Universiteit Leiden
The Netherlands

A +RNA virus diptych : Chikungunya virus-host interactions and arteriviral programmed ribosomal frameshifting

Treffers, Emma Elisabeth

Citation

Treffers, E. E. (2015, December 16). *A +RNA virus diptych : Chikungunya virus-host interactions and arteriviral programmed ribosomal frameshifting*. Retrieved from https://hdl.handle.net/1887/37107

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/37107

Note: To cite this publication please use the final published version (if applicable).

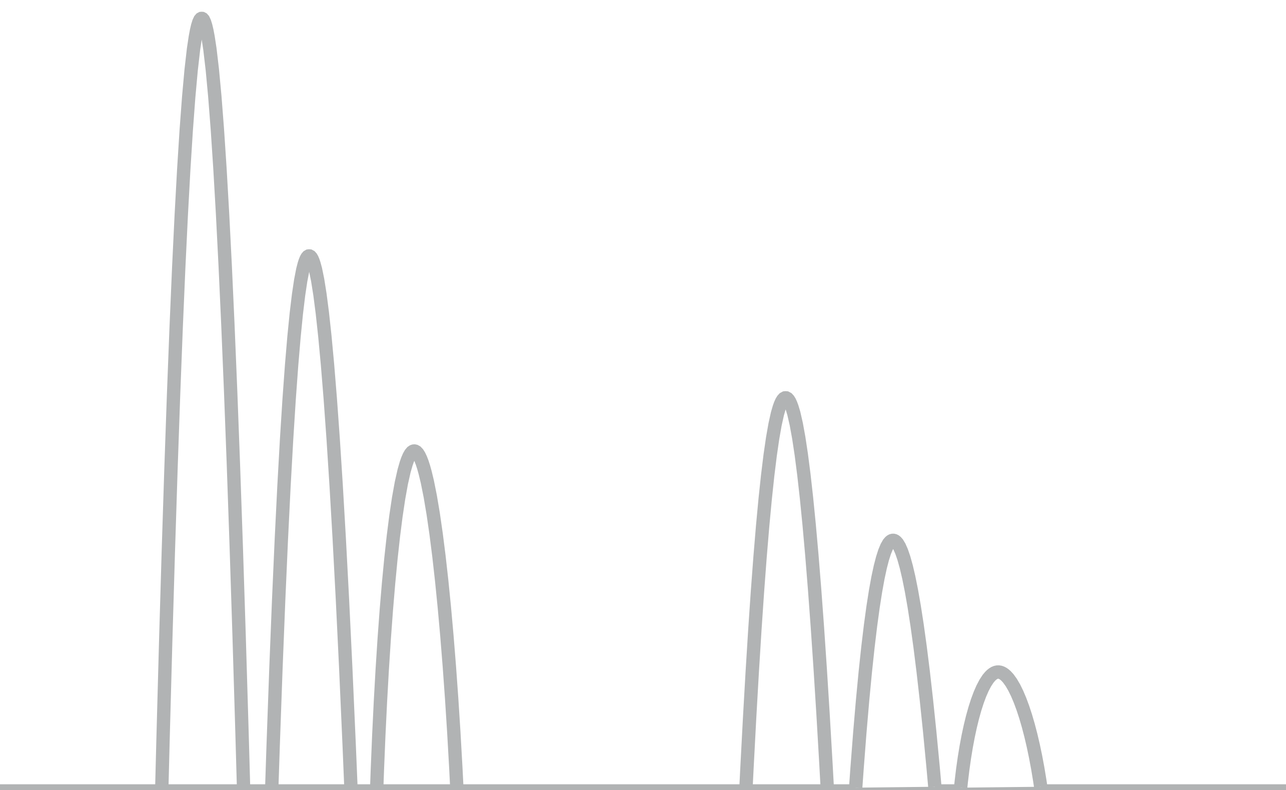
Cover Page

The handle http://hdl.handle.net/1887/37107 holds various files of this Leiden University dissertation.

