmeh *et al.*, 2009, Decroly *et al.*, 2011).

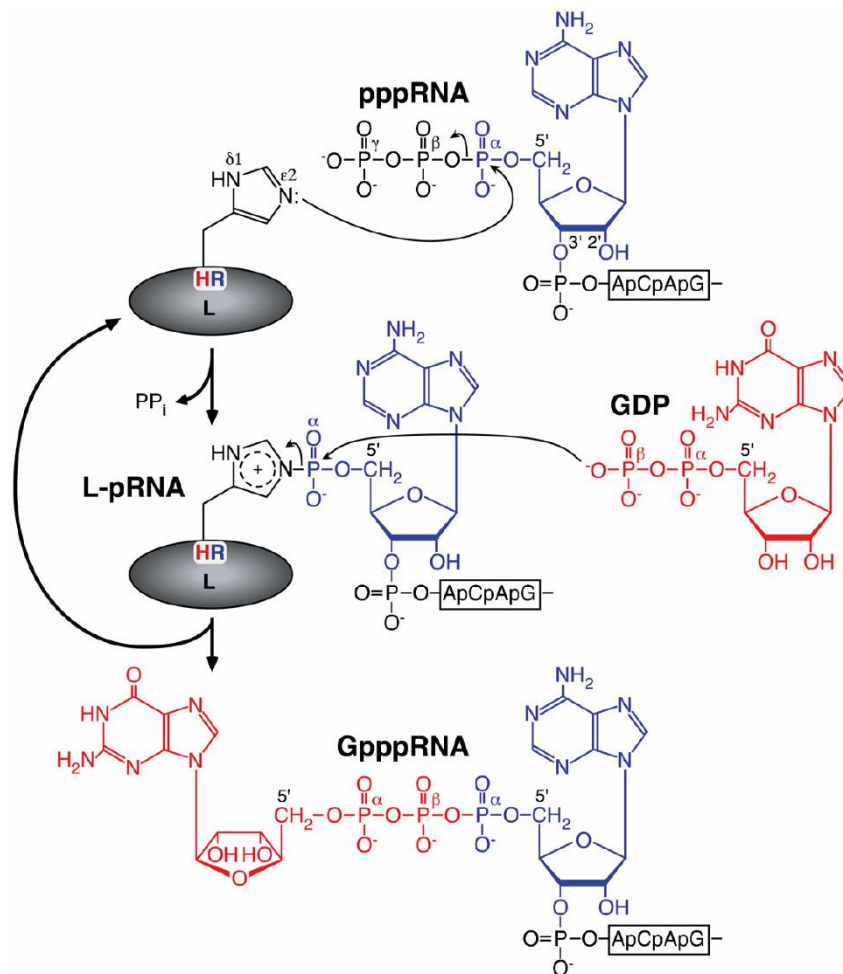


Figure 28. A proposed model of the polyribonucleotidyl transfer reaction catalysed by the unconventional mRNA capping enzyme L protein. For detail, see text. (Ogino *et al.*, 2010)

5.1.3 Unconventional capping pathway with m⁷GMP

This unconventional capping pathway is specific to alphaviruses and alphavirus-like viruses such as Brome mosaic virus, Bamboo mosaic virus, Tobacco mosaic virus and Hepatitis E virus (Li *et al.*, 2001, Magden *et al.*, 2001, Merits *et al.*, 1999, Decroly *et al.*, 2011). The activity of the viral RNA NTPase is necessary to prepare the viral mRNA to accept the m⁷GMP and in alphaviruses this activity is embedded in the nsP2 (Vasiljeva *et al.*, 2000). The GTase and MTase activities are present in a multifunctional enzyme such as alphaviruses-nsP1, Bromo mosaic virus protein-1A, Tobacco mosaic virus-p126, Bamboo mosaic virus nsp and Hepatitis E virus p110 (Li *et al.*, 2001, Magden *et al.*, 2001, Merits *et al.*, 1999, Decroly *et al.*, 2011). In alphaviruses the MTase activity of nsP1 catalyses the transfer of a methyl group from SAM to the N⁷ position of a GTP molecule and produces SAH as a by-product (Ahola & Kaariainen, 1995). The nsP1 of alphaviruses also harbours the GTase activity which in the presence of SAH mediates the formation of a covalent intermediate complex enzyme-GMP (Ahola & Kaariainen,

1995). As alphaviruses do not proceed further than synthesizing a cap-0 structure, the capping is then finalised after nsP1 transfer the m⁷GMP molecule to the diphosphate 5'-end of the alphavirus RNA which leads to the formation of the 5'-5' triphosphate bridge between m⁷G and the nascent transcript (Ahola & Kaariainen, 1995). More insights into alphavirus mRNA capping are **presented immediately in the chapter 5.3.**

5.2. The discovery of non-canonical cap structures

Recently, since 2009, a non-canonical cap was discovered in RNAs from bacterial and eukaryotic cells (Chen *et al.*, 2009, Kowtoniuk *et al.*, 2009, Winz *et al.*, 2017). The discovery of capped RNAs in prokaryotes paved the way to believe that capping is not a unique feature belonging to eukaryotes. In fact, capping is not only conserved among eukaryotes (from yeast to humans) but also is present in other domains of life such as prokaryotes and viruses (Chen *et al.*, 2009, Kowtoniuk *et al.*, 2009, Furuichi, Muthukrishnan, *et al.*, 1975, Furuichi, Morgan, Muthukrishnan, *et al.*, 1975, Furuichi, Morgan, Shatkin, *et al.*, 1975, Wei *et al.*, 1975b). Although the capping is present among several life kingdoms, the structure of the cap, the organization of the capping pathway, the structure and function of enzymes differ between species (Kiledjian, 2018).

A plethora of non-canonical caps at the 5'-end of mRNAs have been identified such as the oxidized form of nicotinamide adenine dinucleotide (NAD⁺) and adenosine residue of other nucleoside metabolites. For example, desphospho-coenzyme A (dpCoA) was already found in bacterial RNAs as a cap structure at the 5'-end of the RNAs (Kowtoniuk *et al.*, 2009). *In vitro* experiments have been used to suggest that the adenosine residue from other nucleoside metabolites can be used as non-canonical initiation nucleotides, this includes dpCoA, the reduced form of nicotinamide adenine dinucleotide (NADH), flavin adenine diphosphate (FAD) and uridine containing metabolites such as; uridine diphosphate glucose (UDP-glucose), uridine diphosphate N-acetyl-muramyl-pentapeptide (MurNAc-pentapeptide) and uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) (Bird *et al.*, 2016, Julius & Yuzenkova, 2017, Malygin & Shemyakin, 1979, Julius & Yuzenkova, 2018).

The non-canonical cap formation pathway is not yet well understood and unravelled. The possible pathways that lead to the formation of the non-canonical caps account with at least two possible models, one being a transcriptional process and the other being post-transcriptional (Kiledjian, 2018). In bacteria the NAD⁺ capping mechanism likely follows a transcriptional process because NAD⁺ can be used by the capping machinery as the initiation nucleotide in place of ATP. In eukaryotic cells both transcriptional and

post-transcriptional processes are also possible (Walters *et al.*, 2017, Jiao *et al.*, 2017, Bird *et al.*, 2016, Dieci *et al.*, 2009, Filipowicz & Pogacic, 2002, Kawaji *et al.*, 2008).

In human cells mRNAs can be present as two distinct populations in cells that differ in the state of their 5'-end cap. It has been reported that one population carries the canonical m⁷G cap, while the other, estimated to comprise approximately 1 to 6 % of the respective mRNA harbours an NAD⁺ cap (Jiao *et al.*, 2017). The function of the canonical cap m⁷G seems to be different from the non-canonical NAD⁺ cap. For example, in human cells the m⁷G cap promotes stability and translation of mRNA, in contrast the NAD⁺ cap apparently is targeting mRNAs for rapid decay and does not support translation. It is curious that in bacteria the NAD⁺ cap functions in a role analogous to the eukaryotic m⁷G cap to stabilize the mRNA (Bird *et al.*, 2016). The mRNA decay promoted by the NAD⁺ cap is mediated in part by the non-canonical decapping enzyme from the decapping and exoribonuclease protein (DXO) family and by a subclass of the nucleoside diphosphate linked to another moiety-X (Nudix) hydrolases family of proteins which start degrading the RNA through the removal of the NAD⁺ cap (deNADing) (Jiao *et al.*, 2017).

Concerning viruses, guanosine seems to be the building block unit of viral caps until now, as only cap structures build up with guanosine and methylations were reported to date. It remains the open question about capping with non-canonical caps among the virus world.

5.3. Unconventional RNA capping pathway in alphaviruses

Alphaviruses replicate only in the cytoplasm and they need to synthesise their own capping enzymes because they cannot use the host capping machinery which is located in the nucleus (Ramanathan, Robb, *et al.*, 2016, Follett *et al.*, 1975, Ahola & Kaariainen, 1995). In alphaviruses, both genomic and sub-genomic RNAs are capped at their 5'-end with a cap-0 structure (m⁷Gp_αp_βp_αNp-RNA) which confers protection to degradation by cellular 5'-exonucleases and enhance and directs translation through the 5'-UTR (Dubin *et al.*, 1977, Wengler *et al.*, 1979, Hefti *et al.*, 1975). The m⁷Gp_αp_βp_αN-RNA cap alone is insufficient to distract the host immune detection due to absence of ribose-2'-O-hydroxyl methylation. Thus, the virus has evolved strategies to prevent the viral cap recognition by the host immune system (Hyde *et al.*, 2015). One of such strategies is the presence of a stem loop in the CSE1 which is near the 5'-end of the viral RNA of positive polarity.

The cap-0 structure was initially addressed in OW alphaviruses such as SINV and SFV (Hefti *et al.*, 1975, Dubin *et al.*, 1977, Wengler *et al.*, 1979). Recently, the alphavirus

capping has been addressed in the most pathogenic OW alphavirus (CHIKV) and in a NW alphavirus (VEEV) (Li *et al.*, 2015, Bullard-Feibelman *et al.*, 2016).

The capping of alphavirus mRNA starts with the cleavage of the phosphoanhydride bond between the β and the γ phospho-groups. This reaction, named 5' RNA triphosphatase activity (RTPase), is executed by nsP2 onto the viral nascent RNA substrate and releases the γ -phosphate and generates a 5'-diphosphate RNA molecule, Fig. 29 (Vasiljeva *et al.*, 2000).

The alphavirus nsP1 presents both MTase and GTase activities and therefore nsP1 displays a key role during the alphavirus RNA capping. The membrane association of the nsP1 is highly required for its activity (Ahola *et al.*, 1999). The nsP1 association to the membrane requires the amphipathic helix located in the middle of the nsP1 sequence, the interaction of hydrophobic amino acids with acyl groups of the membrane, palmitoylated nsP1 cysteines, and the interaction of a set of positively charged amino acids with the polar heads of phospholipids (Ahola *et al.*, 1999, Spuul *et al.*, 2007, Lampio *et al.*, 2000, Ahola *et al.*, 2000, Laakkonen *et al.*, 1996).

Nsp1 is the main player of alphavirus capping enzymes, and its MTase activity is required for the methylation of the N⁷ position of a GTP molecule, Fig. 29. The nsP1 MTase uses SAM as a methyl donor to transfer a methyl group from SAM to the N⁷ position of the GTP substrate, leading to formation of m⁷GTP. The MTase of alphaviruses was early identified in nsP1 by using SINV nsP1 mutant replicating in insect cells depleted of methionine and later confirmed in SINV and SFV by using *in vitro* enzymatic assays (Mi & Stollar, 1991, Laakkonen *et al.*, 1994, Mi *et al.*, 1989, Ahola & Kaariainen, 1995).

Then, the nsP1 proceeds with the alphavirus capping through the GTase, Fig. 29. The first step of the GTase is the guanylation (GT) of the nsP1 protein that leads to the formation of the nsP1-m⁷GMP adduct and early studies were performed on SINV and SFV nsP1 (Ahola & Kaariainen, 1995). The alphavirus nsP1 does not contain either the canonical motifs Lys-Asp-Lys-Glu and Lys-X-Asp-Gly, for MTase and GTase activities respectively (Ghosh & Lima, 2010), and thus requires an alternative mechanism to display these activities. The evaluation of critical residues for MTase and GTase were initially addressed by using SFV (Ahola *et al.*, 1997). From this work it was identified a highly conserved histidine which is crucial for the formation of the nsP1-(histidyl-N)-m⁷GMP intermediate complex instead of the canonical enzyme-(lysyl-N)-GMP, Fig. 29 (Ahola *et al.*, 1997). Indeed, mutation of the His38Ala in SFV prevented the GT activity without disrupting the MTase activity (Ahola *et al.*, 1997). The highly conserved histidine of nsP1 of alphaviruses is also present in the GTase of Bamboo Mosaic Virus and it is

also essential for the GTase activity (Hu *et al.*, 2011, Huang *et al.*, 2004). This suggests that the capping pathway of alphaviruses is extended to alphavirus-like viruses. Regarding the formation of the nsP1-m⁷GMP intermediate complex, the conserved nucleophilic histidine attacks the α -phosphate and induces the cleavage of the phosphoanhydride bond between the β and the α phospho-groups, and pyrophosphate is released, Fig. 29. The nsP1-(histidyl-N)-m⁷GMP adduct is then formed between the N ϵ (or possibly N δ) of the histidine and the α -phosphate of the m⁷GMP (Ahola & Kaariainen, 1995, Ahola *et al.*, 1997). The final step of the GTase is required for the formation of the mature cap-0 structure at the 5' end of the nascent viral RNA. The step consists in the transfer of the m⁷GMP from the nsP1 to the nascent 5' diphosphorylated viral RNA which leads to the formation of the 5'-5' linked m⁷GMP to the diphosphate viral RNA (m⁷Gp α p β p α N-RNA), Fig. 29. Recently, the full GTase reaction of alphaviruses was experimentally reported for the first time using VEEV nsP1 (Li *et al.*, 2015).

Although capping of alphaviruses leads to the formation of a cap-0 structure, it remains to be understood the reason why SINV and SFV seems to present additional methyl groups attached to the N² of the cap structure (Ferron *et al.*, 2012, HsuChen & Dubin, 1976, van Duijn *et al.*, 1986). This additional cap-0 methylations are similar to the ones reported for the 2,2,7-trimethylguanosine (TMG) cap that was discovered on non-coding eukaryotic RNAs such as small nuclear (sn), small nucleolar (sno) RNAs, and telomerase RNA (Ferron *et al.*, 2012, Busch *et al.*, 1982, Seto *et al.*, 1999).

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

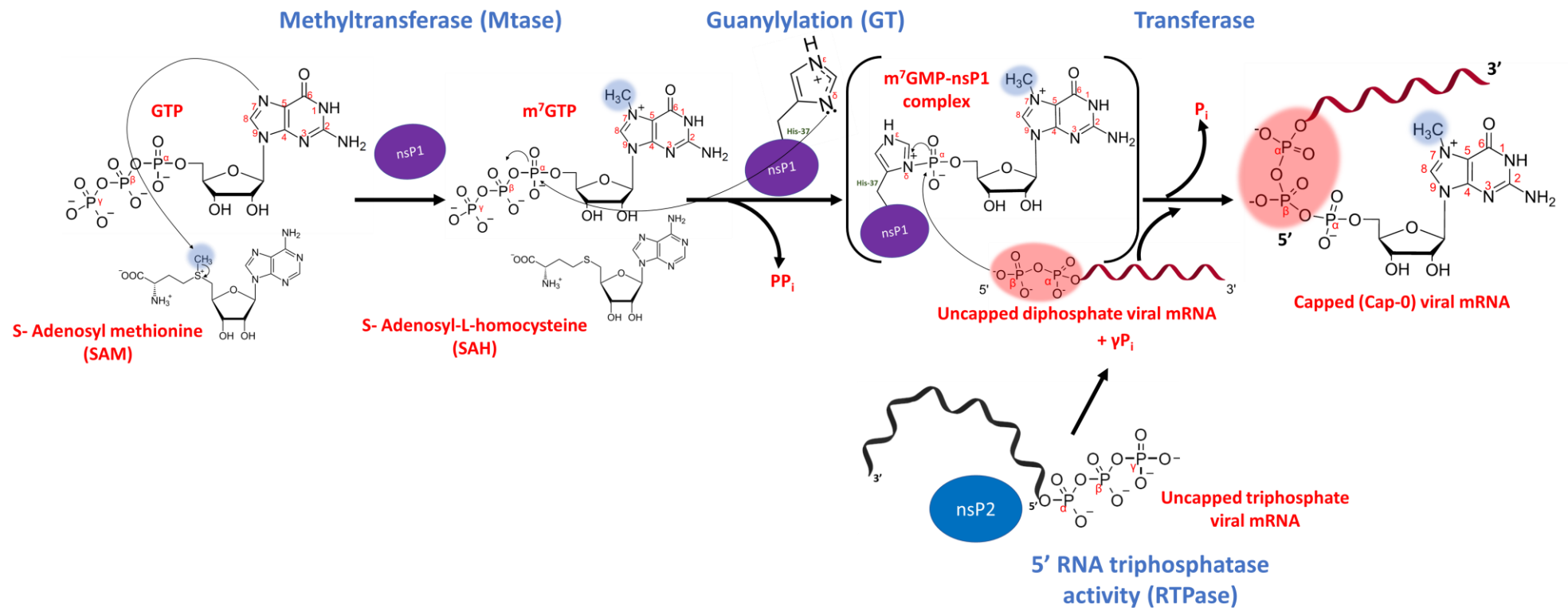


Figure 29. Current understanding of *Alphavirus* mRNA capping. The mRNA of alphaviruses is capped at the 5'-end with a type 0 cap by an unconventional mechanism. The *Alphavirus* capping enzyme is the non-structural protein 1 (nsP1) that presents the three enzymatic activities required for capping for the viral mRNA (genomic and sub-genomic). For the first step of the reaction is required the methyltransferase activity of the nsP1 enzyme. The first step consists in the methylation of a GTP molecule through the transfer of a methyl group from the donor S-adenosyl methionine (SAM) to the N⁷ position a molecule of GTP and generates S-adenosyl-L-homocysteine as a by-product of the reaction. A second step of the cap-0 formation is the attachment of a m⁷Gp (m⁷GMP) moiety to a diphosphate viral mRNA that is previously generated by nsP2 that presents 5' RNA triphosphatase (RTPase) activity. This step requires the guanylyltransferase activity (GTase activity) of nsP1 that involves a highly conserved histidine of nsP1 (histidine 37 in VEEV strain TC83). The GTase activity is a reaction that requires two steps, the first is the guanylylation (GT) of the nsP1 enzyme and the second is the transfer of the m⁷GMP moiety to the diphosphate end of the viral mRNA (transferase reaction). The conserved histidine attacks the α -phosphorus of the m⁷GTP and breaks the pho-pho-anydride phosphoanhydride bond between the β and the α phospho-groups. During this step pyrophosphate is released and the nsP1-(histidyl-N)-m⁷GMP adduct is formed between the N ϵ (or possibly N δ) of the histidine and the α -phosphate of the m⁷GMP. For more details see text.

THIS PAGE IS INTENTIONALLY LEFT BLANK

5.4. The connection of viral RNA capping and the innate immune response

Humans and other jawed vertebrates (Gnathostomata) developed a variety of host-defence mechanisms to fight infectious diseases. These mechanisms are divided in two types of defence which are the innate immunity and the adaptative (also known as acquired) immunity. Interestingly, innate immunity is also present in invertebrates such as insects; they solely rely on innate immunity for their survival against infections. Innate immunity is thus a universal and ancient form of host defence against viral infection (Medzhitov, 2007, Janeway & Medzhitov, 2002).

Although, while the adaptative immunity develops slowly upon infection and becomes more powerful upon re-infections, the innate immunity is initiated immediately and requires the same time either for infection or re-infection. Unlike adaptative immunity, innate immunity is not specific such and is thus used as the first line of defence. Several mechanism belonging to the innate immune defence are constitutively active while others require the detection of pathogen by the host (Medzhitov, 2007, Janeway & Medzhitov, 2002). Pathogens such as viruses own a limited set of molecular structures, such as those imbedded in viral nucleic acids (Yoneyama & Fujita, 2010), which are known as pathogen-associated molecular patterns (PAMPs). The latter are invariant among microorganisms of the same class, are unique to microorganism enabling their distinction as “non-self” from “self” molecular structures, and are essential for the microorganism survival which restricts their adaptative evolution (Medzhitov, 2007, Janeway & Medzhitov, 2002).

The host innate immunity is composed of a limited set of a variety of receptors called pattern recognition receptors (PRRs) which are used by the host to recognize the pathogens through their PAMPs (Medzhitov, 2007, Janeway & Medzhitov, 2002). Nonetheless, PRRs are unable to distinguish pathogenic and symbiotic microorganism because the ligands classified as PAMPs are not unique to pathogens (Rifkin *et al.*, 2005). In some cases, PRRs are also activated by “self” ligands which will cause autoimmunity (Rakoff-Nahoum *et al.*, 2004). After sensing PAMPs, the activated PRRs activates intercellular signalling cascades leading to the production of type I IFN, interleukin-1 (IL-1) and other pro-inflammatory cytokines to induce a host antiviral state to contain the viral propagation (Hansen *et al.*, 2011, Meylan *et al.*, 2006), Fig. 30.

The PRR are homogenously distributed among cells from the same type. A set of this germline-encoded receptors, PRRs, are specialized in sensing “non-self” nucleic acids. Toll-like receptors (TLRs) are a set of PRR specialized in sensing “non-self” DNA or RNA

within intracellular compartments, Fig. 30. Several viruses such as alphaviruses usually replicate in the cytoplasm and several PRRs are specialized in sensing foreign nucleic acids forthwith in the cytoplasm (Wilkins & Gale, 2010). Such PRRs are namely the nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and the retinoic acid-inducible gene (RIG)-like receptors (RLRs) (Brennan & Bowie, 2010, Schlee, 2013), Fig. 30. In order to distinguish the “self” RNA that is present in the cytosol, uncommon RNA structures, which are characteristic of viral pathogens, are sensed by the set of PRRs localized in the cytosol. The detected uncommon RNA structures are usually dsRNA, RNA presenting triphosphate 5'-ends, RNA with an incomplete methylated cap structure, RNA bearing uncommon cap structures or RNA bearing a protein covalently attached to the 5'-end (e.g. VPg-RNA) (Koyama *et al.*, 2008, Takeuchi & Akira, 2007).

TLR family are transmembrane proteins mainly found in endosomal compartments and at the cell surface of cells from the immune system (e.g. macrophages and dendritic cells). The family of TLRs contain receptors specialized in sensing viral nucleic acids or several bacterial products including lipopolysaccharide and lipoteichoic acids (Akira *et al.*, 2006, Medzhitov, 2007). Different types of RNAs can be sensed by TLR3, TLR7 and TLR8. While TLR3 is specialized in sensing dsRNAs, the TLR7 and TLR8 are involved in detecting 5'- triphosphate ends of ssRNA or ssRNA bearing a cap-0 structure (Meylan *et al.*, 2006, Züst *et al.*, 2011), Fig. 30. The activation of the TLRs is followed upon detection of the PAMPs by the leucine-rich repeat (LRR) motif from the ectodomain of the TLR. Therefore, the conformational changes induced in to the cytoplasmic domain, which includes the Toll-IL-1 receptor (TIR) domain, triggers a signalling transduction cascade athwart interactions between the TIR domain with cytoplasmic adaptor proteins such as the myeloid differentiation primary response 88 protein (MyD88) and the TIR-domain-containing adapter-inducing interferon- β (TRIF) (Meylan *et al.*, 2006). Therefore, leading to the production of type I interferon (e.g. INF- α), IL-1 and other pro-inflammatory cytokines, Fig. 30.

RLRs are intracellular cytosolic receptors that function as sensors of intracellular pathogens (Wilkins & Gale, 2010). The RIG-I (alternatively known as DDX58) and the MDA-5 (alternatively known as IFIH1) are intracellular sensors of viral nucleic acids belonging to the family of RLRs (Kato *et al.*, 2006). Both RIG-I and MDA-5 are cytosolic sensors for “non-self” dsRNAs, for example MDA-5 senses long dsRNA (Peisley *et al.*, 2013, Binder *et al.*, 2011, Züst *et al.*, 2011, Bruns *et al.*, 2014, Peisley *et al.*, 2012) while RIG-I senses short dsRNAs (Kolakofsky *et al.*, 2012, Wang *et al.*, 2010, Jiang *et al.*, 2011, Ramanathan, Devarkar, *et al.*, 2016, Goubau *et al.*, 2014). Although RIG-I and MDA-5 distinguish the viral dsRNA from the “self” RNA by their 5'-end (Li *et al.*, 2009,

Gitlin *et al.*, 2006, Yoneyama *et al.*, 2004), the RIG-I is specialized in sensing viral ssRNA with a 5'- triphosphate end (Hornung *et al.*, 2006, Pichlmair *et al.*, 2006), while MDA-5 senses the presence of RNA bearing a cap-0 structure or linked to a protein such as VPg (Zust *et al.*, 2011, Luthra *et al.*, 2011). Interestingly, cap-0 and dsRNA with triphosphate 5'-ends are known to bind RIG-I with similar affinity (Devarkar *et al.*, 2016). However, RIG-I apparently requires base-paired structures in conjunction with the triphosphate 5'-end to trigger an antiviral response (Schmidt *et al.*, 2009). Nevertheless, the binding of dsRNA methylated at the ribose-2'-O-hydroxyl of the first or second nucleotide does not stimulate the RIG-I pathway due to a steric effect within the binding site of the C-terminal regulatory domain CTD (Li *et al.*, 2009, Devarkar *et al.*, 2016, Shiraki *et al.*, 2003).

RIG-I and MDA-5 contain at their N-terminal two caspase-recruitment domains (CARDs), a central Asp-Glu-X-His (DEXH) box-containing RNA helicase-ATPase domain and a regulatory CTD domain (Oshiumi *et al.*, 2010). The binding of nucleic acids to the RNA-binding site of the CTD of RIG-I induces conformational changes in the receptor. Therefore, results in activation of CARDs through its dissociation from inhibitory protein domains and resulting in interactions of CARDs with a mitochondrial-anchored protein, the signalling adaptor INF β promoter stimulator 1 (IPS1; also known as CARDIF, MAVS or VISA) (Meylan *et al.*, 2005, Seth *et al.*, 2005, Kawai *et al.*, 2005, Xu *et al.*, 2005, Kolakofsky *et al.*, 2012), Fig. 30. Transcription factors such as interferon regulatory factors (IRFs) (e.g. IRF3 or IRF7) and nuclear factor-kB (NF-kB) are activated through the recruitment of the signalling complex by IPS1, Fig. 30. Thus, leading to the induction of type I IFN (e.g. IFN β) and production of pro-inflammatory cytokines, such as IL-1 β , in order to maintain the antiviral response.

IFN α/β binds to a common receptor (INFR) and, through an autocrine and paracrine signalling, triggers an antiviral state in surrounding cells. The activation of INFR stimulates the janus kinase-signal transducer and activator of transcription (JAK-STAT) signalling cascade to induce the expression of IFN-stimulated genes (ISGs), among which are antiviral restriction factors that specifically target uncapped viral RNAs, Fig. 30 (Hyde & Diamond, 2015). For example, the interferon induced proteins with tetratricopeptide repeats (IFIT proteins) that specifically sequesters RNA with triphosphate 5'-ends (such as IFIT1) or RNA bearing a cap-0 structure (such as IFIT2) (Hyde & Diamond, 2015, Pichlmair *et al.*, 2011, Daffis *et al.*, 2010), Fig. 30. IFIT1 can also be induced by a mechanism independent of IFN by using an IRF3-dependent pathway, for example (Grandvaux *et al.*, 2002). Other antiviral restriction factor belonging to the IFIT complex is the IFIT5 and strongly interacts with ssRNA bearing a 5'-triphosphate end (Kumar *et al.*, 2014) and it also displays high affinity for ssRNAs with

5'- monophosphate ends and with ssRNAs bearing a cap-0 structure at the 5'-end but not with ssRNAs bearing a cap-1 structure (Abbas *et al.*, 2013). IFIT1 seems to primarily interfere with the interaction of eIF4E with the cap structure, Fig. 30 (Habjan *et al.*, 2013).

The ribose-2'-O-hydroxyl methylation of the viral RNA cap can contribute to the evasion of the host innate antiviral response through avoidance of IFIT-mediated suppression. This is likely the case of the WNV, a flavivirus with positive polarity ssRNA genome bearing a 5'-end cap structure: m⁷Gp_αp_βp_αAm-RNA. It was reported that a mutant WNV lacking the ribose-2'-O-hydroxyl methylation was attenuated in primary cells and mice but was virulent in cells presenting an insufficient IFN signalling pathway (Daffis *et al.*, 2010). This strongly suggests that the ribose-2'-O-hydroxyl methylation is a signature of "self" RNA (Daffis *et al.*, 2010) and that many families of viruses including *Flaviviridae*, *Coronaviridae*, and *Poxviridae*, have evolved to evade the host innate immunity through expression of 2'-O-methyltransferases in order to decorate their caps with ribose-2'-O-hydroxyl methylation (Daffis *et al.*, 2010, Choi *et al.*, 2018).

The detection by IFIT1 of viral RNA with 5'- triphosphate ends can result in sequestration of the viral RNA by IFIT1 into a multiprotein complex containing IFIT2 and IFIT3 (Pichlmair *et al.*, 2011). Recently, it has been proposed that the interaction of the CTD domain of IFIT3 with IFIT1 is required for allosteric regulation of IFIT1 RNA-binding channel and to promote selective recognition of the cap-0 structure over 5'- triphosphate RNA ends or cap-1 structures (Johnson *et al.*, 2018). Notably, recent functional studies on WNV, ZIKV and VEEV, have suggested that the IFIT3-IFIT1 interaction is crucial for stable IFIT1 expression, IFIT1 stability by increasing its half-life and clearance of viruses lacking the ribose-2'-O-hydroxyl methylation in their RNA cap structures (Johnson *et al.*, 2018). This results in translation inhibition of mRNA lacking the ribose-2'-O-hydroxyl methylation and consequently leads to inhibition of viral RNA replication (Pichlmair *et al.*, 2011, Abbas *et al.*, 2013, Hyde & Diamond, 2015, Diamond, 2014, Kumar *et al.*, 2014).

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

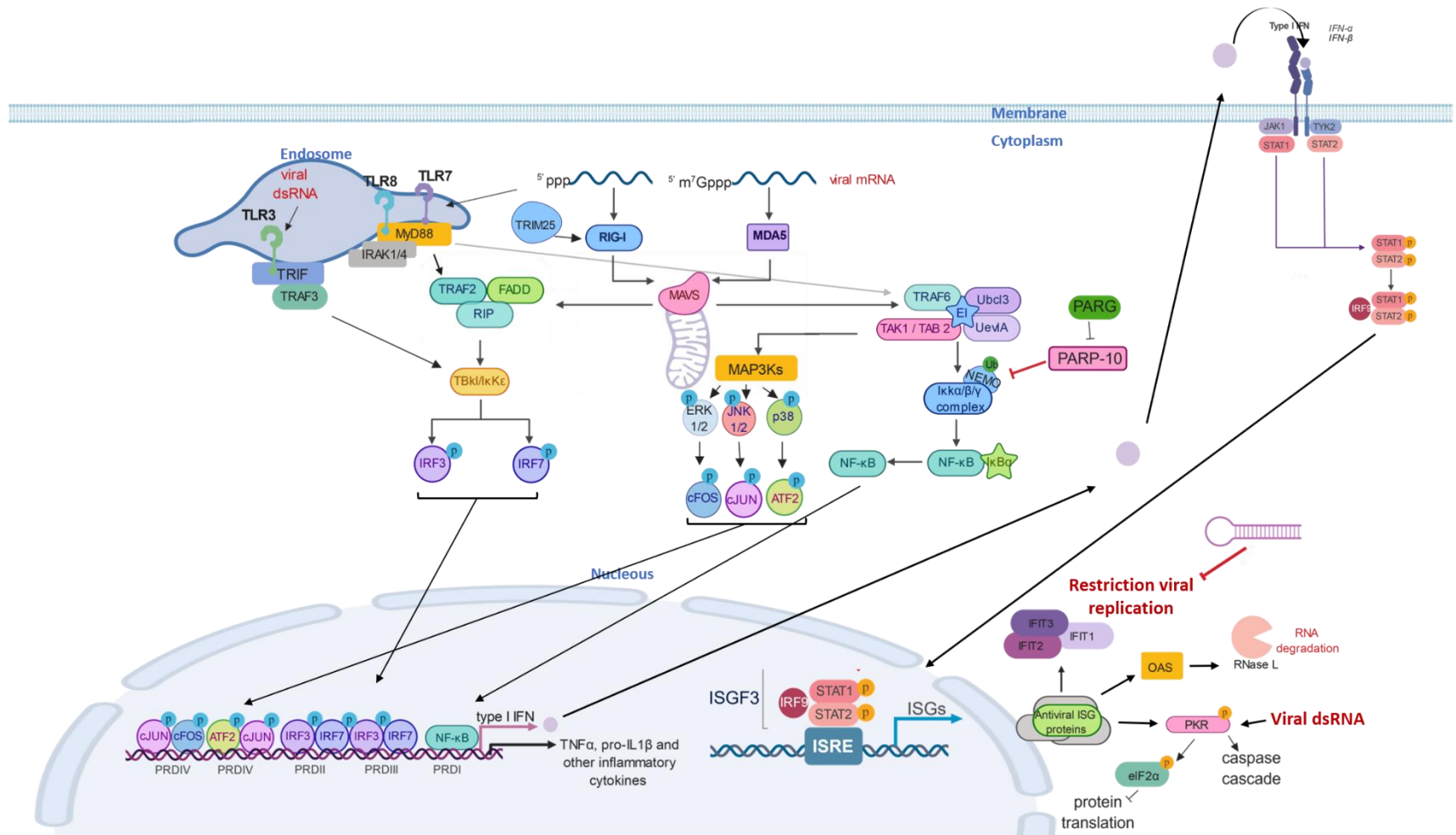


Figure 30. Activation of the innate immune response through sensing of the viral RNA. See legend on the next page.

Figure 30. Activation of the innate immune response through sensing of the viral RNA. The RNA from alphaviruses protected with a type 0 cap structure or non-protected (5' triphosphate ends) can trigger the activation of the innate immune sensors such as RIG-I, MDA5, TLR3, TLR7 and TLR8. RIG-I and MDA5 sensors which are known to activate the mitochondrial-anchored protein, the signalling adaptor IFN β promoter stimulator 1 (IPS1) which is also known as MAVS. At least two signalling cascades are activated by MAVS. One is the same cascade that is also triggered by TLR 3, 7 and 8 that leads to the phosphorylation of IRF3 and IRF7 that will dimerise and after translocation to the nucleus bind positive regulatory domains (PRD) II and III. Therefore, leading to the transcription of type I IFN and inflammatory cytokines such as pro-IL1 β . By another hand, MAVS also activates the TRAF6 signalling cascade. In order to activate the transforming growth factor β -activated kinase (TAK1) complex, an E1 ubiquitin activating enzyme (E1) is recruited to the TRAF6/TAK1/TAB2 complex together with a ubiquitin conjugating enzyme complex (E2) formed by Ubc13 and Uev1A. The TRAF6, an ubiquitin ligase, is ubiquitinated and activates TAK1 complex. Again, two possible pathways are triggered by TAK1 complex, one is the NF κ B pathway and the other is the MAP3Ks cascade. The I κ B α / β / γ complex is stabilized by NEMO in its ribosylated form. However, upon ubiquitination of NEMO, NEMO is degraded by proteasome mediated proteolysis and the I κ B α / β / γ complex is activated and leads to activation of NF κ B by directing for degradation the regulatory subunit (I κ B) of NF κ B and allowing the translocation to the nucleus of the p65-p50 of NF κ B. The p65-p50 of NF κ B usually binds to PRD I and allows the activation of antiviral genes. The activation of the MAP3Ks by the TAK1 complex also leads to downstream events of phosphorylation that will lead to the activation of several transcription factors belonging to the families of cJUN, cFOS and ATF2 and involved in the transcription of antiviral genes. Importantly, the detection of viral RNA leads to the production of type I IFN that through an autocrine and paracrine (and eventually endocrine) leads to the activation of the janus kinase-signal transducer and activator of transcription (JAK-STAT) signalling cascade to induce the expression of IFN-stimulated genes (ISGs). Among several ISGs are the interferon induced proteins with tetratricopeptide repeats (IFIT proteins) that recognise 5' triphosphate RNA and RNA bearing a cap-0 structure. However, alphaviruses have evolved in a way to escape to IFIT proteins because alphaviruses contain a special RNA secondary structure near the cap-0 structure and that blocks the recognition of the cap (or triphosphate ends) by IFIT proteins. Other ISG produced by the JAK-STAT signalling cascade is the protein kinase PKR that can be activated by viral RNA or through components of the JAK-STAT pathway and leads to phosphorylation of eIF2 α and inhibits host protein translation. Another one is the 2',5'oligo(A) oligonucleotides synthesized by oligoadenyl synthases (OASs) that activates RNase L which is involved in degradation of viral RNA (capped and non-capped). See text for more details. Figure designed using <https://biorender.com/>.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Alphaviruses lack ribose-2'-O-hydroxyl methylation or a cap-snatching mechanism and thus their cap-0 structure type at the 5'-end of the genomic and sub-genomic RNA (Hefti *et al.*, 1975, Pettersson *et al.*, 1980, Dubin *et al.*, 1977, Wengler *et al.*, 1979) must be recognized by IFIT1. Surprisingly, it was discovered that an RNA secondary structure from VEEV (strain TC-83) present at the 5'-UTR of the genomic and sub-genomic RNA (Strauss & Strauss, 1994), near of the cap structure, antagonizes IFIT1 binding and its antiviral activity (Roby *et al.*, 2014, Hyde *et al.*, 2014), Fig. 30. Mutations within the 5'-UTR that affect the stable RNA structural elements enabled restriction or antagonism by IFIT1 *in vitro* and *in vivo* by altering binding of IFIT1 to viral RNA (Hyde & Diamond, 2015, Hyde *et al.*, 2014). These evidences were strongly supported by early studies on VEEV where a mutation at the 5'-UTR (G to A at the position 3) attenuated the virus (Kinney *et al.*, 1993, Kinney *et al.*, 1989) and resulted in enhanced sensitivity to type I IFN treatment (White *et al.*, 2001). Indeed, in other alphaviruses, such as SINV and SFV, similar mutations at the 5'-UTR produced a similar effect of altered pathogenicity and sensitivity to type I IFN treatment (Klimstra *et al.*, 1999, Kobilier *et al.*, 1999, Kuhn *et al.*, 1992, Logue *et al.*, 2008). In alphaviruses the translation inhibition induced by type I IFN is determined by both IFIT1 -dependent and -independent mechanisms. The variability of structural elements present at the 5'-UTRs of the different alphaviruses strains are determinant in their ability to replicate in the presence of IFIT1. For example, CHIKV, SINV, SFV and VEEV with natural *WT* 5'-UTRs replicated in IFIT1-expressing cell lines more efficiently than other alphaviruses. Therefore, IFIT1 acts as both an antiviral effector protein and inducer of innate immunity against alphaviruses (Reynaud *et al.*, 2015), Fig. 30.

6. Insights into *Macro* domains: a key module participating in metabolic pathways requiring ADP-ribose (ADPr).

After more than 20 years of research upon the discovery of the first *Macro* domain (Lee *et al.*, 1991), it seems that some functional clues are finally appearing, opening a new and exciting research field. This module was early identified in viruses as the "X domain" that was just next to the putative coronavirus papain-like protease, within the genomes of Mouse Hepatitis Virus (MHV) and Infectious Bronchitis Virus (IBV) (Lee *et al.*, 1991). At that time, *Macro* domains from MHV and IBV were also compared to homologous and conserved domains, present in viruses belonging to the *Togaviridae* family, which were flanking the putative thiol protease. *Macro* domains were thus found to be largely conserved among several ssRNA viruses of positive polarity (Lee *et al.*, 1991, Rack *et al.*, 2016).

It was in 1992 that the first *Macro* domain was identified in vertebrates. On that occasion, the *Macro* domain discovered in *MacroH2A* protein, found in rat liver nucleosomes, was identified as a domain of unknown function (Pehrson & Fried, 1992). Then, only later its hints of function were identified upon their ADP-ribose (ADPr) binding properties (Chakravarthy *et al.*, 2005, Kustatscher *et al.*, 2005, Costanzi & Pehrson, 1998). Thereafter, *Macro* domains were found to be broadly distributed and highly evolutionarily conserved through all kingdoms of life, including a variety of viruses (Karras *et al.*, 2005). *Macro* domain modules were suggested to have co-evolved according to the complexity of the organism. For example, in eukaryotes, the number of *Macro* domains-containing proteins increases with the complexity of the organism (Perina *et al.*, 2014, Rack *et al.*, 2016). Nowadays, a *Macro* domain is a highly conserved ADPr binding module presenting structural conservation, although it remains possible that other classes of *Macro* domains existed but lacking in structural conservation (Rack *et al.*, 2016).

The primordial and diversified family of *Macro* domains are critical for ADPr recycling through continuous breakdown and re-synthesis of individual molecules from the signalling pathways requiring NAD⁺ (ADPr metabolic turnover) (Rack *et al.*, 2016). Indeed, the resulting ADPr can act as both as a precursor of NAD⁺ as well as a secondary messenger (Fliegert *et al.*, 2007). *Macro* domains are recognized to interact with two protein families that make use of NAD⁺, which are the poly-(ADP-ribose) polymerases (PARPs) (Perina *et al.*, 2014, Feijs, Verheugd, *et al.*, 2013, Gibson & Kraus, 2012) and the sirtuins (Denu, 2005, Sauve & Youn, 2012). Although, sirtuins are usually known to promote protein diacylation by using NAD⁺ as a co-substrate and releasing O-acyl-ADP-ribose (OAADPr) (Denu, 2005, Sauve & Youn, 2012), they also catalyse protein ADP-ribosylation (Rack *et al.*, 2015, Kowieski *et al.*, 2008), Fig. 31.

Several PARPs, such as PARP1, PARP2, PARP5a/b and PARP15, are known as “signal writers” because are involved in the post-translational modification of proteins through ADP-ribosylation, Fig. 31. This post-translational modification is catalysed through the transfer of single or multiple ADPr unit(s) from NAD⁺ onto a primary acceptor, usually a glutamate or aspartate residues via an ester bond (Kleine *et al.*, 2008, Ecker *et al.*, 2017, Tao *et al.*, 2009, Ogata *et al.*, 1980, Burzio *et al.*, 1979, Riquelme *et al.*, 1979, Matic *et al.*, 2012, Sharifi *et al.*, 2013), resulting therefore in mono-(ADP)-ribosylation (MARylation) of the target proteins (Perina *et al.*, 2014, Feijs, Verheugd, *et al.*, 2013, Gibson & Kraus, 2012, Barkauskaite *et al.*, 2015, Vyas *et al.*, 2014). Later, the grafted ADPr can further receive other ADPr molecules (poly-[ADP]-ribosylation or PARylation). Lysine (Altmeyer *et al.*, 2009, Messner *et al.*, 2010, Daniels *et al.*, 2014), cysteine (Vyas *et al.*, 2014), arginine (Martello *et al.*, 2016, Daniels *et al.*, 2014, Hottiger, 2015,

Rosenthal & Hottiger, 2014), serine (Bonfiglio *et al.*, 2017, Leidecker *et al.*, 2016) and tyrosine (Leslie Pedrioli *et al.*, 2018) can serve as well as primary acceptor. After the synthesis of the initial ester bond, a set of PARPs are known to catalyse the formation of the 2',1''-O-glycosidic ribose-ribose bond to synthesize linear or branched chains of ADPr polymers (up to 200 ADPr units) (Miwa *et al.*, 1979, Juarez-Salinas *et al.*, 1982, Kanai *et al.*, 1982, Rolli *et al.*, 1997, D'Amours *et al.*, 1999, Barkauskaite *et al.*, 2015, Kleine *et al.*, 2008, Langelier & Pascal, 2013, Miwa *et al.*, 1981), Fig. 31.

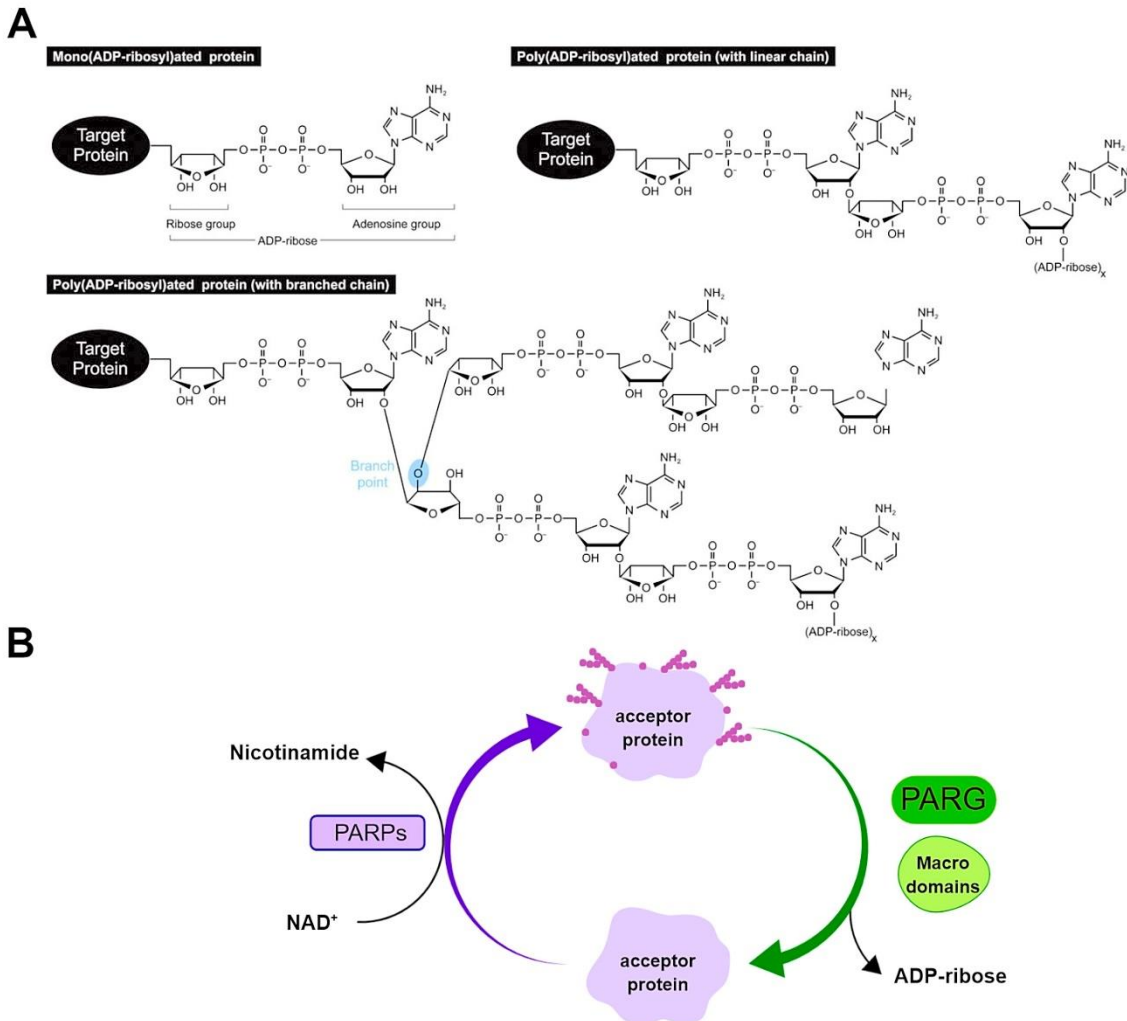


Figure 31. Mono- and poly-ADP ribosylation and reversible post-translational modification. (A) Mono- or poly- (ADP-ribose) is covalently attached mainly to aspartic acid or glutamic acid of acceptor proteins and the ADPr units of the polymer are connected linearly or in a branched fashion (Leung, 2014). **(B)** The metabolism of poly-(ADP-ribose) requires PARPs (poly-[ADP-ribose] polymerases) for biosynthesis of poly-(ADP-ribose) from NAD⁺ while PARG (poly-[ADP-ribose] glycohydrolase) degrades polymer of poly-(ADP-ribose) to ADP-ribose. For more details see text. Adapted from (Tan *et al.*, 2012). Figure 30 B was designed using <https://biorender.com/>.

Macro domain modules efficiently bind free-ADPr, recognize ADPr in derived metabolites and (or) covalently linked to proteins in order to signal and/or control a diversified set of biological events such as DNA damage repair, signal transduction, transcription, maintenance of genomic stability, telomere dynamics, cell differentiation, cell proliferation, necrosis, apoptosis, and immune response (Taverna *et al.*, 2007, Ahel *et al.*, 2009,

Karras *et al.*, 2005, Kraus, 2009, Johnson *et al.*, 2010, Tan *et al.*, 2009). Therefore, several pathologies such as cancer, inflammation, development defects and neurodegeneration might be correlated with abnormal function of *Macro* domains in the cell.

Consequently, *Macro* domain modules are known to control a plethora of events through recognition of protein post-translational modifications. Thus, function as “signal reading” (or “signal interpreting”), “signal erasing” and “signal writers” (Rack *et al.*, 2016). The “signal reading” function of *Macro* domains usually recognize the post-translational modifications such as MARYlation and PARYlation. The “signal erasing” results in removal of the post-translational modification from a protein, thus reversing MARYlation and PARYlation. Several *Macro* domains have the capabilities to hydrolyse of the 2',1'-O-glycosidic ribose-ribose bond in poly-(ADP-ribose) (PAR) chain, the protein-ADPr ester bond, or the acyl-ADPr ester (Chen *et al.*, 2011, Barkauskaite *et al.*, 2015, Jankevicius *et al.*, 2013, Miwa *et al.*, 1981, Peterson *et al.*, 2011, Rosenthal *et al.*, 2013, Sharifi *et al.*, 2013, Slade *et al.*, 2011). The “signal interpreting” may lead to a spatial/temporal recognition of post-translational modifications, which activates a set of downstream events carried-out by effector proteins (Rack *et al.*, 2016).

Due to the wide diversity of *Macro* domains, a phylogenetic analysis subdivided them into six classes which are recognized as *MacroD*-type, *Macro2*-type, *MacroH2A*-like, amplified in liver cancer (ALC1)-like, poly-ADP-ribose glycohydrolase (PARG)-like, and middle of the SARS-unique domain (SUD-M)-like. Although all participate in NAD⁺-utilizing pathways, the PARG-like and SUD-M-like are the more divergent classes. Indeed, the recognition of SUD-M-like as a *Macro* domain occurred only after the determination of the three-dimensional structure of representatives (Tan *et al.*, 2009, Slade *et al.*, 2011, Chatterjee *et al.*, 2009). Humans have at least 12 *Macro* domain containing proteins which are distributed among four classes: *MacroH2A*-like (at least 9 *Macro* domains -containing proteins), ALC1-like (at least 2 *Macro* domains -containing proteins), PARG-like (at least 1 *Macro* domain -containing protein), and *MacroD*-type (at least 3 *Macro* domains -containing proteins) (Rack *et al.*, 2016, Chen *et al.*, 2011, Feijs, Verheugd, *et al.*, 2013). Concerning viruses from the *Togaviridae* family and *Hepeviridae*, only one *MacroD*-type have been reported to date (Rack *et al.*, 2016). However, a member from the *Coronaviridae* family has been shown to present both *MacroD*-type and SUD-M-like (Rack *et al.*, 2016).

Although, some *Macro* domains are known to read both protein ADP-ribosylation as well as PARP-dependent signalling, they also can interpret second messengers such as ADP-ribose and its derivatives. This is known as “signal reading” (or “signal interpreting”).

The derivatives of ADP-ribose can be released either by PARP-dependent activity as well as through sirtuin activation. The *MacroH2A*-like members are included in this category because some members of this class show strong binding to ADP-ribosylated proteins and own no catalytic activity. Despite binding studies having reported that the variant *MacroH2A1.1* interacts with ADP-ribose, O-acetyl-ADP-ribose (AAR) and PAR (Kustatscher *et al.*, 2005), this is not a general rule for all *MacroH2A*-like members (Rack *et al.*, 2016). The *MacroH2A1.1* was found to be crucial for optimal mitochondrial function but not required for gene regulation. The direct binding of *MacroH2A1.1* to the PARP1 induces inhibition of the PARP1 basal activity of PARylation which reduces the nuclear consumption of NAD⁺ allowing the homeostasis of mitochondrial NAD⁺ pools. Thus *MacroH2A1.1* seems to be crucial for the regulation of mitochondrial respiration through limiting the consumption of NAD⁺ (Posavec Marjanovic *et al.*, 2017).

The member from the family ALC1-like are also characterized as “signal reading” (or “signal interpreting”) because some *Macro* domains, like the one from ALC1, own the ability to sense PARP1-generated PAR, and the poly-(ADP)-ribosylated-PARP1 is recruited to the site of DNA damage (Ahel *et al.*, 2009, Gottschalk *et al.*, 2009). Accordingly, a stable ALC1-nucleosome-poly-(ADP)-ribosylated-PARP1 complex is formed and leads to an increase in the chromatin-remodelling activity of ALC1 (Ahel *et al.*, 2009, Forst *et al.*, 2013, Gottschalk *et al.*, 2012). Interestingly, another member from the ALC1-like class is the TARG1 (the terminal ADPr protein glycohydrolase 1, also known by the alternative names: OARD1 or C6orf130) for which experimental work showed that TARG1 can hydrolyse protein ADP ribosylation (Rosenthal *et al.*, 2013, Sharifi *et al.*, 2013). Consequently, it has been suggested that TARG1 function as “signal erasing” because it possesses mono-(ADP-ribosyl) hydrolase activity, through a different mechanism than that used by *MacroD*-type, though (Sharifi *et al.*, 2013). The mono-(ADP-ribosyl) hydrolase carried-out by TARG1 relies on a nucleophilic attack of the C1' atom of the distal ribose through a conserved Lys residue, thus leading to the formation of a lysyl-ADP-ribose intermediate and release of the Glu (or Asp) residue or the target protein (Sharifi *et al.*, 2013). Then, the Lys residue returns to the initial state with the help of a proximal catalytic Asp residue, inducing the release of ADPr from TARG1. Possibly, TARG1 can bind to poly-(ADP)-ribosylated proteins and remove the PAR-chain due to its higher solvent accessibility at the 2'-OH position of the adenosine ribose (Sharifi *et al.*, 2013, Barkauskaite, Brassington, *et al.*, 2013).

The poly-(ADP-ribosyl) glycohydrolases (PARG)-like class members are particularly associated with degradation of the PAR chain (Barkauskaite *et al.*, 2015, Miwa & Sugimura, 1971, Slade *et al.*, 2011). PARG owns a *Macro* domain that forms a catalytic

core together with an accessory domain. Although the loop 2 signature motif of PARG *Macro* domain is the same as in *MacroD*-type, it harbours a specific catalytic motif, Gly-Gly-Gly-X(6-8)-Gln-Glu-Glu, inserted within the loop 1. The folding is different, involving an arrangement of a seven-stranded mixed β -sheet associated with five α -helices (Slade *et al.*, 2011, Dunstan *et al.*, 2012, Tucker *et al.*, 2012, Kim *et al.*, 2012, Lambrecht *et al.*, 2015). Some works have reported that the 2',1'-O-glycosidic ribose-ribose bond is localized near the second catalytic Glu (Rack *et al.*, 2016). The catalytic mechanism of PARG is different from the mechanism used by *MacroD*-type *Macro* domains and it is triggered through acid/base protonation of the ribose 2'-OH in the proximal ADPr moiety, leading to the formation of an oxocarbenium intermediate that is stabilized through a Phe from loop 2. Thereafter, a water molecule through a nucleophilic attack, attacks the oxocarbenium ion, resulting in release of the ADPr and the remaining PAR chain (Rack *et al.*, 2016, Slade *et al.*, 2011, Dunstan *et al.*, 2012, Tucker *et al.*, 2012, Kim *et al.*, 2012). In summary, PARG seems to be mainly an exoglycohydrolase with minor contributions of endocleavage and due to structural constrains it seems unlikely that PARG can cleave at PAR branch points (Tucker *et al.*, 2012, Kim *et al.*, 2012, Slade *et al.*, 2011). However, it seems possible that endocleavage might occur under cellular stress conditions and it could release oligo- or poly (ADPr) fragments (Barkauskaite, Jankevicius, *et al.*, 2013, Barkauskaite, Brassington, *et al.*, 2013).

Concerning *MacroD*-type class, the members from this class are mainly represented by *MacroD1* and *MacroD2* and it is well known that they are involved in reversing mono-(ADP)-ribosylation (Jankevicius *et al.*, 2013, Rosenthal *et al.*, 2013). The structural organization and function of the members of the *MacroD*-type class will be further addressed in the chapter 6.1 intitled "Structure and function of *MacroD*-type".

6.1. Structure and function of *MacroD*-type

The *Macro* domain structure follows a globular $\alpha/\beta/\alpha$ sandwich fold composed of a central six-stranded mixed β -sheet flanked by five α -helices (Kraus, 2009, Allen *et al.*, 2003, Egloff *et al.*, 2006, Kustatscher *et al.*, 2005). Structural and functional studies provided evidences that ADPr and derivatives can bind within a deep, well-formed cavity (Kustatscher *et al.*, 2005, Egloff *et al.*, 2006, Karras *et al.*, 2005). Thus, supporting that *Macro* domains have a nucleotide-binding pocket and processing activity (Rack *et al.*, 2016). The ligands are stabilized within the binding pocket through conserved interactions which are, for example, π - π stacking interactions between the adenosine moiety and a conserved aromatic residue, coordination of the N⁶-adenosine by an Asp residue (Ahel *et al.*, 2009, Karras *et al.*, 2005, Chen *et al.*, 2011, Gottschalk *et al.*, 2009),

side-chain/backbone-pyrophosphate contacts (Karras *et al.*, 2005, Forst *et al.*, 2013), coordination of pyrophosphates with the diphosphate-binding loop (substrate-binding loop 2), and accommodation of the diphosphates and distal-ribose between the diphosphate-binding loop and the catalytic loop (substrate-binding loop 1) (Timinszky *et al.*, 2009, Peterson *et al.*, 2011, Slade *et al.*, 2011, Zaja *et al.*, 2012).

The interactions are crucial to stabilize the *Macro* domain in complex with the ligand, among which the side-chain/backbone-pyrophosphate contacts are important because they induce a closed conformation of the *Macro* domain (Karras *et al.*, 2005, Forst *et al.*, 2013). Also, the catalytic residues from the catalytic loop, mainly those from *Macro* domains presenting hydrolase activity, are important for substrate specificity (Timinszky *et al.*, 2009, Peterson *et al.*, 2011, Slade *et al.*, 2011, Zaja *et al.*, 2012).

Several members from the class *MacroD*-type are known to present mono-(ADP-ribosyl)-hydrolase activity (“erasing *Macro* domains”) and this class is represented by the prototypes *MacroD1* and *MacroD2*. Two specific signature motifs are embedded within the loops 1 and 2 of *MacroD*-type *Macro* domains, which are Asn-X(6)-Gly-Gly-[Val/Leu/Ile] and Gly-[Val/Ile/Ala]-[Tyr/Phe]-Gly respectively. The conserved aromatic residues from both loop 1 and 2 are essential for the adopted constrained conformation of the distal ribose which is bended toward the ADPr α -phosphate group (Jankevicius *et al.*, 2013, Chen *et al.*, 2011). A structural water (positioned between the α -phosphate and distal ribose) playing a coordination role in the groove of the pyrophosphate-binding site distinguish these members from the “reading *Macro* domains”.

To date there are at least two mechanism proposed for the activity of the present mono-(ADP-ribosyl)-hydrolases. One of such is the “substrate-assisted mechanism” in which the structural water molecule becomes activated through the α -phosphate group and carries-out a nucleophilic attack on the protein-ADPr ester bond. However, there is also the possibility in which the constrained conformation of the substrate is likely required (Jankevicius *et al.*, 2013, Barkauskaite, Brassington, *et al.*, 2013). The other mechanism relies in the conserved Asp residue from the active site that acts as general base for activation of the water molecule, which in turn promotes a nucleophilic attack on the C1' atom of the distal ribose (Rosenthal *et al.*, 2013, Chen *et al.*, 2011). The later mechanism seems to be the more widely accepted and suggests a coordinated mechanism involving simultaneous nucleophilic attack and ester bond breakage (Hirsch *et al.*, 2014). However, not all catalytic active *MacroD*-type *Macro* domains have a catalytic Asp residue (Jankevicius *et al.*, 2013).

Therefore, *MacroD*-type *Macro* domains participate in ADPr-dependent signalling pathways that require both recognition and removal of the signal. It has been suggested that *Macro* domains reverse cellular ADP-ribosylation by catalysing signal termination reactions including hydrolysis of mono- and poly (ADP)-ribosylation as well as degradation of NAD⁺- derived second messengers such as OAADPr (Chen *et al.*, 2011, Feijs, Forst, *et al.*, 2013, Barkauskaite, Jankevicius, *et al.*, 2013, Tong & Denu, 2010, Rack *et al.*, 2016). Moreover, *MacroD*-type members such as *MacroD1* and *MacroD2* catalyse the cleavage of the terminal ADPr moiety from proteins, for example after PARG-mediated PAR-chain degradation (Barkauskaite, Jankevicius, *et al.*, 2013, Jankevicius *et al.*, 2013, Barkauskaite, Brassington, *et al.*, 2013). Despite their unknown targets and biological roles, the primary subcellular localizations of both *MacroD1* and *MacroD2* are different which suggest different functions (Neuvonen & Ahola, 2009). For example, the *MacroD1* is localized at the mitochondria and *MacroD2* at the cytoplasm (Neuvonen & Ahola, 2009). Both *MacroD1* and *MacroD2* were reported to be involved in cancer and *MacroD2* seems to reverse the PARP10-mediated mono-(ADP)-ribosylation (Rosenthal *et al.*, 2013).

Macro domains are found among different viruses such as the families: *Coronaviridae*, *Togaviridae*, *Iridoviridae*, *Poxviridae*, and *Hepeviridae* (Egloff *et al.*, 2006, Malet *et al.*, 2009, Malet *et al.*, 2006, Putics *et al.*, 2006). Structural and phylogenetic studies have suggested that viral *Macro* domains are strongly related to cellular *Macro* domains as a result of co-evolution with the host (Neuvonen & Ahola, 2009, Rack *et al.*, 2016), thus, suggesting that viral *Macro* domains might present mono-(ADP-ribosyl)-hydrolase activity (Jankevicius *et al.*, 2013), Fig. 32. This has been supported by studies that showed that alphaviruses *Macro* domains can hydrolyse ADP-1''-phosphate and bind to PAR-chain (Egloff *et al.*, 2006, Malet *et al.*, 2009, Malet *et al.*, 2006, Putics *et al.*, 2005). A *MacroD*-type *Macro* domain is encoded as part of the nsP3 from both corona- and alphaviruses which suggest a limited overlapping function because the other regions of both nsP3 are different (Rack *et al.*, 2016, Neuvonen *et al.*, 2011, Gorbalenya *et al.*, 2006, LaStarza, Lemm, *et al.*, 1994, Neuman *et al.*, 2008). Indeed, the ADP-ribosyl-binding and hydrolase activities of *Alphavirus Macro* domains are required for virus replication (Abraham *et al.*, 2018, McPherson *et al.*, 2017, Li *et al.*, 2016).

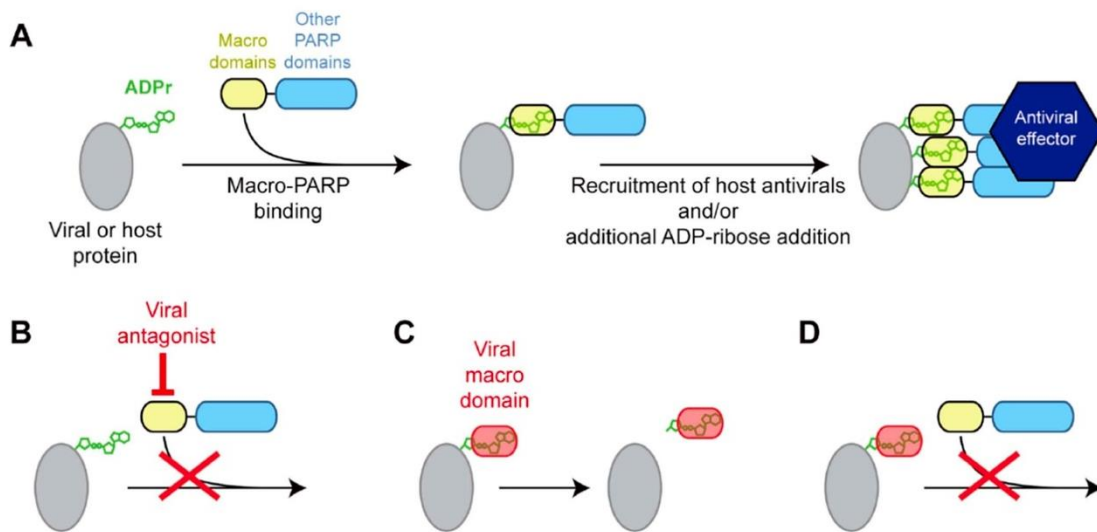


Figure 32. Model for genetic conflict involving PARP Macro domains. (A) Model for *Macro* PARP function. ADP-ribosylated host or viral proteins may be a signal for recruitment of PARP9, 14 or 15, which could facilitate additional recruitment of antiviral effectors, and amplify the initial ADPr signal. (B–D) Three models for how viruses may antagonize *Macro* PARP function. Viruses lacking their own *Macro* domains may use other proteins to directly antagonize *Macro* PARP proteins (B), driving recurrent positive selection in *Macro* PARP genes to escape antagonism. *Macro* domains encoded by viruses (e.g. corona- and togaviruses) may catalyze the removal of ADPr (C) or compete with *Macro* PARPs for binding to ADPr (D) in order to antagonize host ADPr-mediated signalling (Daugherty *et al.*, 2014).

6.2. The connection of *MacroD*-type and the innate immune response

A diversity of viruses, including alphaviruses, have developed strategies to overcome the host immune system. The discussion will be limited here to those having a direct connection to *Macro* domains. Inside the cell, detection of elements of the pathogen (e.g., viral RNAs) is triggered mainly by the RIG-I and MDA-5 sensors leading to the activation of the innate immune response. The activation of the downstream pathways through RIG-I and MDA-5 induce the formation of IPS1 signalling complexes that activates the transcription factors such as the IRFs (e.g. IRF3 or IRF7) and NF- κ B. This transcription factors are known to induce the production of IFN-I and pro-inflammatory cytokines (e.g. IL-1 β) (Meylan *et al.*, 2005, Seth *et al.*, 2005, Kawai *et al.*, 2005, Xu *et al.*, 2005, Kolakofsky *et al.*, 2012), Fig. 32. the cytokine-mediated signalling is an essential component of the innate immunity and aims to inform/alarm neighbouring (and immune) cells of the infection and activates defence mechanisms in an autocrine and paracrine mode. The autocrine and paracrine stimulation of INFR by IFN α/β leads to the activation of the JAK-STAT pathway that induce the expression of ISGs (Hyde & Diamond, 2015). Moreover, IFN-I also triggers the expression of several *PARP* genes, for example the

antiviral PARPs (PARP7, PARP10 and PARP12) (Atasheva *et al.*, 2012, Atasheva *et al.*, 2014), Fig. 33. This allows the host to control the antiviral response.

