The Roles of Template RNA and Replication Proteins in the Formation of Semliki Forest Virus Replication Spherules
The Roles of Template RNA and Replication Proteins in the Formation of Semliki Forest Virus Replication Spherules

Katri Kallio

Department of Food and Environmental Sciences
Division of Microbiology and Biotechnology
Faculty of Agriculture and Forestry

Department of Biosciences
Division of General Microbiology
Faculty of Biological and Environmental Sciences

Integrative Life Science Doctoral Program
University of Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Biological and Environmental Sciences, for public examination in the auditorium 1041 of Biocenter 2, Viikinkaari 5, Helsinki, on May 20th, 2016 at 12 o’clock noon

Helsinki 2016
Supervisor
Docent Tero Ahola
Department of Food and Environmental Sciences
University of Helsinki, Finland

Reviewers
Docent Johan Peränen
Department of Anatomy
University of Helsinki, Finland

Docent Petri Susi
Department of Virology
University of Turku, Finland

Thesis committee
Docent Kristiina Mäkinen
Department of Food and Environmental Sciences
University of Helsinki, Finland

Docent Varpu Marjomäki
Department of Biological and Environmental Sciences
University of Jyväskylä, Finland

Opponent
Associate Professor Gorben Pijlman
Laboratory of Virology
Wageningen University, The Netherlands

Custos
Professor Dennis Bamford
Department of Biosciences and
Institute of Biotechnology
University of Helsinki

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis

ISSN 2342-3161 (print)
ISSN 2342-137X (online)
ISBN 978-951-51-2081-6 (paperback)
ISBN 978-951-51-2082-3 (PDF)

Cover Image: Design by Katri Kallio

Hansaprint 2016
Now look at you wearing a crown
having achieved now all your dreams in just one go
What did I say, you got it made
Still pondering on the spiteful things of long ago?
No!

I said that you'd be so surprised
Seeing what is in for you
Now after all that has been done
You know what you can do
Listen up, come along

- Helloween
All positive-strand RNA viruses replicate their RNA genomes in close association with cellular membranes. A great variety of cellular membranes are utilized by different viruses and those membranes are extensively modified to support viral replication and to protect the viral RNA from host cell defense mechanisms.

Alphaviruses, including Semliki Forest virus (SFV), are positive-strand RNA viruses replicating their RNA on membranes derived from endosomal and lysosomal compartments. SFV induces small invaginations called spherules on plasma membrane and on endosomal membranes.

Viral replication complex assembly, spherule formation and initiation of replication are carefully orchestrated events and are guided by specific sequence elements within the genomic RNA as well as by important enzymatic activities of nonstructural proteins (nsPs). The aim of this research was to study in detail how alphavirus replication complexes are assembled and to define the minimum requirements for spherule formation by using a plasmid-derived trans-replication system mimicking SFV replication.

The role of the genomic RNA in replication was deciphered by using RNA templates, which were either modified or differed in length. Use of RNA templates differing in length clearly showed that they define the spherule diameter suggesting that the template has a significant role in spherule formation. By modifying or deleting specific sequences from the template it was shown that highly conserved RNA elements are important for SFV replication and do not tolerate modifications without compromising replication.

Study with the nsPs of SFV showed that the enzymatic activities essential for virus replication are also needed for spherule formation and that enzymes like helicase, protease and polymerase are absolutely essential for replication. Membrane association of the replication complex is also required to establish virus replication in the cells. The work with mutated nonstructural proteins and modified templates revealed a clear correlation between the minus-strand synthesis and spherule formation.

This work describes the alphavirus replication processes in detail and provides new principles, which may be generally applicable to study the positive-sense RNA virus replication and the formation of virus-induced membranous replication spherules.
# CONTENTS

Abstract ...................................................................................................................... III
List of publications ...................................................................................................... VI
Abbreviations ............................................................................................................. VII

Introduction .............................................................................................................. 1
World of viruses ........................................................................................................ 1
Alphaviruses and their replication cycle ................................................................. 2
Nonstructural proteins of Semliki Forest virus ....................................................... 7
  nsP1 .................................................................................................................. 7
  nsP2 .................................................................................................................. 9
  nsP3 ................................................................................................................ 11
  nsP4 ............................................................................................................... 12

Characteristics of SFV genome ............................................................................. 13
  5’ untranslated region .................................................................................... 13
  3’ untranslated region .................................................................................... 14
  Subgenomic promoter and other important sequences .................................... 15
Virus-induced membrane modifications ............................................................... 15
  Endoplasmic reticulum ................................................................................. 16
  Endosomal membranes ................................................................................. 17
  Peroxisomal membranes .............................................................................. 18
  Mitochondrial membranes ........................................................................... 18

Viral replication systems ....................................................................................... 18
  Replicons ...................................................................................................... 18
  Trans-replication systems ............................................................................ 20

Aims of the study .................................................................................................... 22

Materials and methods .......................................................................................... 23

Results .................................................................................................................... 26
  Characterizing the functions of the viral RNA genome (I, III) ....................... 26
    Replicase can replicate different sizes of RNA templates ......................... 26
    Size of the template determines the diameter of the spherule ................. 28
    Minus-strand RNAs cannot function as a template for the replicase ....... 29
    CSEs of template RNA are needed for efficient replication ................. 29
  
  Studying the functions of the replicase proteins (II) .................................... 31
  
  Minimum requirements for spherule formation (I, II, III) ......................... 32
  
  Recruitment of the RNA template by replicase (II) .................................... 33

Discussion .............................................................................................................. 35
Template RNA has a crucial role in spherule formation............................. 35
Conserved sequence elements ensure replication compatibility of the template.................................................................................................................. 38
Functional nonstructural proteins are the key players in spherule formation .............................................................................................................. 40
Spherules are the sites of replication ........................................................................... 41
Acknowledgements............................................................................................... 43
References................................................................................................................... 44
LIST OF PUBLICATIONS


In addition, some unpublished data is presented

The author’s contribution

I  Katri Kallio mostly designed the study and executed most of the experimental work, analysis and data interpretation, excluding the research done with linear constructs. K. K. has been involved in the writing process.

II  Katri Kallio mostly designed the study. Experimental work was executed together with Dr. Kirsi Hellström. Data analysis was done by the candidate, whereas Westerns were mainly done by Dr. Hellström. K. K. has been involved in the writing process.

III  Katri Kallio has mostly been involved in the design of the experiments and constructs used in the study. K. K. has been involved in some experiments and in the writing process.

Methodological contributions are listed in the materials and methods section.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdoMet</td>
<td>S-Adenosyl methionine</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>AUD</td>
<td>Alphavirus unique domain</td>
</tr>
<tr>
<td>BAR</td>
<td>Bin-Amphipysin-Rvsp</td>
</tr>
<tr>
<td>BFV</td>
<td>Barmah Forest virus</td>
</tr>
<tr>
<td>BHK</td>
<td>Baby hamster kidney cell</td>
</tr>
<tr>
<td>BMV</td>
<td>Brome mosaic virus</td>
</tr>
<tr>
<td>BSR</td>
<td>BHK cells expressing T7 RNA polymerase</td>
</tr>
<tr>
<td>CAAX</td>
<td>Prenylation target sequence (cys-ala-ala-X)</td>
</tr>
<tr>
<td>CHIKV</td>
<td>Chikungunya virus</td>
</tr>
<tr>
<td>CLEM</td>
<td>Correlative light and electron microscopy</td>
</tr>
<tr>
<td>CPV</td>
<td>Cytopathic vacuole</td>
</tr>
<tr>
<td>CSE</td>
<td>Conserved sequence element</td>
</tr>
<tr>
<td>DC-SIGN</td>
<td>Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin</td>
</tr>
<tr>
<td>DMV</td>
<td>Double-membrane vesicle</td>
</tr>
<tr>
<td>dsRNA</td>
<td>Double-stranded RNA</td>
</tr>
<tr>
<td>EEEV</td>
<td>Eastern equine encephalitis virus</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>FHV</td>
<td>Flock House virus</td>
</tr>
<tr>
<td>G3BP</td>
<td>Ras-GAP SH3 domain-binding protein</td>
</tr>
<tr>
<td>GT</td>
<td>Guanylyltransferase</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>HCV</td>
<td>Hepatitis C virus</td>
</tr>
<tr>
<td>HVR</td>
<td>Hypervariable region</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>L-SIGN</td>
<td>Liver/lymph node-specific intracellular adhesion molecules-3 grabbing non-integrin</td>
</tr>
<tr>
<td>m7GMP</td>
<td>7-methylguanylate cap</td>
</tr>
<tr>
<td>MAYR</td>
<td>Mayaro virus</td>
</tr>
<tr>
<td>MT</td>
<td>Methyltransferase</td>
</tr>
<tr>
<td>mTOR</td>
<td>Mechanistic target of rapamycin</td>
</tr>
<tr>
<td>nsP</td>
<td>Nonstructural protein</td>
</tr>
<tr>
<td>NTF2</td>
<td>N-terminal nuclear transport factor 2</td>
</tr>
<tr>
<td>ONNV</td>
<td>O'nyong'nyong virus</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>Pa-</td>
<td>Non-palmitoylated</td>
</tr>
<tr>
<td>PC</td>
<td>Phosphatidylcholine</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol 3-kinase</td>
</tr>
<tr>
<td>Poly(A)</td>
<td>Polyadenylate</td>
</tr>
<tr>
<td>PV</td>
<td>Poliovirus</td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>RLuc</td>
<td>Renilla Luciferase</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>RRV</td>
<td>Ross River virus</td>
</tr>
<tr>
<td>SARS</td>
<td>Severe acute respiratory syndrome</td>
</tr>
<tr>
<td>SFV</td>
<td>Semliki Forest virus</td>
</tr>
<tr>
<td>SG</td>
<td>Stress granule</td>
</tr>
<tr>
<td>SH3</td>
<td>SRC Homology 3</td>
</tr>
<tr>
<td>SINV</td>
<td>Sindbis virus</td>
</tr>
<tr>
<td>SL</td>
<td>Stem loop</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission electron microscopy</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>VEEV</td>
<td>Venezuelan equine encephalitis virus</td>
</tr>
<tr>
<td>WEEV</td>
<td>Western equine encephalitis virus</td>
</tr>
<tr>
<td>WNV</td>
<td>West Nile virus</td>
</tr>
</tbody>
</table>
INTRODUCTION

WORLD OF VIRUSES

Despite the fact that viruses are the smallest self-replicating organisms on the planet they have had a great impact on the history and evolution of life. Virtually all living organisms are infected by different viruses and they have shaped the history of mankind via viral diseases within the human population and in our environment. Viruses are small particles typically made of proteins, genetic material and sometimes lipid layer. The field of virology has evolved quickly and is a relatively new branch of science (Enquist and Racaniello 2013).

In the 19th century microbiology was taking its first steps and the idea of a "microbial world" including bacteria, fungi and protozoa was eventually accepted. Jacob Henle, Louis Pasteur, Robert Koch and Joseph Lister were the pioneers in the field of microbiology and led the way by demonstrating that microorganisms can act as disease causing agents (Enquist and Racaniello 2013).

The first scientist encountering a virus was Adolf Mayer (1843-1942) when studying a tobacco plant disease, which he called tobacco mosaic disease. At that time he was not aware that the disease was caused by a virus. He stated that the causative agent was bacterial, but the infectious agent was not identified. The next step was taken by Dimitri Ivanofsky (1864-1920) who showed that the extract from infected leaves was infectious after microbial filtration process. Filtering was at that time used to remove bacteria from solutions. Even if Ivanofsky was unable to culture any organism from the filtered extract, he stated that the causative agent is of bacterial origin. The work was continued by Martinus Beijerinck (1851-1931), a collaborator of Adolf Mayer, who discovered that the filtered extract can be diluted and when reintroduced to a living plants was able to restore its “strength”. This finding was a remarkable leap forward in virus research by showing that the causative agent of tobacco mosaic disease can reproduce within living tissue. Beijerinck called this agent a *contagium vivum fluidium* or contagious living liquid. Finally, in the year 1917, Félix d'Herelle developed the nowadays commonly used technique called plaque assay, which led to identification of the causative agent for tobacco mosaic disease, as tobacco mosaic virus. The term virus came from Latin word for slimy liquid or poison and was later on restricted to be used only with agents fulfilling the criteria defined by Mayer, Ivanofsky and Beijerinck. This was the starting point for the field of virology and new viruses were quickly discovered and identified (Enquist and Racaniello 2013).

Every year, many new viruses and bacteriophages are discovered, and one can only speculate how many viruses are still waiting to be discovered. Isolated viruses have been studied in detail and many viruses have been found to share similar characteristics.
with each other. Based on certain similarities viruses have been classified into different groups. The Baltimore classification, named after its creator David Baltimore, is one of the most used classification schemes. The Baltimore classification is based on the nature of the viral genome (RNA or DNA), number of strands (double-stranded or single-stranded), polarity of the genome (positive, negative or ambisense), and transcriptase enzyme. This type of classification divides viruses into seven groups.