Author: Treffers, Emma Elisabeth (Emmely)
Title: A +RNA virus diptych : Chikungunya virus-host interactions and arteriviral programmed ribosomal frameshifting
Issue Date: 2015-12-16

PART 1

Quantitative proteomics of chikungunya virus



Chapter 1

General introduction part 1

+RNA VIRUSES

Viruses are obligate intracellular parasites that are highly dependent on many cellular components and processes. Using genome type and replication strategy, the Baltimore classification system distinguishes double-stranded (ds)DNA, single-stranded (ss)DNA, dsRNA, positive (+)ssRNA, and negative (-)ssRNA genomes. Additionally, there are +ssRNA viruses that replicate via a DNA intermediate and dsDNA viruses that replicate through a ssRNA intermediate [1]. The +RNA viruses are the largest virus group, which contains many important human and animal pathogens, such as SARS coronavirus [2-4], poliovirus [5], hepatitis C virus [6], dengue virus [7], chikungunya virus [8], foot and mouth disease virus [9] and porcine reproductive and respiratory syndrome virus [10, 11].

All +RNA viruses encode an RNA-dependent RNA polymerase (RdRp) to replicate their genomes and several virus groups transcribe subgenomic RNAs to express the genes encoding their structural and accessory proteins [12]. +RNA viral genome replication is associated with modified cellular membranes to which components of the viral replication complex are anchored. The origin of these membranes differs between virus groups, but membrane-associated replication might serve the same fundamental purpose(s), as it is a general feature of all eukaryotic +RNA viruses [13].

Chikungunya virus

The work described in this first part of my thesis focuses on chikungunya virus (CHIKV). CHIKV is a reemerging arthropod-borne human pathogen that was first discovered in Tanzania in 1952, during an outbreak on the Makonde Plateau in the Southern region of Tanganyika [8]. Before 2005 only occasional small-scale outbreaks were reported. CHIKV reemerged around 2005 on the east coast of Africa and several Indian Ocean islands, and subsequently spread across the Asian continent while infecting millions of people [14, 15]. This outbreak was closely associated with the occurrence of a single point mutation in the viral genome, resulting in an A226V substitution in the CHIKV envelope protein E1. This mutation increased the epidemic potential of CHIKV, as it allowed the virus to be more efficiently spread by a new vector, *Aedes albopictus* (Asian tiger mosquito). The global distribution of *Ae. albopictus*, which includes urban areas in southern Europe and the USA, is wider than for *Ae. aegypti*, which was the main vector before 2005 [16]. CHIKV spread to the Americas around the end of 2013 and has since caused an explosive outbreak in the Caribbean and South/Central America [17]. Infected travelers have returned to Europe, Australia, the USA and Canada [18-25]. In countries with *Ae. aegypti* or *Ae. albopictus* populations this poses the risk of establishing new CHIKV reservoirs [26, 27], and locally-transmitted infections have already occurred in e.g.

Italy, France and the USA in 2007, 2010 and 2014 [28-30]. During the outbreak in Italy over 200 confirmed cases were reported [28].

In the Makonde language Chikungunya means "that which bends up", which refers to the posture of people suffering from the disease due to the severe and persisting polyarthralgia that characterizes the disease [8]. Other symptoms include a short episode of high fever, myalgia, headache and rash. Asymptomatic infections range from 3-25% [31-34]. Therapy is limited to supportive care as antiviral treatments and vaccines for CHIKV are still being developed [35, 36]. The development of antiviral therapies and vaccines is hampered by the rapidly occurring resistance, caused by the generally high mutation rate in RNA viruses [37-40].

