UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

Biological activity and functions of the capsid protein of Zika virus and related flaviviruses

André Rodrigues Nascimento

Mestrado em Bioquímica
Especialização em Bioquímica Médica

Dissertação orientada por:
Doutor Ivo C. Martins e Prof. Doutor Francisco Pinto

2018

2018

Acknowledgments

I wish to express my humble gratitude to everyone at NSantos Lab for accompanying me in this entire journey for the past 2 years. Every single person at the lab, somehow contributed for the work developed here and for my personal and scientific growth. Especially a big thanks to Prof. Nuno C. Santos, the group leader, which accepted me 1 year early in his lab introducing me into the "Flavivirus" line of work led by Dr. Ivo Martins and provided all the conditions I needed to develop this work and for his amazing support and positive attitude at every moment. To Ivo Martins, my supervisor, colleague and friend, I am grateful for all amazing support, expertise and day-by-day close follow up from which I learned and developed my scientific growth. Also, without him this work would never have been accomplished so, thank you Ivo, you're like a fatherly figure in the world of Science and an exemplary person to follow. A big deep thank you to Ana Martins and André Faustino, who also accompanied me in this journey with an everyday follow-up in the lab, teaching me all I needed to know for this work to be done. They are like my "older brother and sister" in the world of Science, who taught me to further develop my critical side, and were always there when I was in need, even being far. Also, thanks for the measurements made when I couldn't Ivo, André and Ana! To Filomena Carvalho, for the expertise and follow up in the cell-culture/isolation of lipid droplets. To Teresa Freitas, for the amazing job done in teaching me how to do those tricky sucrose gradients, and for always having the time to prepare some of the materials I needed when I didn't have time! To Marcos Domingues, for the DLS tips and ideas for new denaturation experiences and to Barbara Gomes for teaching me how to perform fluorescence assays.

For my friends I also owe a word. Firstly, to BatClan for all the amazing time spent together, for the support during this rough time and for making me feel like home. You know, friends are the family you choose, so I'm grateful to every single person. Also, to my friends in college which provided some amusement times when things got hard, and always found the time to drink a coffee or a beer with me in those sunny afternoons. For the guys in "lel", for all the amazing times spent in those harsh games with much feeding and raging in the mixture. Always have a blast with you all.

A special and deep acknowledgment for my long-standing friends from Santarém, namely Angela António, Bernardo Gabirra, Inês Lucas for all the constant support and friendship, you know, after 7 years you all stood with me. A heartfelt thank you to João Afonso who is like a brother to me, for the amazing time spent together in the past, for all the deep conversations and for the guidance and support throughout this years. A special thanks to some of the great persons that college tied to my life. To Joaquim, for the amazing, constant and true friendship. That this continues for more 7 years you feeder. To Tomás "Beard" Silva, and Tomás "TS - Tómás" Silva, for the amazing support and always finding the time for lunch or coffee in my times of need and for sometimes delivering some reason in harsh times. A heartful and warm acknowledgment to the greatest person that crossed my path, Filipa Simões, whose spirit is so bright that always kept away some of my darkest moments and fears. For all true friendship, laughs and support, I'm in debt to you Filipa. Keep who you are, and I know that I will see you in the top in a few years, alongside myself! To Gerson Lobo, one of my best friends who left this world too soon. Hoping that you are cheering me up like you always did with that smile of yours in your face, wherever you are.

Also, I need to appreciate for all the support and the endlessly motherly love that Margarida Quaresma has for me. For all advice, talks, and close friendship throughout these 7 years I couldn't have picked better in my freshman year to guide my path than you. You are a role model to follow both in life and in Science, for all the effort and hard work you put in every single thing you do. To Madalena

Pinto, for the constant caring and worrisome and for never stopped believing in me. You are the most simple and humble person I've met, thank you for that aunt-love you always had for me.

I want to acknowledge my family for all the trust, patience, hard work and sacrifice they have always made to support me at all times. I want to appreciate especially to my parents, who gave me all the conditions to grow as a person and provided me all I needed both in my Bachelor and especially now in my Masters in Biochemistry. To my father, for all the effort, hard work and support he has always done to ensure that nothing is missing to his sons. To my mother, who taught me to value the good things in life, and to see the good in people and in all the little things and of course for all the emotional support. To my brother, for his spirit, determination and energy and for always bringing some joy and amusement when we are together.

Last but not least, I also need to express my gratitude and heartfelt acknowledgment to the love of my life, Beatriz Silva. For all the constant support, day-by-day care, patience, personality and endless love that are really important for me and from which it gave me the strength I needed to get back on the right track and to carry this through.

This Thesis is dedicated to my father, to my mother and to my brother.

For being my constant pillars in life and reference for the future.

Resumo

Esta tese de mestrado foca-se na importância da interação da proteína da cápside (C) do vírus Zika com sistemas lipídicos do hospedeiro, a qual influência diversos passos importantes do ciclo de vida deste vírus, pertencente à família *Flaviviridae.* Anteriormente, casos de infeção por Zika poderão ter sido mal diagnosticados, uma vez que os sintomas apresentam elevada semelhança com os sintomas da infeção pelo vírus da dengue e que ambos apresentam a mesma distribuição geográfica (África e Oceânia). Nos últimos anos, casos de microcefalia em recém-nascidos e de síndrome Guillain-Barré em adultos foram associados à infeção pelo vírus Zika, nas epidemias de 2015 no Brasil e 2013, na Polinésia Francesa, respetivamente. Dada a extensão do número de afetados e a gravidade, tornou-se numa emergência de saúde pública a nível mundial, conforme reconhecido pela Organização Mundial de Saúde.

O vírus Zika é transmitido pela picada de mosquitos do género *Aedes* spp., em particular *Aedes aegypti* e *Aedes albopictus* sendo que ambos se encontram em zonas tropicais e subtropicais. Contudo, o aquecimento global acompanhado do movimento de pessoas e bens permitiu que em zonas temperadas, cujos invernos tendem agora a ser mais moderados do que no passado, estes vetores se consigam estabelecer, sendo transportados para essas novas zonas através do comércio internacional de pneus usados e de plantas exóticas com pequenos depósitos de água. Esta realidade está presente na Europa (mais propriamente, ao largo do Mediterrâneo), onde cada vez mais se observam colónias de *A. albopictus*, bem como *A. aegypti*. Esta situação já levou a epidemias de dengue na Europa, inclusive em Portugal, uma vez que este vírus é transmitido pelos mesmos vetores. Assim sendo, uma epidemia de Zika na Europa é uma possibilidade real. A infeção pelo vírus Zika apresenta dois panoramas completamente distintos: um relativamente assintomático, com manifestações de febre ligeira, dores nos músculos e articulações, não justificando muitas vezes quaisquer cuidados clínicos, e outro mais agressivo associado a problemas neuropatológicos, podendo resultar na ocorrência de microcefalia em recém-nascidos e de síndrome de Guillain-Barré em adultos, causando paralisia. Não existe ainda qualquer vacina comercial ou terapia específica para o Zika. Tal é devido à falta de conhecimento sobre os mecanismos moleculares por detrás de certos passos importantes no ciclo de vida viral.

O vírus do Zika é semelhante a outros agentes patogénicos relacionados, como a febre amarela, a febre tifóide, o vírus da dengue (DENV) e da hepatite C (HCV). Este vírus, tal como todos os *Flavivirus*, é constituído por um invólucro superficial contendo as proteínas do envelope (E) da membrana (M), que estão ligadas a uma bicamada lipídica viral por domínios transmembranares. Sob essa bicamada encontra-se presente a nucleocápside, sendo esta constituída por uma única molécula de RNA de cadeia simples positiva ($ssRNA^+$) complexada com diversas cópias da proteína da cápside (C). O genoma viral contém uma região codificante a partir do qual toda a maquinaria viral é sintetizada, sendo esta constituída pelas três proteínas estruturais acima referidas e por sete proteínas não estruturais. Estas últimas são responsáveis pela replicação intracelular do vírus, enquanto as proteínas estruturais asseguram principalmente a estrutura do virião. A proteína C, que é também uma proteína estrutural, tem ainda um papel importante no mecanismo de encapsidação e assemblagem viral. Os *Flavivirus* infetam principalmente o fígado e os epitélios vasculares do hospedeiro, e, em alguns casos, o cérebro, coração, músculos e órgãos linfáticos. Durante a infeção por DENV e HCV, o metabolismo lipídico é fortemente influenciado, levando a um aumento do número e do tamanho de corpúsculos lipídicos intracelulares. Existe também uma elevada desordem da estrutura membranar bem como alteração da concentração e composição das lipoproteínas plasmáticas.

A nível molecular os mecanismos da replicação viral ainda não estão inteiramente elucidados. Contudo, o estudo da atividade biológica da proteína C de flavivírus constitui uma linha de investigação promissora com elevado potencial para o desenvolvimento de novos fármacos. As proteínas C de flavivírus são compostas por aproximadamente 100 resíduos de aminoácidos e apresentam uma distribuição não homogénea da sua carga, sendo carregadas positivamente. As estruturas conhecidas de proteínas C de flavivírus transmitidos por picada de mosquito são as correspondentes aos vírus da Dengue, do Nilo Ocidental e mais recentemente do Zika (publicada a 30 de março, depois da realização deste trabalho e entregue da versão provisória desta tese). Todas estas proteínas são homodímeros em solução: cada monómero destas proteínas possui estruturas terciárias idênticas constituídas por quatro regiões em hélice α e um N-terminal não estruturado, menos estudado. Existe uma grande localização de cargas positivas na mesma região (hélice α4), deixando do lado oposto em termos espaciais uma grande superfície hidrófoba (região entre a hélice α1 e α2). Pensa-se então que a região positivamente carregada será responsável pela interação com o RNA viral, enquanto que a região hidrofóbica irá interagir com lípidos e/ou com bicamadas lipídicas. O grupo de investigação onde este trabalho foi desenvolvido demonstrou que, para além da referida região hidrofóbica, a secção N-terminal da proteína C do vírus da Dengue tem um papel fulcral na ligação a corpúsculos lipídicos, tendo desenvolvido e patenteado um péptido designado pep14-23, baseado numa região conservada da mesma. Este péptido inibe a ligação da proteína C de Dengue a corpúsculos lipídicos e a lipoproteínas. Dado que a interação da proteína C do vírus Dengue com corpúsculos lipídicos é essencial para a replicação viral, este é um avanço considerável. O mesmo grupo obteve resultados semelhantes recentemente para a interação da proteína do vírus do Nilo Ocidental com sistemas lipídicos do hospedeiro. Assim, visto que o vírus Zika pertence à mesma família, sendo as proteínas da cápside altamente similares em termos de sequência, é proposto que a proteína da cápside do vírus Zika interaja igualmente com corpúsculos lipídicos intracelulares e/ou lipoproteínas. Esta hipótese é testada neste trabalho recorrendo a técnicas biofísicas, bem como previsões computacionais de forma a ser possível a integração dos dados obtidos em modelos dessas interações.

Para tal, numa primeira fase do trabalho foi necessário produzir proteína C recombinante de Zika. De seguida, medidas de potencial zeta dos corpúsculos mostraram que estes têm carga negativa, na ausência de proteína C. Incubando os corpúsculos com diversas concentrações da proteína viral observa-se que o valor de potencial zeta vai gradualmente aumentando com o aumento da concentração de proteína, estabilizando em valores positivos, sendo indicativo de interação da proteína C com estes sistemas lipídicos. De seguida, as proteínas de superfície destes corpúsculos foram removidas (pela ação da tripsina). Nestas condições, apesar dos corpúsculos se manterem negativamente carregados, o grau de interação é reduzido, sugerindo que a interação entre a proteína C e os corpúsculos não é de apenas natureza electroestática, devendo envolver proteínas específicas da superfície dos corpúsculos (á semelhança do observado com a proteína C do vírus da Dengue).

A interação com lipoproteínas plasmáticas humanas de muito baixa densidade (VLDL) e lipoproteínas de baixa densidade (LDL) também foi estudada, através do uso de espectroscopia de dispersão dinâmica de luz, onde após titulação com proteína C, foi possível medir o aumento do tamanho médio relativo das partículas, correlacionando-se com as previsões realizadas computacionalmente para a dimensão da proteína C do vírus Zika. Com este estudo, observou-se uma interação com LDL, e uma aparente inexistência de interação com VLDL, uma surpresa face ao que é descrito na literatura para o vírus Dengue, que interage com VLDL, mas não com LDL. Dados preliminares indicam que esta interação pode ocorrer quer na presença de iões sódio bem como iões potássio, sendo novamente uma surpresa face ao descrito na literatura, uma vez que esta interação para o vírus Dengue é descrita como dependente da concentração do ião potássio. Estes dados irão ser corroborados por meio de técnicas

adicionais, nomeadamente espectroscopia de força baseada em microscopia de força atómica, tendo em vista a identificação dos alvos moleculares da proteína C de Zika em cada sistema lipídico (corpúsculos lipídicos, VLDL e LDL).

Um estudo da capacidade do péptido pep14-23 de inibir a ligação da proteína C de Zika a LDL foi também realizado, recorrendo à espectroscopia de dispersão dinâmica de luz. Com os dados obtidos, é possível observar que existe inibição por parte do péptido pep14-23. Uma vez que o pep14-23 é baseado na sequência do N-terminal, este irá por sua vez possivelmente competir com a proteína C pelo mesmo alvo molecular. Estes dados podem ser melhorados futuramente, com o teste de derivados do péptido pep14-23, avaliando a capacidade dos mesmos de inibir a atividade biológica da proteína C, o que poderá permitir futuros fármacos contra a infeção pelo vírus Zika e *Flavivirus* semelhantes.

Palavras-chave: Vírus Zika, proteína da cápside, corpúsculos lipídicos, lipoproteínas, péptido inibidor

Abstract

Zika virus (ZIKV) infection is transmitted through the bite of *Aedes* spp. mosquitoes, namely *Aedes aegypti* and *Aedes albopictus*. Recent ZIKV outbreaks were associated with neurological disorders (microcephaly in newborns and Guillain-Barré syndrome in adults), constituting a serious public health threat. The life cycle is poorly understood, and its global spread and disease severity are increasing. Despite this, there is no specific and effective treatment or vaccine against ZIKV infection out in the market. Key steps in the viral life cycle are viral assembly and encapsidation. In related viruses, such as Dengue virus (DENV), these steps are mediated by the capsid (C) protein, which interacts with viral RNA as well as with host lipid structures. Noteworthy, the interaction of DENV C with host lipid droplets (LDs) is essential for viral replication. The host group has characterized DENV C-LDs interaction, as well as DENV C ability to bind lipoproteins. The work led to pep14-23, a peptide patented by the host group that was designed based on a conserved motif of the flavivirus N-terminal region and which inhibits both interactions. Similar findings were obtained concerning West Nile virus (WNV) C protein ability to interact with host lipid systems. Given the similarities of DENV, WNV and ZIKV C proteins, it was therefore hypothesized that ZIKV C may bind the same lipid systems and be inhibited by pep14-23. In this thesis it is shown that ZIKV C is indeed able to bind LDs and low-density lipoproteins (LDL). The interaction of ZIKV C with LDs was studied via zeta potential measurements, showing that the interaction does occur, requiring LDs' surface proteins in potassium buffer. Preliminary data suggests that ZIKV C is able to interact with LDs in the presence of sodium buffer, which was unexpected since for DENV it was reported that DENV C-LDs binding only occurs in the presence of potassium ions. ZIKV C interaction with lipoproteins, namely LDL and very low-density lipoproteins (VLDL) was measured via dynamic light scattering (DLS). ZIKV C seems to bind LDL but not VLDL. In the presence of ZIKV C there is an increase of LDL hydrodynamic radius that correlates with the ZIKV C estimated dimension. Pursuing the goal of developing new drug strategies for ZIKV, the effect of pep14-23 in ZIKV C binding to LDL was determined using DLS. The data acquired indicates that pep14-23 may inhibit ZIKV C protein. This reaffirms the potential ability of pep14-23 as an inhibitor drug lead against ZIKV and closely related flavivirus.

Keywords: Zika virus; capsid protein; lipid droplets; lipoproteins; pep14-23.

