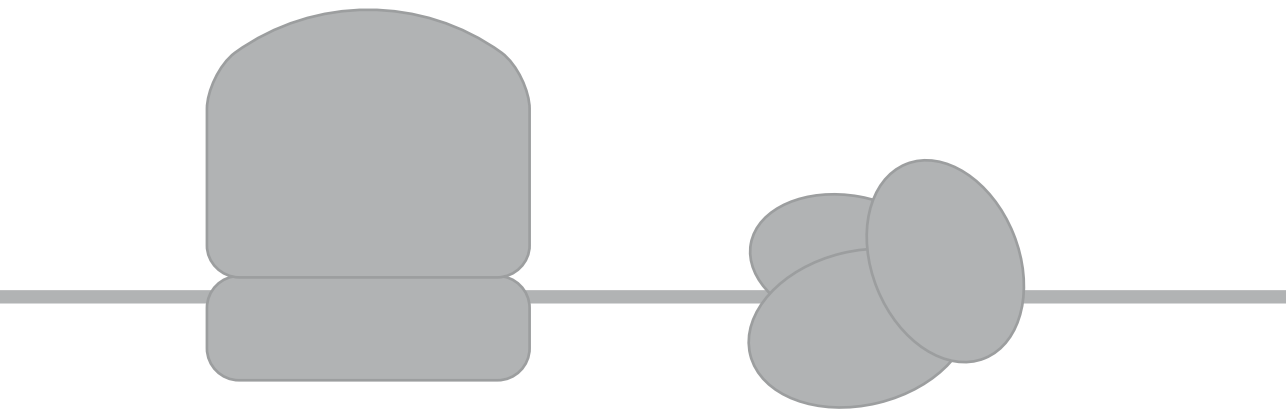
PART 2

–2/–1 programmed ribosomal frameshifting in arteriviruses



Chapter 5

General introduction part 2

CANONICAL TRANSLATION IN EUKARYOTES

Ribosomes convert messenger RNA (mRNA) nucleotide sequences into protein amino acid sequences; this process is called translation. Eukaryotic translation is a complex process that is highly regulated and involves many different factors. It can be subdivided into three phases; initiation, elongation and termination/ribosome recycling [307-311].

Translation initiation

The first step in translation initiation is the formation of a ternary complex (TC) that consists of initiator methionyl-transfer RNA (met-tRNA$_i$) and the GTP-bound form of eukaryotic initiation factor 2 (eIF2). The TC assembles with the small (40S) ribosomal subunit and eIF1, 1A, 3 and 5 to form the 43S preinitiation complex (PIC). Initiation factors eIF4B, 4H and 4F (consisting of 4A, 4E and 4G) bind the 5′ end of the mRNA that in eukaryotes contains a 7-methylguanosine (m7G) cap [309, 312]. Eukaryotic translation is dependent on this 5′ cap [313]. The 3′ end of the mRNA is bound by the poly(A)-binding protein (PAPB). The 5′ and 3′ ends of the mRNA are brought together through the interaction of eIF4G with both eIF4E and PAPB, resulting in a circular activated messenger ribonucleoprotein (mRNP). The PIC is then recruited to the activated mRNP at the 5′ end of the mRNA [309, 312]. After the PIC has been loaded onto the 5′ end, it starts to scan the mRNA in a linear base-by-base fashion until it encounters a start codon in a favorable sequence context (Kozak consensus sequence). In eukaryotes the start codon generally is AUG, although there are some exceptions to this rule [313]. Recognition of the start codon halts mRNA scanning by the PIC. The start codon is recognized through base-pairing between the anticodon of Met-RNA$_i$ and the AUG codon in the peptidyl-tRNA (P)-site of the 40S subunit. This results in the dissociation of eIF1 from the ribosome which triggers full accommodation of the tRNA in the P-site [314] and conversion of eIF2 to its GDP-bound state [315]. Release of GDP-bound eIF2 and most other initiation factors from the PIC and binding of GTP-bound eIF5B to the complex facilitate large (60S) subunit joining to form the 80S initiation complex (IC), after which eIF5B hydrolyzes and dissociates. The last subunit to leave the IC is eIF1A, although eIF3 possibly remains associated with the ribosome during elongation [309, 312]. A schematic representation of the translation initiation phase is depicted in Figure 1.