The virus used in the thesis belongs to the group IV in the Baltimore classification. Viruses of this group carry single-stranded RNA genomes with positive polarity. The plus-strand RNA ([+]RNA) viruses comprise a very large group of viruses including many human pathogens. Epidemic diseases caused by these viruses include e.g. dengue fever, poliomyelitis, and hepatitis C. Single stranded (+)RNA viruses are the largest group of RNA viruses comprising 30 virus families. Seven out of these 30 families include human pathogens like poliovirus, norovirus, hepatitis C virus and dengue virus (Strauss and Strauss 2008)

Nowadays the level of knowledge and the techniques used in the field of virology are on completely different level when compared to the era of Beijerinck and colleagues but there are still many viruses for which we do not have efficient vaccines or antivirals. Viruses have also proven to be valuable tools in scientific research as well in the field of medicine.

To study alphavirus life cycle and infectivity, two well-known members of the family have been extensively used: Sindbis virus (SINV) and Semliki Forest virus (SFV). SINV is a human pathogen causing only mild symptoms whereas SFV infects mostly rodents (Atkins 2013). SFV is easy to handle in the laboratory environment and can infect a great variety of different cell lines, thus being a good model virus for alphavirus research (Atkins 2013)

**ALPHAVIRUSES AND THEIR REPLICATION CYCLE**

The genus *Alphavirus* belongs to the *Togaviridae* family with another genus called *Rubivirus*. The *alphavirus* genus comprises a diverse group of viruses infecting a wide variety of hosts (Atkins 2013) whereas the genus *Rubivirus* has only one member, rubella virus (Powers et al. 2001). As a consequence of broad distribution all over the world and the capability of infecting a wide range of hosts varying from fishes to humans and other vertebrates, alphaviruses have become a great economic and public health concern.

**Alphaviruses** are arthropod-borne viruses (arboviruses) transmitted by hematophagous arthropods such as mosquitoes, ticks, midges, and sandflies (Pfeffer and Dobler 2010). In arthropod-vectors like mosquitoes, they cause a persistent asymptomatic lifelong infection and accumulate in salivary glands from which they are released to the vertebrate host during a blood meal (Strauss and Strauss 1994). **Alphaviruses** have been divided into two classes called the Old World viruses and the New World viruses based on their geographical distribution (Strauss
and Strauss 1994). Epidemics caused by the Old World viruses are typically sporadic but extremely furious like the one caused by chikungunya virus (CHIKV) in the Réunion Island 2005-2006 causing approximately 265,000 clinical cases and 237 deaths (Weaver and Forrester 2015).

The old world alphaviruses including CHIKV, Sindbis virus (SINV), Ross River virus (RRV) and O’nyong’nyong virus (ONNV) are frequently causing epidemics in humans, and the typical symptoms are fever, skin rash, malaise, myalgia, and severe arthralgia (Pfeffer and Dobler 2010, Atkins 2013). In rare cases severe symptoms arise, including debilitating joint pain and hemorrhage (CHIKV), eye and chest pain (ONNV) or splenomegaly and hematuria (RRV). The New World viruses like Eastern equine encephalitis virus (EEEV), Venezuelan equine encephalitis virus (VEEV) and Western equine encephalitis (WEEV) can cause fatal encephalitis in humans (Strauss and Strauss 1994).

Alphavirus particle is made of a spherical capsid with the viral genomic RNA enclosed within. The capsid is surrounded by a host-derived lipid envelope in which the viral glycoprotein spikes are embedded. The capsid is composed of 240 copies of capsid protein, which are arranged in a T=4 lattice to form an icosahedral shell (Strauss and Strauss 1994, Mancini et al. 2000). The spike proteins follow the same lattice on the surface of the enveloped virus particle (Strauss and Strauss 1994). The capsid proteins and the envelope spike proteins are thus forming two protein layers and the host cell derived membrane lies between them (Mancini et al. 2000).

Virus entry to a susceptible cell takes place via receptor mediated endocytosis through clathrin-coated pits (Klimstra et al. 2003). Several different cell surface molecules have been proposed to act as receptors for alphaviruses including laminin receptors, the class I major histocompatibility antigen, α1β1 integrin, and C-type lectins like DC-SIGN (dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin)/ L-SIGN (liver/lymph node-specific intracellular adhesion molecules-3 grabbing non-integrin) (Kielian et al. 2010). It is still unclear what mechanism alphaviruses use to infect both insects and vertebrates but two hypotheses exist: there is either a common receptor present in all cell types or viruses can utilize several different receptors.

Two surface proteins of SFV, called E1 and E2 mediate the internalization of the virus into the cell. These proteins form E1/E2 heterodimers, which are further arranged into 80 trimers. These (E1/E2)₃ complexes are forming an icosahedral protein scaffold around the capsid with triangulation of T=4 (Roussel et al. 2006). Binding of the virus is primarily mediated by E2 protein, whereas E1 is the membrane fusion protein. E2 is closely associated to E1 and is masking most of the E1 counterpart on the virus particle, including the membrane interacting fusion loop (Kielian et al. 2010).
Table 1 Medically important mosquito-borne alphaviruses (Atkins 2013, Zuchi et al. 2014)

<table>
<thead>
<tr>
<th>Virus</th>
<th>Abbreviation</th>
<th>Reservoir host</th>
<th>Human disease</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern equine encephalitis</td>
<td>EEEV</td>
<td>Birds</td>
<td>Encephalitis, high fever, muscle pain, headache, photophobia, seizures</td>
<td>North, Central and South America and the Caribbean</td>
</tr>
<tr>
<td>Venezuelan equine encephalitis</td>
<td>VEEV</td>
<td>Rodents, equines</td>
<td>Encephalitis, high fever, headache</td>
<td>Central and South America</td>
</tr>
<tr>
<td>Western equine encephalitis</td>
<td>WEEV</td>
<td>Birds</td>
<td>Encephalitis</td>
<td>Central and South America</td>
</tr>
<tr>
<td>Chikungunya</td>
<td>CHIKV</td>
<td>Primates</td>
<td>Rash, fever, malaise, arthralgia, myalgia</td>
<td>America, Africa, Asia, Europe, Indian and Pacific Oceans</td>
</tr>
<tr>
<td>O’nyong’nyong</td>
<td>ONNV</td>
<td>Unknown</td>
<td>Rash, fever, arthritis, eye pain, chest pain, lethargy</td>
<td>Central Africa (e.g. Uganda, Kenya, Tanzania, Zaire)</td>
</tr>
<tr>
<td>Ross River</td>
<td>RRV</td>
<td>Marsupials</td>
<td>Rash, arthritis, fever</td>
<td>Australia, Fiji, and other islands in the South Pacific</td>
</tr>
<tr>
<td>Semliki Forest</td>
<td>SFV</td>
<td>Birds</td>
<td>Rash, fever, headache, myalgia, arthralgia</td>
<td>Africa</td>
</tr>
<tr>
<td>Sindbis</td>
<td>SINV</td>
<td>Birds</td>
<td>Rash, arthritis, fever</td>
<td>Africa, Egypt, Israel, Philippines and Australia, Northern Europe</td>
</tr>
<tr>
<td>Mayaro</td>
<td>MAYV</td>
<td>Primates</td>
<td>Rash, fever, headache, myalgia, joint pain</td>
<td>South America</td>
</tr>
<tr>
<td>Barmah Forest</td>
<td>BFV</td>
<td>Marsupials</td>
<td>Rash, fever, malaise, arthralgia, myalgia</td>
<td>Australia</td>
</tr>
</tbody>
</table>
After endocytosis the pH of the virus-containing endosome decreases, which causes the dissociation of E1/E2 dimer exposing the fusion loop of E1 (Kielian et al. 1986, Kielian et al. 2010). The decrease in the pH induces conformational changes in the E1 protein leading to the fusion of viral membrane with the endosomal membrane and release of the capsid to the cytoplasm. (Kielian and Helenius 1985, Gibbons et al. 2004). Recently, it was shown that a host factor TSPAN9, a member of the tetraspanin family of membrane proteins enhances SFV fusion with endosomal membranes (Stiles and Kielian 2016).

The mechanism for the capsid disassembly remains unknown and it is still under debate whether it happens simultaneously with the release (due to
the low pH) or with the help of cellular ribosomes (Wengler 2009). After the disassembly of the capsid, the viral genome is released to the cytoplasm and is directly used for viral protein translation by the cellular translation machinery (Jose et al. 2009).

The viral genome is a single-stranded RNA molecule, approximately 11.5 kb in length, with positive polarity. The RNA is equipped with a 5′ cap structure and a 3′ polyadenylic acid tail (poly[A] tail). Viral untranslated regions (UTRs), involved in the replication and translation of the genomic RNA, are located at the 5′- and 3′-ends of the genome (Liu et al. 2009, Hyde et al. 2015). The coding region of the viral genome is divided into two open reading frames (ORFs). The first ORF is used to produce nonstructural proteins (nsPs), needed for RNA replication, and the second ORF is used to produce the structural proteins (Kaariainen and Ahola 2002). The structural proteins are produced via a subgenomic RNA, which is synthesized only later in the infection (Rupp et al. 2015) (Fig. 1).

The viral replicase proteins are produced as a polyprotein precursor P1234, which is autocatalytically cleaved to individual nsPs by protease enzyme located in the nsP2 region of P1234 (Keranen and Ruohonen 1983, Kaariainen et al. 1987, Lemm and Rice 1993). First cleavage happens in cis and takes place between P123 and nsP4 to release the catalytic subunit of the viral RNA-dependent RNA polymerase (RdRp) (Shirako and Strauss 1994, Vasiljeva et al. 2003). nsP4 together with polyprotein P123 forms the so-called “early polymerase” which is capable of initiating the complementary (minus) strand synthesis (Shirako and Strauss 1994, Vasiljeva et al. 2003). How and where the nsPs recruit the genomic RNA is still unknown but to initiate the minus-strand synthesis nsP4 needs to be cleaved and the complex needs to be membrane-associated (Ahola et al. 1999, Spuul et al. 2007). Next cleavage dissociates the nsP1 creating a complex nsP1/P23/nsP4, which is very short-lived in virus replication but already starts the plus-strand synthesis by using the complementary minus-strand as a template (Vasiljeva et al. 2003). In the final stage all nsPs (nsP1 to nsP4) are cleaved to form a functional replication complex with the viral genome (Vasiljeva et al. 2003) (Fig. 1).

Structural proteins are also translated as a polyprotein precursor, which is proteolytically cleaved to yield five individual structural proteins (C, E3, E2, 6K, and E1) (Roussel et al. 2006). Spike proteins (E1 and E2) are transported to the plasma membrane through the endoplasmic reticulum (ER) and Golgi complex where they are post-translationally modified (Jose et al. 2009). The genomic RNA is packaged into the newly-assembled capsids and the virus budding takes place at the plasma membrane (Jose et al. 2009). Glycoprotein spikes embedded in cell-derived membrane are formed of E1/E2 heterodimers arranged into trimers (Venien-Bryan and Fuller 1994, Cheng et al. 1995).
SFV has four nonstructural proteins (nsPs), which are responsible for viral RNA replication. All nsPs have their own specific role in the replication of viral genomic RNA. nsPs interact with each other as well as with cellular membranes and host factors. The replicase proteins are translated directly from the positive-sense RNA genome and thus SFV does not need to carry any additional proteins e.g. RNA polymerase within the capsid. During virus replication the nsPs are involved in the formation of replication complexes. Additionally after cleavage and accumulation in the cytoplasm they are directed to specific locations within the cell. nsP1 is directed to the cellular membranes (Ahola et al. 1999, Kujala et al. 2001), nsP2 is transported to the nucleus (Peranen et al. 1990), nsP3 is gathered into cytoplasmic stress granules (Panas et al. 2012) and nsP4 is quickly degraded (Lemm et al. 1994).

**nsP1**

nsP1 has two main functions in RNA replication; it is responsible for the capping of the nascent RNA molecules and it serves as the membrane anchor for the replication complex. SFV has a unique way of capping the viral RNA and the reactions differ from the normal cellular capping reactions, which take place in the nucleus. Capping of the cellular mRNAs usually takes place in the following order: the terminal phosphate is removed from the nascent RNA, leaving a bisphosphate group at the 5’ end of the RNA (5’(ppN)[pN]n). GTP is added to the nascent RNA in a reaction releasing a pyrophosphate from the GTP, thus GMP is covalently attached to the mRNA via 5’-5’ triphosphate linkage and the 7-nitrogen of the GMP is methylated (Furuichi and Shatkin 2000). In short, GMP is first conjugated to the nascent RNA and then subsequently methylated.

This type of cap structure is called 7-methylguanylate cap (m7G) or cap-0. Even though the cap structure of the alphavirus RNA is of the cap-0 type, the capping reactions happen in different order. nsP1 binds GTP and a methyl donor called S-adenosylmethionine (AdoMet) and removes two terminal phosphates from the guanine. The 7-nitrogen of GMP is methylated (methyl group from AdoMet) and the nucleotide is bound covalently to nsP1 forming m7GMP-nsP1 complex. The m7GMP is then transferred to the 5’ end of the nascent RNA and bound to it via 5’-5’-linkage (Ahola and Kaariainen 1995, Ahola et al. 1997). Thus, in contrast to cellular capping alphaviruses first methylate the GMP, which is then subsequently transferred to the nascent RNA.