The NF- κ B is a family of 5 proteins that are known to form hetero- and homo- dimers and usually it is formed the heterodimer p50/p65. This transcription factor NF- κ B is important for prevention of apoptosis initiation and to regulate the expression of *tumor necrosis factor- α* (TNF- α), *the inducible nitric oxide synthase (iNOS)*, *interleukines IL-1 β , IL-2, IL-6 and IL-8*, as well as *adhesion molecules ICAM-1 and E-selectin* genes (Shall & de Murcia, 2000), Fig. 33. The pathogenesis of alphaviruses is exacerbated due to the contribution of the excessive production of soluble innate mediators such as IL-6, TNF- α , granulocyte macrophage-colony stimulating factor (GM-CSF), interferon- γ (IFN- γ), macrophage chemoattractant protein-1 (MCP-1) and macrophage migrating inhibitory factor (MIF) (Chow *et al.*, 2011, Daigo & Hamakubo, 2012, Herrero *et al.*, 2011). Interestingly, it has been proposed the involvement of the complement cascade activation in modulating alphavirus pathogenesis due to evidences of over expression of the complement factor 3 (C3) (Morrison *et al.*, 2007) and mannose binding lectins (MBLs) (Gunn *et al.*, 2012) during alphavirus infection.

PARP1 largely contributes to the up-regulation of “danger” signals creating the condition to initiate and sustain the innate immune response (Laudisi *et al.*, 2011). Under some circumstances, PARP-1 is fundamental for the innate immune response, Fig. 33, thus involving ADP-ribosylation for activation of the necessary pathways for production of IL-6, TNF- α and iNOS in monocytes, macrophages and dendritic cells (Hauschildt *et al.*, 1991, Hauschildt *et al.*, 1992, Pellat-Deceunynck *et al.*, 1994, Le Page *et al.*, 1997, Heine *et al.*, 1995). Interestingly, during the process of differentiation of monocytes into dendritic cells, the levels of PARP1 mRNA and protein levels are high (Aldinucci *et al.*, 2007). Furthermore, inhibition of PARP-1 activity was found to reduce NF- κ B and AP-1 DNA binding activity (Aldinucci *et al.*, 2007), Fig. 33. Apart from the innate immune system, it seems that PARP1 also is involved in regulation of other pathways from the adaptative immunity such as differentiation of CD4⁺CD25⁺/Foxp3⁺ regulatory T cells (Treg) (Nasta *et al.*, 2010), differentiation of T cells into T help 2 (T_h2) (Sambucci *et al.*, 2013) and B cell maturation (Arakawa *et al.*, 2002, Paddock *et al.*, 2010).

Some reports have provided evidences that PARP1 forms stable complexes with transcription factors such as octamer-binding transcription factor-1 (Oct-1) (Nie *et al.*, 1998), elongation factor 1 alpha (TEF-1) (Butler & Ordahl, 1999), myeloblastosis viral oncogene homolog (avian)-like2 (B-Myb) (Cervellera & Sala, 2000) and adaptor protein 2 (AP-2) (Kannan *et al.*, 1999) which have important roles in producing chemokines (Bai & Virag, 2012), cytokines, adhesion factors and other inflammatory mediators (Virag &

Szabo, 2002), Fig. 33. It has been suggested that PARP-1 function as a cofactor in the activation of NF- κ B that works as a transcriptional activator of the expression of *iNOS* gene (Le Page *et al.*, 1998, Oliver *et al.*, 1999), Fig. 33. This occurs likely after translocation of NF- κ B to the nucleus which seems to be mediated by PARP-1 (Oliver *et al.*, 1999), Fig. 33. In addition, PARP-1 might be involved in regulation of NF- κ B activity as PARP-1 mediated poly-(ADP)-ribosylation of transcription factors may occur before binding to the promoter and disable transcription factors to bind to the specific DNA recognition sites (Ziegler & Oei, 2001). Indeed, components of NF- κ B are poly-(ADP)-ribosylated by PARP-1, which disable binding of NF- κ B to bind its DNA recognition site (Kameoka *et al.*, 2000), Fig. 33.

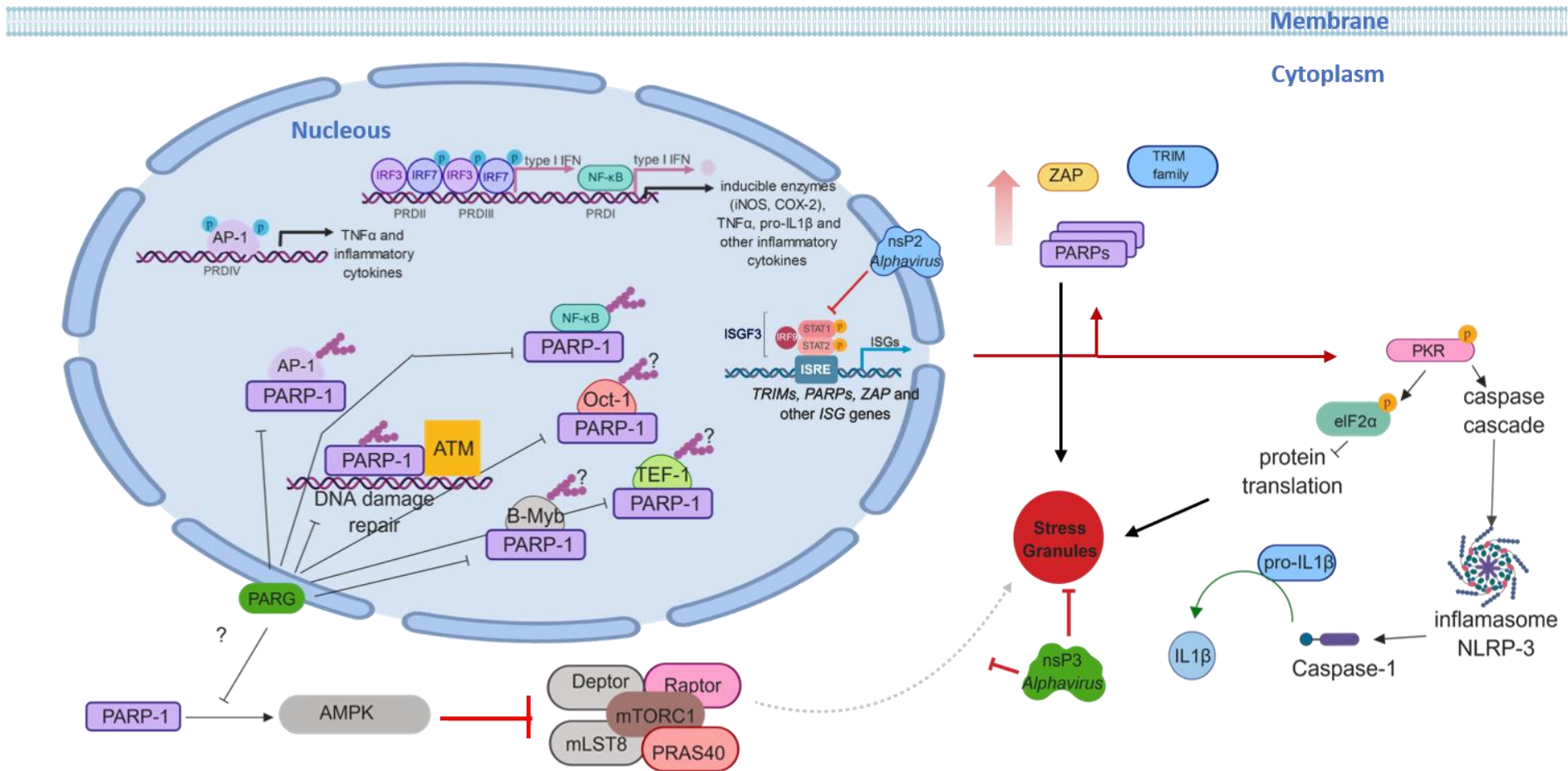


Figure 33. Involvement of PARPs during the innate antiviral response. See legend continuation in the next page.

Figure 33. Involvement of PARPs during the innate antiviral response. The RNA from alphaviruses protected with a type 0 cap structure or non-protected (5' triphosphate ends) can trigger the activation of the innate immune sensors such as RIG-I, MDA5, TLR3, TLR7 and TLR8. RIG-I and MDA5 sensors which are known to activate the mitochondrial-anchored protein, the signalling adaptor IFN β promoter stimulator 1 (IPS1) which is also known as MAVS. The signalling cascades activated were already described in figure 30 and are here presented again. The major new representation in this figure are the involvement of PARP proteins during the antiviral response. PARP-10 is known to be involved in ribosylation of NEMO which confers more stability to the I κ B α / β / γ complex in its inactive state. Most likely the effect of PARP10 can be counterbalanced by PARG. Therefore, the translocation of NF κ B to the nucleus can be modulated by PARP10. In addition, PARP1 seems to be involved in regulation of active NF κ B binding to the positive regulatory domains (PRD) I. Other transcription factors may be regulated in a similar fashion. For example, it has been described that PARP1 can form stable complexes with NF κ B, Oct-1, TEF-1, B-Myb, ATM and AP-1. Although PARP1 is highly present at the nucleus playing several pathways modulations, it has been described that PARP1 is also required for inactivation of mTORC1. To inactivate mTORC1, PARP1 activates AMPK that inhibits mTORC1. The inhibition of mTORC1 leads to its relocation inside stress granules under stress conditions. Importantly, alphaviruses developed original strategies to fight against the innate immune system. One of such is nsP2 that can inactivate the JAK-STAT signalling cascade as previously mentioned. However, another strategy involves the multifunctional role of nsP3. nsP3 has been associated with disassembly of stress granules and it has been associated with activation of mTORC1 complex as well. In a certain way it may be involved in reverse the post-translational modifications of mono- or poly- ADP ribosylation that are carried out by several PARP proteins. See text for more details. The figure was designed using <https://biorender.com/>.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Peroxynitrite (ONOO^-) is formed after iNOS mediated production of nitric oxide (NO) and lead to mitochondrial damage and production of reactive oxygen species (ROS) and inhibits energy production, Fig. 34. Also, diffusion of ONOO^- into the nucleus induces DNA damage followed by PARP-1 activation. Overactivation of PARP-1 consumes NAD^+ in the process of poly-(ADP)-ribosylation. The recycling of the free nicotinamide released during this process is mediated by ATP and thus depletes the cellular energy, Fig. 34. The lack of cellular energy induced by the overactivation of PARP-1 also promotes relocation of apoptosis-inducing factor (AIF) from mitochondria to the nucleus and induces DNA fragmentation and cell death by necrosis [reviewed in (Hong *et al.*, 2004)](Ha & Snyder, 1999, Esposito & Cuzzocrea, 2009), Fig. 34. However, if caspase 3 is activated and cleaves PARP-1, it plays a role in induction of cell death through apoptosis (Halappanavar *et al.*, 1999). PARP1 is also processed during pyroptosis, a cell-death program that occurs upon activation of caspase-1 in inflammasome complexes (Malireddi *et al.*, 2010). PARP1 is also involved in a caspase independent apoptosis in which granzyme A targets PARP1 and compromises DNA damage recognition (Zhu *et al.*, 2009).

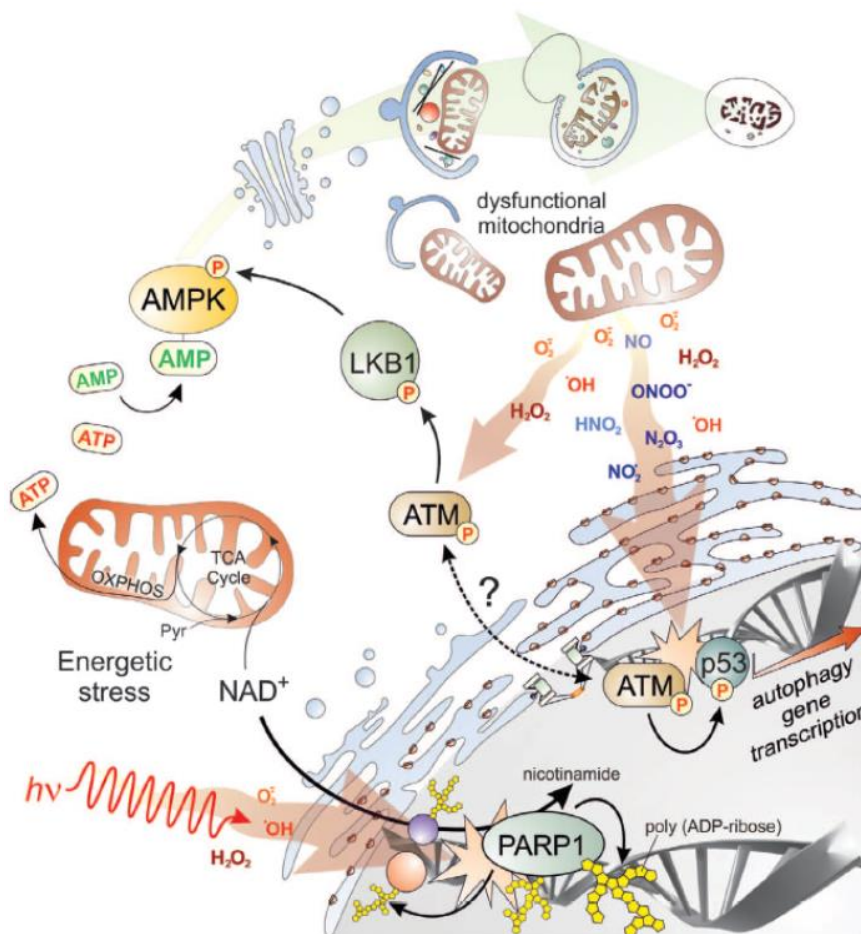


Figure 34. Implication of autophagy in DNA damage repair. Endogenous (e.g., dysfunctional mitochondria, top right) or exogenous (e.g., radiations or genotoxic stimuli, bottom left) sources of ROS and RNS induce DNA damage, whose primary sensors are PARP1 and ATM. Once activated by DNA breaks, PARP1 catalyses poly-ADP ribosylation of itself, as well as of other nuclear proteins, thereby leading to a

massive decrease of NAD⁺ and to a subsequent energetic stress. Upon DNA damage, ATM can activate p53-mediated transcription of autophagic genes. Alternatively, cytosolic pool of ATM could be directly activated by ROS through a still unidentified mechanism and it directly induces the activation of LKB1. The issue of whether cytosolic and nuclear pool of ATM are interconnected still waits to be demonstrated. Both PARP1 and ATM signalling pathways converge on AMPK, whose activation induces the autophagic machinery to remove the main source of DNA damage and contribute to its repair through a negative feedback loop. From (Filomeni *et al.*, 2015).

It has been reported that NO has an important role in protecting mice from fatal encephalitis caused by a neuro-adapted Sindbis strain. The immune response is unlikely involved in this mechanism of NO protection, instead, NO may enhance survival of the infected neurons until the immune response can control virus replication (Tucker *et al.*, 1996). Other works have also suggested that reactive nitrogen species play an important role in VEEV dissemination and survival of the host. Their observation showed that VEEV infection increased levels of NO in resting microglial cultures but VEEV decreased NO production in IFN- γ -stimulated microglia cells and it is known that INF- γ stimulations leads to increase of NO production in microglial cells (Htain *et al.*, 1997, Schoneboom *et al.*, 2000).

The transcription factor NF- κ B is also known to regulate the cyclooxygenase-2 (COX-2) that is involved in production of free radicals as well and prostaglandins that are derived from acid arachidonic and sustain the inflammatory response. It has been suggested that PARP1 binds to the mouse COX-2 promoter region through interaction with the inhibitory element resulting in inhibition of COX-2 expression (Lin *et al.*, 2011).

In spite of induction of DNA repair, PARP1 activation also activates mitotrophic pathways, such as the adenosine monophosphate (AMP)-activated protein kinase (AMPK) and the nuclear respiratory factor 1 (NRF), or autophagy that sustain cell survival, Fig. 33. The cellular process of autophagy recycles and degrade damaged cellular components in order to support cell survival. Autophagy depends on the activation of AMPK that inhibits a serine/threonine kinase complex, called mammalian target of rapamycin complex 1 (mTORC1). PARP1 and APMK activation mutually enhance each other's activity, and active AMPK inhibits mTORC1 through phosphorylation, facilitating autophagy response to DNA-damage-induced cell death (Munoz-Gamez *et al.*, 2009, Bai, 2015), Fig. 33. Interestingly, the HVR of SFV nsP3 is known to induce activation of the PI3K-Akt-mTOR pathway and replication complex internalization from the plasma membrane to form large intracellular CPV-I (Thaa *et al.*, 2015), Fig. 33 and 35.

Efferocytosis is a process that requires the high mobility group box 1 (HMGB1) protein post-translationally modified by PARylated to enhance the uptake of cellular debris (Davis *et al.*, 2012, Virag & Szabo, 2002). This avoid the interaction of debris with

alarmins that induce inflammation that primes local immune cells to produce chemotactic factors, free radicals and activates endothelial cells to facilitate leukocyte homing, diapedesis, and tissue infiltration. Indeed, free radicals can facilitate the movement of immune cells in tissues through activating matrix metalloproteinases (MMPs) that are redox-labile proteins (Bai, 2015).

PARPs other than PARP1 may also modulate the production of chemokines and cytokines, Fig. 33. Deletion of PARP-2 leads to alterations in the transcriptome, including in the expression of genes involved in inflammatory regulation (Bai & Virag, 2012). It has been suggested that exists likely an overlap with the inflammatory gene dysregulated (e.g. IL1 β and TNF- α levels) after knocking out PARP1 or PARP2 (Phulwani & Kielian, 2008, Popoff *et al.*, 2002). Other example is PARP14 that, when knocked out, enhances IL-4-induced STAT6-mediated gene expression that affects T_h2 mediated process.

PARP1, PARP5a (also known as thankyrase 1), PARP7, PARP10, and PARP12 are known to protect against viral infections through inhibiting viral recombination, replication, and translation as it was observed in VEEV (Atasheva *et al.*, 2014, Virag & Szabo, 2002). During VEEV infection the *PARP12* gene expression is up-regulated in a type I IFN dependent manner (Atasheva *et al.*, 2012), and the long isoform of PARP12, PARP10 and PARP14 inhibit replication of several alphaviruses and other RNA viruses (Atasheva *et al.*, 2012, Atasheva *et al.*, 2014). After overexpression of this PARPs (PARP12, PARP10 and PARP14), that are also known as mono-PARPs, the cellular translation is blocked with the finality to inhibit viral replication (Butepage *et al.*, 2015). Interestingly, PARP10 and PARP12 can localize to cytoplasmic structures associated with the autophagic membrane adaptor protein p62, and thus it may be involved in regulation of autophagy and probably modulates pathogen clearance via autophagy (Kleine *et al.*, 2012, Welsby *et al.*, 2014). Both PARP10 and PARP12 may target NF- κ B signalling pathway through MARylation of NF- κ B components leading to a role in modulating the immune response through this pathway (Welsby *et al.*, 2014, Feijs, Forst, *et al.*, 2013, Trocoli & Djavaheri-Mergny, 2011), Fig. 33.

It has been proposed that PARP10 works as a repressor of NF- κ B signalling (Verheugd *et al.*, 2013). PARP10 targets the NF- κ B essential modulator (NEMO), a member of the inhibitor of NF κ B (I κ B) kinase (IKK) complex, which is then MARylated and this leads to a decrease in its poly-ubiquitination, Fig. 33. Thereafter, stability of the subunit I κ B increases and then p65 subunit is less translocated into the nucleus (Verheugd *et al.*, 2013). That way, PARP10 subverts the effect caused by pro-inflammatory chemokines that induces ubiquitination of NEMO that is important for activation of the IKK complex

(Verheugd *et al.*, 2013). Thus, I κ B (an inhibitor of NF- κ B) is not phosphorylated neither degraded, and thus blocks NF- κ B transcription factor subunits, such as p65, to translocate into the nucleus and trigger gene expression (Chen, 2012), Fig. 33. Interestingly, PARP10 also inhibits VEEV replication, less efficiently than PARP12, though, thus suggesting its importance for the immune response (Atasheva *et al.*, 2012).

Recurrent conflicts of the host with pathogens requires also the involvement of other mono-PARPs such as PARP4, the *Macro* domain containing PARPs (e.g. PARP7, PARP8 and PARP9) and the catalytic inactive PARP13 (Butepage *et al.*, 2015). The role of PARP4 is still unknown but likely participates in the anti-viral response as it is supposed to inhibit the replication of certain alphaviruses and other RNA viruses (Daugherty *et al.*, 2014, Bick *et al.*, 2003, Gao *et al.*, 2002, Mao *et al.*, 2013, Muller *et al.*, 2007, Zhu *et al.*, 2011). The PARP13 exists in two isoforms and is also known as a zinc-finger anti-viral protein (ZAP). The short isoform lacks a C-terminal PARP catalytic domain, while the long isoform contains this domain but lacks its catalytic activity (Kleine *et al.*, 2008). PARP13 exerts its antiviral activity through its four Cys-Cys-Cys-His (CCCH)-type zinc finger domains, localized at the N-terminal domain, that binds viral mRNAs to recruit cellular RNA degradation machinery (Mao *et al.*, 2013, Zhu *et al.*, 2011, Jeong *et al.*, 2010, Guo *et al.*, 2007) leading to the prevention of viral RNAs translation and replication. However, the long isoform PARP13 may restrict alphavirus replication through its C-terminal PARP catalytic like domain (Glasker *et al.*, 2014). PARP13 being catalytic inactive, this means that ADP-ribosylation mediated by PARP13 is not involved in this anti-viral mechanism. It is still possible that PARP13 might interact with other PARP members in order to contribute to its cellular functions (Butepage *et al.*, 2015).

PARP16 was reported to localize to ER and plays key functions during signal ER stress. PARP16 participates in regulation of the unfolded protein response (UPR) and can lead to apoptosis (Jwa & Chang, 2012, Di Paola *et al.*, 2012, Hetz, 2012). Under ER stress conditions, PARP16 activity is upregulated and induces auto-MARylation and consequently MARylation of the stress sensors such as inositol requiring the kinase enzyme 1 alpha (IRE1 α) and the proline-rich, extensin-like receptor kinase-1 (PERK1) which is crucial for UPR (Jwa & Chang, 2012).

Interestingly, PARP13, PARP5a, PARP7, PARP8, PARP12 and PARG were reported to localize to SGs (Leung *et al.*, 2011), Fig. 35. SGs are known as dynamic ribonucleoprotein aggregations of translationally stalled mRNAs, ribosomal subunits and different RNA binding proteins (Buchan & Parker, 2009). The formation of SGs can occur after exposure to different forms of cellular stress including oxidative stress, heat shock

or nutrient starvation and regulate the stability or translation of mRNAs (Buchan & Parker, 2009). The overexpression of SGs-associated PARPs induces the assembly of SGs and several SG components suffer ADP-ribosylation (in *cis* and *trans* mono- and poly- ADP-ribosylation) in response to stress (e.g. argonaute [Ago], TIA-1, PARP5a, PARP7, PARP12, PARP13) (Leung *et al.*, 2011, Hottiger *et al.*, 2010) and PARG reverses these effects (Leung *et al.*, 2011). Beyond to different forms of cellular stress that induces SGs formation, SGs can also be triggered during infection of cells with certain viruses. It is likely that SGs can play an important role in the anti-viral defence because some viruses actively block SG assembly during infection (Beckham & Parker, 2008, Onomoto *et al.*, 2014). For example, CHIKV and VEEV can hijack G3BP and FXR, respectively, through its HVR region of nsP3 (Kim *et al.*, 2016, Schulte *et al.*, 2016, Scholte *et al.*, 2015, Fros *et al.*, 2012), Fig. 35. Moreover, the activity of antiviral PARPs (PARP7, PARP10 and PARP12) decreases with the rate of cellular translation and viral replication. Thus, suggesting that the production of viral particles is being neutralized (Atasheva *et al.*, 2014, Vyas *et al.*, 2014).

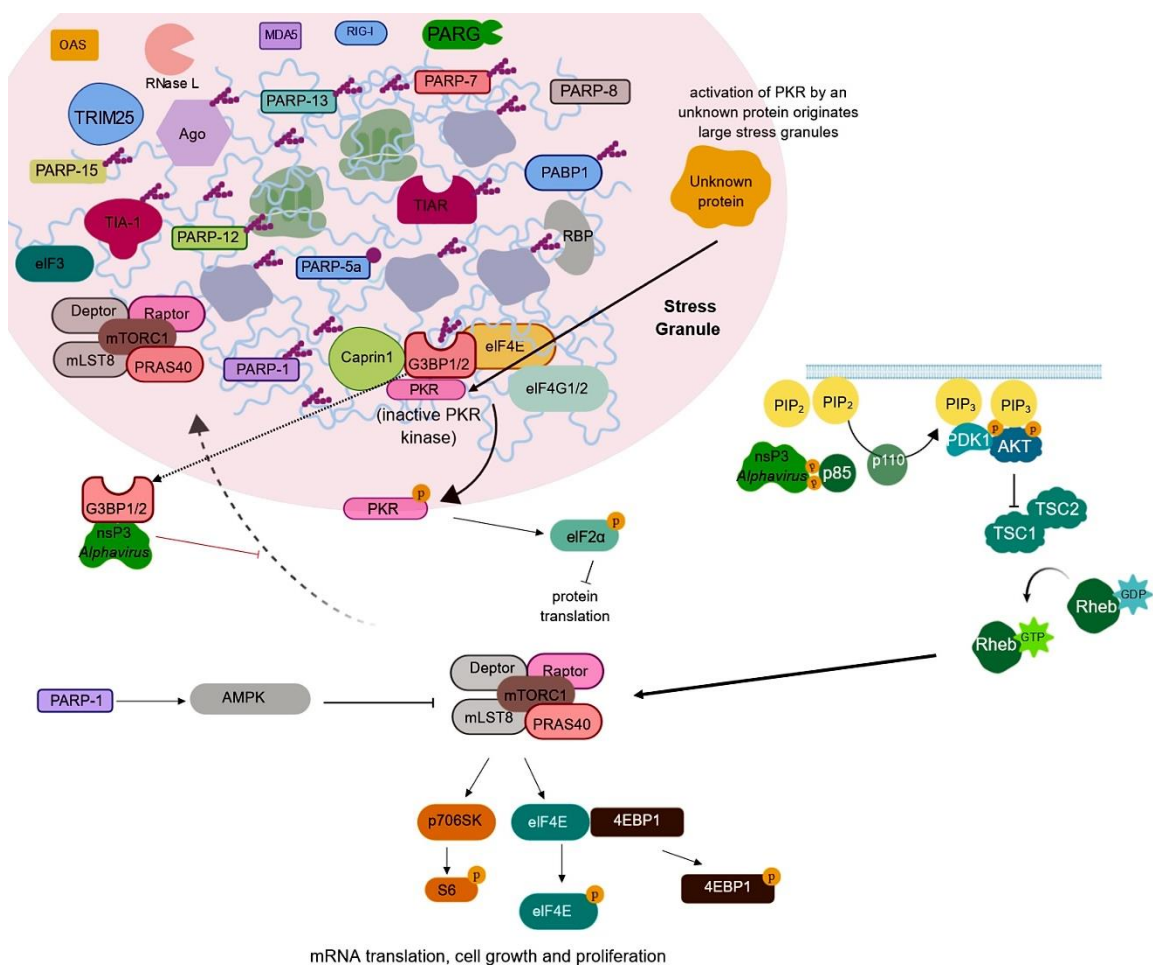


Figure 35. Stress granules formation and composition and the role of nsP3 in stress granules disassembly. Through its YXXM motif, nsP3 interacts with the SH2 domain from PI3K from the p85 subunit. The p85 subunit is the regulatory domain of the PI3K that activates the effector domain p110. Dissociation

of the p110 from p85 leads to its activation and converts the membrane bounded PI(4,5)P₂ to PI(3,4,5)P₃. PKD1 and AKT are recruited to the PM because they harbour a pleckstrin homology domain (PH domain) that binds to PI(3,4,5)P₃. The formation of PKD1/AKT complex leads to activation of AKT through phosphorylation Thr308 and Ser473. The mTORC2 can may exercise some influence in the AKT activation. In addition, the activated AKT triggers the inhibition of TSC1/2 complex which leads to activation of Rheb that converts GDP into GTP and leads to activation of the mTORC1 complex. The mTORC1 complex also induce phosphorylation of downstream targets to ensure efficient host cellular translation. Therefore, nsP3 may have a key role in inhibiting the translocation of inactive mTORC1 to the stress granules (Betz & Hall, 2013). Indeed, nsP3 have been associated with stress granules disassembly and to interact with G3BP1/2 or FXR proteins. Interestingly, under stress conditions G3BP1/2 is located inside stress granules in complex with caprin1, PKR, eIF4E and eIF4G1/2 and this complex ensure the inactive form of PKR. However, in certain conditions, an unknown host factor activates PKR and induces the formation of large stress granules. (Reineke *et al.*, 2015, Leung, 2014). The Figure was designed using <https://biorender.com/>.

Therefore, intracellular mono-PARPs and MARylation, as well as poly-PARPs and PARYlation, presents a key role in host immune defence to counterattack viral infections. In contrast, viruses might have developed strategies to escape host immune defence involving ADP-ribosylation. For this reason, several RNA viruses encode one or more *Macro* domains as part of non-structural proteins that are able of binding to ADP-ribose and PAR (Egloff *et al.*, 2006, Neuvonen & Ahola, 2009). Examples include viruses belonging to the *Togaviridae* (e.g. VEEV, CHIKV, SINV, rubella virus), *Hepeviridae* (HEV) and *Coronaviridae* (e.g. SARS-Coronavirus) families. For some of the viruses it was already demonstrated that their *Macro* domains can hydrolyse ADP-ribose-1''-phosphate (Egloff *et al.*, 2006, Saikatendu *et al.*, 2005). This suggests that viral *Macro* domains most likely can reverse ADP-ribosylation (Li *et al.*, 2016). It was already formally demonstrated, for the first time, that *Macro* domains reverse mono- and poly- ADP-ribosylation by Li *et al.* (Li *et al.*, 2016). Through this work it was reported that HEV *Macro* domain can catalyse the removal of MAR and PAR from modified proteins. Indeed, it was demonstrated that the presence of the HEV helicase in cis induced an increase in removal of PAR and is important for viral replication. Evidences that viral *Macro* domains activity is important to subvert the host immune response were demonstrated in cellular experiments: Mutations in viral *Macro* domains conferred reduced virulence in mice (Eriksson *et al.*, 2008, Park & Griffin, 2009) and mutation of SARS-CoV *Macro* domain increased sensitivity to IFN treatment (Fehr *et al.*, 2015, Kuri *et al.*, 2011, Fehr *et al.*, 2016). The corresponding mutation in HEV *Macro* domain resulted in reduced or no replication in liver cancer cell lines (Eriksson *et al.*, 2008, Li *et al.*, 2016, Parvez, 2015). Due to the importance of *Macro* domains, several authors have explored the identification of important residues for the enzymatic activity of *Macro* domains (Chen *et al.*, 2011, Jankevicius *et al.*, 2013, Rosenthal *et al.*, 2013, Egloff *et al.*, 2006, Malet *et al.*, 2009). For example, mutations within the catalytic loop negatively impact the enzymatic activity of *Macro* domains belonging to HEV and alphaviruses (Parvez, 2015, Abraham *et al.*, 2018, McPherson *et al.*, 2017) such as the mutant Gly32Glu in CHIKV.

In connection with the wide implication of PARPs and their ADP-ribosylated protein targets, one of the most challenging and exciting field of research remains to answer the following questions: which ADP-ribosylated cellular protein is de-ADP-ribosylated by viral macro domains, and which essential innate immunity pathway is thus cut?

7. Outline of this thesis

Alphaviruses are important (re-) emerging viruses with significant human pathogenesis and patients can present severe and long-lasting symptoms. Although patients infected by alphaviruses are treated with palliative treatments, no antivirals to fight alphaviruses are available. It is highly important to develop antiviral molecules to be used to cure human patients infected with alphaviruses.

The non-structural proteins which are involved in the formation of the replication complex have been proposed as excellent targets to eliminate alphaviruses. The work presented in this thesis focus in two non-structural proteins which are nsP1 and nsP3 which are presented in chapter 8 and chapter 9 respectively.

The nsP1 is an enzyme that confers to alphaviruses an original mechanism of capping. Through an unconventional pathway, alphaviruses can cap their mRNAs with a type 0 cap structure. Our group have been involved in the characterization of the mechanism of capping carried out by this enzyme and was pioneer in uncoupling the guanylation and methyltransferase activities. In the present work we developed an enzyme-based high throughput screening to find new molecules with potential antiviral activity by inhibiting the guanylation activity of nsP1. Therefore, chapter 8 is subdivided into 8.1 and 8.2. In the chapter 8.1 we present our work in developing the screening and the screening of a commercial library (Prestwick Chemical Library ®). Concerning the chapter 8.2 we present additional results obtained by testing analogues from MADTPs ([1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones) against the guanylation activity of nsP1 by using the same enzymatic assay presented in chapter 8.1.

Recently it has been discovered that the *Macro* domain from nsP3 is able to reverse the post-translational modification of mono- or/and poly- ADP- ribosylation. Our group was the first in determining the three-dimensional structure of *Alphavirus Macro* domains. Currently, we are interested in understanding the catalytic mechanism used by this *Macro* domains to reverse ADP-ribosylation. We believed that we could find more information about its catalytic mechanism by determining the structure of other related *Alphavirus Macro* domains. The work achieved by determining the three-dimensional structure of Getah virus *Macro* domain highlighted some mechanistic evidences of

scientific relevance. This work is presented in chapter 9 which is subdivided into chapter 9.1 where we present the crystallographic work and chapter 9.2 in which we present additional results showing the ability of Getah nsP3 in the formation of structures -like stress granules that were obtained by using cell-based assays.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

II- RESULTS

THIS PAGE IS INTENTIONALLY LEFT BLANK

8. Antivirals targeting the non-structural protein 1 (nsP1) from alphaviruses

THIS PAGE IS INTENTIONALLY LEFT BLANK

8.1 Approved drugs screening against the nsP1 capping enzyme of Venezuelan equine encephalitis virus using an immunobased assay

Article accepted in Antiviral Research Journal

THIS PAGE IS INTENTIONALLY LEFT BLANK

The viral mRNA capping is a possible target for the development of drugs with antiviral effect. Targeting the viral mRNA capping would abolish the translation of viral proteins and consequently compromise the viral replication. Due to the lack of structural information of the alphavirus nsP1, in this work we have designed and developed a strategy to select small molecules inhibiting one of the steps of the alphavirus capping. We have used VEEV nsP1 as model and among approved compounds we selected, using this method, small molecules with inhibitory potential of the guanylation activity of nsP1 enzyme.

For this work I have produced the recombinant non-structural protein 1 (nsP1) of VEEV using *Escherichia coli* and purified according to previously established protocols (Li *et al.*, 2015). The pre-established assay to monitor the enzymatic activity of guanylation for VEEV nsP1 was optimized for a high throughput ELISA screening. Together with Doctor Changqing Li, I have participated in carrying out the high throughput ELISA screening to screen the Prestwick Chemical Library® at one single concentration (50 μ M) of each compound.

The results obtained from the high throughput ELISA screening were analysed and 18 compounds were identified to inhibit the guanylation activity of nsP1 in approximately 80% or more. I have confirmed the result through determination of the IC_{50} value by WB assays in which was used the anti- m_3G/m^7G -cap monoclonal antibody. I have overcome the limitation of the WB method by increasing the throughput of the method by loading 4 titrations on a single SDS Page gel allowing the determination of the IC_{50} of 4 compounds in a single experiment. The best four hit compounds were selected from the screening and were representative of three series of compounds. The three series of compounds were selected for further analysis and pharmacophore evaluation. In this chapter (8.1) are presented the analysis of two series, the series 1 represented by the best compound which is Prest-37 (pyrimethamine) that shows an IC_{50} of 2.7 ± 0.4 μ M. The series 2 is represented by two other compounds selected from the screening, Prest-392 and Prest-531, both containing a 4-F-benzoylpiperidine and showing IC_{50} values in the 10 μ M range.

Depending on the IC_{50} and chemical nature of the compounds we decided to further investigate additional compounds. Analogue search and structural activity relationship (SAR) were initiated to identify the active pharmacophore features. The compounds were selected by using the scaffold of Prest-37 (series 1) and that of Prest-392/Prest-531 (series 2) for analogue search. I have performed the analogue search at the Prestwick Chemical libraries and in other commercially available databases such as Reaxys and

eMolecules. Like the analogue search, I have also used the 4 initial hit compounds for chemical group validation. Some of these chemical groups and close analogues were commercially available and purchased. Then, I have tested the compounds against the guanylylation activity of nsP1.

Lastly, to complement the MTase activity assays performed by Doctor Etienne Decroly on human N7-MTase and VEEV nsP1 and to gain insight into the mechanism of inhibition, I have tested the inhibition of the head of series 1 and 2 on the VEEV D34S nsP1. Other developed compounds, among which are the 3-aryl-[1,2,3]triazolo[4,5-d]pyrimidin-7[6H]-ones, also known as MADTPs, were previously demonstrated to affect both GT and MTase activity of VEEV nsP1, and VEEV D34S nsP1 was reluctant to MADTP inhibition (Delang et al., 2016; Li et al., 2015). The results obtained from the tested head of series 1 and 2 on VEEV D34S nsP1 were expected to shedding light about a possible common mode of action despite structural differences. Both Prest-37 and 531 showed comparable inhibitory effect on both VEEV D34S and wt nsP1, suggesting that their mechanism of action is likely different from that of the MADTP which targets the N-terminal region of nsP1.

Altogether the results show that this HT enzyme-based assay is a convenient way to select potent and specific hit compounds targeting the viral mRNA capping of Alphaviruses.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK



Contents lists available at ScienceDirect

Antiviral Research

journal homepage: www.elsevier.com/locate/antiviral

Approved drugs screening against the nsP1 capping enzyme of Venezuelan equine encephalitis virus using an immuno-based assay



Ana S. Ferreira-Ramos^a, Changqing Li^a, Cécilia Eydoux^a, Jean Marie Contreras^b, Christophe Morice^b, Gilles Quérat^c, Alba Gigante^d, María-Jesús Pérez Pérez^d, Marie-Louise Jung^b, Bruno Canard^a, Jean-Claude Guillemot^a, Etienne Decroly^{a,1}, Bruno Coutard^{a,c,e,1}

^a Aix Marseille Université, CNRS, ARMB UMR 7257, Marseille, France

^b Prestwick Chemical, ILLKIRCH, Strasbourg, France

^c Unité des Virus Emergents (UVE: Aix-Marseille Univ-IRD 190-Inserm 1207-IHU Méditerranée Infection), Marseille, France

^d Instituto de Química Médica (IQM, CSIC), Madrid, Spain

ARTICLE INFO

Keywords:
Alpha virus
HT screening
Guanylylation
mRNA capping
Approved drugs
Antivirals

ABSTRACT

Alphaviruses such as the Venezuelan equine encephalitis virus (VEEV) are important human emerging pathogens transmitted by mosquitoes. They possess a unique viral mRNA capping mechanism catalyzed by the viral non-structural protein nsP1, which is essential for virus replication. The alphaviruses capping starts by the methylation of a GTP molecule by the N7-guanine methyltransferase (MTase) activity; nsP1 then forms a covalent link with m⁷GMP releasing pyrophosphate (GT reaction) and the m⁷GMP is next transferred onto the 5'-diphosphate end of the viral mRNA to form a cap-0 structure. The cap-0 structure decreases the detection of foreign viral RNAs, prevents RNA degradation by cellular exonucleases, and promotes viral RNA translation into proteins. Additionally, reverse-genetic studies have demonstrated that viruses mutated in nsP1 catalytic residues are both impaired towards replication and attenuated. The nsP1 protein is thus considered an attractive antiviral target for drug discovery. We have previously demonstrated that the guanylylation of VEEV nsP1 can be monitored by Western blot analysis using an antibody recognizing the cap structure. In this study, we developed a high throughput ELISA screening assay to monitor the GT reaction through m⁷GMP-nsP1 adduct quantitation. This assay was validated using known nsP1 inhibitors before screening 1220 approved compounds. 18 compounds inhibiting the nsP1 guanylylation were identified, and their IC₅₀ determined. Compounds from two series were further characterized and shown to inhibit the nsP1 MTase activity. Conversely, these compounds barely inhibited a cellular MTase demonstrating their specificity towards nsP1. Analogues search and SAR were also initiated to identify the active pharmacophore features. Altogether the results show that this HT enzyme-based assay is a convenient way to select potent and specific hit compounds targeting the viral mRNA capping of Alphaviruses.

1. Introduction

Emergence or re-emergence of alphaviruses represents a serious health concern, as exemplified by the worldwide epidemics of Chikungunya virus in recent years (Charrel et al., 2014). Alphavirus genus comprises 31 species among which at least ten of the arthropod-borne alphaviruses are important in terms of public health (Gould et al., 2010). Alphaviruses can be organized into two major groups based on their geographical distribution: Alphaviruses from the Old World (OW), most commonly causing febrile illness and painful arthralgia or

polyarthralgia and New World (NW) alphaviruses mostly leading to neurological diseases. Among NW alphaviruses, Venezuelan equine encephalitis virus (VEEV) is an important pathogen present in the Americas from Texas to Argentina (Adams et al., 2012; Pisano et al., 2013). In 1995, an outbreak in Venezuela and Colombia resulted in roughly 100,000 human cases including more than 300 fatal encephalitis cases (Rivas et al., 1997). Other epidemics were also reported indicating that VEEV represents a real public health concern. Moreover, VEEV infection symptoms resemble those of Dengue fever, likely leading to an underestimation of the number of cases related to this

* Corresponding author. UVE: Aix-Marseille Univ-IRD 190-Inserm 1207-IHU Méditerranée Infection, Marseille, France.
E-mail address: bruno.coutard@univ-amu.fr (B. Coutard).

¹ The authors equally contributed to the work.

<https://doi.org/10.1016/j.antiviral.2019.01.003>

Received 22 October 2018; Received in revised form 28 November 2018; Accepted 9 January 2019

Available online 11 January 2019

0166-3542/ © 2019 Elsevier B.V. All rights reserved.

virus infection in regions where both Dengue virus and VEEV are endemic (Aguilar et al., 2011). In humans, whilst the overall mortality rate is low (< 1%), neurological disease, including disorientation, ataxia, mental depression, and convulsions can be detected in up to 14% of infected individuals, especially in children (Johnson and Martin, 1974).

Alphaviruses belong to the positive stranded RNA viruses replicating their genome in the cytoplasm of host cells. The genome contains two open reading frames (ORFs), encoding for the non-structural proteins (nsPs) and the structural proteins, which are expressed later from the sub-genomic RNA. Following viral entry, the genomic RNA is translated into the non-structural polyprotein P123 and P1234 which are processed into non-structural proteins (nsP1 to nsP4). The non-structural polyproteins and proteins form a complex associated to the plasma membranes, and this complex is embedding the enzymatic activities necessary for the viral replication and transcription. Namely, the nsP4 carries the RNA dependent RNA polymerase activity and drives antigenome synthesis and its subsequent transcription into genomic and sub-genomic mRNAs. The 5' ends of the genomic and sub-genomic RNAs are decorated by a cap-0 structure (m⁷G-pppN-RNA) and RNA are polyadenylated at their 3' extremity. The capping of viral mRNA is a key step for virus replication as cap structures protect viral RNA from cellular 5'-3' exonucleases and promote the initiation of RNA translation into viral proteins. In addition, the cap structure, together with conserved hairpin structure present in the 5'UTR, also hides viral RNA from detection by sensors of the innate immunity, such as RIG-I and/or MDA5 or by antiviral restriction factors (Hyde et al., 2014). The cap structure of alphavirus mRNA is thought to be synthesized by an unconventional capping mechanism involving the methyltransferase/guanylyltransferase activities of nsP1 and the 5' RNA triphosphatase activity of nsP2 (Li et al., 2015; Vasiljeva et al., 2000). Reverse genetic experiments have confirmed that these activities are essential for the viral life cycle. Indeed, the mutation of either MTase or nsP1 guanylylation catalytic residues strongly impairs viral replication (Wang et al., 1996). Capping enzymes are thus considered as promising antiviral drug targets.

Despite important efforts, antiviral drugs against alphaviruses such as VEEV are still lacking and to date, the treatment of VEEV infections is mainly supportive. For instance, the broad spectrum antiviral ribavirin, that was found to be efficient against Chikungunya virus (CHIKV) in resolving joint and soft tissue swelling (Ravichandran and Manian, 2008), has almost no significant activity against VEEV in infected cell cultures (Markdand et al., 2000). Therefore, efforts to discover molecules inhibiting alphaviruses such as VEEV have increased in the last years. Small molecule library screenings were recently developed on cell based assays to discover compounds active against VEEV or other alphaviruses (Chung et al., 2013) (Pohjala et al., 2011; Spurgers et al., 2013) (Delektka et al., 2015; Lucas-Hourani et al., 2014; Seyedi et al., 2016; Varghese et al., 2016). Several compounds were found to be associated to the inhibition of functions carried out by the proteins constituting the viral replication complex. Among the selected compounds, the broad-spectrum antiviral candidate Favipiravir (T-705), initially developed to treat human influenza, shows a potent antiviral effect in small animal models (Furuta et al., 2013). This nucleobase mimetic was proposed to target the polymerase activity of nsP4 after intracellular activation into its 5'-triphosphate ribonucleotide form (Delang et al., 2014). In addition, nucleoside or non-nucleoside analogues have revealed antiviral effects supposedly by targeting nsP2 or nsP4 (Chung et al., 2013, 2014; Urakova et al., 2017; Wada et al., 2017). nsP1 is also considered as an antiviral target: the recent identification of compounds inhibiting the mRNA nsP1 mediated-capping (Delang et al., 2016; Feibelman et al., 2018; Gigante et al., 2014; Gomez-SanJuan et al., 2018) sheds light on this viral enzyme. Screening strategies based on the inhibition of alphavirus nsP1 enzymatic activities can thus be considered for drug discovery.

The enzymatic activities of VEEV nsP1 involved in the formation of

the viral mRNA cap-0 structure were previously characterized (Li et al., 2015). nsP1 first methylates a GTP molecule into its guanine N7 position (MTase activity). The methylation reaction requires the S-adenosylmethionine (SAM) as methyl donor and generates S-adenosylhomocysteine (SAH) as by-product. The m⁷GTP product is next used for nsP1 guanylylation (GT reaction) leading to the formation of an m⁷GMP-nsP1 covalent adduct. m⁷GMP forms a covalent link with a catalytic histidine of VEEV nsP1. The formation of guanylylated-nsP1 adduct is regulated by the SAM/SAH balance, and the release of SAH by the MTase reaction stimulates the GT reaction. Finally, the m⁷GMP molecule is transferred onto the 5'-diphosphate RNA end to yield the capped RNA (GTase reaction). In order to characterize nsP1 functions as well as compounds inhibiting nsP1 we developed a Western blot assay to detect and quantify the guanylylated nsP1. However, this labor intensive assay is not suitable for high throughput screening of guanylylation inhibitors.

In this work, we developed an assay in order to uncouple the capping reactions carried by VEEV nsP1 and to specifically quantify the synthesis of m⁷GMP-nsP1 adduct using ELISA. After validation of the assay using known nsP1 inhibitors, we screened a library containing 1220 approved drugs. The best compounds were selected at 50 μM and further characterized. We determined their IC₅₀ on both the nsP1-mediated MTase and GT reactions. In addition, the specificity of the inhibitors was assessed comparatively using a cellular MTase and viral MTase involved in capping process. Compounds analogues and/or fragments available from commercial libraries were also tested in order to better understand the structure-activity relationships of molecules inhibiting nsP1 activities.

2. Material and methods

Chemicals. Chemicals such as GTP, m⁷GTP, S-adenosylmethionine (SAM) and S-adenosylhomocysteine (SAH) were purchased from Sigma Aldrich. The Prestwick Chemical Library (PCL) is a library of 1280 small molecules, mostly approved drugs (FDA, EMA and other agencies) from which 1220 compounds were picked for the screening (Strasbourg, France). Other compounds used in this study are commercially available and were purchased from Toronto Research Chemicals (TRC), Key Organics, Enamine, Chembridge, Combi Blocks, Enzo Life Science, Sigma Aldrich, Tocris and Alfa Aesar. The synthesis of compounds MADTP-314 (3-(3-acetylphenyl)-5-methyl-3,6-dihydro-7H-[1,2,3]triazolo[4,5-d]pyrimidin-7-one) and MADTP-393 (3-(3-acetylphenyl)-5-ethyl-3,6-dihydro-7H-[1,2,3]triazolo[4,5-d]pyrimidin-7-one) has already been described (Gigante et al., 2014, 2017). The FDA-approved compounds and the compounds purchased from other suppliers were dissolved in DMSO at a stock concentration of 20 mM.

2.1. Expression and purification of VEEV nsP1 wild-type (wt) and mutant D34S

The codon optimized DNA encoding nsP1 of VEEV (strain P676, amino acid 1 to 535) was cloned into the expression vector pET28b (Novagen) to enable the fusion of a 6xHis tag at the C-terminus of nsP1. The protein was produced in *E. coli* Rosetta pLysS (DE3) (Novagen) cells and purified by IMAC chromatography on a 5 ml His TrapFF crude column (GE Healthcare). The proteins eluted 250 mM imidazole, 20 mM Tris pH 7.5, 100 mM NaCl and 5% glycerol were dialysed in a dialysis buffer with the following composition; 20 mM Tris pH 7.5, 100 mM NaCl and 10% Glycerol. The protein purity was assessed on Coomassie blue stained SDS-PAGE gel and protein concentration was determined by Bradford protein assay (Bio Rad). The same procedure was performed for D34S nsP1.

2.2. Methyltransferase filter binding assay

The methyltransferase assays were carried out in a 20 μL reaction

mixture, containing 50 mM Tris (pH 7.0), 2 mM DTT, 10 mM KCl, 2 mM GDP, 330 nM S-adenosyl [methyl-³H] Methionine (83.1 Ci/mmol, PerkinElmer), 10 μM S-adenosylmethionine, 2 μM VEEV nsP1 and increasing concentrations of inhibitors with 5% final concentration of DMSO. After incubation at 30 °C for 1 h, the reaction samples were loaded on DEAE-cellulose filter (PerkinElmer), and the filter was washed twice with 20 mM ammonium formate, once with H₂O, and once with absolute ethanol. The filter was dried, and the radioactivity was measured by scintillation counting with SCINT BETAPLATE solution in a Wallac MicroBeta Trilux 1450 counter (PerkinElmer). Compounds were tested in triplicate at a final concentration of 50 μM for single point analysis.

2.3. Guanylation assay and detection of the m⁷GMP-nsP1 adduct by ELISA

The guanylation reaction (GT) of nsP1 VEEV was performed in 20 μL as previously described (Li et al., 2015). Briefly, 5 μM of nsP1 was incubated at 30 °C in a buffer containing 20 mM HEPES (pH 7.0), 2 mM MgCl₂, 1 mM DTT, 100 μM m⁷GTP, 100 μM S-Adenosylhomocysteine (SAH). Time course experiments using 1 μM, 2.5 μM and 5 μM of nsP1 were performed to optimize the incubation time and enzyme concentration for the ELISA assay. The screening of GT inhibitors was performed in a buffer containing 100 μM m⁷GTP, 100 μM SAH, 5 μM nsP1, and the reaction were incubated during 30 min at 30 °C. Reactions with DMSO but without compound were used as negative control (NG) of inhibition and, as background control (BG) were used reactions without compound and m⁷GTP. For the validation of the assay, 5 μM of enzyme and 1 h incubation of the reaction at 30 °C were selected, speculating that a high signal at OD_{430nm} and a complete reaction would correspond to discriminating and stringent conditions suitable for the selection of the most potent inhibitors.

The ELISA detection of nsP1-m⁷GMP adducts was performed as follows: 8 μL of the GT reaction were transferred into 92 μL of PBS (Phosphate buffered saline, Sigma #P3813) in Nunc MaxiSorp flat-bottom 96 well plate (Ebioscience). The plates were incubated overnight at 4 °C to allow nsP1 capture before plates washing (twice with PBS), and blocking with 2% (w/v) of non-fat milk in PBS for 2 h at 37 °C. The plate was then washed once with PBS, before incubation (1 h at RT) with 100 μL of primary antibody anti-m⁷G-cap monoclonal antibody (Synaptic Systems, Göttingen, Germany) at a dilution of 1:1000 in PBS with 1% (w/v) of non-fat milk. The wells were washed three times with PBS-Tween, and three times with PBS solution before incubation with the peroxidase-conjugated rabbit anti-mouse antibody (Sigma, #A9044) diluted at 1:2000 in PBS with 1% (w/v) of non-fat milk at RT for 1 h. The plate was then washed sequentially three times with PBS-Tween and three times with PBS. Finally, 3,3',5,5'-Tetramethylbenzidine (TMB) solution was prepared by dissolving one 3,3',5,5'-Tetramethylbenzidine tablet (Sigma T5525) in 100 μL of DMSO, which was then added into 9.9 mL of 0.1 M Sodium-acetate buffer, pH 6.0 with 2 μL of fresh 30% hydrogen peroxide. For each well, 100 μL TMB solution were added followed by incubate at 20 min at RT, then 50 μL of 2 M H₂SO₄ were added to stop the reaction, and then the absorbance of each well was read at 430 nm. Each 96-wells reaction and ELISA plate was used to test 80 compounds. The plates also contained two negative controls of the inhibition (reactions with no compound) and two controls for the background of the signal (reactions with no compound and no m⁷GTP). For each plate, the controls were used to define the following values: OD_{NG} represents the mean value of the OD_{430nm} measured for the duplicates for the negative control which is representative of 100% GT activity (reaction without compounds); OD_{BG} represents the mean value of the OD_{430nm} measured for the duplicates for the background control which is representative of 0% GT activity (reaction without compounds and without m⁷GTP).

The percentage of inhibition was calculated according to the following equation (1).

$$Y (\%) = \frac{OD_{NG} - OD_{[sample]}}{OD_{NG} - OD_{BG}} \times 100 \quad (1)$$

Where, Y (%) represents the inhibition percentage of GT activity mediated by nsP1, which is dose dependent of inhibitor concentration used; OD_{sample} represents the OD_{430nm} measured from the sample.

2.4. Hit validation

The hit compounds, showing more than 80% of inhibition at 50 μM were selected for hit validation. The hit validation consisted in the re-determination of the percentage of inhibition in triplicates at 50 μM using fresh powders of compounds and in the determination of the IC₅₀ value for the compounds showing more than 80% of inhibition by Western blot (WB).

The IC₅₀ of the inhibitors was assessed by WB, starting from 500 μM or 100 μM concentration followed by nine 1:2 serial dilutions. For each inhibitor, a negative control for GT (without both m⁷GTP and inhibitor) and positive control for GT (without inhibitor) were included. Reactions were prepared in a final volume of 20 μL, containing 20 mM HEPES pH 7.0, 2 mM MgCl₂, 1 mM DTT, 100 μM SAH, 100 μM m⁷GTP, 2% DMSO and 5 μM of nsP1 enzyme and were incubated at 30 °C for 45 min. The reactions were stopped by adding 2.5 μL of 4x Laemmli Buffer followed by heat at 95 °C for 5 min prior to the loading in 8% (SDS-PAGE) acrylamide Tris-Glycine gels. Reactions were performed in independent duplicates. In order to improve the throughput of the detection by WB, the products of titration of four compounds were loaded on a single 8% acrylamide SDS PAGE gel. 7 μL of each reaction from a same titration were loaded on a gel prior 14 min of electrophoresis at 200 V. Negative and positive controls were loaded on the same gel with each titration for an appropriate quantification. The gel was then re-loaded with second titration. The procedure was repeated for the third and fourth titrations. After the last loading the electrophoresis was performed for 30 min at 200 V. The Western blotting was done using the already described procedure (Li et al., 2015). After the signal acquisition of the WB, the PVDF membranes were stained during 3 min with Coomassie blue and the destained with 50% Ethanol (EtOH) during 5 min.

For each PVDF membrane, a BIP file was acquired with Kodak system, the file was converted in a TIF format that was used for analysis and quantification of the intensity (I) of the bands using ImageJ. After the quantification, the raw data was used to determine the percentage of activity for each compound, using the negative control for activity (NA) and the positive control for activity (MA) as references. The percentage of activity for each inhibitor concentration was determined using equation (2).

$$Y (\%) = \left[\frac{(I_{sample} - I_{NA})}{I_{MA} - I_{NA}} \right] \times 100 \quad (2)$$

Where, Y (%) represents the percentage of remaining guanylation activity mediated by nsP1, which is dose dependent of inhibitor concentration used; I_{sample} is the intensity calculated from the band, of the sample, obtained through the signal acquisition of the WB, I_{NA} represents the average of the duplicates of the intensity calculated for the control representative of 0% guanylation activity (reaction without compounds and without m⁷GTP), I_{MA} represents the average of the duplicates of the intensity calculated for the control representative of 100% guanylation activity (reaction without compound but with m⁷GTP).

IC₅₀ values of selected inhibitors were determined with GraphPad Prism 6, using a non-linear regression dose response inhibition.

2.5. Analogues search and chemical group validation

The analogues search was first achieved through the internal drug approved library (The Prestwick Chemical Library). The second

analogue search was done using the commercial databases Reaxys and eMolecule. In a similar manner to the analogue search, we used the 4 initial hit compounds and from each compound we selected different chemical groups, which were used to search for close analogues at commercial databases. Some of these chemical groups and close analogues were commercially available and purchased for test against the guanylation activity of nsP1.

3. Results

3.1. Monitoring the nsP1 guanylation by ELISA

Alphavirus mRNA are capped by an unconventional mechanism in which nsP1 catalyses first methylation onto the N7 of GTP (MTase) followed by the formation of m⁷GMP-nsP1 adduct (GT). Both MTase and GT reactions precede the transfer of m⁷GMP onto the 5'-end of the ppRNA. In order to identify small molecules inhibiting specifically the GT activity of nsP1, we developed an assay detecting the synthesis of m⁷GMP-nsP1 adduct. The GT reaction assay was performed with purified recombinant VEEV nsP1 expressed in *E. coli*. Since the purified nsP1 protein carries N7 MTase and GT activities, we tried to uncouple the two reactions in order to specifically monitor the GT activity. We thus used m⁷GTP (100 μM) instead of GTP as substrate and the GT reactions were performed in an optimized buffer supplemented with 2 mM MgCl₂ and 100 μM of SAH as both MgCl₂ and SAH have been shown to stimulate the nsP1 mediated-GT activity (Li et al., 2015).

After the GT reaction, the guanylated-nsP1 was detected by ELISA assay. For this purpose, nsP1 was captured on an ELISA plate and the m⁷GMP linked to nsP1 was next detected by the successive addition of an anti-m³G/m⁷G-cap antibody and peroxidase-conjugated rabbit anti-mouse antibody. Fig. 1A shows that using this assay, the m⁷GMP-nsP1 covalent complex resulting from the GT reaction could be detected in presence of m⁷GTP, SAH and nsP1. Conversely, no signal is detected in absence of SAH, m⁷GTP or nsP1, demonstrating the specificity of the ELISA detection system. We next determined the optimal experimental conditions for the screening. The time course experiment presented in Fig. 1B indicates that when using 5 μM of VEEV nsP1 the saturation phase is reached after 30–40 min of reaction whereas at lower nsP1 concentration (i.e. 1 and 2.5 μM) the reaction remains in a linear phase during the 30 min incubation period, but the ELISA signal is lower. Altogether these results indicate that the ELISA detection system provides a suitable tool to follow specifically the VEEV nsP1 mediated-GT reaction in a 96-well format.

The robustness of the GT assay was further assessed by testing already described VEEV nsP1 inhibitors for validation. In these control reactions, we used Sinefungin, a SAM analogue, which is known to inhibit viral and cellular MTases, as well as MADTP compounds, previously demonstrated to inhibit VEEV nsP1 mediated GT reaction

(Delang et al., 2016; Gigante et al., 2017). The inhibitory effect of these compounds was determined at 50 μM and the remaining GTase activity was normalized with the activity detected in absence of inhibitors. Sinefungin inhibits 47.9 ± 10.9% of the GT activity whereas MADTP-393 and MADTP-314 showed 99.2 ± 0.6% and 81.4 ± 2.8% of inhibition on the GT activity, respectively. The inhibition induced by MADTP-393, MADTP-314 and Sinefungin were further confirmed by determination of their IC₅₀. The GT reaction of VEEV nsP1 was performed in the presence of increasing concentration of inhibitor as previously described, and the amount of the m⁷GMP-nsP1 complex produced was quantified by WB, using the anti m³G/m⁷G-cap antibody for detection. Fig. S1 shows that, after WB quantification and Hill plot curve fitting, Sinefungin, MADTP-393, and MADTP-314, inhibit the GT activity of VEEV nsP1 in a dose-dependent manner with IC₅₀ of 96 μM, 69 μM and 177 μM, respectively. Of note, the IC₅₀ determined for Sinefungin in the current study and the previous one differs (29 and 96 μM, respectively) likely because the reaction conditions slightly differ from the original assay (Li et al., 2015). Altogether these results indicate a correlation between inhibitions observed using 50 μM of inhibitors by ELISA assay, and IC₅₀ determined by WB. The latter method was validated in a previous study (Li et al., 2015), confirming that the developed ELISA assay can be used to screen small molecules inhibiting specifically the VEEV nsP1 GT activity.

3.2. Screening of 1220 approved compounds inhibiting GT by ELISA

We next used the ELISA nsP1 GT assay to screen a library containing 1220 compounds from the Prestwick Chemical Library (PCL), a library containing approved drugs. The compounds were incubated with VEEV nsP1 at a final concentration of 50 μM and the GT reaction was started by addition of m⁷GTP. After an incubation period of 30 min, the m⁷GMP-nsP1 adduct was quantitated by ELISA assay and the % of inhibition of each compound was determined (Fig. 2A). The percentage of inhibition follows a right skewed bell-shaped distribution close to the normal distribution (Fig. 2B) with a mean inhibition “σ” of 16.5% close to the median (19.0%) ≈ Mode (20%) and a standard deviation (σ) of 26.8%. 229 compounds present a negative value, most of them yielding to a weak increase of GT activity. The compounds showing more than 40% activation were tested twice in order to determine if some compounds could efficiently stimulate nsP1 GT activity, but none of them showed reproducible activation of the enzyme. We selected 18 compounds presenting at least ~80% of inhibition among which 6 were considered as “strong inhibitors” as they showed more than 3σ above the average (> 96.9% of inhibition) (Table 1).

3.3. Analysis of the screen and hits confirmation

We confirmed the initial screen results by determining the IC₅₀

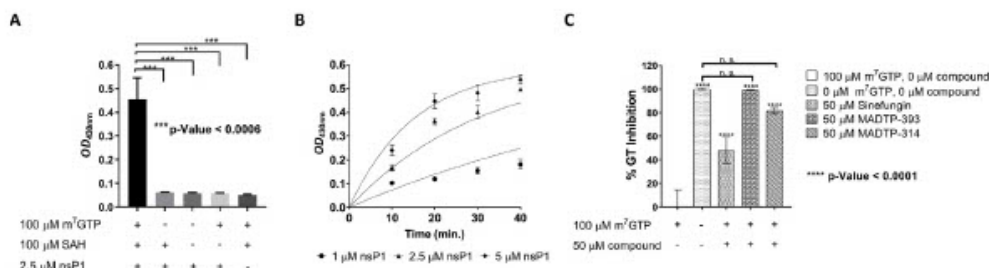


Fig. 1. Setup, optimization and validation of the ELISA monitoring the GT activity of VEEV nsP1. Panel A: Setup of the assay. The m⁷GTP-nsP1 complex is formed with 2.5 μM of VEEV nsP1 in the presence of 100 μM m⁷GTP, and 100 μM SAH. Panel B: Time course experiment performed with intervals of 10 min and the last time point was collected at 40 min. Panel C: percentage of GT inhibition for selected compounds used as reference and tested at 50 μM, namely Sinefungin, MADTP-393 and MADTP-314.

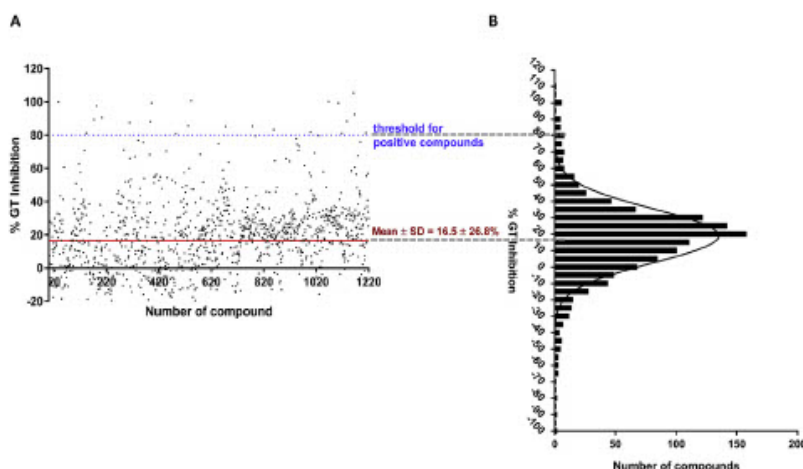


Fig. 2. Screening of approved drugs from Prestwick Chemical Library and hit selection. Panel A: Graph representing the GT inhibition of each tested compound. The average GT inhibition is presented in red and in blue the threshold (80% of GT inhibition) for the hit selection. Panel B: frequency histogram of the GT inhibition with a bin range of 10 and a Gaussian distribution curve fitting performed using GraphPad.

value using the anti- m^7G -cap WB assay on fresh solutions from commercially available powders. The WB assay is more sensitive than the ELISA assay and presents the advantage to detect unambiguously the formation of m^7GMP -nsP1 complex with a low background. The limitation of this assay being the throughput, it was improved by serial loading of 4 titrations on a single SDS Page gel allowing the determination of IC_{50} of 4 compounds in a single experiment (Fig. 3). The IC_{50} value was determined for the selected compounds, except Prest-998 that was not available. 7 compounds display IC_{50} values below $50\mu M$ and 2 show IC_{50} value between 50 and $110\mu M$ (Table 1). The best compound corresponds to Prest-37 (pyrimethamine) and shows an IC_{50} of $2.7 \pm 0.4\mu M$. Interestingly, two other compounds, Prest-392 and Prest-531, both containing a 4-F-benzoylpiperidine, show IC_{50} values in the $10\mu M$ range, highlighting the accuracy of the assay to detect potent inhibitors. Based on the IC_{50} and chemical nature of the compounds we arbitrarily decided to further investigate additional compounds based on the scaffold of Prest-37 (series 1) and that of Prest-392/Prest-531 (series 2).

We first searched for analogues already present in the screened library (PCL). Three compounds having structural and chemical similarities with Prest-37 were identified and their IC_{50} for the GT reaction was determined. The 3 compounds (Prest-347, 858, and 947) share with Prest-37 an aminopyrimidine included in a fused bicyclic ring, and were demonstrated to inhibit the GT reaction less efficiently than the initial hit (Table 2), though. Interestingly, one of these compounds (Prest-858) was picked in the initial screen whereas the two others showed an inhibitory effect at $50\mu M$, but below the 80% threshold. Nine analogues of Prest-392, bearing a piperidine linked to an aromatic moiety through a flexible chain, were selected in the PCL to confirm the hit lead (Table 3). All of them inhibit more than 50% of the nsP1 GT activity at $50\mu M$, highlighting the robustness of the ELISA screening and suggesting the possible presence of active common pharmacophore (s) in the two series. IC_{50} determined by WB showed that none of the compound has a higher potency than the original hit to inhibit nsP1 GT activity. Table 3 also shows that 3 compounds display a weak inhibition effect on the GT reaction (Prest-1117, 115, 1029), suggesting that the detection of m^7GMP -nsP1 adduct by ELISA might yield false positive results.