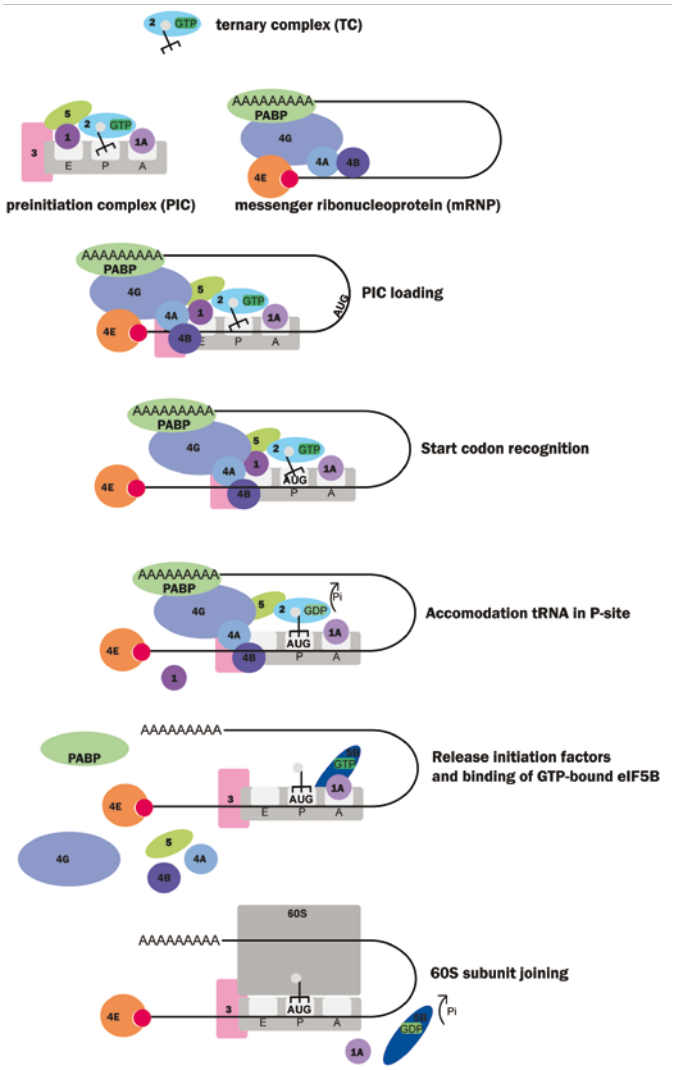


Figure 1: Translation initiation phase. A ternary complex (TC) consisting of initiator methionyl-transfer RNA (met-tRNA) and the GTP-bound form of eIF2 is formed. The TC assembles with the small (40S) ribosomal subunit and eIF1, 1A, 3 and 5 to form the 43S preinitiation complex (PIC). Initiation factors eIF4B, 4H and 4F (consisting of 4A, 4E and 4G) bind the 5′ end of a 7-methylguanosine capped mRNA. The 3′ end of the mRNA is bound by the poly(A)-binding protein (PABP) and the 3′ and 5′ ends of the mRNA are brought together, resulting in a circular activated messenger ribonucleoprotein (mRNP). The PIC is recruited to the activated mRNP at the 5′ end of the RNA and starts to scan until it encounters an AUG start codon. Start codon recognition halts mRNA scanning by the PIC. The start codon is recognized through base-pairing between the anticodon of the Met-tRNA$_i$ and the AUG codon in the ribosomal P-site. This results in dissociation of eIF1 and conversion of eIF2 to its GDP-bound state. Most of the initiation factors dissociate from the ribosome and GTP-bound eIF5B binds. The 60S ribosomal subunit joins to form the 80S initiation complex (IC) which results in eIF5B hydrolysis and dissociation. eIF1A is the last subunit to leave the IC, although it is possible that eIF3 remains associated during the elongation phase.