The capping reactions are catalyzed by two enzymatic activities, the methyltransferase (MT) and guanylyltransferase (GT) both located in the N-terminus of nsP1 (Rozanov et al. 1992) (Fig. 2). Both enzymatic activities lie within the conserved region of nsP1 and two amino acids were shown to be crucial for the capping activity (Ahola and Karlin 2015). These two amino acids were identified as histidine-38 (H38) and aspartate-64.
(D64). H38 is supposedly involved in m7GMP binding and D64 in AdoMet binding (Ahola et al. 1997). Ahola et al. (1997) also showed that the capping reactions of SFV are strictly orchestrated and the methylation of GMP needs to proceed via the covalent binding of GTP to nsP1. It has also been shown that the GMP needs to be methylated when transferred to the nascent RNA, whereas nsP1 is incapable of methylating the cap analog GpppA or methylated RNAs (Laakkonen et al. 1994). It was recently shown that the capping reactions of VEEV follow the same scheme observed with SFV (Li et al. 2015).

The other main function of nsP1 is to bind the replication complex to the cellular membranes. Based on amino acid sequence data nsP1 does not include any hydrophobic regions, which could be interpreted as transmembrane regions and yet it is tightly bound to cellular membranes (Takkinen 1986, Peranen et al. 1995). Membrane association is critical for the activity of nsP1 and is accomplished by two different mechanisms: via a binding peptide located in the central part of nsP1 and palmitoylation of specific residues at the C-terminus (Laakkonen et al. 1996, Ahola et al. 1999, Lampio et al. 2000).

Extensive research on membrane targeting and binding has revealed that the binding peptide consists residues G_{246}STLYTESRKLLRSWHLPSV_{264} and forms an α-helical secondary structure when associated with negatively charged membranes (Ahola et al. 1999, Lampio et al. 2000). In the α-helical form the binding peptide shows amphipathic nature in which the polar and hydrophobic residues are placed on the opposite sides of the helix (Lampio et al. 2000). Lampio et al. (2000) showed that the positively charged residues of the binding peptide form polar interactions with the head groups of anionic phospholipids, and tryptophan-259 (W259) was especially shown to be critical for the membrane association. There is some evidence that the binding peptide region might be involved in the targeting of the replication complex to plasma membrane but this function still remains uncertain (Spuul et al. 2007).

Membrane binding of the replication complex is strengthened by covalent palmitoylation of specific cysteine residues (C418-420) but it is not essential for membrane association (Laakkonen et al. 1996, Zusinaite et al. 2007). Palmitoylation negative (Pa-) nsP1 is targeted to the membranes and is enzymatically active indicating that the conformation is not compromised, although Pa- nsP1 fails to introduce filopodia on the cellular membranes (Ahola et al. 2000). Spuul (Spuul 2010) suggested in her PhD thesis that the electrostatic interactions of the binding peptide first attach nsP1 to plasma membrane, which is followed by palmitoylation of the protein. Palmitoylation then launches conformational change in nsP1 allowing the insertion of the hydrophobic residues into the lipid bilayer. Interactions with lipids promote the formation of the amphipathic helix, which finalizes the strong binding of replication complex onto membranes.
nsP2

nsP2 is a multifunctional protein and has two separate domains called N-terminal helicase domain and C-terminal protease domain (Gomez de Cedron et al. 1999) (Fig. 2). Helicases are motor proteins capable of unwinding double-stranded nucleic acids like dsRNA. Helicase activity is powered by hydrolysis of NTPs or dNTPs by closely associated NTPase (Wu 2012). Helicases are classified into six superfamilies and alphavirus nsP2 helicase belongs to the 1st superfamily (Das et al. 2014).

The N-terminal helicase domain encompasses also RNA triphosphatase activity (Vasiljeva et al. 2000) and NTPase activity (Rikkonen et al. 1994). The RNA triphosphatase is involved in the capping of the viral RNA together with nsP1 by breaking the bond between the $\gamma$ and $\beta$ phosphates of the first nucleotide of the viral RNA prior to the transfer of the $\text{m}^7\text{GMP}$ to the 5`end.

Fig. 2. Schematic view of SFV nsPs. Regions and amino acids important for membrane binding and enzymatic activities necessary for genome replication are shown. Numbers indicate the corresponding amino acid residue in the current nsP (nsP1 to nsP4). Some of the highly conserved amino acids are also underlined or marked with arrows. MT/GT, methyltransferase/guanylyltransferase; SH3, SRC Homology 3; HVR, hypervariable region, G3BP, Ras-GAP SH3 domain–binding protein.
Recent data with CHIKV shows that helicases unwind the dsRNA from 5' to 3' direction and are also capable of rewinding RNA (Das et al. 2014). It was also shown with CHIKV that during genome replication the helicase domain of nsP2 is closely interacting with nsP4, which is the core subunit of the viral RdRp (Stapleford et al. 2015). Stapleford et al. (2015) also showed that multifunctional nsP2 has yet another important role in maintaining the alphavirus replication complex fidelity together with nsP4.

The N-terminus of nsP2 is highly conserved among Alphaviruses and is homologous to many NTPases and helicases (Rikkonen et al. 1994). Amino acid sequence comparisons of NTPase/helicase enzymes have revealed many conserved motifs associated with NTP binding including Walker motif A, which can be found also in SFV nsP2 (Walker et al. 1982, Takkinen 1986). Classical Walker motif A forms a phosphate binding loop and is shown to encompass amino acids GXXXXGKS/T (X marks any amino acid), which completely correlates with the motif found in nsP2 (GVPGSGK<sub>192</sub>S) (Takkinen 1986). Mutation of the highly conserved Lysine-192 (K<sub>192</sub>) residue completely abolishes NTPase activity, RNA triphosphatase activity and helicase activity of nsP2 (Rikkonen et al. 1994, Gomez de Cedron et al. 1999) suggesting that the reaction centers of these activities are overlapping or slightly different but closely connected (Vasiljeva et al. 2000). The native N-terminus of nsP2 was also shown to be important for the function of the NTPase as well as for the helicase (Das et al. 2014).

The viral protease was located to the C-terminus of nsP2 first with SINV and later it was shown to be the case also with SFV. The C-terminal domain forms a papain-like protease, which is solely responsible for the autocatalytic processing of the polyprotein precursor P1234 (Merits et al. 2001, Vasiljeva et al. 2001, Vasiljeva et al. 2003).

Polyprotein processing is a carefully orchestrated cascade of proteolytic cleavages, which are timing the viral RNA replication. nsP4 is the first cleaved protein from the polyprotein and the first mature nsP occurring in the cytoplasm during SFV infection (Takkinen et al. 1991). The complex P123 + nsP4 is known to carry out the minus-strand synthesis and, thus the subsequent processing of P123 determines the timing of the plus-strand genomic RNA replication and subgenomic RNA synthesis (Shirako and Strauss 1994, Merits et al. 2001). nsP2 protease seems to favor slightly different substrates when introduced in different orientations (P12, P123, P23 or P2) (Shirako and Strauss 1990). Later in the infection the concentration of cleaved nsP2 is elevated leading to different cleavage pattern in which the cleavage of P123 is too fast to allow a formation of new replication complexes (Merits et al. 2001). This leads to the decrease of minus-strand synthesis, whereas the synthesis of genomic and subgenomic RNAs continues at constant level.

The C-terminus of nsP2 contains a nuclear localization signal and a great deal of nsP2 is transported to the nucleus during viral infection (Peranen et al. 1990, Rikkonen et al. 1992). It has been shown that SFV nsP2 has the ability to quench the expression of type I interferons (IFN) and thus suppress
the innate immune defense reactions (Breakwell et al. 2007). Type I IFNs are present in most cell types and are the key players in the initiation of innate immune defenses during virus infection through recognition of viral replication intermediates, such as dsRNA (Kato et al. 2005, Kato et al. 2006). It has also been shown with SINV that nsP2 is fully capable of suppressing the type I IFN response during infection (Frolova et al. 2002). Suppressing the innate immunity reactions has been shown to be a common mechanism between viruses, and many viral proteins have been shown to act as antagonists of these reactions (Garcia-Sastre et al. 1998, Ma and Damania 2016).

nsP3

nsP3 can be divided into three different domains, N-terminal macro domain, alphavirus unique domain (AUD) and C-terminal hypervariable region (HVR) (Fig. 2). Of these only the N-terminal structurally conserved macro domain shares sequence homology with some other viruses (LaStarza et al. 1994). The lack of sequence homology is quite extraordinary when compared to the other nsPs, which are highly conserved within RNA virus families (Haseloff et al. 1984, Ahlquist et al. 1985). This may be one of the reasons why nsP3 is still the least understood protein of the SFV replication complex.

The macro domain shows affinity for ADP-ribose derivatives, RNA molecules and it has been shown to be capable of hydrolyzing ADP-ribose-1′ phosphate (Malet et al. 2009, Neuvonen and Ahola 2009). Macro domain could in theory recruit poly(ADPribosyl)ated cellular factors to the replication sites but the importance of this function still remains unknown (Egloff et al. 2006, Neuvonen and Ahola 2009). Although the role of the macro domain is still partially unknown it has been proven to be important for the virus replication. (Park and Griffin 2009).

nsP3 seems to be involved in the virus-host interactions through the non-conserved C-terminal HVR. The size of HVR varies from 150 to 250 amino acid residues between different alphaviruses (Neuvonen et al. 2011). nsP3 is heavily phosphorylated on specific serine and threonine residues and is thus the only phosphoprotein in the replication complex (Peranen et al. 1988, Vihinen et al. 2001). All phosphorylated residues lie within a short stretch of 50 amino-acids (319-368) from the beginning of the HVR.

Phosphorylation is important for the internalization of replication complexes from the plasma membrane as well as for the neuropathogenicity of the virus (Vihinen et al. 2001, Thaa et al. 2015). Recent data from Thaa et al. (2015) showed that nsP3 has a major role in the viral-host interplay via activating cellular signaling pathways and more precisely so called “prosurvival” pathways like Akt-mechanistic target of rapamycin (mTOR) pathway while infecting mammalian cells. It has clearly been shown that nsP3 is capable of activating this signaling pathway only when phosphorylated and in association with cellular membranes (Thaa et al. 2015). It was also shown earlier by Spuul et al. (2010) that the activity of phosphatidylinositol 3-kinase (PI3K) is needed for the internalization of replication complexes, which completely supports the new study; PI3K and Akt-mTOR are both involved in the Akt-mTOR -
pathway where PI3K is the activator (Spuul et al. 2010, Thaa et al. 2015).

HVR contains a highly conserved cluster of proline residues, which was identified as Src homology-3 (SH3) binding motif (Neuvonen et al. 2011). By carrying this SH3 binding motif nsP3 could act as ligand for cellular SH3 domain-containing proteins (Neuvonen et al. 2011). SH3 domains are small, globular protein motifs that specifically bind to proline rich sites of their ligands e.g. proteins involved in cell signaling, membrane trafficking, and cytoskeletal organization (Mayer 2001).

Via this proline rich motif (PIPPPR) nsP3 interacts with amphiphysin-1 and Bin1/amphiphysin-2 and recruits them to the replication sites (Neuvonen et al. 2011). Neuvonen et al. (2011) also showed that mutations affecting the SH3 binding motifs abolished the interaction with amphiphysins leading to the reduced virus replication in cell culture and reduced pathogenicity in mice. It was also shown that SH3 domain of SFV, SINV, and CHIKV nsP3 binds *Drosophila melanogaster* amphiphysin, which is almost identical to the amphiphysins expressed in insect vector e.g. SFV vector *Aedes aegypti*. This observation clearly suggests that nsP3-host interactions may be important also during virus replication in their mosquito vector (Neuvonen et al. 2011).

Amphiphysins are membrane binding proteins that can sense and induce steep positive curvature to the membranes via BAR (Bin/Amphiphysin/Rvsp) domain (Dawson et al. 2006). Although, the spherule body poses a negative curvature the neck structure of a spherule presents a positive curvature that is compatible with the binding characteristics of BAR domains, which is why it is tempting to speculate on its role also in spherule formation (Neuvonen et al. 2011). All amphiphysins have an N-terminal BAR domain and a C-terminal SH3 domain (Dawson et al. 2006). The SH3 binding motif is dispensable for the virus in vitro but results in reduced pathogenicity in mice (Neuvonen et al. 2011). Similar results were obtained with SINV showing that mutations on the macro domain greatly reduced the ability of the virus to infect neurons and/or maintain the infection within central nervous system (Park and Griffin 2009).