Alphavirus genome organization and expression

CHIKV has a single stranded +RNA genome of 12kb and belongs to the alphavirus genus of the Togavirus family [41]. The alphavirus genome contains two large open reading frames (ORFs) that are translated into polyproteins. The genome also serves as the mRNA template for translation of the nonstructural proteins nsP1-4 that are encoded by the first ORF. Translation of nsP4 is dependent on read-through of an opal stop codon at the end of the nsP3-coding region [42]. The polyprotein is proteolytically processed by a protease domain residing in nsP2 [43, 44]. The structural polyprotein is encoded by the second ORF and is expressed from a subgenomic mRNA. The structural polyprotein is processed into the capsid (C) protein, the envelope proteins E1, E2 and E3, and the 6k ion channel protein by a protease domain in C and host proteases in the exocytic pathway [45, 46]. A third small ORF is embedded within the sequence encoding 6k and is translated after a –1 ribosomal frameshift resulting in the synthesis of a transframe (TF) protein that shares the N-terminal sequence with 6k and with the C-terminal amino acids encoded by the -1 ORF [47].

Alphavirus replicative cycle in mammalian cells

The alphavirus nucleocapsid contains a single copy of genomic RNA and is enveloped in a lipid bilayer derived from the host plasma membrane. This bilayer contains glycosylated E1 and E2 heterodimers that are assembled into 80 trimeric spikes [48]. It is still unknown which cellular receptor(s) is required for CHIKV attachment to the host cell. Alphavirus cell entry is initiated by interaction of E2 with cellular proteins. Generally, alphavirus entry is dependent on clathrin-mediated endocytosis and E1-mediated fusion with endosomal membranes [49]. After fusion, the nucleocapsid is released into the cytoplasm and is almost immediately uncoated to release the viral RNA for translation of the nonstructural polyprotein [50].

The nonstructural polyprotein is processed by nsP2 in a sequential manner. The nsP3/nsP4 junction is cleaved first, followed by the nsP1/nsP2 junction and finally the nsP2/

nsP3 junction. This sequential processing regulates the activity of RNA synthesis by the replication and transcription complex (RTC) consisting of the nonstructural proteins [51, 52]. The precursor P123 and nsP4 form a complex that synthesizes -strands exclusively, while the nsP1/nsP2 cleavage converts the complex into one that also synthesizes +strands. After the nsP2/nsP3 junction has been cleaved, the complex can only synthesize +strand genomes and the 26S subgenomic mRNA [53]. Viral replication takes place in spherules at the plasma membrane and in cytoplasmic vacuoles that are derived from modified lysosomal and endosomal membranes [54, 55].

nsP1 has guanyl transferase and guanine-7-methyltransferase enzymatic activities required for capping of viral genomes and subgenomic mRNAs [56-58]. This protein is membrane-associated and is thought to anchor the RTC to the modified membrane structures [59]. Besides the protease activity that resides in a C-terminal domain, nsP2 also has RNA helicase [60] and RNA triphosphatase/nucleoside triphosphatase activities [61, 62] residing in the N-terminal part of the protein. nsP2 is also responsible for the transcriptional and translational shut-off of the host cell during alphavirus infection [63]. nsP2 localizes to cytoplasmic foci and the nucleus [64, 65]. The exact functions of nsP3 still remain unclear but the protein is required for RNA synthesis [43, 66] and the macro domain of CHIKV nsP3 has ADP-ribose 1″-phosphate phosphatase and RNA binding activity [67]. During infection nsP3 localizes to cytoplasmic foci where it might inhibit stress granule formation [65, 68, 69]. nsP4 is the RNA dependent RNA polymerase (RdRp) that synthesizes both + and –stranded RNA [43, 70]. This protein might also have terminal adenylyltransferase (TATase) activity which could be involved in the maintenance of the poly(A)tail [71]. Compared to the other nsPs, nsP4 is very short lived due to ubiquitin-mediated degradation, which is one of the ways by which the nsP4 concentration is regulated during infection [72].