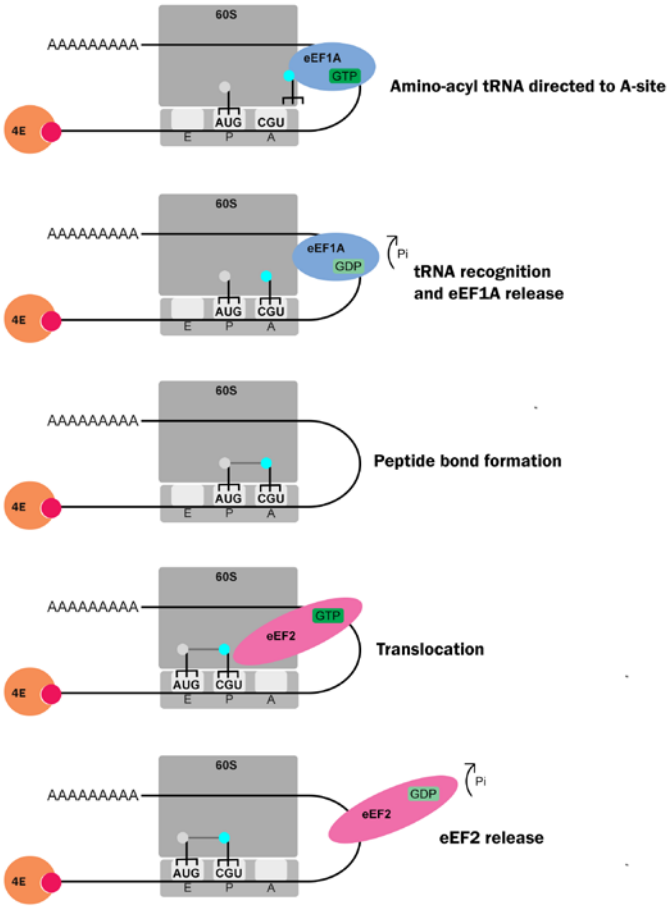


Figure 2: Translation elongation phase. At the start of the elongation phase the 80S ribosome is positioned on the mRNA with the Met-tRNA$_i$ paired to the start condon in the P-site. The A-site contains the second codon of the ORF. Amino-acyl tRNAs are bound by eEF1A in a GTP-dependent manner and directed to the A-site. tRNA recognition results in GTP hydrolysis which releases the elongation factor. Peptide bond formation with the peptidyl-tRNA in the P-site occurs fast. Translocation to the E- and P-site requires binding of eEF2 and GTP. Conformation changes in eEF2 and P$_i$ release following GTP hydrolysis allow movement of tRNA and mRNA. eEF2 is released from the ribosome and the A-site becomes available for the next aminoacyl-tRNA.

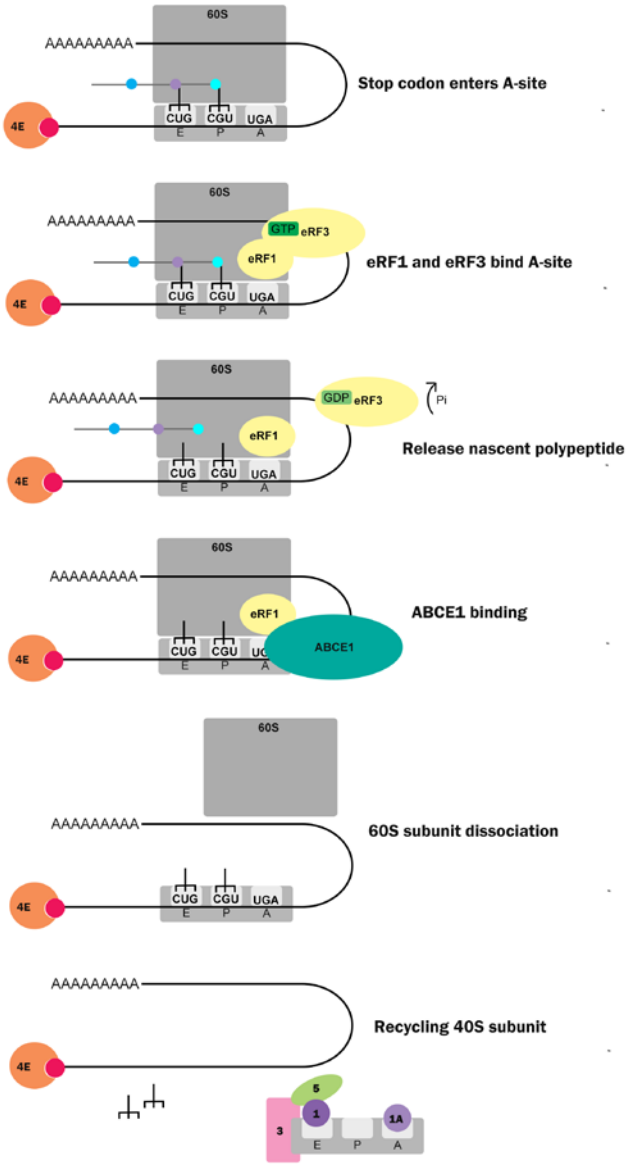


Figure 3: Translation termination/ribosome recycling phase. When one of the stop codons enters the ribosomal A-site, a ternary complex of eRF1 and eRF3 with GTP binds. GTP hydrolysis by eRF3 causes eRF1 to trigger hydrolysis of the polypeptidyl-tRNA in the peptidyl transferase site (PTC) which releases the nascent polypeptide. The 80S ribosome remains associated with the mRNA, the deacetylated tRNAs and eRF1. ABCE1 binds the complex with promotes 60S subunit dissociation. Release of tRNA and 40S subunit is mediated through binding of eIF1A, 1 and 3 to the 40S subunit. These subunits are reused to initiate translation.