The C-terminal domain also carries an L/ITFGDFD motif, which is responsible for binding the Ras-GAP SH3 domain-binding proteins (G3BP). The viral L/ITFGDFD motif is shown to bind to the G3BP N-terminal nuclear transport factor 2 (NTF2)-like domain (Panas et al. 2012, Vognsen et al. 2013). Binding of G3BP suppresses stress granule (SG) formation within infected cells and thus reduces the antiviral effect of SGs (Panas et al. 2015). SGs are aggregates of proteins and translationally silenced mRNA induced upon stress conditions such as virus infection (Panas et al. 2015).

nsP4

nsP4 has been identified as the catalytic core of the RNA-dependent RNA polymerase and it possesses a highly conserved GDD motif close to the C-terminus (Kamer and Argos 1984) (Fig. 2). nsP4 is short lived in cytoplasm compared to other nsPs and is quickly degraded by the N-end rule pathway
It was shown by Takkinen et al. (1991) that only ~20% of the expected amount of P4 was present in the cells. The destabilizing tyrosine (Tyr) residue at the N-terminus of native nsP4 quickly targets the protein to the N-end rule pathway. The N-terminal Tyr residue is crucial for protein function and cannot be changed to any other residue without affecting virus replication (de Groot et al. 1991). The N-terminus of nsP4 is also important for protein-protein interactions in the replication complex, and it has been shown to interact particularly with nsP1 (Fata et al. 2002). Terminal adenylyltransferase activity suggests that nsP4 is also involved in the maintenance and repair of the viral poly(A) tail by catalyzing the addition of adenine to the 3' end of the nascent RNA (Tomar et al. 2006).

The production of nsP4 does not always occur in the same ratio with the other nsPs. This is due to the opal codon, which resides in front of the nsP4 gene in the genome of some alphaviruses. For example SINV has the opal codon, which reduces the production of nsP4 by ~80% compared to other nsPs (Li and Rice 1993, Strauss and Strauss 1994). In contrast, SFV does not have the opal codon and nsP4 is always translated as part of the polyprotein P1234 (Kaariainen and Ahola 2002).

**CHARACTERISTICS OF SFV GENOME**

The genome of SFV is a single-stranded positive-sense RNA molecule approximately 11.5 kb in length. The genome contains several important regions, which are: 1) protecting the RNA from degradation, 2) making the RNA to resemble cellular mRNAs and, 3) making it replication-competent (Furuichi and Shatkin 2000, Frolov et al. 2001). Important elements in the viral genome for protein translation are the 5' cap-0 structure and the 3' poly(A) tail (Strauss and Strauss 1994). These elements make the viral genome applicable for protein translation by cellular ribosomes after the genome has been released into the cytoplasm. Cap structure and the poly(A) tail also greatly enhance the stability of the mRNA in the cellular environment (Furuichi and Shatkin 2000). In addition to these structures, the viral genome includes virus-specific conserved sequence elements (CSEs) which SFV has four (CSE1 to CSE4). CSEs are positioned throughout the SFV genome to ensure the recognition of genomic RNA by replication proteins and to yield maximal replication of the genomic RNA and transcription of subgenomic RNA (Fig. 3).

**5' untranslated region**

Alphavirus 5' untranslated region (5'UTR) varies in length from 27 to 85 nucleotides between different species. SFV carries the longest 5'UTR in the alphavirus family reaching to 85 nucleotides (Hyde et al. 2015). The first CSE (CSE1) lies within the 5'UTR comprising approximately the first 44 nucleotides of the genome. The second one, CSE2, is 51 nucleotides in size and is located in the coding region of nsP1 around nucleotide 155 (Strauss and Strauss 1994, Jrome et al. 2009). Specific secondary stem loop structures are formed within the CSEs called stem loop 1 to 4 (SL1 to SL4). SL1 is formed
by the first 44 nucleotides of the genome (CSE1), SL2 is positioned at
the junction of the 5′UTR and the coding region of nsP1 while stem loops
3 and 4 lie in the coding region of the nsP1 forming the CSE2 (Frolov
et al. 2001, Gorchakov et al. 2003). CSE1 and CSE2 are important
for virus replication and it has been shown that during virus replication
the 5′UTR and 3′UTR interact with each other (Frolov et al. 2001). The
SL1 of 5′UTR was shown to be important for both minus-strand and
plus-strand-synthesis and in double-stranded form the CSE1 structure
in the complementary 3′UTR of the minus strand is proposed to act
as the promoter of plus-strand synthesis (Niesters and Strauss
1990a), although it has been shown that the 51-nt region is
dispensable for virus replication (Levis et al. 1986).

3′ untranslated region

The length of the 3′UTR also varies in size between alphaviruses from 121 to
524 nucleotides (Liu et al. 2009). Even though the sequences of the 3′UTRs
vary a lot they usually share a common core structure with short repeated
sequence elements, and 19-24 nucleotide long CSE4 preceding the
poly(A) tail (Ou et al. 1981, Pfeffer et al. 1998). CSE4 is considered to be the core
promoter for minus-strand synthesis and together with the poly(A) tail
needed for virus replication (Hardy and Rice 2005, Jose et al. 2009).

Research done with SINV clearly pointed out the importance of this sequence
for the minus-strand synthesis by showing that alteration of almost any
nucleotide of the 3′-terminal CSE compromised the minus-strand synthesis
(Hardy and Rice 2005).

It was shown that for minus-strand synthesis the interaction between the
5′UTR and CSE4 of 3′UTR is crucial.
(Frolov et al. 2001). Minimum length for the poly(A) tail is 11 to 12 residues and the correct position after the CSE4 is critical for efficient initiation of minus-strand synthesis (Hardy and Rice 2005). The actual minus-strand synthesis initiation site is located at the -1 position relative to the poly(A) tail (Hardy 2006). It still remains under debate whether there is a complementary polyurinylate tail in the minus-strand RNA (Sawicki and Gomatos 1976, Hardy 2006).

**Subgenomic promoter and other important sequences**

Subgenomic promoter is the third CSE (CSE3) found in the alphavirus genome (Strauss and Strauss 1994). CSE3 is located between the two ORFs and acts as a promoter for the subgenomic RNA. Subgenomic RNA is transcribed from the genomic length minus-strand RNA and is capped and polyadenylated (Levis et al. 1990). Subgenomic RNA does not serve as template for following minus-strand synthesis nor is it packaged into the capsid (Levis et al. 1990).

The CSE3 core sequence is highly conserved among alphaviruses and is positioned around the start site of subgenomic RNA. Core sequence is positioned in a way that 19 nucleotides are positioned upstream and 2 nucleotides are downstream from the transcription start site (Ou et al. 1982, Levis et al. 1990, Strauss and Strauss 1994). Interestingly the termination codons for nonstructural proteins lie within the subgenomic RNA region making the untranslated region of CSE3 relatively short (less than 50 nucleotides). The untranslated region of CSE3 also carries multiple stop codons for replicase protein termination varying from two to four between different alphaviruses (three for SFV) (Ou et al. 1982).

In addition to the replication-related RNA sequences alphavirus genome holds other important regions involved in e.g. genome packaging (Kim et al. 2011), the read-through of opal codons (Firth et al. 2011), and protein translation (Frolov and Schlesinger 1996).

**VIRUS-INDUCED MEMBRANE MODIFICATIONS**

All (+)RNA viruses replicate their RNA genomes in close association with different cellular membranes, which they usually heavily modify during infection (Belov and van Kuppeveld 2012). In addition to the different origin of modified membranes the positive-stranded RNA viruses induce a great variety of unique membranous rearrangements and compartments in the host cell (Miller and Krijnse-Locker 2008, den Boon and Ahlquist 2010, Romero-Brey and Bartenschlager 2014). The membranous structures vary from ~60 nm single membrane invaginations to enormous ~400 nm double-membrane vesicles (DMVs) (Froshauer et al. 1988, Knoops et al. 2008). Viruses also create various membranous webs, convoluted membranes and layered membrane structures.

It is not completely clear why RNA viruses replicate their genomes on
cellular membranes, but this is considered to provide many benefits to the replication. Membranous replication sites e.g. may act as scaffolds for anchoring the replication complexes to a defined position helping to increase the local concentrations of important components needed for replication. Membranous structures may also protect the viral RNA from cellular defense mechanisms during the infection (Neufeldt et al. 2016).

Endoplasmic reticulum

One membranous compartment, which is heavily modified by many viruses, is the endoplasmic reticulum (ER). Poliovirus (PV), a member of the Picornaviridae family, is one of the most studied ER modifying RNA virus. It creates heterogeneously sized vesicles of 70 – 400 nm in diameter in the perinuclear region of the cell. PV does not only modify ER membranes but also reorganizes the Golgi complex and lysosomal membranes to create these extensive membrane modifications (Bienz et al. 1990, Bienz et al. 1992, Schlegel et al. 1996, Belov et al. 2012). Data obtained with PV strongly suggests that there might be a link between polio-induced rearrangements and cellular autophagy pathway (Suhy et al. 2000, Jackson et al. 2005, Drex and Chisari 2014).

Another well-known RNA virus that is modifying the ER membranes is hepatitis C virus (HCV), a member of the Flaviviridae family (genus Hepacivirus). HCV creates membranous webs consisting of clusters of vesicles wrapped inside the ER membrane. These membranes are considered to be the actual sites of replication based on the localization of HCV nonstructural proteins and positive-sense RNA on them (Egger et al. 2002, Romero-Brey et al. 2012).

Dengue virus and West Nile virus, also members of Flaviviridae family, are using ER membranes as the sites of replication. Dengue virus induces large DMVs varying from 40 – 400 nm in diameter composed of two closely opposed membrane bilayers, vesicle packets, convoluted membranes and membranes associated with virus assembly (Mackenzie et al. 1996, Grief et al. 1997, Welsch et al. 2009). The vesicle packets are complex structures in which the outer membrane surrounds a series of inner vesicles (spherules) ~90 nm in diameter that contain most of the replication proteins, dsRNA and nascent RNA (Mackenzie et al. 1996). West Nile virus (WNV) also heavily modifies the ER membrane by inducing large convoluted membranes, paracrystalline arrays and vesicle packets described above. It was shown that dsRNA and especially the RdRp was localizing primarily to the vesicle packets. Interestingly the reconstructions of membranous WNV replication complexes also revealed narrow tubular structures between adjacent spherules in addition to the “normal” neck like openings to cytoplasm (Gillespie et al. 2010).

Severe acute respiratory syndrome (SARS) coronavirus belonging to the Coronaviridae family creates an ER-derived network of interconnected membranes, vesicles and convoluted membranes in which the viral proteins are located (Knoops et al. 2008). SARS induces the formation of big DMVs 200 – 300 nm in diameter in which the dsRNA seems to reside (Knoops et al. 2008). The actual site of the replication
is still under debate because it seems that the DMVs are devoid of openings to the cytoplasm. RNA replication within DMVs would require some kind of openings for transportation of important components like nucleotides and nascent RNA between DMVs and cytoplasm. Interestingly the outer membranes of DMVs are interconnected through narrow tubes and they seem to be often connected to the ER membranes (Knoops et al. 2008). One suggestion is that replication takes place on the convoluted membranes, which are interconnected with the DMVs where the viral replicase proteins accumulate during infection (Knoops et al. 2008).

A plant virus called brome mosaic virus (BMV), a member of the Bromoviridae family, also utilizes ER membranes when replicating but in contrast to PV, HCV, and SARS, BMV creates small vesicular invaginations (or spherules) ~60 nm in size instead of enormous membrane rearrangements. A single BMV replication vesicle contains all necessary components for replication: viral intermediate dsRNA and viral replication proteins (Restrepo-Hartwig and Ahlquist 1996, Schwartz et al. 2002). BMV replication compartments have been proposed to have a protein shell inside the vesicle composed of protein 1a and a neck like opening to the cytoplasm (Díaz and Wang 2014).

Many other (+)RNA plant viruses also induce membrane rearrangements in cells. In tobacco mosaic virus infected cells the replication complexes are associated with cytoplasmic inclusions originating from ER (Mas and Beachy 1999). Recent research with another plant virus called turnip mosaic virus (TuMV), belonging to Potyviridae family, shows also extensive virus-induced membrane rearrangements derived from ER. TuMV was shown to create convoluted membranes connected to ER and later during the infection also single-membrane tubules, which were shown to be the actual replication sites (Wan et al. 2015).

**Endosomal membranes**

RNA synthesis of Togaviridae like SFV takes place in virus-induced membrane structures called cytopathic vacuoles (CPVs). The limiting membrane of SFV-induced CPV is decorated with small invaginations ~60 nm in size called spherules (Froshauer et al. 1988). The nascent RNA, replication intermediate dsRNA and replication proteins localizes to these structures. CPVs are modified endosomal and lysosomal structures of 600 – 2000 nm in size (Grimley et al. 1972, Froshauer et al. 1988, Kujala et al. 2001, Spuul et al. 2010).

With SFV it has been shown that these small invaginations or spherules first arise on the plasma membrane from where they are internalized within small neutral carrier vesicles (Spuul et al. 2010). These neutral carrier vesicles were shown to utilize actin-myosin network to move away from plasma membrane and were eventually transported to perinuclear space close to the microtubule organizing center (MTOC) (Spuul et al. 2010). During the transport the spherule-containing vesicles merge to lysosomes and other endocytic vesicles to give rise to the CPVs (Grimley et al. 1968, Spuul et al. 2010).
**Peroxisomal membranes**

(+)-RNA plant virus called tomato bushy stunt virus (TBSV), belonging to the Tombusviridae family also induces spherule like structures on the site of replication. TBSV replicates on peroxisomal membranes and has been shown to use cellular ESCRT (endosomal sorting complexes required for transport) machinery for spherule formation (Kovalev et al. 2016).