Processing of the alphavirus structural polyprotein C-E3-E2-6k-E1 starts by release of C through autoproteolysis [45, 73]. Subsequently, a signal sequence in the new N-terminus of the polyprotein initiates translocation of pE2 (the E3-E2 precursor [74]) across the endoplasmatic reticulum (ER) membrane [75]. A hydrophobic stretch near the C-terminus of E2 anchors the protein in the ER membrane [76]. The insertion signals for 6k and E1 are located after the E2 anchor sequence and in the C-terminal domain of 6k, respectively. The E2-6k and 6k-E1 junctions are cleaved by signalases in the ER lumen after insertion [46]. In the ER lumen and during transport through the Golgi apparatus pE2 and E1 become glycosylated and palmitoylated [77-79]. pE2 and E1 form a heterodimer in the ER and during transport of the pE2-E1 complex through the *trans* Golgi complex to the plasma membrane pE2 is cleaved into E3 and E2 by furin, a host protease [80, 81]. CHIKV E3 is not incorporated into virions [82]. The alphavirus genomic RNA contains a packaging signal that is recognized by C, which results in the subsequent oligomerization of capsid proteins to generate the nucleocapsid [83, 84]. Budding occurs at the plasma

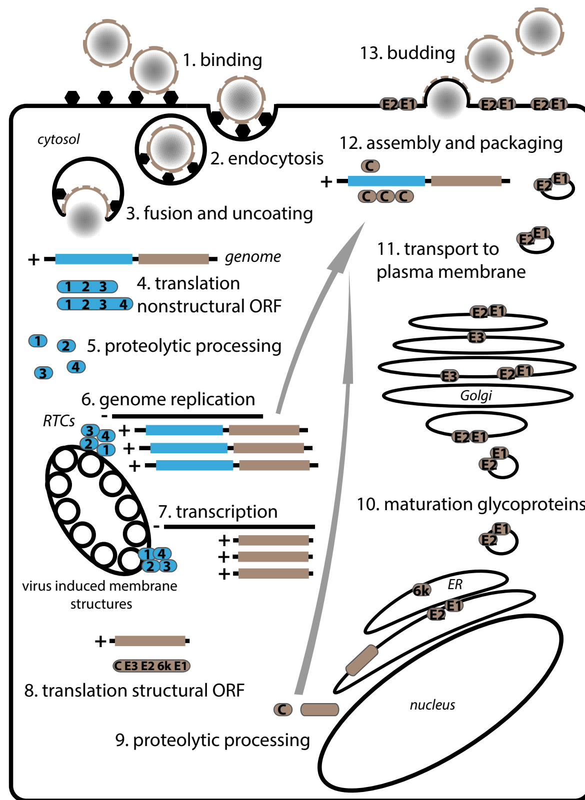


Figure 1: Alphavirus replicative cycle. 1. Alphavirus particles consist of a nucleocapsid that is enveloped in a host plasma membrane-derived lipid bilayer. Alphavirus entry is initiated by binding of the E2 glycoprotein to cellular membrane proteins. **2.** Alphavirus particles generally enter the cell through clathrin-mediated endocytosis. **3.** The E1 glycoprotein mediates fusion with endosomal membranes. The nucleocapsid is released into the cytosol and uncoated to release the +RNA genome. **4.** The nonstructural ORF of the genome is translated by cellular ribosomes into two polyproteins. Translation of nsP4 depends on stop codon readthrough. **5.** nsP2 proteolytically processes the nonstructural polyproteins into the individual subunits. The nsPs assemble into replication and transcription complexes (RTCs) at the plasma membrane and on cytoplasmic vacuoles that are derived from modified lysosomal and endosomal membranes. **6.** During genome replication, nsP4, the RNA-dependent RNA polymerase (RdRp) first synthesizes minus strands which are later used as templates to synthesize new + strand genomes. **7.** – strands are used to transcribe + strand subgenomic mRNAs. **8.** The structural ORF is translated from the +strand subgenomic mRNA to produce the structural polyprotein. **9.** Processing of the structural polyprotein starts by release of C through autoproteolysis. The remaining polyprotein is translocated across the endoplasmatic reticulum (ER) membrane. The E2-6k and 6k-E1 junctions are cleaved by signalases in the ER lumen. **10.** pE2 and E1 form a heterodimer and are glycosylated and palmitoylated in the ER lumen and during transport through the Golgi apparatus. **11.** During transport through the *trans* Golgi complex to the plasma membrane pE2 is cleaved into E3 and E2 by furin. At the plasma membrane E2 and E1 are incorporated into new virions. **12.** The alphavirus genomic RNA contains a packaging signal that is recognized by C, which results in oligomerization of capsid proteins to generate the nucleocapsid. **13.** Budding occurs at the plasma membrane through interaction of the nucleocapsid with the cytoplasmic domain of E2.