Index

List of Figures and Tables

List of Symbols and Abbreviations

ADRP – adipose differentiation-related protein or perilipin 2

AFM – atomic force spectroscopy

apoB100 – apoprotein B100

apoE – apoprotein E

ATP – adenosine triphosphate

BHK-21 – baby hamster kidney cells

C – capsid protein

CHIKV – Chikungunya virus

$C_{1/2}$ (ζ-potential measurements) – C protein concentration at which half of the maximum amplitude of variation of zeta potential has occurred, induced by the protein interaction with LDs (half of ZIKV C protein concentration needed for saturation of LDs)

CD – Circular dichroism spectroscopy

CEs – cholesterol esters

D_H – hydrodynamic diameter

DMEM – Dulbecco's Modified Eagle Medium

DLS – Dinamic Light Scattering spectroscopy

DENV – Dengue virus

DENV C – Dengue vírus capsid protein

E – Envelope protein of *Flavivirus*

EDTA – ethylenediamine tetraacetic acid

EGTA – egtazic acid

EM-clustering – Expectation-maximization clustering algorithm

ER – endoplasmatic reticulm

FBS – fetal bovine serum

f (ka) – Henry's function

g2(t) – second order autocorrelation function

GBS – Guillain-Barre syndrome

HDL – high density lipoproteins

I – Intensity of light scattered

IDPs – intrinsically disordered proteins or regions within a protein

I(t) – Scattering Light intensity fluctuation as a function of time

I-TASSER – web server that predicts the tertiary structure from a protein sequence

IPTG – Isopropyl β-D-1-thiogalactopyranoside

JEV – Japanese Encephalitis virus

KH_2PO_4 – monopotassium phosphate

K2D2 – online server that takes as input the CD data obtained experimentally and returns a predicted value of percentage of secondary structure most present in the protein

K2D3 – an updated version of K2D2

LB medium – Lysogeny broth medium

LDL – low density lipoproteins

LDLR – LDL-receptor pathway

LDs – lipid droplets

LGH – lactate dehydrogenase

LVP – lipoviroparticles

M protein – membrane protein of *flavivirus*

n (DH) – particle number distribution as a function of DH

NMR – nuclear magnetic resonance

NS – non-structural

OD – optic density

ON – overnight

PALS – phase analysis light scattering

PDB – Protein Data Bank

PDB ID – Protein Data Bank identification

PMSF – phenylmethylsulfonyl fluoride

prM – premembrane

RMSD – root-mean-square deviation

RNA – ribonucleic acid

RT-PCR – real time polymerase chain reaction

ssRNA^{+} – single-stranded positive-sense RNA

TAGs – triacylglycerols

TEE – Tris-HCl/EDTA/EGTA Buffer

TIP47 – perilipin 3

Tm – temperature melting point

T_m^{app} – temperature melting point apparent

UK – United Kingdom

USA – United States of America

UV – ultraviolet

VLDL – very low-density lipoproteins

WHO – World Health Organization

WNV – West Nile virus

YFV – Yellow fever virus

ZIKV – Zika virus

ZIKV C – Zika virus' capsid protein

1R6R – Dengue virus PDB ID

2D – two-dimension

3D – three-dimension

1 Introduction

1.1 Flavivirus

The *Flaviviradae* is a large family of viral pathogens that are responsible for several serious diseases and mortality in humans and animals.[1] This family consists in four genera: *Flavivirus, Hepacivirus, Pestivirus* and *Pegivirus*.[2,3] This thesis will focus on the largest genera, the *Flavivirus* genus which contains more than 70 viruses that can be further subdivided into clusters based on the transmission vector, focusing in the mosquito-borne *flavivirus*[3]. The 'flavivirus' name is derived from the Latin word *flavus* meaning yellow, signifying the jaundice caused by the yellow fever virus (YFV).[1]

The genus *Flavivirus* spp. is composed by a group of small enveloped RNA viruses, most of which are arthropod-borne viruses (arboviruses), transmitted to vertebrate hosts by mosquitoes or ticks.[3] Besides causing diseases in animals, some of these viruses are also highly pathogenic to humans, namely Dengue (DENV), Zika (ZIKV), West Nile (WNV), Japanese Encephalitis (JEV) and Yellow Fever (YFV) viruses, among others.[4,5] Flaviviruses share a similar genomic organization and replication strategy yet they can cause a wide range of distinct clinical diseases in humans, *e.g.* ranging from fever, encephalitis, meningitis and hemorrhage, constituting a major international health problem.[6,7] The World Health Organization (WHO) estimates that each year there are more than 50000 cases of Japanese encephalitis and 200000 cases of yellow fever with approximately 30000 deaths worldwide.[6–8] Regarding DENV infection, it has grown worldwide, with the latest estimates suggesting 390 million infections annually, out of which 96 million present clinical manifestations.[9,10] Dengue has spread being now present in 128 countries, with an estimate of 3.9 billion people living in areas with risk of DENV infection.[9,11] As for ZIKV infection, further discussed ahead, it has been recently associated with severe congenital microcephaly in newborns[12,13] and Guillain-Barré syndrome in adults[14]. The 2015 outbreak lead WHO to classify ZIKV as an international public health emergency, as DENV.[12–14]

1.1.1 Flavivirus lifecycle

To initiate its infectious cycle, the virus attaches to the surface of its target cells *via* specific receptors, after which the viral particle enters the infected cell.[15] Then, the viral RNA is released from the viral particle, that is then translated and replicated. The viral RNA is released through a process by which the viral particle, already within the cell, fuses with late endosomes.[7] This fusion is triggered by the acidic pH inside the late endosomes, which leads to a conformational rearrangement of the E protein, which results in the exposure of its hydrophobic protrusion (fusion peptide).[7,16,17] This fusion peptide is inserted into the cellular target membrane and promotes membrane fusion. Then, the genomic RNA is released into the cytoplasm of the cell, being translated into a single polyprotein, containing 3 structural and 7 non-structural (NS) proteins, that is then cleaved.[18,19] Flavivirus remodel intracellular membranes developing organelle-like structures to create adequate platforms for viral replication, like invaginations of vesicle-like structures into the lumen of the endoplasmic reticulum (ER), forming vesicle packets.[20] Then newly synthesized immature viral particles are produced by budding the capsid (C) protein and the associated genomic RNA (nucleocapsid) into ER-derived membranes that are studded with the precursor membrane (prM) and envelope (E) proteins.[21] These *de novo* synthesized particles undergo a maturation process, as they traffic along the secretory pathway across the Golgi complex, in which the prM protein is cleaved by a *trans*-Golgi resident furin-like protease, resulting in the M protein, with the mature viral particle ready to be released[22] (***Figure 1.1***).

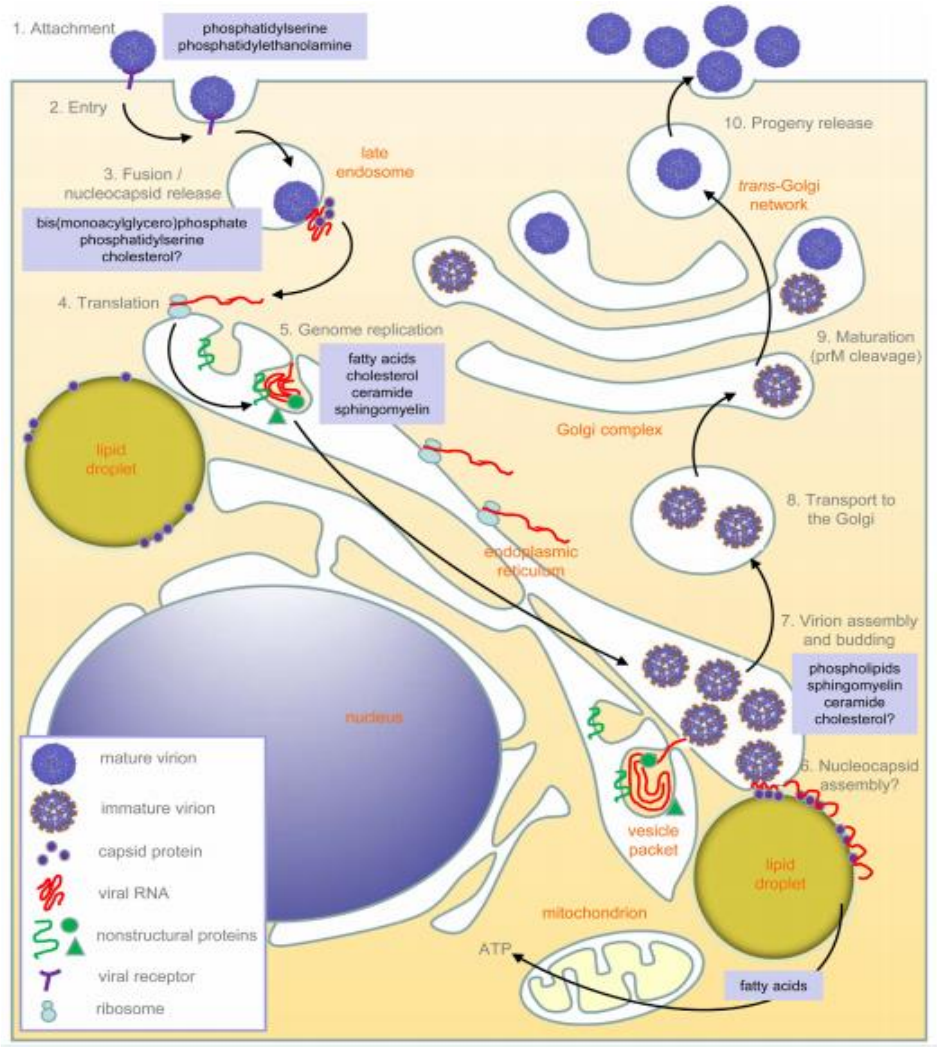


Figure 1.1 – Flavivirus life cycle. Schematics representing viral entry, replication, assembly and release. ER and ER-derived structures produced by invagination of the ER membrane play a major role in flaviviruses replication and assembly. The lipids involved in each step are highlighted. The contribution of lipid droplets to viral infection is also clear. Adapted from [22].

1.1.2 Flavivirus Structure

Mature flavivirus virions are smooth spherical particles with approximately 50 nm diameter that exhibit a herringbone pattern of E protein dimers.[22–24] Below that layer there is the M protein (which is produced by the cleavage of prM upon maturation, as described above). Both proteins are attached to the envelop lipid bilayer by transmembrane domains. Covered by this layer, there is the nucleocapsid core of the virus, which is formed by an array of C protein subunits complexed with the single-stranded positive viral RNA.[6] There is a strong resemblance between the several *Flaviviruses* viral particles for which structures have been described.[1,6,22] The major factor that changes is the spatial organization between the E protein dimers: for example: in WNV the E proteins are widely distributed while in the ZIKV are more compacted, with DENV particles E protein distribution laying somewhere in between.[22,24] Although the mature virion structure is resolved and it is clear that the flavivirus RNA genome encodes for only 10 proteins, with viruses thus exploiting host cell machinery to complete their infectious cycle, the mechanism behind and the molecular basis for viral replication are not fully understood. For details, please consult **Figure *1.2*.**

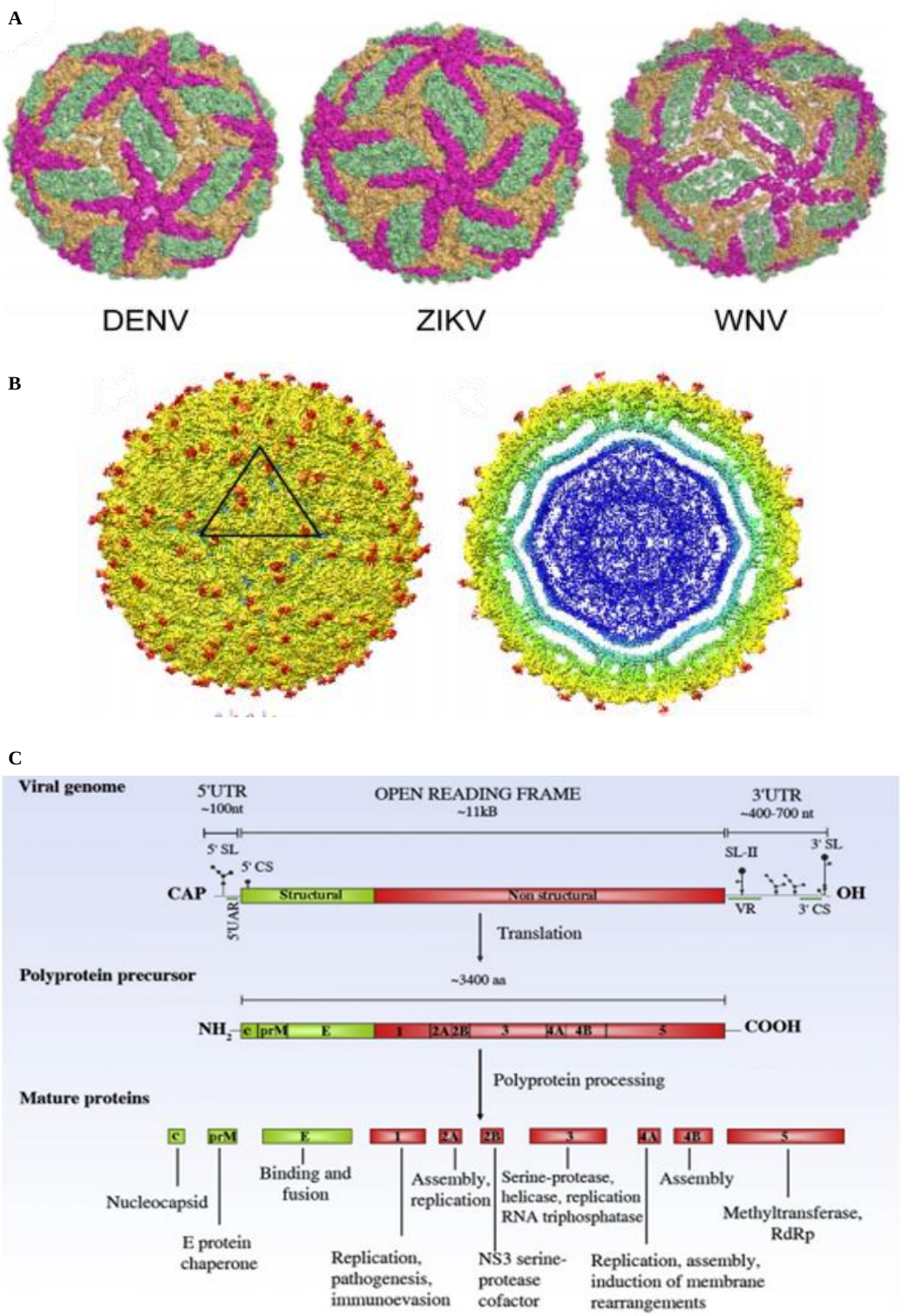


Figure 1.2 – Flavivirus virion organization, ZIKV Cryo-EM structure and ZIKV viral RNA description. (A) DENV, ZIKV and WNV mature particles surface representation: a similar herringbone pattern of E glycoprotein dimers (colored in magenta, green and orange to facilitate interpretation) is visible. **(B)** Cryo-EM structure of Zika virus. The E (yellow) and M (green) proteins are adjacent to the lipid bilayer (cyan), within which is the nucleocapsid (blue), containing the C protein complexed with the viral ssRNA^{+}. **(C)** Flavivirus genome structure, organization and viral proteins function: the single open reading frame encodes a polyprotein precursor that is co-, and post translationally cleaved into three structural proteins (in green) and seven nonstructural proteins (in red). Adapted from [7,22,24].

1.1.3 Flavivirus capsid protein

The capsid protein of flavivirus plays a crucial role in virus biology. During the virus life cycle, several subunits of C protein are assembled to form nucleocapsids which encase the genetic material. [25] Several reports indicated that a cleavage region present in the capsid sequence (dibasic-site) is extremely important for virus growth, and enzymatic activity towards this region may also influence virulence.[26] For this, the C protein is considered a target for drug design. [26,27] In spite of being classically depicted as a structural protein the C protein has other functions that go beyond the virus assembly process.[28] There are reports that the C protein is present in the nucleus of flavivirus-infected cells, raising concerns about the migration of C protein to intracellular compartments.[29–35] In fact, it was also shown that importin-α, a cellular protein specialized in transportation, binds to WNV C protein and transports it from the cytoplasm to the nucleus.[36] Once in the nucleus, it can mimic histones, interacting with the core histones and disrupting nucleosome formation.[37] Despite this, the role of the C protein in the nucleus is unknown. [25] In contrast to other viruses with highly ordered capsid geometries, the nucleocapsid is not well-ordered, being instead visible an amorphous capsid pack that is a peculiar structural feature of flaviviruses. [38] How this amorphous nucleocapsid is formed is yet to be unraveled but it is known for tick-borne encephalitis virus (TBEV) that in the presence of RNA, purified C proteins form capsid-like structures *in vitro*, suggesting that the C protein functions as a basic assembly building block.[39] TBEV and WNV C proteins have a remarkable functional flexibility since they are permissive to many deletions and yet retains its RNA packing activity.[28,40,41] However, despite this functional plasticity, basic residues in the N-terminal of DENV C are strictly required for virus particle formation.[42] These facts suggest the C protein N-terminal region critical importance for viral function and infectivity.[28]

The C protein is relatively small compared to the other flavivirus proteins. [26,43,44] ZIKV C protein has 105 amino acids (11,6 kDa), being an α-helical protein, has judged recently resolved crystallographic structure.[45] The 3D structure of DENV C protein was solved at atomic level through Nuclear Magnetic Resonance (NMR) (PDB ID 1R6R[46]) while WNV C protein was solved via X-Ray crystallization (PDB ID 1SFK[47]). Both WNV and DENV C protein structures are very similar.[48] The recently crystallized ZIKV C protein structure is also structurally similar to these later two, especially to WNV.[45] The flavivirus C protein in solution forms dimers stabilized by internal hydrophobic domains participating in homotypic interactions.[46] The homodimer is largely alpha-helical, organized in a symmetric way with four alpha helices (α1 to α4) in each monomer plus the disordered N-terminal region. The α4-α4' surface is accessible to the solvent and rich in arginine residues, being highly positively charged at physiological pH. On the opposite side there is a α2-α2' region flanked by the α1 helices that creates an hydrophobic cleft.[46,48–51] Thus, it was proposed that the viral genome binds the positively α4-α4' surface region while the hydrophobic α2-α2' interface and α1 helices would bind host lipid systems.[46] (***Figure 1.3***).

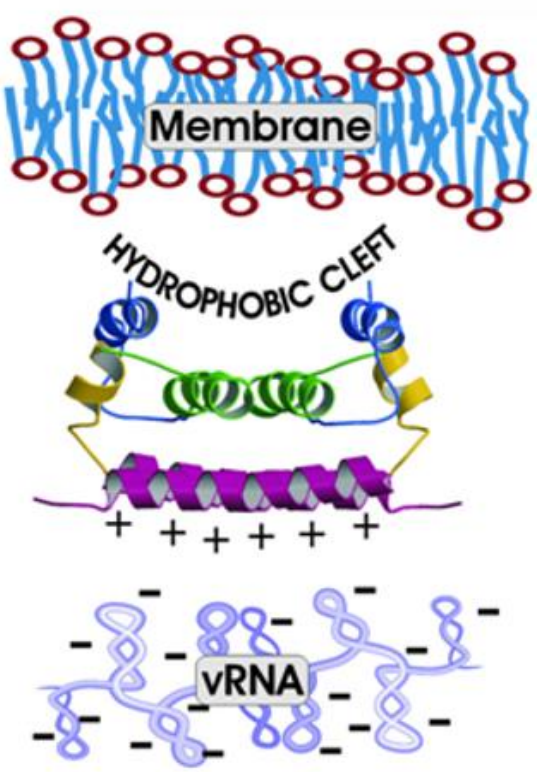


Figure 1.3 – Structural aspects of the flavivirus capsid protein. Proposed orientation of the capsid protein of DENV during viral assembly with α4-α4' interface facing the viral RNA while the hydrophobic cleft points towards the lipid bilayer. Adapted from [45]. The recent publication of ZIKV C structure (March 30, 2018, at the moment that this thesis was delivered[45]) states that ZIKV C is α-helical and with a similar charge distribution as DENV C. Although only the abstract was available at the time of delivery of this thesis, this recent data suggests that ZIKV C behaves in a similar manner to DENV C, as originally hypothesized in this thesis.