Translation elongation

The ribosome has three tRNA binding sites, the acceptor (A)-site for aminoacyl-tRNA, the P-site for peptidyl-tRNA and the exit (E)-site for deacylated tRNA [316]. At the start of elongation, the 80S ribosome is positioned on the mRNA with the Met-tRNA$_i$ anticodon base-paired to the start codon in the P-site. The A-site contains the second codon of the ORF. Amino-acyl tRNAs are bound by eukaryotic elongation factor 1A (eEF1A) in a GTP-dependent fashion and are directed to the A-site of the ribosome. tRNA codon recognition results in GTP hydrolysis by eEF1A, which releases the elongation factor and enables accommodation of the amino-acyl tRNA into the A-site [307, 308]. Peptide bond formation with the peptidyl-tRNA in the P-site occurs fast as the peptidyl transferase center (PTC) in the large ribosomal subunit positions the substrates for catalysis [317]. After peptide bond formation the two ribosomal subunits rotate so that the acceptor ends of the tRNAs are placed in the E- and P-sites while the anticodon loops remain in the P- and A-sites [318]. Translocation to the E and P-sites requires binding of eEF2 (a GTPase) and GTP. Conformational changes in eEF2 and P_i release following GTP hydrolysis allow movement of tRNA and mRNA and locks the subunits in the posttranslocation state [240]. eEF2 is released from the ribosome and the A-site becomes available for binding of the next aminoacyl-tRNA [308]. A schematic representation of the translation elongation phase is depicted in Figure 2.

Translation termination/ribosome recycling

Termination of translation requires two release factors (RFs), eRF1 and eRF3, that form a ternary complex with GTP [319]. When one of the stop codons (UGA, UAG, UAA) enters the ribosomal A-site this complex binds, with eRF1 being responsible for codon recognition [320]. GTP hydrolysis by eRF3 causes eRF1 to trigger hydrolysis of the polypeptidyl-tRNA in the PTC which results in release of the nascent polypeptide [311]. After release of the protein product, the 80S ribosome remains associated with the mRNA, the deacylated tRNA and eRF1. The ribosome dissociates into a free 60S subunit and a 40S subunit that is still associated with the mRNA and tRNA, a process promoted by the ATP-binding cassette protein ABCE1 [321]. Release of tRNA and mRNA from the 40S subunit can be mediated by eIF1A, eIF1 and eIF3 with its loosely associated eIF3j subunit. 40S subunits bound to eIF1A, eIF1 and eIF3 are reused to initiate translation [310, 311]. A schematic representation of the translation termination/ribosome recycling phase is depicted in Figure 3.

Non-canonical translation of RNA virus genomes

Viruses do not encode their own ribosomes and, therefore, for their protein synthesis all viruses are completely dependent on the translational machinery of the host cell. Especially for RNA viruses, this poses a challenge since their genomes are usually short

and often consist of a single strand of RNA from which all replicative, accessory and structural proteins need to be expressed. Additionally, many RNA viruses replicate in the cytoplasm and as a result do not have access to cellular systems for mRNA capping and polyadenylation that reside in the nucleus. Some viruses solve this problem by encoding their own capping enzymes or by snatching caps from cellular mRNAs [322]. Other viruses have evolved non-canonical translation strategies to circumvent the limitations of canonical 5'-end dependent eukaryotic translation, which allows synthesis of just a single protein from a single mRNA. Many +RNA viruses even employ more than one of these mechanisms to expand the possibilities for gene expression from their compact genomes [323-325]. The non-canonical translation strategies that are discussed below are depicted schematically in Figure 4.