As the GT and the MTase reaction involve the same protein domain, we next tested if the selected GT inhibitors also inhibit the VEEV nsP1 MTase activity. The selected compounds were thus incubated with VEEV nsP1 in presence of radio-labelled SAM and the $[3H]$ methyl transfer on the non-hydrolysable GTP (GIDP) was measured by DEAE

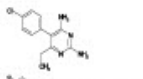

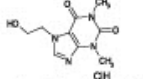
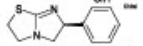
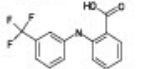
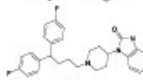
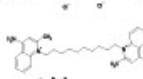
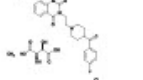
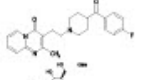
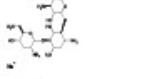
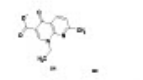
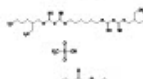
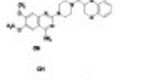
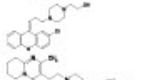
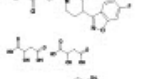
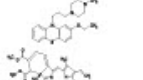
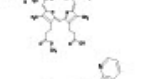
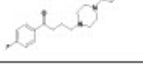
filter binding assay (FBA) (Tables 2 and 3). In this experimental condition uncoupling the MTase activity from the GT, the two heads of series, Prest-37 and Prest-392, display the best inhibitory effect on the MTase activity with $73.5 \pm 2.86\%$ and $79.3 \pm 1.7\%$ of inhibition at $50\mu M$, respectively. This analysis highlights the robustness of the inhibition observed for the two compounds and confirms the interplay between nsP1 mediated GT and MTase activities.

3.4. Hit/series validation beyond the screened library and pharmacophore evaluation of Prest-392

We next looked for analogues of Prest-37 and Prest-392 available in commercial libraries and their IC_{50} for the GT reaction was determined by WB. 5 analogues of Prest-37 were purchased. None of them showed higher or equal inhibitory effect to limit the GT reaction (Table 2). However, the comparative analysis of the inhibitors suggests that the alkyl group (ethyl or methyl) in the 6 position of diaminopyrimidine ring may play a key role in the GT inhibition (Prest 37, E925220 and M338835 vs EN300-183279 and EN300-208764). Indeed, when the alkyl group is missing, the analogues are inactive. The nature of substitution on the phenyl group in the 5 position seems also to modulate the inhibition (Prest 37, E925220 and M338835).

For Prest-392, three commercial compounds were tested, among which Altanserin and Tocris-DV7028 show high chemical similarity (Table 3). The three compounds present pharmacophore similarities to Prest-392: a central piperidine, a bicycle diamide on one side and a phenyl ring on the other side. Altanserin, the direct thio analogue of Prest-392 has a higher to equivalent IC_{50} on the GT activity ($IC_{50} = 9.3 \pm 2.7$) than the one of the head of series confirming the structure related activity of the lead compound. Tocris-DV7028 lacking the quinazoline-2,4-dione motif for the benefit of an unsaturated ring is inactive. This single modification compared to Prest-392 indicates a determinant role of this heteroaromatic bicycle in the inhibitory activity. In Chembridge-9231888 molecular structure, the inversion of flexibility and rigidity on either side of the central piperidine seems detrimental for the activity. This conformational change could be responsible for the complete loss of affinity to the target. To further investigate the role of the structural features in the inhibitory activity, we performed a chemical group assessment on Ketanserin and selected four fragments representative of the main chemical groups present in the molecule. Unfortunately, none of these fragments showed IC_{50} below $100\mu M$. Although these fragments could have an individual affinity to the target, the level of binding energy was not high enough to result in

Table 1
List of compounds presenting more than 80% inhibition on the GT activity carried out by VEEV nsP1. n.d.: not determined.

Compound information			GT reaction	
Prestw nº	Commercial name	Structure	ELISA	WB
			% Inhi. at 50 µM	IC ₅₀ (µM) ± SD
37	Pyrimethamine		100	2.7 ± 0.4
Head of series 1				
143	Chlorhexidine		81.3	> 200
172	Ethofylline		89.4	15.8
182	Levamisole hydrochloride		97.5	> 200
203	Flufenamic acid		90.6	> 200
308	Fimozide		87.4	102.3 ± 40.3
388	Dequalinium dichloride		88.1	15.7 ± 7.5
392	Ketanserin tartrate hydrate		99.4	14.6 ± 2.9
Head of series 2				
531	Firenperone		85.6	39.6 ± 18.3
Head of series 2				
544	Tobramycin		101	> 200
1447	Nalidixic acid sodium salt		85.3	59.8 ± 32.2
777	Alexidine dihydrochloride		85.5	> 200
858	Doxazin mesylate		83.2	29.6 ± 23.9
998	Zincpenthixol dihydrochloride		81.9	n. d.
1029	Risperidone		79.5	> 200
1068	Thiethylperazine dimaleate		100	n. d.
1105	Verteporfin		99.5	13.5 ± 3.1
1117	Azapaperone		81.1	> 200

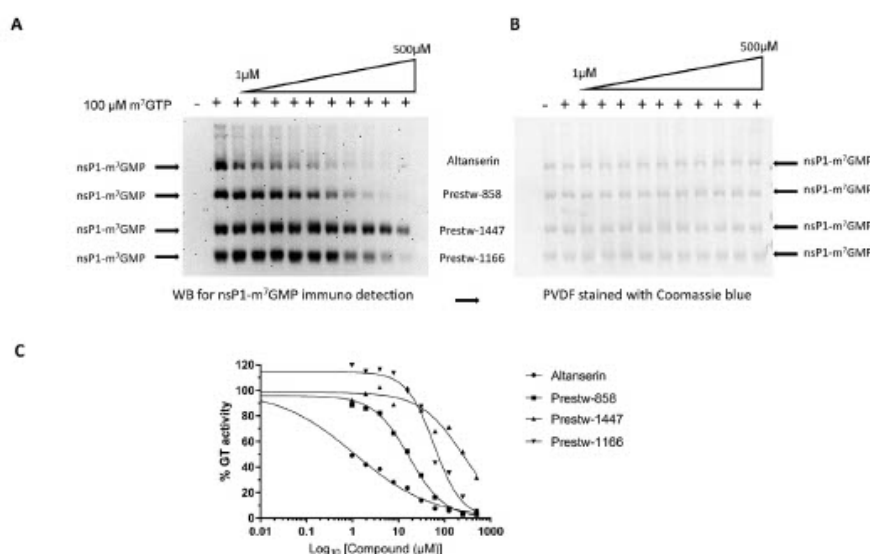


Fig. 3. IC₅₀ determination of compounds on the VEEV nsP1 GT activity determined by Western blot. Panel A: Detection of m⁷GMP-nsP1 complex by Western blot. Panel B: PVDF membrane stained with Coomassie blue. Panel C: Hill plot fitting curve for IC₅₀ determination deduced after Western blot quantification.

Table 2

Selected compounds from the analogue search of the head of series 1 (Prest-37) used for the evaluation of the GT activity carried by nsP1 VEEV. The analogue search was performed using the screened library from Prestwick Chemical and commercially available databases. n.d.: not determined.


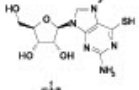

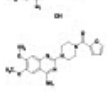
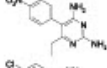
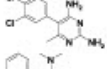
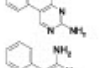
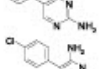
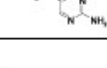
Compound information				GT reaction		MTase reaction
Supplier	Supplier reference	Commercial name or CAS (#)	Structure	ELISA	WB	FBA
				% Inhi. at 50 μM	IC ₅₀ (μM)	% Inhi. at 50 μM
Series 1: Head of series						
Prestwick Chemical	37	Pyrimethamine		100	2.7 ± 0.4	73.5 ± 2.9
Head of serie 1						
Series 1: Analogue search in screened libraries						
Prestwick Chemical	347	Thioguanosine		47.8	115 ± 69.4	n.d.
Prestwick Chemical	858	Doxazoxin mesylate		83.2	29.6 ± 12.6	38.3 ± 3.3
Prestwick Chemical	947	Pransin hydrochloride		74.3	29.5 ± 23.9	37.2 ± 3.0
Series 1: Analogue search in commercial databases						
TRC	E925220	N. A CAS: 71582-34-6		n. d.	39.9 ± 21.3	n.d.
TRC	M338835	Metoprine CAS: 7761-45-7		n. d.	16.1 ± 3.6	n.d.
Key Organics	1B-058	N. A CAS: 320424-61-1		n. d.	> 200	n.d.
Enamine	EN300-183279	N. A CAS: 18588-49-3		n. d.	> 200	n.d.
Enamine	EN300-208764	N. A CAS: 17039-14-4		n. d.	> 200	n.d.

Table 3
Selected compounds from the analogue search of the head of series 2 (Prest-392 and 531) used for the evaluation of the GT activity carried by nsP1 VEEV. The analogue search was performed using the screened library from Prestwick Chemical and commercially available databases.

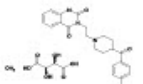
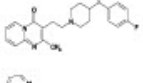
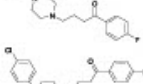
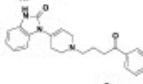
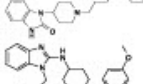
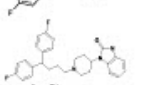

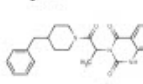
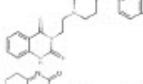

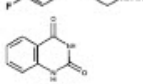
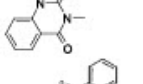
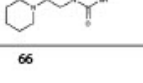


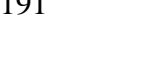

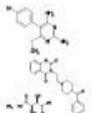
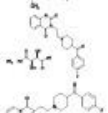
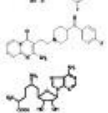
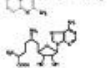
Compound information				GT reaction		MTase reaction
Supplier	Supplier reference	Commercial name CAS (n°)	Structure	ELISA % Inhi. at 50 µM	WB IC ₅₀ (µM)	FBA % Inhi. at 50 µM
Serie 2: Head of series						
Prestwick Chemical	392 Head of serie 3	Ketanserin tartrate hydrate		99.4	14.6 ± 2.9	79.3 ± 1.7
Prestwick Chemical	531 Head of serie 3	Pirenperone		85.6	39.6 ± 18.3	44.6 ± 0.5
Serie 2: Analogue search in screened library						
Prestwick Chemical	1117	Azaparone		81.0	> 200	n.d.
Prestwick Chemical	115	Haloperidol		62.7	> 200	n.d.
Prestwick Chemical	360	Droperidol		75.5	22.5	n.d.
Prestwick Chemical	484	Benperidol		51.2	62.6	n.d.
Prestwick Chemical	136	Astemizole		70.8	97.8 ± 59.2	n.d.
Prestwick Chemical	308	Pimozide		87.4	102 ± 40.3	n.d.
Prestwick Chemical	1029	Risperidone		79.5	> 250	n.d.
Prestwick Chemical	1166	Rebamipide		74.0	94.5 ± 49.8	n.d.
Serie 2: Analogue search in Commercial data bases						
Chembridge	9231888	N. A N. A		n. d.	> 200	n.d.
Aldrich Sigma	A8106	Altanserin CAS: 76330-71-7		n. d.	9.3 ± 2.7	n.d.
Tocris	DV7028	N. A CAS: 133364-62-2		n. d.	> 200	n.d.
Serie 2: Fragments in commercial databases						
Alfa Aesar	B23142	N. A CAS:25519-78-2		n. d.	> 200	n.d.
Alfa Aesar	AS 142026	Benazylneurea CAS: 86-96-4		n. d.	> 200	n.d.
Enamine	EN300-23764	N. A CAS: 607-19-2		n. d.	> 200	n.d.
Chembridge	9195974	N. A CAS:144734-42-9?		n. d.	> 200	n.d.

Table 4

Evaluation of the inhibition of the head of series 1 and 2 in both GT and MTase activity carried by VEEV nsP1 and/or Human N7 methyltransferase. Determination of IC₅₀ for the GT activity was performed by Western blot and the determination of the %age of inhibition of MTase activity at 50 μM was measured by DEAE filter binding assay (FBA). Sinefungin was included as a reference compound. n.d.: not determined. n.i.: no inhibition.

Compound information				GT reaction		MTase reaction	
Hit series	Head compound	Commercial name	Structure	WB		FBA	
				IC ₅₀ (μM)		% Inhi. at 50 μM	
				nsP1(WT) VEEV	nsP1(D34S) VEEV	nsP1 VEEV	HTN
1	37	Pyrimethamine		2.7 ± 0.4	2.8 ⁺	73.5 ± 2.9	n. i.
2	392	Ketanserine tartrate hydrate		14.6 ± 2.9	n.d.	79.3 ± 1.7	2.8 ± 4.0
	531	Pirenperone		39.6 ± 18.3	65.3 ⁺	44.6 ± 0.5	n. i.
Reference compound	-	Sinefungine		29.12 ± 2.6		1.5 ± 0.6	98.2 ± 0.3

i: The value is original from Li, C., Guillen, J., Rabah, N., Blanjoie, A., Debart, F., Vasseur, J. J., Canard, B., Decroly, E. & Coutard, B. (2015). *J Virol* 89, 8292–8303. n. i.: no inhibition; n. d.: not determined; +: data from a single experiment.

detectable inhibition. We also mixed fragments representative of the two distinct chemical groups. The groups tested were the B23142 together with AS 142026, the B23142 together with EN300-23764 and the B23142 together with Chembridge 9195974. None of the combination could give IC₅₀ in the 100 μM range or below (data not shown).

3.5. Prest-37 and Prest-531 specificity and preliminary characterization of their mode of action

A series of compounds with a chemical structure of 3-aryl-[1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones called MADTPs, targeting alpha-virus nsP1 has recently been described (Gigante et al., 2014, 2017) (Delang et al., 2016). A resistance mutation (D34S) to these compounds was found in the N-terminal region of nsP1. The D34S mutation was introduced into the VEEV nsP1 sequence, and allowed the enzyme to keep its GT activity in the presence of MADTP whereas MADTP inhibited the wt VEEV nsP1. To evaluate if some of the identified inhibitors could share the same mode of action (MoA) with MADTPs, we tested Prest-37 and Prest-531 on both wt and D34S VEEV nsP1. The data in Table 4 show that the D34S mutation did not reduce the inhibitory effect of the compounds on the GT activity, suggesting the MoA is different from that of MADTPs series.

The inhibition of the Human N7-MTase (also known as RNMT) was also tested to evaluate the specificity of the compounds. Prest-37 and Prest-392 were tested at 50 μM, and no significant inhibition effect was observed (Table 4), suggesting that the hit compounds are VEEV nsP1 specific inhibitors.

4. Discussion

Arboviruses, among which alphaviruses, are responsible for many viral emergencies causing worldwide health and economic troubles. Many efforts on the development of therapeutics are currently on the way but to date there is no approved treatment dedicated to alphaviruses. One of the first steps toward the development of drugs is the selection and validation of drug targets defined as proteins bearing functions which are essential for the virus replication and which can be modulated by small compounds. Alphavirus nsP1 is the central enzyme of the viral mRNA capping as it is involved in at least three steps of the cap synthesis, namely GTP methylation (MTase), nsP1 guanylation

(GT) and cap transfer on mRNA (GTase). Mutational abrogation of one of these functions leads to loss of viral replication (Kallio et al., 2016). Since the first reports of nsP1 production and purification suitable for inhibition assays, compounds inhibiting MTase and GT reactions were characterized and showed antiviral effect in cell based assays (Feibelman et al., 2018; Tomar et al., 2011) (Delang et al., 2016), thus validating nsP1 as a *bona fide* antiviral target suitable for enzyme-based drug design. Enzyme-based screening assays have recently been developed for CHIKV nsP1, both based on the detection of the formation of m⁷GMP-nsP1 complex (Bullard-Feibelman et al., 2016; Feibelman et al., 2018; Kaur et al., 2018). Our study presents here the development of an ELISA-based assay in which the GT reaction was uncoupled from the MTase. This assay was validated with known nsP1 inhibitors, and then used for the screening of a library of more than 1200 approved drugs on the VEEV nsP1 GT reaction. We identified 18 compounds efficiently inhibiting the GT reaction at 50 μM, and 7 of them were further validated by a specific WB assay already used to determine their mode of action (Delang et al., 2016; Gigante et al., 2017). Thus, our results validated a high throughput ELISA assay to identify compounds blocking the VEEV nsP1-mediated GT reaction. During this work, Kaur and colleagues also reported a similar ELISA assay to detect the CHIKV m⁷GMP-nsP1 adduct, confirming the robustness of this technique for small molecules screening (Kaur et al., 2018). The main difference between the two reports is that we have found experimental conditions uncoupling the GT from the MTase reaction.

From this screening, 18 compounds (1.48% success rate) showed more than 80% inhibition at 50 μM, among which 9 had IC₅₀ values below the 100 μM range, as determined by WB. None of these compounds show structure similarity when compared to compounds selected from a GTP competition assay on CHIKV nsP1 (Feibelman et al., 2018). The best compounds Prest-37 and Prest-392 that were chosen as head of series show a better IC₅₀ on the GT activity of VEEV nsP1 than Sinefungin (29.1 ± 2.6 μM (Li et al., 2015)), and IC₅₀ in the same range as compounds already described to have an antiviral effect on CHIKV (Gigante et al., 2017). The search of analogues for compounds Prest-37 (series 1) and Prest-392 (series 2) was first done within the PCL screened library, and then extended to commercial databases. With the search of compounds in the PCL screened library, 12 structurally related compounds were selected. All of them show a significant inhibition activity on the GT reaction, suggesting that the screening assay

reproducibly detect inhibitors, even when not highly potent. In particular, Prest-531 which is very close to Prest-392 had been selected as a hit, highlighting a potent scaffold for GT inhibition. The second phase of the analogue search was performed in commercial libraries. None of the selected compounds for series 1 gave inhibition higher than or equivalent to that of Prest-37 but the structure analysis revealed the possible key role of an aliphatic group in position 6 of the diaminopyrimidine ring, as the three compounds lacking this group showed no GT inhibition. However, the number of tested compounds remains low (including Prest-37, $n = 6$) and deeper structure activity relationship studies on a larger number of compounds is required to confirm this hypothesis. The number of selected analogues for series 2 was higher ($n = 13$, including hits Prest-392 and 531). Altanserin, known to bind 5-HT_{2A} receptors (5-Hydroxytryptamine (serotonin) 2A receptor), is a more potent inhibitor than the hits of series 2. Altanserin differs from Prest-392 by harbouring a thioetone instead of a ketone in position 2 of the quinazoline-2,4-dione group. A preliminary SAR seems to emerge in this chemical series. Two aromatic moieties are mandatory at each extremity of the molecule with a defined spatial orientation. However, at this stage of the study and regarding the low number of analogues tested, we cannot conclude on the best nature of these aromatic moieties. Moreover, we do not know the importance of the central piperidine ring, or the nature and the length of the chains linking the piperidine to both external rings.

In order to gain insight into the mechanism of inhibition, we assessed the inhibition of head of series 1 and 2 on the MTase activity carried by VEEV nsP1. Prest-37 and 531 can significantly affect the MTase reaction (Table 1). This dual activity is not surprising as it has been already demonstrated that both MTase and GT reactions are coordinated (Li et al., 2015). Other developed compounds were also found to affect both GT and MTase activity of VEEV nsP1 among which compounds from MADTP series or SAM/SAH analogues represented by Sinefungin (Delang et al., 2016; Li et al., 2015), prompting us to evaluate a possible common mode of action despite structural differences. To this aim, we assessed the inhibition of GT activity of VEEV D34S nsP1 reluctant to MADTP inhibition. Both Prest-37 and 531 showed comparable inhibitory effect on both VEEV D34S and wt nsP1, suggesting that their mechanism of action is likely different from that of the MADTP which targets the N-terminal region of nsP1.

One of the key issue in the development of antivirals targeting capping enzymes is the specificity towards the viral protein, as cellular methyltransferases may share common structural features and have conserved substrate binding pockets with the viral target (Lim et al., 2011). The lack of structural information for nsP1 does not allow to claim that the binding pockets such as the SAM/SAH binding site are conserved with cellular MTases, it is therefore useful to address the specificity issue at the early stage of the study by testing the compounds on a cellular MTase of similar activity spectrum (Aouadi et al., 2017) (Coutard et al., 2017). Unlike Sinefungin, none of the tested compounds are able to inhibit the human N7-MTase used as a reference enzyme to address specificity. The development of an initial screening on the GT activity instead of the MTase may have resulted in the selection of compounds targeting an original viral activity, thus with a higher probability to be specific than compounds selected for MTase inhibition.

5. Conclusion

It is now established that capping of viral mRNA is a possible target for the development of molecules with antiviral effect as inhibitors of the capping would abrogate the translation of viral proteins and thereby impair viral replication. As the structure of the alphavirus nsP1 is not known, we developed an enzyme based assay to select among approved compounds small molecules inhibiting one of the step of the capping process using VEEV nsP1 as model. The development of the assay enabled the screening of 1220 compounds and the selection of 18

compounds followed by further characterization of two series of compounds. The analysis of the series suggest that they work through an original mechanism of action when compared to reference molecules, opening new opportunities for the development of novel anti-VEEV compounds.

Acknowledgments

This work was supported by the French research agency ANR “VMTaseIn”, grant ANR-ST14-ASTR-0026, and by the European Union (Seventh Framework Program, Marie Skłodowska-Curie ETN “EUVRINA”, grant agreement number 264286 and Horizon 2020 Marie Skłodowska-Curie ETN “ANTIVIRALS”, grant agreement number 642434), and by the MINECO/FEDER SAF2015-64629-C2-1-R.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.antiviral.2019.01.003>.

References

- Adams, A.P., Navarro-Lopez, R., Ramirez-Aguilar, F.J., Lopez-Gonzalez, I., Leal, G., Flores-Mayorga, J.M., Travassos da Rosa, A.P., Saxton-Shaw, K.D., Singh, A.J., Borland, E.M., Powers, A.M., Tesh, R.B., Weaver, S.C., Estrada-Franco, J.G., 2012. Venezuelan equine encephalitis virus activity in the Gulf Coast region of Mexico, 2003–2010. *PLoS Neglected Trop. Dis.* 6, e1875.
- Aguilar, P.V., Estrada-Franco, J.G., Navarro-Lopez, R., Ferro, C., Haddow, A.D., Weaver, S.C., 2011. Endemic Venezuelan equine encephalitis in the Americas: hidden under the dengue umbrella. *Future Virol.* 6, 721–740.
- Aouadi, W., Eydloux, C., Coutard, B., Martin, B., Debart, F., Vasseur, J.J., Contreras, J.M., Maurice, C., Querat, G., Jung, M.L., Canard, B., Guillemot, J.C., Decroly, E., 2017. Toward the identification of viral cap-methyltransferase inhibitors by fluorescence screening assay. *Antivir. Res.* 144, 330–339.
- Ballard-Felbelman, K.M., Fuller, B.P., Geiss, B.J., 2016. A Sensitive and Robust High-Throughput Screening Assay for Inhibitors of the Chikungunya Virus nsP1 Capping Enzyme. *PLoS One* 11, e0158923.
- Charrel, R.N., Leparc-Goffart, I., Gallian, P., de Lamballerie, X., 2014. Globalization of Chikungunya: 10 years to invade the world. *Clin. Microbiol. Infect.* 20, 662–663.
- Chung, D., Schroeder, C.E., Sotisky, J., Yao, T., Roy, S., Smith, R.A., Tower, N.A., Noah, J.W., McKellip, S., Sosa, M., Rasmussen, L., White, E.L., Aube, J., Golden, J.E., 2013. M1336: Development of Quinazolinone-Based Inhibitors against Venezuelan Equine Encephalitis Virus (VEEV). *Probe Reports from the NIH Molecular Libraries Program*.
- Chung, D.H., Jonsson, C.B., Tower, N.A., Chu, Y.K., Sahin, E., Golden, J.E., Noah, J.W., Schroeder, C.E., Sotisky, J.B., Sosa, M.J., Cramer, D.E., McKellip, S.N., Rasmussen, L., White, E.L., Schmaljohn, C.S., Julander, J.G., Smith, J.M., Filone, C.M., Connor, J.H., Sakurai, Y., Davey, R.A., 2014. Discovery of a novel compound with anti-venezuelan equine encephalitis virus activity that targets the nonstructural protein 2. *PLoS Pathog.* 10, e1004213.
- Coutard, B., Barral, K., Lichiere, J., Selisko, B., Martin, B., Aouadi, W., Lombardia, M.O., Debart, F., Vasseur, J.J., Guillemot, J.C., Canard, B., Decroly, E., 2017. Zika Virus Methyltransferase: structure and Functions for Drug Design Perspectives. *J. Virol.* 91, e02202.
- Delang, L., Li, C., Tas, A., Querat, G., Albulescu, I.C., De Burghgraeve, T., Guerrero, N.A., Gigante, A., Florowski, G., Decroly, E., Jochmans, D., Canard, B., Snijder, E.J., Perez-Perez, M.J., van Hemert, M.J., Coutard, B., Leysen, P., Neyts, J., 2016. The viral capping enzyme nsP1: a novel target for the inhibition of chikungunya virus infection. *Sci. Rep.* 6, 31819.
- Delang, L., Segura Guerrero, N., Tas, A., Querat, G., Pastorino, B., Froeyen, M., Dallmeier, K., Jochmans, D., Herdewijn, P., Belli, F., Snijder, E.J., de Lamballerie, X., Martins, B., Neyts, J., van Hemert, M.J., Leysen, P., 2014. Mutations in the chikungunya virus non-structural proteins cause resistance to favipiravir (T-705), a broad-spectrum antiviral. *J. Antimicrob. Chemother.* 69, 2770–2784.
- Delella, P.C., Raveh, A., Larsen, M.J., Schulz, P.J., Tamayo-Castillo, G., Sherman, D.H., Miller, D.J., 2015. The combined use of alphavirus replicons and pseudoinfectious particles for the discovery of antivirals derived from natural products. *J. Biomol. Screen* 20, 673–680.
- Felbelman, K.M., Fuller, B.P., Li, L., LaBarbera, D.V., Geiss, B.J., 2018. Identification of small molecule inhibitors of the Chikungunya virus nsP1 RNA capping enzyme. *Antivir. Res.* 154, 124–131.
- Furuta, Y., Gowen, B.B., Takahashi, K., Shiraki, K., Smece, D.F., Barnard, D.L., 2013. Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antivir. Res.* 100, 446–454.
- Gigante, A., Canela, M.D., Delang, L., Priego, E.M., Camarasa, M.J., Querat, G., Neyts, J., Leysen, P., Perez-Perez, M.J., 2014. Identification of [1,2,3]triazol[4,5-d]pyrimidin-7(6H)-ones as novel inhibitors of Chikungunya virus replication. *J. Med. Chem.* 57, 4000–4008.
- Gigante, A., Gomez-Sanjuan, A., Delang, L., Li, C., Bueno, O., Gamo, A.M., Priego, E.M., Camarasa, M.J., Jochmans, D., Leysen, P., Decroly, E., Coutard, B., Querat, G., Neyts,

- J., Perez-Perez, M.J., 2017. Antiviral activity of [1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones against chikungunya virus targeting the viral capping mP1. *Antivir. Res.* 144, 216–222.
- Gomez-SanJuan, A., Gamo, A.M., Delang, L., Perez Sanchez, A., Amran, S.N., Abdelnabi, R., Jacobs, S., Priego, E.M., Camarasa, M.J., Jochmans, D., Leysen, P., Ng, L.F.P., Querat, G., Neyts, J., Perez-Perez, M.J., 2018. Inhibition of the replication of different strains of chikungunya virus by 3-aryl-[1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones. *ACS Infect. Dis.* 4, 605–619.
- Gould, E.A., Coutard, B., Malet, H., Morin, B., Jamal, S., Weaver, S., Gorbalenya, A., Moureau, G., Baronsi, C., Delogu, I., Forrester, N., Khasnatov, M., Gröms, T., de Lamballerie, X., Canard, B., 2010. Understanding the alphaviruses: recent research on important emerging pathogens and progress towards their control. *Antivir. Res.* 87, 111–124.
- Hyde, J.L., Gardner, C.L., Kimura, T., White, J.P., Liu, G., Trobaugh, D.W., Huang, C., Tonelli, M., Pressler, S., Takeda, K., Klimstra, W.B., Amarasinghe, G.K., Diamond, M.S., 2014. A viral RNA structural element alters host recognition of nonself RNA. *Science* 343, 783–787.
- Johnson, K.M., Martin, D.H., 1974. Venezuelan equine encephalitis. *Adv. Vet. Sci. Comp. Med.* 18, 79–116.
- Kallio, K., Hellstrom, K., Jokitalo, E., Ahola, T., 2016. RNA replication and membrane modification require the same functions of alphavirus nonstructural proteins. *J. Virol.* 90, 1687–1692.
- Kaur, R., Mudgal, R., Narwal, M., Tomar, S., 2018. Development of an ELISA assay for screening inhibitors against divalent metal ion dependent alphavirus capping enzyme. *Virus Res.* 256, 209–218.
- Li, G., Guillen, J., Rabah, N., Banjole, A., Debart, F., Vasseur, J.J., Canard, B., Decroly, E., Coutard, B., 2015. mRNA capping by Venezuelan equine encephalitis virus nsP1: functional characterization and implications for antiviral research. *J. Virol.* 89, 8292–8303.
- Lim, S.P., Sonntag, L.S., Nohle, C., Nilar, S.H., Ng, R.H., Zou, G., Monaghan, P., Chung, K.Y., Dong, H., Liu, B., Bodenreider, C., Lee, G., Ding, M., Chan, W.L., Wang, G., Jian, Y.L., Chao, A.T., Lescar, J., Yin, Z., Vednanda, T.R., Keller, T.H., Shi, P.Y., 2011. Small molecule inhibitors that selectively block dengue virus methyltransferase. *J. Mol. Chem.* 286, 6233–6240.
- Lucas-Hourani, M., Munier-Lehmann, H., Helynyck, O., Komarova, A., Despres, P., Tangy, F., Vidalain, P.O., 2014. High-throughput screening for broad-spectrum chemical inhibitors of RNA viruses. *J. Vis. Exp.*
- Markland, W., McQuaid, T.J., Jain, J., Kwong, A.D., 2000. Broad-spectrum antiviral activity of the IMP dehydrogenase inhibitor VX-497: a comparison with ribavirin and demonstration of antiviral additivity with alpha interferon. *Antimicrob. Agents Chemother.* 44, 859–866.
- Pisano, M.B., Oria, G., Beskow, G., Aguilar, J., Konigheim, B., Cacace, M.L., Aguirre, L., Stein, M., Contigiani, M.S., 2013. Venezuelan equine encephalitis viruses (VEEV) in Argentina: serological evidence of human infection. *PLoS Neglected Trop. Dis.* 7, e2551.
- Pohjala, L., Utt, A., Varjak, M., Lulla, A., Meris, A., Ahola, T., Tammela, P., 2011. Inhibitors of alphavirus entry and replication identified with a stable Chikungunya replicon cell line and virus-based assays. *PLoS One* 6, e28923.
- Ravichandran, R., Mariani, M., 2008. Ribavirin therapy for Chikungunya arthritis. *J. Infect. Dev. Ctries* 2, 140–142.
- Rivas, F., Diaz, L.A., Cardenas, V.M., Daza, E., Bruzon, L., Alcalá, A., De la Hoz, O., Caceres, F.M., Aristizabal, G., Martinez, J.W., Revelo, D., De la Hoz, F., Boshell, J., Camacho, T., Calderon, L., Olano, V.A., Villarreal, L.I., Roselli, D., Alvarez, G., Ludwig, G., Tsai, T., 1997. Epidemic Venezuelan equine encephalitis in La Guajira, Colombia, 1995. *J. Infect. Dis.* 175, 828–832.
- Seyedi, S.S., Shukri, M., Hassandarvish, P., Oo, A., Shankar, E.M., Abubakar, S., Zandi, K., 2016. Computational approach towards exploring potential anti-chikungunya activity of selected flavonoids. *Sci. Rep.* 6, 24027.
- Spurgens, K.B., Hurt, C.R., Cohen, J.W., Eccleston, L.T., Lind, C.M., Lingappa, V.R., Glass, P.J., 2013. Validation of a cell-based ELISA as a screening tool identifying anti-alphavirus small-molecule inhibitors. *J. Virol. Methods* 193, 226–231.
- Tomar, S., Narwal, M., Harms, E., Smith, J.L., Kuhn, R.J., 2011. Heterologous production, purification and characterization of enzymatically active Sindbis virus nonstructural protein nsP1. *Protein Expr. Purif.* 79, 277–284.
- Urakova, N., Kuznetsova, V., Crossman, D.K., Sokratian, A., Guthrie, D.B., Kolykhalov, A.A., Lockwood, M.A., Natchus, M.G., Crowley, M.R., Painter, G.R., Frolova, E.I., Frolov, I., 2017. beta-D-N(4)-hydroxycytidine is a potent anti-alphavirus compound that induces high level of mutations in viral genome. *J. Virol.* 92, e01965.
- Varghese, F.S., Kaukinen, P., Glasker, S., Beshpalov, M., Hanaki, L., Wemmerberg, K., Kummerer, B.M., Ahola, T., 2016. Discovery of berberine, abamectin and ivermectin as antivirals against chikungunya and other alphaviruses. *Antivir. Res.* 126, 117–124.
- Vasiljeva, L., Meris, A., Auvinen, P., Kaariainen, L., 2000. Identification of a novel function of the alphavirus capping apparatus. RNA 5'-triphosphatase activity of Nsp2. *J. Biol. Chem.* 275, 17281–17287.
- Wach, Y., Orba, Y., Sasaki, M., Kobayashi, S., Carr, M.J., Nohori, H., Sato, A., Hall, W.W., Sawa, H., 2017. Discovery of a novel antiviral agent targeting the nonstructural protein 4 (nsP4) of chikungunya virus. *Virology* 505, 102–112.
- Wang, H.L., O'Rear, J., Stollar, V., 1996. Mutagenesis of the Sindbis virus nsP1 protein: effects on methyltransferase activity and viral infectivity. *Virology* 217, 527–531.

THIS PAGE IS INTENTIONALLY LEFT BLANK

8.2 Additional Results: Hit/series validation beyond the screened library against the VEEV nsP1 capping enzyme and pharmacophore evaluation of Prest-1291

THIS PAGE IS INTENTIONALLY LEFT BLANK

In continuation with the results presented earlier, on chapter 8.1, here are presented the results obtained from the work performed on series 3 represented by Prest-1291 (imiquimod). We have conducted further analysis and pharmacophore evaluation of Prest-1291 that shows an IC_{50} of $92.1 \pm 0.4 \mu\text{M}$.

Analogue search and structural activity relationship (SAR) were initiated to identify the active pharmacophore features. The compounds were selected by using the scaffold of Prest-1291 (imiquimod) for analogue search. I have performed the analogue search at the Prestwick Chemical libraries and in other commercially available databases such as Reaxys and eMolecules. Some of these analogues were commercially available and purchased. Then, I have tested the compounds against the guanylylation activity of nsP1 (table 1).

In a similar manner to the other series, I have also tested the inhibition of the head of series 3 (Prest-1291 [imiquimod]) on the VEEV D34S nsP1. Imiquimod showed comparable inhibitory effect on both VEEV D34S and wt nsP1, suggesting that its mechanism of action is likely different from that of the MADTP which targets the N-terminal region of nsP1 (table 2).

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

RESULTS AND DISCUSSION

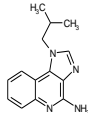
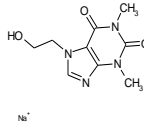
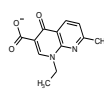
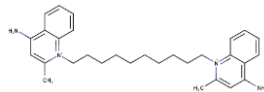
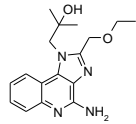
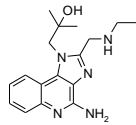
In addition to series 1 and 2 presented earlier, we have also initiated analogue search and structural activity relationship for a third series of compound. The series 3 was represented by Prest-1291 (imiquimod) and the IC₅₀ for the GT reaction was determined by Western blot and it was $92.1 \pm 0.4 \mu\text{M}$. Imiquimod belongs to the imidazoquinolinamine drugs and acts through activation of Toll-like receptor-7 (TLR-7) and thus induce both anti-viral and anti-tumoral activities (Hemmi *et al.*, 2002). Activation of TLR-7 induces amplification of the immune response through stimulation of inflammatory cytokines (Dahl, 2000, Skinner, 2003) leading to production of IFN- α that suppresses replication of viruses.

Analogue search was performed for imiquimod and other two related imidazoquinolinamine drugs were purchased and tested for GT reaction towards VEEV nsP1. Resiquimod, that is a TLR7/8 agonist that is associated with greater stimulation of cytokines and with activation of dendritic cells (Meyer *et al.*, 2013) does not showed higher or equal inhibitory effect to prevent the GT reaction (table 1). The second selected imidazoquinolinamine analogue was gardiquimod which has been described as a specific TLR7 agonist when used at concentrations below $10 \mu\text{M}$ (Buitendijk *et al.*, 2013). It has been suggested that gardiquimod also presents two anti-HIV-1 modes of action, one as an immune system modifier and the other as inhibitor of HIV-1 reverse transcriptase during early step in the life cycle of HIV-1 (Buitendijk *et al.*, 2013). In this specific case gardiquimod could be used for the development of antiviral therapeutics agent to block HIV-1 infection. It was interesting to observe that in our experiments gardiquimod presented an IC₅₀ of $64.6 \mu\text{M}$ (one unique single experiment) and apparently showed higher effect to prevent the GT reaction than imiquimod (table 2). This suggest that gardiquimod may also present a dual mode of action against other viruses such as VEEV. Additional studies are required to understand if gardiquimod presents a dual mode of action against VEEV, one as immune modulator and other as preventing the cap formation of the messenger RNA.

Both resiquimod and gardiquimod are compounds of higher pharmacophore similarity and the core structure is 1H-imidazol[4,5-c]quinoline-4-amine, the same as imiquimod. While imiquimod presents an isobutyl group at the N1 position of the imidazole group, resiquimod has a 2-methylpropan-2-ol and an ethoxymethyl group at the C2 position of the imidazole group. Interestingly, the substitution of the ethoxymethyl group (acceptor group) at the C2 position of the imidazole group to an ethylamino group (donor group) generates the analogue gardiquimod. This substitution conferred higher effect to prevent

the GT reaction than the effect observed in the other imidazoquinolinamine compounds already tested. The pharmacophore similarities observed in gardiquimod and imiquimod and their effect in preventing the GT reaction from VEEV nsP1 confirms the structure related activity of the lead compound. Therefore, the gardiquimod and imiquimod are promising compounds to continue with structure activity relationship studies and further works are needed to improve the specificity of gardiquimod to target the capping enzyme from VEEV.


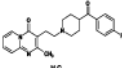



Table 2. Selected compounds from the analogue search of the head of series 3 (Prest-1291) used for the evaluation of the GT activity carried by nsP1 VEEV. The analogue search was performed using the screened library from Prestwick Chemical and commercially available databases. n.d.: not determined

GT					
Supplier	Supplier reference	Commercial name CAS (n°)	Structure	% Inhi. at 50µM	IC ₅₀ (µM)
				ELISA	WB
Serie 3: Head of serie					
Prestwick Chemical	1291 Head of serie 3	Imiquimod		69.2 ± 8.6	92.1 ± 0.4
Serie 3: Analogue search in Prestwick libraries					
Prestwick Chemical	172	Etofylline		44.5 ± 19.6	15.8
Prestwick Chemical	1447	Nalidixic acid sodium salt		17.5 ± 16.6	59.8 ± 32.2
Prestwick Chemical	388	Dequalinium diachloride		99.1 ± 2.1	15.7 ± 7.5
Serie 3: Analogue search in Commercial data bases					
Combi Blocks	QB-9540	Resiquimod CAS: 144875-48-9		n. d.	>200
Enzo Life Science	ALX-420-040	Gardiquimod N. A		n. d.	64.6

The mode of action (MoA) for compounds targeting nsP1 enzyme from VEEV has been described for a series of compounds with a chemical structure of 3-aryl-[1,2,3]triazolo[4,5-d]pyrimidin-7(6H)-ones called MADTPs (Delang *et al.*, 2016, Gigante *et al.*, 2014, Gigante *et al.*, 2017). A resistance mutation (D34S) to these MADTP compounds was found in the N-terminal region of nsP1. While MADTP inhibited the *wt* VEEV nsP1, the introduction of the D34S mutation in the sequence of VEEV nsP1 allowed the enzyme to keep its GT activity in the presence of the MADTP compound. Assessment of the MoA for the series of compounds represented by imiquimod was performed in order to understand if similarities with its MoA are shared with MADTPs. Accordingly, we have tested imiquimod on both *wt* and D34S VEEV nsP1 and the data in table 3 shows that the D34S mutation did not reduce the inhibitory effect of the compounds on the GT activity. Altogether, the data suggests that the MoA is different from that of MADTPs series. Unfortunately, the evaluation of the specificity of imiquimod through the Human N⁷-MTase (also known as RNMT) assays revealed that imiquimod at a concentration of 50 μ M slightly inhibited the Human N⁷-MTase in the 12.0 \pm 17.0 μ M range (table 3). Despite the observed unspecificity of imiquimod, it is still necessary to evaluate gardiquimod specificity as it is a promising imidazoquinolinamine that presented a considerable inhibition of the GT activity of VEEV nsP1.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Table 3. Evaluation of the inhibition of the head of series 1 and 2 in both GT and MTase activity carried by VEEV nsP1. Determination of IC₅₀ for the GT activity was performed by western-blot and the determination of the % of inhibition of MTase activity at 50 μM was measured by DEAE filter binding assay (FBA). Sinefungin was included as a reference compound. n.i.: no inhibition, n.d.: not determined

Compound information				GT reaction		MTase reaction	
Hit series	Head compound	Commercial name	Structure	WB		FBA	
				IC ₅₀ (μM)		% Inhi. at 50 μM	
				nsP1(WT) VEEV	nsP1(D34S) VEEV	nsP1 VEEV	H7N
1	37	Pyrimethamine		2.7 ± 0.4	2.8	73.5 ± 2.9	n. i.
2	531	Pirenperone		39.6 ± 18.3	65.3	44.6 ± 0.5	n. i.
3	1291	Imiquimod		92.1 ± 0.4	69.3	48.0 ± 7.0	12.0 ± 17.0
Reference compound	--	Sinefungine		ⁱ 29.12 ± 2.6	n.d.	ⁱ 1.5 ± 0.6	98.2 ± 0.3
Reference compound	--	MADTP-372		ⁱⁱ 10 ± 0.3	ⁱⁱ >200	n.d.	n.d.

ⁱ The value is original from Li, C., Guillen, J., Rabah, N., Blanjoie, A., Debart, F., Vasseur, J. J., Canard, B., Decroly, E. & Coutard, B. (2015). *J Virol* **89**, 8292-8303.

ⁱⁱ The value is original from Delang, L., Li, C., Tas, A., Querat, G., Albuлесcu, I. C., De Burghgraeve, T., Guerrero, N. A., Gigante, A., Piorkowski, G., Decroly, E., Jochmans, D., Canard, B., Snijder, E. J., Perez-Perez, M. J., van Hemert, M. J., Coutard, B., Leyssen, P. & Neyts, J. (2016). *Sci Rep* **6**, 31819.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

8.3 Additional Results: Enzymatic screening to find more potent MADTP compounds targeting the nsP1 of VEEV

THIS PAGE IS INTENTIONALLY LEFT BLANK

Compounds from the family 3-aryl-[1,2,3]triazolo[4,5-d]pyrimidin-7[6H]-ones, also known as MADTPs, were previously demonstrated to affect both GT and MTase activity of VEEV nsP1, and VEEV D34S nsP1 was resistant to MADTP inhibition (Delang et al., 2016; Li et al., 2015).

An approach was set up to improve the potency of MADTP compounds. Several compounds from the same family were rationally designed and synthesised by our collaborators at CISIC (Madrid, Spain). The compounds were sent to our laboratory for being tested on our enzymatic assay for the guanylation activity of VEEV nsP1 enzyme.

I have determined the IC_{50} for twenty-five MADTP compounds and the results are presented on table 4.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

Table 4. MADTP compounds used for the evaluation of the GT activity carried by nsP1 VEEV. The IC₅₀ was determined in duplicate, except when the standard error is not presented.

GT			
Reference	Compound family	Structure	IC ₅₀ (μM)
			WB
314	MADTP	n.a.	177.6 ± 4.8
410	MADTP	n.a.	38.0 ± 0.9
411	MADTP	n.a.	41.0 ± 1.7
418	MADTP	n.a.	9.3 ± 2.4
428	MADTP	n.a.	13.3 ± 2.7
429	MADTP	n.a.	1.4 ± 0.8
431*	MADTP	n.a.	2.0 ± 0.9
431-2	MADTP	n.a.	0.7
434	MADTP	n.a.	3.7 ± 0.8
439	MADTP	n.a.	418.9 ± 73.2
440	MADTP	n.a.	3.9 ± 2.2
441	MADTP	n.a.	4.1 ± 2.8
443	MADTP	n.a.	2.8 ± 2.2
444	MADTP	n.a.	3.2 ± 1.6
445	MADTP	n.a.	2.1 ± 1.1
446	MADTP	n.a.	8.4 ± 3.1
448	MADTP	n.a.	> 200
452	MADTP	n.a.	5.8 ± 0.4
454	MADTP	n.a.	2.8 ± 1.2
457	MADTP	n.a.	51.3
458	MADTP	n.a.	> 200
459	MADTP	n.a.	24.0
464	MADTP	n.a.	2.7 ± 1.3
465	MADTP	n.a.	167.4 ± 58.7
466	MADTP	n.a.	3.6 ± 0.2

THIS PAGE IS INTENTIONALLY LEFT BLANK

8.4 Discussion

THIS PAGE IS INTENTIONALLY LEFT BLANK

Mosquito-borne viruses are a major global public health concern among which are the (re-) emergency of alphaviruses which are causing worldwide health and economic problems. Currently there are no approved drugs to fight the global burden of alphaviruses. Safe and efficient therapeutics to treat infections of alphaviruses are highly needed and are to date under development.

The selection and validation of viral drug targets is the starting point for the development of drugs. Usually proteins bearing crucial functions for the virus replication and that can be modulated by small compounds are selected as drug targets. For alphaviruses the viral mRNA capping is essential for translation initiation of viral proteins while it is insufficient to scape to the immune detection. Thus, the alphavirus mRNA capping mechanism is an excellent target for drug design as inhibition of this mechanism would abolish the translation of viral proteins. Compromise the viral replication through inhibition of the alphavirus cap formation was one of our aims that we engage with this work in which we designed and developed a strategy to select small molecules inhibiting one of the steps of the alphavirus capping.

Although, alphavirus mRNA capping requires at least two non-structural proteins which are nsP1 and nsP2, the nsP1 is the enzyme that plays a key central role during viral mRNA capping. The nsP1 is a viral capping enzyme required for at least three steps of the cap synthesis, namely GTP methylation (MTase), nsP1 guanylylation (GT) and cap transfer on mRNA (GTase). Recent works have shown the importance of alphavirus capping processing mechanism for the viral replication. Experiments were performed using a plasmid-based *trans*-replication system for SFV and carrying multiple point mutations in the non-structural proteins to destroy their enzymatic activities or post translational modifications. Mutations destroying the enzymatic activities essential for the cap formation were included and were responsible for loss of viral replication (Kallio *et al.*, 2016).

The enzymatic activities carried out by nsP1 have emerged as a *bona fide* antiviral target and appropriated for enzyme-base drug design. To date the production and purification of nsP1 have overcome the limitations to obtain suitable amounts for inhibition assays. Indeed, compounds inhibiting MTase and GT reactions were already characterized and presented antiviral effect in cell-based assays (Feibelman *et al.*, 2018, Tomar *et al.*, 2011, Delang *et al.*, 2016). Detection of the formation of the m⁷GMP-nsP1 complex is being used for the development of enzyme-based screening assays for CHIKV nsP1 (Bullard-Feibelman *et al.*, 2016, Feibelman *et al.*, 2018, Kaur *et al.*, 2018). In our present work and differently from the work reported by Kaur *et. al* (2018), we have uncoupled the

GT reaction from the MTase reaction and thus discovered simpler experimental conditions for an ELISA-based assay to screen small molecules that inhibits the formation of the m⁷GMP-nsP1 adduct. For the presented assay we have used VEEV nsP1 as a model which was validated with known nsP1 inhibitors of the GT reaction. Our experiments together with previous works from others (Bullard-Feibelman *et al.*, 2016, Feibelman *et al.*, 2018, Kaur *et al.*, 2018) confirm that ELISA assays to detect the CHKV m⁷GMP-nsP1 adduct are a robust technique to be use for small molecules screening.

Consequently, we have proceeded with the screening on the VEEV nsP1 GT reaction and tested more than 1200 approved drugs that belong to the Prestwick Chemical library® (PCL®) at a concentration of 50 µM. With a success rate of 1.48 % showing more than 80 % inhibition our screening allowed the identification of 18 compounds that were effective in inhibiting the GT reaction. Among the 18 compounds 7 were further validated. The 7 compounds were selected for validation by using a specific Western blot (WB) assay already used previously (Delang *et al.*, 2016, Gigante *et al.*, 2017). Altogether our results validate high throughput ELISA assay to identify compounds blocking the VEEV nsP1 GT reaction. Among the 18 compounds 9 had IC₅₀ values below the 100 µM range that was determined by WB.

The best 3 compounds were chosen as head of series, the series 1 represented by Prest-37, the series 2 represented by Prest-392 and finally the series 3 represented by Prest-1291. The heads from series 1 and 2 were already presented in (Ferreira-Ramos *et al.*, 2019) and show a better IC₅₀ on the GT activity of VEEV nsP1 than Sinefungin (29.1 ± 2.6 µM) (Li *et al.*, 2015). The IC₅₀ of the heads from series 1 and 2 is in the same range as compounds already described to have antiviral effect on CHIKV (Gigante *et al.*, 2017). Therefore, we search for analogues for compounds Prest-37 (series 1) and Prest-392 (series 2). The analogue search was initially performed at the PCL® screened library, and then extended to commercial databases. The analogue search was also performed in parallel for a third series of compounds that is represented by Prest-1291. While the search of analogue compounds for series 1 and 2 in the PCL® screened library resulted in the selection of 12 structurally related compounds, the analogue search for series 3 resulted in the selection of 3 structurally related compounds. Without exception, the selected compounds from the analogue search in the PCL® screened library were tested for the VEEV nsP1 GT reaction and all presented a significant inhibition activity of the GT reaction. Although even when some inhibitors were not highly potent, the results from the first phase of the analogue search propose that the screening assay reproducibly detect inhibitors. Notably, Prest-531 which is very close to Prest-392 had been selected as a hit, featuring a potent scaffold for GT inhibition.

A second phase of the analogue search was accomplished in commercial libraries for the three series. As already reported in (Ferreira-Ramos *et al.*, 2019), work that resulted from this thesis, despite the low number of tested compounds for series 1 (including Prest-37, n=6) and that none of the selected compounds gave inhibition higher than or equivalent to that of Prestw-37, the structure analysis disclose the possible key role of an aliphatic group in the position 6 of the diaminopyrimidine ring. Indeed, three compounds lacking this group shown no GT inhibition. A more extended structure activity relationship studies on a broader number of compounds is needed to confirm this hypothesis.

Concerning the work achieved for series 2 resulted in a higher number of selected analogues (including hits Prest-392 and 531, n=13) and it was discovered a more potent inhibitor than the hits of series 2. The more potent inhibitor identified during this phase was Altanserin that is recognized to bind 5-HT_{2A} receptors (5-Hydroxytryptamine [serotonin] 2A receptor). Altanserin is distinguished from Prest-392 because Altanserin harbours a thioketone instead of a ketone in position 2 of the quinazoline-2,4-dione group. From a preliminary SAR initiated for this chemical series, two aromatic moieties are compulsory at each extremity of the molecule with a defined spatial orientation. Unfortunately, conclusions on the best nature of these aromatic moieties were not possible to draw at this stage of the study due to the low number of analogues tested. Indeed, we still do not know the significance of the central piperidine ring, or the nature and the length of the chains linking the piperidine to both external rings. The series 3 was also subject of a second phase of analogue search that was performed using the same commercial databases as used for series 1 and 2.

The second phase of analogue search for series 3 resulted in the selection of two commercially available analogues which were resiquimod and gardiquimod. Both resiquimod and gardiquimod are compound sharing the same core structure 1H-imidazol[4,5-c]quinoline-4-amine. While resiquimod showed no inhibition of the VEEV nsP1 GT reaction, gardiquimod seems to be a more potent inhibitor than the head of series 3 (imiquimod). Gardiquimod differs from resiquimod because instead of the acceptor group ethoxymethyl at the C2 position of the imidazole group harbours the donor group ethylamino at the same position. This substitution is not the unique difference between resiquimod although it seems to be important to prevent the GT reaction. Our results suggest that gardiquimod and imiquimod are promising compounds and therefore it is highly important to continue with structure activity relationship studies on this series of compounds.

With the propose to gain a better understanding into the mechanism of inhibition, the head of series were tested on the MTase activity carried by VEEV nsP1. The MTase reaction was significantly affected by the head of series (data not presented for Prest-392). Both GT and MTase activities of VEEV nsP1 are coordinated among them (Li *et al.*, 2015) and other developed compounds were also identified to affect both GT and MTase activities, among which are MADTP series or SAM/SAH analogues represented by Sinefungine (Li *et al.*, 2015, Delang *et al.*, 2016). Therefore, we were motivated to evaluate a possible common mode of action despite structural differences. We evaluated the head of series for the inhibition of GT activity on VEEV D34S nsP1 resistant to MADTP inhibition. Our results showed that Prest-37, 531 and 1291 showed comparable inhibitory effect on both VEEV D34S and *wt* nsP1, suggesting that their mechanism of action is likely different from that of the MADTP which targets the N-terminal region of the nsP1.

Due to the lack of structural information for nsP1 it is not possible to know if the binding cavities such as the SAM/SAH binding site are conserved with cellular MTases, and viral proteins may share common structural features and have conserved substrate binding pockets with the viral target (Lim *et al.*, 2011). Therefore, the specificity issue was addressed at the early stage of the study by testing the compounds on a cellular MTase of similar activity spectrum (Aouadi *et al.*, 2017, Coutard *et al.*, 2017). Unlike Sinefungin, none of the tested compounds inhibited the human N⁷-MTase used as a reference enzyme to address the specificity issue. Altogether, the results suggest that we developed an authentic strategy by implementing an initial screening on the GT activity instead of the MTase activity. The GT activity carried out by VEEV nsP1 is an original viral activity and it is likely the reason that increased the probability to identify specific compounds.

In parallel to the work performed on the PCL® screened library and the studies performed on the selected series we were committed to develop the series of MADTP compounds. To date, the more potent MADTP compound already reported is MADTP-372 and presents an IC₅₀ in the range of 10 µM (Gigante *et al.*, 2014, Gigante *et al.*, 2017). Therefore, to accomplish our aim we have screened several MADTP analogues on the VEEV nsP1 GT reaction by determining the IC₅₀ using the same WB assay. Among the 25 MADTP compounds tested we have identified 12 highly potent MADTP compounds bearing an IC₅₀ below 5 µM.

It is importantly to note that none of these compounds show structure similarities when compared to compounds selected from a GTP competition assay on CHIKV nsP1 (Feibelman *et al.*, 2018).

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

8.5 Conclusions

THIS PAGE IS INTENTIONALLY LEFT BLANK

The capping mechanism of alphavirus mRNA is an authentic target for the development of molecules with antiviral effect. Inhibition of capping carried out by nsP1 would block the translation of viral proteins and consequently result in defective viral replication. Therefore, the discovery of effective, specific and potent inhibitors of alphavirus would allow to overcome the limitation of lack of therapeutics to treat alphavirus infections.

To overcome the gap of structural information of the alphavirus nsP1 we developed an enzyme-based assay to select among approved compounds small molecules inhibiting one of the capping reactions using VEEV nsP1 as model. The strategy implemented to develop the assay was effective to screen more than 1200 compounds and select 18 compounds for further analysis. The analysis permitted the selection of 3 series of compounds that were further characterized. Although the number of analogues tested for series 1 and 3 were low, for series 2 it was possible to identify a more potent analogue than the head of series. This analogue was altanserin and more works are required to better understand the mode of action of the selected series of compounds. The work performed for the selected series of compounds suggest that the selected compounds work through an unconventional mechanism of action when compared to the reference molecules.

Another branch of our work in search for outstanding compounds targeting the alphavirus capping was to contribute for the development of more potent MADTP compounds. Our work pinpointed 12 new MADTP compounds presenting an IC_{50} below the 5 μ M.

Indeed, our results can open new windows of opportunities for the development of novel anti-VEEV compounds.

THIS PAGE IS INTENTIONALLY LEFT BLANK

**9. Understanding the catalytic mechanism behind *Alphavirus*
Macro domains**

THIS PAGE IS INTENTIONALLY LEFT BLANK

9.1 Crystallographic structures of Getah virus *Macro* domain reveal multiple conformation of ADP-ribose in the binding pocket

THIS PAGE IS INTENTIONALLY LEFT BLANK

Macro domains of Alphaviruses are ADP-ribose binding modules that are able to bind ADP-ribose and (or) poly-ADP-ribose chains carried by mono or poly ADP-ribosylated proteins (Li et al., 2016, Egloff et al., 2006, Malet et al., 2009). Alphaviruses can use the *Macro* domain function to evade the host immune surveillance, and thus are attractive targets for drug design aiming to new therapeutics against alphavirus infections.

In our present work we wish to gain insights into the catalytic mechanism played by alphavirus *Macro* domains. Through exploration of the sequence diversity of alphavirus *Macro* domains, we aimed to find substantial information on the catalytic mechanism. I report here the cloning and expression of the *Macro* domain recombinant gene from Getah virus (strain M1), as well as purification and crystallization of the corresponding product that were performed by myself. Thermal shift assays and thermophoresis were performed by myself, as well. X-ray crystallography was the technique used to obtain three-dimensional structures of the native Getah *Macro* domain and its complex with ADP-ribose in different conformations, providing insights into the deMARylation mechanism. Attempts to reverse the deMARylation reaction were performed by myself, and some of the results are presented here.

I have used beam-time available from ESRF and SOLEIL synchrotrons for data collection from crystals produced in-house. The native Getah *Macro* domain was collected during my first experiment at the ESRF synchrotron, experiment under the proposal MX-1696 which was carried out on the beamline ID23-1 on the 11th -12th February 2016. From this experiment, from 4 crystals taken to ESRF only 3 crystals lead to collection of data sets. The first three-dimensional structures of Getah *Macro* domain obtained in complex with ADP-ribose with the distal ribose in open conformation were collected in my second trip to ESRF, experiment under the proposal MX-1786, which was carried out on the beamline ID30B on 22nd -23rd April 2016. From this experiment 9 data sets were collected from 9 of the 15 crystals taken to ESRF.

In my fifth trip to a synchrotron, I have collected at the experiment carried out on beamline PROXIMA-2A on 31st March to 1st April of 2017, under the framework of proposal n^o 20160886, the data sets that lead to the structure of Getah *Macro* domain in complex with ADP-ribose in open conformation and covalently linked to the cysteine 34. For this experiment, 10 crystals of this project were taken to SOLEIL and only from 9 were collected data sets. No visible electron density was observed for glutamic acid or aspartic acid and, the results that I have obtained from the data-sets collected at SOLEIL PROXIMA-2A, lead me to test several different conditions of soaking and co-crystallization for Getah *Macro* domain with ADP-ribose and glutamic acid or aspartic

acid. Thus, in my sixth trip to a synchrotron I went to ESRF for an experiment under the proposal MX-1906 which was carried out on the beamline ID30A-3 on 31st January to 1st February of 2018. For this experiment, 55 crystals were taken to ESRF ID30A-3 and all resulted in collected data sets.

The three-dimensional structure determination was performed by myself and Dr. Gerlind Sulzenbacher.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Crystallographic structures of Getah virus *Macro* domain reveal multiple conformations of ADP-ribose in the binding pocket.

Ferreira-Ramos Ana S.^{A #}, Sulzenbacher Gerlind^{A #}, Papageorgiou Nicolas^A, Canard Bruno^A, Coutard Bruno^{*A, B}

^AAix Marseille Université, CNRS, AFMB UMR 7257, Marseille, France

^BUnité des Virus Emergents (UVE: Aix-Marseille Univ-IRD 190-Inserm 1207-IHU Méditerranée Infection) Marseille, France

[#]The authors equally contributed to the work

*Correspondence should be addressed to Bruno Coutard

E-mail: bruno.coutard@univ-amu.fr

KEYWORDS

Alphavirus; replication; nsP3; *Macro* domain; ribosylation, ADP-ribose; X-domain, ADPR

ABSTRACT

Alphaviruses are important (re-) emerging arboviruses of serious public health concern. Their nsP3 gene product has been reported as one of the major key players during viral replication. NsP3 is organized in three domains, a *Macro* domain at the N-terminal, a zinc binding domain (ZBD), and a C-terminal hypervariable region (HVR). The *Macro* domain is essential at both early and late steps of the replication cycle through ADP-ribose (ADPr) binding and de-ribosylation of ribosylated cellular proteins on surface Asp or Glu residues. The molecular mechanism of de-ribosylation by alphavirus *Macro* domain remains to be elucidated. In order to better understand this mechanism, we initiated a structure-based study for Getah virus (GETV) *Macro* domain which shows a peculiar substitution in one of the conserved residues in the catalytic loop. We produced and purified the GETV *Macro* domain for crystallographic studies and characterized several conformations adopted by ADP-ribose in the binding site. Together, these various conformations observed in the crystallographic structure may represent several snapshots of the de-ribosylation mechanism, highlighting new residues to be further characterized.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

INTRODUCTION

Alphaviruses are arthropod-borne viruses among which several, such as Chikungunya virus (CHIKV) and Venezuelan Equine Encephalitis virus (VEEV), are emerging or re-emerging viruses. According to the 7th report of the International Committee on Taxonomy of Viruses (ICTV), the genus *Alphavirus* is organised into 31 species and at least ten of them are human pathogens (Gould *et al.*, 2010b). Alphavirus can also be classified based on the geographical distribution, with a group of Old World (OW) alphaviruses represented by CHIKV, a group of New World (NW) viruses prototyped by VEEV, and several species corresponding to viruses that most likely appeared from recombination events between OW and NW viruses (Weaver *et al.*, 1993) or viruses of marine origin.

Their genome is a positive strand RNA molecule which is usually 11 to 12 kilobases long. The genome carries two open reading frames (ORFs). The first ORF can be directly translated from the genomic RNA in P123 and P1234 polyproteins which are sequentially processed into four non-structural proteins (nsP) named nsP1 to nsP4. The proteolysis intermediates and the final nsPs constitute the transcription/replication complexes (TRC) organised in spherules at the plasma membrane. Once the TRC is organised, it transcribes a sub-genomic mRNA encoding the structural proteins of the viral particle. Enzymatic functions involved in the viral transcription/replication are unambiguously associated to their corresponding nsPs. The peculiar viral mRNA capping is mediated by nsP1 and nsP2 (Ahola & Kaariainen, 1995, Mi *et al.*, 1989, Li *et al.*, 2015, Vasiljeva *et al.*, 2000), while helicase and protease domains are carried by nsP2 (Das *et al.*, 2014, Strauss *et al.*, 1992, Golubtsov *et al.*, 2006); Nsp4 contains the RNA-dependent RNA polymerase (RdRp), involved in (-) and (+) strand RNA synthesis when associated to other nsPs in membrane spherules (Rubach *et al.*, 2009, Pietila *et al.*, 2018).

Nsp3 has been for a long time the least understood nsP but recent studies revealed some of its functions in the viral replication (Gotte *et al.*, 2018). Nsp3 is organised in three domains: a *Macro* domain (MD) at the N-terminal, a zinc binding domain (ZBD) and a C-terminal hypervariable region (HVR). HVR plays a key role in the viral replication by interacting with several partners among which Ras GTPase-activating protein-binding protein (G3BP) and SH3 containing proteins (Meshram *et al.*, 2018). The N-terminal *Macro* domain binds ADP-ribose (ADPr) binding and has de-ribosylation activity, thought to be essential in both early and late replication steps (Abraham *et al.*, 2018). Ribosylation, a posttranslational modification adding single or poly-ADP-ribose chains onto cellular proteins, is activated during infection by CHIKV virus in an α/β -IFN

independent manner when using the mouse motor neuron-like hybrid (NSC34) cell line (Abraham *et al.*, 2018). It is postulated that de-mono-ribosylation is an alphavirus countermeasure to cellular innate immunity mediated by the viral *Macro* domain (Leung *et al.*, 2018).

The structure of the *Macro* domain from alphavirus species VEEV, CHIKV, Sindbis virus (SINV), and Mayaro virus (MAYV) has been determined using X-ray crystallography and/or NMR (Malet *et al.*, 2009, Shin *et al.*, 2012, Melekis *et al.*, 2015). Structure and sequence analysis together with functional characterization allowed assignment of alphavirus *Macro* domain into a group of domains prototyped by *Macro* D2 (Rosenthal *et al.*, 2013). Viral *Macro* domains were originally found to have ADP-ribose-1"-phosphate phosphatase (A1"Pase) (Putics *et al.*, 2006, Malet *et al.*, 2009). More recently, it was shown that alphavirus *Macro* domains are able to de-ribosylate Asp or Glu acceptors (Li *et al.*, 2016, McPherson *et al.*, 2017, Ecke *et al.*, 2017). Some of the residues involved in ADP-ribose binding and de-ribosylation have been characterized (Leung *et al.*, 2018). However, the molecular mechanism of de-ribosylation by alphavirus *Macro* domain remains to be clarified.

In order to better understand this mechanism, we initiated a structure-based study of the Getah virus (GETV) *Macro* domain. GETV is an alphavirus isolated for the first time in Malaysia in 1955 from *Culex* spp. mosquitoes. GETV is geographically distributed from Asia to the north of Australia and infects mainly horses (Fukunaga *et al.*, 2000). Interestingly, the GETV M1 isolate has oncolytic effects, with increased replication potency in cancer cell lines characterized by loss of expression of zinc-finger antiviral protein (ZAP), a catalytically deficient poly ADP-ribose polymerase (Lin *et al.*, 2014). Furthermore, knocked down of the ZAP gene in M1 resistant cells lead to increased sensitivity, confirming the central role of the ZAP mediated host-virus interaction (Hu *et al.*, 2018). The sequence of GETV *Macro* domain shows a peculiar substitution for one of the largely conserved residues in the catalytic loop (Fig. 36). Given these observations, we speculated that the GETV phenotype could be, at least partially, related to sequence specificity of GETV *Macro* Domain. We thus produced and purified the GETV *Macro* domain for crystallographic studies and documented several conformations adopted by ADP-ribose in the binding site. Taken together, this ensemble of conformations might represent several snapshots of the de-ribosylation mechanism, highlighting new residues to be further characterized.

MATERAIAL AND METHODS

Protein production and purification

The DNA coding sequence of nsP3 *Macro* domain (amino acids [a.a.] 1 to 160) of Getah virus (strain M1, GenBank: ABK32031.1) was optimized for expression in *E. coli* and synthesized by ThermoFischer. The coding sequence was cloned in pDest14 using the “Gateway” cloning procedure (ThermoFischer). A hexa-histidine (6-His) coding sequence was added at the 3' end in order to produce the GETV *Macro* domain in fusion with a C-terminal 6-His tag. A short sequence of two codons coding for methionine and lysine were added to the 5' end in order to allow the translation in *Escherichia coli*. The bacterial growth conditions for optimal expression of GETV *Macro* domain were obtained from an incomplete factorial expression screening (Berrow *et al.*, 2006). GETV *Macro* domain was produced in *E. coli* Rosetta (DE3) pLysS cells (Novagen). Bacteria were grown at 37°C with shaking (200 rpm) in Terrific Broth (TB) medium (Sigma-Aldrich) containing 34 µg·mL⁻¹ chloramphenicol and 100 µg·mL⁻¹ ampicillin. When OD_{600nm} reached a value of 0.4, expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Incubation temperature was dropped to 25°C and for overnight expression. Cells were then harvested by centrifugation at 4,000 g for 15 minutes and each 1 litre of culture pellets were resuspended in 50 mL of lysis buffer (50 mM Tris buffer pH 8.0, 300 mM NaCl, 10 mM Imidazole, 5 % Glycerol, 0.1 % Triton X-100 and 1 tablet of complete EDTA-free protease inhibitor cocktail [Roche]). The resuspended pellets were stored at -80 °C until the time of use.