membrane through interaction of the nucleocapsid with the cytoplasmic domain of E2 [85]. A schematic representation of the alphavirus replicative cycle in mammalian cells is depicted in Figure 1.

Chikungunya virus-host interactions

CHIKV is, like all viruses, very dependent on host factors during its replicative cycle. Assuming cell entry of CHIKV occurs in a similar fashion as for other alphaviruses, it requires binding of virus particles to host receptors [86, 87]. An active endocytic pathway, including endosomal acidification, is required for fusion of the viral and host membranes and release of the nucleocapsid into the cytosol [88]. Uncoating of the nucleocapsid requires interaction with the large ribosomal subunit [89]. Translation of viral mRNA requires the cellular translational machinery since viruses do not encode their own ribosomes [90]. RTCs are associated with modified membrane structures [54] and several host factors have been reported to be recruited to these nsP-containing complexes [65, 91-94]. Viral proteins can be phosphorylated and glycosylated by cellular enzymes [77-79, 95, 96]. During budding the nucleocapsid becomes enwrapped in a host-derived membrane envelope [97]. Several steps in the replication cycle require the presence of host membrane lipids [98]. Alphaviruses inhibit host transcription and translation and suppress (innate) immune responses to promote efficient replication and translation of viral genomes [99, 100].

A better understanding of these complex interactions between a virus and its host can facilitate the development of antiviral therapies. Instead of targeting viral enzymes, host factors may also be targeted to inhibit viral replication. With this strategy resistance is expected to occur less frequently, since cellular antiviral drug targets are unlikely to mutate during infection [101, 102].

Several different methods have been used to study CHIKV-host interactions in both mammalian and insect hosts. RNA interference (RNAi) screens [103], transcriptomics [104-106], yeast two-hybrid assays [91, 107], co-immunoprecipitation with viral proteins [108], computational methods [109, 110] and proteomics [111-120]. In these studies a great number of (potential) interaction partners of viral proteins and cellular processes that are affected by CHIKV infections have been identified. Several CHIKV proteomics studies have been performed so far, in which infected cells, infected mice and patient sera were analyzed. However, a large-scale quantitative proteomics study in which several time points post infection during the first round of replication on synchronously infected cells was still missing. Additionally, a systematic analysis of the posttranslational modifications on host proteins during CHIKV infection has never been performed. The aim of the research in this thesis was to track changes in host protein abundance and phosphorylation status during CHIKV infection in well-characterized cells during a single round of replication before cytopathic effects become apparent.

QUANTITATIVE PROTEOMICS

In this first part of my thesis mass spectrometry-based quantitative proteomics was used to study the consequences of CHIKV infection in mammalian cells. Proteomics is the systematic study of the total set of proteins produced from the genome present in the cell at a given time point [121]. Proteomics is a hypothesis-generating approach and is often used as a starting point to identify proteins and pathways that might be involved in a certain process (such as viral infection). To determine the exact role of identified host factors it is usually necessary to do follow-up experiments. With quantitative proteomics the changes in protein abundance in the cell during viral infection can be studied. Relative changes in host protein abundance can provide novel insights into which host proteins and pathways are important during viral infection. An increase in the abundance of certain proteins can, for example, indicate activation of an innate immune pathway, while a decrease could indicate that the virus targets this protein for degradation [122-125].