1.2 Lipid metabolism during *Flavivirus* infection

The lipid metabolism is greatly affected during Flavivirus infections, resulting in intracellular disordering of the membrane network, imbalanced blood plasma lipoproteins composition and concentration.[52,53] Importantly, DENV C interaction with host LDs is essential for viral replication. DENV C also binds lipoproteins, which may be an important factor for infectivity, leading to formation of lipoviroparticles (LVPs), which are more infective than mature virions and may allow evasion from the host immune system.[54,55]

1.2.1 Lipid Droplets

LDs are the emulsion phase of insoluble oil droplets dispersed in the cytoplasm, originated from the ER, occurring in the cytosol of all eukaryotic cell types.[56] They have a rather unique structure, containing a hydrophobic core and a single layer of amphipathic phospholipids. [57]In its hydrophobic core there are predominantly triacylglycerols (TAGs) and cholesterol esters (CEs).[58] On LDs surface are found several proteins from the PAT family, namely perilipin, adipose differentiation-related protein (ADRP) and tail-interaction protein of 47 kDa (TIP47), also known as perilipin 3 (PLIN3), as well some lipolytic enzymes.[59,60] These particles, in eukaryotes, respond to the increase of cellular fatty acid levels, and emerge from the accumulation of neutral lipids in the ER.[61] In the past, LDs were considered to be simple and passive lipid storage compartments but nowadays are considered highly dynamic and complex, playing a central role in lipid metabolism and being connected to diverse cellular processes like fatty acid trafficking, cellular signaling, protein storage, autophagy, immunity and virus replication.[48] Since LDs are metabolically active organelles, they regulate the balance between the host lipid synthesis and mobilization to maintain cellular homeostasis.[50,60] Due to their proximity to the ER and to the unique structural features, the surface of LDs can serve as transient storage depots for proteins which are destined for degradation.[62] There is a molecular link between LDs and the ubiquitination machinery (a signal for protein degradation pathways).[60,62,63]

Regarding the association of LDs with the viral replication, this dynamic cellular lipid storage organelle participates in the viral lifecycle by providing intracellular membrane surface area (acting as platform for the virus assembly), lipids, energy and proteins.[64] In the infected cell, massive intracellular membrane rearrangements are induced by perturbing lipid biosynthetic pathways to spatially segregate the replication and assembly sites.[59,65] These two sites need to be separated to avoid competition between the capsid protein and the viral replicase complex at the level of RNA binding.[62,65] On the other hand, the newly synthesized Flavivirus' $ssRNA^+$ needs to be transported from the replication to the assembly site, where the capsid protein is concentrated.[59,62] Thus, the two sites require close proximity to each other to reach maximum efficiency in virus assembly.[59,65] In accordance with the described above, LDs have been reported to associate with virus-induced membrane bound compartments, which are considered to be possible replication sites.[59,62]

As such, DENV infection induces the formation of single-membrane in-folding into the ER lumen, and unstructured convoluted membranes.[52,65] These vesicle-like structures contain the viral replicase and $ssRNA^+$ and also pore-like openings that enable the release of newly synthesized viral RNA, facilitating replication and efficient encapsidation.[53,65] Besides this, LDs are close to mitochondria, with the fatty acids that are mobilized by the activation of autophagy pathways upon viral replication being oxidized, generating ATP required for viral RNA replication.[52,53,61] DENV C protein interacts and accumulates on the surface of LDs through the hydrophobic cleft and with the N-terminal disordered region.[48,51] This interaction is associated with a protein in the surface of LDs, PLIN3, which

occurs in a potassium ion dependent manner.[48,50] Through the regulation of the potassium ion concentration the capsid protein regulates its binding and releases itself from LDs.[50] Thus, DENV and probably other members of the flavivirus manipulate specific intracellular ion concentrations to favor viral replication.[50,62] Moreover, the LDs might function as a scaffold for DENV assembly by exposure of the protein cationic surface towards the aqueous environment, prompt to interact with viral RNA (**Figure *1.4***).[50,62]

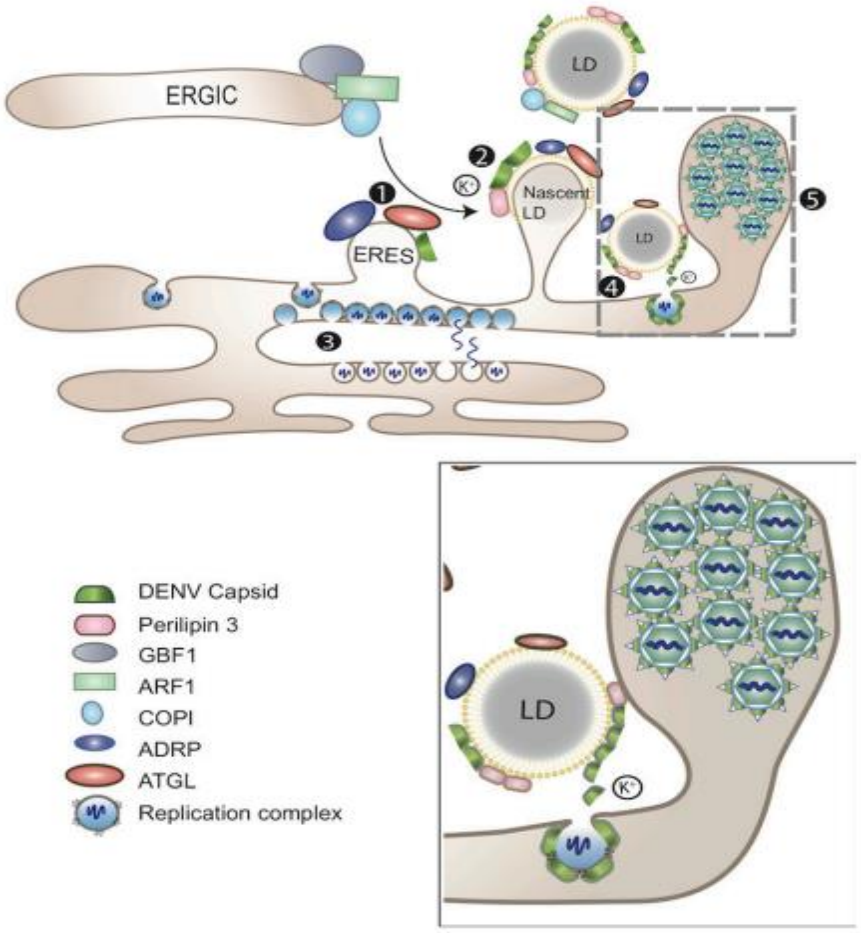


Figure 1.4 – LDs as platforms for virion assembly in DENV infection. (1) At the ER–Golgi intermediate compartment, DENV subverts physiological processes for the transportation of its C protein to the LDs surface. (2) Accumulation of DENV C protein on LDs, along with cellular perilipin 3 (and intracellular K^+ ions). (3) Replicated DENV genomes are released through the vesicle pore and then engaged into nucleocapsids that bud through the ER membrane. (4) The C protein can be released from LDs to the cytosol or other cellular compartments for viral assembly (gray-dashed frame and enlarged panel). (5) Packed virions accumulate within the lumen of vesicle packets-containing ER network before transported to Golgi. Adapted [62]

1.2.2 Lipoproteins (VLDL and LDL)

Lipoproteins circulate in the bloodstream in suspension in the blood plasma and can be classified in 4 types, according to their density and function: Chylomicron, very low-density lipoproteins (VLDL), low density lipoproteins (LDL), and high-density lipoproteins (HDL), ordered by decreasing average size and increasing density.[63,64] Chylomicron collects the dietary lipids absorbed in the intestine and redirects the majority to the liver.[63] The liver releases HDL to the blood almost without lipids and as it circulates the bloodstream, collect the excess of cholesterol in the organism and grow in size and are later recovered by the liver.[63] VLDL are formed in the liver and distribute lipids from there to all over the body.[63] To supply the body tissues, VLDL are metabolized and lose their lipids, decreasing in size and as a result, become LDL. LDL are recovered by the liver by entering hepatic cells through the LDLR pathway. [63,64] Flavivirus may interact with lipoproteins, which can lead to the formation of lipoviroparticles (LVPs).[59,66,67] Upon formation, these particles are excreted to the bloodstream as normal lipoproteins and enter new cells by following the LDL-receptor pathway (LDLR), achieving high efficiency.[68,69] Also, these LVPs are much more infective than mature virions and with this lipoprotein camouflage, they can easily evade detection by the immune system.[66] In fact, VLDL are quite similar to LDs in composition.[63,64] They are composed by a hydrophobic core containing TAGs and CE, that is surrounded by a phospholipids and cholesterol monolayer and thus, the study of these lipoproteins in the context of the C protein might provide better grounds for the understanding of Flavivirus infection.[63,64,70] (**Figure *1.5***).

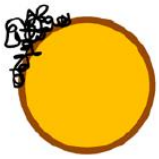
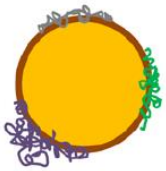
Lipid system	LDL	VLDL	LDs	LVP[a]
Structure				
Density (g/mL)	1.019 – 1.063	0.930 – 1.006	–	~1.0
Diameter (nm)	18 – 25	30 – 80[b]	100 – 1000[c]	20 – 80
Cholesterol (%)	50	20	20	15
Triglycerides (%)	10	50	50	45
Phospholipids (%)	20	20	20	15
Proteins (%)	20	10	10	25
Proteins	ApoB100	ApoB100, ApoE, ApoC-I, ApoC-II, ApoC-III	Perilipin 1, Perilipin 2, Perilipin 3	ApoB100, ApoE, ApoC-I, C-II, C-III, Envelope, Capsid

Figure 1.5 – Structure, physical properties, biochemical composition and proteins of LDL, VLDL, LDs and LVPs. All these systems are constituted by a core of neutral lipids (mainly triglycerides and cholesteryl esters; colored yellow), surrounded by a monolayer of polar lipids (mainly phospholipids and cholesterol; colored brown) where specific proteins are embedded. However, they have different lipid compositions, giving them different densities. Their biochemical and protein compositions are explained by its formation pathway and also by its function. All these systems are ER-derived. LDL, VLDL and LVP are extracellular structures that have similar proteins: apoB100 (black) is present in all of them, VLDL and LVP also contain apoE (red) and apoCs (dark green). LDs main proteins are perilipins 1, 2 and 3 (violet, grey and light green). LVPs, as chimerical structures related with virions and with VLDL, also contain viral envelope proteins (E; cyan) and a nucleocapsid in its core, composed by viral genomic RNA (orange) condensed with the capsid protein (C; blue). As it can be observed, VLDL are very similar to LDs and LVP in terms of lipid composition, composed of a hydrophobic core, surrounded by an external phospholipids and cholesterol monolayer, with proteins embedded. The main VLDL proteins are the apolipoproteins B100, E, CI, C-II and C-III. As VLDL are metabolized in the bloodstream, they loose lipids and some of the proteins attached to their surface and become LDL. Drawings are representative and not at the same scale. Adapted from [70].

1.3 Zika virus epidemics and global impact

ZIKV was accidently discovered 70 years ago, during investigations for the vector responsible for the non-human cycle of yellow fever in Uganda.[71] The first and second virus isolation were made from the serum of febrile rhesus monkey (**Figure *1.6A***) and from *Aedes africanus* mosquitoes (**Figure *1.6C***) in the canopy of Zika forest near Lake Victoria (**Figure *1.6B***)[72,73]. Therefore, ZIKV received its designated name from the geographical area where the initials isolations were made from. [72]

Figure 1.6 –Zika virus wild reservoirs, location of isolation and mosquito vector. (A) Rhesus monkey where the first isolation of Zika virus was made from, extracted from the Zika forest. **(B)** Zika forest near Lake Victoria, Uganda – Localization of the first Zika virus identification. **(C)** Mosquito *Aedes africanus* from which the second isolation of Zika virus was made.

The first human cases related to ZIKV are detected in 1952 in Uganda and Tanzania.[74] Since then, several human cases were confirmed through blood tests, despite no reports of deaths or hospitalizations.[71] Studies of the distribution of ZIKV report the wide circulation of the virus, showing a probable widespread human exposure to the virus.[75] In 1964, a researcher was infected with ZIKV classifying the illness as "mild" thus confirming that ZIKV causes human diseases.[76] The fact that a low number of human cases had been related to ZIKV in Africa might be due to under diagnosis since in this region DENV and Chikungunya virus (CHIKV) also circulates and have similar symptomatology to ZIKV's.[73] A geographical expansion across a wide area of Africa was observed and mosquitoes containing ZIKV were also detected in equatorial Asia including India, Indonesia, Malaysia and Pakistan. [74,76]

The first large outbreak of ZIKV was reported at the Island of Yap, Micronesia in 2007.[77] Prior to this outbreak, only 14 human cases were reported worldwide. During the outbreak, an estimated 73% of Yap's residents were infected with ZIKV but not a single hospitalization, hemorrhagic manifestation or death were reported.[73,77] The clinical condition was characterized by rash, conjunctivitis and arthralgia and initially thought it was caused by dengue virus.[78,79] How the virus was introduced is still unknown, but it was proposed that it was due to an infected mosquito or to an asymptomatic person with undetected infection.[78] This also suggests a lack of immunity in the islands' population, since a regular exposure to infection (like the populations in Africa and Asia) might have prevented this large outbreak as well as the one seen in the Americas.[73,80] Also, under-reporting due to the clinical similarities of (mild) illness symptoms associated with ZIKV, DENV and CHIKV might account for previous ZIKV outbreaks, being overlooked and miss-classified.[73,81]

In 2008 the first case of sexual transmission is documented, in which a US scientist was conducting field work in Senegal and got infected with ZIKV.[82] When returned home to Colorado, his wife got infected, being the first documented case of sexual transmission of a disease that was usually transmitted by insects.[83] Later 2012, researchers identified 2 distinct lineages strains of the ZIKV, one African and the other Asian.[71]

During 2013-2014, several outbreaks happened in distinct groups of Pacific islands: the French Polynesia, Easter Island, the Cook Islands and New Caledonia.[73,80] Regarding the outbreak in the French Polynesia, up to 11% of the population was affected.[79,84] Most clinical cases presented low fever, asthenia, wrist and fingers arthralgia, headache, rash and one patient presented Guillain-Barre syndrome (GBS).[80] These reports indicated a possible association between ZIKV, congenital malformations and severe neurological and autoimmune complications. [79] In fact, it is noteworthy that after this outbreak, there was an increase of 20-fold in the incidence of GBS in the French Polynesia.[73,80] Also, the subsequent observed outbreaks in the Easter Island and in the Cook Islands were imported infections from the French Polynesia.[78] First autochthonous case in New Caledonia was then also reported. Since then several autochthonous cases were reported in the Oceania.[84]

In early 2015, the first autochthonous case of transmission of ZIKV in Brazil was reported and, since then, the virus quickly spread across the Americas.[82] Brazilian Ministry of Health estimated that between 400000 to 1300000 cases of ZIKV C infections occurred in 2015 in the country.[85] Since the initial report in Brazil, 29 countries or territories of the America also reported autochthonous cases of ZIKV infections.[86] Despite these outbreaks, imported cases in the USA and Canada are starting to show, with, respectively, 82 and 3 cases reported since then in these countries.[87] The risk of ZIKV establishment in the Canada and USA is very low due to the absence of the transmission vectors but it cannot be discarded.[73] Actually, autochthonous transmission was observed with other related flaviviruses like WNV. [82,88]

Currently in Europe, there is no evidence of autochthonous ZIKV infection.[73] All reported cases were imported from travelers returning from affected countries. There is evidence of colonization of *A. albopticus* mosquito in Europe, mainly in the Mediterranean area, which can also transmit ZIKV.[78,89] There is also the possibility of the establishment of the main transmission vector (*A. aegypti* mosquito spp.) in the continent, which might lead to a spread and to autochthonous case of ZIKV infections.[73] In fact, autochthonous DENV and CHIKV infections due to this vector were documented in France, Croatia and Italy.[90–92] The same could occur with ZIKV transmission in Europe.

1.3.1 ZIKV transmission

ZIKV natural transmission cycle is assured by arthropod vectors, namely mosquitoes of the genus *Aedes*.[93] Usually most arboviruses are perpetuated in transmission cycles independent of human hosts, but those with sylvatic cycles tend to only infect people who accidentally intrude on their natural habitats.[73,94] Nonetheless, humans are dead-end hosts in complex transmission cycles that involve different wild and domestic vertebrate hosts.[73,94] The virus was first isolated from *A. africanus* but, since then, several other *Aedes* spp. revealed to be competent vectors, like the *A. aegypti* (the main vector of ZIKV nowadays) and *A. albopictus* which has preference for humans.[72,95,96] This indicates that humans probably serve as primary amplification hosts when their viremia is sufficient in duration and magnitude.[71,73] It is still unknown if ZIKV overwinters in some geographical areas but, as in other flaviviral infections (such as that caused by WNV), epidemics are more related to specific mosquito species, its population density, competence and behavior in each area.[73,88] This can shape the virus dynamics. [71] *Aedes aegypti* is currently distributed in Asia, Oceania, the Americas and in some regions of Africa and Europe, including Portugal, in Madeira island.[73,97] Recent predictions suggest that this species might soon colonize southern Europe, as well as North America and Australia.[73,97] This mosquito does not overwinter but can be sheltered in domestic settings that provide protection against environmental conditions.[73,88] Currently this species does not circulate in Europe, but once it is

introduced in the continent there are no climatic reasons to believe that it cannot become widely established in southern Europe.[91,97] Looking at the geographical evolution of the distribution of this species (**Figure *1.7***) it is possible to observe established colonies in Turkey (red areas), and new-introduced colonies in Egypt (yellow). This new introduced colonies in the Egypt might be a bridge for further implementation of *A.aegypti* in the Mediterranean.