Pellets were then thawed and 0.25 mg·mL⁻¹ of lysozyme (Sigma-Aldrich), 10 µg·mL⁻¹ of DNase I (Sigma-Aldrich) and 1 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich) were added. The samples were incubated at 4 °C for 30 min and then sonicated. Samples were then centrifuged at 13,000 g for 1 hour. The supernatant was loaded onto a 5 mL His prep column (GE Healthcare) for Immobilized Metal Affinity Chromatography (IMAC) using the ÄKTA Xpress system (GE Healthcare). After loading of the supernatant, the column was washed with a buffer containing 50 mM Tris, 300 mM NaCl, 50 mM imidazole, pH 8.0. The proteins were then eluted in 50 mM Tris, 300 mM NaCl, and 250 mM imidazole, pH 8.0. Size exclusion chromatography was then performed on a Superdex 200 HiLoad 16/600 column (GE Healthcare) preequilibrated in 20 mM HEPES, 300 mM NaCl, pH 7.4. The protein was then concentrated up to 14 mg·mL⁻¹ using an Ultracel regenerated cellulose membrane with a 3 kDa molecular weight cut-off (Amicon Ultra-15 Centrifugal Filter, Merck Millipore).

Purity and quality assessment of the purified protein

The purity was assessed on SDS-PAGE gels stained with Coomassie blue and quality of the protein was additionally checked by thermal shift assay prior to crystallization (Geerlof *et al.*, 2006). To do so, thermal shift assay (TSA) was performed using the quantitative PCR machine ICycler IQ (Bio-Rad) and in a 96-well thin-wall PCR plate (Bio-Rad). The protein at a final concentration of 0.3 mg·mL⁻¹ was tested with the following concentrations of ADP-ribose: 1 mM, 0.5 mM, 0.25 mM, 0.125 mM, 0.063 mM, 0.031 mM, 0.016 mM and 0 mM. The assays were performed following previously described TSA experiments and using the quantitative PCR machine ICycler IQ (Bio-Rad) and in a 96-well thin-wall PCR plate (Bio-Rad). The proteins at a final concentration of 0.3 mg·mL⁻¹ were mixed with the ligand and with a SYPRO orange solution at concentrations recommended by the manufacturer (ThermoFisher Scientific) in a final volume of 25 µL. The plates were sealed with Optical-Quality Sealing Tape (Bio-Rad). Accumulative steps of temperature from 20 to 90 °C were applied to the samples. The increase of the fluorescence emitted by the probe that binds the exposed hydrophobic regions of the denatured protein was used to monitor the denaturation of the protein. A melting temperature (T_m) was also calculated as the mid-log of the transition phase from the native to the denatured protein using a Boltzmann model (Origin software).

Crystallization, co-crystallization and crystal soaking experiments

Initial crystallization trials were carried out with GETV *Macro* domain at 14 mg·mL⁻¹ using the commercial screens Wizard™ Classic 1 & 2 HT96, Structure 1 & 2 HT96, and Stura FootPrint Screen 48 conditions (Molecular Dimensions Limited) in 3-well sitting-drop Swissci crystallization plates (TPP Labtech). 100 nL of GETV *Macro* domain were added to 100, 200 or 300 nL of the crystallization solutions using a Mosquito Robot™ (TTP Labtech). The conditions where crystal hits were obtained were then optimized using either 96-well Swissci crystallization plates or 24-well hanging drop Limbro plates.

Co-crystallization experiments were set-up with GETV *Macro* domain complemented with 1) ADP-ribose, 2) ADP-ribose and glutamic acid or 3) ADP-ribose and aspartic acid, with final concentrations of 3 mM for ADP-ribose, 3 mM or 50 mM for glutamic acid and 3 mM or 30 mM for aspartic acid. In general, crystals appeared after one day and were grown for approximately two weeks. Soaking experiments were performed with crystals of GETV *Macro* domain co-crystallized with ADP-ribose by adding an equal volume of mother liquor supplemented with 30 mM of aspartic acid or 50 mM of aspartic acid to the crystallization droplet followed by 2-3 hours or overnight incubation at 20 °C. Additional

soaking experiments were made for crystals with apo GETV *Macro* domain by adding 0.5 μL of an ADP-ribose stock solution of 10 mM directly into the 2 μL crystallization drops (final concentration of 3 mM for ADP-ribose) and incubating the droplets for 2-3 hours or overnight at 20 °C. We also performed soaking experiments with a final concentration of 3 mM (or 1.5 mM) of ADP-ribose together with glutamic acid or aspartic acid and using the same concentrations and incubation times for the amino acids as described above. Afterwards crystals were then flash-cooled in liquid nitrogen using 20% of glycerol in the mother liquid as cryo-protectant.

Data collection and three-dimensional structure determination

Diffraction intensities were recorded on beamlines ID23-1, ID30-B and ID30A-3 at the European Synchrotron Radiation Facility (Grenoble, France) and on beamline Proxima-2 at Soleil Synchrotron (Gif-sur-Yvette, France).

Indexing and integration of the different datasets was performed using MOSFLM (Battye et al., 2011) or XDS (Kabsch, 2010). Data were scaled and merged using the CCP4 (Winn et al., 2011) suite of programs Pointless (Evans, 2006), Aimless (Evans & Murshudov, 2013) and Truncate (French & Wilson, 1978). Random sets of approximately 5 % of reflections, depending on the resolution limit, were set aside for FreeR cross-validation purposes. Where data-sets of ligand complexes were in the same space group as the native data set, the composition of cross-validation data sets was systematically taken over from the parent data set. The structure of native *Macro* domain GETV was determined by molecular replacement with MolRep (Vagin & Teplyakov, 1997) using the *Macro* domain of Chikungunya virus (CHIKV *Macro* Domain, PDB 3GPG) as a search model. The structure of GETV *Macro* domain in complex with ligands was determined either by molecular replacement using the native GETV *Macro* domain as a search model or by difference Fourier synthesis.

Refinement was performed using REFMAC (Murshudov et al., 1997), interspersed with cycles of manual model adjustments with COOT (Emsley et al., 2010). Ligands were fitted into unbiased $F_o - F_c$ difference electron density maps calculated after 10 cycles of rigid body refinement. Hydrogens were added in the riding position. Coordinates and restraints for ADP-ribose in the close conformation were retrieved from the CCP4 ligand dictionary and a model and restraints for ADP-ribose in the open conformation were generated with the CCP4 Monomer Library Sketcher. Model quality was assessed with internal modules of Coot and with the Molprobtity server (Chen et al., 2010). Crystallographic models are of good quality with 99.4–100% of the residues in favoured

regions of the Ramachandran plot without outliers. Data collection and refinement statistics are summarized in Table 6 with representative electron density in figures 38, 40, 42 and 43. Figures representing structural renderings were generated with the PyMOL Molecular Graphics System (DeLano, W.L. The PyMOL Molecular Graphics on <http://www.pymol.org/>). Sequence alignments were made using Clustal omega (Sievers & Higgins, 2014) and graphical rendering of the alignments, considering structural information, were made with ESPript (Gouet *et al.*, 2003, Robert & Gouet, 2014). The atomic coordinates and structure factors will be soon deposited in the Protein Data Bank for native GETV *Macro* domain, GETV *Macro* domain with ADP-ribose in closed conformation “pose 1”, GETV *Macro* domain with ADP-ribose in closed conformation “pose 2”, GETV *Macro* domain with ADP-ribose in open conformation, GETV *Macro* domain with ADP-ribose open ring in double conformation and GETV *Macro* domain with ADP-ribose covalently bound.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

RESULTS AND DISCUSSION

Protein production, crystallization and structure determination.

The sequence motif corresponding to the putative *Macro* domain on GETV nsp3 sequence (GenBank reference ABK32032.1) is encompassing aa 1333 to 1492 (positions in the polyprotein; hereafter numbered 1-160 for convenience). Its sequence identity to alphavirus homologues with known structures such as CHIKV and VEEV *Macro* domains is 69% and 57%, respectively (Fig. 36). *Macro* domains harbour the consensus motif Gly(Asp/Gly/Gly)Gly(Val/Leu) (Li *et al.*, 2016). A focused analysis of these residues (amino acids 26 to 33) of *Macro* domains revealed that the consensus motif is followed by a togavirus-specific cysteine. It can be also noticed that GETV *Macro* domain is an exception regarding the consensus motif, as its sequence harbours a serine instead of a glycine at the first position of the consensus motif. In order to evaluate the importance of this special feature, we endeavoured a structural study of GETV *Macro* domain by X-ray crystallography.

The recombinant GETV *Macro* domain was produced in *E. coli* and purified under non-denaturing conditions. After purification, the integrity of the protein was controlled by thermal shift assay (TSA). TSA experiments shows that the protein could be denatured by heat in a classical folded-to-denatured transition phase (Fig. 37) with a melting temperature (T_m) of 46.5 ± 0.2 °C. Addition of 0.5 mM ADP-ribose to the protein solution leads to a 10 °C increase of the T_m suggesting that the recombinant GETV *Macro* domain is folded and has the ability to bind ADP-ribose (Fig. 37). The study of the structure of the di-manganese mono-ADP-ribosylhydrolase DraG in complex with a trapped reaction intermediate, Lys-ADP-ribose, originating from a neighbouring DraG chain in the crystal (Berthold *et al.*, 2009) prompted us to test glutamic or aspartic acids in the crystallization trials because ribosylated glutamate and aspartate residues are substrates of alphavirus *Macro* domain. Glutamic or aspartic acids were thus added with or without ADP-ribose to co-crystallization or soaking solutions at five to eight-times higher concentrations than that of ADP-ribose. All the optimal crystallization and co-crystallization solutions converged to the following buffer composition: 0.2 M imidazole malate pH 5.9 ± 0.2 and 34 ± 4 % of PEG 4K. The crystallization and soaking procedures that led to the structures discussed in this study are summarize in Table 5, and data collection and refinement statistics are presented in Table 6.

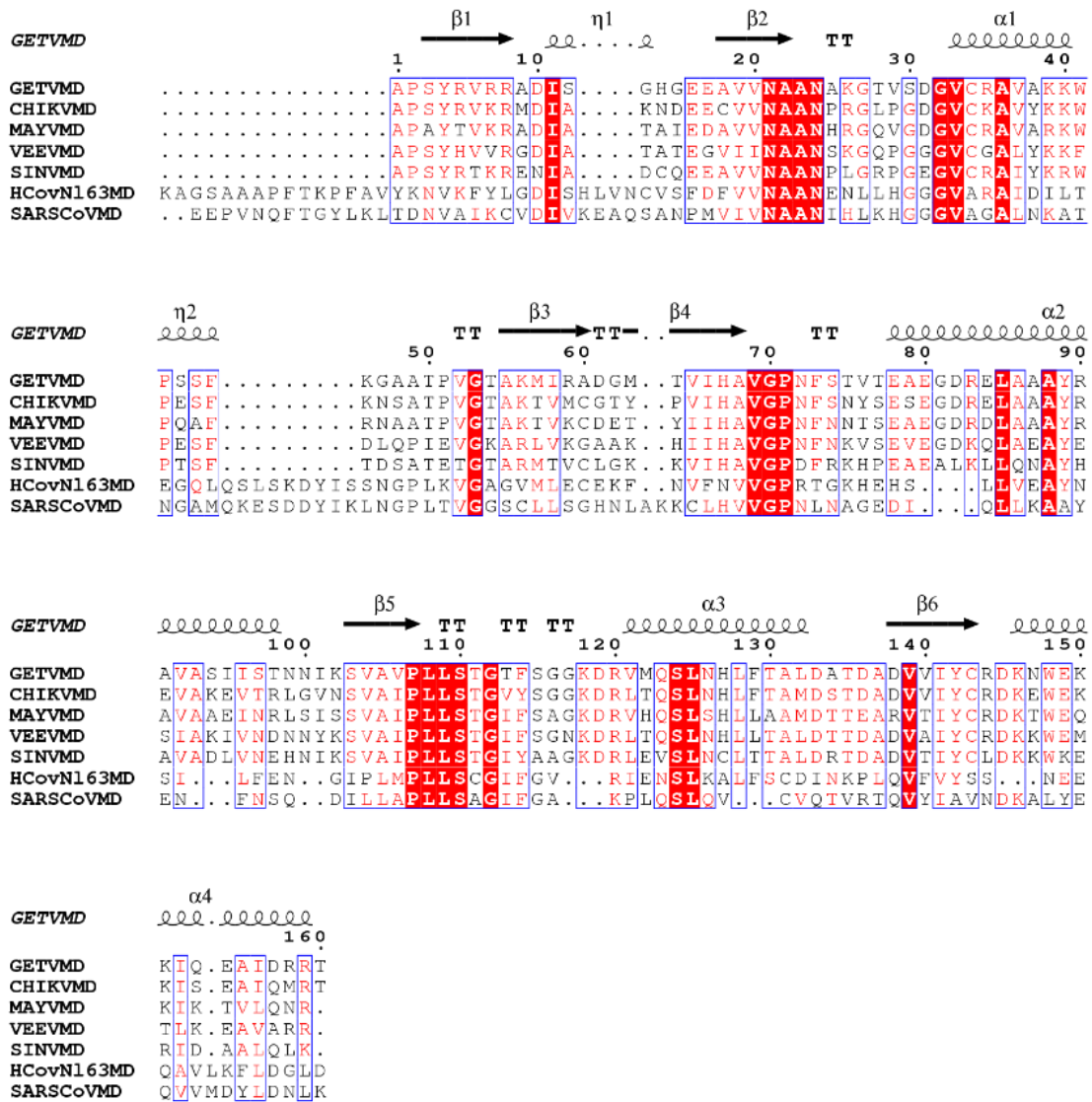


Figure 36. Sequence alignment for GETV Macro domain against other viral Macro domains. The OW group from the genus *Alphavirus* is represented in the alignment by CHIKV, MAYV and SINV Macro domains, while NW group is represented by VEEV Macro domain. The human Coronavirus NL63 Macro domain as well as SARS Coronavirus are represented in the alignment. The alignment picture was obtained using ESPrnt server and residues highlighted with red boxes are strictly conserved, red residues represent similarity in a group, and blue frame represent similarity across groups. Secondary structure elements from the GETV Macro domain crystal structure obtained is represented above the alignment.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

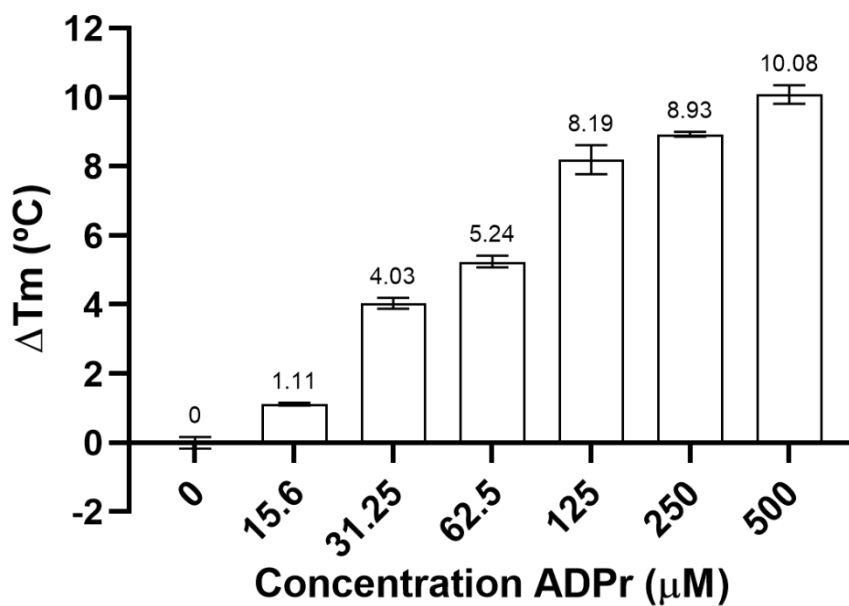


Figure 37. Effect of ADP-ribose concentrations on the thermostability of GETV *Macro* domain. Several concentrations of ADP-ribose at the μM range, presented in x axis, were tested on the thermostability of GETV *Macro* domain. For each concentration of ADP-ribose titrated onto the GETV *Macro* domain the melting temperature was calculated (T_m). The changes in T_m respective to the GETV *Macro* domain without ADP-ribose were determined and are presented in the graphic Y axis.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Table 5. Summary of the crystallization and soaking procedures used to produce the crystals of GETV *Macro* domain that led to the structures discussed in this work.

Data set	Method/Component	Crystallization solution	Conformation of ADP-ribose
Xtal-4	Crystallization	0.2 M Imidazole, Malate pH=6 and	No ADP-ribose
	GETV <i>Macro</i> domain 16 mg·mL ⁻¹	30 % PEG 4K	
Xtal-13	Co-crystallization	0.2 M Imidazole	ADP-ribose with open distal ribose
	GETV <i>Macro</i> domain 13 mg·mL ⁻¹ , 2.4 mM ADP-ribose	Malate pH=5.9 and 38 % PEG 4K	
Xtal-15	Crystallization	0.2 M Imidazole	ADP-ribose with open distal ribose
	GETV <i>Macro</i> domain 13 mg·mL ⁻¹	Malate pH=6.0 and 38 % PEG 4K	
Xtal-44	Soaking: 2.4 mM ADP-ribose		ADP-ribose with open distal ribose
	Co-crystallization	0.2 M Imidazole	
Xtal-44	GETV <i>Macro</i> domain 13 mg·mL ⁻¹ , 3 mM ADP-ribose and 3 mM aspartic acid	Malate pH=5.9 and 38 % PEG 4K	Covalent link Cys34-distal ribose of ADP-ribose
	Co-crystallization	0.2 M Imidazole	ADP-ribose with open distal ribose (2)
Xtal-55	GETV <i>Macro</i> domain 13 mg·mL ⁻¹ , 3 mM ADP-ribose and 50 mM glutamic acid	Malate pH=5.7 and 30 % PEG 4K	
Xtal-62	Co-crystallization	0.2 M Imidazole	ADP-ribose with closed distal ribose in "Pose 2"
	GETV <i>Macro</i> domain 13 mg·mL ⁻¹ , 3 mM ADP-ribose and 50 mM glutamic acid	Malate pH=6,0 and 30 % PEG 4K	
Xtal-68	Co-crystallization	0.2 M Imidazole	ADP-ribose with open distal ribose in double conformation
	GETV <i>Macro</i> domain 13 mg·mL ⁻¹ , 3 mM ADP-ribose and 30 mM aspartic acid	Malate pH=6 and 32 % PEG 4K	
Xtal-72	Co-crystallization	0.2 M Imidazole	ADP-ribose with open distal ribose (1)
	GETV <i>Macro</i> domain 13 mg·mL ⁻¹ , 3 mM ADP-ribose and 30 mM aspartic acid	Malate pH=5.9 and 34 % PEG 4K	
Xtal-91	Co-crystallization	0.2 M Imidazole	ADP-ribose with closed distal ribose in "Pose 1"
	GETV <i>Macro</i> domain 13 mg·mL ⁻¹ and 3 mM ADP-ribose	Malate pH=5.9 and 34 % PEG 4K	
	Soaking: 15 mM aspartic acid		

Table 6. Data collection and refinement statistics

	Native	Closed ribose Pose 1	Closed ribose Pose 2	Open ribose	Open ribose double conformation	Covalent bond ADP-ribose /Cys
Data collection						
Beam line	ESRF ID23-1	ESRF ID30A3	ESRF ID30A3	ESRF ID30A3	ESRF ID30A3	SOLEIL Proxima2
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	C2
Cell dimensions						
<i>a, b, c</i> (Å)	46.55, 71.36, 94.95	46.73, 71.57, 98.69	46.88, 71.65, 98.80	46.67, 71.44, 98.97	46.70, 71.45, 99.50	64.01, 46.80, 51.00; β=104.16
Resolution (Å)	41.80-2.0 (2.05- 2.00)	42.23-2.05 (2.11-2.05)	42.35-1.70 (1.73-1.70)	40.68-1.85 (1.89-1.85)	40.83-1.60 (1.63- 1.60)	37.37-1.45 (1.47- 1.45)
<i>R_{merge}</i>	0.134 (0.748)	0.144 (1.328)	0.062 (1.012)	0.078 (1.221)	0.059 (1.268)	0.046 (0.826)
<i>R_{pim}</i>	0.079 (0.448)	0.073 (0.660)	0.040 (0.664)	0.039 (0.591)	0.034 (0.710)	0.019 (0.484)
CC(1/2)	0.989 (0.603)	0.994 (0.468)	0.999 (0.527)	0.998 (0.627)	0.998 (0.559)	0.999 (0.352)

	Native	Closed ribose Pose 1	Closed ribose Pose 2	Open ribose	Open ribose double conformation	Covalent bond ADP-ribose /Cys
I/σ	7.3 (2.1)	8.5 (1.5)	15.6 (1.6)	13.1 (1.40)	13.1 (1.2)	20.6 (2.1)
Completeness (%)	99.8 (99.8)	98.6 (99.0)	97.5 (97.4)	99.6 (99.8)	98.7 (97.5)	99.9 (100)
Redundancy	4.5 (4.3)	4.8 (5.1)	5.0 (5.3)	4.8 (5.1)	4.0 (4.1)	6.6 (6.5)
Wilson B (Å ²)	13.3	25.50	18.9	18.34	15.36	17.39
Refinement						
Resolution (Å)	57.05-2.00	40.66-2.05	40.70-1.70	33.98-1.85	36.41-1.60	37.40-1.45
No. reflections	20700	19838	34198	27122	41597	24542
R_{work}	19.39 (27.70)	19.31 (32.30)	18.11 (34.20)	16.84 (30.90)	16.94 (48.50)	17.01 (52.60)
R_{free}	23.11 (28.70)	23.55 (32.60)	20.75 (33.50)	19.90 (34.00)	19.35 (51.70)	19.63 (54.70)
No. atoms						

	Native	Closed ribose Pose 1	Closed ribose Pose 2	Open ribose	Open ribose double conformation	Covalent bond ADP-ribose /Cys
Protein	2386	2386	2388	2388	2388	1210
ADP-ribose	-	72	72	72	72	36
Water/ligands	190/11	185/-	231/8	190/24	279/8	186/4
B-factors (Å ²)						
Protein	12.11	37.92	27.57	46.05	26.41	19.67
ADP-ribose	-	48.20	29.20	48.32	27.06	19.50
Water/ligands	34.72/47.06	36.94/-	33.42/40.84	50.42/59.22	34.40/27.28	28.98/28.99
R.m.s. deviations						
Bond lengths (Å)	0.009	0.007	0.005	0.007	0.006	0.010
Bond angles (°)	1.304	1.458	1.250	1.454	1.446	1.625
Ramachandran favoured	99.68	99.68	100	99.68	99.68	99.38

	Native	Closed ribose Pose 1	Closed ribose Pose 2	Open ribose	Open ribose double conformation	Covalent bond ADP-ribose /Cys
Data set	Xtal-4	Xtal-91	Xtal-62	Xtal-72	Xtal-68	Xtal-44
PDB ID	6QZU	6R0F	6R0G	6R0T	6R0P	6R0R

Values in parentheses are for the highest-resolution shell.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Structure of the GETV *Macro* domain and comparison with other (alphavirus) *Macro* domains

The structure of the GETV *Macro* domain was determined at 2.0 Å resolution by molecular replacement using CHIKV *Macro* domain (3GPG) as a template. The crystals of the GETV *Macro* domain belong to the space group P2₁2₁2₁ and two molecules are present in the asymmetric unit. All residues from Ala1 to Thr160 are well defined in the electron density and the model has excellent stereochemistry with 99.2% of side-chain rotamers in the favoured conformation and 99.7% of residues in the favoured Ramachandran plot regions. The refined structure contains one diethylene glycol, one ethylene glycol molecule and 190 solvent molecules.

As expected, the structure of GETV *Macro* domain consists of a central twisted six-stranded β sheet (strand order β1, β6, β5, β2, β4, β3) sandwiched between one α-helix (α1) at one side and two α-helices (α2 and α3) at the opposite site (Fig. 38). The two GETV *Macro* domain chains present in the asymmetric unit are virtually identical, with an r.m.s.d. between the two chains of 0.22 Å for pairwise alignment of 160 C^α positions. Structural differences between the two chains arise mainly from differences in rotamer conformations of flexible surface exposed Lys and Arg side-chains. A structural

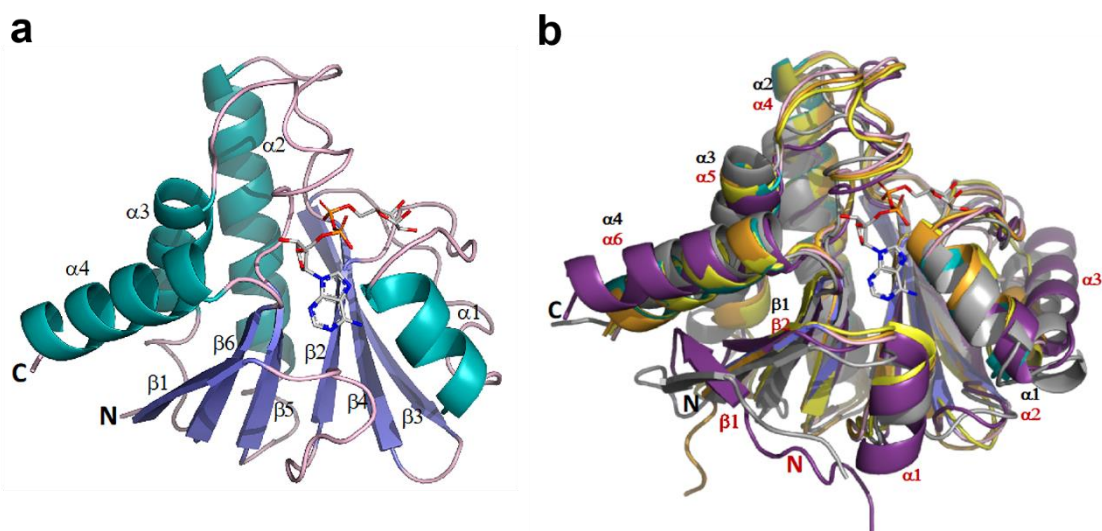


Figure 38. Cartoon representation of the structures for GETV *Macro* domain alone (a) and with homologues (b). (a) Overall structure of native GETV *Macro* domain with APD-ribose depicted as seen in the GETV *Macro* domain with ADP-ribose complex “pose1”. β-sheets are coloured in slate, α-helices in teal and loops in pink. Secondary structure elements and the N- and C-terminal of GETV *Macro* domain are labelled. ADP-ribose is presented in stick-mode, with carbon atoms coloured in grey, oxygens in red, nitrogen in blue and phosphorus atoms in orange. ADP-ribose was added for sake reference for the ADP-ribose binding pocket. (b) Overlap of native GETV *Macro* domain, same colour coding as in (a), APD-ribose depicted as seen in the GETV *Macro* domain with ADP-ribose complex “pose 1” and colour-coded as in (a) with the structures of CHIKV *Macro* domain (3GPO) in orange, VEEV *Macro* Domain (3GQ) in yellow, SARS *Macro* domain (2FAV) in purple and the non-histone domain of the histone variant MacroH2A1.1 (1YD9) in grey. The N- and C-terminus and secondary structure elements are labelled in black for GETV *Macro* domain and in red for SARS *Macro* domain. For clarity, only the first β-sheets are labelled. ADP-ribose was added for sake reference for the ADP-ribose binding pocket.

homology search with the DALI server (Holm & Sander, 1993) revealed closest homology with *Macro* domains from the alphavirus family such as CHIKV (3GPG), SINV (4GUA), VEEV (3GQO) and MAYV (5IQ5) with Z-scores in the range of 32.2 to 27.8 and r.m.s.d. values of 0.7 to 1.4 Å for 158-160 pairwise aligned C α positions.

A second group of close structural homologues, with Z-scores in the range of 21.4 to 19.0 and r.m.s.d values of 1.8 to 2.2 Å for 151-154 pairwise aligned C α positions comprises the structures of human *Macro* domains (histone *Macro*H2A1.1 [5LNC], Parp14 *Macro* domain 1 and *Macro* domain 2 [3Vfq], MacroD2 [4IQY]), and a putative phosphatase (1SPV) and YmdB (5CB5), a O-acetyl-ADP-ribose deacetylase, from *Escherichia coli*. Interestingly, lower structural homologies than for the above protein structures are calculated by DALI for the viral *Macro* domain structures from the coronavirus family (SARS-CoV, 2FAV, Human-CoV NL63, 2VRI, Human-CoV 229E, 3EJG, MERS-CoV, 5DUS, feline-CoV, 3ETI, and Avian infectious bronchitis virus (IBV) strains Beaudette, 3EKE, and M41, 3EWO), with Z-scores in the range of 16.4 to 18.8 and r.m.s.d values of 1.8 to 2.3 Å for 136-143 pairwise aligned C α positions. Finally, more distant structural homologies could be detected for *Macro* domains from *Mycobacterium tuberculosis* (5M3I) and *Thermus aquaticus* (5M3E) (Z-scores around 17) and yeast YMX7 (1TXZ), an ADP-ribose-1"-monophosphatase (Z-score 10.3).

The comparison of the overall structure of GETV *Macro* domain with the structures of *Macro* domains from other alphaviruses shows that the alphavirus *Macro* domain fold is very well conserved, Fig. 38 B. In particular, the C α chains of three Old World alphaviruses GETV *Macro* domain, CHIKV *Macro* domain and SINV *Macro* domain (not shown in Fig. 38 B) superimpose well, as illustrated by the low root mean square deviations along the ~160 amino acids of the domains (r.m.s.d.= 0.7 Å). Only minor structural changes are observed at the level of the β 3 β 4 loop, the r.m.s.d.s between pairwise aligned C α positions of GETV *Macro* domain and the *Macro* domains of the New-World alphavirus VEEV and the Old-World alphavirus MAYV are slightly higher (0.9 and 1.4 Å, respectively). The main structural differences between GETV *Macro* domain and these two structures are in the loops branching structural elements i) α 1 and β 3, ii) β 3 and β 4 and iii) β 5 and α 3. The β 2 α 1 loop, referred to as the catalytic loop, is structurally well conserved among alphavirus *Macro* domain structures. With respect to the non-histone motif of the histone variant *Macro*H2A, from which the *Macro* domain was originally characterized (Chakravarthy *et al.*, 2005), and coronavirus *Macro* domains, chiefly consisting of a central seven-stranded β -sheet flanked on either sides by three α -helices, alphavirus *Macro* domains lack the first β -strand, and α -helices α 1

and $\alpha 2$ (CHIKV annotation) are degenerated to 3_{10} -helices in other alphavirus *Macro* domains structures. They reduce to loops in GETV *Macro* domains, Fig. 38 B.

Structures of the GETV *Macro* domain in complex with ADP-ribose

The early evaluation of the integrity of the protein by monitoring the thermostability using thermal shift assay suggested that addition of 0.5 mM ADP-ribose to the protein solution leads to a 10 °C increase of the T_m (Fig. 37). Thus, it was proposed that in a similar fashion to other alphavirus *Macro* domains, the GETV *Macro* domain is also able to bind ADP-ribose. To prove it, we produced crystals of GETV *Macro* domain in the presence of ADP-ribose. For that experiments, we used co-crystallization or soaking experiments with ADP-ribose. One crystal structure was determined from diffraction data with resolution extending to 2.0 Å and collected from a crystal obtained by co-crystallization of GETV *Macro* domain with 2.4 mM ADP-ribose. Another similar crystal structure was determined with resolution extending to 2.0 Å, collected from a crystal of GETV *Macro* domain which was soaked overnight in 2.4 mM ADP-ribose. Both crystal structures belong to the same space group $P2_12_12_1$ and no significant structural differences were observed between both.

The initial electron density maps were calculated before incorporation of the ligand and the unbiased F_o-F_c difference electron density maps indicated the presence of ADP-ribose (data not shown). Intriguingly, when we incorporated the ligand ADP-ribose with the “classical” conformation, presenting the distal ribose in close conformation, the final weighted difference maps F_o-F_c and $2F_o-F_c$ suggested that the ADP-ribose adopts a different conformation than the “classical” conformation, Fig 39 A. We then calculated new electron density maps after incorporating the ligand ADP-ribose containing the distal ribose in an open conformation: Interestingly, the resulting weighted maps F_o-F_c and $2F_o-$ clearly suggested that, indeed, the distal ribose was not cyclic anymore Fig. 39 A and B.

The ring opening observed here might well be part of the catalytic mechanism of the deMARylation reaction. It suggests that a nucleophile (*eg.*, amino acid side chain or water molecule) is at some point in the vicinity of this scissile bond. In the deMARylation reaction, ADP-ribose moieties grafted to surface Asp of Glu residues of MARylated proteins are first bound by the *Macro* domain, and the ester bond between the carboxyl group and the 3'-hydroxyl of the distal ribose of ADP-ribose is cleaved, releasing a free Asp or Glu on the surface of the target protein, and ADP-ribose. We reasoned that the availability of carboxyl groups might trigger visible conformational changes in our

structures, and that might help to better characterize this still elusive deMARylation mechanism. Consequently, aspartic and glutamic acids were added in the crystallization solutions together with ADP-ribose, several *Macro* domain crystals were grown, and their structure determined. Five different conformations could be observed for ADP-ribose bound to GETV *Macro* Domain.

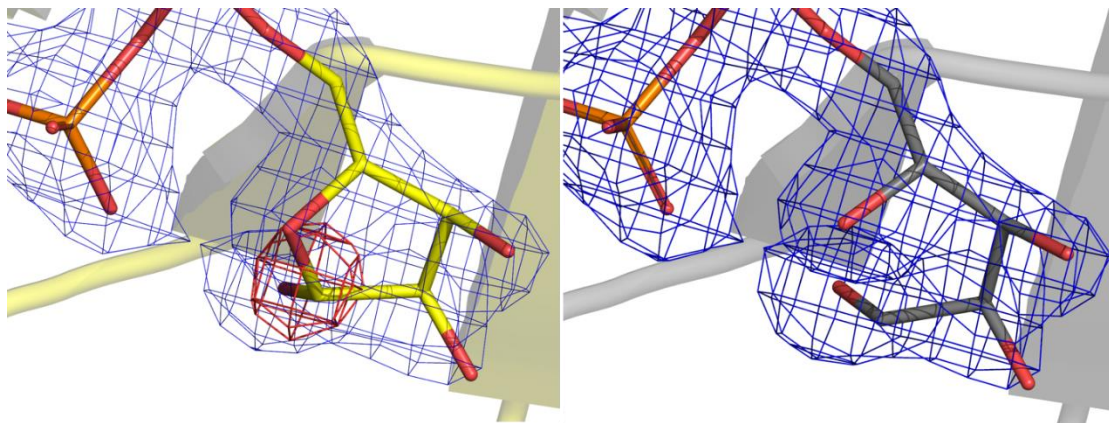


Figure 39. The binding mode of ADP-ribose adopted a new conformation in the distal ribose. From diffraction data collected from Xtal-15, the initial electron density maps were calculated before incorporation of the ligand and the unbiased F_o-F_c difference electron density maps indicated the presence of ADP-ribose (data not shown). **(a)** After incorporation of the ligand ADP-ribose with the distal ribose in close conformation, the final weighted difference maps F_o-F_c (red and green) countered at 3σ level and the $2F_o-F_c$ (blue) countered at 1σ level suggested that distal ribose was not in closed conformation. **(b)** Then, the ligand was incorporated with the distal ribose in open conformation, and the final weighted difference maps F_o-F_c (red and green) countered at 3σ level and the $2F_o-F_c$ (blue) countered at 1σ level supported the hypothesis that the distal ribose was in an open conformation. The cartoon represents the GETV *Macro* domain from Xtal-15 (a) in yellow, for the model with ADP-ribose with the distal ribose in close conformation. (b) the grey colour represents the model of Xtal-15 with the ligand ADP-ribose with the distal ribose in open conformation.

Structure of the *Macro* domain GETV with ADP-ribose complex presenting the distal ribose in “pose 1”

In a first instance we obtained diffraction data extending to 2.05 \AA resolution for GETV *Macro* domain in complex with ADP-ribose from a crystal obtained by co-crystallization of GETV *Macro* domain with 3 mM ADP-ribose and subsequently soaked in a mother-liquor solution containing 15 mM of aspartic acid, and the structure was solved by difference Fourier synthesis. As for the native structure, all residues are well defined in the electron density and the final model has excellent stereochemistry. 185 solvent molecules are present in the refined structure. Clear difference electron density could be observed for ADP-ribose bound to each of the two *Macro* domain chains present in the asymmetric unit. GETV *Macro* domain binds ADP-ribose in a deep cleft and the distal ribose ring adopts a pose as observed in other *Macro* domain ADP-ribose complexes, henceforth called “pose 1”, Fig. 40.

The adenine moiety binds *via* the N-1 atom to the main-chain of Ile11 and *via* the N-6 atom to the side-chain of Asp10 and makes stacking interactions at one side with the side-chains of Ile11 and Val33 and at the other side with the side-chain of Arg144, Fig. 40. The interaction N-6 with Asp10 is well conserved in most *Macro* domain /ADP-ribose complexes and it had been shown to be crucial for ADP-ribose binding in the *Archaeoglobus fulgidus Macro* domain (Karras *et al.*, 2005). At the other hand, Arg144 is not conserved, not even within the alphavirus *Macro* domain family, and its side-chain is rather disordered in all the structures described in this work, indicating that the stacking interaction with the adenine ring might contribute only weakly to the binding energy.

The proximal ribose of ADP-ribose interacts with GETV *Macro* domain through a single hydrogen bond between 3'-OH and the side-chain of Thr111, Fig. 40. It is noteworthy that Trp148, conserved in almost all alphavirus *Macro* domains, protrudes into the ADP-

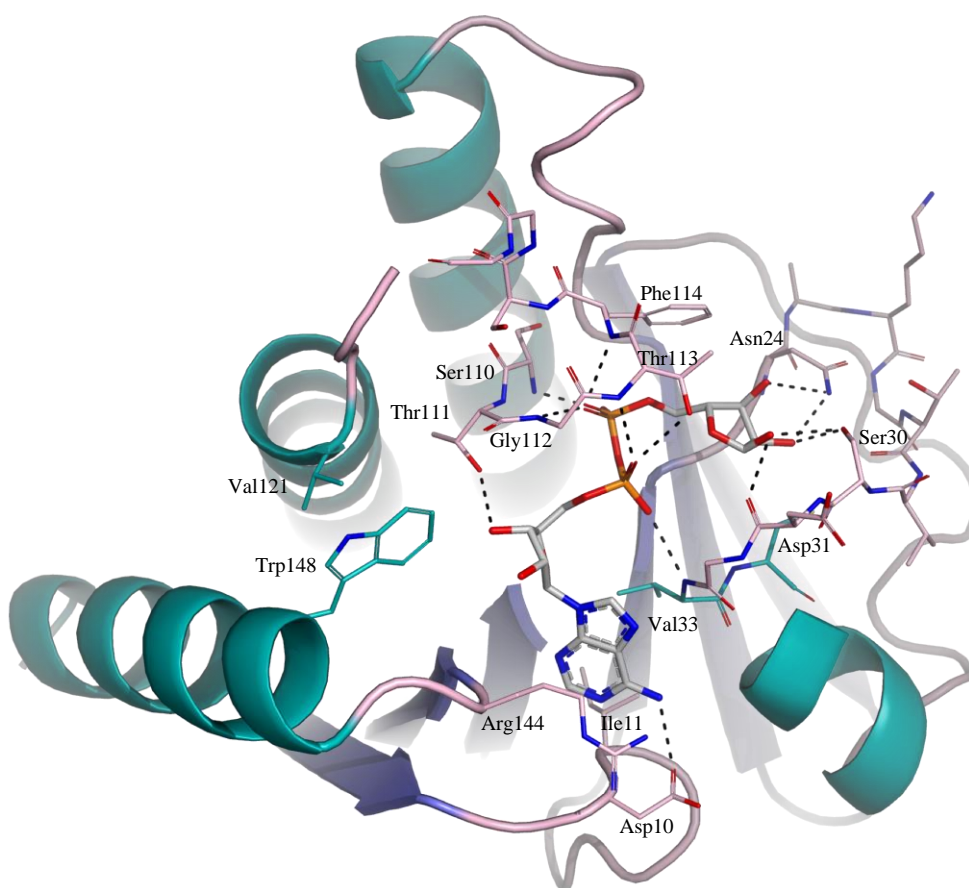


Figure 40. Interaction network between ADP-ribose in “pose 1” with GETV *Macro*. The ADP-ribose binding to GETV *Macro* domain as observed in the complex of GETV *Macro* domain with ADP-ribose in “pose 1”. ADP-ribose and residues interacting with ADP-ribose are depicted in sticks. Atoms of ADP-ribose and carbon atoms of GETV *Macro* domain residues are colour-coded as in Fig. 38.

ribose binding site and makes a steric clash with the 3'-OH, Fig. 40. In the other alphavirus *Macro* domain structures, the side-chains of the equivalent Trp residues are shifted about 1.5-2 Å away from the ADP-ribose binding site with respect to the position

of Trp148, Fig. 40. However, in GETV *Macro* domain, a Val residue pointing from the back towards the Trp148 indole ring, Val121, impedes an equivalent rearrangement of the Trp148 side-chain and this residue appears to be highly dynamic and could not be modelled in a satisfactory fashion in all the structures described in this work, Fig. 40.

Similarly, to what had been observed in other *Macro* domain /ADP-ribose complexes, the phosphate groups of ADP-ribose are lined by the catalytic loop $\beta 2\alpha 1$ and loop $\beta 5\alpha 3$ and tightly coordinated by hydrogen bonds with main-chain atoms of residues Val33, Ser110, Gly112, Thr113 and Phe114 and the side-chain of residue Thr113, Fig. 40. Finally, the distal ribose, present in the anomeric α -configuration, interacts with GETV *Macro* domain through hydrogen bonds contracted between 3''-OH and the side-chain of Asn24, between 2''-OH and the side-chains of Asn24 and Ser30, and between 1''-OH and the side-chain Ser30 and the main-chain of Asp31, and establishes a stacking interaction with the side-chain of Phe114, Fig. 40. The Ser30 of the catalytic $\beta 2\alpha 1$ loop is unique to GETV *Macro* domain and substituted by Gly in other viral *Macro* domains harbouring the so-called glycine-rich motif ${}_{30}\text{Gly}(\text{Asp}/\text{Glu}/\text{Gly})\text{Gly}(\text{Val}/\text{Leu})_{33}$. The tight interactions established between the distal ribose and Ser30 point towards a role in substrate binding or catalysis for this residue, Fig. 40.

Not surprisingly, the most important structural changes occurring in GETV *Macro* domain upon ADP-ribose binding can be observed in the catalytic loop $\beta 2\alpha 1$ and loop $\beta 5\alpha 3$, which close-up over the ligand in order to establish tight binding interactions.

Structure of the GETV *Macro* domain /ADPr complex with the distal ribose in “pose 2”

By adding 3 mM ADP-ribose and 50 mM glutamic acid to the crystallization medium of GETV *Macro* domain we obtained a second crystal for a GETV *Macro* domain /ADP-ribose complex for which diffraction data extending to 1.7 Å resolution were collected. The structure was determined by difference Fourier synthesis and the final model, with excellent stereochemistry, comprises two acetate molecules and 231 solvent molecules. Thr160 in chain A could not be modelled in satisfactory fashion and in chain B electron density revealed the presence of Lys0, originating from the cloning strategy that includes an AAA triplet (Lys codon) to promote efficient translation expression (Care *et al.*, 2008).

In this second GETV *Macro* domain /ADP-ribose complex structure the interactions between GETV *Macro* domain and the adenine moiety, the proximal ribose and the phosphate groups are virtually identically to the interactions described above. However, the distal ribose is tilted by approximately 90° with respect to the “classical” pose of the

distal ribose as observed in the above described GETV *Macro* domain /ADP-ribose complex (Fig. 41 and Fig. 42 A and B) and all other *Macro* domain /ADP-ribose complexes described so far, with exception of the crystal structure of the *Macro* domain of human histone *MacroH2A1.1* in complex with ADP-ribose in form B (3IIF) (Timinszky *et al.*, 2009). Yet, in this latter crystal structure the distal ribose is tilted in the opposite direction with respect to the distal ribose observed in the present structure, lying face to face at an angle of approximately 120°.

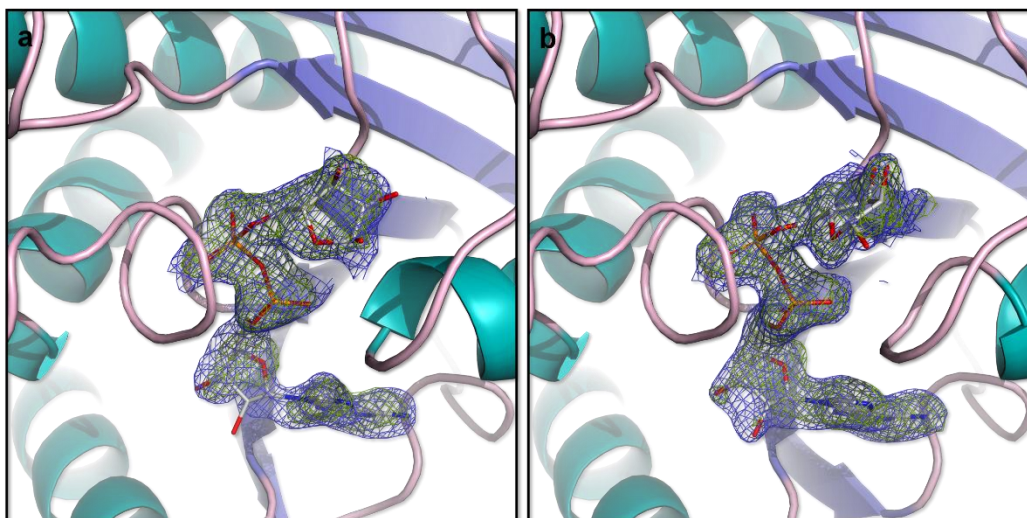


Figure 41. Comparison between GETV *Macro* domain in “pose 1” (a) and in “pose 2” (b). The interaction between GETV *Macro* domain with the adenine moiety, the proximal ribose and the phosphate groups are virtually identically for both “pose 1” (a) and “pose 2” (b). While the distal ribose in “pose 1” is tilted by approximately 90° with reference to the “conventional” pose of the distal ribose observed in the already described “pose 1”. Unbiased $F_o - F_c$ difference electron density maps, calculated before incorporation of the ligands into the models and contoured at 3.0 σ , are shown in green, and final weighted $2F_o - F_c$ maps and contoured at 1.0 σ are shown in blue.

In the novel and unusual pose 2 the distal ribose in the GETV *Macro* domain /ADP-ribose complex structure, here in the β -configuration, establishes hydrogen bonds with Ala22 and Asn24 main-chain atoms *via* 3'-OH, with the side-chain of Asn24 *via* 2'-OH and with the Ser30 side-chain and the Asp31 main-chain *via* the 1'-OH hydroxyl group, Fig. 42 B. The ribose ring lines up in perfectly parallel fashion with the side-chain of Phe114. Interestingly, one of the two aforementioned acetate molecules coordinate the 1'-OH and 2'-OH hydroxyl groups of ADP-ribose bound to chain A, Fig. 42 B. The acetate molecule is further stabilized by a H-bond interaction with Ser30, Fig. 42 B. This acetate represents in reality the carboxyl function of glutamic acid that had been added to the crystallization solution. As only a smear of electron density was visible for the rest of the amino acid, we decided to model only the carboxyl function, or more exactly, an acetate molecule. Though, the positions of the acetate molecule relative to the distal ribose ring is reminiscent of a putative ADP-ribose- glutamate conjugate, a possible substrate for

GETV *Macro* domain, as extrapolated from the function of other alphaviruses. The second acetate molecule is found at the interface between chains A and B, remote from the ADP-ribose binding site.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

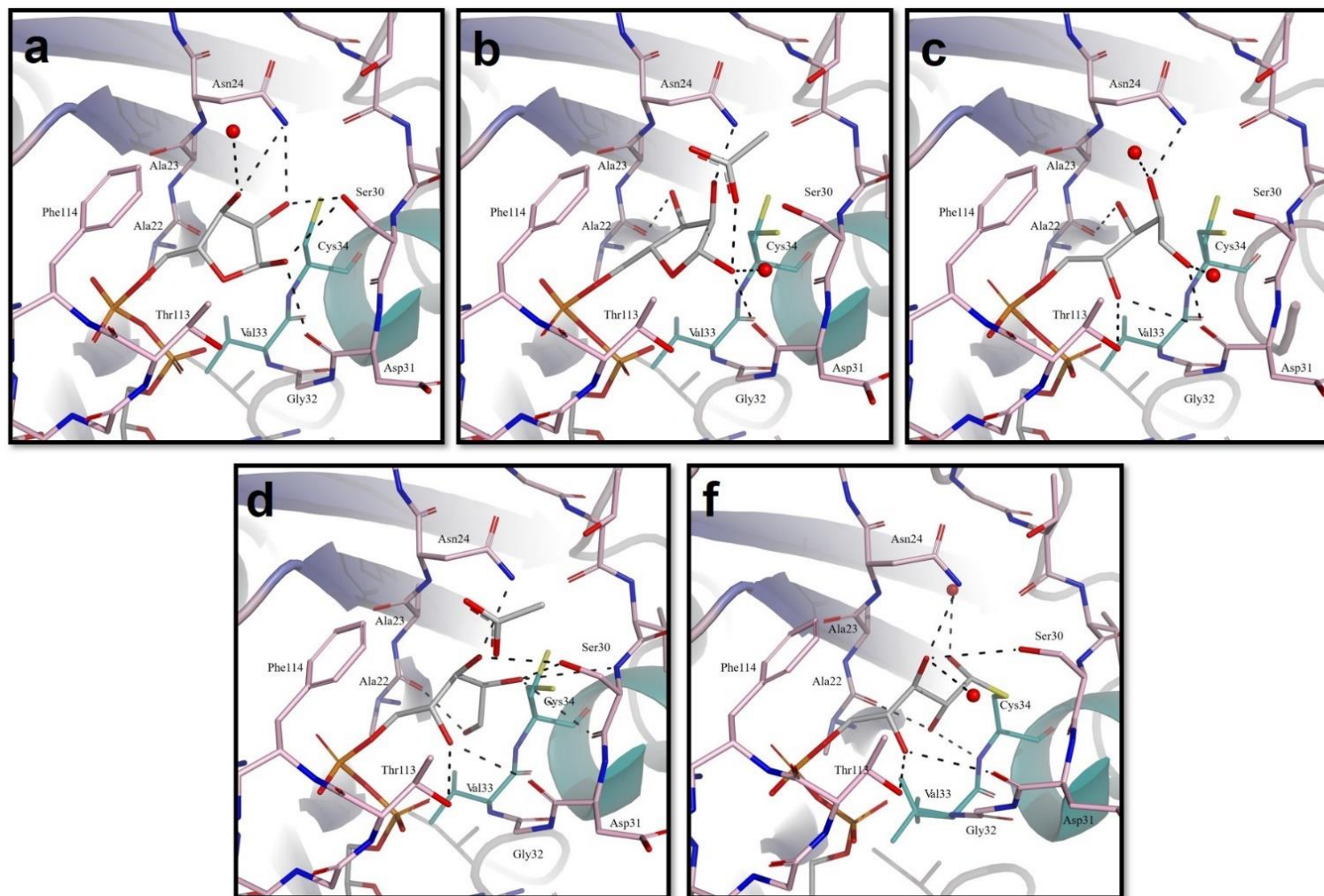


Figure 42. Conformational trajectory triggered by the binding of ADP-ribose to GETV *Macro* domain in the presence of aspartic or glutamic acid. (a) see continuation of the legend on the next page.

Figure 42. Conformational trajectory triggered by the binding of ADP-ribose to GETV *Macro* domain in the presence of aspartic or glutamic acid. (a) Three-dimension structure of GETV *Macro* domain in complex with ADP-ribose presenting the distal ribose in close conformation "Pose 1" obtained by soaking a crystal obtained by co-crystallization with 3 mM ADP-ribose in mother liquor supplemented with 15 mM aspartic acid (Xtal-91). (b) Three-dimension structure of GETV *Macro* domain in complex with ADP-ribose presenting the distal ribose in close conformation "Pose 2" obtained from diffraction data collected from a crystal obtained by co-crystallization with 3 mM ADP-ribose and 50 mM glutamic acid (Xtal-62). (c) A three-dimension structure was determined from diffracting data collected from a crystal obtained by co-crystallization of 3 mM ADP-ribose and 30 mM aspartic acid (Xtal-68) and in this three-dimension structure the distal ribose adopted a double conformation. In this panel we present the conformer A. The conformation of conformer B is the same as the one obtained for the three-dimensional structure containing a single conformation for ADP-ribose with the distal ribose in open conformation. (d) In this panel is illustrated the conformer B that is the same conformation adopted by the ADP-ribose from the three-dimension structure presenting the ADP-ribose with a single open conformation (Xtal-72). (f) The three-dimension structure representing the covalent bond established between 1"-C and the Cys34 SG. This three-dimension structure was obtained from diffracting data collected from a crystal obtained by co-crystallization with 3 mM ADP-ribose and 3 mM of aspartic acid (Xtal-44). ADP-ribose and residues interacting with ADP-ribose are depicted in sticks and colour-coded as in Fig. 38 and 40.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Structure of the GETV *Macro* domain /ADP-ribose complex with the distal ribose in the open ring conformation

We further obtained two additional GETV *Macro* domain /ADP-ribose complexes by co-crystallizing the protein in the presence of 3 mM ADP-ribose and 30 mM aspartic acid. Diffraction data were recorded for the two complexes extending to 1.85 and 1.6 Å, respectively. Again, in these two GETV *Macro* domain /ADP-ribose complexes the interactions between GETV *Macro* domain and the adenine moiety, the proximal ribose and the phosphate groups are comparable to those seen in the above described complex structures. Conversely, preliminary electron density maps, calculated before incorporation of the ligand, clearly indicated that the distal ribose was present in the open conformation, Fig. 43. Such *Macro* domain with ADP-ribose in the open conformation had not been observed before.

In the complex structure at resolution of 1.85 Å, the distal ribose is present in a single conformation, Fig. 43 A, whereas the best way to account for difference electron density in the 1.6 Å data set was to model the distal ADP-ribose in a double open conformation, Fig. 43 B. Both structures were determined by difference Fourier synthesis and the final models present excellent stereochemistry. As in the structure of the GETV *Macro* domain /ADP-ribose complex with the distal ribose in pose 2, in both structures with ADP-ribose in the open conformation no clear electron density could be observed for Thr160 in

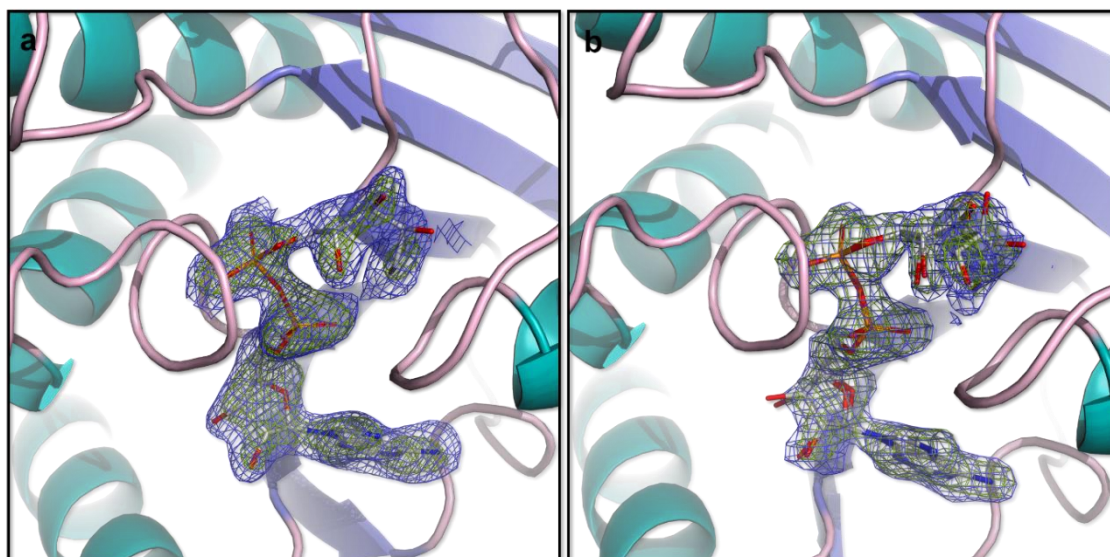


Figure 43. GETV *Macro* domain in complex with ADP-ribose with the distal ribose in open conformation. (a) the three-dimensional structure of the complex at resolution 1.85 Å where the open distal ribose was modelled in a single conformation. **(b)** the three-dimensional structure of the complex at resolution 1.6 Å where the distal ribose in open conformation was modelled in a double conformation. Unbiased $F_o - F_c$ difference electron density maps, calculated before incorporation of the ligands into the models and contoured at 3.0σ , are shown in green, and final weighted $2F_o - F_c$ maps and contoured at 1.0σ are shown in blue.

chains A and the additional Lys0, originating from the expression vector, was present in chains B. The final model of GETV *Macro* domain with the distal ribose in a single open conformation comprises two acetate molecules, Fig. 42 D, reminiscence of the aspartic acid added to crystallization solution, four ethylene glycol molecules and 190 solvent molecules. The final model of GETV *Macro* domain with the distal ribose in the double open conformation comprises one acetate molecule, one ethylene glycol molecule and 279 solvent molecules. Again, in these two structures the interactions of GETV *Macro* domain with the adenine group, the proximal ribose and the phosphate groups are practically identically to the interactions described above.

In one of the conformations of ADP-ribose in the double open conformation, conformation A, Fig. 42 C, the carbon atoms of the distal ribose superpose with the carbon atoms of the distal ribose ring as seen in the GETV *Macro* domain /ADP-ribose complex with ADP-ribose in pose 2 and the hydrogen-bonding interactions are to a large extent comparable. However, due to the ring opening the 4''-OH is now free to make hydrogen-bonding interactions with the side-chain of Thr113 and the main-chain of Asp31. The acetate molecule, as observed in the GETV *Macro* domain /ADP-ribose complex with ADP-ribose in pose 2, could not be spotted in the ADP-ribose double open ring complex and only the acetate molecules at the interface between chains A and B is present, Fig- 42 B. The conformations of the distal ADP-ribose in conformation B and the distal ribose in the GETV *Macro* domain /ADP-ribose single conformation complex are virtually identical, Fig. 42 D. Here the linear ribose chain underwent a $\sim 100^\circ$ rotation around the bond connecting carbon atoms 4''-C and 3''-C, and consequently the hydrogen-bonding pattern between APD-ribose and GETV *Macro* domain changed drastically. In this new conformation the 4''-OH hydroxyl hydrogen-bonds the side-chain of Thr113 and the main-chain of Asp31, 3''-OH interacts with the side-chains of Asn24 and Ser30, the 2''-OH hydroxyl group interacts with the side-chain of Ser30 and 1''-OH is coordinated by the main-chain atoms of Ala22 and Cys34. An aspartate molecule is isosteric to the one observed in the GETV *Macro* domain /ADP-ribose complex with the distal ribose in pose 2 and interacts with the 3''-OH hydroxyl of ADP-ribose and the side-chain of Ser30, Fig. 42.

Structure of the GETV *Macro* domain /ADP-ribose complex with the distal ribose covalently bound to Cys34

Finally, we obtained a last GETV *Macro* domain /ADP-ribose complex by co-crystallizing GETV *Macro* domain with 3 mM ADP-ribose and 3 mM aspartic acid and diffraction data were collected to 1.45 Å resolution. For no apparent reason the space group in this

complex changed from $P2_12_12_1$ (as observed in all the previously described structures) to $C2$ and only one molecule of the GETV *Macro* domain /ADP-ribose complex is present in the asymmetric unit. We verified cautiously whether the position of ADP-ribose in this novel complex could have influenced the crystal packing and could not retrieve any plausible explanation for the change of space group.

The structure was solved by molecular replacement, using the native GETV *Macro* domain structure as search model, and the final model, of very good stereochemistry, comprises one ethylene glycol molecule and 186 solvent molecules. The interactions between GETV *Macro* domain and the adenine moiety, the proximal ribose and the phosphate groups of ADP-ribose are specular to the interactions observed in all the GETV *Macro* domain /ADP-ribose complexes described so far. To our surprise, ADP-ribose in this structure is found in the open conformation and a covalent bond is established between 1'-C and to Cys34 SG, Fig. 42 F and Fig. 44. It should be noted that in all the above described GETV *Macro* domain /ADP-ribose structures the residues of the catalytic loop $\beta 2\alpha 1$ and of the loop $\beta 5\alpha 3$ are isosteric. The only exception is represented by Cys34, which adopts a double conformation in the complexes of GETV *Macro* domain /ADP-ribose with ADP-ribose in pose 2 and ADP-ribose in the open conformation, with the alternate conformation of Cys34 pointing towards the distal ribose, Fig. 42.

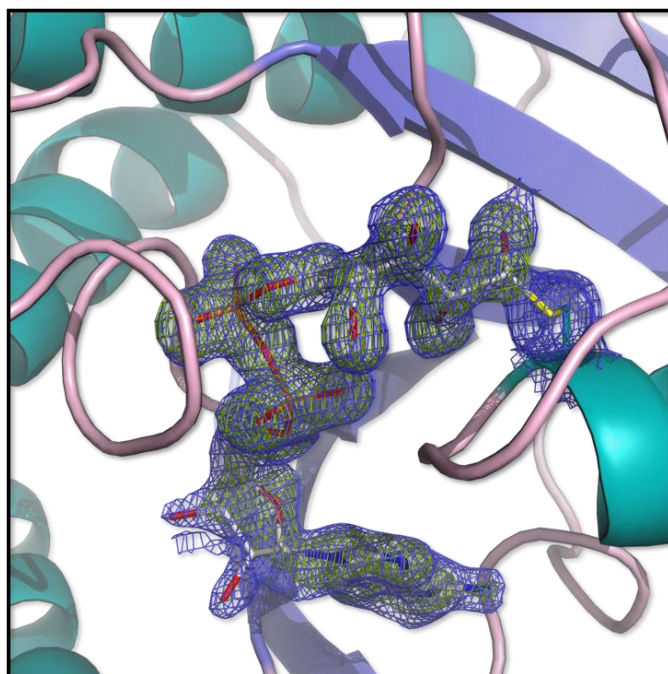


Figure 44. GETV *Macro* domain in complex with ADP-ribose with the distal ribose in open conformation and establishing a covalent bond with the lateral chain of cysteine 34. Unbiased $F_o - F_c$ difference electron density maps, calculated before incorporation of the ligands into the models and contoured at 3.0σ , are shown in green, and final weighted $2F_o - F_c$ maps and contoured at 1.0σ are shown in blue.

In the GETV *Macro* domain structure with ADP-ribose covalently bound to Cys34 a movement of approximately 1 Å could be observed for residues Ser30-Val33 of the catalytic loop $\beta 2\alpha 1$. This structural rearrangement of the catalytic loop is most likely due to a slight positional shift of the open ADP-ribose chain and a rotation of the of about 100° around the 3'-C and 2'-C bond, with respect to ADP-ribose as observed in the single open conformation. In the new conformation the hydrogen bond between 4'-OH and the side-chain of Thr113 is still maintained, but the interactions between 4'-OH and the main-chain of Asp31 and between 3'-OH and the side-chain of Ser30 are lost, and therefore the tightness of the ADP-ribose binding cleft is loosened, Fig. 42.

Due to the rearrangement of the open ADP-ribose chain the 2'-OH hydroxyl group hydrogen bonds the main-chain of Cys34 and the 1'-OH group coordinates a water molecule. We can't exclude that the covalent intermediate observed in this structure is an artefact, owing the pronounced nucleophilic character of the cysteine residue. An X-ray induced effect should be excluded, due to the change in space group, occurring intrinsically before any exposure to X-rays. Advocating for a possible role of Cys34 in catalysis is the observation that this residue is isosteric to the putative catalytic residues: Asp35 in the putative *E. coli phosphatase* (1SPV), Asp102 in the human *MacroD2* (4IQY) (Jankevicius *et al.*, 2013) and Glu114 in poly(ADP-ribose) glycohydrolase (PARG) from *Thermomonospora curvata* (3SIG) (Slade *et al.*, 2011).

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

DISCUSSION and CONCLUSIONS

In this study, we described by mean of crystallographic structures different poses adopted by a molecule of ADP-ribose in the binding site of GETV *Macro* domain. The different structures were compared to the structures of homologues available in the PDB. In addition to the pose of ADP-ribose found in other structures of alphavirus *Macro* domain, this work reveals original features such as the opening of the distal ribose, and its stabilization by a Ser30, a peculiar residue at this position. To understand how this could be part of the catalytic mechanism, we co-crystallized or soaked *Macro* domains structures with increasing concentrations of Asp or Glu, which could mimick the ADP-ribosylated side chain. This strategy was inspired by the structural study of the mono-ADP-ribosylhydrolase DraG in which an amino acid- ADP-ribose intermediate was trapped (Berthold *et al.*, 2009). No clear electron density could be detected for aspartic or glutamic acid in the crystal structures. However, addition of the amino acids was associated to conformational changes in the ADP-ribose binding site, in the catalytic loop, and of ADP-ribose itself. In our case, this allowed us to identify a covalent link with a cysteine located in the catalytic loop, as well as several poses of ADP-ribose susceptible to give clues about the catalytic mechanism. Since this cysteine is conserved only in alphaviruses, this finding would deserve to be co-related to specific deMARylation activity of alphaviruses specifically. Although the relevance of this covalent link remains to be validated by enzymatic assays and reverse genetics, these results suggest an important role, and perhaps target specificity, of this alphavirus specific cysteine on the functions of alphavirus *Macro* domains.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

REFERENCES

- Abraham, R., Hauer, D., McPherson, R. L., Utt, A., Kirby, I. T., Cohen, M. S., Merits, A., Leung, A. K. L. & Griffin, D. E. (2018). *Proc Natl Acad Sci U S A* **115**, E10457-E10466.
- Ahola, T. & Kaariainen, L. (1995). *Proc Natl Acad Sci U S A* **92**, 507-511.
- Berrow, N. S., Bussow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G. E., Levy, N., Lieu, V., Owens, R. J., Peleg, Y., Pinaglia, C., Quevillon-Cheruel, S., Salim, L., Scheich, C., Vincentelli, R. & Busso, D. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1218-1226.
- Berthold, C. L., Wang, H., Nordlund, S. & Hogbom, M. (2009). *Proc Natl Acad Sci U S A* **106**, 14247-14252.
- Care, S., Bignon, C., Pelissier, M. C., Blanc, E., Canard, B. & Coutard, B. (2008). *Nucleic Acids Res* **36**, e6.
- Chakravarthy, S., Gundimella, S. K., Caron, C., Perche, P. Y., Pehrson, J. R., Khochbin, S. & Luger, K. (2005). *Mol Cell Biol* **25**, 7616-7624.
- Das, P. K., Merits, A. & Lulla, A. (2014). *J Biol Chem* **289**, 5635-5653.
- Eckei, L., Krieg, S., Butepage, M., Lehmann, A., Gross, A., Lippok, B., Grimm, A. R., Kummerer, B. M., Rossetti, G., Luscher, B. & Verheugd, P. (2017). *Sci Rep* **7**, 41746.
- Egloff, M. P., Malet, H., Putics, A., Heinonen, M., Dutartre, H., Frangeul, A., Gruez, A., Campanacci, V., Cambillau, C., Ziebuhr, J., Ahola, T. & Canard, B. (2006). *J Virol* **80**, 8493-8502.
- Evans, P. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.
- French, S. & Wilson, K. (1978). *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **34**, 517-525.
- Fukunaga, Y., Kumanomido, T. & Kamada, M. (2000). *Vet Clin North Am Equine Pract* **16**, 605-617.
- Geerlof, A., Brown, J., Coutard, B., Egloff, M. P., Enguita, F. J., Fogg, M. J., Gilbert, R. J., Groves, M. R., Haouz, A., Nettleship, J. E., Nordlund, P., Owens, R. J., Ruff, M., Sainsbury, S., Svergun, D. I. & Wilmanns, M. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1125-1136.