In this thesis, the stable isotope labeling by amino acids in cell culture (SILAC) method (Figure 2) was used to study the full spectrum of host proteins during CHIKV infection. Cells are grown in media lacking certain essential amino acids, which are supplemented with non-radioactive, isotopically labeled forms of these amino acids [126]. In the heavy state, several ^{12}C and/or ^{14}N atoms are replaced with ^{13}C and ^{15}N isotopes to increase the mass of the amino acid. Cells are grown in SILAC media for at least five cell-doublings to allow >96% incorporation of the heavy amino acids. When labeled arginine and lysine are used in combination with trypsin digestion, all peptides, except for the C-terminal one, will contain at least one labeled amino acid [127]. There is no chemical difference between the heavy isotope-labeled amino acids and their natural counterparts and as a result cell growth and behavior in the light and heavy isotope-labeled conditions is the same. Light- and heavy-labeled peptides co-elute from the high performance liquid chromatography (HPLC) column and are analyzed together in the mass spectrometer [126]. Consequently, the relative quantification of peptide and protein ratios with SILAC is very accurate [128].

An advantage of SILAC over other quantitative proteomics methods such as iTRAQ [129], ICAT [130], dimethyl labeling [131] or label-free approaches [132-139] is that experimental and control samples can be mixed directly after harvesting, so before further sample processing, because every protein in the sample has already been labeled. This reduces the amount of variability between experimental and control conditions caused by sample handling [126] and increases reproducibility [140].

SILAC is most often applied in cell culture but has also been used successfully to label complete animals such as mouse [141], fly [142] and zebrafish [143].

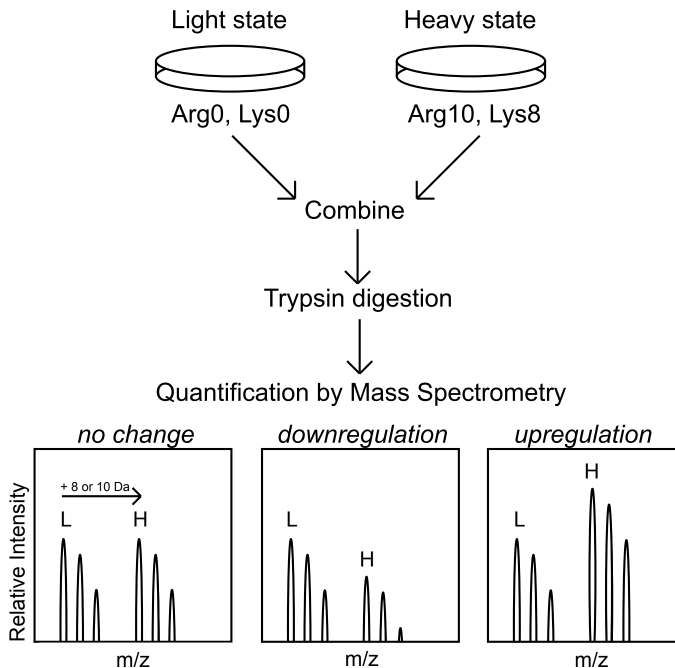


Figure 2: SILAC labeling strategy. Cells are grown in media containing stable isotope-encoded amino acids until the amino acids have been fully incorporated into the cells. After the experiment (e.g. a viral infection) has been performed, the light and heavy samples are mixed and further processed as a single sample. Peptides originating from the light and heavy experimental conditions can be distinguished through their mass difference by mass spectrometry and the relative intensities are used to determine whether the abundance of a protein did not change, was downregulated or upregulated as a result of the experimental treatment.

Quantification of post translational modifications

Most proteins undergo post-translational modifications (PTMs). These modifications play an important role in regulating biological processes and add another layer of complexity to the proteome. PTMs allow the cell to respond rapidly to extra- and intracellular stimuli by changing protein activity, subcellular localization, binding partners and stability [144-146]. Notably, the function of a certain protein can change significantly due to the addition of a specific PTM, while the abundance of that protein does not change at all. Over 400 different PTMs have been described, the most common being phosphorylation, acetylation, N-linked glycosylation and amidation [147]. A single protein can often be modified by distinct PTMs at the same or multiple residues and crosstalk between different PTMs generates an even more complex proteome [146, 148].