**Mitochondrial membranes**

FHV is an insect virus belonging to the Nodaviridae family and utilizes mitochondrial membranes when replicating its RNA in Drosophila cells (Lanman et al. 2008). FHV is another spherule-inducing virus and creates numerous spherule-like structures of 40-60 nm in diameter between the inner and outer mitochondrial membranes (Miller et al. 2001). FHV protein A and the FHV RNA localize into these structures suggesting that the spherules are the actual site of replication (Miller et al. 2001). FHV-induced spherules also have an internal protein shell similarly to BMV. FHV-induced vesicles remain open to the cytoplasm via narrow neck structure as do the spherule structures induced by SFV and BMV (Kopek et al. 2007).

---

**VIRAL REPLICATION SYSTEMS**

To explore virus replication in more detail different kinds of replication systems have been created (Liljestrom and Garoff 1991, Khromykh et al. 1998, Spuul et al. 2011, Gomes et al. 2015). Depending on the mode of replication and the tolerance of the virus for modifications, replication can be artificially created to function either in cis or in trans (Fig. 4). Replication systems are extremely useful when studying e.g. lethal mutations or other aspects of virus life cycle, which cannot be studied with wild type virus. Replication systems have been shown to be valuable tools when studying e.g. pathogenic viruses, which would need to be otherwise handled in high level biosafety laboratory (Utt et al. 2016), mutations, which would be lethal in wild type virus (II), antivirals (Varghese et al. 2016) or protein functions (Khromykh et al. 1999).

**Replicons**

Replicons are widely used in virus research to mimic the replication. Alphavirus replicons are typically introduced to the cells in the form of in vitro transcribed RNA but also plasmid-based replicons are used. Replicons consists of the full length viral replicase with all necessary CSE elements for replication. Replicons express the replicase proteins (nsPs) but no structural proteins (pSFV1), thus the second ORF can be used to express other genes e.g. marker genes (Liljestrom and Garoff 1991). Replicons were originally developed for protein expression studies but as a self-replicating and self-transcribing entity they have been a powerful tool in biological research (Liljestrom and Garoff 1991, Spuul et al. 2011).
Two ways of replication typically exists, in trans replication presented on the left (trans-replication system) and in cis replication presented on the right (virus infection). In trans replication the replicase proteins are translated independently from different origin e.g. plasmid, not from the replicated RNA. The replicated template RNA is thus introduced from "outside" (meaning in trans). Contrary, in the in cis replication the replicase protein is translated from the same RNA that what is replicated by the replicase proteins. Major difference between the trans-replication system and virus infection is that in the trans-replication system no nsPs are produced from the RNA molecules produced by the system (Template RNA and Subgenomic RNA). In this system those RNA molecules express different replication markers e.g. fluorescent proteins. In virus infection multiplied genomic RNA molecules are used for translation and are thus producing more nsPs whereas structural proteins are translated from the Subgenomic RNA molecules.
Replicon RNA can be combined with the helper vector system (pSFV-Helper), which encodes all structural proteins. The helper vector sequence contains all necessary CSEs for replication but lacks the packaging signal (Liljestrom and Garoff 1991). This results in production of virus particles having replicon RNA but no helper RNA, which are thus incapable of producing infectious particles in the second round of infection.

Replicons have also been used to create replicon-containing cell lines (Pohjala et al. 2011, Utt et al. 2015). Cell lines have been generated by introduction of constitutively replicating replicon RNA into the cells. Replicon cell lines are powerful tools in virus research and have been successfully used in e.g. antiviral research (Varghese et al. 2016).

**Trans-replication systems**

Trans-replication systems are typically plasmid based systems, which mimic virus replication in mammalian cells (Spuul et al. 2011, Utt et al. 2016). Trans-replication system enables the expression of viral replicase in high levels independently from replication. Trans-replication system resembles other expression system called a trans-complementation system in which the functional copy of defected protein is recruited in trans (Khromykh et al. 1998, Khromykh et al. 1999), whereas in trans-replication system the template RNA is recruited by the replicase in trans (Spuul et al. 2011). Different trans-complementation systems have been designed but the main principle, to rescue the defective form, remains the same.

Two different trans-replication systems were recently created for CHIKV (Utt et al. 2016). In these systems two different promoters were introduced into the constructs in order to transcribe the mRNAs for nonstructural proteins, the T7 promoter and cytomegalovirus (CMV) promoter. The replicase expression and RNA replication are uncoupled and the mRNA transcription is achieved either by T7 RNA polymerase or by cellular RNA polymerase II (via CMV promoter). In cells co-transfected with constructs expressing template RNA and CHIKV replicases, formation of characteristic replicase spherules was observed (Utt et al. 2016).

In SFV trans-replication system, used in this thesis, the whole viral replicase polyprotein (P1234) is expressed in BHK cells, which express T7 RNA polymerase; these cells are called BSR T7/5 (Spuul et al. 2011). The replicase polyprotein is introduced from a plasmid construct carrying a promoter for the T7 polymerase as well as an internal ribosome entry site (IRES) element of encephalomyocarditis virus. The mRNA for the replicase polyprotein is synthesized by the T7 polymerase while the IRES element enhances the protein translation by cellular ribosomes. Replicase proteins in this system are fully functional and are capable of replicating specific RNA molecules provided in trans (Spuul et al. 2011).

The replication competent RNA molecules are designed to resemble viral genomes and are equipped with the CSEs needed for replication. The template holds the entire 5′UTR and 222 first nucleotides of the first ORF, 61 nucleotides from the end of the 3′UTR, subgenomic promoter, and poly(A) tail.
The RNA templates are also provided from plasmids and are originally synthesized by T7 polymerase. In this system the production of the replicase proteins is independent of the RNA replication, which makes the system more flexible for modifications (I, II, and III).
AIMS OF THE STUDY

The aim of this research was to study in detail how alphavirus replication complexes are assembled and to define the minimum requirements for the formation of the membranous replication spherule. In order to study the functions of both nsPs and the RNA genome, a plasmid-derived trans-replication system developed in our laboratory was used (Spuul et al. 2011). This system is designed to mimic the replication of SFV, a well-known member of the alphavirus family.

The first aim of this research was to study the replication of RNA templates differing in length. This research was inspired by the initial work of Spuul et al. (2011) in which it was suggested that the size of the RNA template may influence the size of the membranous replication spherule. The aim was to confirm the replication efficiency of all templates differing in length and to show that the spherules containing these RNA templates also differ in diameter. (I)

Secondly, by using the trans-replication system the essential functions and activities of SFV nsPs were studied in more detail. The well-known enzymatic activities of alphaviruses (e.g. helicase and RNA polymerase) were inhibited in order to study the replication initiation and spherule formation in more detail. (II)

Thirdly, the importance of the virus-specific sequences for the genomic RNA was assessed. In this study the aim was to identify the important RNA regions for SFV by modifying the template RNA in the trans-replication system. (III)
MATERIALS AND METHODS

The materials and methods used in this research are described in the original publications as indicated in Table 2, Table 3 and Table 4. The asterisks indicate the candidate’s contribution to the method listed, (**) indicates that the candidate has a strong knowledge of the method and has performed it herself and (*) indicates that the candidate is familiar with the technique but has not used them herself. Methods, plasmid constructs and antibodies used are listed in separate tables.

Table 2. Methods used in the current study

<table>
<thead>
<tr>
<th>METHODS</th>
<th>Used in publication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cells and viruses</strong></td>
<td></td>
</tr>
<tr>
<td>BSR T7/5 **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>BHK **</td>
<td>I</td>
</tr>
<tr>
<td>wtSFV **</td>
<td>I</td>
</tr>
<tr>
<td>SFV-Rluc **</td>
<td>I</td>
</tr>
<tr>
<td><strong>DNA and RNA transfections for</strong></td>
<td></td>
</tr>
<tr>
<td>CLEM **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>RNA labeling **</td>
<td>I</td>
</tr>
<tr>
<td>luciferase measurements **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>linearized plasmids *</td>
<td>I</td>
</tr>
<tr>
<td><strong>DNA techniques</strong></td>
<td></td>
</tr>
<tr>
<td>cloning **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>PCR **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>synthetic sequences</td>
<td>III</td>
</tr>
<tr>
<td><strong>RNA techniques</strong></td>
<td></td>
</tr>
<tr>
<td>in vivo [3H] labeling of cells **</td>
<td>I</td>
</tr>
<tr>
<td>isolation and purification **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Northern blotting **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>RNase treatments **</td>
<td>II</td>
</tr>
<tr>
<td><strong>Microscopy</strong></td>
<td></td>
</tr>
<tr>
<td>Confocal microscopy **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>TEM **</td>
<td>I, II, III</td>
</tr>
<tr>
<td>CLEM **</td>
<td>I, II, III</td>
</tr>
<tr>
<td><strong>Data analysis</strong></td>
<td></td>
</tr>
<tr>
<td>Image Pro plus **</td>
<td>I</td>
</tr>
</tbody>
</table>

23
<table>
<thead>
<tr>
<th>Plasmid Constructs</th>
<th>Used in Publication</th>
<th>Originally Described In</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Polyprotein constructs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P123Z4</td>
<td>I, II, III</td>
<td>(Spuul et al. 2011)</td>
</tr>
<tr>
<td>P123Z</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P123Z4(m)</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>P1^H38A^23Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^D64A^23Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^Y249A^23Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^R253E^23Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^W259A^23Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^C418-420A^23Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^K192N^3Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^C478A^3Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P1^2^3Z4</td>
<td>II</td>
<td>(Spuul et al. 2011)</td>
</tr>
<tr>
<td>P123Δ50Z4</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>P123Z4^GAA</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td><strong>Template constructs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC18 templ+</td>
<td>I, III</td>
<td>(Spuul et al. 2011)</td>
</tr>
<tr>
<td>Tshort</td>
<td>I, II, III</td>
<td>(Spuul et al. 2011)</td>
</tr>
<tr>
<td>TshortCh</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>Tshort_24loop</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>TshortΔ51nt</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>TshortΔnsP1</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>TshortΔSGP</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Tshort5exc</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Tshort3exc</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Tshort(m)</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Tmed</td>
<td>I, II, III</td>
<td>(Spuul et al. 2011)</td>
</tr>
<tr>
<td>Tmed_Vis</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>TmedR</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>TmedRΔ51nt</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>TmedRΔnsP1</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>TmedRΔSGP</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>TmedR5exc</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>TmedR3exc</td>
<td>III</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. List of plasmid constructs created and used in the study. The constructs created previously are indicated by the reference.
<table>
<thead>
<tr>
<th>Table 4. List of primary and secondary antibodies used in Western blot analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
</tr>
<tr>
<td>anti-β-actin</td>
</tr>
<tr>
<td>IRDye®800CW donkey anti-rabbit / anti-mouse IgG</td>
</tr>
<tr>
<td>Alexa Fluor 680 anti-mouse / anti-rabbit IgG</td>
</tr>
<tr>
<td>anti-nsP1</td>
</tr>
<tr>
<td>anti-nsP3</td>
</tr>
<tr>
<td>anti-nsP4</td>
</tr>
<tr>
<td>anti-mCherry</td>
</tr>
</tbody>
</table>
RESULTS

An efficient plasmid-based trans-replication system was recently created to study alphavirus replication in mammalian cells (Spuul et al. 2011). In this system both the viral replicase proteins and the RNA template molecules are encoded by separate plasmids, which allows the modification of both components of the system individually. Spuul et al. (2011) created an extensive set of RNA template constructs and different polyprotein constructs, which were proven to be functional and comparable to wild type SFV replication (I). These constructs were then further modified to fit our experiments (I, II, III). The basic structure of the expression plasmids were kept unmodified except in the experiments specifically focusing on defining the functions of the virus-specific sequences (III). In the experiments described below the polyprotein plasmid was always co-transfected with the template plasmid if not specified otherwise.

CHARACTERIZING THE FUNCTIONS OF THE VIRAL RNA GENOME (I, III)

In trans-replication system the template molecules are provided from plasmids, which is why modifying them is relatively easy. Templates differing in size and with deleted or exchanged conserved sequence elements were created to decipher the role of the viral genome in the replication of SFV. Plasmid constructs expressing the complementary minus-strand were created in order to study whether they are replication-competent.

Replicase can replicate different sizes of RNA templates

The template construct set created by Spuul et al (2011) included three templates varying in size (I, Fig 1). The templates were named Tshort, Tmed and Tlong according to their lengths (1.5 kb, 3 kb and 6 kb respectively). All templates share the same basic structure having the T7 promoter followed by viral 5'UTR and the first 222 nucleotides from the coding region of nsP1, subgenomic promoter and partial 3'UTR with poly(A) tail (Spuul et al. 2011). Replicase protein constructs were designed in a way that replicase proteins can be expressed either from one plasmid carrying the whole replicase protein ORF (P1234) or so that nsP4 is expressed from another plasmid. When expressed separately, nsP4 is fused with ubiquitin to ensure the native Tyr residue at the N-terminus of the protein (Lemm et al. 1994). In this system the mRNA for the replicase proteins is not replication-competent, ensuring that all replicase proteins are only expressed from the plasmid-derived mRNA. This way the expression of the replicase is not dependent of replication. The template set created by Spuul (2011) was complemented with two additional template constructs termed TshortCh (1.3kb) and Tmax (11.2kb) (I, Fig 1). The TshortCh template was created solely for the correlative light and electron microscopy (CLEM) experiments while the Tmax was
created to correlate with the size of the wild-type SFV genome (Hellstrom et al. 2015). When tested in fluorescence microscopy both templates gave a red signal indicating ongoing replication when co-transfected with the polyprotein construct (I, Fig 3). The replication efficiency of the template Tmax was extremely low based on luciferase assay but high enough to be used in the CLEM experiments (I, Fig 3).