- Golubtsov, A., Kaariainen, L. & Caldentey, J. (2006). *FEBS Lett* **580**, 1502-1508.
- Gotte, B., Liu, L. & McInerney, G. M. (2018). *Viruses* **10**.
- Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res* **31**, 3320-3323.
- Gould, E. A., Coutard, B., Malet, H., Morin, B., Jamal, S., Weaver, S., Gorbalenya, A., Moureau, G., Baronti, C., Delogu, I., Forrester, N., Khasnatinov, M., Gritsun, T., de Lamballerie, X. & Canard, B. (2010). *Antivir Res* **87**, 111-124.
- Holm, L. & Sander, C. (1993). *J Mol Biol* **233**, 123-138.
- Hu, C., Liu, Y., Lin, Y., Liang, J. K., Zhong, W. W., Li, K., Huang, W. T., Wang, D. J., Yan, G. M., Zhu, W. B., Qiu, J. G. & Gao, X. (2018). *Cell Death Dis* **9**, 274.
- Jankevicius, G., Hassler, M., Golia, B., Rybin, V., Zacharias, M., Timinszky, G. & Ladurner, A. G. (2013). *Nat Struct Mol Biol* **20**, 508-514.
- Karras, G. I., Kustatscher, G., Buhecha, H. R., Allen, M. D., Pugieux, C., Sait, F., Bycroft, M. & Ladurner, A. G. (2005). *EMBO J* **24**, 1911-1920.
- Leung, A. K. L., McPherson, R. L. & Griffin, D. E. (2018). *PLoS Pathog* **14**, e1006864.
- Li, C., Debing, Y., Jankevicius, G., Neyts, J., Ahel, I., Coutard, B. & Canard, B. (2016). *J Virol* **90**, 8478-8486.
- Li, C., Guillen, J., Rabah, N., Blanjoie, A., Debart, F., Vasseur, J. J., Canard, B., Decroly, E. & Coutard, B. (2015). *J Virol* **89**, 8292-8303.
- Lin, Y., Zhang, H., Liang, J., Li, K., Zhu, W., Fu, L., Wang, F., Zheng, X., Shi, H., Wu, S., Xiao, X., Chen, L., Tang, L., Yan, M., Yang, X., Tan, Y., Qiu, P., Huang, Y., Yin, W., Su, X., Hu, H., Hu, J. & Yan, G. (2014). *Proc Natl Acad Sci U S A* **111**, E4504-4512.
- Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X. & Canard, B. (2009). *J Virol* **83**, 6534-6545.
- McPherson, R. L., Abraham, R., Sreekumar, E., Ong, S. E., Cheng, S. J., Baxter, V. K., Kistemaker, H. A., Filippov, D. V., Griffin, D. E. & Leung, A. K. (2017). *Proc Natl Acad Sci U S A* **114**, 1666-1671.
- Melekis, E., Tsika, A. C., Lichiere, J., Chasapis, C. T., Margiolaki, I., Papageorgiou, N., Coutard, B., Bentrop, D. & Spyroulias, G. A. (2015). *Biomol NMR Assign* **9**, 191-195.

- Meshram, C. D., Agback, P., Shiliaev, N., Urakova, N., Mobley, J. A., Agback, T., Frolova, E. I. & Frolov, I. (2018). *J Virol*.
- Mi, S., Durbin, R., Huang, H. V., Rice, C. M. & Stollar, V. (1989). *Virology* **170**, 385-391.
- Pietila, M. K., van Hemert, M. J. & Ahola, T. (2018). *J Virol* **92**.
- Putics, A., Slaby, J., Filipowicz, W., Gorbalenya, A. E. & Ziebuhr, J. (2006). *Adv Exp Med Biol* **581**, 93-96.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res* **42**, W320-324.
- Rosenthal, F., Feijs, K. L., Frugier, E., Bonalli, M., Forst, A. H., Imhof, R., Winkler, H. C., Fischer, D., Caffisch, A., Hassa, P. O., Luscher, B. & Hottiger, M. O. (2013). *Nat Struct Mol Biol* **20**, 502-507.
- Rubach, J. K., Wasik, B. R., Rupp, J. C., Kuhn, R. J., Hardy, R. W. & Smith, J. L. (2009). *Virology* **384**, 201-208.
- Shin, G., Yost, S. A., Miller, M. T., Elrod, E. J., Grakoui, A. & Marcotrigiano, J. (2012). *Proc Natl Acad Sci U S A* **109**, 16534-16539.
- Sievers, F. & Higgins, D. G. (2014). *Curr Protoc Bioinformatics* **48**, 3 13 11-16.
- Slade, D., Dunstan, M. S., Barkauskaite, E., Weston, R., Lafite, P., Dixon, N., Ahel, M., Leys, D. & Ahel, I. (2011). *Nature* **477**, 616-620.
- Strauss, E. G., De Groot, R. J., Levinson, R. & Strauss, J. H. (1992). *Virology* **191**, 932-940.
- Timinszky, G., Till, S., Hassa, P. O., Hothorn, M., Kustatscher, G., Nijmeijer, B., Colombelli, J., Altmeyer, M., Stelzer, E. H., Scheffzek, K., Hottiger, M. O. & Ladurner, A. G. (2009). *Nat Struct Mol Biol* **16**, 923-929.
- Vasiljeva, L., Merits, A., Auvinen, P. & Kaariainen, L. (2000). *J Biol Chem* **275**, 17281-17287.
- Weaver, S. C., Hagenbaugh, A., Bellew, L. A., Netesov, S. V., Volchkov, V. E., Chang, G. J., Clarke, D. K., Gousset, L., Scott, T. W., Trent, D. W. & et al. (1993). *Virology* **197**, 375-390.

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

**9.2. Additional results: Evaluation of the effect of the viral
Macro domains catalysis in the cell using cell-based assays**

THIS PAGE IS INTENTIONALLY LEFT BLANK

The work presented in this chapter was initiated after preliminary analysis of the structures presented in chapter 9.1. As it was already mentioned in previous chapters, the alphavirus nsP3 is organized in three domains, the *Macro* domain, the ZBD and the HVR. We hypothesised that *Macro* domain and the HVR of nsP3 function in a synergistic manner. Indeed, it has been proposed that alphavirus *Macro* domains mediate the viral countermeasures to reverse ribosylation of cellular proteins (Leung et al., 2018), while the HVR plays a key role in the viral replication by interacting with several partners among which G3BP proteins and SH3 containing proteins (Meshram et al., 2018).

Interestingly, it has been proposed that when the cells are under stress or viral infections, G3BP1 is involved in the formation of membrane-less structures composed of mRNA and protein aggregates which are usually recognized as SGs (Kim et al., 2016, Panas et al., 2012, Fros et al., 2012, Leung et al., 2018, Scholte et al., 2015, Moujaber et al., 2017). G3BP1 is known to be subject of several post-translational modifications. It is a protein that harbours several motifs, proline rich (PxxP) motifs, the glutamine and glycine rich regions, a nuclear transfer factor 2 (NTF2)-like domain localized at the N-terminal, an RNA recognition motif (RRM domain), and an Arg-Gly-rich (RGG) region where the arginine can be methylated (Bikkavilli & Malbon, 2011, Vognsen et al., 2013). Moreover, decrease in methylation has been suggested to promote SG assembly (Tsai et al., 2016). Stress induced with arsenite reversibly decreases asymmetric arginine methylation of G3BP1 and thus it has been suggested that arginine methylation in the RGG domain prevents large stress granule assembly and demethylation is a novel signal that regulates stress granule assembly (Tsai et al., 2016). Apart of methylation modifications, G3BP1 can also be phosphorylated at the serine 149 and this post-translational modification is believed to be required for regulation of G3BP1 oligomerization, stress granule assembly, and the RNase activity intrinsic to G3BP1 (Reineke et al., 2017).

G3BP1 like other RNA-binding proteins, such as Ago2 and TIA-1, are increasingly poly-ADP-ribosylated and enriched in stress granules that are formed upon stress (Leung et al., 2011). The increase in poly-ADP-ribosylation during stress could be associated to increase in PARP activity and/or decrease in PARG activity. For example, the local concentration of poly-ADP-ribosylation seems to be regulated by diverse families that contain poly-ADP-ribose binding *Macro* domains, for example: Tankyrase PARP-5a; RNA-binding PARP-12 and PARP-13 isoforms; and PARP-14 and PARP-15 (Leung et al., 2011). As already reported, G3BP-mediated stress granules assembly is dependent on poly-(ADP-ribose) (Isabelle et al., 2012), the poly-(ADP-ribose) is probably involved in modulating the nuclear translocation of G3BP and it has been demonstrated that poly-(ADP-ribose) binds to the Gly-Arg rich domain of G3BP in a non-covalent form (Isabelle

et al., 2012). The G3BP members are known to interact with the HVR of nsP3 through their amino terminal domain NTF2-like. Although the carboxyterminal domain of G3BP members are not required for the binding to the HVR of nsP3, this domain is essential for the protein function during viral replication. Interestingly the carboxyterminal domain of G3BP members has been suggested to harbour the ability to bind RNA through an unknown mechanism. Therefore, we suppose that the HVR contributes to the substrate recognition and *Macro* domain for catalysis. Thus, through cell-based assays, we decided to evaluate the effect of catalytic mutations in the nsP3 *Macro* domain in order to validate our hypothesis.

The first objective to achieve for this chapter was to demonstrate that the wt GETV nsP3 also induces the formation of G3BP containing cytoplasmic foci upon stress induced with arsenite. The hypothesis that both HVR and *Macro* domain function in a synergistic function was never demonstrated. Here we initiated studies to evaluate if catalytic mutations in the *Macro* domain could affect the recruitment of G3BP to the cytoplasmic foci observed upon arsenite treatment. The G3BP1 containing cytoplasmic foci observed in the wt and mutants of nsP3 presented different morphology and then we tried to investigate the reasons that could lead to difference morphologies. One of the hypotheses already initiated was to investigate the possible involvement of G3BP1 recruitment by nsP3 in the modulation of the mTOR pathway. Here, the experiment initiated were dedicated to the contribution of the development of an assay that could be used to investigate this theory.

For these experiments, each sequence *wt* and mutant were cloned into the plasmid vector for expression in cell-based assays. Among the mutations performed, one was the cysteine 34 discovered in the work presented in the chapter 9.1. All the plasmids, for expression in cell-based assays, containing the wt or mutant nsP3 sequence of GETV or VEEV, were sent to Leiden University Medical Center, Netherlands. There, I have performed my secondment and performed the cell-based assays presented in this chapter 9.2.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

ABSTRACT

Currently, the role of the alphavirus nsP3 remains to be clarified. In the present work we show that the GETV nsP3 interacts with G3BP1 in a similar manner as the previously observations for CHIKV infected cells. The morphology, protein composition and behaviour of the induced G3BP1-containing cytoplasmic foci induced by the expression of GETV nsP3 *wt* and mutants was also subject of investigation. Here we show that these characteristics are different from the G3BP1-containing *bona-fide* stress granules (SG) induced by arsenite and are highly similar with what observed in early works for CHIKV infected cells. We also initiated the evaluation of the effect of catalytic mutations in the nsP3 *Macro* domain of GETV. More experiments are required to understand the importance of the residues within the catalytic loop of the GETV *Macro* domain. Lastly, we could not find experimental evidences that *Macro* domain and the HVR of nsP3 function in a synergistic manner.

KEYWORDS

Bona-fide stress granules, G3BP1-containing cytoplasmic foci, alphavirus, nsP3, Getah virus, *Macro* domain, catalytic mutations

INTRODUCTION

The HVR of nsP3s are known to be rich in acidic residues, in serine, threonine and proline among which is hyperphosphorylated in serine and threonines and likely the hyperphosphorylated form of nsP3 have a role during RNA synthesis. (Peranen *et al.*, 1988)(Vihinen & Saarinen, 2000, Vihinen *et al.*, 2001). The nsP3 through its HVR critically interacts with several partners among which are G3BP, the fragile X syndrome (FXR) protein family and SH3 containing proteins (Meshram *et al.*, 2018). Until now G3BP members have been found to specifically interact with nsP3s belonging to OW alphaviruses such as SINV, SFV and CHIKV, though Phe-Gly-Asp-Phe (FGDF) sequence motifs (Cristea *et al.*, 2006, Frolova *et al.*, 2006, Scholte *et al.*, 2015, Panas *et al.*, 2012, Panas *et al.*, 2014).

Regarding NW alphaviruses, it was initially supposed that the HVR of their nsP3 specifically interacts with FXR members due to the findings reported for VEEV (Kim *et al.*, 2016, Foy *et al.*, 2013). Although, later reports concerning some members belonging to the NW group revolutionized the current opinion about the specificity of the nsP3 HVR for FXR or G3BP family members with respect to their dependency on the NW or OW alphavirus group (Frolov *et al.*, 2017). These works reported that the nsP3 HVR from EEEV, a member of the NW alphavirus group, was able to interact with both FXR and G3BP family members (Frolov *et al.*, 2017). Both G3BP and FXR family members have the ability to bind RNAs and to self-assemble into higher order complexes known as cellular stress granules (SGs) (Panas *et al.*, 2014). Interestingly, the HVR of nsP3 from CHIKV and SINV cannot utilize FXRs even if the binding sequences are inserted into their HVRs (Frolov *et al.*, 2017, Kim *et al.*, 2016, Foy *et al.*, 2013).

It has been proposed that G3BP-mediated *bona fide* SG assembly is poly-(ADP-ribose) dependent (Isabelle *et al.*, 2012) and likely the disassembly occurs when PARG is activated to perform dePARylation in order to modulate the levels of chained ADPr proteins in the cell (Leung *et al.*, 2011). The *bona fide* SGs are host intracellular membrane-less organelles in a dynamic phase transition into liquid droplets (White & Lloyd, 2012, Kedersha *et al.*, 2016, Calabretta & Richard, 2015, Kedersha *et al.*, 2013, Kato *et al.*, 2012). SGs are composed of cellular mRNAs and a variety of cellular proteins such as stalled translation pre-initiation complexes, translation initiation factors (e.g., eIF3 and eIF4B), the small ribosomal subunit, RNA-binding proteins (such as T-cell restricted antigen 1 [TIA-1], TIA-1 related protein [TIAR] and G3BP members), (Kimball *et al.*, 2003, Tourriere *et al.*, 2003, Kedersha *et al.*, 1999). Several cellular kinases (Williams, 2001, Harding *et al.*, 2000, Kimball, 2001, Han *et al.*, 2001) when sense the

environmental stress induce the phosphorylation of the α -subunit of eukaryotic translation initiation factor 2 (eIF2 α) and this leads to dephosphorylation, multimerization of G3BP1 that seems to be crucial for nucleation of SGs (Tourriere *et al.*, 2003).

CHIKV infection inhibits the assembly of *bona-fide* stress granule by recruiting G3BP into cytoplasmic foci through its HVR of nsP3 (Fros *et al.*, 2012, Scholte *et al.*, 2015) and the same was observed for SFV infection (Panas *et al.*, 2012). It was also demonstrated that these cytoplasmic foci are different from *bona-fide* SGs in terms of morphology, composition, and behaviour (Scholte *et al.*, 2015).

Structure-based study of GETV *Macro* domain was initiated in chapter 9.1. We hypothesised that *Macro* domain and the HVR of nsP3 function in a synergistic manner. Therefore, the HVR contributes to the substrate recognition and *Macro* domain for catalysis. In the present work we evaluated the potential of nsP3 to recruit G3BP1 and we initiated a work to evaluation the effect of catalytic mutations on the *Macro* domain catalysis by using cell-based assays. Our results provided for the first-time evidences that the nsP3 from GETV can recruit G3BP1 in a similar fashion as CHIKV *Macro* domain. We have also observed that the presence of mutations in amino acids from the catalytic loop affect the morphology of the cytoplasmic foci formed by G3BP1 in the presence of nsP3 GETV. Unfortunately, we were not able to find more evidences that could support our hypothesis that suggests that both *Macro* domain and the HVR of nsP3 function in a synergistic manner.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

MATERIAL AND METHODS

Reagents

Unless otherwise stated, enzymes were purchased from New England Biolabs (NEB) and primers or oligonucleotides from Eurofins-GATC. Plasmids used in this work were previously sequenced at Eurofins-GATC. The recombinant plasmid pNIC28-Bsa4-PARP15a was a gift from Nicola Burgess-Brown (Addgene plasmid # 42398) (Diefenbach & Burkle, 2005, Schreiber *et al.*, 2006). The synthetic *wt* sequence of nsP3 of Getah virus South Korean strain (GenBank accession number: YP_164438.1) was obtained from Thermofisher Scientific cloned into a pMTK vector. The *wt* sequence of nsP3 from CHIKV Leiden Synthetic 3 (LS3) (GenBank accession number KC149888) cloned into an IRES-pCAGGS vector and the empty pGFPmx vector were a gift from Martijn van Hemert (Leiden University Medical Center, Leiden, Netherlands) (Scholte *et al.*, 2015). The sequence of nsP3 from VEEV strain TC-83 (GenBank accession number: L01443.1) already cloned in pDonor201 was originated from Malet *et al.*, 2009.

Cell lines and cell culture

Cultures of Vero E6 cells (ATCC® CRL-1586™, passage 31) were maintained in an incubator at 37 °C under an atmosphere of 5 % CO₂ and using Dulbecco's modified Eagle's medium (DMEM) from Lonza, containing 4.5 g·L⁻¹ glucose, 4 mM L-glutamine and supplemented with 10 % heat-inactivated fetal calf serum (iFCS), 100 U·mL⁻¹ of penicillin and 100 U·mL⁻¹ of streptomycin.

Briefly, a monolayer culture of Vero E6 cells with approximately 80 % confluence was washed with sterile phosphate buffered saline (PBS) (Braun). Cells were then trypsinized by incubation at 37 °C in 5 % CO₂ atmosphere until the cells detached from the bottom of the flask. Later, trypsin was inactivated with complete DMEM medium and cells were then counted. To count the cells, an aliquoted of 100 µL of cells was diluted using a dilution factor of 1:100 and counted using an automated cell counter (CASY cell counter). When using standard TC flask T75 cm² (Sarstedt), cells were seeded at 2.6×10⁵ or 1.7×10⁵ cells·mL⁻¹ in a final volume of 10 mL of complete DMEM medium.

(Multi-) Site- directed mutagenesis performed on alphavirus nsP3s

Recombinant plasmid VEEV nsP3^{wt}-pDonor201 and GETV nsP3^{wt}-pMTK (Invitrogen) were used as templates and were obtained using the maxiprep kit from Macherey-Nagel. The mutants created for VEEV nsP3^{wt}-pDonor201 were the following; nsP3^{Asn24Ala}, nsP3^{Cys34Ala} and nsP3^{Gly32Ala}. For GETV nsP3^{wt}-pMTK the mutants created were nsP3^{Asn24Ala}, nsP3^{Cys34Ala} and nsP3^{Asp31Ala}. The primers designed to generate each mutant are presented in table S1. The (multi-) site-directed mutagenesis reactions were performed using a “in house” protocol derived from Quikchange (Agilent). Briefly, reactions were prepared in a final volume of 50 µL by adding 2 U of Phusion DNA polymerase (NEB), 1x Phusion GC buffer, 15 ng of plasmid DNA template, 2.5 µM of sense primer, 2.5 µM of anti-sense primer, and 300 µM dNTPs (NEB). Then, the reactions were incubated in the thermocycler with the following protocol: an initial denaturation cycle of 30 sec at 98 °C, followed by 30 cycles of 10 sec at 98 °C for denaturation, 30 sec with an annealing temperature of 50 °C and 7 min for extension at 72 °C. An additional extension cycle of 10 min at 12 °C was added after the 30 cycles. Then, the samples were incubated on ice (or 4 °C in the thermocycle) for 5 minutes and after DpnI was added to each reaction, followed by incubation at 37 °C for 4 hours.

The digestion of the template plasmid was confirmed by DNA agarose electrophoresis. Lastly, 5 µL of reaction, containing only the newly amplified plasmid, was used to transform 50 µM of RapidTrans TAM1 competent *E. coli* cells (Activemotif). The cells containing the new amplified plasmid were selected in LB agarose plates supplemented with 50 µg·mL⁻¹ of kanamycin for mutants generated with the template plasmids VEEV nsP3^{wt}-pDonor201 and GETV nsP3^{wt}-pMTK (Invitrogen). The selected colonies were grown in liquid LB media and the new amplified plasmid was obtained using the miniprep kit from Macherey-Nagel.

(Multi-) Site- directed mutagenesis performed on alphavirus *Macro* domain for *E. coli* expression

Mutants MD^{Asn24Ala}, MD^{Ser30Gly}, MD^{Cys34Ala} and MD^{Asp31Ala} were generated using the recombinant plasmid template GETV MD^{wt}-pDest14. The procedure was the same as described for reactions using the recombinant plasmid templates VEEV nsP3^{wt}-pDonor201 and GETV nsP3^{wt}-pMTK (Thermo Fisher Scientific). However, the selection of the colonies was performed using 100 µg·mL⁻¹ of ampicillin. The primers designed to create each mutant are presented in table S1.

***In vitro* cell-based assays for visualization of stress granules (SG) formation in the context of alphavirus nsP3 expression**

In vitro cell-based assays were performed for visualization of stress granules formation induced by expressing the alphavirus nsP3s (*wt* and mutants). Then, cells were stained for visualization of markers of *bona-fide* SG. Vero E6 cells were seeded on glass coverslips (Ø 13 mm) in 24-well clusters using complete DMEM medium. Briefly, coverslips were sterilized with 100 % ethanol and placed in the wells or were UV-sterilized for 30 min. Into the wells containing the coverslips were added 0.5 mL of cell suspension at a concentration of 1.5×10^5 cells·mL⁻¹ or 0.8×10^5 cells·mL⁻¹.

Transfection of Vero E6 cells seeded on coverslips in 24-well clusters was performed by using Lipofectamine™ 2000 (Thermo Fisher Scientific) following the manufacturer's instructions. Initially, an optimization was made using two control plasmids, namely the empty plasmid IRES-pCAGGS and the pGFPmx. In addition, the IRES-pCAGGS containing the full-length nsP3^{wt} of CHIKV, GETV and VEEV were used. To optimize the optimal amount of lipofectamine, the transfection for each plasmid was made in duplicate, using three different concentrations of lipofectamine, 1 µL, 1.7 µL or 2.5 µL. Briefly, 500 ng of plasmid DNA were mixed in 50 µL of Opti-MEM and other 50 µL of Opti-MEM were used to dilute desired amounts of lipofectamine. The mixes were incubated for 5 min at room temperature and then the plasmids diluted in Opti-MEM were mixed gently in a ratio 1:1 with lipofectamine diluted in Opti-MEM. The samples were incubated at room temperature for 20 min. Then, each 100 µL of oligomer-lipofectamine complexes were added to the respective well containing a monolayer of approximately 90% of confluent Vero E6 cells in 0.5 mL of complete DMEM medium. The 24-well clusters were then gently mixed by rocking the plate back and forth. Lastly, cells were incubated at 37 °C in a 5 % CO₂ incubator until the next day. Then, the samples were processed for immune fluorescence analysis (IFA) and the optimal condition obtained contained 2.5 µL of lipofectamine.

Induction of *bona-fide* SG formation and disassembly was performed according to Scholte *et al.* (Scholte *et al.*, 2015). Briefly, Vero E6 cells were grown on coverslips as previously described. To induce the formation of *bona-fide* SG, confluent monolayers of cells that had grown on coverslips were treated for 1 hour with 0.5 mM sodium arsenite (Sigma-Aldrich). The disassembly of *bona-fide* SG was induced by treatment with 100 µg·mL⁻¹ of cycloheximide (CHX) for 30 min. Then, cells were treated for indirect immune fluorescence analysis as described previously (Scholte *et al.*, 2015) by using 3 % of paraformaldehyde (PFA) to fix the cells and then stored at 4 °C until the time of use.

Detection of markers from *bona-fide* SG during alphavirus nsP3 expression by Immune fluorescence analysis (IFA)

Visualize the *bona-fide* SG markers and the viral nsP3s proteins (*wt* and mutants) were performed by immune fluorescence analysis (IFA). To detect nsP3 of CHIKV were used the primary antibody α -CHIKV-nsP3 (Rabbit IgG) (1:200) and the secondary antibody D α R-Cy3 (Donkey) (1:1000). The labelling of nsP3 of VEEV and GETV was made by using a mixture of primary antibodies (α -CHIKV-nsP3 [Rabbit IgG] [1:200] with α -SFV-nsP3 [Rabbit IgG] [1:200]) and the secondary antibody was the same as the one used to label nsP3 of CHIKV. G3BP1 (Ras GTPase-activating protein-binding protein 1) was detected with the primary antibody α -G3BP1 (Mouse, BD Biosciences) using a dilution of 1:200 and the secondary antibody was D α M-Cy3 (Donkey) at a dilution 1:1000. EIF3 (eukaryotic translation initiation factor) was detected using the primary antibody α -EIF3 (Goat) diluted 1:500 and the secondary antibody was the α -Goat-Cy3 (Rabbit) diluted 1:1000. For the marker TIA (T-cell intracellular antigen-1) the primary antibody was α -TIA (Goat) at a dilution 1:200 and the secondary antibody α -Goat-Cy3 (Rabbit) diluted in 1:1000 were used. The staining for PABP was made using the primary antibody α -PABP (Mouse) in 1:200 and the secondary antibody was D α M-Cy3 (Donkey) at a dilution 1:1000. To stain the Poly-ADPr chain (PAR chain) the primary antibody α -PAR (Mouse) diluted 1:200 and the secondary antibody D α R-Cy3 (Donkey) diluted 1:1000 were used.

The procedure to stain the coverslips with the transfected Vero E6 cells is described below. Samples were fixed for 1 hour at room temperature using 3 % PFA pre-warmed at 37 °C. Three subsequent washes were performed using 1 mL of PBS-glycine. The cells were then permeabilized for 10 min using 1 mL of 0.1 % Triton X-100 in PBS. Incubation with the primary antibody was made for more than 1 hour at room temperature by using 200 μ L of antibody diluted in PBS containing 5 % bovine serum albumin (BSA) (0.5 % BSA-PBS). The dilutions of the antibody were made according to the supplier's instructions. After, the primary antibody was removed, and the samples washed three times with 1 mL of PBS. Then, samples were incubated with the fluorescent secondary antibody, diluted in 0.5 % BSA-PBS. The incubation of the secondary antibody was made for more than 30 min in the dark at room temperature. After, the secondary antibody was removed, and the samples were washed with 1 mL of PBS. Next, 200 μ L of 1 x Hoechst diluted in PBS was used to stain the nucleus of the cells. The incubation with the Hoechst solution was performed for 10 min at room temperature in the dark. The solution of Hoechst was then removed, and the samples washed three times with PBS. The excess of liquid was removed from the coverslips, which were then fixed with ProLong Antifade reagent in a pre-cleaned (with 100 % ethanol) microscope slide. The samples were

temporarily stored at 4 °C in the dark. The imaging of the samples was performed using a fluorescence microscope (Zeiss Axioscope).

Protocol development for assessment of the direct effect of nsP3 *wt* and mutants in the mTOR pathway

The G3BP1 containing cytoplasmic foci observed in cell-based assays for the *wt* and mutants of nsP3 Getah presented different morphologies. In order to evaluate if these differences of morphology are due to the effect on the recruitment of G3BP1 due to the mutations performed in the catalytic region of *Macro* domain, a second objective of this work was to evaluate if differences in the activation of the mTOR pathway could be observed between *wt* and mutants. The evaluation of the mTOR pathway activation on cell-based assays for testing the *wt* and mutants of Getah nsP3 requires adaptation of previous protocols. As a starting point, the following experiment aims to reproduce an existent protocol for Vero E6 cells infected with SFV, strain SFV4. In the future, the protocol will be adapted. To evaluate the effect of alphavirus nsP3 in the mTOR pathway, Vero E6 cells were infected with SFV (strain SFV4) and samples were analysed by Western-blot in order to detect the total S6 and the phosphorylated Ser240/244-S6.

Briefly, Vero E6 cells were seeded in a 12-well cluster by adding 1 mL per well of cell suspension at 1.75×10^5 cells·mL⁻¹ using complete DMEM medium. A time course of infection under starvation and non-starvation conditions was performed at 4 hours, 6 hours and 8 hours post-infection of Vero E6 cells with SFV. A stock of SFV at a concentration of 8.9×10^8 pfu·mL⁻¹ was used to prepare a suspension of SFV in Eagle's Minimum Essential Medium (EMEM) (Lonza) at a final concentration of 12×10^6 pfu·mL⁻¹. The Vero E6 cells seeded in the 12-well cluster were infected using a MOI of 0.1. After 3 hours post- infection, the starvation was induced in the selected wells by changing the medium of the cells to Hank's Balanced Salt Solution (HBSS) medium (Lonza), which is a minimal medium containing Mg²⁺ and Ca²⁺. During the change of medium, an intermediate wash with HBSS medium was included. The plates were again incubated at 37 °C under an atmosphere of 5 % CO₂ for 1 hour. Then, cells were collected for WB at 4 hours, 6 hours and 8 hours post-infection which corresponds to 1 hour, 3 hours and 5 hours of starvation, respectively.

Before collecting the samples, cells were washed with sterile PBS (Sigma-Aldrich) and resuspended in 300 µL of 4 x Laemmli buffer composed of 100 mM Tris-HCl pH 6.8, 40 % glycerol, 8 % SDS, 40 mM DTT, 0.04 mg·mL⁻¹ bromophenol blue. Lysates were then collected within a 1.5 mL Eppendorf tube with O-ring and rubber lid. The lysates were

stored at -20 °C until the time of use. Lysis was completed by defrosting the samples and heating at 95 °C for 10 minutes.

The proteins S6 and phosphor-Ser240/244-S6 were detected by WB using Hybond-LFP membranes (GE Healthcare). The membranes were blocked using 1 % casein (Sigma) with Tris-buffered saline containing 0.1 % Tween-20 (TBS-T) for 1 hour at room temperature. Incubation with the primary antibody diluted in 0.5 % casein-TBS-T was performed overnight at room temperature. To detect nsP1 and nsP3 of SFV it was used rabbit antiserum α - SFV nsP1 or α - SFV nsP3. As loading controls were used antibodies to detect the transferrin receptor produced in mouse (Zymed), the β -actine also produced in mouse (Sigma) and the SFV capsid produced in rabbit. The monoclonal phospho-Ser240/244-S6 primary antibody and the monoclonal total S6 antibody was obtained from CellSignal. Biotin conjugated swine- α -rabbit (DAKO) or goat- α -mouse (DAKO), and Cy3-conjugated mouse- α -biotin (Jackson) were used for fluorescent detection of the primary antibodies with a Typhoon-9410 scanner (GE Healthcare).

***In vitro* functional enzymatic assays for the ADP-ribose hydrolysis activity of alphavirus *Macro* domains**

The deMARylation and dePARylation reactions were performed for alphavirus *Macro* domains. The catalytic domain of ARTD7 (also known as PARP15a) was used to produce the substrate for deMARylation experiments and to perform the dePARylation reactions was used the catalytic domain of hTNKS1 (also known as PARP5a) to produce the substrate.

The catalytic domain of ARTD7, corresponding to the amino acids 460 up to 656 of ARTD7, was produced using the recombinant plasmid pNIC28-Bsa4-PARP15a. The recombinant catalytic domain of ARTD7 containing a hexa-histidine sequence tag at the N-terminus was produced, with some modifications, according to Venkannagari *et al.* (Venkannagari *et al.*, 2013). Briefly, the catalytic domain of ARTD7 was produced in *E. coli* Rosetta (DE3) pLysS cells (Novagen) using TB media supplemented with 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin and 34 $\mu\text{g}\cdot\text{mL}^{-1}$ of chloramphenicol. Protein production was induced with 0.5 mM IPTG when the OD_{600} reached a value between 0.4 and 0.6 and cultures were incubated at 18 °C overnight with shaking (200 rpm). The next day, cells were harvested by centrifugation at 4,000 g for 15 min and pellets were resuspended in lysis buffer made of 100 mM Hepes buffer (pH 7.5) containing 500 mM NaCl, 10 % glycerol and 10 mM imidazole. The pellets were then stored at -80 °C until the time of use.

The catalytic domain of hTNKS1 (also known as PARP5a) was produced. The procedure was similar as described for ARTD7 catalytic domain. Briefly, cells were grown in TB media containing 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin and 34 $\mu\text{g}\cdot\text{mL}^{-1}$ of chloramphenicol and protein production was induced using 0.5 mM IPTG when the OD_{600} reached a value between 0.4 and 0.6. Then cultures were grown at 16 °C overnight with shaking (200 rpm). Cells were harvested by centrifugation at 4,000 g during 15 min and pellets were resuspended in lysis buffer composed of 50 mM Hepes buffer (pH 7.5) containing 300 mM NaCl, 10 mM imidazole, 10 % glycerol and complete EDTA-free protease inhibitor cocktail tablets (Roche). The resuspended pellets were stored at -80 °C until the time of use.

Both ARTD7 and hTNKS1 catalytic domain were purified by immobilized metal affinity chromatography (IMAC) as an initial purification step by using a 5 mL His Prep column (GE Healthcare). The purification was then refined by size exclusion chromatography using a Superdex 200 HiLoad 16/600 column (GE Healthcare). Briefly, the soluble fractions of ARTD7 catalytic domain were loaded onto a 5 mL His Prep column (GE Healthcare) using an ÄKTA Xpress with 20 mM Hepes buffer (pH 7.5) containing 500 mM NaCl, 10 mM imidazole and 10 % glycerol. The protein was eluted in 20 mM Hepes (pH 7.5) containing 500 mM NaCl, 250 mM imidazole and 10 % glycerol. The purification was then refined by size exclusion chromatography using a Superdex 200 HiLoad 16/600 column (GE Healthcare) preequilibrated in 20 mM Hepes (pH 7.5) containing 300 mM NaCl and 10 % glycerol. The fractions were analysed by SDS-PAGE electrophoresis and the pure recombinant protein was collected and concentrated as described above. The protein concentration was determined spectrophotometrically at 280 nm. Aliquots of 50 μL of the catalytic domain of ARTD7 at 4 $\text{mg}\cdot\text{mL}^{-1}$ were flash-frozen in liquid nitrogen prior to storage at -80 °C.

Concerning the catalytic domain of hTNKS1, the soluble fractions were loaded onto a 5 mL His Prep column (GE Healthcare) using 30 mM Tris buffer (pH 7.5) containing 500 mM NaCl, 50 mM imidazole and 10 % glycerol. The recombinant protein was then eluted in 30 mM Tris buffer (pH 7.5) containing 500 mM NaCl, 10 mM imidazole and 10 % glycerol. A refinement was performed using size exclusion chromatography in a Superdex 200 HiLoad 16/600 column (GE Healthcare) preequilibrated in 30 mM Hepes (pH 7.5) containing 300 mM NaCl and 5 % glycerol. The collected fractions were analysed by SDS-PAGE electrophoresis and the fractions containing the pure protein were pooled together and concentrated up to 1.3 $\text{mg}\cdot\text{mL}^{-1}$ (measured spectrophotometrically at 280 nm). To store purified hTNKS1 catalytic domain, aliquots were flash-frozen in liquid nitrogen prior storage at -80 °C at a concentration of 1.3

mg·mL⁻¹ and other aliquots at a concentration of 0.7 mg·mL⁻¹ were stored in a final concentration of 50 % glycerol.

The catalytic domains of ARTD7 or hTNKS1 were ribosylated by incubating 2 μM of each protein individually with 100 mM Tris buffer (pH 8.0) containing 10 mM MgCl₂, 1 mM DTT and 10 μM NAD⁺-biotin. Incubation at room temperature was performed for 30 min, 1 hour and 30 min, 2 hours and 30 min and 3 hours. Detection of ribosylated protein was performed by WB. Before incubating the substrate (ARTD7-mono[ADP]ribosylated or hTNKS1-poly[ADP]ribosylated) with alphavirus *Macro* domains (GETV, VEEV and CHIKV), produced and purified and mentioned in chapter 9.1 for GETV and as mentioned in Malet *et al.* 2009 for VEEV and CHIKV, the samples were cleaned by using desalting columns Zeba™ Spin Desalting Columns, 7K MWCO, 0.5 mL (Thermo Scientific) or Ni²⁺ charged sepharose beads. Then, after incubation for 1 or 2 hours with *Macro* domains in 50 mM Tris (pH 7.5) containing 50 mM NaCl and 0.5 mM DTT, samples were subject to SDS-PAGE electrophoresis and WB. The detection of biotinylated protein was performed by using Strep-Tactin HRP conjugated (IBA). Poly-ADP ribosylation of hTNKS1 was also tested by incubation of 5 μM of recombinant protein in 100 mM Tris buffer (pH 8.0) containing 10 mM MgCl₂, 1 mM DTT and 1 mM βNAD⁺ (Sigma-Aldrich). Poly-ADP ribosylated hTNKS1 catalytic domain was detected by WB using as primary antibody the anti-Poly (ADP-ribose) [PAR], mAb (10H) [Mouse IgG3κ] (Adipogen).

The deMARylation and dePARylation assays using [³²P] NAD⁺ were performed by Bruno COUTARD and Etienne DECROLY using a slightly modified protocol from Li *et al.* (Li *et al.*, 2016).

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

RESULTS AND DISCUSSION

Alphaviruses are known to induce the formation of cytoplasmic foci that resemble *bona-fide* SGs (Kim *et al.*, 2016, Panas *et al.*, 2012, Fros *et al.*, 2012, Leung *et al.*, 2018, Scholte *et al.*, 2015, Moujaber *et al.*, 2017). Independent of belonging to NW or OW the nsP3 of alphaviruses plays a critical role for the formation of these cytoplasmic foci. The HVR of nsP3 is a domain known to be involved in the interaction with several partners among which are G3BP, the fragile X syndrome (FXR) protein family and SH3 containing proteins (Meshram *et al.*, 2018).

For NW alphaviruses their nsP3 HVR can specifically interact with FXR members such as observed for VEEV (Kim *et al.*, 2016, Foy *et al.*, 2013) or in some specific cases, such as in EEEV, the nsP3 HVR can interact with both FXR and G3BP family members (Frolov *et al.*, 2017). The current knowledge suggests that alphaviruses belonging to the OW harbour FGDF sequence motifs in the HVR of nsP3 and through these sequences' motifs can only interact with G3BP members. For some OW alphaviruses such as SINV, SFV and CHIKV it has been already proved (Cristea *et al.*, 2006, Frolova *et al.*, 2006, Scholte *et al.*, 2015, Panas *et al.*, 2012, Panas *et al.*, 2014). Even inserting the sequences to bind FXRs into the HVRs of nsP3 from CHIKV and SINV it was not observed the recruitment of FXRs by nsP3 (Frolov *et al.*, 2017, Kim *et al.*, 2016, Foy *et al.*, 2013).

It was never proved that the nsP3 of GETV follows the same trend observed for other OW alphaviruses such as SINV, SFV or CHIKV. Interestingly the sequence FGDF is not totally conserved as instead of FGDF it was possible to observe the sequence motifs FGDL and FGDI into the HVR of nsP3 belonging to the Getah virus South Korean strain (GenBank: YP_164438.1) (alignment not shown). Thus, we have first investigated if the nsP3 of GETV South Korean strain is able to recruit G3BP members to induce the formation of cytoplasmic foci that resemble the *bona-fide* SGs. We also aimed to understand if the HVR of nsP3 function in a synergistic manner with *Macro* domain. To assess the contribution of HVR in the substrate recognition for de-ribosylation carried out by the *Macro* domain, we initiated a work to evaluate the effect of catalytic mutations on the *Macro* domain. For this we explored the effect of the expression of nsP3 mutants to induce the formation of cytoplasmic foci that resemble SGs.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

Expression of Getah virus nsP3^{wt} in Vero E6 cells induced the formation of G3BP-containing cytoplasmic foci that resemble SGs but are different in composition.

We explored the effect of expression of GETV nsP3 *wt* in the the formation of cytoplasmic foci that resemble SGs by localizing G3BP1 contained cytoplasmic foci. The G3BP1 is one of the markers of the *bona-fide* SGs which can be generated by oxidative stress using arsenite. SG-associated G3BP1 cannot be detected in the non-transfected and non arsenite-treated cells (Mock) (data not shown), whereas after 1-hour treatment with arsenite, the conventional G3BP1-containing foci with the usual round morphology can be observed, Fig. 45F. In the cells transfected with the control empty plasmids, IRES-pCAGGS and pGFPmx (data not shown), we observed a diffuse cytoplasmic distribution of G3BP1.

Interestingly, in the Vero E6 cells transfected with the IRES-pCAGGS carrying the GETV nsP3^{wt} sequence, the G3BP1-containing foci were observed, Fig. 45B. These G3BP1-containing cytoplasmic foci did not possess the conventional rounded morphology of arsenite-induced SGs instead presented more a rod-like morphology. This rod-like morphology was identical to the morphology already observed for the SGs induced by nsP3^{wt} during CHIKV infection (Scholte *et al.*, 2015).

Previous works performed for CHIKV, SFV and SINV suggested that the composition of the G3BP1-containing foci was different from the genuine SGs induced by arsenite (Cristea *et al.*, 2006, Frolova *et al.*, 2006, Scholte *et al.*, 2015, Panas *et al.*, 2012, Panas *et al.*, 2014). Thus, we investigated the composition of the G3BP1-containing cytoplasmic foci induced by the expression in Vero E6 cells of the GETV nsP3^{wt}. Scholte *et al.* (Scholte *et al.*, 2015) showed that in arsenite-induced SGs, G3BP1, G3BP2, TIA-1, TIAR, PABP and eIF3 could readily be detected.

In consequence of the previous observation from Scholte *et al.* (Scholte *et al.*, 2015) we also initiated the characterization of the protein composition for the cytoplasmic foci induced by expression of the GETV nsP3^{wt}. We have investigated the presence of markers of the *bona-fide* SGs such as; G3BP1, TIA-1, TIAR, PABP and eIF3 containing cytoplasmic foci. However, the quality of our staining experiments only allowed to formulate conclusions for G3BP1, eIF3 and PABP, Fig. 45. From the results where we monitored the presence of cytoplasmic foci associated to G3BP1, eIF3 and PABP we were able to demonstrate that the protein composition of the cytoplasmic foci induced by the expression of GETV nsP3^{wt} is different from the *bona-fide* SGs induced by arsenite. Our results suggest that these cytoplasmic foci contain G3BP1 but not eIF3 and PABP

and might not be related to the aggregation of stalled RNA preinitiation complexes, Fig. 45.

Further experimental confirmations are required for TIA-1, TIAR and G3BP2. However, we can already suggest that GETV nsP3^{wt} induced cytoplasmic foci differ from the *bona-fide* SGs induced by arsenite. Similar experiments would be highly valuable for the GETV nsP3 mutants. For the GETV nsP3 mutants only investigations for the presence of eIF3 were performed but the results were like what has been observed for the GETV nsP3^{wt} (data not shown). We have also tried to investigate the presence of poly-ADPr chains but the preliminary results that were obtained did not allow us to raise conclusions.

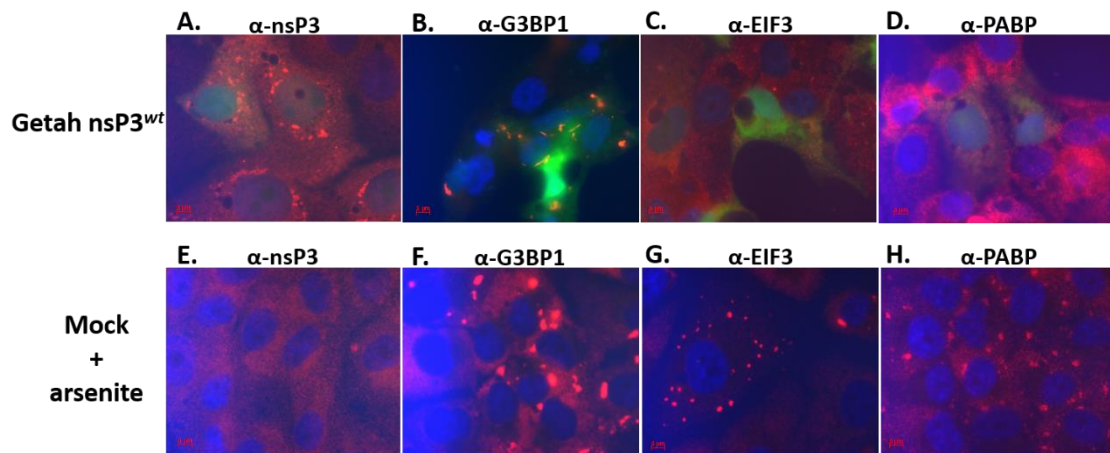


Figure 45. Getah nsP3^{wt} induced cytoplasmic foci that do not contain markers from the genuine SGs induced by arsenite induced. In red is represented the staining with nsP3 (Panel A and E), G3BP1 (Panel B and F), eIF3 (Panel C and G) and PABP (Panel D and H), Green is the expression of GFP in transfected cells and blue is the nucleus stained with Hoechst.

Getah nsP3^{wt} and mutants induced the formation of G3BP1-containing cytoplasmic foci that were identical to the G3BP1-containing cytoplasmic foci induced by CHIKV.

In our experiments we also transfected Vero E6 cells with the nsP3^{wt} from CHIKV cloned into the IRES-pCAGGS vector and the rod-like morphology of the G3BP1-containing foci were reproducible, Fig. 46D. These cytoplasmic foci were identical to the rod-like G3BP1-cytoplasmic foci observed for the transfected Vero E6 cells with the GETV nsP3^{wt} cloned into the IRES-pCAGGS, Fig. 46D. As expected accordingly to the literature, when Vero E6 cells were transfected with the VEEV nsP3^{wt} cloned into the IRES-pCAGGS, the G3BP1 -containing foci were not observed, Fig. 46B (Kim *et al.*, 2016, Foy *et al.*, 2013).

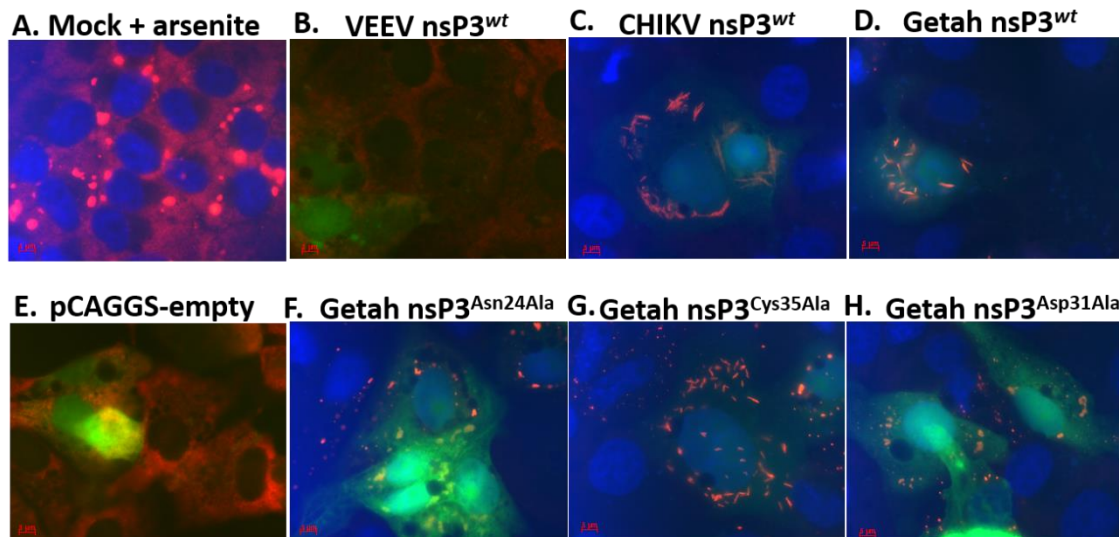


Figure 46. Getah nsP3^{wt} (D) and mutants nsP3^{Asn24Ala} (E), nsP3^{Cys35Ala} (F) and nsP3^{Asp31Ala} (G) can recruit G3BP1 to cytoplasmic foci that resemble like SGs in a similar fashion as nsP3^{wt} CHIKV (C). The VEEV^{wt} (B) was used as a negative control for G3BP1 cytoplasmic foci, the empty vector pCAGGS was used to demonstrate that expression of GFP do not induce G3BP1 cytoplasmic foci (E) and untransfected cells (Mock control) treated with arsenite were used as positive control for *bona fide* G3BP1 cytoplasmic foci (A) which are characteristic of *bona fide* SGs. In red is represented the staining with G3BP1, Green is the expression of GFP in transfected cells and blue is the nucleus stained with Hoechst.

The effect of catalytic mutations in the nsP3 *Macro* domains was also addressed by mutational study. The selection of the catalytic mutations emerged from the analysis of previous structural and functional works on *Macro* domains from VEEV, CHIKV, SINV, MAYV (Malet *et al.*, 2009, Shin *et al.*, 2012, Melekis *et al.*, 2015) and recently from GETV (results were presented in chapter 9.1). The Asn24 together with Tyr114 accommodate the distal ribose and the de-ribosylation catalytic residues are probably located in the $\beta 2\alpha 1$ loop (amino acid 26 to 33). In addition, it is likely that a catalytic water molecule plays a critical role and this water can be detected beyond alphavirus *Macro* domains (Zapata-Perez *et al.*, 2017). Thus, we decided to mutate into Ala the Asn24 and the Asp31 residues for further works on cell-based assays.

Recently and for the first time we have observed through structural works that a highly conserved cysteine (Cys34) from the catalytic loop is likely crucial for the several conformations adopted by ADPr in the binding site of GETV *Macro* domain. Thus, we decided also to include this Cys34 for further characterization and in the present work we have also mutated this cysteine to alanine. In a similar manner to what was performed for the *wt* nsP3s, we transfected the IRES-pCAGGS carrying the sequence of GETV nsP3 mutants into Vero E6 cells and the formation of G3BP1-containing cytoplasmic foci were monitored. Accordingly, to our observations, all the mutants (nsP3^{Asn24Ala}, nsP3^{Cys34Ala} and nsP3^{Asp31Ala}) induced the formation of G3BP1-containing cytoplasmic foci, Fig. 46.

Upon treatment with cycloheximide (CHX), the G3BP-containing cytoplasmic foci induced by expression of nsP3^{wt} and mutants of Getah virus, in Vero E6 cells, presented a behaviour different from the arsenite induced G3BP-containing cytoplasmic foci.

Scholte *et al.* have demonstrated that in CHIKV the induced cytoplasmic foci differ in morphology, protein composition and also in behaviour to CHX treatment (Scholte *et al.*, 2015). It is documented that the *bona-fide* SGs are dispersed upon cycloheximide (CHX) treatment. The CHX treatment stabilizes polysomes and prevents their disassembly, a crucial step in SG formation.

Thus, the observed G3BP1-containing foci induced by expression of nsP3^{wt} and mutants in Vero E6 cells was further characterized. As already observed in works like from Scholte *et al.* (Scholte *et al.*, 2015) the arsenite-induced G3BP1 SGs readily dispersed upon CHX treatment, Fig 47G. Interestingly, like in CHIKV infected cells, the G3BP1-cytoplasmic foci formed in cells transfected with plasmids expression GETV nsP3^{wt} and nsP3^{Asn24Ala}, nsP3^{Cys35Ala} and nsP3^{Asp31Ala} mutants were not dispersed by CHX treatment, Fig.47 B to E.

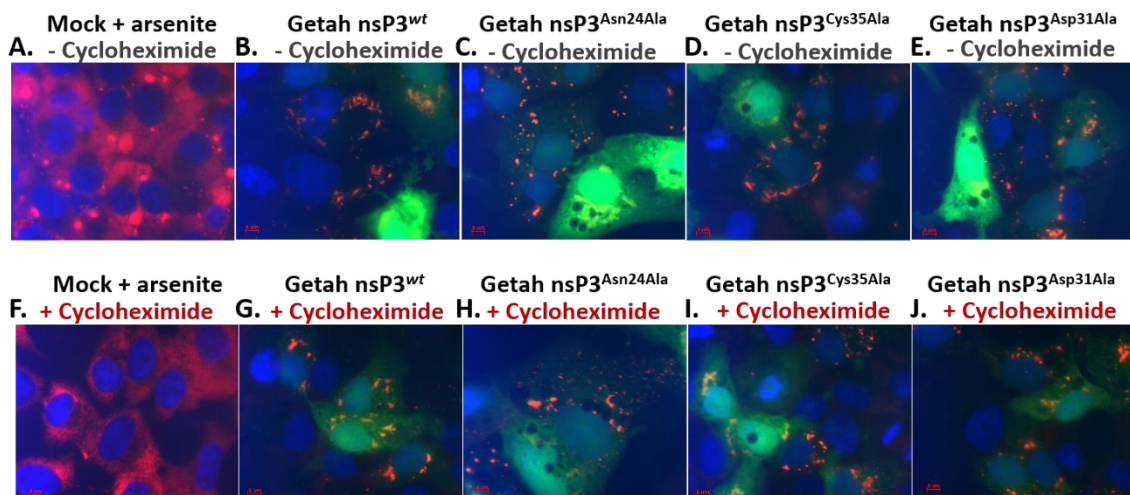


Figure 47. Getah nsP3^{wt} and mutants induced the formation of G3BP1-containing cytoplasmic foci that were resistant to cycloheximide (CHX) treatment. Panels on the top (A to E) are representing the control conditions without cycloheximide treatment. The panels on the bottom are representing the conditions in which cycloheximide was used. Arsenite induced stress granules were easily dispersed upon cycloheximide treatment (A and F). Getah nsP3^{wt} induced cytoplasmic foci that contain G3BP1 and that were resistant to cycloheximide treatment (B and G). For the nsP3 Getah mutants [nsP3^{Asn24Ala} (C and H), nsP3^{Cys35Ala} (D and I), nsP3^{Asp31Ala} (E and J)] it was observed the same effect observed for the Getah nsP3^{wt}. In red is represented the staining with G3BP1, Green is the expression of GFP in transfected cells and blue is the nucleus stained with Hoechst.

However, it was not possible to conclude if in the mutants the G3BP1-containing cytoplasmic foci were weakly dispersed. Interestingly we observed that nsP3^{Cys35Ala} upon CHX treatment, Fig. 47I, showed some alteration in the morphology of the G3BP1-

containing cytoplasmic foci that become more like round-like morphology in contrast to the GETV nsP3^{wt} that present a rod-like morphology, Fig.47B.

Thus, we could conclude that GETV nsP3^{wt} can recruit G3BP1 to cytoplasmic foci that resemble bona-fide SGs but that present significant differences concerning morphology and composition and behaviour. The GETV nsP3 mutants presented weak differences when comparing to the behaviour and morphology of the GETV nsP3^{wt}.

Protocol development to characterize the involvement of GETV nsP3 in the mTOR pathway

The recruitment of G3BP1 by nsP3 to form nsP3 complexes containing G3BP1 are the reason that likely leads to disruption of the *bona-fide* SGs in alphavirus infected cells (Fros et al., 2012, Panas et al., 2012). We may speculate that disruption of bona-fide SGs may occur before internalization of the spherule into the “body” of the infected cell. The replication complex (spherules) are actively internalised in a process that correlates with the activation of phosphoinositide-3-kinase (PI3K)-Akt-mammalian target of rapamycin (mTOR) pathway (Thaa et al., 2015, Spuul et al., 2010). However, it seems that the PI3K-AKT-mTOR pathway is not required for production of virions, reviewed in (Lark et al., 2017). Alphaviruses are known to modulate the PI3K-Akt-mTOR pathway, also known as the “pro-survival” pathway of rapamycin to prevent an apoptotic response. Experiments with nsP3 from SFV showed that nsP3 activates directly Akt only when associated with the plasma membrane though the hyperphosphorylated HVR which induces the internalization of the replication complex (Thaa et al., 2015). In contrast, it has been reported that nsP3 from CHIKV only moderately activated Akt which was dependent on PI3K which did not stimulate the internalization of the replication complex for CHIKV (Thaa et al., 2015, Spuul et al., 2010). It was proposed that the HVR of nsP3 works by mimicking activated growth factor receptors at the plasma membrane and results in the activation of the PI3K-Akt-mTOR pathway, Fig. 18 and 21 (Gotte et al., 2018).

In non-infectious situations the G3BP1 has been suggested to interact with RAPTOR and ASTRIN, also known as SPAG5 (Thedieck et al., 2013). RAPTOR is one of the proteins from the mTORC1 complex. It is likely the competition of ASTRIN with mTOR for RAPTOR-binding that leads to a decrease in the mTORC1 formation. The complex ASTRIN-RAPTOR is increased by oxidative stress and is probably present within *bona-fide* SGs induced by arsenite, thus preventing mTORC1-hyperactivation-induced apoptosis (Thedieck et al., 2013).

It remains not understandable the process and the reason why nsP3 recruits G3BP1. Accordingly, to the functions of G3BP1 and the known modulation of PI3K-Akt-mTOR pathway during alphavirus infection it may direct us in searching for a co-relation between PI3K-Akt-mTOR pathway and G3BP1. Interestingly, the FXR protein members might also regulate the mTOR/S6K signalling pathway (Huang *et al.*, 2016). Some works have suggested that their overexpression experiments of FXR prevented the growth of cells and induced cell cycle arrest, which was enhanced by the mTOR/S6K inhibitor rapamycin (Huang *et al.*, 2016). Thus, it may reduce the phosphorylated protein S6K (Huang *et al.*, 2016). In agreement with these evidences, their experiments of FXR upregulation intensified the inhibition of cell growth by rapamycin and the downregulation of FXR produced the opposite effect (Huang *et al.*, 2016).

We can speculate that for some alphaviruses to increase the formation and activation of mTORC1, G3BP1 needs to interact with ASTRIN-RAPTOR and that nsP3 may be involved in the disruption of the ASTRIN-RAPTOR complex by hijacking the G3BP1. To address this hypothesis, we need to develop starvation assays that could allow us to monitor the composition of the cytoplasmic foci induced by nsP3. It is possible that these cytoplasmic foci may contain proteins that play a role in the PI3K-Akt-mTOR pathway. In the present work we have tested the protocols for starvation experiments already available in Leiden for infection by SFV (data not shown). Currently we are trying to find a possible adaptability of the available protocols to answer our questions regarding the effect of GETV nsP3 in the mTOR pathway. One possibility would be the development of a synthetic virus for GETV that could be used for similar experiments. Alternatively, we could also adapt the starvation protocols for cell-based assays involving transfection of plasmids expressing the alphavirus nsP3 and monitored the presence of proteins from the mTOR pathway in the cytoplasmic foci.

***In vitro* functional enzymatic characterization of the amino acids critical for the ADP-ribose hydrolysis activity of alphavirus *Macro* domains**

To understand if the *Macro* domain acts in a synergistic manner with the HVR of nsP3 we need to characterize the effect of mutations in critical residues for ADP-ribose hydrolysis. Also, a newly amino acid was identified as critical for the function of the deMARylation reaction. This amino acid was identified from the work presented in chapter 9.1 and is a highly conserved cysteine present in the catalytic loop.

We have initiated the production of the necessary proteins to prepare the substrates. However, the GETV *Macro* domain mutants were not yet expressed. The initial

deMARylation experiments did not produced the expected results, including for the controls. A process of protocol adaptation is in curse and it will be used to access the effect of the catalytic mutations in the activity of the GETV *Macro* domain. Further experiments are thus required.

CONCLUSIONS

Some of the enigmatic functions of the alphavirus nsP3 were already discovered and others are still under investigation. Here we tried to understand some of the peculiar hidden functions of nsP3. In the present work we discovered that the GETV nsP3 interacts with G3BP1 in a similar manner as the previously observations for CHIKV infected cells. We have also noted that the morphology, protein composition and behaviour of the induced G3BP1-containing cytoplasmic foci induced by expression of nsP3 are different from the G3BP1-containing *bona-fide* SGs induced by arsenite. The evaluation of catalytic mutations in the nsP3 *Macro* domain of GETV were inconclusive, however we observed possible differences of morphology in the G3BP1-containing cytoplasmic foci induced by expression of the GETV nsP3 mutants. In order to conclude that residues from the catalytic loop of the GETV *Macro* domain are crucial for the morphology and stability of the G3BP1-containing cytoplasmic foci, additional experiments are required.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

REFERENCES

- Bikkavilli, R. K. & Malbon, C. C. (2011). *J Cell Sci* **124**, 2310-2320.
- Diefenbach, J. & Burkle, A. (2005). *Cell Mol Life Sci* **62**, 721-730.
- Fros, J. J., Domeradзка, N. E., Baggen, J., Geertsema, C., Flipse, J., Vlak, J. M. & Pijlman, G. P. (2012). *J Virol* **86**, 10873-10879.
- Han, A. P., Yu, C., Lu, L., Fujiwara, Y., Browne, C., Chin, G., Fleming, M., Leboulch, P., Orkin, S. H. & Chen, J. J. (2001). *EMBO J* **20**, 6909-6918.
- Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M. & Ron, D. (2000). *Mol Cell* **6**, 1099-1108.
- Huang, X., Zeng, Y., Wang, X., Ma, X., Li, Q., Li, N., Su, H. & Huang, W. (2016). *Biochem Biophys Res Commun* **474**, 351-356.
- Isabelle, M., Gagne, J. P., Gallouzi, I. E. & Poirier, G. G. (2012). *J Cell Sci* **125**, 4555-4566.
- Kedersha, N. L., Gupta, M., Li, W., Miller, I. & Anderson, P. (1999). *J Cell Biol* **147**, 1431-1442.
- Kim, D. Y., Reynaud, J. M., Rasaloukaya, A., Akhrymuk, I., Mobley, J. A., Frolov, I. & Frolova, E. I. (2016). *PLoS Pathog* **12**, e1005810.
- Kimball, S. R. (2001). *Prog Mol Subcell Biol* **26**, 155-184.
- Kimball, S. R., Horetsky, R. L., Ron, D., Jefferson, L. S. & Harding, H. P. (2003). *Am J Physiol Cell Physiol* **284**, C273-284.
- Leung, A. K., Vyas, S., Rood, J. E., Bhutkar, A., Sharp, P. A. & Chang, P. (2011). *Mol Cell* **42**, 489-499.
- Leung, A. K. L., McPherson, R. L. & Griffin, D. E. (2018). *PLoS Pathog* **14**, e1006864.
- Li, C., Debing, Y., Jankevicius, G., Neyts, J., Ahel, I., Coutard, B. & Canard, B. (2016). *J Virol* **90**, 8478-8486.
- Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X. & Canard, B. (2009). *J Virol* **83**, 6534-6545.
- Melekis, E., Tsika, A. C., Lichiere, J., Chasapis, C. T., Margiolaki, I., Papageorgiou, N., Coutard, B., Bentreop, D. & Spyroulias, G. A. (2015). *Biomol NMR Assign* **9**, 191-195.
- Meshram, C. D., Agback, P., Shiliaev, N., Urakova, N., Mobley, J. A., Agback, T., Frolova, E. I. & Frolov, I. (2018). *J Virol*.
- Moujaber, O., Mahboubi, H., Kodiha, M., Bouttier, M., Bednarz, K., Bakshi, R., White, J., Larose, L., Colmegna, I. & Stochaj, U. (2017). *Biochim Biophys Acta Mol Cell Res* **1864**, 475-486.
- Panas, M. D., Varjak, M., Lulla, A., Eng, K. E., Merits, A., Karlsson Hedestam, G. B. & McInerney, G. M. (2012). *Mol Biol Cell* **23**, 4701-4712.

- Reineke, L. C., Tsai, W. C., Jain, A., Kaelber, J. T., Jung, S. Y. & Lloyd, R. E. (2017). *Mol Cell Biol* **37**.
- Scholte, F. E., Tas, A., Albulescu, I. C., Zusinaite, E., Merits, A., Snijder, E. J. & van Hemert, M. J. (2015). *J Virol* **89**, 4457-4469.
- Schreiber, V., Dantzer, F., Ame, J. C. & de Murcia, G. (2006). *Nat Rev Mol Cell Biol* **7**, 517-528.
- Shin, G., Yost, S. A., Miller, M. T., Elrod, E. J., Grakoui, A. & Marcotrigiano, J. (2012). *Proc Natl Acad Sci U S A* **109**, 16534-16539.
- Thedieck, K., Holzwarth, B., Prentzell, M. T., Boehlke, C., Klasener, K., Ruf, S., Sonntag, A. G., Maerz, L., Grellscheid, S. N., Kremmer, E., Nitschke, R., Kuehn, E. W., Jonker, J. W., Groen, A. K., Reth, M., Hall, M. N. & Baumeister, R. (2013). *Cell* **154**, 859-874.
- Tourriere, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J. M., Bertrand, E. & Tazi, J. (2003). *J Cell Biol* **160**, 823-831.
- Tsai, W. C., Gayatri, S., Reineke, L. C., Sbardella, G., Bedford, M. T. & Lloyd, R. E. (2016). *J Biol Chem* **291**, 22671-22685.
- Venkannagari, H., Fallarero, A., Feijs, K. L., Luscher, B. & Lehtio, L. (2013). *Eur J Pharm Sci* **49**, 148-156.
- Vognsen, T., Moller, I. R. & Kristensen, O. (2013). *PLoS One* **8**, e80947.
- Williams, B. R. (2001). *Sci STKE* **2001**, re2.
- Zapata-Perez, R., Gil-Ortiz, F., Martinez-Monino, A. B., Garcia-Saura, A. G., Juanhuix, J. & Sanchez-Ferrer, A. (2017). *Open Biol* **7**.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

SUPPLEMENTARY INFORMATION

Table S1 | List of sense and anti-sense primers designed to obtain the mutants of nsP3 for VEEV and GETV by (multi-) site- directed mutagenesis.