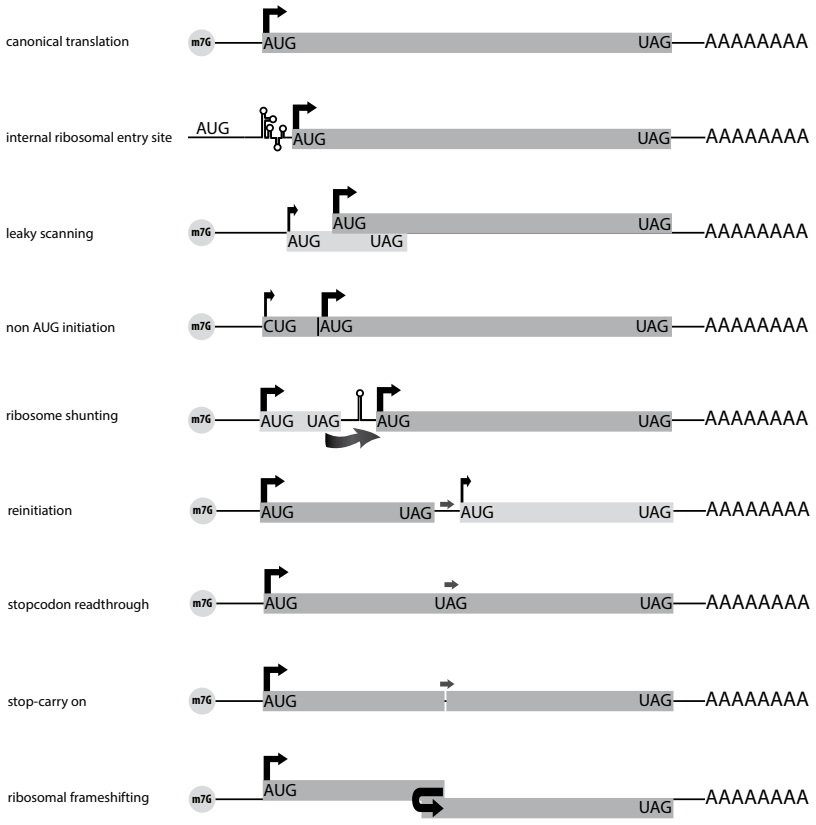


Figure 4: Non-canonical translation mechanisms. Eukaryotic canonical translation is shown at the top. ORFs are indicated as grey bars on the mRNA. The darker grey indicates the main route taken by translating ribosomes. Arrows at the start of the ORF indicate ribosome translation initiation. Horizontal arrows indicate where the ribosome moves non-canonically.

IRES

Several +RNA virus groups employ cap-independent translation via an internal ribosomal entry site (IRES). An RNA structure, usually located in the 5′-untranslated region (UTR) of the genome, replaces the cap and, depending on the type of IRES, some or all of the initiation factors [326], and serves as a binding site for ribosomes. Some viruses can initiate translation at multiple AUGs in the same or different reading frames from one IRES [327, 328]. Other viruses express two polyproteins from separate IRESes [329].

Leaky Scanning

In principle, ribosomes scan mRNAs in a linear fashion, starting at the 5′ end and initiating translation at the first AUG start codon they encounter. However, when the context for recognition of that AUG start codon is not optimal (GCCRCCaugG, R is A or G, -3 R and +4 G are most important for efficient initiation) ribosomes can continue scanning and initiate translation at a downstream AUG instead. If the upstream AUG occurs in a very weak context nearly all ribosomes will continue scanning, if it is only somewhat suboptimal most ribosomes will initiate translation [330]. Leaky scanning is widely employed by viruses and allows the translation of two proteins from a single mRNA in the same or a different reading frame [323].

Initiation at non-AUG codons

In a strong context, near-cognate codons, such as CUG and ACG, can be recognized by the Met-tRNA$_i$ to initiate translation. Non-AUG initiation is often inefficient and, as a result, frequently occurs in combination with leaky scanning [330]. In viruses these combined mechanisms can enable the expression of 3 or 4 different proteins from a single strand of RNA [323].

Ribosome shunting

Ribosome shunting allows ribosomes to translate downstream ORFs in a 5′ cap-dependent manner. The RNA usually contains a first short ORF followed by a stem-loop structure and a downstream main ORF. Translation of the first ORF is initiated at the start codon and terminates just before the stem-loop. It is thought that the small subunit of the ribosome retains certain initiation factors during translation of the first ORF. Ribosomes are then capable of bypassing the stem-loop and resume scanning at the landing site 3′ of the stem-loop [311, 331].