All template lengths were replication-competent but the replication efficiency varied between them (Spuu et al. 2011). It was obvious that the longer the template the lower the luciferase counts obtained (I, Fig. 3). The transfection efficiency was also greatly reduced with longer templates (I, Fig 3). To confirm that the newly synthesized RNA with different templates was correct in size Northern blot analysis were performed with radioactively labelled probe against the luciferase gene. All template RNAs produced in the trans-replication system were correct in size and the levels of accumulated RNA correlated with the luciferase results: Tshort gave the highest levels and Tlong the lowest (Fig 3). We detected some extra bands in the plus-strand blots, which were proven to be caused by the T7 polymerase as shown with the linearized templates (I, Fig. 4). In the minus-strand blots only the bands corresponding to the actual plus strand template bands were detected (I, Fig 3 and 4).

**Fig. 5.** In CLEM experiments the coordinates in the EM thin sections can be visualized only from the first few thin sections. These sections are cut from the bottom of the cell. In these sections spherules are often observed in the orientation showed in the figure B. In this orientation the spherule structure resembles a closed sphere because the neck-like opening is not seen. Most of the counted spherules lie in this orientation but when thin sections are cut higher from the cell, more vertically cut spherules are seen (A). From spherules cut in this vertical orientation the complete outline of the spherule can be observed including the thin neck-like opening (A, black box). A small set (n = 40) of vertically cut spherules from three independent experiments were measured to obtain the widest diameter. Images in the white boxes are showing the orientation and the average of the widest diameters for SFV induced spherules either vertically cut (A) or horizontally cut (B).
Size of the template determines the diameter of the spherule

Preliminary data obtained with those three original templates suggested that the size of the spherule structure correlates with the size of the replicated template RNA. To further study this observation all template lengths; TshortCh (1.3 kb), Tmed (3 kb), Tlong (6 kb), and Tmax (11.2 kb), were analyzed by using CLEM technique and average spherule sizes were calculated based on more than 4,200 individually measured spherules (I, Fig 2). Templates differing in size induced clearly different sizes of spherules compared to SFV (11.5 kb) except the Tmax, which induced spherules almost similar in size with the wild type virus (I, Fig 2). The average spherule sizes with different templates were 40 ± 6 nm for Tshort, 39 ± 4 nm for Tmed, 45 ± 5 nm for Tlong and 57 ± 5 nm for Tmax whereas SFV-induced spherules were 58 nm in size (I, Fig 2).

In addition, diameters were measured from vertically cut spherules. The data obtained clearly correlated with the data obtained by measuring mostly horizontally-cut spherules. The average of widest diameters of spherules induced by SFV was 57 nm (Fig. 5).

Different sizes of templates were also tested in linearized form. Linearization was performed to avoid the synthesis of extended template transcripts produced by T7 polymerase. Linearized plasmid constructs were replication-compatible and gave relatively high luciferase values (I, Fig 4). Average diameters of the spherules were 32 ± 5 nm for Tshort, 42 ± 5 nm for Tmed, and 48 ± 7 nm for Tlong (I, Fig 2). The linearized form of Tshort gave rise to spherules smaller in diameter when compared to the non-linearized template plasmid.

Fig. 6. Replication of replicon RNA in different cell lines induced spherules different in size. In BHK cells the average diameter of replicon-induced spherules was 57 nm whereas average diameter in BSR cells was 50 nm (graph on left). It seems that the BHK cells are for reason unknown inducing wider spherules because when replicon data was compared with other spherule measurements in BSR cells (right graph) the size of the replicon made spherules correlated with the size of the spherule similarly to other templates and SFV (graph on right).
The influence of RNA size on the spherule size was further tested by using SFV replicons. Replicons are capable of replicating in cis but are not producing virus particles because of the lack of the structural ORF. Replicon RNAs were transfected into the cells alone or together with in vitro transcribed Tmed template RNA and analyzed with CLEM. For replicon alone more than 2,800 spherules were analyzed to determine the average diameter to be 56 ± 8 nm in BHK and 50 ± 3 nm in BSR (n = 749) (Fig. 6). When co-transfected with Tmed RNA and analyzed with CLEM, two populations of spherules clearly different in size were observed within one cell. Spherules induced by Tmed replication showed in average diameter of ~43 nm and the in cis replication of replicon ~58 nm (I, Fig 5).

Minus-strand RNAs cannot function as a template for the replicase

With a system as flexible as the trans-replication system, it was relatively easy to decipher the roles of the minus strand RNA. It has been suggested that the minus-strand RNAs only exist inside the replication spherules and are never released into the cytoplasm. In order to study whether a minus-strand can initiate the replication a template construct carrying the template sequence in reverse complementary orientation was created. This way the T7 polymerase transcribes an RNA transcript resembling the minus strand, normally produced only by the replicase. This minus template was incapable of initiating the viral replication, thus no newly synthesized RNA nor increased luciferase values were detected (III, Fig 5).

In order to study whether any proteins can be expressed from the minus-strand a template carrying the luciferase gene under the subgenomic promoter in reverse complementary orientation was created. In this orientation the luciferase gene can be translated from the minus-strand RNA only. This template was efficiently replicated by replicase but no detectable luciferase values were seen meaning no luciferase was translated from the minus-strand during replication. Thus the minus-strand RNA is not available for translation during replication.

To ensure the capability of cellular translational machinery to use these artificially created minus templates for translation the luciferase gene was inserted in plus sense orientation in the minus template ([−]StLucREV). With this template the luciferase values were at the same level with the control template (CFP-StLuc) carrying CFP in the 1st ORF and luciferase in the 2nd ORF (III, Fig. 6).

CSEs of template RNA are needed for efficient replication

To address the roles of the conserved viral RNA sequences in the genome, deletions and exchanges of segments were made to the template constructs Tshort and TmedIR (III, Fig 1 and 3, respectively) (Fig. 7). Several new templates were constructed where either important elements were deleted (Tshort∆SGP, Tshort∆51, and Tshort∆nsP1) or where the UTRs were changed to unrelated sequences of a human gene (Tshort5exc and
The replication efficiency for all modified templates was addressed in luciferase assay when co-transfected with the polyprotein construct. Deletion of the subgenomic promoter (ΔSGP) had no effect on the replication and the deletion of the 51-nt element (ΔS1) only reduced the efficiency to some extent (III, Fig 1). These two templates (ΔSGP and ΔS1) were also capable of producing relatively high levels of both the plus- and the minus-strand RNA (III, Fig 2). Template with the changed 5’UTR but unmodified 3’UTR was also applicable for minus-strand synthesis but not for the following plus-strand synthesis, or the levels of newly synthesized plus-strands were too low to be distinguished from the background. Other modifications in the template construct (ΔnsP1 and 3exc) abolished both minus-strand and plus-strand synthesis. No minus-strand was produced in the absence of the replicase (III, Fig 2).

The first 222 nucleotides of the coding region of nsP1 inserted in the template (same region as deleted in ΔnsP1) were also intolerant for nucleotide changes. Alteration of the nucleotide sequence in the template without affecting the amino acid sequence of the nsP1 region caused significant reduction in the replication (III, Fig 4).
Several mutations were introduced to specific sites in replicase proteins of SFV in the trans-replication system (II, Table 1) (Fig. 8). These mutations have proven to be lethal in SFV or to generate many compensatory mutations when introduced in the SFV genome (Spaull et al. 2007, Zusinaite et al. 2007). All the mutant constructs were tested in luciferase assay and were shown to be non-replicating except those termed Δ50 and 1^2^3Z4 (II, Table 1).

The mutant Δ50 lacks all phosphorylation sites of nsP3 and when this deletion is introduced to SFV genome (SFV-Δ50) it slows down the spherule intake from the plasma membrane into the carrier vesicles. (Thaa et al. 2015). In trans-replication system the effect of the Δ50 deletion was not that clear and the results obtained were bit puzzling. In trans-replication system the replicase with the Δ50 mutation was not accumulating on the plasma membrane similarly to SFV when imaged by confocal microscopy and the spherule formation could not be detected. For reasons still unknown luciferase values raised above the background but no detectable amounts of plus or minus sense RNA were detected when using the Δ50 mutant.

The 1^2^3Z4 is quickly processed in cells into a form of P123 + nsP4, which is responsible for the minus-strand synthesis and is possibly capable of generating reduced amounts of plus-strands (Lemm et al. 1994). Other mutations were affecting the replication either at transcriptional level (capping, helicase, and polymerase) or at post-translational level (cleavage and membrane binding). When analyzed by Northern blot, four of these mutants were capable of synthesizing the minus-strand. Minus-strands were synthesized by the ones involved in the capping reactions (H38A, D64A and Y249A) and the P1^2^3Z4 as being the “early

![Fig. 8. Schematic of the SFV polyprotein containing fluorescent protein gene ZsGreen (ZsG) in fusion to nsP3 gene. Mutations are marked on the polyprotein. Mutations with similar type of action are highlighted by colored circles; light red is grouping mutations involved in the capping reactions, light blue is showing the modifications involved in the polyprotein processing and light green is showing the ones involved in the membrane association.](image-url)
polymerase” complex responsible for the minus-strand synthesis (II, Fig 1). Minus-strand synthesis was not detected with the other mutants. The constant protein levels and the correct proteolytic processing of mutated replication proteins was confirmed by Western blotting (II, Fig 1).

Data clearly shows that nsP4 is absolutely needed for replication and cannot be eliminated from the replication machinery as was shown with P123 construct, which was completely incapable of replicating the template RNA (II, Table 1 and Fig. 1). The uncleaved polyprotein P123Z4 (CA) was completely replication-negative suggesting that nsP4 needs to be cleaved from the polyprotein in order to form functional replication complex. Introduction of additional nsP4 into the system with P12ca3Z4 did not rescue the phenotype (unpublished data) (Fig. 9).

Fig. 9. Undeaved P123Z4 (12ca3Z4) was shown to be replication negative when measured with luciferase based assay. Co-transfection of the cells with (ubi)nsP4 construct and polyprotein P123Z4 carrying an inactivated protease (ca) together with Tmed template did not create functional replication complexes. The level of replication stays at the background level whenever the nsP4 is not cleaved; 12ca3Z4 and 12ca3Z4+nsP4. Introduction of additional nsP4 into the system did not rescue the observed phenotype. Values presented in x-axis are relative luciferase values.

In order to study the requirements for spherule formation all modified constructs were analyzed in the EM level. To study the non-replicating replicase proteins it was crucial to be able to identify the cells containing both constructs, i.e. the template plasmid and the polyprotein plasmid. Polyprotein plasmid expression was easily recognized using nsP3-ZsG fusion protein as a marker but the fluorescent protein cloned under the control of SGP was not expressed because of the lack of replication.

To overcome this problem we introduced a new ORF into Tmed template construct behind the original template sequence. Fluorescent marker mCherry with nuclear localization signal was cloned to the second ORF from which it was independently expressed (Fig. 7). This new template, which was named Tmed_Vis, was confirmed to be replication-compatible by luciferase assay and by visual
observation using confocal microscopy (II, Fig 3). With this new Tmed_Vis construct the spherule formation with the various mutated replicases could be analyzed. Both plasmids were expressing fluorescent marker, which was easily visualized by confocal microscopy.

To summarize, it was clearly shown that the same mutants, which were capable of initiating the replication by minus-strand synthesis (H38A, D42A, Y249A and 1^2^3Z4), were also inducing spherules. Even after extensive analysis spherules were not detected with other mutated replicases (II, Fig 3). The same phenomenon was observed with modified template constructs; only the ones synthesizing minus-strand RNA were generating spherules. Thus, all of the data showed a strong correlation between the minus-strand synthesis and the spherule formation.

**RECRUITMENT OF THE RNA TEMPLATE BY REPLICASE (II)**

In order to study RNA template recruitment to replication complexes flotation assay was used. In this experiment transfected cells were lysed and subjected to flotation assay in a sucrose gradient ultracentrifugation. During the centrifugation the cellular membranes carrying the replication spherules start to float and concentrate to certain sucrose density. After flotation the membranous fractions can be isolated and further studied.

To confirm that the system can be used to study membrane bound proteins and RNA a fusion protein capable of binding the membranes as well as an RNA molecule was created. This dual functionality was achieved by fusing the fluorescent protein mCherry carrying C-terminal CAAX motif with bacteriophage MS2 coat protein (Fusco et al. 2003, van Rheenen et al. 2007). The CAAX motif directs proteins to the cellular membranes via binding a hydrophobic farnesyl group. The farnesyl group is added to the CAAX motif by a cellular farnesyltransferase. The MS2 bacteriophage capsid protein binds irreversibly to specific RNA loops, which were included in the template RNA. The flotation assay and cell fractioning was combined with Northern blot analysis to confirm the recruitment of the RNA templates. With this arrangement it was possible to bind the RNA template artificially to the plasma membrane showing that the system itself is functional (II, Fig 2).