Recombinant Plasmid DNA Template	Mutation	Primer	Sequence primer (5'-3')	
(VEEV)nsP3 ^{WT} -pDonor201	nsP3 ^{N25A}	sense	GGAGTGATTATAAATGCTGCTGCCAGCAAAGGACAA CCTGGCGG	
		anti-sense	CCGCCAGGTTGTCTTTGCTGGCAGCAGCATTATA ATCACTCC	
	nsP3 ^{C35A}	sense	GGACAACCTGGCGGAGGGGTGGCCGGAGCGCTGT ATAAGAAATTCC	
		anti-sense	GGAATTTCTTATACAGCGCTCCGGCCACCCCTCCGC CAGGTTGTCC	
	nsP3 ^{G31S}	sense	GCTAACAGCAAAGGACAACCTAGCGGAGGGGTG	
		anti-sense	CACCCCTCCGCTAGGTTGTCTTTGCTGTTAGC	
(GETV)nsP3 ^{WT} -pMTK	nsP3 ^{N25A}	sense	CACAGTGAGGAAGCGGTCGTAGCTGCTGCCAATGC CAAAGGTACC	
		anti-sense	GGTACCTTTGGCATTGGCAGCAGCTACGACCGCTT CCTCACTGTG	
	nsP3 ^{C35A}	sense	GGTACCGTGAGCGACGGAGTGCCAGGGCGGTCCG CTAAGAAGTGG	
		anti-sense	CCACTTCTTAGCGACCGCCCTGGCCACTCCGTCGC TCACGGTACC	
	nsP3 ^{S31G}	sense	GCCAATGCCAAAGGTACCGTGGGCGACGGAGTGTG CAGGGCGGTCCG	
		anti-sense	CGACCGCCCTGCACACTCCGTCGCCACGGTACCT TTGGCATTGGC	
	nsP3 ^{D32A}	sense	GCCAAAGGTACCGTGAGCGCCGGAGTGTGCAGGG CGGTCCGTAAG	
		anti-sense	CTTAGCGACCGCCCTGCACACTCCGGCGCTCACGG TACCTTTGGC	
	MD(GETV)nsP3 ^{WT} -pDest14	MDnsP3 ^{N25A}	sense	GAAGCGGTGGTGAATGCGGCGCCGCGAAAGGTA CCGTTAGCGATG
			anti-sense	CATCGCTAACGGTACCTTTTCGCGGCCGCCGATTC ACCACCGCTTC
		MDnsP3 ^{C35A}	sense	GGTACCGTTAGCGATGGTGTGGCCCGTGGCGTTGC GAAGAAATGG
			anti-sense	CCATTTCTTCGCAACCGCACGGCCACACCATCGCT AACGGTACC
MDnsP3 ^{D32A}		sense	GAATGCGAAAGGTACCGTTAGCGCCGGTGTGTGCC GTGCGGTTGCG	
		anti-sense	CGCAACCGCACGGCACACACCGGCGCTAACGGTAC CTTTCGATTC	
MDnsP3 ^{S31G}		sense	GCACACCATCGCCAACGGTACCTTTTCGATTC	
		anti-sense	GAATGCGAAAGGTACCGTTGGCGATGGTGTGTGC	

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

9.3 Discussion

THIS PAGE IS INTENTIONALLY LEFT BLANK

Arboviruses are a major global public health concern due to their emergency and re-emergency and they are causing a global burden of health and economic problems worldwide. Among which are at least 11 species belonging to the *Alphavirus* genus and for those there are no specific treatments to cure infected patients (Gould *et al.*, 2010b). Safe, specific and efficient therapeutics to treat infections caused by alphaviruses are highly needed because until now only treatments are available to ameliorate the symptoms of patients, such as analgesics and antipyretic drugs. The starting point for drug development is the selection and validation of viral drug targets, usually are proteins that display key functions for virus replication and that can be modulated by small compounds. With the presented work we aim to gain more insight in the function of nsP3 and to understand how nsP3 would fit as a viral drug target for drug development.

Since many years the function of nsP3 was a mystery and only recently some studies succeeded to elucidate some of its functions during viral replication (Gotte *et al.*, 2018). Interestingly, it has been suggested that the multifunctional role of nsP3 is explained in part by the specific function of their three domains. One of such is the *Macro* domain localized at the N-terminal region and that is involved in ADP-ribose binding and de-ribosylation during both early and late steps of replication (Abraham *et al.*, 2018). Thus, it has been suggested that the activity of *Macro* domain function as a viral countermeasure (Leung *et al.*, 2018).

Our initial objective was to understand the catalytic mechanism carried by *Macro* domain through natural sequence diversity. Thus, we have initiated structure-based studies for GETV *Macro* domain. Infecting mainly horses, GETV is geographically distributed from Asia to the north of Australia (Fukunaga *et al.*, 2000). It has been reported by others that the strain M1 has a feature of increased replication potency in cancer cell lines that lack expression of zinc-finger antiviral protein (ZAP), a catalytically deficient poly ADP-ribose polymerase (Lin *et al.*, 2014). Moreover, the *Macro* domain from GETV strain M1 presents 69 % identity with the CHIKV *Macro* domain and in the catalytic loop is present a conspicuous substitution of the conserved residues, as the case of the glycine to serine in the catalytic loop.

We have determined the three-dimensional structure for GETV *Macro* domain, for the first time. To explore the binding capacity of ADP-ribose and to better understand its catalytic mechanism of ADP-ribose removal from ribosylated protein substrates, we have characterized several conformations adopted by ADPr in the binding site that may represent several features of the de-ribosylation mechanism. Structure analysis and comparisons with homologues for which the three-dimensional structure was available at the PDB databank was performed and revealed original features belonging to the GETV

Macro domain. However, the overall fold of GETV *Macro* domain is very well conserved, in specially with *Macro* domains belonging to the Old-World alphavirus group, but minor structural changes can be observed at the level of the $\beta 3\beta 4$ loop. Moreover, the catalytic loop (the $\beta 2\alpha 1$ loop) is structurally well conserved among alphavirus *Macro* domain structures. When comparing with of the histone variant *MacroH2A* and coronavirus *Macro* domains that present mainly a central seven-stranded β -sheet flanked on either side by three α -helices, the alphavirus *Macro* domains lack the first β -strand and α -helices $\alpha 1$ and $\alpha 2$ (CHIKV annotation) which are degenerated to 3_{10} -helices in other alphavirus *Macro* domains structures. Interestingly, in the GETV *Macro* domain are reduced to loops.

We have identified within this work that when co-crystallizing the GETV *Macro* domain only with ADPr, the obtained structure presents the distal ribose in an unusual open conformation. Several interactions with residues from the catalytic site with the ADP-ribose were mapped in the work presented in this manuscript. Intriguingly, it was the observation of different poses adopted by ADP-ribose when the *Macro* domain was co-crystallized with ADP-ribose and aspartic/ glutamic acid, or even when the co-crystallized *Macro* domain with ADP-ribose was soaked with aspartic acid. Two interesting poses were observed for ADP-ribose in close conformation, “pose1” virtually identical to the pose adopted by ADP-ribose in other alphavirus *Macro* domains and “pose 2” corresponding to a new pose observed in which the distal ribose in β -configuration is tilted by approximately 90° with respect to the “classical” pose 1 of the distal ribose.

Also, other poses were observed referring that the ADP-ribose also adopted an open conformation of the distal ribose. Interestingly, with the presence of aspartic or glutamic acid, it was obtained a three-dimensional structure where was identified a double conformation of the distal ribose in open conformation, while other three-dimensional structure presented a typical single open conformation of the distal ribose. Despite the presence of residual electron density was not possible to model the electron density for aspartic or glutamic acid for any of the collected diffraction data. Interestingly, in one of our conditions we have obtained a three-dimensional structure of GETV *Macro* domain /ADP-ribose complex in which a covalent bond is established between $1''$ -C and to Cys34 SG. In this complex it was noticed that the space group changed from $P2_12_12_1$ (as observed in all the previously described structures) to $C2$ and only one molecule of the GETV *Macro* domain /ADP-ribose complex is present in the asymmetric unit. Any plausible explanation is available at the moment, but we can exclude an X-ray induced effect, however we can't exclude that the covalent intermediate observed in this structure is an artefact.

It is curious that the Cys34 is isosterical to the putative catalytic residue Asp35 in the *E. coli phosphatase* (1SPV), Asp102 in the human MacroD2 (4IQY) (Jankevicius *et al.*, 2013) and Glu114 in poly(ADP-ribose) glycohydrolase (PARG) from *Thermomonospora curvata* (3SIG) (Slade *et al.*, 2011). Thus, reinforcing the possible role of Cys34 for the catalysis of *Macro* domain and for this reason we started to address the role of the Cys34 by using cell-based assays. Following our initial objective, to understand the catalytic mechanism carried by *Macro* domain through natural sequence diversity, we tried to understand this mechanism in the context of the full length nsP3 in cell-based assays. We suppose that both HVR and *Macro* domains function in a synergistic manner, likely the HVR contributes to the substrate recognition and *Macro* domain for catalysis.

An interactive role of nsP3 was described for the formation of cytoplasmic foci that resemble *bona-fide* SGs (Cristea *et al.*, 2006, Frolova *et al.*, 2006, Scholte *et al.*, 2015, Panas *et al.*, 2012, Panas *et al.*, 2014)(Kim *et al.*, 2016, Foy *et al.*, 2013)(Frolov *et al.*, 2017). The described cytoplasmic foci are known to result from the interaction of the HVR of nsP3 with several partners among which are G3BP, the fragile X syndrome (FXR) protein family and SH3 containing proteins (Meshram *et al.*, 2018). The interaction of the HVR with the fragile X syndrome (FXR) protein family is a characteristic of NW alphaviruses that use this interaction to form cytoplasmic foci that resemble *bona-fide* SGs (Kim *et al.*, 2016, Foy *et al.*, 2013)(Frolov *et al.*, 2017). While the interactions with G3BP are typical characteristic of OW alphaviruses, however it has been already described to occur with EEEV, a NW alphavirus (Frolov *et al.*, 2017).

Studies were lacking for GETV nsP3 in what concern to the formation of the cytoplasmic foci that resemble *bona-fide* SGs. The sequence FGDF to which is expected to bind G3BP is not totally conserved and our sequence analysis suggested that are FGDL and FGDI for GETV South Korean strain (GenBank: YP_164438.1) (alignment not shown). Thus, the first step was to understand if the *wt* GETV nsP3 can recruit G3BP members to induce the formation of cytoplasmic foci that resemble the *bona-fide* SGs. Our work with cell-based assays suggested that similarly to CHIKV (Fros *et al.*, 2012, Scholte *et al.*, 2015) and SFV (Panas *et al.*, 2012), the *wt* GETV nsP3 inhibits the *bona-fide* stress granule assembly by recruiting G3BP into cytoplasmic foci. Likely involving the interaction of the HVR of nsP3 with the FGDL and FGDI motifs. Interestingly, the cytoplasmic foci formed by overexpression of nsP3 in Vero E6 cells are different from *bona-fide* SGs in terms of morphology, composition, and behaviour as it was observed for CHIKV (Scholte *et al.*, 2015). In our experiments we observed a rod-like morphology of the G3BP1-containing foci that was reproducible (Fig. 46 C). As expected accordingly to the literature (Kim *et al.*, 2016, Foy *et al.*, 2013), when Vero E6 cells were transfected

with the VEEV nsP3^{wt} cloned into the IRES-pCAGGS, the G3BP1 -containing foci were not observed (Fig. 46 B).

Then, we started to address the effect of catalytic mutations in the nsP3 *Macro* domain. Using the same *in vitro* system assay to monitor the formation of G3BP-containing cytoplasmic foci we tested several full length nsP3 mutants harbouring mutations localized in the *Macro* domain. The decision to mutate Asn24 and Asp31 to alanine (nsP3^{Asn24Ala} and nsP3^{Asp31Ala} respectively) was taken based on analysis from previous structural works on *Macro* domains VEEV, CHIKV, SINV, MAYV (Malet *et al.*, 2009, Shin *et al.*, 2012, Melekis *et al.*, 2015) and recently from GETV (results were presented in chapter 9.1). The Asn 24 is involved in interactions with the distal ribose and the residue Asp31 is included within the β 2 α 1 loop (amino acid 26 to 33) and it is likely one of the catalytic residues. For the first time we have identified the highly conserved Cys34 as a key player within the catalytic loop. We found that this residue was highly important for the different conformations adopted by ADPr in our structural work presented in the chapter 9.1. Thus, we also included the Cys34 for mutation to alanine (nsP3^{Cys34Ala}). Surprisingly, all the GETV mutants nsP3^{Asn24Ala}, nsP3^{Cys34Ala} and nsP3^{Asp31Ala} induced the formation of G3BP-containing cytoplasmic foci, however the morphology for the mutants nsP3^{Asn24Ala} and nsP3^{Asp31Ala} (Fig.46 E and G) was different from the nsP3^{wt} (Fig. 46 D). The morphology of those mutants resembles the morphology observed for SGs induced with arsenite (Fig.46 A) as it were more round-like G3BP1-cytoplasmic foci, although presenting inferior size. While for the mutant nsP3^{Cys34Ala} (Fig.46 F) the morphology is still like the nsP3^{wt} (Fig.46 D). the reason why different morphologies were observed remains obscure.

The formation of *bona-fide* SG requires the disassembly of polysomes (Scholte *et al.*, 2015) and to further characterize the G3BP-containing foci from Vero E6 cells transfected with plasmids expressing *wt* nsP3 or nsP3 mutants, after 1 hour of transfection, we have stabilized the polysomes by treating the cells with cycloheximide (CHX). Thus, if the G3BP-containing cytoplasmic foci behave like the arsenite-induced *bona-fide* SGs (Fig. 47 A and F) our expectation was to observe the disassembly of the G3BP-containing cytoplasmic foci. Interestingly, the CHX treatment did not dispersed the G3BP-containing cytoplasmic foci for the assays with GETV nsP3^{wt} (Fig. 47 B and G) and the mutants nsP3^{Asn24Ala} (Fig. 47 C and H), nsP3^{Cys35Ala} (Fig. 47 D and I), and nsP3^{Asp31Ala} (Fig. 47 E and J). However, it remains to confirm if for mutants the G3BP1-containing cytoplasmic foci were weakly dispersed. Interestingly for GETV nsP3^{Cys35Ala} upon CHX treatment (Fig. 47 D and I) it was observed some alteration in the morphology of the G3BP1-containing cytoplasmic foci that become more like round-like morphology

in contrast to the GETV nsP3^{wt} that present a rod-like morphology. All these evidences suggest that GETV nsP3^{wt} can recruit G3BP1 to cytoplasmic foci that resemble *bona-fide* SGs but significative differences were observed in what concerns morphology and behaviour. The alterations observed in the cell-based assays performed with the mutants suggested that possible the mutated residues are required for some interaction with partner proteins required for the formation of the characteristic G3BP-containing cytoplasmic foci.

The composition of the G3BP-containing cytoplasmic foci was also evaluated for all the GETV nsP3 mutants and nsP3^{wt}. Thus, we have investigated the presence of markers of the *bona-fide* SGs such as; G3BP1, TIA-1, TIAR, PABP and eIF3. Then, the results were compared with arsenite-induced SGs for which G3BP1, G3BP2, TIA-1, TIAR, PABP and eIF3 could readily be detected as it was already demonstrated by (Scholte *et al.*, 2015). Despite the quality of our staining, we could observe that among G3BP1, eIF3 and PABP only G3BP-cytoplasmic foci were identified. This led us to suggests that the cytoplasmic foci formed GETV nsP3 mutants and nsP3^{wt} do not contain eIF3 and PABP (Fig. 45). Thus, the protein composition of the cytoplasmic foci induced by the expression of GETV nsP3^{wt} or nsP3 mutants is different from the *bona-fide* SGs induced by arsenite. Our results identical to what has been observed for CHIKV (Fros *et al.*, 2012, Scholte *et al.*, 2015) and SFV (Panas *et al.*, 2012). The novelty of this work was the inclusion of mutations in the *Macro* domain and interestingly further works would be an asset to better elucidate the role of the selected mutated amino acids during the formation of the G3BP-containing cytoplasmic foci. Attempts to characterize the effect of mutations in critical residues for ADP-ribose hydrolysis through enzymatic assays did not produced the expected results and thus further attempts are required.

In non-infectious situations the G3BP1 has been suggested to interact with RAPTOR that belongs to the mTORC1 complex and ASTRIN also known as SPAG5. It is likely the competition of ASTRIN with mTOR for RAPTOR-binding that leads to a decrease in the mTORC1 formation. The complex ASTRIN-RAPTOR is increased by oxidative stress and is probably present within *bona-fide* SGs induced by arsenite, thus preventing mTORC1-hyperactivation-induced apoptosis (Thedieck *et al.*, 2013). Interestingly, the FXR protein members might also regulate the mTOR/S6K signalling pathway (Huang *et al.*, 2016). It remains not completely understandable the process and the reason why nsP3 recruits G3BP1. Accordingly, to the functions of G3BP1 and the known modulation of PI3K-Akt-mTOR pathway during alphavirus infection it may direct us in searching for a co-relation between PI3K-Akt-mTOR pathway and G3BP1.

We can speculate that for some alphaviruses to increase the activation of mTORC1, the G3BP1 needs to interact with ASTRIN-RAPTOR and that nsP3 may be involved in the disruption of the SPAG5-RAPTOR complex by hijacking the G3BP1. Although, this hypothesis seems to have never been addressed so far in the literature. To address this hypothesis, we need to develop starvation assays that could allow us to monitor the composition of the cytoplasmic foci induced by nsP3. It is possible that these cytoplasmic foci may contain proteins that play a role in the PI3K-Akt-mTOR pathway. In the present work we have tested the protocols for starvation experiments already available in the laboratory of our collaborators in Leiden for infection with SFV4 (data not shown). Currently we are trying to find a possible adaptability of the available protocols to answer to our questions regarding the effect of GETV nsP3 in the mTOR pathway. One possibility would be the development of a synthetic virus for GETV that could be used for similar experiments. Alternatively, we could adapt the starvation protocols for *in vitro* assays involving transfection of plasmids expressing the alphavirus nsP3 and monitoring the presence of proteins from the mTOR pathway in the cytoplasmic foci.

We have observed that the presence of mutations in amino acids from the catalytic loop affect the morphology of the cytoplasmic foci formed by G3BP1 in the presence of nsP3 GETV. Unfortunately, we were not managed to find more evidences that could support our original hypothesis and it still inconclusive if the *Macro* domain and the HVR of nsP3 function in a synergistic manner.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

9.4 Conclusion

THIS PAGE IS INTENTIONALLY LEFT BLANK

Although, the mysterious multifunctional role of nsP3 has started to be revealed, there are still many characteristics of nsP3 to remain to be elucidated. A better understanding of the multi-played roles of nsP3 could benefit the search of potent and specific antivirals targeting alphaviruses. Exploring the sequence diversity of alphaviruses, we tried to find evidences that could better propose a catalytic mechanism for the *Macro* domain of nsP3. Our work described different poses adopted by a molecule of ADP-ribose in the binding cleft of GETV *Macro* domain. After comparison with other structure homologues available from PDB databank we found original features such as opening of the distal ribose, stabilization of the opening of the distal ribose by a Ser30 and a covalent link with a Cys34 located in the catalytic loop. Our results pinpointed an important role of this alphavirus specific Cysteine on the functions of alphavirus *Macro* domains, but the relevance of the covalent link remains to be validated by enzymatic assays and reverse genetics. Intriguingly, the cysteine is highly conserved among alphavirus *Macro* domains with exception of insect specific alphaviruses such as Eilat, Mwinilunga and Tai-Forest viruses for which at the place of the cysteine is present a tyrosine (data not shown).

The results from our structural and functional work on GETV *Macro* domain conducted us to try to understand some other peculiar hidden functions of nsP3. We tried to understand if the catalytic residues identified as important for the catalytic activity of *Macro* domain would be also important for the formation of G3BP-containing cytoplasmic foci. These results would be valuable to understand if both *Macro* domain and HVR function in a synergistic manner, likely the HVR contributes to the substrate recognition and *Macro* domain for catalysis. Unfortunately, we were not managed to find strong supportive evidences for our original hypothesis and it still inconclusive if the *Macro* domain and the HVR of nsP3 function in a synergistic manner.

Altogether our results suggest that GETV nsP3 interacts with G3BP1 in a similar manner as the previously observations for CHIKV infected cells (Scholte *et al.*, 2015). The morphology, protein composition and behaviour of the induced G3BP1-containing cytoplasmic foci induced by expression of nsP3 showed to differ from the G3BP1-containing *bona-fide* SGs induced by arsenite. Although, the evaluation of catalytic mutations in the nsP3 *Macro* domain were inconclusive, we observed possible differences of morphology in the G3BP1-containing cytoplasmic foci induced by expression of the GETV nsP3 mutants. In order to conclude that residues from the catalytic loop of the GETV *Macro* domain are crucial for the morphology and stability of the G3BP1-containing cytoplasmic foci, more experiments are highly required.

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

III- General Discussion

THIS PAGE IS INTENTIONALLY LEFT BLANK

Outbreaks caused by alphaviruses are difficult to prevent due to the high mobility of enzootic hosts such as birds and mammals, expansion of bridge vectors and the global trade of domestic animals across the world (Gloria-Soria et al., 2014). Arboviruses, with 11 species from the Alphavirus genus, are thus a major concern of public health. Human pathogenic alphaviruses have emerged and re-emerged during the last decades (CDC, 2018). These viruses can cause debilitating symptoms in patients and are consequently responsible for a significant economic and clinical burden. Notably amongst those are VEEV infections causing neurologic complications in humans (Armstrong & Andreadis, 2013) and CHIKV infections leading sometimes to highly debilitating chronic symptoms (Powers, 2019).

Currently, effective prophylaxis of the major pathogenic alphaviruses such as VEEV and CHIKV are still under development because effective and safe vaccines are still lacking to protect all the individuals at risk of infection. Therefore, safe and efficient therapeutics are highly desirable since treatments are only available to treat the symptoms. The development of new drugs requires a suitable selection of appropriated drug targets which should constitute the basis of the success of drug development. Appropriate drug targets are usually proteins bearing essential functions for virus replication and that can be modulated by small compounds. Accordingly, the presented work was divided in two directions, one reporting compounds targeting a previously validated target, and the second section investigating original alphavirus mechanism that might be latter explored for drug design.

The target used for the drug design work is the mRNA capping machinery, mainly driven by nsP1. The viral mRNA capping is important for alphaviruses to initiate viral translation, while the conserved sequence element CSE1 has been suggested to hide the viral cap type-0 structure from recognition by the host interferon-induced protein with tetratricopeptide repeats 1 (IFIT1) (Daffis *et al.*, 2010, Reynaud *et al.*, 2015, Hyde *et al.*, 2014, Hyde & Diamond, 2015, Hyde *et al.*, 2015). Inhibiting the mRNA capping mechanism would likely induce impaired translation of viral proteins and by consequence compromise viral replication. Critical residues involved in the guanylyl-transferase and methyltransferase activities of the alphavirus nsP1 were earlier identified by using the nsP1 of SFV (Ahola *et al.*, 1997, Laakkonen *et al.*, 1994, Mi & Stollar, 1991, Rozanov *et al.*, 1992). More recent works have demonstrated, by using plasmid-based trans-replication system for SFV, that mutations affecting essential enzymatic activities for the cap formation were responsible for the loss of viral replication (Kallio et al., 2016). Therefore, the enzymatic activities carried out by nsP1 have emerged as a genuine antiviral target and appropriated for enzyme-based drug design.

This nsP1 enzyme is required for at least three steps of the cap synthesis, namely GTP methylation (MTase), nsP1 guanylylation (GT) and cap transfer on mRNA (GTase). Thus, we developed a strategy to select small molecules inhibiting the original guanylylation step of the alphavirus capping mechanism through the detection of the m⁷GMP-nsP1 complex formation. The detection of the formation of the m⁷GMP-nsP1 complex is been used for the development of enzyme-based screening assays for CHIKV nsP1 (Bullard-Feibelman et al., 2016, Feibelman et al., 2018, Kaur et al., 2018). Using VEEV nsP1 as a model, simpler experimental conditions were found by uncoupling the GT reaction from the MTase reaction. The GT reaction conditions were then adapted for an ELISA-based assay to screen small molecules that inhibits the formation of the m⁷GMP-nsP1 adduct. Our work together with works from others (Bullard-Feibelman et al., 2016, Feibelman et al., 2018, Kaur et al., 2018) supports the robustness of the ELISA assays to detect the m⁷GMP-nsP1 adduct to be used for small molecules screening.

Using this methodology, we screened on the VEEV nsP1 GT reaction more than 1200 approved drugs that belong to the Prestwick Chemical library® (PCL®) at a concentration of 50 µM. The results from our screening resulted in the identification of 18 compounds that were effective in inhibiting the GT reaction, among them 9 had IC₅₀ values below the 100 µM range that was determined by a specific Western blot (WB) assay, a technique already used previously (Delang et al., 2016, Gigante et al., 2017).

The screening resulted in the selection of 3 series and each one was represented by one of the best compounds selected from the screening as a head of series. The heads from series 1 and 2 showed a better IC₅₀ on the GT activity of VEEV nsP1 than Sinefungin (29.1 ± 2.6 µM) (Li et al., 2015). The IC₅₀ of the heads from series 1 and 2 is in the same range as compounds already described to have antiviral effect on CHIKV (Gigante et al., 2017). Therefore, search for analogues was performed for compounds Prest-37 (series 1) and Prest-392 (series 2). The selected compounds from the analogue search in the PCL® screened library were tested for the VEEV nsP1 GT reaction. Interestingly the pharmacophore evaluation for Prest-37 highlighted a possible key role of an aliphatic group in the position 6 of the diaminopyrimidine ring. However, a more extended structure activity relationship studies is required due to the low number of analogues tested for this series. Whereas specificity of inhibitors selected against other viral MTase (Flavivirus, Coronavirus) is an issue as they also inhibit cellular MTases (Aouadi *et al.*, 2017), the compounds selected against VEEV nsP1 are selective, highlighting the peculiar capping mechanism of mRNA alphavirus

Altogether, the results suggest that we developed an authentic strategy by implementing an initial screening on the GT activity instead of the MTase activity. The GT activity

carried out by VEEV nsP1 is an original viral activity and it is likely the reason that increased the probability to identify specific compounds. It is importantly to note that none of these compounds show structure similarities when compared to compounds selected from a GTP competition assay on CHIKV nsP1 (Feibelman et al., 2018). In parallel to the work already presented, we have continued with the development of the series of MADTP compounds. Among the 25 MADTP compounds tested we have identified 12 highly potent MADTP compounds bearing an IC₅₀ below 5 µM. These MADTP compounds were better compounds than the MADTPs already reported in other works with an IC₅₀ in the range of 10 µM (Gigante et al., 2014, Gigante et al., 2017).

The multifunctional role of nsP3 was a mystery for decades and only recently were presented some evidences of its functions during viral replication (Gotte *et al.*, 2018). A better understanding of nsP3 function during infection could provide potential efficient targets to be further explored for drug design. Along nsp3, only for its *Macro* domain localized at the N-terminal region (Malet *et al.*, 2009, Li *et al.*, 2016, Park & Griffin, 2009, Egloff *et al.*, 2006, Abraham *et al.*, 2018) and its hyper variable region (Cristea et al., 2006, Frolova et al., 2006, Scholte et al., 2015, Panas et al., 2012, Panas et al., 2014)(Kim et al., 2016, Foy et al., 2013)(Frolov et al., 2017) has been provided functional clues.

One of our aims was to understand the catalytic mechanism carried by *Macro* domain through natural sequence diversity. Thus, we have initiated structure-based studies for GETV *Macro* domain, and it has been suggested that alphavirus *Macro* domains function as a viral countermeasure (Leung *et al.*, 2018) because it seems to be required for ADPr binding and de-ribosylation during both early and late steps of replication (Abraham *et al.*, 2018). Infecting mainly horses, GETV is geographically distributed from Asia to the north of Australia (Fukunaga *et al.*, 2000). It has been reported by others that the strain M1 have a feature of increased replication potency in cancer cell lines that lack expression of zinc-finger antiviral protein (ZAP), a catalytically deficient poly ADP-ribose polymerase (Lin *et al.*, 2014). Indeed, the *Macro* domain from GETV strain M1 presents 69 % identity with the CHIKV *Macro* domain and in the catalytic loop is present a conspicuous substitution of the conserved residues, a glycine to serine in the catalytic loop.

Within our structural work we reported several poses adopted by ADP-ribose among which several ones were suggested to represent several features of the de-ribosylation mechanism. Despite original features belonging to the GETV *Macro* domain reported in this work, the overall fold of GETV *Macro* domain is very well conserved, especially compared to *Macro* domains belonging to the Old-World alphavirus group, and only

minor structural changes can be observed at the level of the β 3- β 4 loop. Among the diversity of different conformations adopted by ADP-ribose obtained from different diffraction data sets, the first obtained was the structure of the distal open ring of the ribose. Initially this conformation was identified when co-crystallizing or soaking the native GETV *Macro* domain with ADP-ribose only.

During the deMARylation reaction, the ADP-ribosylated residues of MARylated proteins are recognized by *Macro* domains that cleave the ester bond between the carboxyl group and the 1''-hydroxyl of the distal ribose of the ADP-ribose. Thus, the Asp or Glu on the surface of the target protein becomes free, and ADP-ribose is released. Expecting that the availability of carboxyl groups might trigger visible conformational changes in our structure, we have added to the crystallization solutions aspartic and glutamic acids together with ADP-ribose. The results of five different conformations helps to better characterize the deMARylation mechanism. Two interesting poses were observed for ADP-ribose in close conformation, "pose1" virtually identical to the pose adopted by ADP-ribose in other alphavirus *Macro* domains and a new "pose 2" in which the distal ribose in β -configuration is tilted by approximately 90° with respect to the "classical" pose 1 of the distal ribose.

Also, in the presence of aspartic or glutamic acid we have characterized a double conformation of the distal ribose in open conformation aside another three-dimensional structure with a single conformation of the ADP-ribose in open conformation as well. It was not possible to model aspartic or glutamic acid in the electron density. In one of our conditions we surprisingly observed the formation of a covalent bond that was established between 1''-C and the SG of the Cys34, a residue conserved in almost all alphavirus *Macro* domains. Remains to be elucidated the reason behind the modification in space group, from P2₁2₁2₁ (as observed in all the previously described structures) to C2, and only one molecule of the GETV *Macro* domain /ADP-ribose complex present in the asymmetric unit. An X-ray induced effect most likely will not promote the change in space group, however we cannot exclude that the covalent intermediate observed in this structure could be a result of an artefact.

Our results on GETV *Macro* domain highlighted a key role for Cys34 in a function carried by alphavirus *Macro* domain. These hypothesis is supported by the finding that Cys34 is isosteric to the putative catalytic residue Asp35 in the *E. coli* phosphatase (1SPV), Asp102 in the human *MacroD2* (4IQY) (Jankevicius *et al.*, 2013) and Glu114 in poly-(ADP)-ribose glycohydrolase (PARG) from *Thermomonospora curvata* (3SIG) (Slade *et al.*, 2011). Thus, we started to address this role by expressing recombinant nsP3 in cells and look for phenotypes. We also postulated that both HVR and *Macro* domains could

function in a synergistic manner because enzymes are known to present regions involved in substrate recognition and catalysis and, most likely in this case we expected that the HVR contributes to the substrate recognition and *Macro* domain for catalysis.

Studies were not available for GETV nsP3 in what concerns to the formation of the cytoplasmic foci that resemble bona-fide SGs. The bona-fide SGs are storage sites of cytoplasmic mRNAs that are released from disassembled polysomes under stress conditions. Thus, the bona-fide SGs are important for selective translation of stress-induced genes and G3BP1/2 are components from these cytoplasmic foci that initiates the assembly of the bona-fide SGs through stress-induced multimerization (Matsuki *et al.*, 2013). Our work with cell-based assays suggested that similarly to CHIKV (Fros *et al.*, 2012, Scholte *et al.*, 2015) and SFV (Panas *et al.*, 2012), the wt GETV nsP3 inhibits the bona-fide stress granule assembly by recruiting G3BP1 into cytoplasmic foci.

Accordingly to what has been described so far in the literature, OW alphaviruses such as CHIKV and SFV are known to use short repeating a.a. sequences located at the carboxy terminus of the HVR, the FGDF motifs, to interact with G3BP1/2 proteins (Vognsen *et al.*, 2013)(Schulte *et al.*, 2016, Panas *et al.*, 2014). Thus, likely the interaction of the wt GETV nsP3 with G3BP1 is mediated by the FGDL and FGDI motifs which are present in the carboxy terminus of the HVR in the GETV nsP3. Interestingly, the cytoplasmic foci formed by overexpression of nsP3 in Vero E6 cells are different from bona-fide SGs in terms of morphology, composition, and behaviour as it was observed for CHIKV (Scholte *et al.*, 2015). In our experiments we observed a rod-like morphology of the G3BP1-containing foci which were reproducible (Fig. 46 C). As expected, and accordingly to the literature (Kim *et al.*, 2016, Foy *et al.*, 2013), when Vero E6 cells were transfected with the VEEV nsP3^{wt} cloned into the IRES-pCAGGS, the G3BP1 -containing foci were not observed (Fig. 46 B).

The effect of catalytic mutations in the nsP3 *Macro* domain was also addressed. The amino acids selected for this work were the Asn24, Asp31 and Cys34. The Asn24 is involved in interactions with the distal ribose and the residue Asp31 is included within the $\beta 2\alpha 1$ loop (amino acid 26 to 33) and it is likely one of the catalytic residues. All the GETV mutants nsP3^{Asn24Ala}, nsP3^{Cys34Ala} and nsP3^{Asp31Ala} induced the formation of G3BP1-containing cytoplasmic foci, however the morphology for the mutants nsP3^{Asn24Ala} and nsP3^{Asp31Ala} (Fig.46 E and G) was different from the nsP3^{wt} (Fig. 46 D). The morphology of those mutants resembles the morphology observed for SGs induced with arsenite (Fig.46 A) as they were more round-like G3BP1-cytoplasmic foci, although presenting inferior sizes. However, for the mutant nsP3^{Cys34Ala} (Fig.46 F) the morphology is still like the nsP3^{wt} (Fig.45 D).

As the formation of bona-fide SG requires the disassembly of polysomes, we stabilized the polysomes by treating the cells with cycloheximide (CHX) as it was performed in works from others (Scholte et al., 2015). Interestingly, the CHX treatment did not dispersed the G3BP-containing cytoplasmic foci for the assays with GETV nsP3^{wt} and nsP3^{Asn24Ala}, nsP3^{Cys35Ala} and nsP3^{Asp31Ala} mutants (Fig.47). However, it remains to confirm if for mutants the G3BP1-containing cytoplasmic foci were weakly dispersed, and alterations in the morphology were observed for GETV nsP3^{Cys35Ala} (Fig. 47 D and I). Therefore, it is possible that the selected residues are involved at some point with the formation of the G3BP-containing cytoplasmic foci through interaction with partner proteins, for example. However, all these experiments are preliminary and remain to be confirmed.

We have investigated the presence of markers of the bona-fide SGs such as; G3BP1, TIA-1, TIAR, PABP and eIF3 as it was already performed for works from others (Scholte et al., 2015). Our works were comparable to what has been observed for CHIKV (Fros et al., 2012, Scholte et al., 2015) and SFV (Panas et al., 2012). The protein composition of the cytoplasmic foci induced by the expression of GETV nsP3^{wt} or nsP3 mutants is different from the *bona-fide* SGs induced by arsenite, Fig. 45. Other attempts to characterize the effect of mutations in critical residues for ADP-ribose hydrolysis through enzymatic assays did not produced the expected results and thus further attempts are required.

In non-infectious situations the G3BP1 has been suggested to interact with RAPTOR, that belongs to the mTORC1 complex, and with ASTRIN, also known as SPAG5 (Thedieck et al., 2013). It is likely the competition of ASTRIN with mTOR for RAPTOR-binding that leads to a decrease in the mTORC1 formation. The complex ASTRIN-RAPTOR is increased by oxidative stress and is probably present within bona-fide SGs induced by arsenite, thus preventing mTORC1-hyperactivation-induced apoptosis (Thedieck et al., 2013). Interestingly, the FXR protein members might also regulate the mTOR/S6K signalling pathway (Huang et al., 2016). The process remains not completely understandable as well as why nsP3 recruits G3BP1. The functions of G3BP1 and the known modulation of PI3K-Akt-mTOR pathway during alphavirus infection may point to searching for a co-relation between PI3K-Akt-mTOR pathway and G3BP1.

One can speculate that for some alphaviruses G3BP1 interacts with ASTRIN-RAPTOR to activate mTORC1 and that nsP3 may be involved in the disruption of the ASTRIN-RAPTOR complex by hijacking G3BP1. Although, this hypothesis seems to have never been addressed so far in the literature. To address this hypothesis, we need to develop starvation assays that could allow us to monitor the composition of the cytoplasmic foci

induced by nsP3. It is possible that these cytoplasmic foci may contain proteins that play a role in the PI3K-Akt-mTOR pathway. In the present work we have tested the protocols for starvation experiments already available in the Leiden for infection with SFV (data not shown). Currently we are trying to find a possible adaptability of the available protocols to answer to our questions regarding the effect of GETV nsP3 in the mTOR pathway. One possibility would be the development of a reverse genetic system for GETV that could be used for similar experiments. Alternatively, we could adapt the starvation protocols for in vitro assays involving transfection of plasmids expressing the alphavirus nsP3 and monitoring the presence of proteins from the mTOR pathway in the cytoplasmic foci.

The multifunctional role of nsP3 has been suggested to be very important for minus-strand or sub-genomic RNA synthesis, viral replication and pathogenesis. (LaStarza, Lemm, et al., 1994, Rupp et al., 2011, Wang et al., 1994)(Hahn, Strauss, et al., 1989, Lemm et al., 1994, Shirako & Strauss, 1994). Interestingly, experiments from others showed that mutations to Ala of residues localized in the ADP-ribose binding region of SINV *Macro* domain impaired SINV replication and viral RNA synthesis in neurons without alter the binding of poly-(ADP)-ribose (Park & Griffin, 2009). Such residues were the Asn10 and Asn24 and interestingly when the Asn10 was mutated to Ala it was attenuated the virulence in 2-week-old mice. It was curious the they observed a compensatory mutation with the mutant Asn10 that occurred at the residue 31 from a Glu to a Gly but without increase the plate size (Park & Griffin, 2009). Altogether their work suggested that *Macro* domain from SINV is important for viral replication and neurovirulence in mice (Park & Griffin, 2009).

Interestingly, one of our GETV mutation was performed also the Asn24 to Ala and in our assays, we observed that despite inducing the formation of G3BP1-containing cytoplasmic foci, the morphology of these cytoplasmic foci was different than the cytoplasmic foci induced by the wt GETV nsP3. Like the GETV nsP3 Asn24 mutant, the same was observed for the mutant nsP3 Asp31 to Ala. Interestingly the morphology of mutant GETV nsP3 Cys34 to Ala presented similar morphology to the nsP3^{wt}.

Other works have reported that defective mutants in nsP3 phosphorylation exhibited a decrease rate of RNA synthesis and a strong reduction of pathogenicity in mice (Vihinen & Saarinen, 2000). The HVR is also likely a determinant of pathogenesis through interactions with cell-type-specific factors, reviewed in (Rupp et al., 2015), and the interaction with G3BP1/2 in OW alphaviruses reduces the translation of viral RNAs to allow the switch to genome amplification during early infection (Rupp *et al.*, 2015, Fros *et al.*, 2012, Scholte *et al.*, 2015). Similarly, to OW alphaviruses, the NW alphaviruses

through FXR proteins are known to interact with the HVR of nsP3 which is also likely crucial for infectivity, RNA replication initiation and rates of replication (Kim et al., 2016).

We have observed that the presence of mutations in amino acids from the catalytic loop may affect the morphology of the cytoplasmic foci formed by G3BP1 in the presence of nsP3 GETV. Unfortunately, we did not manage to find more evidence that could support our hypothesis and it remains still inconclusive whether the *Macro* domain and the HVR of nsP3 would function in a synergistic manner. Interestingly, our results from the structural work performed for the GETV *Macro* domain also highlighted a key role for Cys34 in a function carried by alphavirus *Macro* domain. Further investigation on alphavirus nsP3 are highly necessary to better understand its function during replication.

Thus, nsP3 is a potential attractive target for drug design and a better elucidation of its role during the viral replication would be an advantage to select the appropriated nsP3 activity/function to be targeted by drug design. Efficiently targeting the nsP3 could lead to the switch off the viral replication and clearance of the viral infection.

FROM THIS POINT FORWARD THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

V- General Conclusions

THIS PAGE IS INTENTIONALLY LEFT BLANK

Overall, our results together with work from others are supporting the use of the alphavirus mRNA capping mechanism as an excellent target for the development of molecules with antiviral effect. We have implemented our assay using an effective strategy to screen more than 1200 compounds on the GT reaction of VEEV nsP1, and through the inclusion of complementary assays we overcame the lack of structural information of the alphavirus nsP1. The analysis of the screening resulted in the selection of 3 series of compounds that were further characterized. Only for series 2 was identified a more potent analogue, altanserin, than the series head. More efforts are required to better understand the mode of action of the selected series of compounds. Interestingly, the selected compounds seem to possess an unconventional mechanism of action when compared to reference molecules. Thus, altanserin seems to be a promising compound for drug repurposing with a potential to treat alphavirus infections.

With our structural and functional work, we have identified new important characteristics of alphavirus *Macro* domains that can help us to better understand their catalytic mechanism. A better understanding of these mechanism could help us to design a strategy to screen compounds targeting the *Macro* domains. This could be an asset for the drug design of compounds inhibiting alphaviruses. The structural and functional work performed on GETV *Macro* domain revealed different conformations of the ADPr in the catalytic site. For the first time we presented a three-dimensional structure of an alphavirus *Macro* domain with an ADP-ribose presenting the distal ribose in an open conformation inside the catalytic site. Comparison with other structure homologues available from PDB databank highlighted original features for GETV *Macro* domain such as opening of the distal ribose and stabilization of the opening of the distal ribose by Ser30. As it was possible to observe from the snapshots of GETV *Macro* domain of the ADP-ribose in open conformation and from snapshots with the ADP-ribose covalently linked to the Cys34 of GETV *Macro* domain. Our structural work revealed new conformations adopted by the ADPr molecule and highlighted for the first time a highly conserved cysteine in the $\beta 2\alpha 1$ loop that might be important for the catalytic mechanism carried out by *Macro* domain. Intriguingly, the cysteine is highly conserved among alphavirus *Macro* domains with exception of insect specific alphaviruses such as Eilat, Mwinilunga and Tai-Forest viruses for which the cysteine is replaced by a tyrosine (data not shown). Thus, it is possible that the cysteine could be involved in specific pathways specific of vertebrate hosts.

These results suggested that it could be important to understand some other peculiar hidden functions of nsP3. We tried to understand if the catalytic residues identified as important for the catalytic activity of *Macro* domain would be also important for the

formation of G3BP-containing cytoplasmic foci. Altogether our results suggest that the wt GETV nsP3 interacts with G3BP1 in a similar manner as the previously observations for CHIKV infected cells (Scholte et al., 2015). The morphology, protein composition and behaviour of the induced G3BP1-containing cytoplasmic foci induced by expression of nsP3 showed to differ from the G3BP1-containing bona-fide SGs induced by arsenite. Although, further validation assays are required. Indeed, more work is necessary to understand if both the *Macro* domain and the HVR of nsP3 function in a synergistic manner. By exploring the available natural sequence diversity of alphavirus nsP3 it was possible to select an alphavirus which possess a *Macro* domain with peculiar differences compared to the ones already reported. The information obtained from the structural data obtained with this work revealed for the first time new structural information that might be a piece of the puzzle of the catalytic mechanism carried out by alphavirus *Macro* domains. This new structural information might contribute for a better understanding of the *Macro* domain catalytic mechanism and which it will be important for an efficient drug design. A better understand of the catalytic mechanism of alphavirus *Macro* domain it will contribute for the development of more effective, specific and potent inhibitors. For the specificity requirement it would be also important to better understand the differences between the catalytic mechanism carried out by the alphavirus and human *Macro* domains. The development of drugs targeting the functions of alphavirus nsP3 are promising and this would be an important hallmark towards overcoming the lack of therapeutics to treat alphavirus infections.

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

VI- Annex

THIS PAGE IS INTENTIONALLY LEFT BLANK

10. Additional collaboration

THIS PAGE IS INTENTIONALLY LEFT BLANK

10.1. Computer-aided design, synthesis and evaluation of novel compounds against Chikungunya nsP3

THIS PAGE IS INTENTIONALLY LEFT BLANK

The scope of this collaboration under the ANTIVIRALS consortium was to provide to other ESRs opportunities to have a chance to use the techniques available in our laboratory and to help them to develop their own projects.

To do so, I have provided to the Cardiff team the first TSA results of the compounds that they have selected from their work. Despite the results showing no binding to the Chikungunya *Macro* domain we welcome the ESR 8 for a secondment in our laboratory to test the compounds by TSA and thermophoresis. I have supported the ESR 8 during her secondment and showed to her our established protocols for plasmid extraction, protein production, protein purification, TSA, thermophoresis, protein crystallization and co-crystallization. After her secondment of two weeks, I have finalized the work by taking the crystals to the synchrotron and collected the data. Unfortunately, the compounds were not present in the crystal structure.

The article of this work is currently under preparation and one option of journal to submit is the “Antiviral Research” journal.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Computer-aided discovery, synthesis and *in vitro* mode-of-action studies of novel inhibitors against chikungunya virus.

Birgit Zonsics,^A Rana Abdelnabi,^B Leen Delang,^B Ana-Sofia Ferreira-Ramos,^C Gilles Querat,^D Bruno Canard,^C Bruno Coutard,^{C,D} Johan Neyts,^B Andrea Brancale^{A*}

^A Department of Medicinal Chemistry, Welsh School of Pharmacy and Pharmaceutical Sciences, Cardiff University, King Edward VII avenue, CF10 3NB, Cardiff, United Kingdom

^B Rega Institute for Medical Research, Laboratory of Virology and Chemotherapy, KU Leuven, B-3000, Leuven, Belgium

^C Aix Marseille Université, CNRS, AFMB UMR 7257, Marseille, France

^D Unité des Virus Emergents (UVE) Aix-Marseille Univ-IRD 190-Inserm 1207-IHU Méditerranée Infection, Marseille, France

*Correspondence should be addressed to Andrea Brancale

E-mail: brancalea@cardiff.ac.uk

Tel: +44 29 2087 4485

keywords

Alphavirus; antivirals; chikungunya;

Abstract

Chikungunya virus (CHIKV) is an arthropod-borne (arbovirus) (+)ssRNA virus belonging to the genus Alphavirus, family *Togaviridae*. CHIKV causes chikungunya fever, which is mostly characterized by fever, arthralgia and, sometimes, a maculopapular rash. Although a CHIKV infection is rarely fatal, the disease proceeds in 15-60% of infected patients into a chronic persistent disabling polyarthritis. Subjects can be severely incapacitated for many weeks up to several years. Due to its re-emergence and the considerable disease burden associated with the infection, CHIKV has become a substantial global health threat. A vaccine is currently under development but selective and potent antiviral therapy against CHIKV still urgently needed.

The CHIKV genome encodes four non-structural proteins (nsP1-4) that form the viral replication complex responsible of the synthesis of the viral RNA. The crystallized macro domain is part of the nsP3 protein and was chosen as a target for structure-based pharmacophore modelling and docking studies to identify new potential CHIKV inhibitors. The *in-silico* hits were purchased and tested in a cell-based antiviral assay revealing four active compounds the best of which in the low micromolar range. A small library of analogues of the most potent hit was synthesized to highlight the essential features of the hit compound and to find molecules with improved potency. In parallel, mechanistic studies were performed with the hit compound and the most promising analogues to confirm whether nsP3 is their molecular target. In summary, we here report a series of chemically novel anti-CHIKV compounds, with some preliminary studies to clarify their mode-of-action.

THIS PAGE IS INTENTIONALLY LEFT BLANK

THIS PAGE IS INTENTIONALLY LEFT BLANK

VII- REFERENCES

THIS PAGE IS INTENTIONALLY LEFT BLANK

- Aaskov, J., Jones, A., Choi, W., Lowry, K. & Stewart, E. (2011). *Virology* **410**, 353-359.
- Abbas, Y. M., Pichlmair, A., Gorna, M. W., Superti-Furga, G. & Nagar, B. (2013). *Nature* **494**, 60-64.
- Abdelnabi, R., Neyts, J. & Delang, L. (2015). *Antiviral Res* **121**, 59-68.
- Abraham, R., Hauer, D., McPherson, R. L., Utt, A., Kirby, I. T., Cohen, M. S., Merits, A., Leung, A. K. L. & Griffin, D. E. (2018). *Proc Natl Acad Sci U S A* **115**, E10457-E10466.
- Abu Bakar, F. & Ng, L. F. P. (2018). *Viruses* **10**.
- AbuBakar, S., Sam, I. C., Wong, P. F., MatRahim, N., Hooi, P. S. & Roslan, N. (2007). *Emerg Infect Dis* **13**, 147-149.
- Adams, A. P., Navarro-Lopez, R., Ramirez-Aguilar, F. J., Lopez-Gonzalez, I., Leal, G., Flores-Mayorga, J. M., Travassos da Rosa, A. P., Saxton-Shaw, K. D., Singh, A. J., Borland, E. M., Powers, A. M., Tesh, R. B., Weaver, S. C. & Estrada-Franco, J. G. (2012). *PLoS Negl Trop Dis* **6**, e1875.
- Adams, J. M. & Cory, S. (1975). *Nature* **255**, 28-33.
- Adkins, S., Stawicki, S. S., Faurote, G., Siegel, R. W. & Kao, C. C. (1998). *RNA* **4**, 455-470.
- Aguilar, P. V., Adams, A. P., Suarez, V., Beingolea, L., Vargas, J., Manock, S., Freire, J., Espinoza, W. R., Felices, V., Diaz, A., Liang, X., Roca, Y., Weaver, S. C. & Kochel, T. J. (2009). *PLoS Negl Trop Dis* **3**, e514.
- Aguilar, P. V., Estrada-Franco, J. G., Navarro-Lopez, R., Ferro, C., Haddow, A. D. & Weaver, S. C. (2011). *Future Virol* **6**, 721-740.
- Aguilar, P. V., Greene, I. P., Coffey, L. L., Medina, G., Moncayo, A. C., Anishchenko, M., Ludwig, G. V., Turell, M. J., O'Guinn, M. L., Lee, J., Tesh, R. B., Watts, D. M., Russell, K. L., Hice, C., Yanoviak, S., Morrison, A. C., Klein, T. A., Dohm, D. J., Guzman, H., Travassos da Rosa, A. P., Guevara, C., Kochel, T., Olson, J., Cabezas, C. & Weaver, S. C. (2004). *Emerg Infect Dis* **10**, 880-888.
- Ahel, D., Horejsi, Z., Wiechens, N., Polo, S. E., Garcia-Wilson, E., Ahel, I., Flynn, H., Skehel, M., West, S. C., Jackson, S. P., Owen-Hughes, T. & Boulton, S. J. (2009). *Science* **325**, 1240-1243.
- Ahola, T. & Kaariainen, L. (1995). *Proc Natl Acad Sci U S A* **92**, 507-511.
- Ahola, T. & Karlin, D. G. (2015). *Biol Direct* **10**, 16.
- Ahola, T., Kujala, P., Tuittila, M., Blom, T., Laakkonen, P., Hinkkanen, A. & Auvinen, P. (2000). *J Virol* **74**, 6725-6733.
- Ahola, T., Laakkonen, P., Vihinen, H. & Kaariainen, L. (1997). *J Virol* **71**, 392-397.
- Ahola, T., Lampio, A., Auvinen, P. & Kaariainen, L. (1999). *EMBO J* **18**, 3164-3172.

- Akahata, W., Yang, Z. Y., Andersen, H., Sun, S., Holdaway, H. A., Kong, W. P., Lewis, M. G., Higgs, S., Rossmann, M. G., Rao, S. & Nabel, G. J. (2010). *Nat Med* **16**, 334-338.
- Akhrymuk, I., Frolov, I. & Frolova, E. I. (2016). *Virology* **487**, 230-241.
- Akhrymuk, I., Kulemzin, S. V. & Frolova, E. I. (2012). *J Virol* **86**, 7180-7191.
- Akira, S., Uematsu, S. & Takeuchi, O. (2006). *Cell* **124**, 783-801.
- Aldinucci, A., Gerlini, G., Fossati, S., Cipriani, G., Ballerini, C., Biagioli, T., Pimpinelli, N., Borgognoni, L., Massacesi, L., Moroni, F. & Chiarugi, A. (2007). *J Immunol* **179**, 305-312.
- Alevizatos, A. C., McKinney, R. W. & Feigin, R. D. (1967). *Am J Trop Med Hyg* **16**, 762-768.
- Aliperti, G. & Schlesinger, M. J. (1978). *Virology* **90**, 366-369.
- Allen, M. D., Buckle, A. M., Cordell, S. C., Lowe, J. & Bycroft, M. (2003). *J Mol Biol* **330**, 503-511.
- Altmeyer, M., Messner, S., Hassa, P. O., Fey, M. & Hottiger, M. O. (2009). *Nucleic Acids Res* **37**, 3723-3738.
- Anderson, P. & Kedersha, N. (2008). *Trends Biochem Sci* **33**, 141-150.
- Anishchenko, M., Bowen, R. A., Paessler, S., Austgen, L., Greene, I. P. & Weaver, S. C. (2006). *Proc Natl Acad Sci U S A* **103**, 4994-4999.
- Anthony, R. P., Paredes, A. M. & Brown, D. T. (1992). *Virology* **190**, 330-336.
- Aouadi, W., Blanjoie, A., Vasseur, J. J., Debart, F., Canard, B. & Decroly, E. (2017). *J Virol* **91**.
- Arakawa, H., Hauschild, J. & Buerstedde, J. M. (2002). *Science* **295**, 1301-1306.
- Arankalle, V. A., Shrivastava, S., Cherian, S., Gunjekar, R. S., Walimbe, A. M., Jadhav, S. M., Sudeep, A. B. & Mishra, A. C. (2007). *J Gen Virol* **88**, 1967-1976.
- Arias-Goeta, C., Mousson, L., Rougeon, F. & Failloux, A. B. (2013). *PLoS One* **8**, e57548.
- Armstrong, P. M. & Andreadis, T. G. (2013). *N Engl J Med* **368**, 1670-1673.
- Ashbrook, A. W., Lentscher, A. J., Zamora, P. F., Silva, L. A., May, N. A., Bauer, J. A., Morrison, T. E. & Dermody, T. S. (2016). *MBio* **7**.
- Atasheva, S., Akhrymuk, M., Frolova, E. I. & Frolov, I. (2012). *J Virol* **86**, 8147-8160.
- Atasheva, S., Frolova, E. I. & Frolov, I. (2014). *J Virol* **88**, 2116-2130.
- Atasheva, S., Gorchakov, R., English, R., Frolov, I. & Frolova, E. (2007). *J Virol* **81**, 5046-5057.
- Bachmair, A., Finley, D. & Varshavsky, A. (1986). *Science* **234**, 179-186.
- Bai, P. (2015). *Mol Cell* **58**, 947-958.
- Bai, P. & Virag, L. (2012). *FEBS Lett* **586**, 3771-3777.

- Barkauskaite, E., Brassington, A., Tan, E. S., Warwicker, J., Dunstan, M. S., Banos, B., Lafite, P., Ahel, M., Mitchison, T. J., Ahel, I. & Leys, D. (2013). *Nat Commun* **4**, 2164.
- Barkauskaite, E., Jankevicius, G. & Ahel, I. (2015). *Mol Cell* **58**, 935-946.
- Barkauskaite, E., Jankevicius, G., Ladurner, A. G., Ahel, I. & Timinszky, G. (2013). *FEBS J* **280**, 3491-3507.
- Barrera, R., Ferro, C., Navarro, J. C., Freier, J., Liria, J., Salas, R., Ahumada, M., Vasquez, C., Gonzalez, M., Kang, W., Boshell, J. & Weaver, S. C. (2002). *Am J Trop Med Hyg* **67**, 324-334.
- Barton, D. J., Sawicki, S. G. & Sawicki, D. L. (1991). *J Virol* **65**, 1496-1506.
- Bassetto, M., De Burghgraeve, T., Delang, L., Massarotti, A., Coluccia, A., Zonta, N., Gatti, V., Colombano, G., Sorba, G., Silvestri, R., Tron, G. C., Neyts, J., Leyssen, P. & Brancale, A. (2013). *Antiviral Res* **98**, 12-18.
- Beck, C. E. & Wyckoff, R. W. (1938). *Science* **88**, 530.
- Beckham, C. J. & Parker, R. (2008). *Cell Host Microbe* **3**, 206-212.
- Beier, H. & Grimm, M. (2001). *Nucleic Acids Res* **29**, 4767-4782.
- Ben-Ishai, Z., Goldblum, N. & Becker, Y. (1968). *J Gen Virol* **2**, 365-375.
- Berrow, N. S., Bussow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G. E., Levy, N., Lieu, V., Owens, R. J., Peleg, Y., Pinaglia, C., Quevillon-Cheruel, S., Salim, L., Scheich, C., Vincentelli, R. & Busso, D. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1218-1226.
- Berthold, C. L., Wang, H., Nordlund, S. & Hoggom, M. (2009). *Proc Natl Acad Sci U S A* **106**, 14247-14252.
- Betz, C. & Hall, M. N. (2013). *J Cell Biol* **203**, 563-574.
- Bhalla, N., Sun, C., Matthew Lam, L. K., Gardner, C. L., Ryman, K. D. & Klimstra, W. B. (2016). *Virology* **496**, 147-165.
- Bick, M. J., Carroll, J. W., Gao, G., Goff, S. P., Rice, C. M. & MacDonald, M. R. (2003). *J Virol* **77**, 11555-11562.
- Bigler, W. J., Ventura, A. K., Lewis, A. L., Wellings, F. M. & Ehrenkranz, N. J. (1974). *Am J Trop Med Hyg* **23**, 513-521.
- Binder, M., Eberle, F., Seitz, S., Mucke, N., Huber, C. M., Kiani, N., Kaderali, L., Lohmann, V., Dalpke, A. & Bartenschlager, R. (2011). *J Biol Chem* **286**, 27278-27287.
- Bird, J. G., Zhang, Y., Tian, Y., Panova, N., Barvik, I., Greene, L., Liu, M., Buckley, B., Krasny, L., Lee, J. K., Kaplan, C. D., Ebright, R. H. & Nickels, B. E. (2016). *Nature* **535**, 444-447.
- Birdwell, C. R., Strauss, E. G. & Strauss, J. H. (1973). *Virology* **56**, 429-438.

- Boivin, S., Cusack, S., Ruigrok, R. W. & Hart, D. J. (2010). *J Biol Chem* **285**, 28411-28417.
- Bonatti, S. & Blobel, G. (1979). *J Biol Chem* **254**, 12261-12264.
- Bonatti, S., Migliaccio, G., Blobel, G. & Walter, P. (1984). *Eur J Biochem* **140**, 499-502.
- Bonfiglio, J. J., Fontana, P., Zhang, Q., Colby, T., Gibbs-Seymour, I., Atanassov, I., Bartlett, E., Zaja, R., Ahel, I. & Matic, I. (2017). *Mol Cell* **65**, 932-940 e936.
- Bouloy, M., Plotch, S. J. & Krug, R. M. (1978). *Proc Natl Acad Sci U S A* **75**, 4886-4890.
- Bowen, G. S., Fashinell, T. R., Dean, P. B. & Gregg, M. B. (1976). *Bull Pan Am Health Organ* **10**, 46-57.
- Brault, A. C., Powers, A. M., Ortiz, D., Estrada-Franco, J. G., Navarro-Lopez, R. & Weaver, S. C. (2004). *Proc Natl Acad Sci U S A* **101**, 11344-11349.
- Breakwell, L., Dosenovic, P., Karlsson Hedestam, G. B., D'Amato, M., Liljestrom, P., Fazakerley, J. & McInerney, G. M. (2007). *J Virol* **81**, 8677-8684.
- Brennan, K. & Bowie, A. G. (2010). *Curr Opin Microbiol* **13**, 503-507.
- Brighton, S. W. (1984). *S Afr Med J* **66**, 217-218.
- Briolant, S., Garin, D., Scaramozzino, N., Jouan, A. & Crance, J. M. (2004). *Antiviral Res* **61**, 111-117.
- Brizzi, K. (2017). *Curr Infect Dis Rep* **19**, 6.
- Brown, R. S., Wan, J. J. & Kielian, M. (2018). *Viruses* **10**.
- Bruns, A. M., Leser, G. P., Lamb, R. A. & Horvath, C. M. (2014). *Mol Cell* **55**, 771-781.
- Buchan, J. R. & Parker, R. (2009). *Mol Cell* **36**, 932-941.
- Buitendijk, M., Eszterhas, S. K. & Howell, A. L. (2013). *AIDS Res Hum Retroviruses* **29**, 907-918.
- Bujnicki, J. M. & Rychlewski, L. (2002). *Protein Eng* **15**, 101-108.
- Bullard-Feibelman, K. M., Fuller, B. P. & Geiss, B. J. (2016). *PLoS One* **11**, e0158923.
- Burge, B. W. & Strauss, J. H., Jr. (1970). *J Mol Biol* **47**, 449-466.
- Burzio, L. O., Riquelme, P. T. & Koide, S. S. (1979). *J Biol Chem* **254**, 3029-3037.
- Busch, H., Reddy, R., Rothblum, L. & Choi, Y. C. (1982). *Annu Rev Biochem* **51**, 617-654.
- Butepage, M., Ecke, L., Verheugd, P. & Luscher, B. (2015). *Cells* **4**, 569-595.
- Butler, A. J. & Ordahl, C. P. (1999). *Mol Cell Biol* **19**, 296-306.
- Calabretta, S. & Richard, S. (2015). *Trends Biochem Sci* **40**, 662-672.
- Calisher, C. H., Murphy, F. A., France, J. K., Lazuick, J. S., Muth, D. J., Steck, F., Lindsey, H. S., Bauer, S. P., Buff, E. E. & Schneider, N. J. (1980). *South Med J* **73**, 1548.
- Cao-Lormeau, V. M. & Musso, D. (2014). *Lancet* **384**, 1571-1572.

- Care, S., Bignon, C., Pelissier, M. C., Blanc, E., Canard, B. & Coutard, B. (2008). *Nucleic Acids Res* **36**, e6.
- Carey, D. E. (1971). *J Hist Med Allied Sci* **26**, 243-262.
- Carrasco, L. (1995). *Adv Virus Res* **45**, 61-112.
- Casamassima, A. C., Hess, L. W. & Marty, A. (1987). *Teratology* **36**, 287-289.
- Cassadou, S., Boucau, S., Petit-Sinturel, M., Huc, P., Leparac-Goffart, I. & Ledrans, M. (2014). *Euro Surveill* **19**.
- Caton, A. J. & Robertson, J. S. (1980). *Nucleic Acids Res* **8**, 2591-2603.
- Cavrini, F., Gaibani, P., Pierro, A. M., Rossini, G., Landini, M. P. & Sambri, V. (2009). *J Infect Dev Ctries* **3**, 744-752.
- Cecilia, D., Kakade, M., Alagarasu, K., Patil, J., Salunke, A., Parashar, D. & Shah, P. S. (2015). *Arch Virol* **160**, 323-327.
- Centers for Disease, C. & Prevention (1995). *MMWR Morb Mortal Wkly Rep* **44**, 721-724.
- Centers for Disease Control and Prevention [CDC] (2018). "Chikungunya for health care providers", <https://www.cdc.gov/chikungunya/hc/index.html>.
- Cervellera, M. N. & Sala, A. (2000). *J Biol Chem* **275**, 10692-10696.
- Chakravarthy, S., Gundimella, S. K., Caron, C., Perche, P. Y., Pehrson, J. R., Khochbin, S. & Luger, K. (2005). *Mol Cell Biol* **25**, 7616-7624.
- Chamberlain, R. W., Sudia, W. D., Coleman, P. H. & Work, T. H. (1964). *Science* **145**, 272-274.
- Chamberlain, R. W., Sudia, W. D., Work, T. H., Coleman, P. H., Newhouse, V. F. & Johnston, J. G., Jr. (1969). *Am J Epidemiol* **89**, 197-210.
- Chang, J. H., Jiao, X., Chiba, K., Oh, C., Martin, C. E., Kiledjian, M. & Tong, L. (2012). *Nat Struct Mol Biol* **19**, 1011-1017.
- Chang, L. J., Dowd, K. A., Mendoza, F. H., Saunders, J. G., Sitar, S., Plummer, S. H., Yamshchikov, G., Sarwar, U. N., Hu, Z., Enama, M. E., Bailer, R. T., Koup, R. A., Schwartz, R. M., Akahata, W., Nabel, G. J., Mascola, J. R., Pierson, T. C., Graham, B. S., Ledgerwood, J. E. & Team, V. R. C. S. (2014). *Lancet* **384**, 2046-2052.
- Changela, A., Ho, C. K., Martins, A., Shuman, S. & Mondragon, A. (2001). *EMBO J* **20**, 2575-2586.
- Chatterjee, A., Johnson, M. A., Serrano, P., Pedrini, B., Joseph, J. S., Neuman, B. W., Saikatendu, K., Buchmeier, M. J., Kuhn, P. & Wuthrich, K. (2009). *J Virol* **83**, 1823-1836.

- Chen, D., Vollmar, M., Rossi, M. N., Phillips, C., Kraehenbuehl, R., Slade, D., Mehrotra, P. V., von Delft, F., Crosthwaite, S. K., Gileadi, O., Denu, J. M. & Ahel, I. (2011). *J Biol Chem* **286**, 13261-13271.
- Chen, K. C., Kam, Y. W., Lin, R. T., Ng, M. M., Ng, L. F. & Chu, J. J. (2013). *Virology* **10**, 169.
- Chen, R., Wang, E., Tsetsarkin, K. A. & Weaver, S. C. (2013). *PLoS Pathog* **9**, e1003591.
- Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y. & Liu, D. R. (2009). *Nat Chem Biol* **5**, 879-881.
- Chen, Z. J. (2012). *Immunol Rev* **246**, 95-106.
- Cheng, R. H., Kuhn, R. J., Olson, N. H., Rossmann, M. G., Choi, H. K., Smith, T. J. & Baker, T. S. (1995). *Cell* **80**, 621-630.
- Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. (1991). *Nature* **354**, 37-43.
- Choi, Y. J., Bowman, J. W. & Jung, J. U. (2018). *Immunity* **48**, 474-476.
- Chopra, A., Saluja, M. & Venugopalan, A. (2014). *Arthritis Rheumatol* **66**, 319-326.
- Chow, A., Her, Z., Ong, E. K., Chen, J. M., Dimatatac, F., Kwek, D. J., Barkham, T., Yang, H., Renia, L., Leo, Y. S. & Ng, L. F. (2011). *J Infect Dis* **203**, 149-157.
- Chretien, J. P., Anyamba, A., Bedno, S. A., Breiman, R. F., Sang, R., Sergon, K., Powers, A. M., Onyango, C. O., Small, J., Tucker, C. J. & Linthicum, K. J. (2007). *Am J Trop Med Hyg* **76**, 405-407.
- Chung, B. Y., Firth, A. E. & Atkins, J. F. (2010). *J Mol Biol* **397**, 448-456.
- Chung, D. H., Jonsson, C. B., Tower, N. A., Chu, Y. K., Sahin, E., Golden, J. E., Noah, J. W., Schroeder, C. E., Sotsky, J. B., Sosa, M. I., Cramer, D. E., McKellip, S. N., Rasmussen, L., White, E. L., Schmaljohn, C. S., Julander, J. G., Smith, J. M., Filone, C. M., Connor, J. H., Sakurai, Y. & Davey, R. A. (2014). *PLoS Pathog* **10**, e1004213.
- Clarkson, C., Jacobs, Z., Marwick, B., Fullagar, R., Wallis, L., Smith, M., Roberts, R. G., Hayes, E., Lowe, K., Carah, X., Florin, S. A., McNeil, J., Cox, D., Arnold, L. J., Hua, Q., Huntley, J., Brand, H. E. A., Manne, T., Fairbairn, A., Shulmeister, J., Lyle, L., Salinas, M., Page, M., Connell, K., Park, G., Norman, K., Murphy, T. & Pardoe, C. (2017). *Nature* **547**, 306-310.
- Coffey, L. L., Failloux, A. B. & Weaver, S. C. (2014). *Viruses* **6**, 4628-4663.
- Coffey, L. L., Forrester, N., Tsetsarkin, K., Vasilakis, N. & Weaver, S. C. (2013). *Future Microbiol* **8**, 155-176.
- Coombs, K. M. & Brown, D. T. (1989). *J Virol* **63**, 883-891.
- Costanzi, C. & Pehrson, J. R. (1998). *Nature* **393**, 599-601.