Reinitiation

Certain initiation factors (most likely eIF3 and eIF4G) remain briefly associated with the 40S ribosomal subunit after joining of the 40S and 60S subunits. If translation of a short ORF is terminated before these factors are released, the 40S subunit can resume scan-

ning after dissociation of the 60S subunit. The rescanning ribosome initially does not have an associated TC and successful reinitiation of translation depends on the distance between stop codon and reinitiation site and TC availability [311]. In mammalian systems it is very rare for translation reinitiation to occur after translation of a long ORF. Mammalian caliciviruses have bicistronic subgenomic mRNAs with a short overlap region of the two ORFs. Translation reinitiation at the second ORF is dependent on translation of the upstream ORF. It requires the presence of a sequence element upstream of the second initiation site that is called the "termination codon upstream ribosome-binding site" (TURBS) [332]. Motifs within the TURBS sequence hybridize with 18S rRNA [333]. The TURBS also interacts with eIF3 and 40S ribosomal subunits. It is thought that the interaction with the TURBS secures 40S subunits that have terminated translation of the upstream ORF to the mRNA. Subsequent recruitment of the TC results in translation initiation to express the downstream ORF [334].

Stop codon read-through

The occurrence of a stop codon signals translation termination and induces the release of the polypeptide from the ribosome. However, the efficiency of translation termination is influenced by the type of stop codon (UGA and UAG are more leaky than UAA) and its context. When stop codon read-through occurs the stop codon is decoded by a near-cognate or "suppressor" tRNA, and translation continues until the next stop codon enters the ribosomal A-site. The UGA and UAA codons can induce the incorporation of tryptophan, arginine or cysteine. The UAG codon can induce the incorporation of tyrosine, glutamine or leucine [335]. Read-through results in the synthesis of a C-terminally extended polypeptide. Read-through occurs at a defined frequency and many viruses use it to express their polymerase at a lower level than other replicative proteins or to produce an extended version of their coat protein [323, 335].

Stop-carry on

Many viruses translate proteins from a single ORF as a polyprotein that is subsequently cleaved into the individual protein products by host and/or viral proteases. Stop-carry on is an alternative to proteolytic cleavage that also allows the expression of multiple proteins from a single ORF with near to 100% efficiency. The amino acid motif D(V/I)ExNPGP and the upstream amino acids that are present in the ribosome exit tunnel prevent formation of the peptide bond between glycine and the final proline. Instead, the nascent peptide is released from the ribosome after which the proline tRNA can bind the A-site and translation continues with proline as the N-terminal amino acid of the downstream-encoded product [210, 323].

Ribosomal frameshifting

In certain contexts, ribosomes can shift 1 or 2 nucleotides and continue translation in a different reading frame, which is called programmed ribosomal frameshifting (PRF). –1 PRF is widely used by +RNA viruses and allows them to produce proteins at a defined ratio or translate transframe proteins that share the N-terminal sequence [323]. A bipartite signal in the mRNA determines the frameshifting efficiency. A slippery sequence X_XXY_YYZ (_ separates 0 frame codons, X is any 3 identical nucleotides, YYY is AAA or UUU, Z is A, C or U) on which the ribosome backs up 1 nucleotide and a downstream stimulatory element, in most cases either a pseudoknot or a very stable RNA stem-loop structure [336]. The distance between slippery sequence and downstream element (6-8 nt) is also important [336]. It has been proposed that the resistance of the stimulatory RNA structure to unwinding induces tension in the mRNA. This causes ribosomes to pause at the slippery sequence with XXY in the P-site and YYZ in the A-site. Unpairing of mRNA and tRNA could release this tension, which then causes the ribosome to move to the –1 frame while still retaining base-pairing in the non-wobble positions [337, 338]. +1 and –2 frameshifting mechanisms are less common and their efficiency is usually very low [323].

PRRSV

The work in this second part of my thesis focuses on porcine reproductive and respiratory syndrome virus (PRRSV), a single-stranded enveloped +RNA virus belonging to the family *Arteriviridae* of the order *Nidovirales [339]*. For the expression of its polycistronic genome PRRSV and other arteriviruses employ a combination of strategies, including multiple non-canonical translation mechanisms.

PRRSV currently is one of the economically most important viruses in the swine industry [340]. In the late 1980's, distantly related PRRSVs, which are likely to be reclassified as separate viral species in the near future, simultaneously emerged in Europe [10] and the United States [11, 341]. The two current genotypes, type 1 (European) and type 2 (North-American), probably share a common ancestor but diverged extensively, retaining a nucleotide identity of only 60-70% [342, 343]. Within each genotype the nucleotide similarity is >80% [344].