The Tmed template was tested together with the replicase and after Northern analysis both minus-strands and plus-strands were seen in the membranous fractions as well as in the bottom fractions (II, Fig. 2). However, the template RNA did not associate with the membranous fractions when expressed alone or with replication-defective replicase.

Western blotting clearly showed that nsP1, being the sole membrane-bound nsP of SFV, was exclusively located to the membrane fractions after flotation and fractionation. The membrane-targeted mCherry (decorated with membrane binding CAAX motif) was also mostly found in the membrane fractions (II, Fig 2). Experiment clearly showed a strong binding of nsP1 and
CAAX-fusion protein on cellular membranes and that both RNA strands float with membranes presumably inside the spherule structures.

The experimental setup was also used for RNase treatments in order to study the stability of the RNA. The stability tests were performed with wild type replicase and with certain mutated replication defective replicase. The data obtained clearly shows that in the case of full replication both strands are protected in the membrane fraction but not in the soluble fraction (II, Fig 2). Protection was further studied with membrane destabilizing agents and, indeed, Triton X-100 treatment together with RNase treatment degraded the RNA completely.
DISCUSSION

TEMPLATE RNA HAS A CRUCIAL ROLE IN SPHERULE FORMATION

Preliminary data obtained by the use of trans-replication system (Spul et al. 2013) suggested that the size of the replicating template influences the size of the replication-induced spherules. This phenomenon was novel and challenged the existing view of spherule formation. Generation of differently-sized spherules was not a new discovery. Previously, smaller spherules have been generated e.g. by altering the membrane binding domain of viral proteins (Liu et al. 2009), by depleting important host factors like reticulon homology protein from BMV spherules (Diaz et al. 2010), or by altering the host lipid metabolism (Zhang et al. 2012). However, different sizes of spherules have not previously been generated by alternating the length of the replicating RNA. In contrast, Kopek et al. (2010) showed that FHV-induced spherules can tolerate even ten-fold changes in the length of the RNA without changes in size. The sizes of spherules were repeatedly documented to be quite constant within infected cells and some viruses were shown to have a protein lining the internal side of the spherule structure (Kopek et al. 2007, Diaz and Wang 2014). With this in mind it was easy to imagine spherules as non-dynamic structures with defined dimensions achieved by viral proteins together with host factors.

To study this phenomenon a set of templates varying from 1.3 kb to 11.2 kb were used. The template set included Tshort (1.5 kb) template, which was designed to be used in luciferase assays and thus did not include a fluorescent marker necessary for CLEM. In order to create as small spherules as possible we created a new template called TshortCh in which the fluorescent marker was placed under the control of subgenomic promoter. A fluorescent protein located under the subgenomic promoter proved to be an excellent tool when studying replicating cells in EM, because markers from the second ORF are only expressed when the RNA template is replicated. Surprisingly, we were able to see spherules clearly different in size.

To eliminate the possibility that the size-difference is caused by the trans-replication system itself another template called Tmax (11.2 kb) was created. This template, similar in size with the SFV genome, was used to compare spherules induced by SFV and the trans-replication system. As expected, spherules from both experiments were very similar in size (SFV 58 nm Ø and Tmax 57 nm Ø). Taking into consideration that Tmax is only 300 nucleotides shorter than SFV genome one can state that the spherules are similar in size and the system itself does not influence spherule formation.

To further study this novel finding the self-replicating replicon RNA was used together with shorter template RNA (Tmed). The default in this experiment was that the replicon RNA (8.8 kb) would replicate in cis whereas the Tmed RNA (3 kb) would replicate in
trans. nsPs would be provided from the replicon mRNA only. Both constructs were co-transfected into BHK cells as RNA and the samples were processed for CLEM. The in trans replication of Tmed RNA was easily identified based on red fluorescence and those cells were selected for EM. Different sizes of spherules were indeed formed within one cell and the spherules measured formed two separate pools of spherules with different diameters. These diameters correlated with the data obtained with the trans-replication system. For reason unknown spherules formed in BHK cells seemed to be somewhat larger than in BSR cells. To confirm that the size difference is due to the different cell line the replicon RNAs were transfected also into BSR cells. The data clearly showed that spherules formed in these two cell lines were slightly different in size (Fig. 6).

One very interesting question is the maximum size of the replication spherule. Longer templates were not included in the study because the Tmax template (11.2 kb) was already very challenging to work with presumably due to the low transfection efficiency and/or RNA stability. It would be very interesting to find out whether there is a maximum size for the SFV-induced spherule.

Interestingly, recent study with BMV showed by modifying the lipid composition of the host cell that larger spherules can indeed be induced. The study shows that BMV like other positive-strand RNA viruses promotes host lipid synthesis and specifically phosphatidylcholine (PC) synthesis/accumulation at viral replication sites (Zhang et al. 2016). After BMV infection the total level of PC was ~29% higher when compared to non-infected cells (Zhang et al. 2016). PC is typically synthesized in yeast cells via CDP (phospholipase D)-DAG (phosphorylation of diacylglycerol) pathway (Carman and Henry 1999). The two MTs catalyzing the PC synthesis are Cho2p and Opi3p. BMV 1a protein is shown to interact with Cho2p enzymes and recruits them to replication sites. Depletion of Cho2p from cells greatly reduces the replication efficiency of BMV but does not block the spherule formation. Interestingly, the spherule size in Cho2p-depleted cells was increased by ~25% when compared to spherules formed in wild type cells (Zhang et al. 2016). In the study by Zhang et al. (2016) the spherules induced in modified cells were greater than spherules induced by virus in non-modified cells. This result clearly shows that many factors are involved in the spherule formation in addition to the virus itself and e.g. the correct membrane composition is important for the replication and spherule formation (Zhang et al. 2016).

In addition to the Tmax related difficulties, the use of the smallest template Tshort also raised some technical issues. Tshort-induced small spherules were very difficult to find and identify as spherules in EM. Secondly, when using circular plasmid constructs, spherules induced by Tshort and Tmed templates were similar in size. Northern blot result clearly showed that the termination site of T7 polymerase is leaky and that the polymerase introduced extended plasmid-derived templates of different size into the cell. This phenomenon was encountered by linearizing the template plasmids right after the actual template sequence, which led to
production of templates with correct size. Interestingly, spherules induced by the linearized Tshort plasmid were smaller than the ones induced with the circular plasmid. It is possible that longer templates (i.e. leaky products of T7 polymerase activity) were recruited to the replication complexes to some extent and induced spherules with wider diameter. As these small spherules were difficult to find the imaging and measuring was possibly biased by the larger spherules. With linearized plasmid the spherule diameters were smaller and correlated with the size of the template.

The fact that the length of the replicated RNA has an influence on spherule size gives new perspective to the research of spherule formation and supports the emerging view that spherules are formed in different ways and their morphology varies depending on the virus. It also indicates that the template RNA has an important role in spherule formation. It was recently shown that RNA also defines the spherule size of tombusviruses. This supports these data and emphasizes the role of the RNA in spherule formation (Kovalev et al. 2016).

It has been shown with SFV that the replicase proteins are not sufficient to induce spherules when expressed alone. For SFV it is crucial to have active RdRp and replication-compatible RNA template present in the cell to induce spherules. In contrast, replicase protein 1a of BMV is capable of producing spherules without the genomic RNA or replication (Schwartz et al. 2002). The 1a protein forms a protein shell inside the spherule structure, which presumably defines spherule size (Diaz and Wang 2014). FHV also forms a protein shell inside the spherule but in contrast to BMV it needs active RdRp and genomic RNA to induce spherules. Based on the known spherule sizes and the lengths of the genomes, it has been speculated what is the amount of RNA that can be sealed into a single spherule. This question remains open for discussion (Kopek et al. 2007).

How the RNA is measured during the spherule formation is another question to be discussed. SFV possesses only one linear single-stranded genome and for every plus-strand RNA presumably only one complementary negative-sense RNA molecule is synthesized and yet it is not known whether there is only one or several dsRNAs within one spherule. It is also under debate whether minus and plus strand RNAs are always together in the same spherule in the form of dsRNA. The fact that there is always complementary RNA molecules in the replicating cells has complicated the studies of the nature of RNA. Complementary RNA molecules are likely to form RNA duplexes and thus isolating the RNA in its native form is challenging. Isolation of RNAs from purified membranous spherules might give a hint about the ratios of plus-strand and minus-strand RNAs but does not reveal the form of the RNA.
Conserved sequence elements ensure replication compatibility of the template

As mentioned above these data and the data obtained with tombusviruses clearly show that the length of the replicated RNA determines the size of the spherule at least in the case of some viruses. Viral genomes are often very compact and do not allow many modifications. Only two of the known CSEs are dispensable in the case of SFV; the 51-nt CSE2 and the subgenomic promoter (CSE3). In trans-replication system the deletion of the subgenomic promoter (SGP) had no effect on the replication of genomic RNA although in the virus lack of SGP would compromise the production of structural proteins (Strauss and Strauss 1994). The 51-nt region includes two stem loop structures called SL3 and SL4 and they have been shown to be dispensable also with SINV (Frolov et al. 2001). With SINV it was shown that deletion or mutation of these stem loops reduced replication efficiency and they were thus considered to act as replication enhancers (Frolov et al. 2001). Exactly the same effect was seen in our experiments with the Δ51 mutant; replication was not inhibited but only reduced (II, Fig 1).

In the template, after the 5′UTR there is a short stretch of nucleotides from the coding region of nsP1. These 222 nucleotides comprise the only region in the template identical to the polyprotein mRNA but it is shown to be important for the replication. SL3 and SL4 (CSE2), discussed above, lie within this nsP1 region and there is a third stem loop located right at the junction of the 5′UTR and the coding region of nsP1. Interestingly the deletion of the CSE2 (Δ51) only reduced the replication efficiency whereas deleting the whole nsP1 region (ΔnsP1) abolished the

![Fig. 10 Schematic of Tmed template. Important CSEs from virus RNA are implemented into the N-terminus of the template including 5′UTR and 222 nucleotides from the beginning of the nsP1 coding region (purple). Marker genes (I and II) can be chosen according to the experiment. Deletions and modifications are marked in the brackets above or below the RNA region, which they influence. CSE, conserved sequence element; SL, stem loop; SGP, subgenomic promoter; Δ, deletion; exc, original sequence changed to non-virus sequence.](image-url)
replication completely. SL2 is most probably also destroyed in $\Delta$nsP1 construct and it is difficult to state whether the complete shut-off of replication is due to the loss of all three stem loops or because of some other (lost) features of the nsP1 region.

In order to study the importance of the SL structures within the beginning of the nsP1 coding region (added into the template sequence) the nucleotide sequence was modified to disturb the correct folding of SL3 and SL4. Construct with modified nucleotide sequence was capable of replicating indicating that the modified region was enough for the template to be used in the replication. This template was less efficient when compared to the Tshort but was still functional unlike the $\Delta$nsP1 template.

It has been shown with alphaviruses e. g. SINV, VEEV and now with SFV that this short 51-nt region is important for replication (Niesters and Strauss 1990b, Frolov et al. 2001, Michel et al. 2007). It was shown with SINV that mutations in this region (destabilizing the loop structures) only prevented the replication in insect cells but the virus was viable in mammalian cells (Frolov et al. 2001). Hyde et al. (2015) showed that this particular region is involved in replication in insect vectors. Research done with VEEV 51-nt region showed that loss of one SL within this region was insufficient to inhibit the replication whereas destabilizing both SLs were suppressing the replication to great extent. VEEV with deleted 51-nt CSE quickly introduced compensatory mutations into nsP2 and nsP3 to restore the capability of replication complexes to bind the genomic RNA (Michel et al. 2007). Silent mutations within this region strongly inhibited the replication of the virus in both, vertebrate and invertebrate cells, which was also seen with SINV (Michel et al. 2007, Hyde et al. 2015). Michel et al. (2007) state that SLs may be involved in the core promoter recognition and act as a binding site for replicase proteins and, thus, are important in the replication.

In this study the construct was only tested in mammalian cells but the effect was similar to SINV, showing reduced levels of replication. Interestingly, in SINV the deletion of SL2 alone increased the replication at least in mammalian cells (Frolov et al. 2001). In these experiments there were no constructs lacking the SL2, but it seems that the stem loops two to four in alphaviruses have a role in the fine-tuning of replication efficiency especially in different hosts.

Both UTRs have a significant role in the virus replication. Exchange of either UTR to non-viral UTR abolished the replication completely. Interestingly the template in which the 5'UTR was exchanged (5exc) was still applicable for minus-strand synthesis and spherule formation likely due to the remaining partial viral 3'UTR and nsP1 region. Intact virus-derived 3'UTR together with the nsP1 region is enough for the recruitment of the template by replicase and to be used in the minus-strand synthesis. The newly synthesized complementary minus-strand anyhow was incapable of acting as a template for the plus-strand synthesis. It has been reported earlier that the 3'UTR is needed for minus-strand synthesis (Niesters and Strauss 1990a, Frolov et al. 2001), which is in line with the data obtained in the current study. With template construct carrying exchanged 3'UTR (3exc) no
minus-strand or spherule formation were detected but it is not evident from the data whether the template was even recruited by the replicase or not.