- Couderc, T., Khandoudi, N., Grandadam, M., Visse, C., Gangneux, N., Bagot, S., Prost, J. F. & Lecuit, M. (2009). *J Infect Dis* **200**, 516-523.
- Coutard, B., Barral, K., Lichiere, J., Selisko, B., Martin, B., Aouadi, W., Lombardia, M. O., Debart, F., Vasseur, J. J., Guillemot, J. C., Canard, B. & Decroly, E. (2017). *J Virol* **91**.
- Couturier, E., Guillemin, F., Mura, M., Leon, L., Virion, J. M., Letort, M. J., De Valk, H., Simon, F. & Vaillant, V. (2012). *Rheumatology (Oxford)* **51**, 1315-1322.
- Cristea, I. M., Carroll, J. W., Rout, M. P., Rice, C. M., Chait, B. T. & MacDonald, M. R. (2006). *J Biol Chem* **281**, 30269-30278.
- Cristea, I. M., Rozjabek, H., Molloy, K. R., Karki, S., White, L. L., Rice, C. M., Rout, M. P., Chait, B. T. & MacDonald, M. R. (2010). *J Virol* **84**, 6720-6732.
- Cross, R. K. (1983). *Virology* **130**, 452-463.
- Cross, R. K. & Gomas, P. J. (1981). *Virology* **114**, 542-554.
- Cullen, B. R. (2009). *Cell* **136**, 592-597.
- D'Amours, D., Desnoyers, S., D'Silva, I. & Poirier, G. G. (1999). *Biochem J* **342 (Pt 2)**, 249-268.
- Daffis, S., Szretter, K. J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T. Y., Schneller, S., Zust, R., Dong, H., Thiel, V., Sen, G. C., Fensterl, V., Klimstra, W. B., Pierson, T. C., Buller, R. M., Gale, M., Jr., Shi, P. Y. & Diamond, M. S. (2010). *Nature* **468**, 452-456.
- Dahl, M. V. (2000). *J Am Acad Dermatol* **43**, S1-5.
- Daigo, K. & Hamakubo, T. (2012). *Front Immunol* **3**, 378.
- Dalbey, R. E., Lively, M. O., Bron, S. & van Dijk, J. M. (1997). *Protein Sci* **6**, 1129-1138.
- Daniels, C. M., Ong, S. E. & Leung, A. K. (2014). *J Proteome Res* **13**, 3510-3522.
- Das, P. K., Merits, A. & Lulla, A. (2014). *J Biol Chem* **289**, 5635-5653.
- Das, P. K., Puusepp, L., Varghese, F. S., Utt, A., Ahola, T., Kananovich, D. G., Lopp, M., Merits, A. & Karelson, M. (2016). *Antimicrob Agents Chemother* **60**, 7382-7395.
- Das, T., Jaffar-Bandjee, M. C., Hoarau, J. J., Krejbich Trotot, P., Denizot, M., Lee-Pat-Yuen, G., Sahoo, R., Guiraud, P., Ramful, D., Robin, S., Alessandri, J. L., Gauzere, B. A. & Gasque, P. (2010). *Prog Neurobiol* **91**, 121-129.
- Dash, P. K., Parida, M. M., Santhosh, S. R., Verma, S. K., Tripathi, N. K., Ambuj, S., Saxena, P., Gupta, N., Chaudhary, M., Babu, J. P., Lakshmi, V., Mamidi, N., Subhalaxmi, M. V., Lakshmana Rao, P. V. & Sekhar, K. (2007). *Vector Borne Zoonotic Dis* **7**, 519-527.
- Daugherty, M. D., Young, J. M., Kerns, J. A. & Malik, H. S. (2014). *PLoS Genet* **10**, e1004403.

- Davis, K., Banerjee, S., Friggeri, A., Bell, C., Abraham, E. & Zerfaoui, M. (2012). *Mol Med* **18**, 359-369.
- Davis, N. L., Pence, D. F., Meyer, W. J., Schmaljohn, A. L. & Johnston, R. E. (1987). *Virology* **161**, 101-108.
- Davis, N. L., Willis, L. V., Smith, J. F. & Johnston, R. E. (1989). *Virology* **171**, 189-204.
- de Curtis, I. & Simons, K. (1988). *Proc Natl Acad Sci U S A* **85**, 8052-8056.
- de Groot, R. J., Hardy, W. R., Shirako, Y. & Strauss, J. H. (1990). *EMBO J* **9**, 2631-2638.
- de Groot, R. J., Rumenapf, T., Kuhn, R. J., Strauss, E. G. & Strauss, J. H. (1991). *Proc Natl Acad Sci U S A* **88**, 8967-8971.
- De, I., Fata-Hartley, C., Sawicki, S. G. & Sawicki, D. L. (2003). *J Virol* **77**, 13106-13116.
- De, I., Sawicki, S. G. & Sawicki, D. L. (1996). *J Virol* **70**, 2706-2719.
- De Lamballerie, X., Boisson, V., Reynier, J. C., Enault, S., Charrel, R. N., Flahault, A., Roques, P. & Le Grand, R. (2008). *Vector Borne Zoonotic Dis* **8**, 837-839.
- de Lamballerie, X., Leroy, E., Charrel, R. N., Ttsetsarkin, K., Higgs, S. & Gould, E. A. (2008). *Virol J* **5**, 33.
- De Mucha Macias, J. (1963). *Gac Med Mex* **93**, 415-420.
- Deardorff, E. R. & Weaver, S. C. (2010). *Am J Trop Med Hyg* **82**, 1047-1052.
- Decroly, E., Ferron, F., Lescar, J. & Canard, B. (2011). *Nat Rev Microbiol* **10**, 51-65.
- Deeba, F., Islam, A., Kazim, S. N., Naqvi, I. H., Broor, S., Ahmed, A. & Parveen, S. (2016). *Pathog Dis* **74**.
- Delang, L., Li, C., Tas, A., Querat, G., Albuлесcu, I. C., De Burghgraeve, T., Guerrero, N. A., Gigante, A., Piorkowski, G., Decroly, E., Jochmans, D., Canard, B., Snijder, E. J., Perez-Perez, M. J., van Hemert, M. J., Coutard, B., Leyssen, P. & Neyts, J. (2016). *Sci Rep* **6**, 31819.
- Delang, L., Segura Guerrero, N., Tas, A., Querat, G., Pastorino, B., Froeyen, M., Dallmeier, K., Jochmans, D., Herdewijn, P., Bello, F., Snijder, E. J., de Lamballerie, X., Martina, B., Neyts, J., van Hemert, M. J. & Leyssen, P. (2014). *J Antimicrob Chemother* **69**, 2770-2784.
- Delisle, E., Rousseau, C., Broche, B., Leparc-Goffart, I., L'Ambert, G., Cochet, A., Prat, C., Foulongne, V., Ferre, J. B., Catelinois, O., Flusin, O., Tchernonog, E., Moussion, I. E., Wiegandt, A., Septfons, A., Mendy, A., Moyano, M. B., Laporte, L., Maurel, J., Jourdain, F., Reynes, J., Paty, M. C. & Golliot, F. (2015). *Euro Surveill* **20**.
- Denu, J. M. (2005). *Curr Opin Chem Biol* **9**, 431-440.
- Desrosiers, R., Friderici, K. & Rottman, F. (1974). *Proc Natl Acad Sci U S A* **71**, 3971-3975.

- Devarkar, S. C., Wang, C., Miller, M. T., Ramanathan, A., Jiang, F., Khan, A. G., Patel, S. S. & Marcotrigiano, J. (2016). *Proc Natl Acad Sci U S A* **113**, 596-601.
- Di Paola, S., Micaroni, M., Di Tullio, G., Buccione, R. & Di Girolamo, M. (2012). *PLoS One* **7**, e37352.
- Diallo, M., Thonnon, J., Traore-Lamizana, M. & Fontenille, D. (1999). *Am J Trop Med Hyg* **60**, 281-286.
- Diamond, M. S. (2014). *Cytokine Growth Factor Rev* **25**, 543-550.
- Dias, A., Bouvier, D., Crepin, T., McCarthy, A. A., Hart, D. J., Baudin, F., Cusack, S. & Ruigrok, R. W. (2009). *Nature* **458**, 914-918.
- Dieci, G., Preti, M. & Montanini, B. (2009). *Genomics* **94**, 83-88.
- Diefenbach, J. & Burkle, A. (2005). *Cell Mol Life Sci* **62**, 721-730.
- Ding, F., Fu, J., Jiang, D., Hao, M. & Lin, G. (2018). *Acta Trop* **178**, 155-162.
- Dobos, P. & Faulkner, P. (1969). *J Virol* **4**, 429-438.
- Dryga, S. A., Dryga, O. A. & Schlesinger, S. (1997). *Virology* **228**, 74-83.
- Dubin, D. T., Stollar, V., Hsueh, C. C., Timko, K. & Guild, G. M. (1977). *Virology* **77**, 457-470.
- Dunstan, M. S., Barkauskaite, E., Lafite, P., Knezevic, C. E., Brassington, A., Ahel, M., Hergenrother, P. J., Leys, D. & Ahel, I. (2012). *Nat Commun* **3**, 878.
- Dupont-Rouzeyrol, M., Caro, V., Guillaumot, L., Vazeille, M., D'Ortenzio, E., Thiberge, J. M., Baroux, N., Gourinat, A. C., Grandadam, M. & Failloux, A. B. (2012). *Vector Borne Zoonotic Dis* **12**, 1036-1041.
- Durand, B., Lecollinet, S., Beck, C., Martinez-Lopez, B., Balenghien, T. & Chevalier, V. (2013). *PLoS One* **8**, e70000.
- Eckei, L., Krieg, S., Butepage, M., Lehmann, A., Gross, A., Lippok, B., Grimm, A. R., Kummerer, B. M., Rossetti, G., Luscher, B. & Verheugd, P. (2017). *Sci Rep* **7**, 41746.
- Eckels, K. H., Harrison, V. R. & Hetrick, F. M. (1970). *Appl Microbiol* **19**, 321-325.
- Edelman, R., Tacket, C. O., Wasserman, S. S., Bodison, S. A., Perry, J. G. & Mangiafico, J. A. (2000). *Am J Trop Med Hyg* **62**, 681-685.
- Edwards, C. J., Welch, S. R., Chamberlain, J., Hewson, R., Tolley, H., Cane, P. A. & Lloyd, G. (2007). *J Clin Virol* **39**, 271-275.
- Egloff, M. P., Malet, H., Putics, A., Heinonen, M., Dutartre, H., Frangeul, A., Gruez, A., Campanacci, V., Cambillau, C., Ziebuhr, J., Ahola, T. & Canard, B. (2006). *J Virol* **80**, 8493-8502.
- Ehteshami, M., Tao, S., Zandi, K., Hsiao, H. M., Jiang, Y., Hammond, E., Amblard, F., Russell, O. O., Merits, A. & Schinazi, R. F. (2017). *Antimicrob Agents Chemother* **61**.

- Eriksson, K. K., Cervantes-Barragan, L., Ludewig, B. & Thiel, V. (2008). *J Virol* **82**, 12325-12334.
- Esposito, E. & Cuzzocrea, S. (2009). *Front Biosci (Landmark Ed)* **14**, 263-296.
- Estrada-Franco, J. G., Navarro-Lopez, R., Freier, J. E., Cordova, D., Clements, T., Moncayo, A., Kang, W., Gomez-Hernandez, C., Rodriguez-Dominguez, G., Ludwig, G. V. & Weaver, S. C. (2004). *Emerg Infect Dis* **10**, 2113-2121.
- Evans, P. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.
- Fabrega, C., Shen, V., Shuman, S. & Lima, C. D. (2003). *Mol Cell* **11**, 1549-1561.
- Fata, C. L., Sawicki, S. G. & Sawicki, D. L. (2002). *J Virol* **76**, 8641-8649.
- Fayzuln, R. & Frolov, I. (2004). *J Virol* **78**, 4953-4964.
- Fehr, A. R., Athmer, J., Channappanavar, R., Phillips, J. M., Meyerholz, D. K. & Perlman, S. (2015). *J Virol* **89**, 1523-1536.
- Fehr, A. R., Channappanavar, R., Jankevicius, G., Fett, C., Zhao, J., Athmer, J., Meyerholz, D. K., Ahel, I. & Perlman, S. (2016). *MBio* **7**.
- Feibelman, K. M., Fuller, B. P., Li, L., LaBarbera, D. V. & Geiss, B. J. (2018). *Antiviral Res* **154**, 124-131.
- Feijs, K. L., Forst, A. H., Verheugd, P. & Luscher, B. (2013). *Nat Rev Mol Cell Biol* **14**, 443-451.
- Feijs, K. L., Verheugd, P. & Luscher, B. (2013). *FEBS J* **280**, 3519-3529.
- Ferreira-Ramos, A. S., Li, C., Eydoux, C., Contreras, J. M., Morice, C., Querat, G., Gigante, A., Perez Perez, M. J., Jung, M. L., Canard, B., Guillemot, J. C., Decroly, E. & Coutard, B. (2019). *Antiviral Res* **163**, 59-69.
- Ferro, C., Boshell, J., Moncayo, A. C., Gonzalez, M., Ahumada, M. L., Kang, W. & Weaver, S. C. (2003). *Emerg Infect Dis* **9**, 49-54.
- Ferron, F., Decroly, E., Selisko, B. & Canard, B. (2012). *Antiviral Res* **96**, 21-31.
- Ferron, F., Longhi, S., Henrissat, B. & Canard, B. (2002). *Trends Biochem Sci* **27**, 222-224.
- Filipowicz, W. & Pogacic, V. (2002). *Curr Opin Cell Biol* **14**, 319-327.
- Filomeni, G., De Zio, D. & Cecconi, F. (2015). *Cell Death Differ* **22**, 377-388.
- Firth, A. E., Chung, B. Y., Fleeton, M. N. & Atkins, J. F. (2008). *Virology* **5**, 108.
- Firth, A. E., Wills, N. M., Gesteland, R. F. & Atkins, J. F. (2011). *Nucleic Acids Res* **39**, 6679-6691.
- Fischer, M., Staples, J. E., Arboviral Diseases Branch, N. C. f. E. & Zoonotic Infectious Diseases, C. D. C. (2014). *MMWR Morb Mortal Wkly Rep* **63**, 500-501.
- Fischer, W. B. & Kruger, J. (2009). *Int Rev Cell Mol Biol* **275**, 35-63.
- Fliegert, R., Gasser, A. & Guse, A. H. (2007). *Biochem Soc Trans* **35**, 109-114.
- Follett, E. A., Pringle, C. R. & Pennington, T. H. (1975). *J Gen Virol* **26**, 183-196.

- Forrester, N. L., Palacios, G., Tesh, R. B., Savji, N., Guzman, H., Sherman, M., Weaver, S. C. & Lipkin, W. I. (2012). *J Virol* **86**, 2729-2738.
- Forrester, N. L., Wertheim, J. O., Dugan, V. G., Auguste, A. J., Lin, D., Adams, A. P., Chen, R., Gorchakov, R., Leal, G., Estrada-Franco, J. G., Pandya, J., Halpin, R. A., Hari, K., Jain, R., Stockwell, T. B., Das, S. R., Wentworth, D. E., Smith, M. D., Kosakovsky Pond, S. L. & Weaver, S. C. (2017). *PLoS Negl Trop Dis* **11**, e0005693.
- Forshey, B. M., Guevara, C., Laguna-Torres, V. A., Cespedes, M., Vargas, J., Gianella, A., Vallejo, E., Madrid, C., Aguayo, N., Gotuzzo, E., Suarez, V., Morales, A. M., Beingolea, L., Reyes, N., Perez, J., Negrete, M., Rocha, C., Morrison, A. C., Russell, K. L., Blair, P. J., Olson, J. G., Kochel, T. J. & Group, N. F. S. W. (2010). *PLoS Negl Trop Dis* **4**, e787.
- Forst, A. H., Karlberg, T., Herzog, N., Thorsell, A. G., Gross, A., Feijs, K. L., Verheugd, P., Kursula, P., Nijmeijer, B., Kremmer, E., Kleine, H., Ladurner, A. G., Schuler, H. & Luscher, B. (2013). *Structure* **21**, 462-475.
- Foy, N. J., Akhrymuk, M., Akhrymuk, I., Atasheva, S., Bopda-Waffo, A., Frolov, I. & Frolova, E. I. (2013). *J Virol* **87**, 1997-2010.
- Franck, P. T. & Johnson, K. M. (1970). *Am J Trop Med Hyg* **19**, 860-865.
- French, S. & Wilson, K. (1978). *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **34**, 517-525.
- Frey, T. K. & Strauss, J. H. (1978). *Virology* **86**, 494-506.
- Friedman, R. M., Levy, H. B. & Carter, W. B. (1966). *Proc Natl Acad Sci U S A* **56**, 440-446.
- Frolov, I., Agapov, E., Hoffman, T. A., Jr., Pragai, B. M., Lippa, M., Schlesinger, S. & Rice, C. M. (1999). *J Virol* **73**, 3854-3865.
- Frolov, I., Frolova, E. & Schlesinger, S. (1997). *J Virol* **71**, 2819-2829.
- Frolov, I., Garmashova, N., Atasheva, S. & Frolova, E. I. (2009). *J Virol* **83**, 9031-9044.
- Frolov, I., Hardy, R. & Rice, C. M. (2001). *RNA* **7**, 1638-1651.
- Frolov, I., Kim, D. Y., Akhrymuk, M., Mobley, J. A. & Frolova, E. I. (2017). *J Virol* **91**.
- Frolov, I. & Schlesinger, S. (1994). *J Virol* **68**, 8111-8117.
- Frolova, E., Frolov, I. & Schlesinger, S. (1997). *J Virol* **71**, 248-258.
- Frolova, E., Gorchakov, R., Garmashova, N., Atasheva, S., Vergara, L. A. & Frolov, I. (2006). *J Virol* **80**, 4122-4134.
- Frolova, E. I., Fayzulin, R. Z., Cook, S. H., Griffin, D. E., Rice, C. M. & Frolov, I. (2002). *J Virol* **76**, 11254-11264.
- Frolova, E. I., Gorchakov, R., Pereboeva, L., Atasheva, S. & Frolov, I. (2010). *J Virol* **84**, 11679-11695.

- Fros, J. J., Domeradzka, N. E., Baggen, J., Geertsema, C., Flipse, J., Vlak, J. M. & Pijlman, G. P. (2012). *J Virol* **86**, 10873-10879.
- Fros, J. J., Liu, W. J., Prow, N. A., Geertsema, C., Ligtenberg, M., Vanlandingham, D. L., Schnettler, E., Vlak, J. M., Suhrbier, A., Khromykh, A. A. & Pijlman, G. P. (2010). *J Virol* **84**, 10877-10887.
- Fros, J. J., Major, L. D., Scholte, F. E. M., Gardner, J., van Hemert, M. J., Suhrbier, A. & Pijlman, G. P. (2015). *J Gen Virol* **96**, 580-589.
- Fros, J. J. & Pijlman, G. P. (2016). *Viruses* **8**.
- Froshauer, S., Kartenbeck, J. & Helenius, A. (1988). *J Cell Biol* **107**, 2075-2086.
- Fukunaga, Y., Kumanomido, T. & Kamada, M. (2000). *Vet Clin North Am Equine Pract* **16**, 605-617.
- Fuller, S. D. (1987). *Cell* **48**, 923-934.
- Furuichi, Y. (1974). *Nucleic Acids Res* **1**, 809-822.
- Furuichi, Y. (2015). *Proc Jpn Acad Ser B Phys Biol Sci* **91**, 394-409.
- Furuichi, Y., LaFiandra, A. & Shatkin, A. J. (1977). *Nature* **266**, 235-239.
- Furuichi, Y., Morgan, M., Muthukrishnan, S. & Shatkin, A. J. (1975). *Proc Natl Acad Sci U S A* **72**, 362-366.
- Furuichi, Y., Morgan, M., Shatkin, A. J., Jelinek, W., Salditt-Georgieff, M. & Darnell, J. E. (1975). *Proc Natl Acad Sci U S A* **72**, 1904-1908.
- Furuichi, Y., Muthukrishnan, S. & Shatkin, A. J. (1975). *Proc Natl Acad Sci U S A* **72**, 742-745.
- Gaedigk-Nitschko, K., Ding, M. X., Levy, M. A. & Schlesinger, M. J. (1990). *Virology* **175**, 282-291.
- Gaedigk-Nitschko, K. & Schlesinger, M. J. (1990). *Virology* **175**, 274-281.
- Gaedigk-Nitschko, K. & Schlesinger, M. J. (1991). *Virology* **183**, 206-214.
- Galbraith, S. E., Sheahan, B. J. & Atkins, G. J. (2006). *J Gen Virol* **87**, 937-947.
- Galindo, P. & Grayson, M. A. (1971). *Science* **172**, 594-595.
- Gao, G., Guo, X. & Goff, S. P. (2002). *Science* **297**, 1703-1706.
- Garmashova, N., Atasheva, S., Kang, W., Weaver, S. C., Frolova, E. & Frolov, I. (2007). *J Virol* **81**, 13552-13565.
- Garmashova, N., Gorchakov, R., Frolova, E. & Frolov, I. (2006). *J Virol* **80**, 5686-5696.
- Garmashova, N., Gorchakov, R., Volkova, E., Paessler, S., Frolova, E. & Frolov, I. (2007). *J Virol* **81**, 2472-2484.
- Garneau, N. L., Sokoloski, K. J., Opyrchal, M., Neff, C. P., Wilusz, C. J. & Wilusz, J. (2008). *J Virol* **82**, 880-892.
- Garoff, H., Huylebroeck, D., Robinson, A., Tillman, U. & Liljestrom, P. (1990). *J Cell Biol* **111**, 867-876.

- Garoff, H., Simons, K. & Dobberstein, B. (1978). *J Mol Biol* **124**, 587-600.
- Geerlof, A., Brown, J., Coutard, B., Egloff, M. P., Enguita, F. J., Fogg, M. J., Gilbert, R. J., Groves, M. R., Haouz, A., Nettleship, J. E., Nordlund, P., Owens, R. J., Ruff, M., Sainsbury, S., Svergun, D. I. & Wilmanns, M. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1125-1136.
- Gerardin, P., Barau, G., Michault, A., Bintner, M., Randrianaivo, H., Choker, G., Lenglet, Y., Touret, Y., Bouveret, A., Grivard, P., Le Roux, K., Blanc, S., Schuffenecker, I., Couderc, T., Arenzana-Seisdedos, F., Lecuit, M. & Robillard, P. Y. (2008). *PLoS Med* **5**, e60.
- Gerardin, P., Samperiz, S., Ramful, D., Boumahni, B., Bintner, M., Alessandri, J. L., Carbonnier, M., Tiran-Rajaoefera, I., Beullier, G., Boya, I., Noormahomed, T., Okoi, J., Rollot, O., Cotte, L., Jaffar-Bandjee, M. C., Michault, A., Favier, F., Kaminski, M., Fourmaintraux, A. & Fritel, X. (2014). *PLoS Negl Trop Dis* **8**, e2996.
- Ghosh, A. & Lima, C. D. (2010). *Wiley Interdiscip Rev RNA* **1**, 152-172.
- Gibson, B. A. & Kraus, W. L. (2012). *Nat Rev Mol Cell Biol* **13**, 411-424.
- Gigante, A., Canela, M. D., Delang, L., Priego, E. M., Camarasa, M. J., Querat, G., Neyts, J., Leyssen, P. & Perez-Perez, M. J. (2014). *J Med Chem* **57**, 4000-4008.
- Gigante, A., Gomez-SanJuan, A., Delang, L., Li, C., Bueno, O., Gamo, A. M., Priego, E. M., Camarasa, M. J., Jochmans, D., Leyssen, P., Decroly, E., Coutard, B., Querat, G., Neyts, J. & Perez-Perez, M. J. (2017). *Antiviral Res* **144**, 216-222.
- Gingras, A. C., Raught, B. & Sonenberg, N. (1999). *Annu Rev Biochem* **68**, 913-963.
- Gitlin, L., Barchet, W., Gilfillan, S., Cella, M., Beutler, B., Flavell, R. A., Diamond, M. S. & Colonna, M. (2006). *Proc Natl Acad Sci U S A* **103**, 8459-8464.
- Glasker, S., Toller, M. & Kummerer, B. M. (2014). *J Gen Virol* **95**, 816-822.
- Gloria-Soria, A., Brown, J. E., Kramer, V., Hardstone Yoshimizu, M. & Powell, J. R. (2014). *PLoS Negl Trop Dis* **8**, e3029.
- Goff, S. P. (2007). *Nat Rev Microbiol* **5**, 253-263.
- Golubtsov, A., Kaariainen, L. & Caldentey, J. (2006). *FEBS Lett* **580**, 1502-1508.
- Gomatos, P. J., Kaariainen, L., Keranen, S., Ranki, M. & Sawicki, D. L. (1980). *J Gen Virol* **49**, 61-69.
- Gomez de Cedron, M., Ehsani, N., Mikkola, M. L., Garcia, J. A. & Kaariainen, L. (1999). *FEBS Lett* **448**, 19-22.
- Gonda, D. K., Bachmair, A., Wunning, I., Tobias, J. W., Lane, W. S. & Varshavsky, A. (1989). *J Biol Chem* **264**, 16700-16712.
- Gonzalez-Salazar, D., Estrada-Franco, J. G., Carrara, A. S., Aronson, J. F. & Weaver, S. C. (2003). *Emerg Infect Dis* **9**, 161-168.
- Gonzalez, M. E. & Carrasco, L. (2001). *Virology* **279**, 201-209.

- Gonzalez, M. E. & Carrasco, L. (2003). *FEBS Lett* **552**, 28-34.
- Goodfellow, I., Chaudhry, Y., Gioldasi, I., Gerondopoulos, A., Natoni, A., Labrie, L., Laliberte, J. F. & Roberts, L. (2005). *EMBO Rep* **6**, 968-972.
- Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. (2006). *Virus Res* **117**, 17-37.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1988). *FEBS Lett* **235**, 16-24.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989). *Nucleic Acids Res* **17**, 4713-4730.
- Gorchakov, R., Frolova, E. & Frolov, I. (2005). *J Virol* **79**, 9397-9409.
- Gorchakov, R., Frolova, E., Sawicki, S., Atasheva, S., Sawicki, D. & Frolov, I. (2008). *J Virol* **82**, 6218-6231.
- Gorchakov, R., Garmashova, N., Frolova, E. & Frolov, I. (2008). *J Virol* **82**, 10088-10101.
- Gorchakov, R., Hardy, R., Rice, C. M. & Frolov, I. (2004). *J Virol* **78**, 61-75.
- Gotte, B., Liu, L. & McInerney, G. M. (2018). *Viruses* **10**.
- Gottschalk, A. J., Timinszky, G., Kong, S. E., Jin, J., Cai, Y., Swanson, S. K., Washburn, M. P., Florens, L., Ladurner, A. G., Conaway, J. W. & Conaway, R. C. (2009). *Proc Natl Acad Sci U S A* **106**, 13770-13774.
- Gottschalk, A. J., Trivedi, R. D., Conaway, J. W. & Conaway, R. C. (2012). *J Biol Chem* **287**, 43527-43532.
- Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A. J., Zillinger, T., Goldeck, M., Schuberth, C., Van der Veen, A. G., Fujimura, T., Rehwinkel, J., Iskarpatyoti, J. A., Barchet, W., Ludwig, J., Dermody, T. S., Hartmann, G. & Reis e Sousa, C. (2014). *Nature* **514**, 372-375.
- Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res* **31**, 3320-3323.
- Gould, E. A., Coutard, B., Malet, H., Morin, B., Jamal, S., Weaver, S., Gorbalenya, A., Moureau, G., Baronti, C., Delogu, I., Forrester, N., Khasnatinov, M., Gritsun, T., de Lamballerie, X. & Canard, B. (2010a). *Antiviral Res* **87**, 111-124.
- Gould, E. A., Coutard, B., Malet, H., Morin, B., Jamal, S., Weaver, S., Gorbalenya, A., Moureau, G., Baronti, C., Delogu, I., Forrester, N., Khasnatinov, M., Gritsun, T., de Lamballerie, X. & Canard, B. (2010b). *Antivir Res* **87**, 111-124.
- Grakoui, A., Levis, R., Raju, R., Huang, H. V. & Rice, C. M. (1989). *J Virol* **63**, 5216-5227.
- Grandadam, M., Caro, V., Plumet, S., Thiberge, J. M., Souares, Y., Failloux, A. B., Tolou, H. J., Budelot, M., Cosserat, D., Leparac-Goffart, I. & Despres, P. (2011). *Emerg Infect Dis* **17**, 910-913.

- Grandvaux, N., Servant, M. J., tenOever, B., Sen, G. C., Balachandran, S., Barber, G. N., Lin, R. & Hiscott, J. (2002). *J Virol* **76**, 5532-5539.
- Grayson, M. A. & Galindo, P. (1968). *Am J Epidemiol* **88**, 80-96.
- Greaves, J. & Chamberlain, L. H. (2007). *J Cell Biol* **176**, 249-254.
- Greene, I. P., Paessler, S., Austgen, L., Anishchenko, M., Brault, A. C., Bowen, R. A. & Weaver, S. C. (2005). *J Virol* **79**, 9128-9133.
- Griffin, D. E. (2005). *Curr Top Microbiol Immunol* **289**, 57-77.
- Grimley, P. M., Berezsky, I. K. & Friedman, R. M. (1968). *J Virol* **2**, 1326-1338.
- Gu, W., Gallagher, G. R., Dai, W., Liu, P., Li, R., Trombly, M. I., Gammon, D. B., Mello, C. C., Wang, J. P. & Finberg, R. W. (2015). *RNA* **21**, 2067-2075.
- Gudo, E. S., Black, J. F. & Cliff, J. L. (2016). *PLoS Negl Trop Dis* **10**, e0005001.
- Gunn, B. M., Morrison, T. E., Whitmore, A. C., Blevins, L. K., Hueston, L., Fraser, R. J., Herrero, L. J., Ramirez, R., Smith, P. N., Mahalingam, S. & Heise, M. T. (2012). *PLoS Pathog* **8**, e1002586.
- Guo, X., Ma, J., Sun, J. & Gao, G. (2007). *Proc Natl Acad Sci U S A* **104**, 151-156.
- Guo, Y. C., Zhou, Y. F., Zhang, X. E., Zhang, Z. P., Qiao, Y. M., Bi, L. J., Wen, J. K., Liang, M. F. & Zhang, J. B. (2006). *Nucleic Acids Res* **34**, e62.
- Gutierrez, E., Monath, T. P., Alava, A., Uriguen, D., Arzube, M. & Chamberlain, R. W. (1975). *Am J Epidemiol* **102**, 400-413.
- Gyawali, N., Bradbury, R. S., Aaskov, J. G. & Taylor-Robinson, A. W. (2017). *Microbes Infect* **19**, 388-401.
- Ha, H. C. & Snyder, S. H. (1999). *Proc Natl Acad Sci U S A* **96**, 13978-13982.
- Habjan, M., Hubel, P., Lacerda, L., Benda, C., Holze, C., Eberl, C. H., Mann, A., Kindler, E., Gil-Cruz, C., Ziebuhr, J., Thiel, V. & Pichlmair, A. (2013). *PLoS Pathog* **9**, e1003663.
- Hahn, C. S., Lustig, S., Strauss, E. G. & Strauss, J. H. (1988). *Proc Natl Acad Sci U S A* **85**, 5997-6001.
- Hahn, C. S., Strauss, E. G. & Strauss, J. H. (1985). *Proc Natl Acad Sci U S A* **82**, 4648-4652.
- Hahn, C. S. & Strauss, J. H. (1990). *J Virol* **64**, 3069-3073.
- Hahn, Y. S., Grakoui, A., Rice, C. M., Strauss, E. G. & Strauss, J. H. (1989). *J Virol* **63**, 1194-1202.
- Hahn, Y. S., Strauss, E. G. & Strauss, J. H. (1989). *J Virol* **63**, 3142-3150.
- Hakansson, K., Doherty, A. J., Shuman, S. & Wigley, D. B. (1997). *Cell* **89**, 545-553.
- Halappanavar, S. S., Rhun, Y. L., Mounir, S., Martins, L. M., Huot, J., Earnshaw, W. C. & Shah, G. M. (1999). *J Biol Chem* **274**, 37097-37104.
- Hamm, J. & Mattaj, I. W. (1990). *Cell* **63**, 109-118.

- Hammon, W. M., Rudnick, A. & Sather, G. E. (1960). *Science* **131**, 1102-1103.
- Han, A. P., Yu, C., Lu, L., Fujiwara, Y., Browne, C., Chin, G., Fleming, M., Leboulch, P., Orkin, S. H. & Chen, J. J. (2001). *EMBO J* **20**, 6909-6918.
- Hansen, J. D., Vojtech, L. N. & Laing, K. J. (2011). *Dev Comp Immunol* **35**, 886-897.
- Hapuarachchi, H. C., Bandara, K. B., Sumanadasa, S. D., Hapugoda, M. D., Lai, Y. L., Lee, K. S., Tan, L. K., Lin, R. T., Ng, L. F., Bucht, G., Abeyewickreme, W. & Ng, L. C. (2010). *J Gen Virol* **91**, 1067-1076.
- Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M. & Ron, D. (2000). *Mol Cell* **6**, 1099-1108.
- Hardy, R. W. (2006). *Virology* **345**, 520-531.
- Hardy, R. W. & Rice, C. M. (2005). *J Virol* **79**, 4630-4639.
- Hardy, W. R. & Strauss, J. H. (1989). *J Virol* **63**, 4653-4664.
- Harley, D., Sleigh, A. & Ritchie, S. (2001). *Clin Microbiol Rev* **14**, 909-932, table of contents.
- Harrison, V. R., Eckels, K. H., Bartelloni, P. J. & Hampton, C. (1971). *J Immunol* **107**, 643-647.
- Hashimoto, K., Erdei, S., Keranen, S., Saraste, J. & Kaariainen, L. (1981). *J Virol* **38**, 34-40.
- Hauschildt, S., Scheipers, P. & Bessler, W. G. (1991). *Biochem Biophys Res Commun* **179**, 865-871.
- Hauschildt, S., Scheipers, P., Bessler, W. G. & Mulsch, A. (1992). *Biochem J* **288 (Pt 1)**, 255-260.
- Hefti, E., Bishop, D. H., Dubin, D. T. & Stollar, V. (1975). *J Virol* **17**, 149-159.
- Heidner, H. W., McKnight, K. L., Davis, N. L. & Johnston, R. E. (1994). *J Virol* **68**, 2683-2692.
- Heine, H., Ulmer, A. J., Flad, H. D. & Hauschildt, S. (1995). *J Immunol* **155**, 4899-4908.
- Hellen, C. U. & Sarnow, P. (2001). *Genes Dev* **15**, 1593-1612.
- Hemmi, H., Kaisho, T., Takeuchi, O., Sato, S., Sanjo, H., Hoshino, K., Horiuchi, T., Tomizawa, H., Takeda, K. & Akira, S. (2002). *Nat Immunol* **3**, 196-200.
- Hermanns, K., Zirkel, F., Kopp, A., Marklewitz, M., Rwegu, I. B., Estrada, A., Gillespie, T. R., Drosten, C. & Junglen, S. (2017). *J Gen Virol* **98**, 43-49.
- Hermon, Y. E. (1967). *Ceylon Med J* **12**, 81-92.
- Herrero, L. J., Nelson, M., Srikiatkachorn, A., Gu, R., Anantapreecha, S., Fingerle-Rowson, G., Bucala, R., Morand, E., Santos, L. L. & Mahalingam, S. (2011). *Proc Natl Acad Sci U S A* **108**, 12048-12053.
- Hetz, C. (2012). *Nat Rev Mol Cell Biol* **13**, 89-102.
- Hill, K. R., Hajjou, M., Hu, J. Y. & Raju, R. (1997). *J Virol* **71**, 2693-2704.

- Hinman, A. R., McGowan, J. E., Jr. & Henderson, B. E. (1971). *Am J Epidemiol* **93**, 130-136.
- Hirsch, B. M., Burgos, E. S. & Schramm, V. L. (2014). *ACS Chem Biol* **9**, 2255-2262.
- Ho, C. K. & Shuman, S. (1999). *Mol Cell* **3**, 405-411.
- Holm, L. & Sander, C. (1993). *J Mol Biol* **233**, 123-138.
- Hong, S. J., Dawson, T. M. & Dawson, V. L. (2004). *Trends Pharmacol Sci* **25**, 259-264.
- Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K. K., Schlee, M., Endres, S. & Hartmann, G. (2006). *Science* **314**, 994-997.
- Horwood, P. F. & Buchy, P. (2015). *Rev Sci Tech* **34**, 479-489.
- Horwood, P. F., Reimer, L. J., Dagina, R., Susapu, M., Bande, G., Katusele, M., Koimbu, G., Jimmy, S., Ropa, B., Siba, P. M. & Pavlin, B. I. (2013). *Emerg Infect Dis* **19**, 1535-1538.
- Hottiger, M. O. (2015). *Annu Rev Biochem* **84**, 227-263.
- Hottiger, M. O., Hassa, P. O., Luscher, B., Schuler, H. & Koch-Nolte, F. (2010). *Trends Biochem Sci* **35**, 208-219.
- Hsieh, P. & Robbins, P. W. (1984). *J Biol Chem* **259**, 2375-2382.
- HsuChen, C. C. & Dubin, D. T. (1976). *Nature* **264**, 190-191.
- Htain, W. W., Leong, S. K. & Ling, E. A. (1997). *Neurosci Lett* **223**, 53-56.
- Hu, C., Liu, Y., Lin, Y., Liang, J. K., Zhong, W. W., Li, K., Huang, W. T., Wang, D. J., Yan, G. M., Zhu, W. B., Qiu, J. G. & Gao, X. (2018). *Cell Death Dis* **9**, 274.
- Hu, R. H., Lin, M. C., Hsu, Y. H. & Meng, M. (2011). *Virology* **411**, 15-24.
- Hu, X., Compton, J. R., Leary, D. H., Olson, M. A., Lee, M. S., Cheung, J., Ye, W., Ferrer, M., Southall, N., Jadhav, A., Morazzani, E. M., Glass, P. J., Marugan, J. & Legler, P. M. (2016). *Biochemistry* **55**, 3007-3019.
- Huang, X., Zeng, Y., Wang, X., Ma, X., Li, Q., Li, N., Su, H. & Huang, W. (2016). *Biochem Biophys Res Commun* **474**, 351-356.
- Huang, Y. L., Han, Y. T., Chang, Y. T., Hsu, Y. H. & Meng, M. (2004). *J Virol* **78**, 1271-1280.
- Huet, S., Avilov, S. V., Ferbitz, L., Daigle, N., Cusack, S. & Ellenberg, J. (2010). *J Virol* **84**, 1254-1264.
- Hyde, J. L., Chen, R., Trobaugh, D. W., Diamond, M. S., Weaver, S. C., Klimstra, W. B. & Wilusz, J. (2015). *Virus Res* **206**, 99-107.
- Hyde, J. L. & Diamond, M. S. (2015). *Virology* **479-480**, 66-74.
- Hyde, J. L., Gardner, C. L., Kimura, T., White, J. P., Liu, G., Trobaugh, D. W., Huang, C., Tonelli, M., Paessler, S., Takeda, K., Klimstra, W. B., Amarasinghe, G. K. & Diamond, M. S. (2014). *Science* **343**, 783-787.

- Iranpour, M., Moghadam, A. R., Yazdi, M., Ande, S. R., Alizadeh, J., Wiechec, E., Lindsay, R., Drebot, M., Coombs, K. M. & Ghavami, S. (2016). *Expert Rev Mol Med* **18**, e1.
- Irvine, K., Stirling, R., Hume, D. & Kennedy, D. (2004). *Int J Dev Biol* **48**, 1065-1077.
- Isabelle, M., Gagne, J. P., Gallouzi, I. E. & Poirier, G. G. (2012). *J Cell Sci* **125**, 4555-4566.
- Ivanova, L. & Schlesinger, M. J. (1993). *J Virol* **67**, 2546-2551.
- Jain, S. K., DeCandido, S. & Kielian, M. (1991). *J Biol Chem* **266**, 5756-5761.
- Janeway, C. A., Jr. & Medzhitov, R. (2002). *Annu Rev Immunol* **20**, 197-216.
- Jankevicius, G., Hassler, M., Golia, B., Rybin, V., Zacharias, M., Timinszky, G. & Ladurner, A. G. (2013). *Nat Struct Mol Biol* **20**, 508-514.
- Jeong, M. S., Kim, E. J. & Jang, S. B. (2010). *Biochem Biophys Res Commun* **396**, 696-702.
- Jia, F., Miao, H., Zhu, X. & Xu, F. (2017). *J Neurovirol* **23**, 205-215.
- Jiang, F., Ramanathan, A., Miller, M. T., Tang, G. Q., Gale, M., Jr., Patel, S. S. & Marcotrigiano, J. (2011). *Nature* **479**, 423-427.
- Jiao, X., Chang, J. H., Kilic, T., Tong, L. & Kiledjian, M. (2013). *Mol Cell* **50**, 104-115.
- Jiao, X., Doamekpor, S. K., Bird, J. G., Nickels, B. E., Tong, L., Hart, R. P. & Kiledjian, M. (2017). *Cell* **168**, 1015-1027 e1010.
- Jiao, X., Xiang, S., Oh, C., Martin, C. E., Tong, L. & Kiledjian, M. (2010). *Nature* **467**, 608-611.
- Johnson, B., VanBlargan, L. A., Xu, W., White, J. P., Shan, C., Shi, P. Y., Zhang, R., Adhikari, J., Gross, M. L., Leung, D. W., Diamond, M. S. & Amarasinghe, G. K. (2018). *Immunity* **48**, 487-499 e485.
- Johnson, K. M. & Martin, D. H. (1974). *Adv Vet Sci Comp Med* **18**, 79-116.
- Johnson, K. M., Shelokov, A., Peralta, P. H., Dammin, G. J. & Young, N. A. (1968). *Am J Trop Med Hyg* **17**, 432-440.
- Johnson, M. A., Chatterjee, A., Neuman, B. W. & Wuthrich, K. (2010). *J Mol Biol* **400**, 724-742.
- Jones, J. E., Long, K. M., Whitmore, A. C., Sanders, W., Thurlow, L. R., Brown, J. A., Morrison, C. R., Vincent, H., Peck, K. M., Browning, C., Moorman, N., Lim, J. K. & Heise, M. T. (2017). *MBio* **8**.
- Jones, P. H., Maric, M., Madison, M. N., Maury, W., Roller, R. J. & Okeoma, C. M. (2013). *Virology* **438**, 37-49.
- Jose, J., Przybyla, L., Edwards, T. J., Perera, R., Burgner, J. W., 2nd & Kuhn, R. J. (2012). *J Virol* **86**, 2585-2599.
- Jose, J., Snyder, J. E. & Kuhn, R. J. (2009). *Future Microbiol* **4**, 837-856.

- Jose, J., Taylor, A. B. & Kuhn, R. J. (2017). *MBio* **8**.
- Juarez-Salinas, H., Levi, V., Jacobson, E. L. & Jacobson, M. K. (1982). *J Biol Chem* **257**, 607-609.
- Julander, J. G., Smee, D. F., Morrey, J. D. & Furuta, Y. (2009). *Antiviral Res* **82**, 169-171.
- Julius, C. & Yuzenkova, Y. (2017). *Nucleic Acids Res* **45**, 8282-8290.
- Julius, C. & Yuzenkova, Y. (2018). *Wiley Interdiscip Rev RNA*, e1512.
- Jupp, P. G. & Kemp, A. (1996). *S Afr Med J* **86**, 35-37.
- Jwa, M. & Chang, P. (2012). *Nat Cell Biol* **14**, 1223-1230.
- Kallio, K., Hellstrom, K., Jokitalo, E. & Ahola, T. (2016). *J Virol* **90**, 1687-1692.
- Kameoka, M., Ota, K., Tetsuka, T., Tanaka, Y., Itaya, A., Okamoto, T. & Yoshihara, K. (2000). *Biochem J* **346 Pt 3**, 641-649.
- Kanai, M., Miwa, M., Kuchino, Y. & Sugimura, T. (1982). *J Biol Chem* **257**, 6217-6223.
- Kanai, Y., Dohmae, N. & Hirokawa, N. (2004). *Neuron* **43**, 513-525.
- Kannan, P., Yu, Y., Wankhade, S. & Tainsky, M. A. (1999). *Nucleic Acids Res* **27**, 866-874.
- Karo-Astover, L., Sarova, O., Merits, A. & Zusinaite, E. (2010). *Virus Res* **153**, 277-287.
- Karpe, Y. A., Aher, P. P. & Lole, K. S. (2011). *PLoS One* **6**, e22336.
- Karras, G. I., Kustatscher, G., Buhecha, H. R., Allen, M. D., Pugieux, C., Sait, F., Bycroft, M. & Ladurner, A. G. (2005). *EMBO J* **24**, 1911-1920.
- Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K. J., Yamaguchi, O., Otsu, K., Tsujimura, T., Koh, C. S., Reis e Sousa, C., Matsuura, Y., Fujita, T. & Akira, S. (2006). *Nature* **441**, 101-105.
- Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R. & McKnight, S. L. (2012). *Cell* **149**, 753-767.
- Kaur, P., Thiruchelvan, M., Lee, R. C., Chen, H., Chen, K. C., Ng, M. L. & Chu, J. J. (2013). *Antimicrob Agents Chemother* **57**, 155-167.
- Kaur, R., Mudgal, R., Narwal, M. & Tomar, S. (2018). *Virus Res* **256**, 209-218.
- Kawai, T., Takahashi, K., Sato, S., Coban, C., Kumar, H., Kato, H., Ishii, K. J., Takeuchi, O. & Akira, S. (2005). *Nat Immunol* **6**, 981-988.
- Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C. O., Kai, C., Kawai, J., Yasuda, J., Carninci, P. & Hayashizaki, Y. (2008). *BMC Genomics* **9**, 157.

- Kawashima, K. D., Suarez, L. A., Labayo, H. K., Liles, V. R., Salvoza, N. C., Klinzing, D. C., Daroy, M. L., Matias, R. R. & Natividad, F. F. (2014). *Genome Announc* **2**.
- Kedersha, N., Ivanov, P. & Anderson, P. (2013). *Trends Biochem Sci* **38**, 494-506.
- Kedersha, N., Panas, M. D., Achorn, C. A., Lyons, S., Tisdale, S., Hickman, T., Thomas, M., Lieberman, J., McInerney, G. M., Ivanov, P. & Anderson, P. (2016). *J Cell Biol* **212**, 845-860.
- Kedersha, N. L., Gupta, M., Li, W., Miller, I. & Anderson, P. (1999). *J Cell Biol* **147**, 1431-1442.
- Kendra, J. A., de la Fuente, C., Brahms, A., Woodson, C., Bell, T. M., Chen, B., Khan, Y. A., Jacobs, J. L., Kehn-Hall, K. & Dinman, J. D. (2017). *J Virol* **91**.
- Kennelly, P. J. & Krebs, E. G. (1991). *J Biol Chem* **266**, 15555-15558.
- Khan, A. H., Morita, K., Parquet Md Mdel, C., Hasebe, F., Mathenge, E. G. & Igarashi, A. (2002). *J Gen Virol* **83**, 3075-3084.
- Kielian, M. (2010). *Nature* **468**, 645-646.
- Kiledjian, M. (2018). *Trends Cell Biol* **28**, 454-464.
- Kim, D. Y., Atasheva, S., Frolova, E. I. & Frolov, I. (2013). *J Virol* **87**, 4202-4213.
- Kim, D. Y., Firth, A. E., Atasheva, S., Frolova, E. I. & Frolov, I. (2011). *J Virol* **85**, 8022-8036.
- Kim, D. Y., Reynaud, J. M., Rasaloukaya, A., Akhrymuk, I., Mobley, J. A., Frolov, I. & Frolova, E. I. (2016). *PLoS Pathog* **12**, e1005810.
- Kim, I. K., Kiefer, J. R., Ho, C. M., Stegeman, R. A., Classen, S., Tainer, J. A. & Ellenberger, T. (2012). *Nat Struct Mol Biol* **19**, 653-656.
- Kim, K. H., Rumenapf, T., Strauss, E. G. & Strauss, J. H. (2004). *Virology* **323**, 153-163.
- Kimball, S. R. (2001). *Prog Mol Subcell Biol* **26**, 155-184.
- Kimball, S. R., Horetsky, R. L., Ron, D., Jefferson, L. S. & Harding, H. P. (2003). *Am J Physiol Cell Physiol* **284**, C273-284.
- Kinney, R. M., Chang, G. J., Tsuchiya, K. R., Sneider, J. M., Roehrig, J. T., Woodward, T. M. & Trent, D. W. (1993). *J Virol* **67**, 1269-1277.
- Kinney, R. M., Johnson, B. J., Welch, J. B., Tsuchiya, K. R. & Trent, D. W. (1989). *Virology* **170**, 19-30.
- Kinney, R. M., Tsuchiya, K. R., Sneider, J. M. & Trent, D. W. (1992a). *J Gen Virol* **73** (Pt 12), 3301-3305.
- Kinney, R. M., Tsuchiya, K. R., Sneider, J. M. & Trent, D. W. (1992b). *Virology* **191**, 569-580.
- Kirsch, M. I., Hulseweh, B., Nacke, C., Rulker, T., Schirrmann, T., Marschall, H. J., Hust, M. & Dubel, S. (2008). *BMC Biotechnol* **8**, 66.

- Kleine, H., Herrmann, A., Lamark, T., Forst, A. H., Verheugd, P., Luscher-Firzlaff, J., Lippok, B., Feijs, K. L., Herzog, N., Kremmer, E., Johansen, T., Muller-Newen, G. & Luscher, B. (2012). *Cell Commun Signal* **10**, 28.
- Kleine, H., Poreba, E., Lesniewicz, K., Hassa, P. O., Hottiger, M. O., Litchfield, D. W., Shilton, B. H. & Luscher, B. (2008). *Mol Cell* **32**, 57-69.
- Klimstra, W. B., Ryman, K. D., Bernard, K. A., Nguyen, K. B., Biron, C. A. & Johnston, R. E. (1999). *J Virol* **73**, 10387-10398.
- Knight, R. L., Schultz, K. L., Kent, R. J., Venkatesan, M. & Griffin, D. E. (2009). *J Virol* **83**, 5640-5647.
- Kobiler, D., Rice, C. M., Brodie, C., Shahar, A., Dubuisson, J., Halevy, M. & Lustig, S. (1999). *J Virol* **73**, 10440-10446.
- Kolakofsky, D., Kowalinski, E. & Cusack, S. (2012). *RNA* **18**, 2118-2127.
- Kolokoltsov, A. A., Wang, E., Colpitts, T. M., Weaver, S. C. & Davey, R. A. (2006). *Am J Trop Med Hyg* **75**, 702-709.
- Koonin, E. V. & Dolja, V. V. (1993). *Crit Rev Biochem Mol Biol* **28**, 375-430.
- Kostyuchenko, V. A., Jakana, J., Liu, X., Haddow, A. D., Aung, M., Weaver, S. C., Chiu, W. & Lok, S. M. (2011). *J Virol* **85**, 9327-9333.
- Kowieski, T. M., Lee, S. & Denu, J. M. (2008). *J Biol Chem* **283**, 5317-5326.
- Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. (2009). *Proc Natl Acad Sci U S A* **106**, 7768-7773.
- Koyama, S., Ishii, K. J., Coban, C. & Akira, S. (2008). *Cytokine* **43**, 336-341.
- Kozlov, M. M., McMahon, H. T. & Chernomordik, L. V. (2010). *Trends Biochem Sci* **35**, 699-706.
- Kraemer, M. U., Sinka, M. E., Duda, K. A., Mylne, A. Q., Shearer, F. M., Barker, C. M., Moore, C. G., Carvalho, R. G., Coelho, G. E., Van Bortel, W., Hendrickx, G., Schaffner, F., Elyazar, I. R., Teng, H. J., Brady, O. J., Messina, J. P., Pigott, D. M., Scott, T. W., Smith, D. L., Wint, G. R., Golding, N. & Hay, S. I. (2015). *Elife* **4**, e08347.
- Kraus, W. L. (2009). *Nat Struct Mol Biol* **16**, 904-907.
- Kubes, V. & Rios, F. A. (1939). *Science* **90**, 20-21.
- Kuhn, R. J., Griffin, D. E., Zhang, H., Niesters, H. G. & Strauss, J. H. (1992). *J Virol* **66**, 7121-7127.
- Kuhn, R. J., Hong, Z. & Strauss, J. H. (1990). *J Virol* **64**, 1465-1476.
- Kujala, P., Ikaheimonen, A., Ehsani, N., Vihinen, H., Auvinen, P. & Kaariainen, L. (2001). *J Virol* **75**, 3873-3884.
- Kulasegaran-Shyliani, R., Atasheva, S., Gorenstein, D. G. & Frolov, I. (2009). *J Virol* **83**, 8327-8339.

- Kulasegaran-Shylini, R., Thiviyanathan, V., Gorenstein, D. G. & Frolov, I. (2009). *Virology* **387**, 211-221.
- Kumar, N. P., Joseph, R., Kamaraj, T. & Jambulingam, P. (2008). *J Gen Virol* **89**, 1945-1948.
- Kumar, P., Sweeney, T. R., Skabkin, M. A., Skabkina, O. V., Hellen, C. U. & Pestova, T. V. (2014). *Nucleic Acids Res* **42**, 3228-3245.
- Kumar, S., Kumar, A., Mamidi, P., Tiwari, A., Kumar, S., Mayavannan, A., Mudulli, S., Singh, A. K., Subudhi, B. B. & Chattopadhyay, S. (2018). *Sci Rep* **8**, 1045.
- Kuri, T., Eriksson, K. K., Putics, A., Züst, R., Snijder, E. J., Davidson, A. D., Siddell, S. G., Thiel, V., Ziebuhr, J. & Weber, F. (2011). *J Gen Virol* **92**, 1899-1905.
- Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A. G. (2005). *Nat Struct Mol Biol* **12**, 624-625.
- Laakkonen, P., Ahola, T. & Kaariainen, L. (1996). *J Biol Chem* **271**, 28567-28571.
- Laakkonen, P., Auvinen, P., Kujala, P. & Kaariainen, L. (1998). *J Virol* **72**, 10265-10269.
- Laakkonen, P., Hyvonen, M., Peranen, J. & Kaariainen, L. (1994). *J Virol* **68**, 7418-7425.
- Lahariya, C. & Pradhan, S. K. (2006). *J Vector Borne Dis* **43**, 151-160.
- Lakshmi, V., Neeraja, M., Subbalaxmi, M. V., Parida, M. M., Dash, P. K., Santhosh, S. R. & Rao, P. V. (2008). *Clin Infect Dis* **46**, 1436-1442.
- Lambrecht, M. J., Brichacek, M., Barkauskaite, E., Ariza, A., Ahel, I. & Hergenrother, P. J. (2015). *J Am Chem Soc* **137**, 3558-3564.
- Lamond, A. L. (1990). *Trends Biochem Sci* **15**, 451-452.
- Lampio, A., Kilpelainen, I., Pesonen, S., Karhi, K., Auvinen, P., Somerharju, P. & Kaariainen, L. (2000). *J Biol Chem* **275**, 37853-37859.
- Lanciotti, R. S., Kosoy, O. L., Laven, J. J., Panella, A. J., Velez, J. O., Lambert, A. J. & Campbell, G. L. (2007). *Emerg Infect Dis* **13**, 764-767.
- Lanciotti, R. S., Ludwig, M. L., Rwaguma, E. B., Lutwama, J. J., Kram, T. M., Karabatsos, N., Cropp, B. C. & Miller, B. R. (1998). *Virology* **252**, 258-268.
- Lanciotti, R. S. & Valadere, A. M. (2014). *Emerg Infect Dis* **20**, 1400-1402.
- Langelier, M. F. & Pascal, J. M. (2013). *Curr Opin Struct Biol* **23**, 134-143.
- Lark, T., Keck, F. & Narayanan, A. (2017). *Front Microbiol* **8**, 2652.
- Lastarza, M. W., Grakoui, A. & Rice, C. M. (1994). *Virology* **202**, 224-232.
- LaStarza, M. W., Lemm, J. A. & Rice, C. M. (1994). *J Virol* **68**, 5781-5791.
- Laudisi, F., Sambucci, M. & Pioli, C. (2011). *Endocr Metab Immune Disord Drug Targets* **11**, 326-333.
- Laurent, P., Le Roux, K., Grivard, P., Bertil, G., Naze, F., Picard, M., Staikowsky, F., Barau, G., Schuffenecker, I. & Michault, A. (2007). *Clin Chem* **53**, 1408-1414.

- Le Page, C., Pellat-Deceunynck, C., Drapier, J. C. & Wietzerbin, J. (1997). *Adv Exp Med Biol* **419**, 203-207.
- Le Page, C., Sanceau, J., Drapier, J. C. & Wietzerbin, J. (1998). *Biochem Biophys Res Commun* **243**, 451-457.
- Ledermann, J. P., Guillaumot, L., Yug, L., Saweyog, S. C., Tided, M., Machieng, P., Pretrick, M., Marfel, M., Griggs, A., Bel, M., Duffy, M. R., Hancock, W. T., Ho-Chen, T. & Powers, A. M. (2014). *PLoS Negl Trop Dis* **8**, e3188.
- Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A. & Lai, M. M. (1991). *Virology* **180**, 567-582.
- Lee, S., Owen, K. E., Choi, H. K., Lee, H., Lu, G., Wengler, G., Brown, D. T., Rossmann, M. G. & Kuhn, R. J. (1996). *Structure* **4**, 531-541.
- Leidecker, O., Bonfiglio, J. J., Colby, T., Zhang, Q., Atanassov, I., Zaja, R., Palazzo, L., Stockum, A., Ahel, I. & Matic, I. (2016). *Nat Chem Biol* **12**, 998-1000.
- Lemm, J. A., Bergqvist, A., Read, C. M. & Rice, C. M. (1998). *J Virol* **72**, 6546-6553.
- Lemm, J. A. & Rice, C. M. (1993a). *J Virol* **67**, 1905-1915.
- Lemm, J. A. & Rice, C. M. (1993b). *J Virol* **67**, 1916-1926.
- Lemm, J. A., Rumenapf, T., Strauss, E. G., Strauss, J. H. & Rice, C. M. (1994). *EMBO J* **13**, 2925-2934.
- Leparc-Goffart, I., Nougairede, A., Cassadou, S., Prat, C. & de Lamballerie, X. (2014). *Lancet* **383**, 514.
- Lequime, S., Paul, R. E. & Lambrechts, L. (2016). *PLoS Pathog* **12**, e1005548.
- Leroy, E. M., Nkoghe, D., Ollomo, B., Nze-Nkogue, C., Becquart, P., Grard, G., Pourrut, X., Charrel, R., Moureau, G., Ndjoyi-Mbiguino, A. & De-Lamballerie, X. (2009). *Emerg Infect Dis* **15**, 591-593.
- Lescar, J., Roussel, A., Wien, M. W., Navaza, J., Fuller, S. D., Wengler, G., Wengler, G. & Rey, F. A. (2001). *Cell* **105**, 137-148.
- Leslie Pedrioli, D. M., Leutert, M., Bilan, V., Nowak, K., Gunasekera, K., Ferrari, E., Imhof, R., Malmstrom, L. & Hottiger, M. O. (2018). *EMBO Rep* **19**.
- Leta, S., Beyene, T. J., De Clercq, E. M., Amenu, K., Kraemer, M. U. G. & Revie, C. W. (2018). *Int J Infect Dis* **67**, 25-35.
- Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J. & Bork, P. (2006). *Nucleic Acids Res* **34**, D257-260.
- Leung, A. K. (2014). *J Cell Biol* **205**, 613-619.
- Leung, A. K., Vyas, S., Rood, J. E., Bhutkar, A., Sharp, P. A. & Chang, P. (2011). *Mol Cell* **42**, 489-499.
- Leung, A. K. L., McPherson, R. L. & Griffin, D. E. (2018). *PLoS Pathog* **14**, e1006864.
- Levinson, R. S., Strauss, J. H. & Strauss, E. G. (1990). *Virology* **175**, 110-123.

- Li, C., Debing, Y., Jankevicius, G., Neyts, J., Ahel, I., Coutard, B. & Canard, B. (2016). *J Virol* **90**, 8478-8486.
- Li, C., Guillen, J., Rabah, N., Blanjoie, A., Debart, F., Vasseur, J. J., Canard, B., Decroly, E. & Coutard, B. (2015). *J Virol* **89**, 8292-8303.
- Li, G. & Rice, C. M. (1993). *J Virol* **67**, 5062-5067.
- Li, G. P., La Starza, M. W., Hardy, W. R., Strauss, J. H. & Rice, C. M. (1990). *Virology* **179**, 416-427.
- Li, G. P. & Rice, C. M. (1989). *J Virol* **63**, 1326-1337.
- Li, J., Fontaine-Rodriguez, E. C. & Whelan, S. P. (2005). *J Virol* **79**, 13373-13384.
- Li, M. L. & Stollar, V. (2004). *Proc Natl Acad Sci U S A* **101**, 9429-9434.
- Li, M. L. & Stollar, V. (2007). *J Virol* **81**, 4371-4373.
- Li, M. L., Wang, H. & Stollar, V. (2010). *J Virol* **84**, 2732-2739.
- Li, X., Lu, C., Stewart, M., Xu, H., Strong, R. K., Igumenova, T. & Li, P. (2009). *Arch Biochem Biophys* **488**, 23-33.
- Li, Y. I., Chen, Y. J., Hsu, Y. H. & Meng, M. (2001). *J Virol* **75**, 782-788.
- Liang, B., Li, Z., Jenni, S., Rahmeh, A. A., Morin, B. M., Grant, T., Grigorieff, N., Harrison, S. C. & Whelan, S. P. J. (2015). *Cell* **162**, 314-327.
- Liljestrom, P. & Garoff, H. (1991). *J Virol* **65**, 147-154.
- Liljestrom, P., Lusa, S., Huylebroeck, D. & Garoff, H. (1991). *J Virol* **65**, 4107-4113.
- Lim, S. P., Sonntag, L. S., Noble, C., Nilar, S. H., Ng, R. H., Zou, G., Monaghan, P., Chung, K. Y., Dong, H., Liu, B., Bodenreider, C., Lee, G., Ding, M., Chan, W. L., Wang, G., Jian, Y. L., Chao, A. T., Lescar, J., Yin, Z., Vedananda, T. R., Keller, T. H. & Shi, P. Y. (2011). *J Biol Chem* **286**, 6233-6240.
- Lin, H. Y., Yu, C. Y., Hsu, Y. H. & Meng, M. (2012). *FEBS Lett* **586**, 2326-2331.
- Lin, Y., Tang, X., Zhu, Y., Shu, T. & Han, X. (2011). *Arch Biochem Biophys* **505**, 123-129.
- Lin, Y., Zhang, H., Liang, J., Li, K., Zhu, W., Fu, L., Wang, F., Zheng, X., Shi, H., Wu, S., Xiao, X., Chen, L., Tang, L., Yan, M., Yang, X., Tan, Y., Qiu, P., Huang, Y., Yin, W., Su, X., Hu, H., Hu, J. & Yan, G. (2014). *Proc Natl Acad Sci U S A* **111**, E4504-4512.
- Lindh, E., Argentini, C., Remoli, M. E., Fortuna, C., Faggioni, G., Benedetti, E., Amendola, A., Marsili, G., Lista, F., Rezza, G. & Venturi, G. (2019). *Open Forum Infect Dis* **6**, ofy321.
- Linger, B. R., Kunovska, L., Kuhn, R. J. & Golden, B. L. (2004). *RNA* **10**, 128-138.
- Litzba, N., Schuffenecker, I., Zeller, H., Drosten, C., Emmerich, P., Charrel, R., Kreher, P. & Niedrig, M. (2008). *J Virol Methods* **149**, 175-179.
- Lobigs, M. & Garoff, H. (1990). *J Virol* **64**, 1233-1240.

- Lobigs, M., Zhao, H. X. & Garoff, H. (1990). *J Virol* **64**, 4346-4355.
- Loewy, A., Smyth, J., von Bonsdorff, C. H., Liljestrom, P. & Schlesinger, M. J. (1995). *J Virol* **69**, 469-475.
- Logue, C. H., Sheahan, B. J. & Atkins, G. J. (2008). *Virus Genes* **36**, 313-321.
- Lord, R. D. (1974). *Bull Pan Am Health Organ* **8**, 100-110.
- Lorono-Pino, M. A., Garcia-Rejon, J. E., Machain-Williams, C., Gomez-Carro, S., Nunez-Ayala, G., Najera-Vazquez Mdel, R., Losoya, A., Aguilar, L., Saavedra-Rodriguez, K., Lozano-Fuentes, S., Beaty, M. K., Black, W. C. t., Keefe, T. J., Eisen, L. & Beaty, B. J. (2013). *Am J Trop Med Hyg* **89**, 385-397.
- Lu, Y. E., Eng, C. H., Shome, S. G. & Kielian, M. (2001). *J Virol* **75**, 8329-8339.
- Lu, Y. E. & Kielian, M. (2000). *J Virol* **74**, 7708-7719.
- Lucas-Hourani, M., Lupan, A., Despres, P., Thoret, S., Pamlard, O., Dubois, J., Guillou, C., Tangy, F., Vidalain, P. O. & Munier-Lehmann, H. (2013). *J Biomol Screen* **18**, 172-179.
- Lulla, A., Lulla, V. & Merits, A. (2012). *J Virol* **86**, 553-565.
- Lulla, A., Lulla, V., Tints, K., Ahola, T. & Merits, A. (2006). *J Virol* **80**, 5413-5422.
- Lulla, V., Sawicki, D. L., Sawicki, S. G., Lulla, A., Merits, A. & Ahola, T. (2008). *J Virol* **82**, 9236-9244.
- Lumsden, W. H. (1955). *Trans R Soc Trop Med Hyg* **49**, 33-57.
- Lusa, S., Garoff, H. & Liljestrom, P. (1991). *Virology* **185**, 843-846.
- Luthra, P., Sun, D., Silverman, R. H. & He, B. (2011). *Proc Natl Acad Sci U S A* **108**, 2118-2123.
- Madan, V., Castello, A. & Carrasco, L. (2008). *Cell Microbiol* **10**, 437-451.
- Madan, V., Sanz, M. A. & Carrasco, L. (2005). *Virology* **332**, 307-315.
- Magden, J., Takeda, N., Li, T., Auvinen, P., Ahola, T., Miyamura, T., Merits, A. & Kaariainen, L. (2001). *J Virol* **75**, 6249-6255.
- Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X. & Canard, B. (2009). *J Virol* **83**, 6534-6545.
- Malet, H., Dalle, K., Bremond, N., Tocque, F., Blangy, S., Campanacci, V., Coutard, B., Grisel, S., Lichiere, J., Lantez, V., Cambillau, C., Canard, B. & Egloff, M. P. (2006). *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**, 405-408.
- Malireddi, R. K., Ippagunta, S., Lamkanfi, M. & Kanneganti, T. D. (2010). *J Immunol* **185**, 3127-3130.
- Malvy, D., Ezzedine, K., Mamani-Matsuda, M., Autran, B., Tolou, H., Receveur, M. C., Pistone, T., Rambert, J., Moynet, D. & Mossalayi, D. (2009). *BMC Infect Dis* **9**, 200.