Order Nidovirales

The viruses belonging to the *Nidovirales* have a similar genome organization and expression strategy and their key replicative enzymes, including the RNA-dependent RNA polymerase (RdRp) and helicase, presumably share a common ancestor [345, 346]. The order *Nidovirales* currently contains four families, the *Coronaviridae*, *Roniviridae*,

Mesoniviridae and *Arteriviridae* [347, 348]. Its coronavirus branch includes several human pathogens, in particular four "established" human coronaviruses (HCoVs), which generally cause common colds [349], and highly pathogenic zoonotic agents like severe acute respiratory syndrome coronavirus (SARS-CoV) [2, 4] and Middle East respiratory syndrome coronavirus (MERS-CoV) [350].

The arterivirus family currently consists of four recognized species that infect mammals: equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) of mice, simian hemorrhagic fever virus (SHFV) and PRRSV [339]. The recently discovered wobbly possum disease virus (WPDV) will most likely be classified as an arterivirus [351] and several newly discovered monkey viruses appear to be distantly related to SHFV [352].

The PRRSV replicative cycle

The 13-16 kb arterivirus genome is encapsidated by the N protein and wrapped in a lipid envelope that is cell-derived and contains seven viral membrane proteins. The two major envelope proteins are the ORF5-encoded glycoprotein (GP5) and the membrane (M) protein. The five minor envelope proteins are GP2, GP3, GP4, envelope (E) protein and the ORF5a product [353-355]. For PRRSV all envelope proteins are required for the production of infectious virus particles [356, 357]. Cell entry occurs through clathrin-mediated endocytosis. Following endosome acidification and membrane fusion, the viral nucleocapsid is released into the cytosol [358, 359]. It is still uncertain which of the viral envelope proteins mediates fusion but host receptors CD163 and CD169 are likely required for PRRSV internalization and uncoating [360].

Arterivirus genomes are poly-adenylated and presumed to have a 5′ cap. Upon cell entry the PRRSV replicase proteins are expressed from two 5′ ORFs (ORF1a and ORF1b) as two precursor polyproteins (pp1a and pp1ab) that are proteolytically processed into 14 individual nonstructural proteins (nsps) by four viral proteases residing in nsp1α, nsp1β, nsp2 and nsp4 [361-364]. There is a small overlap between ORF1a and ORF1b and translation of pp1ab depends on a –1 PRF event that is directed by a "slippery sequence" and a downstream pseudoknot structure [342, 365]. The structural and accessory proteins are translated from the 3′ORFs 2-7 as a nested (Latin nidus = nest) set of 5′- and 3′-coterminal subgenomic mRNAs (sgmRNAs) [218, 366]. The termini of most of the ORFs 2a to 7 overlap with neighboring genes and several sets of proteins are translated from the same sgmRNA, presumably through leaky scanning [354, 366, 367].

The arteriviral replication and transcription complex (RTC) is associated with membrane structures. These are complex networks of modified ER-derived structures in the perinuclear region of the cell that consist of double-membrane vesicles [368-370]. The RTC first synthesizes full-length and subgenome-length minus strand RNAs, which are used as templates for the synthesis of new genomes and the sgmRNAs, respectively [371]. Newly synthesized genomes become encapsidated by N protein [372]. Subse-

quently, nucleocapsids are enwrapped by membranes from smooth ER and/or Golgi complex [373, 374]. Virions are then transported to the plasma membrane by the exocytic pathway and released from the cell [375]. A schematic overview of the arterivirus replicative cycle is shown in figure 5.