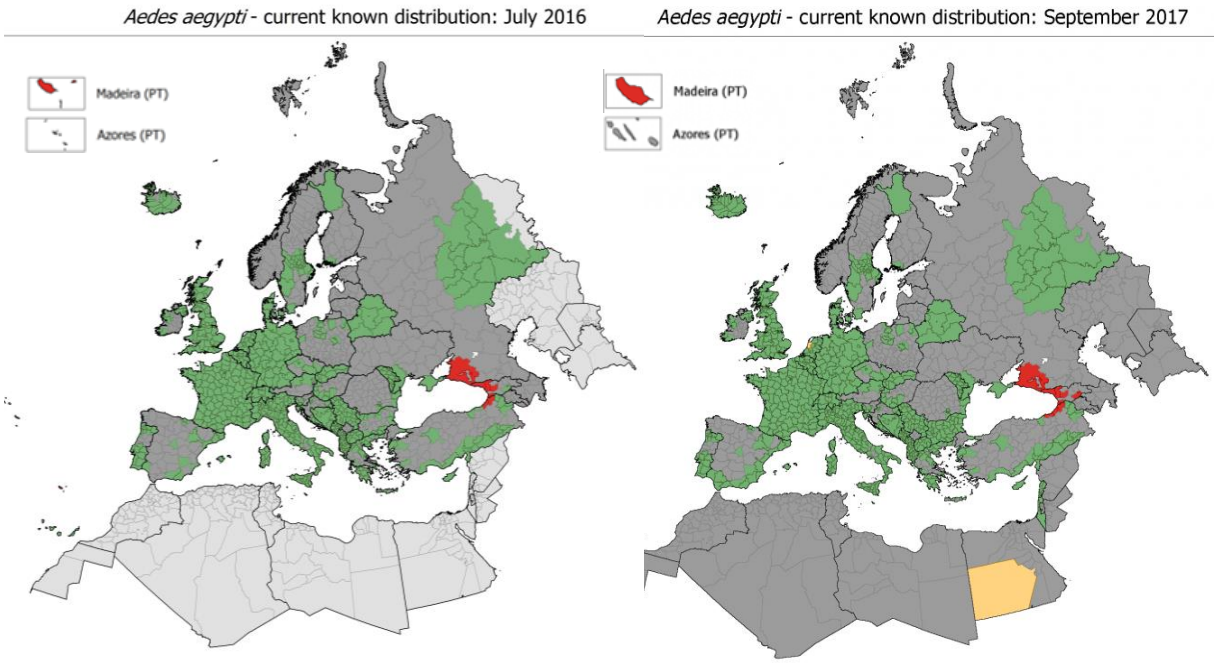


Figure 1.7 – European distribution of *Aedes aegypti* changes in a year (2016 to 2017). It is observable an increase in already established area surrounding Turkey (red) and newly introduced colonies in the Egypt area (yellow). Green areas are where this species is currently absent. It is also found in Madeira island, Portugal. Data extracted from European Centre for Disease Prevention and Control.

Aedes albopictus is also called Asian tiger mosquito and is widely distributed. This species is currently circulating in Asia, North, Central and South America, northern Australia, some areas of Africa and in southern Europe, where it has spread to France, Germany, Italy, Spain and, possibly, mainland Portugal.[97–99] This species can hibernate and survive in temperate regions and is the most invasive mosquito species in the world.[96,100,101] It can adapt to different climates through the production of cold-resistant eggs, with temperate strains surviving cold winters in northern latitudes.[73,100] On top of that, it prefers container habitats (e.g., tires and vases) in domestic settings which resulted in the increase potential for this species to contact with humans.[73,96] Looking at the geographical evolution of the distribution of this species between 2016 and 2017 (**Figure *1.8***) it is possible to observe that the green areas the *A.albopictus* is absent, the yellow areas is recently introduced and red areas is already established.

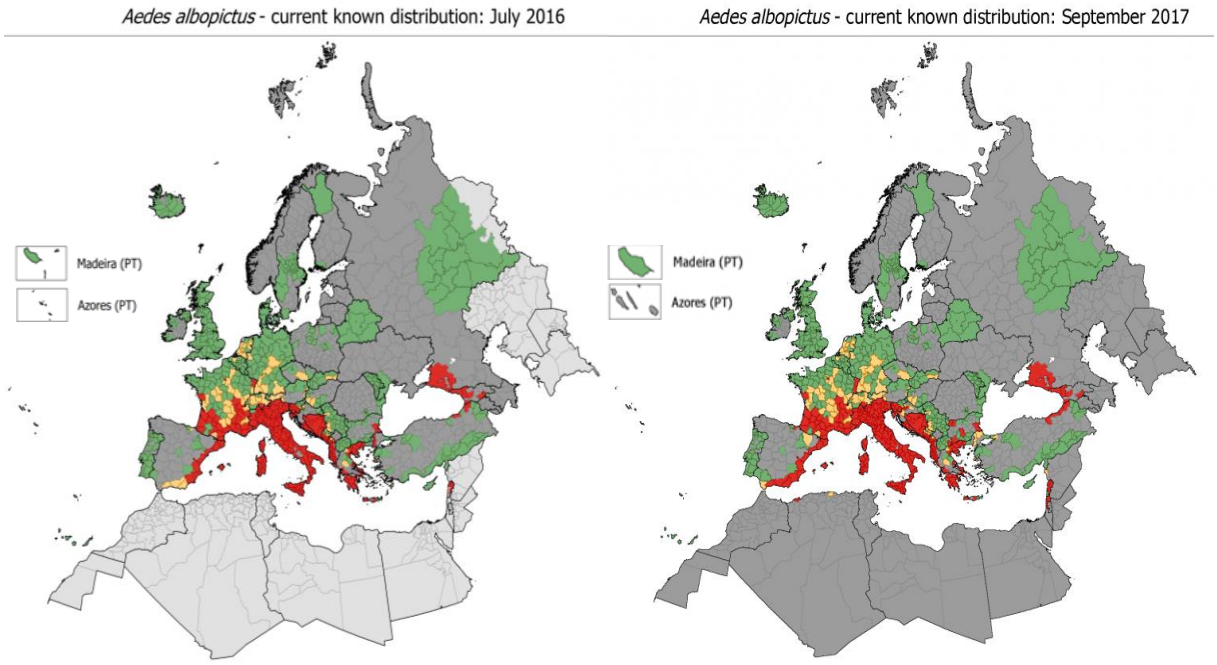


Figure 1.8 – Europe distribution of *Aedes albopictus* changes in a year (2016 to 2017). Established areas are red, introduced areas are yellow and absent areas are green. An increase of established colonies is observable around the Mediterranean and the Black sea. Data extracted from European Centre for Disease Prevention and Control.

This ability of ZIKV to be efficiently transmitted by both mosquito species further complicates the control of the virus.[73] Both species grow close to human populations but *A. aegypti* feeds almost exclusively on humans in daylight hours and typically rest indoors.[102] *A. albopictus* is usually exophagic and bites humans and domestic and livestock animals, although under some circumstances it preferentially feed on humans thus having an anthropophilic behavior like *A. aegypti*.[98] This brings some underlying difficulties in the control of both species, since a method of control for a species may not serve to control the other one.[73] Furthermore, upon reducing populations of *A. aegypti*, the opportunistic invasive *A. albopictus* may rapidly move into the area.[101]

In the recent outbreak in Brazil, nearly 80% of ZIKV infected people were asymptomatic.[103] Viral RNA has been identified in brain, placenta and amniotic fluid specimens.[73] The presence of RNA in those places is associated to microcephaly in infants and miscarries during pregnancy.[86,87] Sporadic reports of direct human-to-human transmission have been reported to occur perinatally, sexually and through breastfeeding and blood transfusion.[73] Perinatal transmission from two mothers to their newborns during the French Polynesia outbreak were documented, although contamination during delivery was not discarded.[104] The sera from the mothers were positive for ZIKV within 2 days post-delivery and those of their newborns within 4 days after birth.[73,104] High ZIKV's RNA load was detected in breast milk samples from both mothers.[73,104] Besides these sporadic cases of non-vector transmission, one of the most surprising phenomenon on ZIKV infection is the unexpected number of infants born with microcephaly during the Brazilian outbreak.[73,101] Around 4000 cases of ZIKV-related microcephaly have been reported in that country since 2015.[73,101] Furthermore, ZIKV was recently found in fetal brain tissue of a baby with microcephaly after termination of the pregnancy.[13] Also, reports in samples of semen and urine of a patient suggest that ZIKV could be sexually transmitted.[83,105] Potential transmission through blood transfusion was demonstrated during the French Polynesia outbreak, in which 3% of asymptomatic blood donors were positive for acute ZIKV infection by RT-PCR (real time – polymerase chain reaction).[73,105]

1.3.2 Symptoms and diagnosis of Zika infection

In early reports, ZIKV infection was clinically manifested as mild and self-limiting symptoms, as such febrile illness, rash, arthralgia and conjunctivitis without severe complications and with a low hospitalization rate, often mistaken with other arboviral infections like DENV or CHIKV.[73] In fact, only around 18 % of the cases have been reported to be symptomatic.[12,73] Although, since the French Polynesia outbreak that perception was changed, and reports of severe neurological complications started to appear.[73,87] An unexpectedly high number of GBS cases were observed in the French Polynesia outbreak, about 20-fold higher incidence than expected.[80,106] All GBS patients developed neurological symptoms following a "Zika-like syndrome" episode (low-grade fever, myalgia, rash and conjunctivitis).[73,80,106] GBS is an autoimmune disease causing acute or subacute flaccid paralysis that can even cause death and it has been previously associated with other flaviviruses infections (DENV, WNV or CHIKV).[80,81,106] Besides the French Polynesia outbreak, during the outbreak of Colombia, 86 cases of GBS have been associated to ZIKV infection, thus being hypothesized that ZIKV is a trigger to GBS.[13,73,107] These changes in the clinical overview of the ZIKV infection point to a worrisome increase in the potential clinical severity of the disease. [81,85,106,108,109]

Similarly to GBS, even more disturbing is the astonishing rise in the number of newborns with microcephaly and neurological disorders since the ZIKV outbreak in Brazil. [13,86,87] The annual number of reported cases of microcephaly in Brazil ranged from 150 to 200, but since the outbreak this number increased to near 4000 cases reported. These congenital infections due to ZIKV exposure have also been associated with an increase in vision-threatening findings (bilateral macular and perimacular lesions and optic nerve abnormalities). Successful RT-PCR amplification of the complete ZIKV genome from a fetal brain tissue and amniotic fluid samples has been recently described. Also, an expectant mother presented a febrile illness with rash, while she was living in Brazil. Ultrasonography performed at 29 weeks of gestation revealed microcephaly with calcifications in the fetal brain and placenta. After request of termination of the pregnancy, the fetal autopsy revealed microcephaly with almost complete agyria (presence of silver compounds in the body), hydrocephalus (accumulation of cerebrospinal fluid in the brain), and multifocal dystrophic calcifications in the cortex and subcortical white matter, associated with cortical displacement and mild focal inflammation. No virus or pathological changes were found in any other organs, suggesting that the virus is strongly neurotropic, which means it preferentially attacks the nervous system. Electron microscopy analysis revealed spherical virus particles with morphologic characteristics consistent with ZIKV. Facing this risk of microcephaly upon ZIKV infection, WHO declared a public health emergency of international concern on February 1st, 2016. [107,110,111]

Currently, different arboviral infections can have similar clinical presentations thus their circulation may be underreported if specific diagnostic tool has not been implemented.[73] In the case of ZIKV infections, diagnosis presents several drawbacks since antibodies often cross-reacts between flaviviruses (which limits the use of serology), viral culture is not routinely performed and so far, there is no antigenic detection test available.[112] Besides this, at the present, ZIKV infection diagnosis is mainly made through molecular commercial tests by RT-PCR and serologic assays (IgM ELISA).[73,113] Relatively to the RT-PCR method, the detection of the virus should be performed during the first 3 to 5 days after the onset of symptoms, since the viremic period is short.[113] Besides blood, saliva and urine samples are also suitable for the detection of the virus, being recommendable the collection of blood and saliva for the molecular diagnosis of ZIKV during the acute phase of Zika fever and urine for the later stages of the disease.[73,113] Thus, a combination of those three biological samples is recommended for an increase of sensitivity of the detection of ZIKV infections. [73,85,113] At the present day, there are no specific commercial antiviral agents or vaccine for ZIKV.[73] Patient are commonly treated for symptom relief. There are more than 40 vaccine candidates in pipeline and 5 are entering Phase I of clinical trials.[73] Despite this fact, it will take time until any drugs or vaccines against ZIKV to be commercially available.[73]

1.4 Aims and goals

DENV and WNV viruses are closely related to ZIKV, inclusively in terms of the sequence of their highly homologous capsid protein.[48] In terms of structure, both DENV and WNV C protein have been determined, by nuclear magnetic resonance (NMR) and X-Ray crystallography, respectively.[46,47] Both proteins display alpha-helical structure, with 4 helices. The recently published ZIKV C truncated structure is similar to WNV and DENV C proteins structure.[45] There is a common "conserved" fold that, roughly, superimposes from residues 33rd to the C terminus and corresponds to 3 alpha helices.[48] The first alpha-helix does not superimpose and the N-terminal region, roughly defined as amino acids 1 to 26 is disordered in solution. A peptide based on a conserved motif between 16 mosquito-borne flavivirus, pep14-23 [48], and corresponding to conserved residues 14 to 23 of the N-terminal region, gains α-helical structure in the presence of negatively charged phospholipids. [114] This region is involved in DENV C binding to host lipid systems.[67,70,114] The N-terminal region may play a similar role in other flaviviruses, being relevant to study its possible conformations and relating biological activity, since it has been proposed to have a autoinhibitory mechanism/function.[114] Given that is a disordered region, studying it by traditional X-Ray or NMR based approaches is difficult since these regions usually are cleaved, losing the information and relevance of the N-terminal region in a biological context of the capsid.

The advances observed in the past years constitute an important development in the field of *Flaviviridae* research particularly in the understanding of flavivirus life cycle, namely the closely related DENV and WNV. Nonetheless, ZIKV C full length protein structure is still unknown and the proposed interactions with host lipid systems need to be further investigated. Similarities with DENV, namely C protein interaction with host lipid systems (LDs and lipoproteins), are expected. If so, it is possible to postulate strategies to inhibit crucial key steps in ZIKV replication, using the potential pep14-23 and its derivates as inhibitors. This is a line of research being developed at the host laboratory (NSantos Lab, *Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa*, Lisbon, Portugal).

The present thesis aims to unfold the molecular details behind the ZIKV C protein interactions with host lipid systems, since it may elucidate key processes occurring *in vivo*. Thus, a combination of several *in silico* and *in vitro* studies will result in beneficial advances of fundamental knowledge of Zika virus and provide ground for future improvement of therapies for ZIKV. A computer-aided analysis and computational study was performed to the ZIKV C sequence and structure by using algorithms already implemented in the host lab. Also, a new computational method was developed to inquire if the process of dimerization occurs for the predicted structure. To study the interactions between ZIKV C and host lipid systems firstly it was necessary to express and purify the ZIKV C protein, with circular dichroism spectroscopy (CD) and mass spectroscopy used to characterize the expressed ZIKV C protein proprieties. Then, zeta potential (ζ-potential) was used to analyze the binding of ZIKV C protein to LDs (isolated from a kidney cell line to mimic *in vivo* conditions as closely as possible) and evaluate the role of charges in these interactions. Dynamic light scattering (DLS) spectroscopy allowed to determine the size increment of ZIKV C – lipoproteins complexes formed, by allowing ZIKV C and lipoproteins isolated from blood plasma to interact. DLS was also used to investigate the ability of pep14-23 to inhibit the ZIKV C – lipoproteins interaction. Finally, all the information gathered was crossed and models describing these interactions are proposed (**Figure *1.9***).

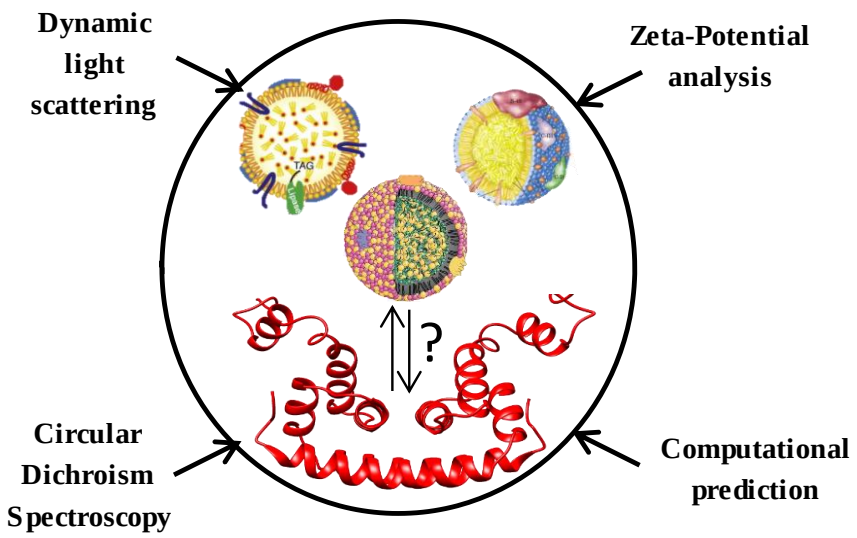


Figure 1.9 – Proposed study of ZIKV C interaction with host lipid systems. It was hypothesized that ZIKV C protein interacts with lipoproteins and lipid droplets. Biophysical and computational approaches allow the characterization of the protein in the context of these interactions.