**FUNCTIONAL NONSTRUCTURAL PROTEINS ARE THE KEY PLAYERS IN SPHERULE FORMATION**

In addition to the template, functional nsPs are also needed for spherule formation. In this study the functionality of viral replicase was tested from which activities of several well-known enzymes were inhibited by mutations (II, Table 1). Some of the introduced mutations were known to be lethal when introduced in SFV genome but some of them were provoking the mutated virus to generate compensatory mutations to overcome the initial defect (Rikkonen 1996, Zusinaite et al. 2007). After introducing the mutations into the nsPs in trans-replication system the ability of these crippled nsPs to function in the replication machinery was assessed. The key functions of replicase studied were capping of viral RNA (H38A, D64A and Y249A), membrane association (R253E, W259A and C418-420A), helicase activity (GNS), protease activity (CA) and polymerase activity (GAA). In addition the functions of non-phosphorylated replicase (Δ50) and a construct having the protease cleavage sites destroyed between nsPs, marked with ^ symbol (P1^2^34) were addressed. Enzymatic activities needed for the initiation of replication were determined by co-transfecting the Tmed template together with these mutated nsPs. It was clear that the lack of capping activity did not prevent the recruitment or minus-strand synthesis and all three capping related mutants produced similar levels of minus-strand RNA. Subsequent plus-strand synthesis could not be detected but that does not exclude the possibility that the plus-strand levels were too low to be detected. Interestingly, the mutant Y249A behaved similarly to the mutants previously shown to be involved in the capping event even though the mutation itself is located in the membrane binding peptide region of the nsP1 (Ahola et al. 1997). Based on the data this highly conserved residue (Y249) is more important in the capping of viral RNA than in the membrane association (Spuul et al. 2007, Ahola and Karlin 2015).

The mutants involved in the membrane association of the replicase were behaving as expected. Spuul et al. (2007) showed that nsP1 carrying the mutations in the binding peptide region (either R253E or W259A) were diffusely spread into the cytoplasm in contrast to the wild type nsP1, which localized to the plasma membrane. These two mutants were also shown to be incapable of synthesizing the complementary minus-strand, presumable due to the lack of membrane association and failure to form functional replication complex. Membrane association has been shown to be one requirement for viral replication, which is in agreement with our data (Spuul et al. 2007). The residue W259 was shown to be the key residue in the binding peptides interaction with membranes by penetrating deep into the lipid bilayer, whereas other residues maintain ionic
interactions with membrane lipid head groups and acyl chains (Lampio et al. 2000).

The lack of membrane binding was also seen in Western blot analysis, where lower levels of nsP1 were seen with these mutants (II, Fig 1). Surprisingly, the palmitoylation negative mutant (C418-420A) was completely replication-negative, which may be due to the trans-replication system. In palmitoylation negative virus compensatory mutations arise quickly and virus growth is restored. In trans-replication system compensatory mutations do not arise because the replicase proteins and template RNAs are expressed from plasmids and there is no pressure for mutations (Zusinaite et al. 2007).

The data obtained with replicase having inactive protease or inactivated cleavage sites supports the fact that the cleavage of nsP4 is crucial for minus-strand synthesis (Lemm et al. 1994, Shirako and Strauss 1994). Replicase polyprotein carrying inactive protease was shown to be completely replication-negative when the nsP4 is not cleaved and remains in the polyprotein. This phenotype could not be rescued even by introduction of additional nsP4 into the system (unpublished data). One can speculate whether the conformation of the replication complex is compromised or the activity of the core subunit of RdRP is not functional when it is still bound to the nsP3. It is quite possible that the cleaved nsP4 cannot be introduced to the replication complex when nsP4 is bound to the polyprotein. On the other hand, nsP4 needs to be present with P123 replication to take place. It has been shown earlier with SINV that nsP4 is needed (Rubach et al. 2009), which was also clearly seen in here (II, Fig 1).

The cleavage of nsP4 is the first step in polyprotein processing and transforms the polyprotein from its inactive form to active form by creating a complex P123/nsP4. This cleavage launches the replication process, which begins with the minus-strand synthesis. This complex, which is called an early polymerase complex, was indeed synthesizing high levels of minus-strand RNAs (II, Fig 1) but was less efficient in making the next transition; switch from minus-strand synthesis to plus-strand synthesis (Vasiljeva et al. 2003). Helicase and polymerase, as highly conserved and important enzymes in the viral replicase, were absolutely essential for replication. Both mutants (GNS and GAA) were shown to be completely replication-negative when introduced in the trans-replication system and no replication was observed at any level.

**Spherules are the sites of replication**

These data strongly suggest that replication of viral RNA is needed for induction of spherule structures on cellular membranes. It does not matter whether the replicase is inactivated by mutating important activities or if the template is crippled by e.g. deletions; the outcome is always the same. If there is no replication, meaning that not even the minus-strand is synthesized, no spherules arise. However, with completely inactive constructs it is impossible to say whether the template is recruited to the spherule? It cannot
be stated whether the recruitment itself would be enough for spherule formation, because even after extensive trials completely reliable system for RNA recruitment detection was not achieved (II, Fig 2). The aim was to purify cellular membranes by cell fractionation and gradient flotation in order to study the recruitment but the sensitivity remained a problem throughout the whole experimental setup. It was clearly seen that both strands concentrated on the membranous fractions with full replication and both strands were protected by the membranous spherule but not in soluble fraction. It was shown that a template artificially bound to plasma membrane was concentrating on membranous fractions but in the end the experiment was not sensitive enough to show the recruitment of the RNA to the cellular membranes with replication negative constructs.

It has been previously shown that the replicase proteins co-localize with the dsRNA and it has been shown that nascent RNA is located in the close proximity with spherule structures (Grimley et al. 1968, Kujala et al. 2001, Kopek et al. 2007, Spuul et al. 2011). We have shown here that spherules arise only when replication is initiated and at least minus strands have been synthesized. In further experiments it would be interesting to see whether the recruitment is enough to induce spherules.

The fact that the membranous spherule protects both strands from RNases and that relative high amounts of both strands can be rescued one can assume that the spherule structures arise before the viral RNA is in its double-stranded form. In addition, if spherules are mainly formed after the synthesis of minus-strand RNA, which is followed by translocation of the dsRNA into the spherules, few dsRNA molecules should be protected by the spherules.

Even though alphavirus replication has been extensively studied many decades it is not known where and when the replication complex is formed. There is no data describing whether the complex is formed directly on the plasma membrane or if some kind of pre complexes form in the cytoplasm prior to the membrane association. It is also not known how the spherules are precisely formed; simultaneously with the minus-strand synthesis, after the membrane association prior to the minus-strand synthesis or after synthesis of dsRNA. Another open question is how the membrane is shaped and how the spherule structure itself is formed?

To study the replication complex assembly and spherule formation in detail obviously requires either nanoscale EM studies, advanced light microscopy techniques e.g. Förster resonance energy transfer (FRET) or biochemical assays e.g. RNA pull-down. It would also be advantageous to create a reliable protocol for detecting the recruitment of the template RNA in order to solve the mysteries of spherule formation.
ACKNOWLEDGEMENTS

I would like to warmly thank all of you, who have been involved in this incredible journey through the world of Semliki Forest virus. My journey from the Institute of Biotechnology to the Faculty of Agriculture and Forestry has been the most interesting and full of happy moments. I would like to thank the Graduate Program in Biotechnology and Molecular Biology, currently the Integrative Life Science Doctoral Program for the financial support as well as for providing high level education and opportunities to meet new people and create contacts.

The biggest thanks definitely go to my supervisor Dr. Tero Ahola who made all this happen in the first place. Thank you for choosing me into this project which I have liked from the day one and thank you for all the support and advice during this amazing journey. Special thanks to Dr. Maarit Neuvonen, Dr. Pirjo Spuul and Dr. Giuseppe Balistreri for helping me during my first steps in the SFV lab. I would also like to express my sincere gratitude to Dr. Kirsi Helström who has always helped me, supported me and has always been there for me. I would like to thank all the other members of the SFV-group, past and present; Pasi, Yaseen, Leena, Antti, Ilkka, Hanski, Maija, Finny and Tania, you all have a special place in my heart. I will miss our conversations and all the fun we had during these years.

I would like to express my sincere gratitude to Dr. Gorben Pijlman for agreeing to be my opponent and to my thesis committee members Dr. Kristiina Mäkinen and Dr. Varpu Marjomäki, for their critical comments, advises, and help during my PhD studies. Besides my thesis committee, I would like to thank the reviewers of my thesis Dr. Johan Peränen and Dr. Petri Susi for great comments and advises. I would like to warmly thank also my custodian Prof Dennis Bamford for creating the most pleasant and relaxed atmosphere around the sometimes so stressful PhD life. I would like to acknowledge also Anita Tienhaara for helping me with all “graduation-related things” – you are so important.

My sincere thanks go also to the talented personnel in the Light Microscopy unit and in the Electron Microscopy unit. Special thanks to Dr. Eija Jokitalo, Mervi Lindman, and Arja Strandell for helping me so much with all EM-related issues and to Dr. Kimmo Tanhuanpää, Mika Molin, and Marko Crivaro for all the help and support in light microscopy related issues. Without your support and help this thesis would not have happened!

Last but not the least, I would like to thank my mom and dad for supporting me in all possible ways during my whole life and my brother for being there for me. Special thanks to Hannu and Oona – you are my life. Eli suuri kiitos isille ja äidille, kun olette tukenet ja auttanet kaikilla mahdollisilla tavoin läpi koko elämän ja kiitos Jono kaikesta. Suurin kiitos omalle pienelle perheelleni, Hannulle ja Oonalle – minun rakkaille.
REFERENCES


replication is supported by a reticulovesicular network of modified endoplasmic reticulum. PLoS Biol 6: e226.


**Park, E., and Griffin, D. E. 2009.** The nsP3 macro domain is important for Sindbis virus replication in neurons and neurovirulence in mice. Virology 388: 305-314.


**Pfeffer, M., and Dobler, G. 2010.** Emergence of zoonotic arboviruses by animal trade and migration. Parasit Vectors 3: 35.


The Roles of Template RNA and Replication Proteins in the Formation of Semliki Forest Virus Replication Spherules

KATRI KALLIO

Recent Publications in this Series

6/2016 Teija Ojala
Lactobacillus crispatus and Propionibacterium freudenreichii: A Genomic and Transcriptomic View

7/2016 César Araujo
Prostatic Acid Phosphatase as a Regulator of Endo/Exocytosis and Lysosomal Degradation

8/2016 Jens Verbeeren
Regulation of the Minor Spliceosome through Alternative Splicing and Nuclear Retention of the U1/U12-65K mRNA

9/2016 Xiang Zhao
HMGB1 (Amphoterin) and AMGO1 in Brain Development

10/2016 Tarja Pääkkönen (Jokinen)
Benign Familial Juvenile Epilepsy in Lagotto Romognolo Dogs

11/2016 Nora Hiivala
Patient Safety Incidents, Their Contributing and Mitigating Factors in Dentistry

12/2016 Juho Heinonen
Intravenous Lipid Emulsion for Treatment of Local Anaesthetic and Tricyclic Antidepressant Toxicity

13/2016 Riikka Jokinen
Activation of Innate Immune Responses by Toll-like Receptors and Influenza Viruses

14/2016 Sanna Mäkelä
Immunological Boosting and Personalization of Oncolytic Virotherapies for Cancer Treatment

15/2016 Mari Hirvinen
Screening of Marine Natural Products and Their Synthetic Derivatives for Antimicrobial and Antiproliferative Properties

16/2016 Sofia Montalvão
Bioinformatic Tools for Analysis, Mining and Modelling Large-Scale Gene Expression and Drug Testing Datasets

17/2016 Mpindi John Patrick
Bioinformatic and Genomic Approaches to Study Cardiovascular Diseases

18/2016 Hilla Sumanen
Bioinformatic and Genomic Approaches to Study Cardiovascular Diseases

19/2016 Oyediran Olulana Akinrinade
Bioinformatic and Genomic Approaches to Study Cardiovascular Diseases

20/2016 Prasanna Sakha
Development of Microfluidic Applications to Study the Role of Kainate Receptors in Synaptogenesis

21/2016 Neha Shrestha
Mesoporous Silicon Systems for Oral Protein/Peptide-Based Diabetes Mellitus Therapy

22/2016 Tanja Holopainen
Targeting Endothelial Tyrosine Kinase Pathways in Tumor Growth and Metastasis

23/2016 Jussi Leppilähti
Variability of Gingival Crevicular Fluid Matrix Metalloproteinase-8 Levels in Respect to Point-of-Care Diagnostics in Periodontal Diseases

24/2016 Niina Markkula
Prevalence, Predictors and Prognosis of Depressive Disorders in the General Population

DEPARTMENT OF FOOD AND ENVIRONMENTAL SCIENCES
DIVISION OF MICROBIOLOGY AND BIOTECHNOLOGY
FACULTY OF AGRICULTURE AND FORESTRY

DEPARTMENT OF BIOSCIENCES
DIVISION OF GENERAL MICROBIOLOGY
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES
DOCTORAL PROGRAMME IN INTEGRATIVE LIFE SCIENCE
UNIVERSITY OF HELSINKI