- Malygin, A. G. & Shemyakin, M. F. (1979). *FEBS Lett* **102**, 51-54.
- Manock, S. R., Jacobsen, K. H., de Bravo, N. B., Russell, K. L., Negrete, M., Olson, J. G., Sanchez, J. L., Blair, P. J., Smalligan, R. D., Quist, B. K., Espin, J. F., Espinoza, W. R., MacCormick, F., Fleming, L. C. & Kochel, T. (2009). *Am J Trop Med Hyg* **81**, 146-151.
- Mao, R., Nie, H., Cai, D., Zhang, J., Liu, H., Yan, R., Cuconati, A., Block, T. M., Guo, J. T. & Guo, H. (2013). *PLoS Pathog* **9**, e1003494.
- Markland, W., McQuaid, T. J., Jain, J. & Kwong, A. D. (2000). *Antimicrob Agents Chemother* **44**, 859-866.
- Marquardt, M. T., Phalen, T. & Kielian, M. (1993). *J Cell Biol* **123**, 57-65.
- Martello, R., Leutert, M., Jungmichel, S., Bilan, V., Larsen, S. C., Young, C., Hottiger, M. O. & Nielsen, M. L. (2016). *Nat Commun* **7**, 12917.
- Martin, J. L. & McMillan, F. M. (2002). *Curr Opin Struct Biol* **12**, 783-793.
- Martinez-Gil, L. & Mingarro, I. (2015). *Viruses* **7**, 3462-3482.
- Martinez, M. G. & Kielian, M. (2016). *PLoS Pathog* **12**, e1006061.
- Martinez, M. G., Snapp, E. L., Perumal, G. S., Macaluso, F. P. & Kielian, M. (2014). *J Virol* **88**, 6922-6933.
- Matic, I., Ahel, I. & Hay, R. T. (2012). *Nat Methods* **9**, 771-772.
- Matsuki, H., Takahashi, M., Higuchi, M., Makokha, G. N., Oie, M. & Fujii, M. (2013). *Genes Cells* **18**, 135-146.
- Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., Linder, B., Pickering, B. F., Vasseur, J. J., Chen, Q., Gross, S. S., Elemento, O., Debart, F., Kiledjian, M. & Jaffrey, S. R. (2017). *Nature* **541**, 371-375.
- Mavalankar, D., Shastri, P. & Raman, P. (2007). *Lancet Infect Dis* **7**, 306-307.
- Mayuri, Geders, T. W., Smith, J. L. & Kuhn, R. J. (2008). *J Virol* **82**, 7284-7297.
- Mazzon, M., Castro, C., Thaa, B., Liu, L., Mutso, M., Liu, X., Mahalingam, S., Griffin, J. L., Marsh, M. & McInerney, G. M. (2018). *PLoS Pathog* **14**, e1006835.
- McClain, D. J., Pittman, P. R., Ramsburg, H. H., Nelson, G. O., Rossi, C. A., Mangiafico, J. A., Schmaljohn, A. L. & Malinoski, F. J. (1998). *J Infect Dis* **177**, 634-641.
- McInerney, G. M., Kedersha, N. L., Kaufman, R. J., Anderson, P. & Liljestrom, P. (2005). *Mol Biol Cell* **16**, 3753-3763.
- McPherson, R. L., Abraham, R., Sreekumar, E., Ong, S. E., Cheng, S. J., Baxter, V. K., Kistemaker, H. A., Filippov, D. V., Griffin, D. E. & Leung, A. K. (2017). *Proc Natl Acad Sci U S A* **114**, 1666-1671.
- Medina, G., Garzaro, D. J., Barrios, M., Auguste, A. J., Weaver, S. C. & Pujol, F. H. (2015). *Am J Trop Med Hyg* **93**, 7-10.
- Medzhitov, R. (2007). *Nature* **449**, 819-826.

- Mehta, R., Gerardin, P., de Brito, C. A. A., Soares, C. N., Ferreira, M. L. B. & Solomon, T. (2018). *Rev Med Virol*, e1978.
- Melancon, P. & Garoff, H. (1986). *EMBO J* **5**, 1551-1560.
- Melancon, P. & Garoff, H. (1987). *J Virol* **61**, 1301-1309.
- Melekis, E., Tsika, A. C., Lichiere, J., Chasapis, C. T., Margiolaki, I., Papageorgiou, N., Coutard, B., Bentrop, D. & Spyroulias, G. A. (2015). *Biomol NMR Assign* **9**, 191-195.
- Melton, J. V., Ewart, G. D., Weir, R. C., Board, P. G., Lee, E. & Gage, P. W. (2002). *J Biol Chem* **277**, 46923-46931.
- Merbs, C. F. (1992). *Yearb Phys Anthropol* **35**, 3-42.
- Merits, A., Kettunen, R., Makinen, K., Lampio, A., Auvinen, P., Kaariainen, L. & Ahola, T. (1999). *FEBS Lett* **455**, 45-48.
- Meshram, C. D., Agback, P., Shiliaev, N., Urakova, N., Mobley, J. A., Agback, T., Frolova, E. I. & Frolov, I. (2018). *J Virol*.
- Messner, S., Altmeyer, M., Zhao, H., Pozivil, A., Roschitzki, B., Gehrig, P., Rutishauser, D., Huang, D., Caflisch, A. & Hottiger, M. O. (2010). *Nucleic Acids Res* **38**, 6350-6362.
- Meyer, T., Surber, C., French, L. E. & Stockfleth, E. (2013). *Expert Opin Investig Drugs* **22**, 149-159.
- Meyer, W. J. & Johnston, R. E. (1993). *J Virol* **67**, 5117-5125.
- Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R. & Tschopp, J. (2005). *Nature* **437**, 1167-1172.
- Meylan, E., Tschopp, J. & Karin, M. (2006). *Nature* **442**, 39-44.
- Mi, S., Durbin, R., Huang, H. V., Rice, C. M. & Stollar, V. (1989). *Virology* **170**, 385-391.
- Mi, S. & Stollar, V. (1991). *Virology* **184**, 423-427.
- Michel, G., Petrakova, O., Atasheva, S. & Frolov, I. (2007). *Virology* **362**, 475-487.
- Miwa, M., Ishihara, M., Takishima, S., Takasuka, N., Maeda, M., Yamaizumi, Z., Sugimura, T., Yokoyama, S. & Miyazawa, T. (1981). *J Biol Chem* **256**, 2916-2921.
- Miwa, M., Saikawa, N., Yamaizumi, Z., Nishimura, S. & Sugimura, T. (1979). *Proc Natl Acad Sci U S A* **76**, 595-599.
- Miwa, M. & Sugimura, T. (1971). *J Biol Chem* **246**, 6362-6364.
- Molloy, S. S., Anderson, E. D., Jean, F. & Thomas, G. (1999). *Trends Cell Biol* **9**, 28-35.
- Mombouli, J. V., Bitsindou, P., Elion, D. O., Grolla, A., Feldmann, H., Niama, F. R., Parra, H. J. & Munster, V. J. (2013). *Emerg Infect Dis* **19**, 1542-1543.

- Monath, T. P., Lazuick, J. S., Cropp, C. B., Rush, W. A., Calisher, C. H., Kinney, R. M., Trent, D. W., Kemp, G. E., Bowen, G. S. & Francly, D. B. (1980). *Am J Trop Med Hyg* **29**, 969-983.
- Montgomery, S. A., Berglund, P., Beard, C. W. & Johnston, R. E. (2006). *J Virol* **80**, 7729-7739.
- Moreira, L. A., Iturbe-Ormaetxe, I., Jeffery, J. A., Lu, G., Pyke, A. T., Hedges, L. M., Rocha, B. C., Hall-Mendelin, S., Day, A., Riegler, M., Hugo, L. E., Johnson, K. N., Kay, B. H., McGraw, E. A., van den Hurk, A. F., Ryan, P. A. & O'Neill, S. L. (2009). *Cell* **139**, 1268-1278.
- Morrison, A. C., Forshey, B. M., Notyce, D., Astete, H., Lopez, V., Rocha, C., Carrion, R., Carey, C., Eza, D., Montgomery, J. M. & Kochel, T. J. (2008). *PLoS Negl Trop Dis* **2**, e349.
- Morrison, A. C., Gray, K., Getis, A., Astete, H., Sihuincha, M., Focks, D., Watts, D., Stancil, J. D., Olson, J. G., Blair, P. & Scott, T. W. (2004). *J Med Entomol* **41**, 1123-1142.
- Morrison, T. E., Fraser, R. J., Smith, P. N., Mahalingam, S. & Heise, M. T. (2007). *J Virol* **81**, 5132-5143.
- Morrison, T. E., Oko, L., Montgomery, S. A., Whitmore, A. C., Lotstein, A. R., Gunn, B. M., Elmore, S. A. & Heise, M. T. (2011). *Am J Pathol* **178**, 32-40.
- Moujaber, O., Mahboubi, H., Kodiha, M., Bouttier, M., Bednarz, K., Bakshi, R., White, J., Larose, L., Colmegna, I. & Stochaj, U. (2017). *Biochim Biophys Acta Mol Cell Res* **1864**, 475-486.
- Mousson, L., Martin, E., Zouache, K., Madec, Y., Mavingui, P. & Failloux, A. B. (2010). *Mol Ecol* **19**, 1953-1964.
- Mukhopadhyay, S., Zhang, W., Gabler, S., Chipman, P. R., Strauss, E. G., Strauss, J. H., Baker, T. S., Kuhn, R. J. & Rossmann, M. G. (2006). *Structure* **14**, 63-73.
- Muller, S., Moller, P., Bick, M. J., Wurr, S., Becker, S., Gunther, S. & Kummerer, B. M. (2007). *J Virol* **81**, 2391-2400.
- Munoz-Gamez, J. A., Rodriguez-Vargas, J. M., Quiles-Perez, R., Aguilar-Quesada, R., Martin-Oliva, D., de Murcia, G., Menissier de Murcia, J., Almendros, A., Ruiz de Almodovar, M. & Oliver, F. J. (2009). *Autophagy* **5**, 61-74.
- Myles, K. M., Kelly, C. L., Ledermann, J. P. & Powers, A. M. (2006). *J Virol* **80**, 4992-4997.
- Nasar, F., Palacios, G., Gorchakov, R. V., Guzman, H., Da Rosa, A. P., Savji, N., Popov, V. L., Sherman, M. B., Lipkin, W. I., Tesh, R. B. & Weaver, S. C. (2012). *Proc Natl Acad Sci U S A* **109**, 14622-14627.

- Nasta, F., Laudisi, F., Sambucci, M., Rosado, M. M. & Pioli, C. (2010). *J Immunol* **184**, 3470-3477.
- Neil, S. J. (2013). *Curr Top Microbiol Immunol* **371**, 67-104.
- Nelson, J., Waggoner, J. J., Sahoo, M. K., Grant, P. M. & Pinsky, B. A. (2014). *J Clin Microbiol* **52**, 3459-3461.
- Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., Liao, L., Klaus, J. P., Yates, J. R., 3rd, Wuthrich, K., Stevens, R. C., Buchmeier, M. J. & Kuhn, P. (2008). *J Virol* **82**, 5279-5294.
- Neuvonen, M. & Ahola, T. (2009). *J Mol Biol* **385**, 212-225.
- Neuvonen, M., Kazlauskas, A., Martikainen, M., Hinkkanen, A., Ahola, T. & Saksela, K. (2011). *PLoS Pathog* **7**, e1002383.
- Ng, L. C., Tan, L. K., Tan, C. H., Tan, S. S., Hapuarachchi, H. C., Pok, K. Y., Lai, Y. L., Lam-Phua, S. G., Bucht, G., Lin, R. T., Leo, Y. S., Tan, B. H., Han, H. K., Ooi, P. L., James, L. & Khoo, S. P. (2009). *Emerg Infect Dis* **15**, 1243-1249.
- Nhan, T. X., Claverie, A., Roche, C., Teissier, A., Colleuil, M., Baudet, J. M., Cao-Lormeau, V. M. & Musso, D. (2014). *Emerg Infect Dis* **20**, 1773-1774.
- Nhan, T. X. & Musso, D. (2015). *Clin Microbiol Infect* **21**, e47-48.
- Nickens, D. G. & Hardy, R. W. (2008). *Virology* **370**, 158-172.
- Nie, J., Sakamoto, S., Song, D., Qu, Z., Ota, K. & Taniguchi, T. (1998). *FEBS Lett* **424**, 27-32.
- Niesters, H. G. & Strauss, J. H. (1990a). *J Virol* **64**, 1639-1647.
- Niesters, H. G. & Strauss, J. H. (1990b). *J Virol* **64**, 4162-4168.
- Nikonov, A., Molder, T., Sikut, R., Kiiver, K., Mannik, A., Toots, U., Lulla, A., Lulla, V., Utt, A., Merits, A. & Ustav, M. (2013). *PLoS Pathog* **9**, e1003610.
- Nkoghe, D., Kassa, R. F., Caron, M., Grard, G., Mombo, I., Bikie, B., Paupy, C., Becquart, P., Bisvigou, U. & Leroy, E. M. (2012). *PLoS Negl Trop Dis* **6**, e1517.
- Nunes, M. R., Faria, N. R., de Vasconcelos, J. M., Golding, N., Kraemer, M. U., de Oliveira, L. F., Azevedo Rdo, S., da Silva, D. E., da Silva, E. V., da Silva, S. P., Carvalho, V. L., Coelho, G. E., Cruz, A. C., Rodrigues, S. G., Vianez, J. L., Jr., Nunes, B. T., Cardoso, J. F., Tesh, R. B., Hay, S. I., Pybus, O. G. & Vasconcelos, P. F. (2015). *BMC Med* **13**, 102.
- Nuss, D. L., Furuichi, Y., Koch, G. & Shatkin, A. J. (1975). *Cell* **6**, 21-27.
- O'Reilly, E. K. & Kao, C. C. (1998). *Virology* **252**, 287-303.
- Oberste, M. S., Fraire, M., Navarro, R., Zepeda, C., Zarate, M. L., Ludwig, G. V., Kondig, J. F., Weaver, S. C., Smith, J. F. & Rico-Hesse, R. (1998). *Am J Trop Med Hyg* **59**, 100-107.
- Oberste, M. S., Parker, M. D. & Smith, J. F. (1996). *Virology* **219**, 314-320.

- Oberste, M. S., Schmura, S. M., Weaver, S. C. & Smith, J. F. (1999). *Am J Trop Med Hyg* **60**, 630-634.
- Oberste, M. S., Weaver, S. C., Watts, D. M. & Smith, J. F. (1998). *Am J Trop Med Hyg* **58**, 41-46.
- Ogata, N., Ueda, K. & Hayaishi, O. (1980). *J Biol Chem* **255**, 7610-7615.
- Ogino, T. & Banerjee, A. K. (2007). *Mol Cell* **25**, 85-97.
- Ogino, T., Yadav, S. P. & Banerjee, A. K. (2010). *Proc Natl Acad Sci U S A* **107**, 3463-3468.
- Oliver, F. J., Menissier-de Murcia, J., Nacci, C., Decker, P., Andriantsitohaina, R., Muller, S., de la Rubia, G., Stoclet, J. C. & de Murcia, G. (1999). *EMBO J* **18**, 4446-4454.
- Onomoto, K., Yoneyama, M., Fung, G., Kato, H. & Fujita, T. (2014). *Trends Immunol* **35**, 420-428.
- Oruetxebarria, I., Guo, D., Merits, A., Makinen, K., Saarma, M. & Valkonen, J. P. (2001). *Virus Res* **73**, 103-112.
- Oshiumi, H., Sakai, K., Matsumoto, M. & Seya, T. (2010). *Eur J Immunol* **40**, 940-948.
- Ou, J. H., Strauss, E. G. & Strauss, J. H. (1981). *Virology* **109**, 281-289.
- Ou, J. H., Strauss, E. G. & Strauss, J. H. (1983). *J Mol Biol* **168**, 1-15.
- Ou, J. H., Trent, D. W. & Strauss, J. H. (1982). *J Mol Biol* **156**, 719-730.
- Paddock, M. N., Buelow, B. D., Takeda, S. & Scharenberg, A. M. (2010). *PLoS Biol* **8**, e1000428.
- Paesen, G. C., Collet, A., Sallamand, C., Debart, F., Vasseur, J. J., Canard, B., Decroly, E. & Grimes, J. M. (2015). *Nat Commun* **6**, 8749.
- Paessler, S., Ni, H., Petrakova, O., Fayzulin, R. Z., Yun, N., Anishchenko, M., Weaver, S. C. & Frolov, I. (2006). *J Virol* **80**, 2784-2796.
- Panas, M. D., Ahola, T. & McInerney, G. M. (2014). *J Virol* **88**, 5888-5893.
- Panas, M. D., Varjak, M., Lulla, A., Eng, K. E., Merits, A., Karlsson Hedestam, G. B. & McInerney, G. M. (2012). *Mol Biol Cell* **23**, 4701-4712.
- Panning, M., Grywna, K., van Esbroeck, M., Emmerich, P. & Drosten, C. (2008). *Emerg Infect Dis* **14**, 416-422.
- Paquet, C., Quatresous, I., Solet, J. L., Sissoko, D., Renault, P., Pierre, V., Cordel, H., Lassalle, C., Thiria, J., Zeller, H. & Schuffnecker, I. (2006). *Euro Surveill* **11**, E060202 060203.
- Paredes, A. M., Brown, D. T., Rothnagel, R., Chiu, W., Schoepp, R. J., Johnston, R. E. & Prasad, B. V. (1993). *Proc Natl Acad Sci U S A* **90**, 9095-9099.
- Paredes, A. M., Simon, M. N. & Brown, D. T. (1992). *Virology* **187**, 329-332.
- Park, E. & Griffin, D. E. (2009). *Virology* **388**, 305-314.

- Parrott, M. M., Sitarski, S. A., Arnold, R. J., Picton, L. K., Hill, R. B. & Mukhopadhyay, S. (2009). *J Virol* **83**, 2584-2591.
- Parvez, M. K. (2015). *Gene* **566**, 47-53.
- Pastorino, B., Muyembe-Tamfum, J. J., Bessaud, M., Tock, F., Tolou, H., Durand, J. P. & Peyrefitte, C. N. (2004). *J Med Virol* **74**, 277-282.
- Paterson, H. E. & McIntosh, B. M. (1964). *Ann Trop Med Parasitol* **58**, 52-55.
- Pehrson, J. R. & Fried, V. A. (1992). *Science* **257**, 1398-1400.
- Pehrson, J. R. & Fuji, R. N. (1998). *Nucleic Acids Res* **26**, 2837-2842.
- Peisley, A., Jo, M. H., Lin, C., Wu, B., Orme-Johnson, M., Walz, T., Hohng, S. & Hur, S. (2012). *Proc Natl Acad Sci U S A* **109**, E3340-3349.
- Peisley, A., Wu, B., Yao, H., Walz, T. & Hur, S. (2013). *Mol Cell* **51**, 573-583.
- Pellat-Deceunynck, C., Wietzerbin, J. & Drapier, J. C. (1994). *Biochem J* **297 (Pt 1)**, 53-58.
- Peranen, J., Laakkonen, P., Hyvonen, M. & Kaariainen, L. (1995). *Virology* **208**, 610-620.
- Peranen, J., Rikkonen, M., Liljestrom, P. & Kaariainen, L. (1990). *J Virol* **64**, 1888-1896.
- Peranen, J., Takkinen, K., Kalkkinen, N. & Kaariainen, L. (1988). *J Gen Virol* **69 (Pt 9)**, 2165-2178.
- Perina, D., Mikoc, A., Ahel, J., Cetkovic, H., Zaja, R. & Ahel, I. (2014). *DNA Repair (Amst)* **23**, 4-16.
- Peterson, F. C., Chen, D., Lytle, B. L., Rossi, M. N., Ahel, I., Denu, J. M. & Volkman, B. F. (2011). *J Biol Chem* **286**, 35955-35965.
- Petitdemange, C., Wauquier, N. & Vieillard, V. (2015). *J Allergy Clin Immunol* **135**, 846-855.
- Pettersson, R. F., Soderlund, H. & Kaariainen, L. (1980). *Eur J Biochem* **105**, 435-443.
- Peyrefitte, C. N., Bessaud, M., Pastorino, B. A., Gravier, P., Plumet, S., Merle, O. L., Moltini, I., Coppin, E., Tock, F., Daries, W., Ollivier, L., Pages, F., Martin, R., Boniface, F., Tolou, H. J. & Grandadam, M. (2008). *J Med Virol* **80**, 430-433.
- Phuc, H. K., Andreasen, M. H., Burton, R. S., Vass, C., Epton, M. J., Pape, G., Fu, G., Condon, K. C., Scaife, S., Donnelly, C. A., Coleman, P. G., White-Cooper, H. & Alpey, L. (2007). *BMC Biol* **5**, 11.
- Phulwani, N. K. & Kielian, T. (2008). *J Neurochem* **106**, 578-590.
- Pialoux, G., Gauzere, B. A., Jaureguiberry, S. & Strobel, M. (2007). *Lancet Infect Dis* **7**, 319-327.
- Pichlmair, A., Lassnig, C., Eberle, C. A., Gorna, M. W., Baumann, C. L., Burkard, T. R., Burckstummer, T., Stefanovic, A., Krieger, S., Bennett, K. L., Rulicke, T., Weber, F., Colinge, J., Muller, M. & Superti-Furga, G. (2011). *Nat Immunol* **12**, 624-630.

- Pichlmair, A., Schulz, O., Tan, C. P., Naslund, T. I., Liljestrom, P., Weber, F. & Reis e Sousa, C. (2006). *Science* **314**, 997-1001.
- Pietila, M. K., Hellstrom, K. & Ahola, T. (2017). *Virus Res* **234**, 44-57.
- Pietila, M. K., van Hemert, M. J. & Ahola, T. (2018). *J Virol* **92**.
- Pisano, M. B., Torres, C., Re, V. E., Farias, A. A., Sanchez Seco, M. P., Tenorio, A., Campos, R. & Contigiani, M. S. (2014). *Infect Genet Evol* **26**, 72-79.
- Pletnev, S. V., Zhang, W., Mukhopadhyay, S., Fisher, B. R., Hernandez, R., Brown, D. T., Baker, T. S., Rossmann, M. G. & Kuhn, R. J. (2001). *Cell* **105**, 127-136.
- Plotch, S. J., Bouloy, M. & Krug, R. M. (1979). *Proc Natl Acad Sci U S A* **76**, 1618-1622.
- Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. (1981). *Cell* **23**, 847-858.
- Popoff, I., Jijon, H., Monia, B., Tavernini, M., Ma, M., McKay, R. & Madsen, K. (2002). *J Pharmacol Exp Ther* **303**, 1145-1154.
- Posavec Marjanovic, M., Hurtado-Bages, S., Lassi, M., Valero, V., Malinverni, R., Delage, H., Navarro, M., Corujo, D., Guberovic, I., Douet, J., Gama-Perez, P., Garcia-Roves, P. M., Ahel, I., Ladurner, A. G., Yanes, O., Bouvet, P., Suelves, M., Teperino, R., Pospisilik, J. A. & Buschbeck, M. (2017). *Nat Struct Mol Biol* **24**, 902-910.
- Powers, A. M. (2015). *J Gen Virol* **96**, 1-5.
- Powers, A. M. (2018). *Clin Microbiol Rev* **31**.
- Powers, A. M. (2019). *Lancet* **392**, 2660-2661.
- Powers, A. M., Brault, A. C., Tesh, R. B. & Weaver, S. C. (2000). *J Gen Virol* **81**, 471-479.
- Powers, A. M. & Logue, C. H. (2007). *J Gen Virol* **88**, 2363-2377.
- Powers, A. M., Oberste, M. S., Brault, A. C., Rico-Hesse, R., Schmura, S. M., Smith, J. F., Kang, W., Sweeney, W. P. & Weaver, S. C. (1997). *J Virol* **71**, 6697-6705.
- Powers, A. M. & Waterman, S. H. (2017). *PLoS Negl Trop Dis* **11**, e0005421.
- Prokic, I., Cowling, B. S. & Laporte, J. (2014). *J Mol Med (Berl)* **92**, 453-463.
- Pushko, P., Parker, M., Ludwig, G. V., Davis, N. L., Johnston, R. E. & Smith, J. F. (1997). *Virology* **239**, 389-401.
- Putics, A., Filipowicz, W., Hall, J., Gorbalenya, A. E. & Ziebuhr, J. (2005). *J Virol* **79**, 12721-12731.
- Putics, A., Slaby, J., Filipowicz, W., Gorbalenya, A. E. & Ziebuhr, J. (2006). *Adv Exp Med Biol* **581**, 93-96.
- Qi, X., Lan, S., Wang, W., Schelde, L. M., Dong, H., Wallat, G. D., Ly, H., Liang, Y. & Dong, C. (2010). *Nature* **468**, 779-783.
- Quiroz, E., Aguilar, P. V., Cisneros, J., Tesh, R. B. & Weaver, S. C. (2009). *PLoS Negl Trop Dis* **3**, e472.

- Rack, J. G., Morra, R., Barkauskaite, E., Kraehenbuehl, R., Ariza, A., Qu, Y., Ortmayer, M., Leidecker, O., Cameron, D. R., Matic, I., Peleg, A. Y., Leys, D., Traven, A. & Ahel, I. (2015). *Mol Cell* **59**, 309-320.
- Rack, J. G., Perina, D. & Ahel, I. (2016). *Annu Rev Biochem* **85**, 431-454.
- Rahmeh, A. A., Li, J., Kranzusch, P. J. & Whelan, S. P. (2009). *J Virol* **83**, 11043-11050.
- Rajapakse, S., Rodrigo, C. & Rajapakse, A. (2010). *Trans R Soc Trop Med Hyg* **104**, 89-96.
- Raju, R., Hajjou, M., Hill, K. R., Botta, V. & Botta, S. (1999). *J Virol* **73**, 2410-2419.
- Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. (2004). *Cell* **118**, 229-241.
- Ramakrishnan, C., Kutumbarao, N. H. V., Suhitha, S. & Velmurugan, D. (2017). *Chem Biol Drug Des* **89**, 772-782.
- Ramanathan, A., Devarkar, S. C., Jiang, F., Miller, M. T., Khan, A. G., Marcotrigiano, J. & Patel, S. S. (2016). *Nucleic Acids Res* **44**, 896-909.
- Ramanathan, A., Robb, G. B. & Chan, S. H. (2016). *Nucleic Acids Res* **44**, 7511-7526.
- Ramsey, J. & Mukhopadhyay, S. (2017). *Viruses* **9**.
- Ramsey, J., Renzi, E. C., Arnold, R. J., Trinidad, J. C. & Mukhopadhyay, S. (2017). *J Virol* **91**.
- Ravichandran, R. & Manian, M. (2008). *J Infect Dev Ctries* **2**, 140-142.
- Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crepin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R. W. & Cusack, S. (2014). *Nature* **516**, 361-366.
- Reineke, L. C., Kedersha, N., Langereis, M. A., van Kuppeveld, F. J. & Lloyd, R. E. (2015). *MBio* **6**, e02486.
- Renault, P., Josseran, L. & Pierre, V. (2008). *Emerg Infect Dis* **14**, 1327.
- Reynaud, J. M., Kim, D. Y., Atasheva, S., Rasalouslykaya, A., White, J. P., Diamond, M. S., Weaver, S. C., Frolova, E. I. & Frolov, I. (2015). *PLoS Pathog* **11**, e1004863.
- Rezza, G., Nicoletti, L., Angelini, R., Romi, R., Finarelli, A. C., Panning, M., Cordioli, P., Fortuna, C., Boros, S., Magurano, F., Silvi, G., Angelini, P., Dottori, M., Ciufolini, M. G., Majori, G. C., Cassone, A. & group, C. s. (2007). *Lancet* **370**, 1840-1846.
- Rianthavorn, P., Prianantathavorn, K., Wuttirattanakowit, N., Theamboonlers, A. & Poovorawan, Y. (2010). *Int J Infect Dis* **14 Suppl 3**, e161-165.
- Rice, C. M., Bell, J. R., Hunkapiller, M. W., Strauss, E. G. & Strauss, J. H. (1982). *J Mol Biol* **154**, 355-378.
- Rice, C. M., Levis, R., Strauss, J. H. & Huang, H. V. (1987). *J Virol* **61**, 3809-3819.
- Rice, C. M. & Strauss, J. H. (1981). *Proc Natl Acad Sci U S A* **78**, 2062-2066.
- Rico-Hesse, R., Roehrig, J. T., Trent, D. W. & Dickerman, R. W. (1988). *Am J Trop Med Hyg* **38**, 195-204.

- Rieder, E., Paul, A. V., Kim, D. W., van Boom, J. H. & Wimmer, E. (2000). *J Virol* **74**, 10371-10380.
- Rifkin, I. R., Leadbetter, E. A., Busconi, L., Viglianti, G. & Marshak-Rothstein, A. (2005). *Immunol Rev* **204**, 27-42.
- Rikkonen, M., Peranen, J. & Kaariainen, L. (1994a). *Arch Virol Suppl* **9**, 369-377.
- Rikkonen, M., Peranen, J. & Kaariainen, L. (1994b). *J Virol* **68**, 5804-5810.
- Riquelme, P. T., Burzio, L. O. & Koide, S. S. (1979). *J Biol Chem* **254**, 3018-3028.
- Rivas, F., Diaz, L. A., Cardenas, V. M., Daza, E., Bruzon, L., Alcalá, A., De la Hoz, O., Caceres, F. M., Aristizabal, G., Martinez, J. W., Revelo, D., De la Hoz, F., Boshell, J., Camacho, T., Calderon, L., Olano, V. A., Villarreal, L. I., Roselli, D., Alvarez, G., Ludwig, G. & Tsai, T. (1997). *J Infect Dis* **175**, 828-832.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res* **42**, W320-324.
- Roby, J. A., Clarke, B. D. & Khromykh, A. A. (2014). *Trends Microbiol* **22**, 171-173.
- Rolli, V., O'Farrell, M., Menissier-de Murcia, J. & de Murcia, G. (1997). *Biochemistry* **36**, 12147-12154.
- Rosenthal, F., Feijs, K. L., Frugier, E., Bonalli, M., Forst, A. H., Imhof, R., Winkler, H. C., Fischer, D., Caflisch, A., Hassa, P. O., Luscher, B. & Hottiger, M. O. (2013). *Nat Struct Mol Biol* **20**, 502-507.
- Rosenthal, F. & Hottiger, M. O. (2014). *Front Biosci (Landmark Ed)* **19**, 1041-1056.
- Ross, R. W. (1956). *J Hyg (Lond)* **54**, 177-191.
- Rossi, A. L. (1967). *Prog Med Virol* **9**, 176-203.
- Roth, A., Mercier, A., Lepers, C., Hoy, D., Duituturaga, S., Benyon, E., Guillaumot, L. & Soares, Y. (2014). *Euro Surveill* **19**.
- Rothan, H. A., Bahrani, H., Mohamed, Z., Teoh, T. C., Shankar, E. M., Rahman, N. A. & Yusof, R. (2015). *PLoS One* **10**, e0126360.
- Rozanov, M. N., Koonin, E. V. & Gorbalenya, A. E. (1992). *J Gen Virol* **73 (Pt 8)**, 2129-2134.
- Rubach, J. K., Wasik, B. R., Rupp, J. C., Kuhn, R. J., Hardy, R. W. & Smith, J. L. (2009). *Virology* **384**, 201-208.
- Ruigrok, R. W., Crepin, T., Hart, D. J. & Cusack, S. (2010). *Curr Opin Struct Biol* **20**, 104-113.
- Ruiz-Guillen, M., Gabev, E., Quetglas, J. I., Casales, E., Ballesteros-Briones, M. C., Poutou, J., Aranda, A., Martisova, E., Bezunartea, J., Ondiviela, M., Prieto, J., Hernandez-Alcoceba, R., Abrescia, N. G. & Smerdou, C. (2016). *Cell Mol Life Sci* **73**, 3897-3916.
- Rumenapf, T., Strauss, E. G. & Strauss, J. H. (1994). *J Virol* **68**, 56-62.
- Rupp, J. C., Jundt, N. & Hardy, R. W. (2011). *J Virol* **85**, 3449-3460.

- Rupp, J. C., Sokoloski, K. J., Gebhart, N. N. & Hardy, R. W. (2015). *J Gen Virol* **96**, 2483-2500.
- Russo, A. T., White, M. A. & Watowich, S. J. (2006). *Structure* **14**, 1449-1458.
- Ryman, K. D., Klimstra, W. B. & Johnston, R. E. (2004). *Virology* **322**, 1-12.
- Saikatendu, K. S., Joseph, J. S., Subramanian, V., Clayton, T., Griffith, M., Moy, K., Velasquez, J., Neuman, B. W., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2005). *Structure* **13**, 1665-1675.
- Salas, R. A., Garcia, C. Z., Liria, J., Barrera, R., Navarro, J. C., Medina, G., Vasquez, C., Fernandez, Z. & Weaver, S. C. (2001). *Am J Trop Med Hyg* **64**, 84-92.
- Salditt-Georgieff, M., Jelinek, W., Darnell, J. E., Furuichi, Y., Morgan, M. & Shatkin, A. (1976). *Cell* **7**, 227-237.
- Salminen, A., Wahlberg, J. M., Lobigs, M., Liljestrom, P. & Garoff, H. (1992). *J Cell Biol* **116**, 349-357.
- Salonen, A., Ahola, T. & Kaariainen, L. (2005). *Curr Top Microbiol Immunol* **285**, 139-173.
- Salonen, A., Vasiljeva, L., Merits, A., Magden, J., Jokitalo, E. & Kaariainen, L. (2003). *J Virol* **77**, 1691-1702.
- Sam, I. C. & AbuBakar, S. (2006). *Med J Malaysia* **61**, 264-269.
- Sam, I. C., Chan, Y. F., Chan, S. Y., Loong, S. K., Chin, H. K., Hooi, P. S., Ganeswrie, R. & Abubakar, S. (2009). *J Clin Virol* **46**, 180-183.
- Sambucci, M., Laudisi, F., Novelli, F., Bennici, E., Rosado, M. M. & Pioli, C. (2013). *ScientificWorldJournal* **2013**, 375024.
- Sanchez, J. L., Takafuji, E. T., Lednar, W. M., LeDuc, J. W., Macasaet, F. F., Mangiafico, J. A., Rosato, R. R., Driggers, D. P. & Haecker, J. C. (1984). *Mil Med* **149**, 618-621.
- Sang, R. C., Ahmed, O., Faye, O., Kelly, C. L., Yahaya, A. A., Mmadi, I., Toilibou, A., Sergon, K., Brown, J., Agata, N., Yakouide, A., Ball, M. D., Breiman, R. F., Miller, B. R. & Powers, A. M. (2008). *Am J Trop Med Hyg* **78**, 77-82.
- Sanmartin-Barberi, C., Groot, H. & Osorno-Mesa, E. (1954). *Am J Trop Med Hyg* **3**, 283-293.
- Santhosh, S. R., Dash, P. K., Parida, M., Khan, M. & Rao, P. V. (2009). *Viol J* **6**, 172.
- Sanz, M. A. & Carrasco, L. (2001). *J Virol* **75**, 7778-7784.
- Sanz, M. A., Madan, V., Carrasco, L. & Nieva, J. L. (2003). *J Biol Chem* **278**, 2051-2057.
- Sauter, D. (2014). *Front Microbiol* **5**, 163.
- Sauve, A. A. & Youn, D. Y. (2012). *Curr Opin Chem Biol* **16**, 535-543.
- Sawicki, D. L. & Gomatos, P. J. (1976). *J Virol* **20**, 446-464.
- Sawicki, D. L., Kaariainen, L., Lambek, C. & Gomatos, P. J. (1978). *J Virol* **25**, 19-27.

- Sawicki, D. L., Perri, S., Polo, J. M. & Sawicki, S. G. (2006). *J Virol* **80**, 360-371.
- Sawicki, D. L. & Sawicki, S. G. (1980). *J Virol* **34**, 108-118.
- Sawicki, D. L. & Sawicki, S. G. (1994). *Arch Virol Suppl* **9**, 393-405.
- Sawicki, D. L., Silverman, R. H., Williams, B. R. & Sawicki, S. G. (2003). *J Virol* **77**, 1801-1811.
- Scherer, W. F. & Anderson, K. (1975). *Am J Epidemiol* **101**, 356-361.
- Scherer, W. F. & Chin, J. (1983). *Am J Trop Med Hyg* **32**, 871-876.
- Scherer, W. F., Dickerman, R. W., Chia, C. W., Ventura, A., Moorhouse, A. & Geiger, R. (1964). *Science* **145**, 274-275.
- Scherer, W. F., Madalengoitia, J., Flores, W. & Acosta, M. (1975). *Am J Epidemiol* **101**, 347-355.
- Schilte, C., Staikowsky, F., Couderc, T., Madec, Y., Carpentier, F., Kassab, S., Albert, M. L., Lecuit, M. & Michault, A. (2013). *PLoS Negl Trop Dis* **7**, e2137.
- Schlee, M. (2013). *Immunobiology* **218**, 1322-1335.
- Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. (1995). *J Mol Biol* **247**, 16-20.
- Schmidt, A., Schwerd, T., Hamm, W., Hellmuth, J. C., Cui, S., Wenzel, M., Hoffmann, F. S., Michallet, M. C., Besch, R., Hopfner, K. P., Endres, S. & Rothenfusser, S. (2009). *Proc Natl Acad Sci U S A* **106**, 12067-12072.
- Schmidt, M. F. (1982). *Virology* **116**, 327-338.
- Schmidt, M. F., Bracha, M. & Schlesinger, M. J. (1979). *Proc Natl Acad Sci U S A* **76**, 1687-1691.
- Schoenberg, D. R. & Maquat, L. E. (2009). *Trends Biochem Sci* **34**, 435-442.
- Scholte, F. E., Tas, A., Albulescu, I. C., Zusinaite, E., Merits, A., Snijder, E. J. & van Hemert, M. J. (2015). *J Virol* **89**, 4457-4469.
- Schoneboom, B. A., Lee, J. S. & Grieder, F. B. (2000). *J Interferon Cytokine Res* **20**, 205-215.
- Schreiber, V., Dantzer, F., Ame, J. C. & de Murcia, G. (2006). *Nat Rev Mol Cell Biol* **7**, 517-528.
- Schroeder, C. E., Yao, T., Sotsky, J., Smith, R. A., Roy, S., Chu, Y. K., Guo, H., Tower, N. A., Noah, J. W., McKellip, S., Sosa, M., Rasmussen, L., Smith, L. H., White, E. L., Aube, J., Jonsson, C. B., Chung, D. & Golden, J. E. (2014). *J Med Chem* **57**, 8608-8621.
- Schuffenecker, I., Itean, I., Michault, A., Murri, S., Frangeul, L., Vaney, M. C., Lavenir, R., Pardigon, N., Reynes, J. M., Pettinelli, F., Biscornet, L., Diancourt, L., Michel, S., Duquerroy, S., Guigon, G., Frenkiel, M. P., Brehin, A. C., Cubito, N., Despres, P., Kunst, F., Rey, F. A., Zeller, H. & Brisse, S. (2006). *PLoS Med* **3**, e263.

- Schulte, T., Liu, L., Panas, M. D., Thaa, B., Dickson, N., Gotte, B., Achour, A. & McInerney, G. M. (2016). *Open Biol* **6**.
- Schwartz, O. & Albert, M. L. (2010). *Nat Rev Microbiol* **8**, 491-500.
- Sefton, B. M. (1977). *Cell* **10**, 659-668.
- Selvam, P., Vijayalakshimi, P., Smee, D. F., Gowen, B. B., Julander, J. G., Day, C. W. & Barnard, D. L. (2007). *Antivir Chem Chemother* **18**, 301-305.
- Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. (2005). *Cell* **122**, 669-682.
- Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L. & Cech, T. R. (1999). *Nature* **401**, 177-180.
- Severini, F., Boccolini, D., Fortuna, C., Di Luca, M., Toma, L., Amendola, A., Benedetti, E., Minelli, G., Romi, R., Venturi, G., Rezza, G. & Remoli, M. E. (2018). *PLoS Negl Trop Dis* **12**, e0006435.
- Shall, S. & de Murcia, G. (2000). *Mutat Res* **460**, 1-15.
- Sharifi, R., Morra, R., Appel, C. D., Tallis, M., Chioza, B., Jankevicius, G., Simpson, M. A., Matic, I., Ozkan, E., Golia, B., Schellenberg, M. J., Weston, R., Williams, J. G., Rossi, M. N., Galehdari, H., Krahn, J., Wan, A., Trembath, R. C., Crosby, A. H., Ahel, D., Hay, R., Ladurner, A. G., Timinszky, G., Williams, R. S. & Ahel, I. (2013). *EMBO J* **32**, 1225-1237.
- Shatkin, A. J. (1976). *Cell* **9**, 645-653.
- Shin, G., Yost, S. A., Miller, M. T., Elrod, E. J., Grakoui, A. & Marcotrigiano, J. (2012). *Proc Natl Acad Sci U S A* **109**, 16534-16539.
- Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajaska, A., Harbers, M., Kawai, J., Carninci, P. & Hayashizaki, Y. (2003). *Proc Natl Acad Sci U S A* **100**, 15776-15781.
- Shirako, Y., Strauss, E. G. & Strauss, J. H. (2000). *Virology* **276**, 148-160.
- Shirako, Y. & Strauss, J. H. (1990). *Virology* **177**, 54-64.
- Shirako, Y. & Strauss, J. H. (1994). *J Virol* **68**, 1874-1885.
- Shirako, Y. & Strauss, J. H. (1998). *J Virol* **72**, 2310-2315.
- Shope, R. E., Causey, O. R. & De Andrade, A. H. (1964). *Am J Trop Med Hyg* **13**, 723-727.
- Shuman, S. (1995). *Prog Nucleic Acid Res Mol Biol* **50**, 101-129.
- Shuman, S. (2002). *Nat Rev Mol Cell Biol* **3**, 619-625.
- Shuman, S. & Lima, C. D. (2004). *Curr Opin Struct Biol* **14**, 757-764.
- Shuman, S. & Schwer, B. (1995). *Mol Microbiol* **17**, 405-410.
- Sievers, F. & Higgins, D. G. (2014). *Curr Protoc Bioinformatics* **48**, 3 13 11-16.
- Silva, L. A. & Dermody, T. S. (2017). *J Clin Invest* **127**, 737-749.

- Sjoberg, M., Lindqvist, B. & Garoff, H. (2011). *J Virol* **85**, 5644-5650.
- Skinner, R. B., Jr. (2003). *Dermatol Clin* **21**, 291-300.
- Skoging, U., Vihinen, M., Nilsson, L. & Liljestrom, P. (1996). *Structure* **4**, 519-529.
- Slade, D., Dunstan, M. S., Barkauskaite, E., Weston, R., Lafite, P., Dixon, N., Ahel, M., Leys, D. & Ahel, I. (2011). *Nature* **477**, 616-620.
- Smith, T. J., Cheng, R. H., Olson, N. H., Peterson, P., Chase, E., Kuhn, R. J. & Baker, T. S. (1995). *Proc Natl Acad Sci U S A* **92**, 10648-10652.
- Snead, W. T., Hayden, C. C., Gadok, A. K., Zhao, C., Lafer, E. M., Rangamani, P. & Stachowiak, J. C. (2017). *Proc Natl Acad Sci U S A* **114**, E3258-E3267.
- Snyder, A. J. & Mukhopadhyay, S. (2012). *J Virol* **86**, 13609-13620.
- Snyder, J. E., Berrios, C. J., Edwards, T. J., Jose, J., Perera, R. & Kuhn, R. J. (2012). *J Virol* **86**, 12372-12383.
- Snyder, J. E., Kulcsar, K. A., Schultz, K. L., Riley, C. P., Neary, J. T., Marr, S., Jose, J., Griffin, D. E. & Kuhn, R. J. (2013). *J Virol* **87**, 8511-8523.
- Sokoloski, K. J., Dickson, A. M., Chaskey, E. L., Garneau, N. L., Wilusz, C. J. & Wilusz, J. (2010). *Cell Host Microbe* **8**, 196-207.
- Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J. & et al. (1993). *Cell* **72**, 767-778.
- Sonnabend, J. A., Dalgarno, L., Friedman, R. M., and Martin, E. M., (1964). *Biochem. Biophys. Res. Commun.*, **17**, 6.
- Sonnabend, J. A., Martin, E. M. & Mecs, E. (1967). *Nature* **213**, 365-367.
- Soonsawad, P., Xing, L., Milla, E., Espinoza, J. M., Kawano, M., Marko, M., Hsieh, C., Furukawa, H., Kawasaki, M., Weerachatanukul, W., Srivastava, R., Barnett, S. W., Srivastava, I. K. & Cheng, R. H. (2010). *J Virol* **84**, 11145-11151.
- Spuul, P., Balistreri, G., Hellstrom, K., Golubtsov, A. V., Jokitalo, E. & Ahola, T. (2011). *J Virol* **85**, 4739-4751.
- Spuul, P., Balistreri, G., Kaariainen, L. & Ahola, T. (2010). *J Virol* **84**, 7543-7557.
- Spuul, P., Salonen, A., Merits, A., Jokitalo, E., Kaariainen, L. & Ahola, T. (2007). *J Virol* **81**, 872-883.
- Sreevalsan, T. & Lockart, R. Z., Jr. (1966). *Proc Natl Acad Sci U S A* **55**, 974-981.
- Srihongse, S., Scherer, W. F. & Galindo, P. (1967). *Am J Trop Med Hyg* **16**, 519-524.
- Staples, J. E., Breiman, R. F. & Powers, A. M. (2009). *Clin Infect Dis* **49**, 942-948.
- Strauss, E. G., De Groot, R. J., Levinson, R. & Strauss, J. H. (1992). *Virology* **191**, 932-940.
- Strauss, E. G., Levinson, R., Rice, C. M., Dalrymple, J. & Strauss, J. H. (1988). *Virology* **164**, 265-274.

- Strauss, E. G., Rice, C. M. & Strauss, J. H. (1983). *Proc Natl Acad Sci U S A* **80**, 5271-5275.
- Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984). *Virology* **133**, 92-110.
- Strauss, E. G., Stec, D. S., Schmaljohn, A. L. & Strauss, J. H. (1991). *J Virol* **65**, 4654-4664.
- Strauss, J. H. & Strauss, E. G. (1994). *Microbiol Rev* **58**, 491-562.
- Sudia, W. D., Lord, R. D., Newhouse, V. F., Miller, D. L. & Kissling, R. E. (1971). *Am J Epidemiol* **93**, 137-143.
- Sugiyama, K., Obayashi, E., Kawaguchi, A., Suzuki, Y., Tame, J. R., Nagata, K. & Park, S. Y. (2009). *EMBO J* **28**, 1803-1811.
- Suhrbier, A., Jaffar-Bandjee, M. C. & Gasque, P. (2012). *Nat Rev Rheumatol* **8**, 420-429.
- Suomalainen, M., Liljestrom, P. & Garoff, H. (1992). *J Virol* **66**, 4737-4747.
- Suopanki, J., Sawicki, D. L., Sawicki, S. G. & Kaariainen, L. (1998). *J Gen Virol* **79** (Pt 2), 309-319.
- Suthar, M. S., Shabman, R., Madric, K., Lambeth, C. & Heise, M. T. (2005). *J Virol* **79**, 4219-4228.
- Sutton, L. S. & Brooke, C. C. (1954). *J Am Med Assoc* **155**, 1473-1476.
- Takagi, T., Moore, C. R., Diehn, F. & Buratowski, S. (1997). *Cell* **89**, 867-873.
- Takeuchi, O. & Akira, S. (2007). *Immunol Rev* **220**, 214-224.
- Takkinen, K. (1986). *Nucleic Acids Res* **14**, 5667-5682.
- Tamm, K., Merits, A. & Sarand, I. (2008). *J Gen Virol* **89**, 676-686.
- Tan, E. S., Krukenberg, K. A. & Mitchison, T. J. (2012). *Anal Biochem* **428**, 126-136.
- Tan, J., Vonrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., Hansen, G., Mesters, J. R., Schmidt, C. L. & Hilgenfeld, R. (2009). *PLoS Pathog* **5**, e1000428.
- Tao, Z., Gao, P. & Liu, H. W. (2009). *J Am Chem Soc* **131**, 14258-14260.
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. (2007). *Nat Struct Mol Biol* **14**, 1025-1040.
- Taylor, G. M., Hanson, P. I. & Kielian, M. (2007). *J Virol* **81**, 13631-13639.
- Tellinghuisen, T. L. & Kuhn, R. J. (2000). *J Virol* **74**, 4302-4309.
- Thaa, B., Biasiotto, R., Eng, K., Neuvonen, M., Gotte, B., Rheinemann, L., Mutso, M., Utt, A., Varghese, F., Balistreri, G., Merits, A., Ahola, T. & McInerney, G. M. (2015). *J Virol* **89**, 11420-11437.
- Thedieck, K., Holzwarth, B., Prentzell, M. T., Boehlke, C., Klasener, K., Ruf, S., Sonntag, A. G., Maerz, L., Grellscheid, S. N., Kremmer, E., Nitschke, R., Kuehn, E. W., Jonker, J. W., Groen, A. K., Reth, M., Hall, M. N. & Baumeister, R. (2013). *Cell* **154**, 859-874.

- Thiberville, S. D., Moyen, N., Dupuis-Maguiraga, L., Nougairede, A., Gould, E. A., Roques, P. & de Lamballerie, X. (2013). *Antiviral Res* **99**, 345-370.
- Timinszky, G., Till, S., Hassa, P. O., Hothorn, M., Kustatscher, G., Nijmeijer, B., Colombelli, J., Altmeyer, M., Stelzer, E. H., Scheffzek, K., Hottiger, M. O. & Ladurner, A. G. (2009). *Nat Struct Mol Biol* **16**, 923-929.
- Tomar, S., Hardy, R. W., Smith, J. L. & Kuhn, R. J. (2006). *J Virol* **80**, 9962-9969.
- Tomar, S., Narwal, M., Harms, E., Smith, J. L. & Kuhn, R. J. (2011). *Protein Expr Purif* **79**, 277-284.
- Tong, L. & Denu, J. M. (2010). *Biochim Biophys Acta* **1804**, 1617-1625.
- Topisirovic, I., Svitkin, Y. V., Sonenberg, N. & Shatkin, A. J. (2011). *Wiley Interdiscip Rev RNA* **2**, 277-298.
- Torii, S., Orba, Y., Hang'ombe, B. M., Mweene, A. S., Wada, Y., Anindita, P. D., Phongphaew, W., Qiu, Y., Kajihara, M., Mori-Kajihara, A., Eto, Y., Harima, H., Sasaki, M., Carr, M., Hall, W. W., Eshita, Y., Abe, T. & Sawa, H. (2018). *Virus Res* **250**, 31-36.
- Tossavainen, H., Aitio, O., Hellman, M., Saksela, K. & Permi, P. (2016). *J Biol Chem* **291**, 16307-16317.
- Tourriere, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J. M., Bertrand, E. & Tazi, J. (2003). *J Cell Biol* **160**, 823-831.
- Trocoli, A. & Djavaheri-Mergny, M. (2011). *Am J Cancer Res* **1**, 629-649.
- Trotman, J. B. & Schoenberg, D. R. (2018). *Wiley Interdiscip Rev RNA*, e1504.
- Tsetsarkin, K. A., Chen, R., Leal, G., Forrester, N., Higgs, S., Huang, J. & Weaver, S. C. (2011). *Proc Natl Acad Sci U S A* **108**, 7872-7877.
- Tsetsarkin, K. A., Chen, R., Sherman, M. B. & Weaver, S. C. (2011). *Curr Opin Virol* **1**, 310-317.
- Tsetsarkin, K. A., Chen, R. & Weaver, S. C. (2016). *Curr Opin Virol* **16**, 143-150.
- Tsetsarkin, K. A., Chen, R., Yun, R., Rossi, S. L., Plante, K. S., Guerbois, M., Forrester, N., Perng, G. C., Sreekumar, E., Leal, G., Huang, J., Mukhopadhyay, S. & Weaver, S. C. (2014). *Nat Commun* **5**, 4084.
- Tsetsarkin, K. A., McGee, C. E., Volk, S. M., Vanlandingham, D. L., Weaver, S. C. & Higgs, S. (2009). *PLoS One* **4**, e6835.
- Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E. & Higgs, S. (2007). *PLoS Pathog* **3**, e201.
- Tsetsarkin, K. A. & Weaver, S. C. (2011). *PLoS Pathog* **7**, e1002412.
- Tucker, J. A., Bennett, N., Brassington, C., Durant, S. T., Hassall, G., Holdgate, G., McAlister, M., Nissink, J. W., Truman, C. & Watson, M. (2012). *PLoS One* **7**, e50889.

- Tucker, P. C., Griffin, D. E., Choi, S., Bui, N. & Wesselingh, S. (1996). *J Virol* **70**, 3972-3977.
- Tucker, P. C., Lee, S. H., Bui, N., Martinie, D. & Griffin, D. E. (1997). *J Virol* **71**, 6106-6112.
- Tuittila, M. & Hinkkanen, A. E. (2003). *J Gen Virol* **84**, 1525-1533.
- Tuittila, M. T., Santagati, M. G., Roytta, M., Maatta, J. A. & Hinkkanen, A. E. (2000). *J Virol* **74**, 4579-4589.
- Turell, M. J., Dohm, D. J., Fernandez, R., Calampa, C. & O'Guinn, M. L. (2006). *J Am Mosq Control Assoc* **22**, 70-75.
- Turell, M. J., Jones, J. W., Sardelis, M. R., Dohm, D. J., Coleman, R. E., Watts, D. M., Fernandez, R., Calampa, C. & Klein, T. A. (2000). *J Med Entomol* **37**, 835-839.
- Uchime, O., Fields, W. & Kielian, M. (2013). *J Virol* **87**, 10255-10262.
- Urakova, N., Kuznetsova, V., Crossman, D. K., Sokratian, A., Guthrie, D. B., Kolykhalov, A. A., Lockwood, M. A., Natchus, M. G., Crowley, M. R., Painter, G. R., Frolova, E. I. & Frolov, I. (2017). *J Virol*.
- van den Hurk, A. F., Hall-Mendelin, S., Pyke, A. T., Frentiu, F. D., McElroy, K., Day, A., Higgs, S. & O'Neill, S. L. (2012). *PLoS Negl Trop Dis* **6**, e1892.
- van Duijn, L. P., Kasperaitis, M., Ameling, C. & Voorma, H. O. (1986). *Virus Res* **5**, 61-66.
- Varshavsky, A. (1992). *Cell* **69**, 725-735.
- Vashishtha, M., Phalen, T., Marquardt, M. T., Ryu, J. S., Ng, A. C. & Kielian, M. (1998). *J Cell Biol* **140**, 91-99.
- Vasiljeva, L., Merits, A., Auvinen, P. & Kaariainen, L. (2000). *J Biol Chem* **275**, 17281-17287.
- Vasiljeva, L., Merits, A., Golubtsov, A., Sizemskaja, V., Kaariainen, L. & Ahola, T. (2003). *J Biol Chem* **278**, 41636-41645.
- Vazeille, M., Moutailler, S., Coudrier, D., Rousseaux, C., Khun, H., Huerre, M., Thiria, J., Dehecq, J. S., Fontenille, D., Schuffenecker, I., Despres, P. & Failloux, A. B. (2007). *PLoS One* **2**, e1168.
- Vazeille, M., Zouache, K., Vega-Rua, A., Thiberge, J. M., Caro, V., Yebakima, A., Mousson, L., Piorkowski, G., Dauga, C., Vaney, M. C., Manni, M., Gasperi, G., de Lamballerie, X. & Failloux, A. B. (2016). *Sci Rep* **6**, 29564.
- Venkannagari, H., Fallarero, A., Feijs, K. L., Luscher, B. & Lehtio, L. (2013). *Eur J Pharm Sci* **49**, 148-156.
- Verheugd, P., Forst, A. H., Milke, L., Herzog, N., Feijs, K. L., Kremmer, E., Kleine, H. & Luscher, B. (2013). *Nat Commun* **4**, 1683.

- Vihinen, H., Ahola, T., Tuittila, M., Merits, A. & Kaariainen, L. (2001). *J Biol Chem* **276**, 5745-5752.
- Vihinen, H. & Saarinen, J. (2000). *J Biol Chem* **275**, 27775-27783.
- Vilcarromero, S., Aguilar, P. V., Halsey, E. S., Laguna-Torres, V. A., Razuri, H., Perez, J., Valderrama, Y., Gotuzzo, E., Suarez, L., Cespedes, M. & Kochel, T. J. (2010). *Emerg Infect Dis* **16**, 553-556.
- Vilcarromero, S., Laguna-Torres, V. A., Fernandez, C., Gotuzzo, E., Suarez, L., Cespedes, M., Aguilar, P. V. & Kochel, T. J. (2009). *Emerg Infect Dis* **15**, 323-325.
- Vina-Rodriguez, A., Eiden, M., Keller, M., Hinrichs, W. & Groschup, M. H. (2016). *Biomed Res Int* **2016**, 8543204.
- Virag, L. & Szabo, C. (2002). *Pharmacol Rev* **54**, 375-429.
- Vogel, R. H., Provencher, S. W., von Bonsdorff, C. H., Adrian, M. & Dubochet, J. (1986). *Nature* **320**, 533-535.
- Vognsen, T., Moller, I. R. & Kristensen, O. (2013). *PLoS One* **8**, e80947.
- Volk, S. M., Chen, R., Tsetsarkin, K. A., Adams, A. P., Garcia, T. I., Sall, A. A., Nasar, F., Schuh, A. J., Holmes, E. C., Higgs, S., Maharaj, P. D., Brault, A. C. & Weaver, S. C. (2010). *J Virol* **84**, 6497-6504.
- von Bonsdorff, C. H. & Harrison, S. C. (1978). *J Virol* **28**, 578-583.
- Voss, J. E., Vaney, M. C., Duquerroy, S., Vonnrhein, C., Girard-Blanc, C., Crublet, E., Thompson, A., Bricogne, G. & Rey, F. A. (2010). *Nature* **468**, 709-712.
- Vyas, S., Matic, I., Uchima, L., Rood, J., Zaja, R., Hay, R. T., Ahel, I. & Chang, P. (2014). *Nat Commun* **5**, 4426.
- Wada, Y., Orba, Y., Sasaki, M., Kobayashi, S., Carr, M. J., Nobori, H., Sato, A., Hall, W. W. & Sawa, H. (2017). *Virology* **505**, 102-112.
- Wahid, B., Ali, A., Rafique, S. & Idrees, M. (2017). *Int J Infect Dis* **58**, 69-76.
- Wahlberg, J. M., Boere, W. A. & Garoff, H. (1989). *J Virol* **63**, 4991-4997.
- Walters, R. W., Matheny, T., Mizoue, L. S., Rao, B. S., Muhlrad, D. & Parker, R. (2017). *Proc Natl Acad Sci U S A* **114**, 480-485.
- Walton, T. E., Alvarez, O., Jr., Buckwalter, R. M. & Johnson, K. M. (1973). *J Infect Dis* **128**, 271-282.
- Wang, E., Barrera, R., Boshell, J., Ferro, C., Freier, J. E., Navarro, J. C., Salas, R., Vasquez, C. & Weaver, S. C. (1999). *J Virol* **73**, 4266-4271.
- Wang, E., Bowen, R. A., Medina, G., Powers, A. M., Kang, W., Chandler, L. M., Shope, R. E., Weaver, S. C. & Cysticercosis Working Group in, P. (2001). *Am J Trop Med Hyg* **65**, 64-69.

- Wang, E., Paessler, S., Aguilar, P. V., Smith, D. R., Coffey, L. L., Kang, W., Pfeffer, M., Olson, J., Blair, P. J., Guevara, C., Estrada-Franco, J. & Weaver, S. C. (2005). *Am J Trop Med Hyg* **72**, 805-810.
- Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., Tuschl, T., Hartmann, G. & Patel, D. J. (2010). *Nat Struct Mol Biol* **17**, 781-787.
- Wang, Y. F., Sawicki, S. G. & Sawicki, D. L. (1991). *J Virol* **65**, 985-988.
- Wang, Y. F., Sawicki, S. G. & Sawicki, D. L. (1994). *J Virol* **68**, 6466-6475.
- Wang, Z. & Kiledjian, M. (2001). *Cell* **107**, 751-762.
- Wangchuk, S., Chinnawirotpisan, P., Dorji, T., Tobgay, T., Dorji, T., Yoon, I. K. & Fernandez, S. (2013). *Emerg Infect Dis* **19**, 1681-1684.
- Watts, D. M., Callahan, J., Rossi, C., Oberste, M. S., Roehrig, J. T., Wooster, M. T., Smith, J. F., Cropp, C. B., Gentrau, E. M., Karabatsos, N., Gubler, D. & Hayes, C. G. (1998). *Am J Trop Med Hyg* **58**, 35-40.
- Watts, D. M., Lavera, V., Callahan, J., Rossi, C., Oberste, M. S., Roehrig, J. T., Cropp, C. B., Karabatsos, N., Smith, J. F., Gubler, D. J., Wooster, M. T., Nelson, W. M. & Hayes, C. G. (1997). *Am J Trop Med Hyg* **56**, 661-667.
- Weaver, S. C. (2013). *Trends Microbiol* **21**, 360-363.
- Weaver, S. C. & Barrett, A. D. (2004). *Nat Rev Microbiol* **2**, 789-801.
- Weaver, S. C., Bellew, L. A. & Rico-Hesse, R. (1992). *Virology* **191**, 282-290.
- Weaver, S. C., Ferro, C., Barrera, R., Boshell, J. & Navarro, J. C. (2004). *Annu Rev Entomol* **49**, 141-174.
- Weaver, S. C. & Forrester, N. L. (2015). *Antiviral Res* **120**, 32-39.
- Weaver, S. C., Hagenbaugh, A., Bellew, L. A., Netesov, S. V., Volchkov, V. E., Chang, G. J., Clarke, D. K., Gousset, L., Scott, T. W., Trent, D. W. & et al. (1993). *Virology* **197**, 375-390.
- Weaver, S. C., Kang, W., Shirako, Y., Rumenapf, T., Strauss, E. G. & Strauss, J. H. (1997). *J Virol* **71**, 613-623.
- Weaver, S. C. & Lecuit, M. (2015). *N Engl J Med* **372**, 1231-1239.
- Weaver, S. C., Osorio, J. E., Livengood, J. A., Chen, R. & Stinchcomb, D. T. (2012). *Expert Rev Vaccines* **11**, 1087-1101.
- Weaver, S. C., Pfeffer, M., Marriott, K., Kang, W. & Kinney, R. M. (1999). *Am J Trop Med Hyg* **60**, 441-448.
- Weaver, S. C. & Reisen, W. K. (2010). *Antiviral Res* **85**, 328-345.
- Weaver, S. C., Salas, R., Rico-Hesse, R., Ludwig, G. V., Oberste, M. S., Boshell, J. & Tesh, R. B. (1996). *Lancet* **348**, 436-440.
- Wei, C., Gershowitz, A. & Moss, B. (1975a). *Nature* **257**, 251-253.

- Wei, C. M., Gershowitz, A. & Moss, B. (1975b). *Cell* **4**, 379-386.
- Weinbren, M. P. (1958). *Trans R Soc Trop Med Hyg* **52**, 258-259.
- Weiss, B., Nitschko, H., Ghattas, I., Wright, R. & Schlesinger, S. (1989). *J Virol* **63**, 5310-5318.
- Weissenhorn, W., Poudevigne, E., Effantin, G. & Bassereau, P. (2013). *Curr Opin Virol* **3**, 159-167.
- Welsby, I., Hutin, D., Gueydan, C., Kruys, V., Rongvaux, A. & Leo, O. (2014). *J Biol Chem* **289**, 26642-26657.
- Wengler, G. & Wengler, G. (1984). *Virology* **134**, 435-442.
- Wengler, G., Wengler, G. & Gross, H. S. (1979). *Nature* **282**, 754-756.
- Westaway, E. G., Brinton, M. A., Gaidamovich, S., Horzinek, M. C., Igarashi, A., Kaariainen, L., Lvov, D. K., Porterfield, J. S., Russell, P. K. & Trent, D. W. (1985). *Intervirology* **24**, 125-139.
- White, J. P. & Lloyd, R. E. (2012). *Trends Microbiol* **20**, 175-183.
- White, L. J., Wang, J. G., Davis, N. L. & Johnston, R. E. (2001). *J Virol* **75**, 3706-3718.
- Wilkins, C. & Gale, M., Jr. (2010). *Curr Opin Immunol* **22**, 41-47.
- Wilkinson, T. A., Tellinghuisen, T. L., Kuhn, R. J. & Post, C. B. (2005). *Biochemistry* **44**, 2800-2810.
- Willems, W. R., Kaluza, G., Boschek, C. B., Bauer, H., Hager, H., Schutz, H. J. & Feistner, H. (1979). *Science* **203**, 1127-1129.
- Williams, B. R. (2001). *Sci STKE* **2001**, re2.
- Winz, M. L., Cahova, H., Nubel, G., Frindert, J., Hofer, K. & Jaschke, A. (2017). *Nat Protoc* **12**, 122-149.
- Wu, D., Zhang, Y., Zhouhui, Q., Kou, J., Liang, W., Zhang, H., Monagin, C., Zhang, Q., Li, W., Zhong, H., He, J., Li, H., Cai, S., Ke, C. & Lin, J. (2013). *Virol J* **10**, 174.
- Xu, L. G., Wang, Y. Y., Han, K. J., Li, L. Y., Zhai, Z. & Shu, H. B. (2005). *Mol Cell* **19**, 727-740.
- Yanoviak, S. P., Aguilar, P. V., Lounibos, L. P. & Weaver, S. C. (2005). *J Med Entomol* **42**, 404-408.
- Yap, M. L., Klose, T., Urakami, A., Hasan, S. S., Akahata, W. & Rossmann, M. G. (2017). *Proc Natl Acad Sci U S A* **114**, 13703-13707.
- Yergolkar, P. N., Tandale, B. V., Arankalle, V. A., Sathe, P. S., Sudeep, A. B., Gandhe, S. S., Gokhle, M. D., Jacob, G. P., Hundekar, S. L. & Mishra, A. C. (2006). *Emerg Infect Dis* **12**, 1580-1583.
- Yondola, M. & Carter, C. (2011). *Viruses* **3**, 26-31.
- Yoneyama, M. & Fujita, T. (2010). *Rev Med Virol* **20**, 4-22.

- Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S. & Fujita, T. (2004). *Nat Immunol* **5**, 730-737.
- Young, N. A. & Johnson, K. M. (1969). *Am J Epidemiol* **89**, 286-307.
- Young, N. A., Johnson, K. M. & Gauld, L. W. (1969). *Am J Trop Med Hyg* **18**, 290-296.
- Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D. & Shatkin, A. J. (1997). *Proc Natl Acad Sci U S A* **94**, 12898-12903.
- Zaja, R., Mikoc, A., Barkauskaite, E. & Ahel, I. (2012). *Biomolecules* **3**, 1-17.
- Zapata-Perez, R., Gil-Ortiz, F., Martinez-Monino, A. B., Garcia-Saura, A. G., Juanhuix, J. & Sanchez-Ferrer, A. (2017). *Open Biol* **7**.
- Zayed, A., Awash, A. A., Esmail, M. A., Al-Mohamadi, H. A., Al-Salwai, M., Al-Jasari, A., Medhat, I., Morales-Betoulle, M. E. & Mnzava, A. (2012). *Acta Trop* **123**, 62-66.
- Zhang, R., Hryc, C. F., Cong, Y., Liu, X., Jakana, J., Gorchakov, R., Baker, M. L., Weaver, S. C. & Chiu, W. (2011). *EMBO J* **30**, 3854-3863.
- Zhang, W., Heil, M., Kuhn, R. J. & Baker, T. S. (2005). *Virology* **332**, 511-518.
- Zhang, W., Mukhopadhyay, S., Pletnev, S. V., Baker, T. S., Kuhn, R. J. & Rossmann, M. G. (2002). *J Virol* **76**, 11645-11658.
- Zhang, X., Fugere, M., Day, R. & Kielian, M. (2003). *J Virol* **77**, 2981-2989.
- Zhao, H., Lindqvist, B., Garoff, H., von Bonsdorff, C. H. & Liljestrom, P. (1994). *EMBO J* **13**, 4204-4211.
- Zheng, Y. & Kielian, M. (2013). *J Virol* **87**, 9579-9589.
- Zhu, P., Martinvalet, D., Chowdhury, D., Zhang, D., Schlesinger, A. & Lieberman, J. (2009). *Blood* **114**, 1205-1216.
- Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y. T. & Gao, G. (2011). *Proc Natl Acad Sci U S A* **108**, 15834-15839.
- Ziegler, M. & Oei, S. L. (2001). *Bioessays* **23**, 543-548.
- Zusinaite, E., Tints, K., Kiiver, K., Spuul, P., Karo-Astover, L., Merits, A. & Sarand, I. (2007). *J Gen Virol* **88**, 1977-1985.
- Zust, R., Cervantes-Barragan, L., Habjan, M., Maier, R., Neuman, B. W., Ziebuhr, J., Szretter, K. J., Baker, S. C., Barchet, W., Diamond, M. S., Siddell, S. G., Ludewig, B. & Thiel, V. (2011). *Nat Immunol* **12**, 137-143.

THIS PAGE IS INTENTIONALLY LEFT BLANK