2 Materials & Methods

2.1 Materials

The first batch of VLDL (batch 1) was obtained from Kalen Biomedical LLC (Montgomery Village, MD, USA). The second batch of VLDL (batch 2) and LDL were obtained from Merck Milipore (Merck KgaA, Darmstadt, Germany). All lipoproteins were isolated from human plasma by ultracentrifugation. VLDL from Kalen Biomedical LLC were delivered at 1 mg/mL in a solution of 154 mM NaCl, 5.6 mM Na2HPO4, 1.1 mM KH2PO4 and 0.34 mM EDTA, pH 7.4 and stored at 4 ºC. VLDL from Merck Milipore were delivered at 5 mg/mL concentration in 150 mM NaCl, 0,01% EDTA, pH 7.4 and stored at 4 ºC. LDL from Merck Milipore were delivered at 4,6 mg/mL concentration in 150 mM NaCl, 0,01% EDTA, pH 7.4 and stored at 4 ºC. The peptide inhibitor pep14-23(H-NMLKRARNRV-NH2), a peptide rationally designed at the host lab and protected under an international patent application (WO 2012159187) was custom synthesized by JPT Technologies GmbH (Manheim, Germany) with > 95% purity (confirmed by reverse-phase HPLC and ESI-MS analysis). ZIKV C Brazil sequence (GenBank: KU497555.1) was selected due to the association of ZIKV Brazil strain with microcephaly[110] and pET21a was to maintain the expression plasmid of DENV C protein. The ZIKV C protein gene (encoding residues 1 to 104) was commercially cloned into a plasmid pET21a (NZYTech, Lda, Lisbon, Portugal). ZIKV C was then expressed and purified in *E. coli* via recombinant protein expression following the protocol optimized and described in the Results section. Two different buffers were used in all experiments: TEE-KCl buffer (20 mM Tris-HCl, 100 mM KCl, 1 mM EDTA and 1 mM EGTA, pH 7.4) and TEE-NaCl buffer (20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA and 1 mM EGTA, pH 7.4). Cell culture was performed with the high-glucose Dulbecco's modified Eagle's medium (DMEM) with 0.01% sodium pyruvate and 4 mM L-glutamine, supplemented with 10% fetal bovine serum, 100 U/mL penicillin and 100 µg/ml streptomycin. Oleic acid was used at a concentration of 10 mM. To disrupt cells a cell disruption vessel model 4639, from Parr Instrument Company (Moline, IL, USA) was used. Unless otherwise stated, all other chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2 Prediction of flavivirus C protein tertiary structure

To predict the tertiary structure of ZIKV C from its amino acid sequence, the online server I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER) was used. I-TASSER *ab initio* prediction suggested 5 monomeric conformations. The same experiment was run, excluding as templates DENV and WNV C structures. In both cases, the same 5 monomeric conformers were predicted (only minor differences were observed when comparing predictions with and without C protein templates). The conformers resulting from the prediction excluding DENV and WNV C template structures were used. From these 5 monomers, a conformational study was performed (described further), where 3 conformations were discarded due to high stereochemical constrains and clashes in the context of the formation of a dimer structure. The remaining conformations were superimposed with DENV C homodimer partial structure to generate the homodimer structures presented. This was performed via UCSF Chimera 1.8.1 software package.

The same was conducted for 16 highly homologous mosquito borne flavivirus C protein sequences, previously identified by the host group[48]. At each iteration, the server does a homologous comparison between the given sequence and the database and associates a proper structure available for that specific sequence. After that, energy minimization of the free energy of the conformer structure is performed. An output of 80 different predicted structures was obtained, five for each flavivirus. In a biological context, the flavivirus' capsid protein is expected to form a dimer, so all 80 predicted

structures needed to be analyzed to determine if they allowed dimerization. Structures were analyzed regarding stereochemical constrains by superimposing monomers of each virus with DENV C dimer structure,[114] in line with a similar approach for DENV C alternative' N-terminal conformations[114], monomers that do not permit a dimer to be formed in the "conserved fold" region were discarded. Based on analysis of N-terminal backbone clashes with the other monomer, conformers which had more than 6 clashes (not allowing a homodimer structure) were discarded.

2.3 Protein structure visualization

For protein structure visualization and preparation, the UCSF Chimera v1.9 software was used. PDB file 1R6R was used as a experimentally determined DENV C structure (lacking the first 20 residues – N-terminal region) and as a template for all comparisons made in here.

2.4 Circular dichroism

Circular dichroism (CD) spectroscopy is based on the molecular property named chirality, which is a property of many biological active molecules such as proteins and sugars. Two molecules are chiral if they have the same chemical composition, the same functional groups and similar/equal physical properties but a different special orientation of its constituting chemical groups, being the mirror image of each other but not superposing (such as the left and right human hands). Good examples of chirality are the peptide bond linking amino acids together to form peptides and proteins. Peptide bonds react differently to polarized oriented light. This light can be circularly polarized to the right or to the left and these compounds absorb the light differently, causing an angular difference in the sum of the incident light vector and this angular distortion generates an elliptical shift (ellipticity), normally expressed in millidegree (mdeg). This is the basis of CD spectroscopy. [115,116]

Hence proteins are very active chiral compounds due to the peptide bound, CD spectroscopy is widely used in protein studies. In particular, far-UV CD spectroscopy is a powerful technique to inquire about protein secondary structure due to the conformational constrains of the peptide bond and to the spatial occupancy and arrangement (stereochemistry) of the surrounding atoms and bound. In this manner amino acid residues in the protein can only occupy restricted regions of the ϕ and ψ dihedral angles. These spatial constrains prohibit some photo-excitation transitions of the peptide bond delocalized electrons, originating different contributions to the CD spectra according to the protein structural motif, as demonstrated in **Figure *2.1A***. α-helix motifs are characterized by electronic transitions $\pi \rightarrow \pi^*$ (causing a positive peak around 192 nm and negative peak around 208 nm) and $n \rightarrow \pi^*$ (negative peak around 222 nm) (**Figure *2.1B***, yellow), β-sheet motifs have transitions $\pi \rightarrow \pi^*$ (positive peak, around 195 nm) and $n \rightarrow \pi^*$ (negative peak, around 217 nm) (**Figure *2.1B***, blue). Non α-helix nor β-sheet oriented regions, named disordered regions or random coil, have transitions $\pi \rightarrow \pi^*$ (negative peak, around 195 nm) and a small or inexistent transition $n \rightarrow \pi^*$ (**Figure *2.1B***, red). CD is a quick technique routinely used to determine protein secondary structure. It also allows the study of structural changes on the protein backbone as a result of different conditions (temperature, pH), to access the state of folding and the thermodynamic stability of proteins, being employed in this project with that aim.

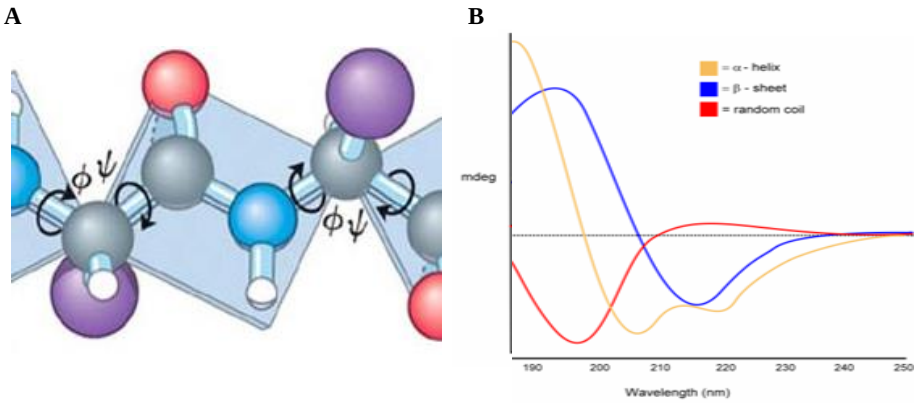


Figure 2.1– Circular dichroism spectroscopy of proteins. (A) The peptide bound is chiral and a chromophore in the UV region of the electromagnetic spectrum. Its conformational restrictions affect the allowed electronic transitions, making him a good sensor of the protein local fold. **(B)** The different protein secondary structure has characteristic CD spectra: α-helix motifs have a CD spectrum similar to the yellow spectra; β-sheet motifs have a CD spectrum similar to the blue spectra and disordered regions/random coil have a CD spectrum similar to the red spectra. Adapted from [64,112].

2.4.1 Assessment of ZIKV C stability by temperature denaturation followed by CD

To characterize the ZIKV C stability, studies of denaturation through temperature were performed using CD spectroscopy. CD measurements were performed in a JASCO J-815 (Tokyo, Japan), using 0,1 cm path length quartz cuvettes, data pitch of 0,5 nm, velocity of 200 nm/min with a data integration time (DIT) of 2 s and performing 3 accumulations. Spectra were acquired in the far-UV region, between 200 nm and 260 nm, with 1 nm bandwidth. Temperature was controlled by a JASCO PTC-432S/15 Peltier equipment and the range of temperatures studied was from 6 ºC to 96 ºC, in steps of 2 ºC, increasing at a rate of 8 ºC/min. Before and after denaturation, spectra were obtained at 25 ºC, to confirm a process of reversibility. ZIKV C concentration was 15 µM (monomer) in 50 mM KH_2PO_4, 200 mM KCl, pH 6,0 with 220 µL of final volume. The thermal profile obtained experimentally was fitted according to a two-state unfolding model, in line with previous work[117], and expressed by:

$$Y = \frac{(y_n + m_n T) + (y_d + m_d T) \times e^{\left(\frac{\Delta H_m}{R} \times \left(\frac{1}{T_m} - \frac{1}{T}\right)\right)}}{1 + e^{\left(\frac{\Delta H_m}{R} \times \left(\frac{1}{T_m} - \frac{1}{T}\right)\right)}} \quad (1)$$

where Y is the measured ellipticity, DHm is the enthalpy at the unfolding transition, Tm is the melting temperature, T is temperature in Kelvin, and R is the universal gas constant. The pretransition baseline slope and intercept is described by mn and yn, respectively, while md and yd are the slope and intercept of the post-transition baseline.

2.5 Cell culture and production of lipid droplets

LDs were isolated from baby hamster kidney cells (BHK-21) through a method established at the host laboratory.[50] BHK-21 cell line was maintained in high-glucose Dulbecco's modified Eagle's medium (DMEM) with 0,01 % sodium pyruvate and 4 mM L-glutamine, supplemented with 10 % fetal bovine serum (FBS), 1 % penicillin/streptomycin in a T75 culture flask. Cells were grown at 37 ºC, in a humidified 5 % carbon dioxide incubator. When approximately 80 % confluence was reached (after 24 h) the cells were treated with trypsin and divided to three T75 culture flasks. When 80 % confluence was again reached in the three T75 culture flasks (approximately 48h after), each culture flask was treated with trypsin and the cells were divided to ten T75 culture flasks. The resultant thirty T75 culture flasks were maintained at 37 ºC in a humidified 5% carbon dioxide incubator. After 72 h, when

approximately 80 % confluence was reached in all culture flasks, cells medium was replaced for DMEM without FBS or antibiotics, and cells were treated with 10 μM oleic acid.

2.6 Purification of lipid droplets

LDs were isolated after 24 h incubation with oleic acid. To isolate LDs, cells were washed twice and resuspended in 3 mL TEE buffer 100 mM KCl with a protease inhibitor cocktail (Roche Diagnostics GmbH, Mannheim, Germany). Cells were disrupted by nitrogen cavitation at 700 lb/in^2 for 20 min at 4ºC using a cell disruption vessel. The resultant lysate was centrifuged at 1,500 × g for 10 min to remove the nuclei, and the supernatant was collected and mixed with an equal volume of TEE-KCl buffer containing 1,08 M sucrose. The sample was transferred into a 12 mL ultracentrifuge tube and a sucrose gradient was formed by sequentially overlaying 2 mL of 0,270 M sucrose in TEE-KCl buffer, 0,135 M sucrose in TEE-KCl buffer and TEE-KCl buffer without sucrose (***Figure 2.2A***). The sample in the sucrose gradient was centrifuged at 250,000 × g for 70 min at 4ºC (***Figure 2.2B***). Three fractions were collected from the top to the bottom of the gradient, fraction 1 (1 mL), fraction 2 (1,5 mL) and fraction 3 (1 mL) (***Figure 2.2C***). The gradient fractions were kept at 4ºC before use. Only fraction 2 was used in the related experiments, as it was previously proven to be absent of cytosolic contamination, lowest at lactate dehydrogenase activity and enriched in the major LDs protein components. Before use in experiments, fraction 2 was tested to ascertain that it had it showed a ζ-potential value within the ranges previously reported.[50]

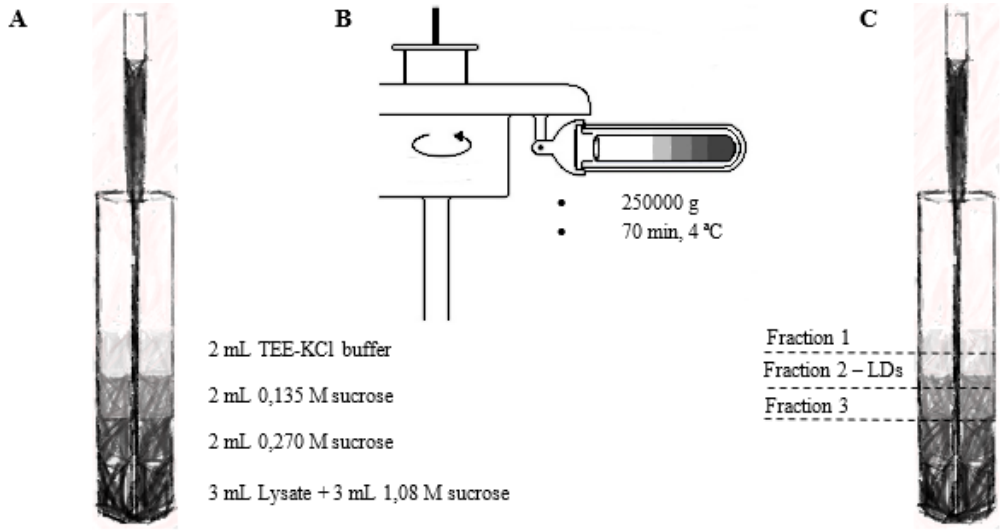


Figure 2.2 – Schematic representation LDs purification protocol. (A) Lysate was transferred to 1,08M of sucrose in TEE-KCl buffer, then the remaining layers, were subsequently overlaid with sucrose gradients (B) The tube was centrifuged at 250,000 g for 70 min at 4ºC and (C) fractions were collected. Adapted from [118].

2.7 Zeta-potential

Particles with charge in a solution attract ions of opposite charge to their surface, forming a layer, covering the particle surface, commonly named by the Stern layer (**Figure *2.3A***). Another layer is formed beyond the Stern layer, where ions diffuse more freely and are not strongly bound to the particle surface as in the Stern layer. As the particle travels through the solution, the ions in the Stern layer move with it, in contrast with those from the diffuse layer, which do not move along with the particle. The potential in the boundary between the Stern layer and the diffuse layer is called ζ-potential.

[119] (**Figure *2.3A***). The ζ-potential is calculated based on the electrophoretic mobility of the particles in solution to the electrode of the opposite charge, in the presence of an electric field (**Figure *2.3B***).

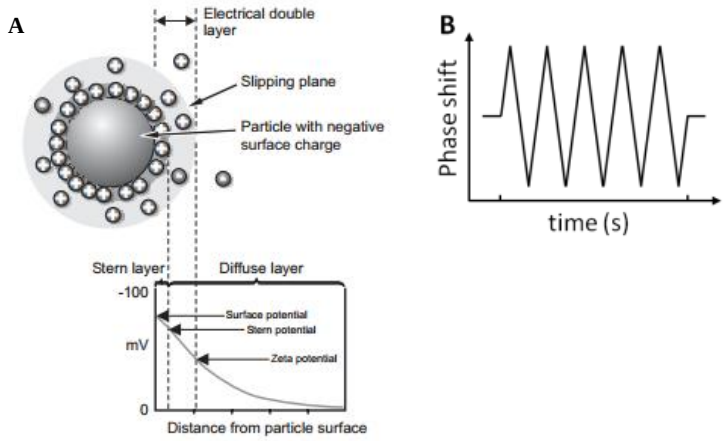


Figure 2.3 – The principle of ζ-Potential. (A) The ζ-potential of a particle is proportional to its surface charge density. **(B)** By performing an electrophoresis in solution, with an electric field that changes polarity, the particles in solution move back and forward, causing a phase shift in the scattered beam that is measurable over time, to extract the electrophoretic mobility of the scattering particle (*u*), proportional to ζ-potential (equation 1). Adapted from [117].

The movement of a certain particle can be opposed by viscous forces in the solvent, until equilibrium, in which a constant velocity can be reached. There are two light scattering spectroscopy-based methods to calculate the electrophoretic mobility of the scattering particles in a certain sample, which are the laser Doppler velocimetry and phase analysis light scattering (PALS). The first one (laser Doppler velocimetry) relates the frequency measured by intensity fluctuation of the scattered light. This method has a very low sensitivity (due to Doppler effect) to the low mobility of larger particles, implying some difficulties in the calculus of the electrophoretic mobility. PALS measures differences in phases between the reference beam and the sample-scattered shifted beam thus being a more sensitive method **Figure *2.3B***. This phase shift can be related to the position of the particle. ζ-potential can be accurately determined for measurements with samples containing particle sizes between 5 nm and 10 μm, though this range can be smaller for some equipment.[119] Thus, the electrophoretic motion can be related to the mean phase change with time.[119] The ζ-potential can be calculated through the Henry's relation (more detailed description in [119]):

$$\zeta = \frac{3\eta u}{2\varepsilon f(ka)} \quad (2)$$

where u is the electrophoretic mobility, η the viscosity of the solvent, ε the dielectric constant of the solvent and $f(ka)$ is the Henry's function. When particles are suspended in aqueous solutions with high ionic strength, the value of Henry's function is 1,5 (through Smoluchowski's approximation, that fits well to physiological conditions), and 1 when non-aqueous media are used (Huckel approximation). With this simple technique, the surface charge potential of molecules in solution and the change in this parameter can be used to easily assess if there is any interaction with biologically relevant ligands and evaluate the role of charges in biological interactions.

2.7.1 Analysis of LD's surface charge by ζ-potential

ζ-potential measurements were performed on a Malvern Zetasizer Nano ZS (Malvern, UK) with forward (13º) scattering detection apparatus, equipped with a He-Ne laser (632,8 nm). For the analysis of the ζ-potential measurements the following physical constants were used: n_0 = 1,330; η = 0,8872 cP;

T = 298,15 K, λ = 632,8 nm and θ = 13º. LDs samples were titrated independently with different ZIKV C protein concentrations in TEE-KCl buffer (buffer containing 20mM of Tris-HCl, 1mM of EDTA, 1mM of EGTA, and 100mM KCl, pH 7.4). Samples were equilibrated for 15 min at 25ºC in the device. ζ-potential was measured 15 times, in which each measure contained 100 runs (being the value obtained the average of these runs), with 90 s of waiting between measurements. Disposable zeta cells with gold electrodes (Malvern Instruments) were used, with 40 V of applied current. The variation of the zeta potential for each condition was calculated relative to LDs measured without protein. All conditions were measured independently and in triplicate. Following previous approaches[48,50] LDs were incubated with 10 μM trypsin (to perform a limited proteolysis of the surface proteins) in TEE-KCl buffer for 15 min at room temperature. 1mM of phenylmethylsulfonyl fluoride (PMSF) was added to the mixture for 5 min at room temperature, to stop the reaction. ζ-potential of trypsinized LDs was determined by the measure of independent LDs in a final volume of 842,5 μL. The variation of ζ-potential (Δζ) was calculated by subtracting the value of the ζ-potential of the LDs without ZIKV C from the value of ζ-potential of LDs with ZIKV C, at a certain concentration. The variation in the ζ-potential, can be expressed as a function of ZIKV C concentration, allowing the fit of the experimental data, taking in consideration the cooperativity effect or not, with the following equations:

$$\Delta\zeta = \frac{\Delta\zeta_{max} \cdot [ZIKV\ C]}{C_{1/2} + [ZIKV\ C]} \quad (3)$$

and

$$\Delta\zeta = \frac{\Delta\zeta_{max} \cdot [ZIKV\ C]^n}{C_{1/2} + [ZIKV\ C]^n} \quad (4)$$

where $\Delta\zeta_{max}$ is the maximum amplitude of variation of ζ-potential caused by the interaction between LDs and ZIKV C; the $C_{1/2}$ is where half the concentration of ZIKV C needed to saturate all the LDs is reached and where *n* expresses the effect of cooperativity (if the value is lower than 1 there is a poor cooperativity and if the value is higher than 1 there is cooperativity).