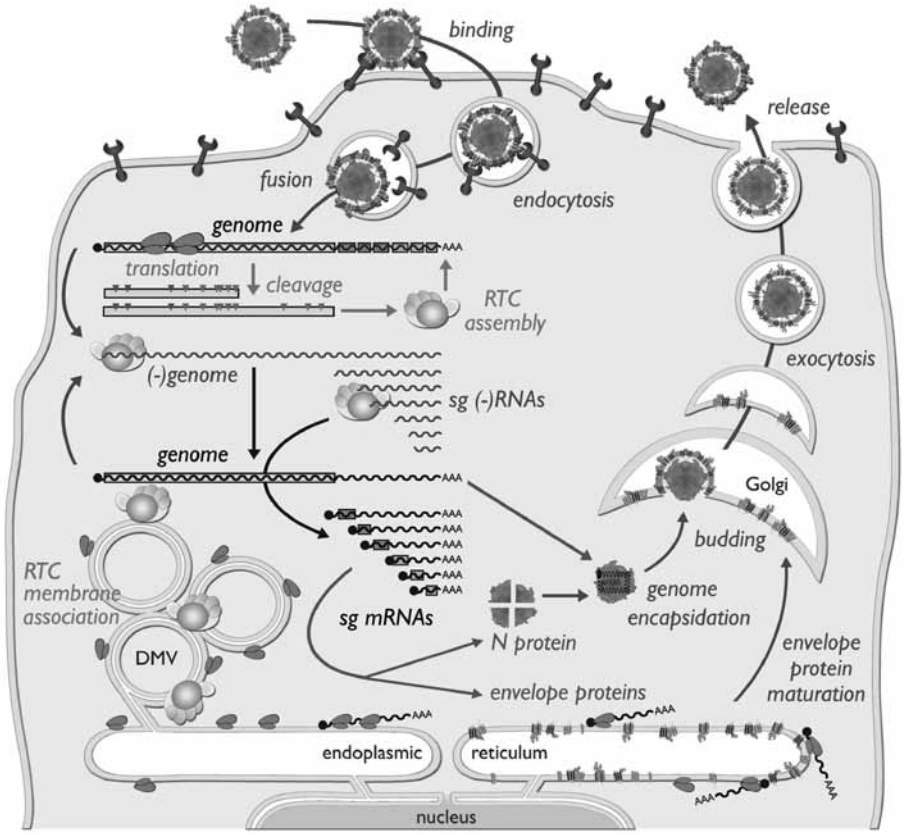


Figure 5: Overview of the arterivirus replicative cycle. Following entry by receptor-mediated endocytosis and endosomal membrane fusion the nucleocapsid is released into the cytosol. Genome translation yields replicase polyproteins pp1a and pp1ab that are cleaved by internal proteinases. The viral nonstructural proteins assemble into a replication and transcription complex (RTC) that first engages in minus-strand RNA synthesis. Both full-length and subgenome-length minus strands are produced, the latter serving as templates for the synthesis of sg mRNAs required to express the structural protein genes, which reside in the 3′-proximal quarter of the genome. Novel genomes are packaged into nucleocapsids that become enveloped by budding from smooth intracellular membranes, after which the new virions leave the cell using the exocytic pathway. For further details, see text. Reprinted with permission from [392].

Pathogenesis of PRRSV infection in swine

PRRSV can be transmitted between swine through semen, milk, direct contact and aerosol transmission [376-378]. Type 1 PRRSV mainly causes reproductive failure in sows late during gestation, while type 2 also causes respiratory disease in growing pigs [379]. The virulence of different PRRSV isolates can vary from very mild to highly pathogenic [380, 381]. Primary target cells for viral replication are porcine alveolar macrophages (PAMs) [10, 382] and the virus replicates and persists mostly in lungs and lymphoid organs [383, 384]. The acute phase of PRRSV infection is usually characterized by a high viral load and the presence of clinical symptoms that can last up to one month post infection. This is followed by a phase with low replication levels of persisting virus, until the virus is eventually cleared, which can take up to 150 days [385]. Both innate and adaptive immune responses against PRRSV are generally weak. PRRSV modulates the innate immune response by suppressing the production of type I interferons [386, 387] and regulating the expression of other cytokines [388]. The production of neutralizing antibodies occurs only late and their levels are generally low [385]. Development of cell-mediated immunity is weak and slow [388, 389]. The immunity acquired after clearance is often not effective against reinfection with a heterologous strain [389]. Vaccines against PRRSV are available, but are not completely effective [390, 391].

OUTLINE PART 2

In part 2 of my thesis, the identification of a new PRF mechanism in arteriviruses is described, which was explored in detail using PRRSV. **Chapter 6** presents the discovery of this mechanism, a highly efficient form of –2 PRF that ensures the expression of a previously unknown transframe protein, nsp2TF, that shares its N-terminal sequence with nsp2 but has a different C-terminal domain. In infected cells, nsp2 and nsp2TF localize to different cellular compartments and virus mutants that are incapable of expressing nsp2TF are seriously crippled. In **chapter 7** –2 PRF in PRRSV is shown to depend on the expression of the PRRSV nsp1β protein. This chapter further shows that –1 PRF takes place at the same frameshift site, resulting in the translation of a truncated form of nsp2, nsp2N. **Chapter 8** shows that –2/–1 PRF in PRRSV is also dependent on the presence of specific host proteins, namely poly (C) binding proteins (PCBP) 1 and 2, which appear to form a PRF-inducing protein complex together with nsp1β. In **chapter 9** the findings of part 2 of this thesis are summarized.