The analysis of PTMs poses several technical challenges. Modifications are frequently transient, which makes the timing of sampling during an experiment crucial. The abun-

dance of modified proteins is often very low compared to their unmodified counterparts, which makes it essential to enrich for modified proteins or peptides prior to analysis [149]. Modified peptides can be lost during sample preparation, either due to instability of the modification or because modified peptides preferentially adsorb to metal or plastics [150, 151]. The identification and quantification of a modified site is often based on a single peptide, while multiple peptide identifications are used when protein abundance is quantified. This makes PTM quantification less accurate than the quantification of relative protein abundance. The software search engine that is used to identify peptides and proteins from mass spectrometry data, tests each possible arrangement of amino acids with and without the variable modifications of interest (such as phosphorylation) to find the best peptide match in the database that is used for the search. The addition of each extra variable modification during the search greatly increases search complexity and generates more false assignments [151, 152]. For peptides that contain multiple residues that could contain the modification it is not always possible to identify the exact modified site [152]. When a change in modified peptide abundance is observed, it is important to verify that the total protein abundance did not change as well [151].

In this thesis, SILAC was used to study changes in host protein phosphorylation during CHIKV infection. One third of eukaryotic proteins is estimated to become phosphorylated [153]. Proteins are phosphorylated by the transfer of a phosphoryl group from ATP or GTP to either a side chain of a serine or threonine residue by protein serine/threonine kinases or to a tyrosine residue by protein tyrosine kinases [153]. Humans express ±520 kinases of which around 90 are tyrosine kinases [154, 155]. Protein phosphorylation is a reversible modification and can be removed by protein phosphatases. The human genome contains genes for 107 protein tyrosine phosphatases [156] and around 30 protein serine/threonine phosphatases [157]. The distribution of phosphotyrosine, phosphothreonine and phosphoserine sites in the cell is ± 2%, 12%, and 86% [158].

Several features of phosphorylated peptides aid their identification by mass spectrometry. The addition of HPO_3 results in an increase in amino acid residue mass of 80 Da. Sites on phosphopeptides can be identified from mass shifts in fragment ions generated by MS/MS [151]. Peptides containing phosphotyrosine can often be identified by a fragment ion of 216 Da which derives from peptide bond cleavage on either side of the phosphotyrosine residue [151, 159]. Additionally, peptides containing phosphoserine and phosphothreonine often undergo a 98 Da neutral loss corresponding to the loss of H_3PO_4 [151].

OUTLINE PART 1

In this first part of my thesis, changes in the host cellular proteome following CHIKV infection were studied to better understand the interplay between viral replication and host cell infrastructure and metabolism. **Chapter 2** describes a SILAC-based quantitative proteomics study to analyze temporal changes in the proteome of CHIKV-infected cells. This study revealed that changes in protein abundance during CHIKV infection are relatively small and that most of the proteins that showed significantly changed abundance were downregulated. Four proteins that were significantly downregulated during CHIKV infection, Rnd3, DDX56, UbcH10 and Plk1, were overexpressed from plasmids in host cells and this inhibited CHIKV infection. **Chapter 3** describes a SILAC-based quantitative phosphoproteomics study of CHIKV-infected cells. Eukaryotic elongation factor 2 (eEF2) was identified as a factor that becomes phosphorylated in various cell lines early during infection with CHIKV, Semliki forest virus (SFV) or Sindbis virus (SINV). Infection with coxsackie virus B3, a picornavirus, also triggered eEF2 phosphorylation. eEF2 phosphorylation might reflect part of the antiviral response of the cell, but it is not activated via one of the 'classical' pathways that are generally involved in the innate immune response to virus infections. In **chapter 4** the findings of part 1 are summarized.