2.8 Dynamic light scattering

Dynamic light scattering (DLS) is a technique that aims the measurement of the particle size [119]. There are other methods that could give a more accurate and complete information about the system (*e.g.*, Atomic Force Microscopy (AFM), electron microscopy (EM), small angle X-ray scattering and small angle neutron scattering), but these have the drawback of requiring expensive equipment. Besides that, the EM technique does not allow measurements of the sample in solution and the remaining techniques needs high values of concentration of the sample, some unfeasible to reach in biological systems.

DLS offers an alternative way to obtain accurate values for particle size ranging from 1 nm to 1 μm and, although sample concentration may affect the measurements, it still supports highly diluted samples, being for this reason, commonly used.[119] **Figure *2.4*** demonstrates the basis of DLS measurements. It measures the scattered light intensity (*I*) fluctuations on a small volume in the time-scale of the molecular diffusion (microseconds) (**Figure *2.4A***).

These fluctuations happen due to the Brownian motion of the particles, caused by the random collision of molecules, and are dependent of the molecule size. The determination of the size of the particle is possible by the measure of the scattering light intensity fluctuations as a function of time (*I(t)*).

This is achieved by employing the second order autocorrelation function ($g_2(t)$), demonstrated in the equation below, which correlates points of the $I(t)$ function, as shown in the equation below:

$$g_2(t) = < I_t . I_{t+\Delta t} > (5)$$

where I_t is a point of the $I(t)$ function and $I_{t+\Delta t}$, another point next (spaced Δt) in the same function. The Siegert principle relates the $g_2(t)$ with the first order autocorrelation function ($g_1(t)$)[120]):

$$g_2(t) = < I_t^2 > . \beta . g_1^2(t) \ + < I_t >^2 (6)$$

where β is a parameter that express the shift from the ideality (where it would take the value of 1).

Through $g_1(t)$, the diffusion coefficient (D) of the particle is obtained by the decay rate (Γ) and the scattering vector (q) of the intensity function using the relation[121]:

$$g_1(t) = e^{-\Gamma t} (7)$$

with:

$$\Gamma = Dq^2 (8)$$

and:

$$q = \frac{4\pi n_0}{\lambda} \sin\left(\frac{\theta}{2}\right) (9)$$

where t is the time, n_0 is the refractive index of the scattering material, λ and θ are the wavelength and the angle of the scattered light, respectively.

Thus, the correlation kinetics (expressed by Γ) depends on the intensity-weighted D and can be calculated using several methods, such as Cumulants or CONTIN. Cumulants use a monoexponential correlogram fit to get the information about an average D, while CONTIN uses multiexponential correlogram fit to assess D distribution in solution [122–125].

By the Stokes-Einstein equation, the hydrodynamic diameter (D_H) is determined from D [121]:

$$D_H = \frac{\kappa T}{3\pi\eta D} (10)$$

where η is the dispersant viscosity, κ is the Boltzmann constant and T the absolute temperature.

By the Mie theory, a scattering intensity distribution function of D_H ($I(D_H)$) can be obtained, and converted to $n(D_H)$, the particle number distribution function of D_H (**Figure *2.4B***). $n(D_H)$ expresses the quantity of particles (with a certain diameter) scatters light, while $I(D_H)$ gives the quantity of light scattered by particles in a certain diameter. Hence the scattering intensity of a particle is proportional to the sixth power of its D_H (Rayleigh's approximation), the conversion can be done by applying the transformation:

$$n(D_H) \approx \frac{I(D_H)}{{D_H}^6} (11)$$

For each measurement performed, the D_H values may be obtained from the peak of the $n(D_H)$ and analyzed statistically using the Mann-Whitney (or Wilcoxon) U unilateral test.

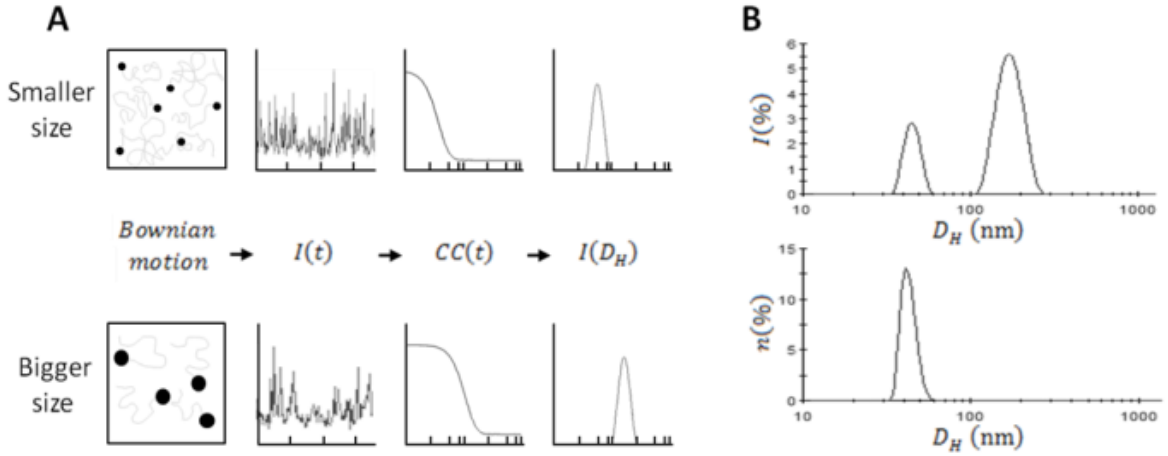


Figure 2.4 – The principle of Dynamic Light Scattering and a flowchart of its analysis. (A) DLS measurements revolve around on the microsecond fluctuations observed in the intensity (*I*) of the light scattered by the sample over a time interval, being related to the sample particles Brownian motion. **(B)** Each intensity value is correlated with the others by the autocorrelation function (*CC(t)* is the correlation coefficient of the *I(t)* function). Larger particles generate higher correlation than the smaller ones, since their motion is slower. The D_H is extracted through the correlation function and the Stokes-Einstein equation. The intensity function of D_H translates into the particle sizes distribution of D_H.

2.8.1 Assessment of lipoproteins interaction with ZIKV C and pep14-23 by DLS

To assess the interaction between lipoproteins with ZIKV C and pep14-23, DLS experiments were carried out on a Malvern Zetasizer Nano ZS (Malvern, UK) with a backscattering detection angle at 173º, equipped with a He-Ne laser (632,8 nm), at 25ºC and using glass cuvettes with round aperture. Lipoproteins (VLDL and LDL) were diluted to a concentration of 50 µg/mL in TEE-KCl buffer. The size of VLDL and LDL were first measured without ZIKV C in the medium. Subsequently, successive additions of ZIKV C were added before each measure, to reach a range of concentration from 0,5 to 5,0 µM and was calculated the D_H for each ZIKV C concentration. Samples were equilibrated for 15 min at 25ºC in the equipment before each set of measurements (10 measurements in which each measure contained 10 runs (being the value obtained the average of these runs), with 10 s *per* run). On the experiments with pep14-23, 100 µM of peptide was incubated with LDL at 50 µg/mL for 15 min, and then incubated for 15 min after the addition of 1, 3 and 5 µM of ZIKV C. Control experiments with DENV C and LDL (shown to have no interaction with this lipoprotein) were also conducted. The correlograms obtain from DLS were used to observe details of ZIKV C interaction with LDL, VLDL and in the presence of pep14-23. The normalized intensity autocorrelation function was analyzed using the CONTIN method. For each measurement, the values of D_H were extracted from the peak of $n(D_H)$ and then they were statistically analyzed. The data points of the lipoproteins size were analyzed statistically intra-group (of the 10 measurements) by average and standard deviation, rejecting outliers. The average without outliers was near to the value of the median in all the size data points. Presented values are mean ± standard error. Data sets were compared against the first set of measurements (lipoproteins only, without ZIKV C) using the Mann-Whitney U (or Wilcoxon) unilateral test. A size data set was considered significantly different from the set without ZIKV C if $p < 0{,}05$. Each experiment, for each concentration of ZIKV C was performed at least three times, with independent lipoproteins samples.

3 Results

In this thesis a combination of *in silico* and *in vitro* studies was used to gain insight into the capsid protein structure of mosquito borne *Flavivirus*, focused in ZIKV C protein and its expected interactions with host lipid systems.

First, a computer-aided analysis and a computational study were performed to compare (and group in clusters) the C protein of *Flavivirus* with the DENV C protein, to understand if/how they can relate in terms of structure, dimerization process, possible ways of interaction with lipid systems and inhibition by pep14-23. Secondly, it was necessary to express and purify the ZIKV C protein, as well as fully characterize it. Finally, experiments were performed to inquire the binding of ZIKV C protein to LDs and lipoproteins. The ability of pep14-23 to inhibit the ZIKV C-lipoproteins interaction was also investigated.

3.1 Similarities between homologous flavivirus C proteins

Ongoing work at the host laboratory allowed predicting the C protein structure, including alternative N-terminal conformations of 16 mosquito-borne flavivirus, identified as having homologous C protein sequences.[48] This was achieved by, briefly, submitting the C proteins sequence of each virus to an online server (I-Tasser, see Materials & Methodologies), which predicts tertiary protein structures. The resulting predicted structures were superimposed with DENV C dimer. Conformers with 6 or more clashes in the homodimer were excluded. (**Table *3.1***).

Table 3.1 – Summary of the study performed in terms of clashes and conformers for each flavivirus that allows dimer formation

Virus	Conformer 1		Conformer 2		Conformer 3		Conformer 4		Conformer 5	
	Clashes	Dimer	Clashes	Dimer	Clashes	Dimer	Clashes	Dimer	Clashes	Dimer
ALFV_C	8	✗	18	✗	0	✓	0	✓	8	✗
AROAV_C	0	✓	0	✓	10	✗	0	✓	10	✗
BAGV_C	5	✓	12	✗	0	✓	3	✓	5	✓
DENV_C	2	✓	0	✓	0	✓	6	✗	0	✓
IGUV_C	0	✓	0	✓	13	✗	0	✓	0	✓
ILHV_C	5	✓	4	✓	0	✓	4	✗	0	✓
JEV_C	9	✗	18	✗	0	✓	0	✓	0	✓
KEDV_C	12	✗	8	✗	0	✓	5	✓	0	✓
KOKV_C	0	✓	0	✓	5	✓	7	✗	0	✗
MVEV_C	3	✓	19	✗	0	✓	11	✗	0	✓
ROCV_C	1	✓	5	✗	4	✓	1	✓	0	✓
SLEV_C	0	✓	12	✗	0	✓	12	✗	0	✓
SPOV_C	0	✓	0	✓	10	✗	0	✓	0	✓
USUV_C	2	✓	13	✗	0	✓	5	✓	9	✗
WNV_C	2	✓	11	✗	0	✓	6	✗	2	✓
ZIKV_C	11	✗	9	✗	9	✗	0	✓	1	✓

Overall, as previously observed for DENV C, visual inspection shows that the predicted conformations for residues 33 to 100, that includes α2 and α4 regions (which are jointly the main drivers of dimerization), is essentially similar. Thus, dimers are likely the final observed structure. Contrarily, in the disordered N-terminal and the α1 regions there are alternative conformations, with previous work with DENV C showing 4 possible conformations for their N-terminal regions[114], classified according to their position: "open conformation", where the N-terminal region and the α1 region are exposed to the solvent (three conformers) and another, a "closed conformation", where the N-terminal region blocks the access of the solvent to the α2-α2' interface (one conformer). If the C protein is in an "open conformation" it is more likely to be able to interact with host lipid systems such as lipid droplets through the α2-α2' region, involved in that interaction.[48] Otherwise the adoption of a "closed conformation" might reflect in a autoinhibitory mechanism which prevents such interactions. Such alternative conformers may be representative of flavivirus C biological activity shedding light into, for example, the ZIKV C protein biological activity. To classify the predicted structures, an RMSD study was performed, comparing each flavivirus C protein conformer to each of the 4 viable predicted conformers of DENV C previously described (**Figure *3.1*** of next page).

This study allows determining which predictions are more similar to each dengue conformers, since by proximity (low RMSD) a classification of those predictions into a "open" or "closed" conformation can be made. More than a third of flavivirus C conformers N-terminal residues are within 10 Å of the N-terminal region of DENV conformer 1 (25.4%, 17 out of 48), with 6 less than 5 Å of distance. So, a large majority are in a "closed conformation", which is thus a likely structure (**Figure *3.1A***). Against one of the open conformation (DENV conformer 2) there is no clear distinction of the distribution of the conformers in groups or clusters, hence the majority (91.7 %, 44 out of 48) of the flavivirus conformations N-terminal residues are more than 10 Å away of that DENV conformer N-terminal, with only four conformers between 5 and 10 Å, and none less than 5 Å (**Figure *3.1B***). Such conformer is thus less likely to occur. Against DENV 3, another "open conformation", a similar distribution is observed: none within 5 Å, six within 5 and 10 Å and 89.4% (42 out of 48) more than 10 Å away (**Figure *3.1C***). Lastly, observing all virus conformations against DENV 5 (**Figure *3.1D***) (the last conformer in an "open conformation"), 20.8% are within 10 Å, while the remaining is further away.

From this analysis it is observed that it might be possible to group the predicted flavivirus into different clusters according to their position of the N-terminal. A distribution of the predicted conformations into clusters had already been performed in the host laboratory (Faustino, AF (2016) PhD Thesis – unpublished data) although this distribution was made by a visual analysis which might be faulty and biased (**Figure *3.2***, page 28). For this reason, an analytical method was needed.

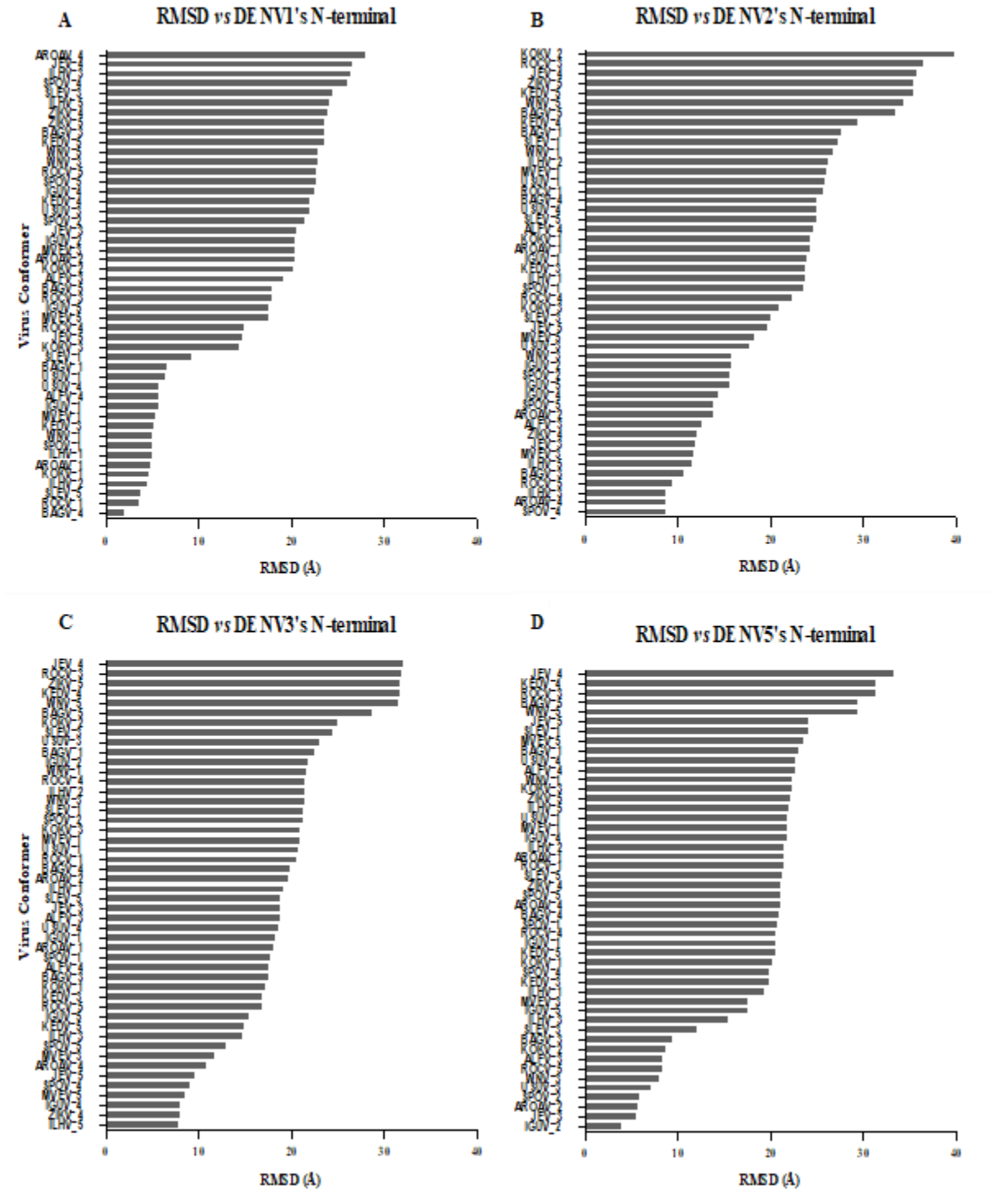


Figure 3.1 – RMSD of the N-terminal of all flavivirus against the N-terminal of DENV C predictions. (A) RMSD of the N-terminal of all flavivirus against the region 6-23 of DENV_1 conformer; **(B)** RMSD of the N-terminal of all flavivirus against the region 6-23 of DENV_2 conformer; **(C)** RMSD of the N-terminal of all flavivirus against the 6-23 region of DENV_3 conformer; **(D)** RMSD of the N-terminal of all flavivirus against the 6-23 region of DENV_5 conformer.

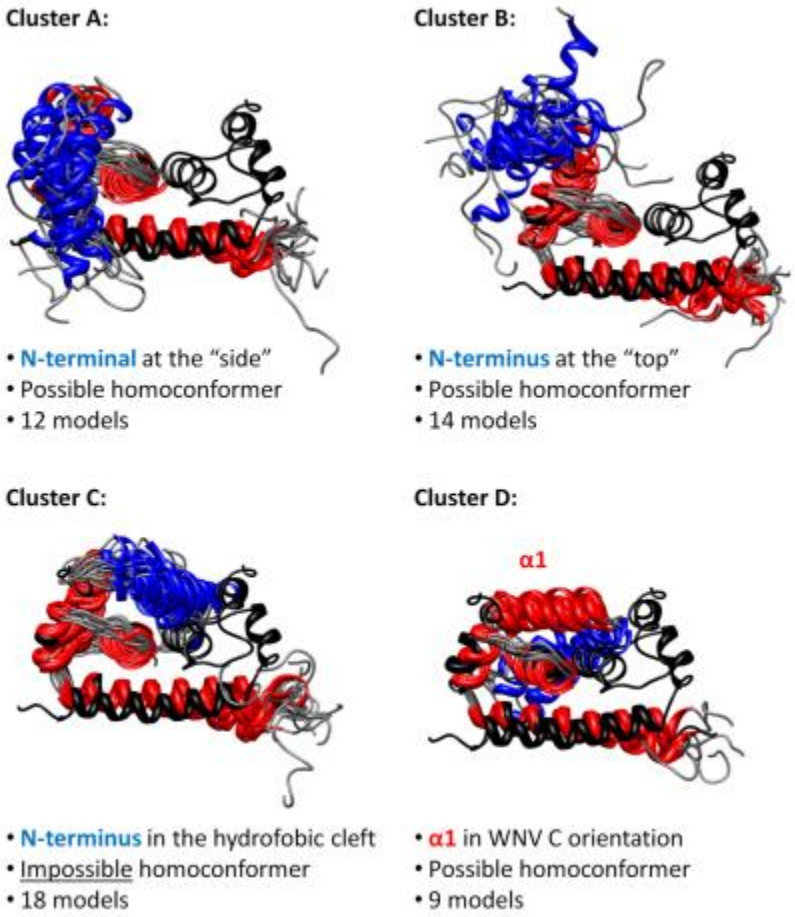


Figure 3.2 – Possible cluster formation based on the predicted conformations from I-TASSER and superimposing with the different conformations of DENV C predicted structures. Adapted from Faustino, André F. 2016 PhD Thesis.

This is a two-step method: first the residue which has more variability in the N-terminal region was determined for all predicted structures (by calculating the RMSD of all residues in the N-terminal, the residue that has more variability is the one that has higher RMSD); secondly, a EM-clustering algorithm with that residue's coordinates was performed for all predicted structures. This algorithm takes in consideration *a priori* the pre-defined numbers of clusters (in this case, the four clusters from the visual analysis), which the human eye can distinguish with ease without any faulty or biased assumptions.

The EM-algorithm assigns the data set to the *a priori* pre-defined clusters and computes the probability of each data point belonging to a corresponding cluster and not belonging to the remaining. It starts from a random input of the data set and, using the variances of each data point, it iterates and converges, trying to fit to the data's normal distribution. This was analyzed via the XLSTAT add-on of MS Excel, using the EM (Expectation Maximization) algorithm of the Cluster Distribution function. Briefly, the method uses the fitting of different Gaussian distributions, mixing them to discover the most likely distribution. Mixture modeling of Gaussian distributions of parameters is mainly due to the EM algorithm developments.[126] This is a probabilistic method for obtaining a fuzzy classification of observed parameters into clusters. The probability of the studied parameter(s) belonging to each of the proposed clusters is calculated and a classification is usually achieved by assigning each observation to the most likely cluster (in the case proposed, by comparing a model structure position with the distance to another, to find the closest and exclude the farthest). These probabilities can also be used to interpret suspected classifications. Mixture modeling is very flexible and suitable for this purpose. Overall, what the EM-algorithm achieves is structuring the dataset into several clusters, as described, with XLSTAT mixture of Gaussian distributions.

The models studied here being in three-dimension (3D) represented a huge difficulty to clearly reproduce the data obtained from the cluster analysis. To surpass this difficulty, it was necessary to transform the 3D data into a two-dimension (2D) visualization by projecting the spatial coordinates (x, y, and z) into x' and y', according to the Hammer Projection[127] (which is the 2D projection that less distorts the 3D reality) where:

$$x' = \frac{2\sqrt{2}\cos(\sin^{-1}(zy))\sin(\frac{2\tan^{-1}(\frac{zx}{2(2z^2-1)})}{2})}{\sqrt{1+\cos(\sin^{-1}(zy))\cos(\frac{2\tan^{-1}\left(\frac{zx}{2(2z^2-1)}\right)}{2})}} \quad (12)$$

$$y' = \frac{\sqrt{2}\sin(\sin^{-1}(zy))}{\sqrt{1+\cos(\sin^{-1}(zy))\cos(\frac{2\tan^{-1}(\frac{zx}{2(2z^2-1)})}{2})}} \quad (13)$$

Applying the formulas for the Hammer projection for the coordinates for all viruses most variable residue transformation, it was possible to reproduce the data obtained from the cluster analysis into a "planisphere"-like plot represented in the following figure (**Figure *3.3***).

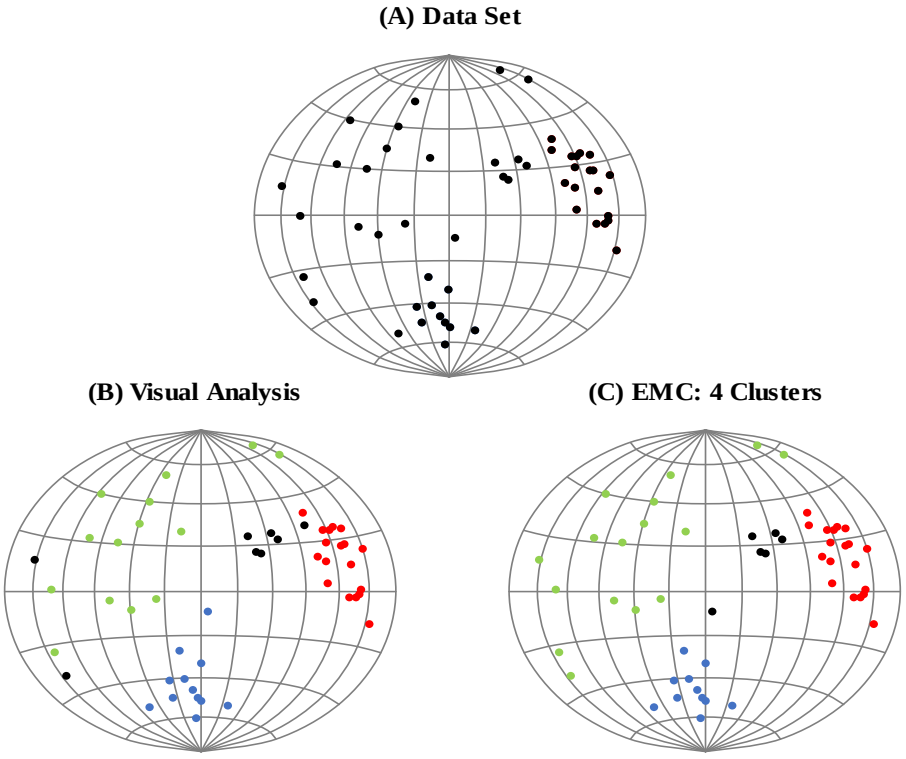


Figure 3.3 – Hammer plot projections for the original data set, and the data set distributed in clusters by the visual analysis previously reported and the new-method EM-clustering. (A) Original data set without a distribution into clusters; (B) Representation of the clusters defined by the visual analysis; (C) Representation of the new cluster definition by the EM-clustering algorithm. The colors represent the several clusters defined by the visual analysis where blue: Cluster A; green: Cluster B, red: Cluster C and black: Cluster D.

Comparing the EM-cluster results to the visual analysis it is possible to observe that three predicted conformers changed from Cluster D (black) to Cluster C (red) (one structure) and from Cluster D (black) to Cluster B (green) (two structures). It is also possible to observe that one structure changed from cluster A (blue) to cluster D (black). Visual inspections of the structures in the context of their new-cluster assignment confirmed it was in accordance.

This type of analysis is a major contribution for the understanding of the role of capsid protein in all flavivirus, since it can distinguish which conformers are more likely to be in a "open conformation" and able to interact and perform they biological propose from those that are more likely to be in a "closed conformation" which might be explained by a autoinhibitory mechanism.

In the context of this thesis, the major focus is the Zika virus. Therefore, in the next section, a computational approach was performed to determine the predicted full-length structure of the capsid protein for Zika virus.

3.1.1 ZIKV C protein predicted structure and dimer conformations

The next step of this thesis was to determine the structure of the ZIKV C protein as a dimer, which was not previously described. To do so, the allowed conformations of ZIKV C were superimposed with the crystalized structure of DENV C as described in Materials & Methods section. The resulting conformations are shown in (**Figure *3.4***) from a "front" and "top" perspective. The predicted conformations resulted in three possible dimer conformations: one fully open exposing the N-terminal (**Figure *3.4A***), one fully closed and compact (**Figure *3.4C***) and one partially closed (**Figure *3.4B***). The "fully open conformation" might allow interaction with lipid systems. The fully closed and the partially closed conformations might promote a autoinhibitory mechanism. A comparison with pep14-23 experimental structure shows that the peptide overlays with ZIKV C homologous N-terminal residues (**Figure *3.4* D**). Given that and since DENV C protein interacts with host lipid systems such as intracellular LDs and circulatory lipoproteins, namely VLDL, ZIKV C might also bind these host lipid systems and be inhibited by pep14-23. To later on test this hypothesis, we proceeded to express and purify ZIKV C protein, as described in the next section.

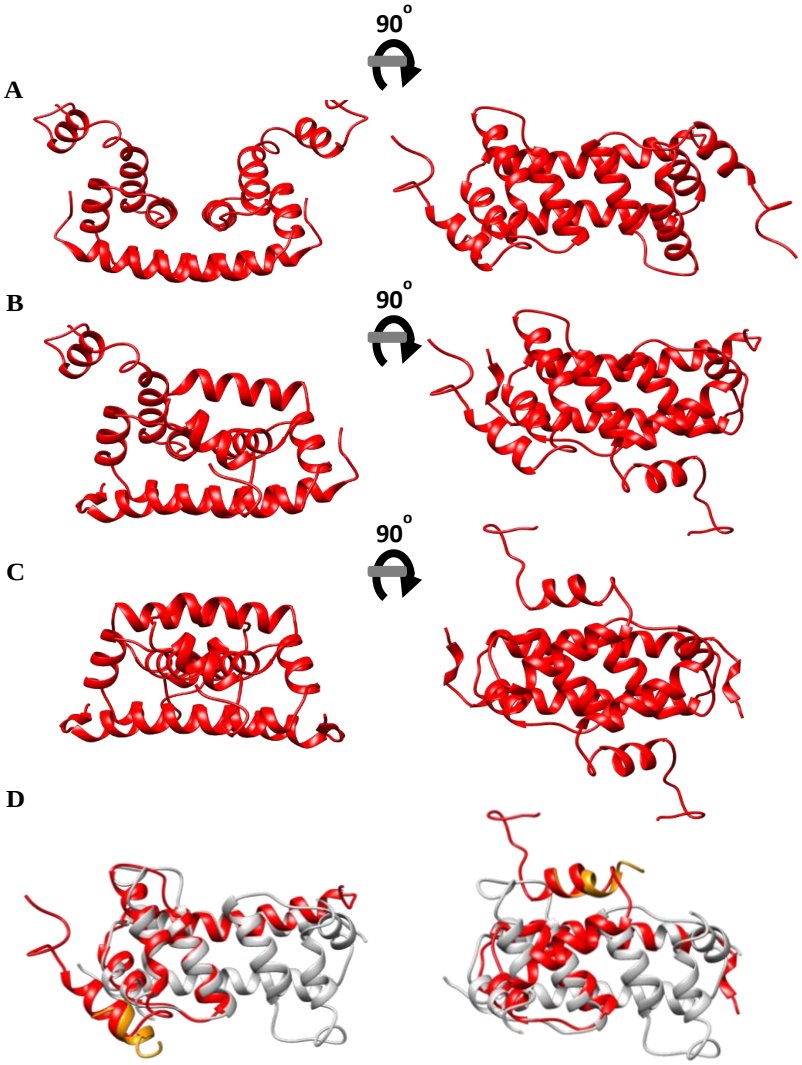
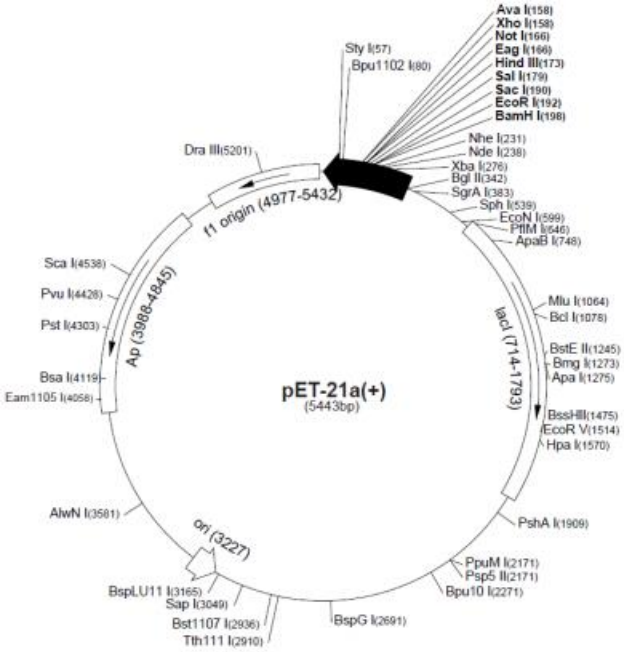


Figure 3.4 – **Predicted tertiary structure for ZIKV C showing 3 possible dimer conformations in a "front" and "top" perspective. (A)** Open conformation where the N-terminal is faced to the exterior; **(B)** Conformation where 1 of the subunits is in the interior of the dimer, possible an intermediary conformation of fully "open" conformer (A) and fully "closed" conformer (C); **(C)** Fully closed conformation where the both N-terminal region are in the interior of the dimer; **(D)** Conformer showing where the pep14-23 would interact and inhibit the interaction of the capsid with key targets. Color code: ZIKV C conformers in red, DENV C structure in gray and pep14-23 structure in yellow.

3.2 ZIKV C protein characterization

3.2.1 ZIKV C expression and purification

The ZIKV C protein used in this thesis to study the interaction with host lipid systems was expressed and purified in the host laboratory. As no methods have been published to express and purify the ZIKV C protein, a DENV C expression protocol already established at the host laboratory[50] was adapted and optimized for ZIKV C expression.[50] ZIKV C Brazil sequence (GenBank: KU497555.1) was selected due to the association of ZIKV Brazil strain with microcephaly[110] and pET21a was to maintain the expression plasmid of DENV C protein. The ZIKV C protein gene (encoding residues 1 to 104) was cloned into a plasmid pET21a by NZYTech, Lda (**Figure *3.5***).



Recombinant Protein Sequence

>Zika_capside

MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAME
IIKKFKKDLAAMLRIINARKEKKRR

Figure 3.5 – Illustrative figure of the final construct – the plasmid pET21a + ZIKV C.
The plasmid pET21a + ZIKV C is an *E. coli* expression vector that contains a sequence conferring resistance to ampicillin. The recombinant protein sequence was inserted in the plasmid by Nde I and Xho I restriction sites. This construct does not contain any N- or C-terminal His-tag.

Purification of pET21a+ZIKV C

To purify pET21a+ZIKV C we transformed a strain of competent cells (DH5-α, engineered *E. coli* cells to maximize transformation efficiency). 200 µL of competent DH5-α were thawed on ice and divided to 2 eppendorf, 100 µL to each. 1 µL of pET21a+ZIKV C from NZYTech was added to one of the eppendorfs. For the incorporation of the plasmid in the cells, we performed a heat shock transformation protocol: 30 min in ice, 90 s at 42 ºC, and 2 min on ice. 800 µl of LB medium were added to each tube following an incubation of 1 hour at 37°C, 220 rpm. To selectively grow isolated colonies

of transformed cells, we plumbed the cells suspension in LB agar plates with 100 µg/mLampicillin. Since the plasmid confers resistance to ampicillin, only the cells that were successfully transformed can grow in such medium. Controls were also made, cells transformed with plasmid were plumbed in LB agar plates without ampicillin and cells not transformed were plumbed in LB agar plates with and without ampicillin. As expected, DH5-α grew only in the absence of ampicillin and DH5-α transformed with pET21a+ZIKV C grew in the presence and absence of ampicillin. To isolate and purify the plasmid, one single colony of DH5-α transformed with pET21a+ZIKV C from the LB agar plate with ampicillin was inoculated in 10 mL of LB medium in the presence of 100 µg/mL ampicillin and incubated overnight at 37°C, 220 rpm. The cell culture was centrifuged at 4,618 x *g* for 20 min at 4°C. The plasmid was purified from the cells *pellet* with the Thermo Scientific Protocol "Protocol A: Plasmid DNA purification using centrifuges".

Screening of different *E. coli* strains

For a better expression and purification of ZIKV C protein it was necessary to test different strains of *E. coli* to select the most efficient strain. Three different *E. coli* strains were transformed, C41, C43 and BL21(DE3), using the pET21a+ZIKV C purified and the same protocol mentioned above for the transformation of DH5-α.Cells were grown overnight at 37°C and 220 rpm in LB medium with 100 µg/mL ampicillin. The overnight cultures were diluted (1:50 dilution) into a freshly LB medium in the presence of the same ampicillin concentration. Protein expression was induced at an optical density at 600 nm (OD_{600}) of approximately 0.9 with 0,5 mM of IPTG (Isopropyl β-D-1-thiogalactopyranoside) and the cell cultures were grown overnight (ON) at room temperature (approximately 26°C), 220 rpm. Aliquots of 20 µL were collected before induction and 2h after induction. Afterwards, the cultures were centrifuged at 4,618 x *g* for 30 min at 4 °C and 20 µL of the supernatants and *pellets* were collected. These aliquots were applied to an SDS-Page electrophoresis to inquire which strain is more efficient to express protein. From this SDS-Page (*data not shown*) we observed that the band corresponding to ZIKV C protein was intense for strain C4 (*pellet*); dim for strain C43 (*pellet*) and was not visible for strain BL21(DE3). From this it can be concluded that the strains C41 and C43 express the ZIKV C protein and the BL21(DE3) strain doesn't. Since the band is more intense for the C41 strain, this strain was selected for the expression at larger scale of ZIKV C protein. A representative schematic of the protocol implement in the host lab is thereby presented in **Figure *3.6***.

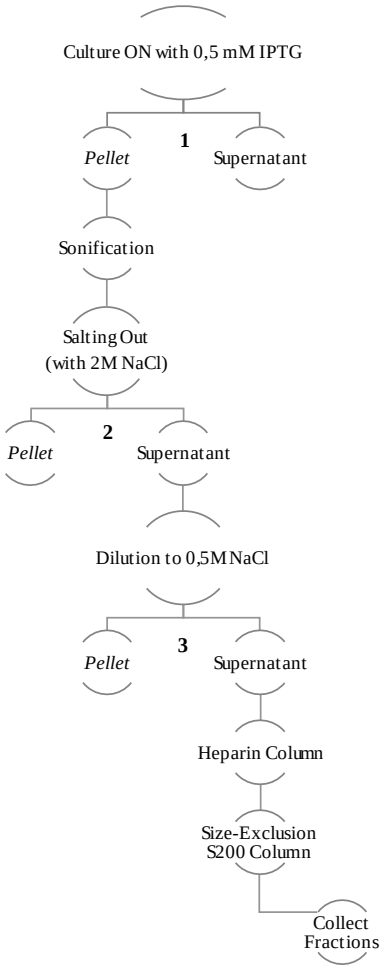


Figure 3.6 – **Schematics of the ZIKV C expression and purification protocol implemented in the host lab.**
Protocol utilized for purification of ZIKV C protein to further studies. Key steps: Sonication – 6 cycles 2 min each; **(1)** centrifugation at 4618 x *g* during 30 min at 4 ºC; **(2)** centrifugation at 16110 × g during 20 min at 4 ºC; **(3)** centrifugation at 16110 × g during 10 min at 4 ºC.

ZIKV C expression and purification at large scale

Proceeding to the expression and purification of ZIKV C at a larger scale, 500 mL of transformed C41 – pET21a + ZIKV C were grown to an $OD_{600\ nm}$ of approximately 0,9 and induced with 0,5 mM IPTG. Aliquots of 20 µL were collected after 1h, 2h, 4h and ON of induction. Then the culture was centrifuged at 4618 x *g* for 30 min at 4 ºC (first centrifugation) and aliquots of 20 µL of the resulting *pellet* and supernatant were collected. Afterwards the *pellets* were resuspended in a final volume of 50 mL TEE-NaCl 0,2M buffer and sonicated (6 cycles for 2 min each with an interval of 2 min). To increase the yield of ZIKV C protein, it was added NaCL to the sonicated *pellet* to a final concentration of 2 M of NaCl with agitation on ice for at least 1h (salting out the protein). Subsequently another centrifugation was made at 16110 × *g*, for 20 min at 4 ºC (second centrifugation) where aliquots of 20 µL were collected in the resulting *pellet* (which was discarded after) and supernatant. To confirm the success of the process, an intermediate quality control step was performed in the form of an SDS-Page to ensure if there was no loss of protein within the steps (**Figure 3.7**).

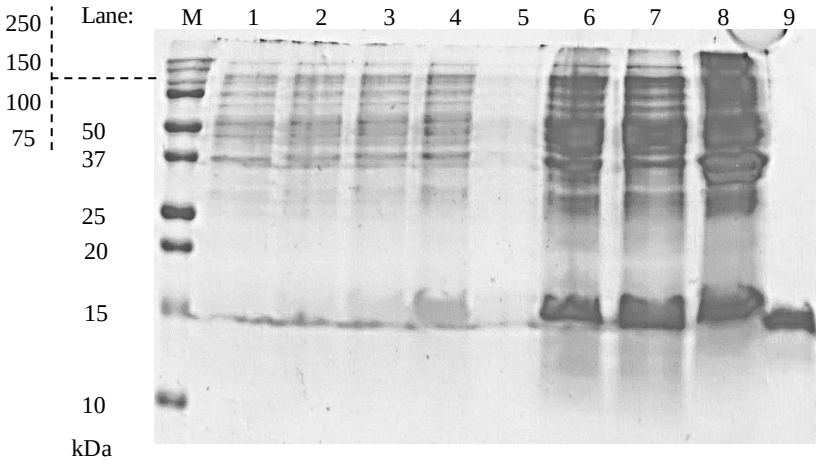


Figure 3.7 – Electrophoresis SDS-PAGE of the protocol for protein expression.
Result for a electrophoresis in denaturant conditions (SDS-PAGE) for several steps of the protein expression where: **(M)** Protein marker ranging from 10 to 250 kDa; **(1-4)** Lysed bacteria collected after **(1)** 1h, **(2)** 2h, **(3)** 4h and **(4)** ON incubation with 0,5 nM IPTG; **(5)** Supernatant resultant of the first centrifugation; **(6)** *Pellet* resultant of the first centrifugation; **(7)** Supernatant resultant of the second centrifugation; **(8)** *Pellet* resultant of the second centrifugation; **(9)** Recombinant DENV C protein.

A time-dependent band increase at 15 kDa can been seen from lanes 1 to 4 (1h to ON expression with IPTG): For this reason, ON was chosen as the optimal time for IPTG incubation. However, surprisingly, after sonication there is still a great amount of protein in the *pellet*, which was not expected since all soluble protein should be in the supernatant after sonication.

After the third centrifugation the supernatant was diluted to 4 fold of the volume acquired with NaCl 0,2M (50 mL of supernatant for 200 mL of final volume) and was used to purify the ZIKV C protein, resorting to an affinity chromatography (since the protein is positively charged, and heparin is negatively charged) (**Figure 3.8**).

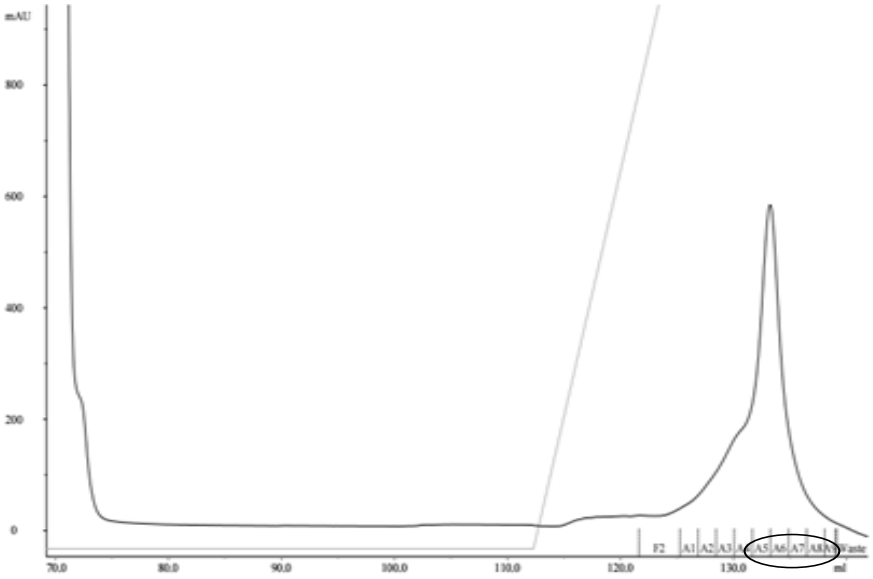


Figure 3.8 – Chromatogram obtained for protein purification with an affinity chromatography by heparin column.
To purify the ZIKV C protein an affinity chromatography with a heparin column was performed. Aliquots with 1,5 mL were collected correspondent to the fractions A5-A9 (black circle). The gray line is representative of a change in the buffer, from a buffer with low ionic strength (0,2M KCl) to a buffer with high ionic strength (2M KCl).

Fractions A5 to A9 were collected and another SDS-Page was performed (**Figure 3.9**).

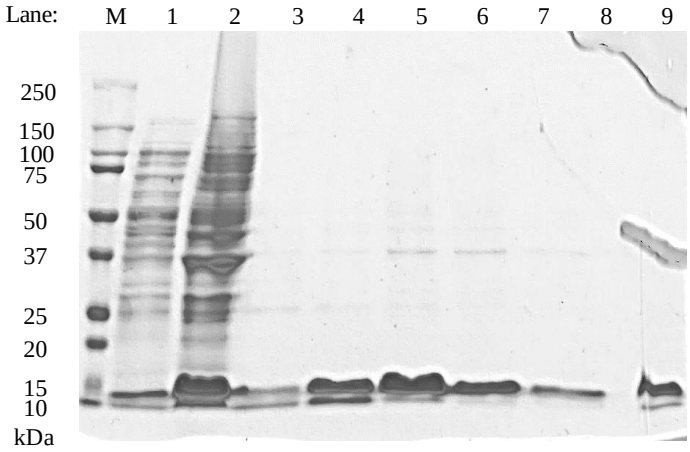


Figure 3.9 – Electrophoresis SDS-PAGE of the protocol for protein purification step – Heparin column
Result for a electrophoresis in denaturant conditions (SDS-PAGE) for the protein purification with heparin column where: **(M)** Protein marker ranging from 10 to 250 kDa; **(1)** Supernatant resultant of the third centrifugation, which entered in the heparin column ; **(2)** *Pellet* resultant of the third centrifugation of the protein expression protocol; **(3)** Eluted fraction A5 from heparin column; **(4)** Eluted fraction A6 from heparin column; **(5)** Eluted fraction A7 from heparin column; **(6)** Eluted fraction A8 from heparin column **(7)** Eluted fraction A9 from heparin column; **(8)** Empty lane; **(9)** Recombinant DENV C protein.

A large amount of protein in the *pellet* resultant of the third centrifugation (lane 2) is still observable, which again was not expected. However, after the elution, much "cleaner" lanes are observable, with all visible bands corresponding to the ZIKV C protein (lanes 3 to 7).

All fractions with protein (A5-A9) were then submitted to a size exclusion chromatography (since the protein has 11,6 kDa of size, a S200 column covers this range of size) to further purify the ZIKV C protein expressed (**Figure *3.10***).

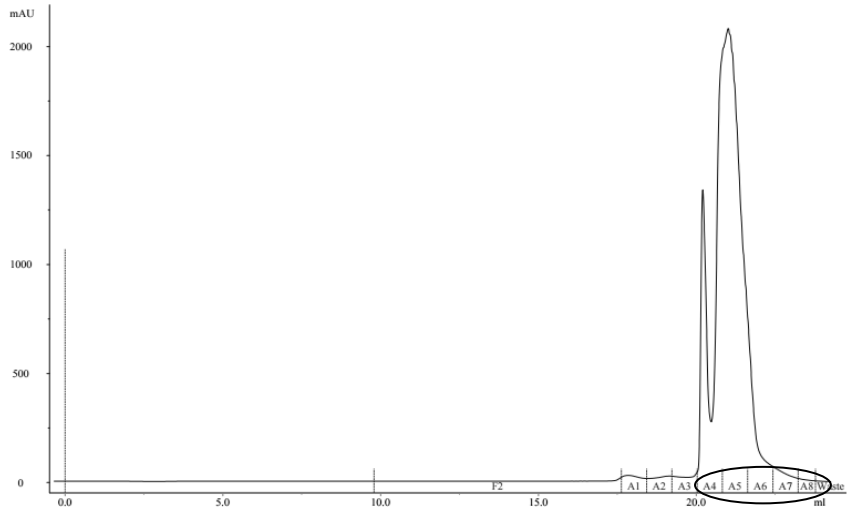


Figure 3.10 – Size-exclusion chromatography with a S200 column.
To further purify the ZIKV C protein a size-exclusion chromatography via S200 column was performed. The protein has a size of approximately 11,6 kDa. Aliquots with 1,5 mL were collected correspondent to the fractions A4-A8 (black circle). The gray line is representative of a change in the buffer, from a buffer with low ionic strength (0,2M KCl) to a buffer with high ionic strength (2M KCl).

Fractions A4-A8 were collected as well smaller aliquots for each fraction and a final SDS-Page was performed. (**Figure *3.11***).

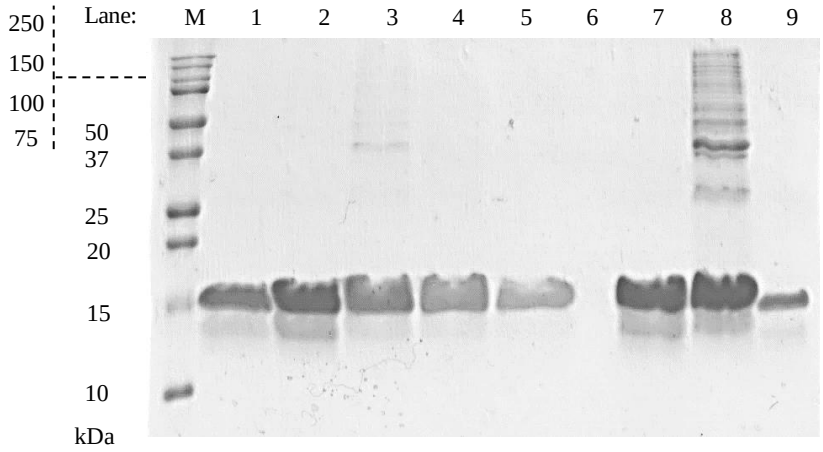


Figure 3.11 – Electrophoresis SDS-PAGE of the protocol for protein purification step – Size Exclusion S200 column
Result for a electrophoresis in denaturant conditions (SDS-PAGE) for the protein purification step with a size exclusion S200 column aliquots where: **(M)** Protein marker ranging from 10 to 250 kDa; **(1)** Eluted fraction A4 from the size exclusion S200 column; **(2)** Eluted fraction A5 from the size exclusion S200 column; **(3)** Eluted fraction A6 from the size exclusion S200 column; **(4)** Eluted fraction A7 from the size exclusion S200 column; **(5)** Eluted fraction A8 from the size exclusion S200 column; **(6)** Empty lane; **(7)** Fraction A5 supernatant **(8)** Fraction A5 *pellet* **(9)** Recombinant DENV C protein.

Protein aggregated immediately after elution from the size exclusion column at room temperature or at 4 ºC, and that's why for fraction A5 (more intense band hence more protein present) a centrifugation was performed to evaluate if there was ZIKV C protein in the supernatant and the resulting *pellet* (corresponding to lanes 7 and 8 in the **Figure *3.11***).

At the same time, it was necessary to prevent the aggregation observed. Several steps were attempted to optimize this part of the protocol:

- Reducing the volume of the initial C41 – pET21a + ZIKV C culture (e.g., to 100-200 mL, instead of 500 mL);
- Collecting less volume of fractions eluted from the heparin and size exclusion columns, (e.g., 0,8 mL instead of 1,5 mL);
- Varying conditions for the storing buffer (e.g., increasing the ionic strength of the buffer from 0,55 M KCl to 1,5M KCl).

However, upon decrease of the initial volume of the culture to 100 mL no relevant expression of ZIKV C protein was observed. In addition, even after increasing the ionic strength of the storing buffer from 0,55 M to 1,5 M of KCl, aggregation still occurred. This indicates that the forces promoting this aggregation are more complex and do not only involve electrostatic forces. Despite this, if the aggregates are centrifuged there is still ZIKV C protein in the supernatant which should be good for use for the further studies. To support this notion, as a final quality control, samples of the 0,55 M KCl fraction 5 and 1.5 M KCl fraction 5 supernatants were sent to Mass Spectrometry to evaluate if the protein had the expected mass and if there was any degradation of the N-terminal region. Results are shown in **Appendix I** and **Appendix II**.

Analyzing the obtained mass spectrum for the resulting supernatant and *pellet* for fraction A5 in 0,55 M KCl (**Appendix I**) we observe that the expressed ZIKV C protein has a mass of approximately 11647 Da, which is concordant to the predicted in the computational approach. There is no sign of

degradation of the N-terminal in both samples, therefore this protein can be used for further studies. Analyzing the **Appendix II**, which is the mass spectrum for the resulting supernatant and *pellet* for fraction A5 in 1.5 M KCl is observable again that the base peak corresponds to the predicted mass for the ZIKV C protein. But it is also seen some degree of degradation of the N-terminal region, which results into the lesser peaks left from the base peak (**Appendix IIA**). This indicates that this protein is not good to use in the further studies since the region of interest (N-terminal) is degraded.

To further analyze the expressed and purified ZIKV C protein in a quality control manner, a CD spectrum of ZIKV C protein 15 μM at 25 ºC was acquired. The spectrum obtained (***Figure 3.12***) shows two minimal peaks at 208 nm and 222 nm, characteristic of an α-helix secondary structure confirming the predicted structure in the computational section.

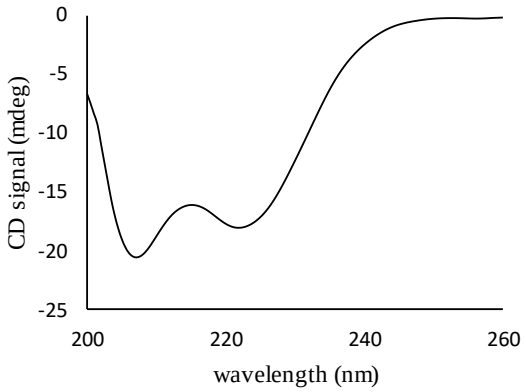


Figure 3.12 – CD spectrum obtained for ZIKV C protein at 25 ºC.
Typical spectra for a α-helix secondary structure obtained for ZIKV C protein at 25 ºC. This data was used in the online deconvolution servers K2D2/K2D3 were the results obtained is shown in ***Table 3.2***.

This experimental data was then deconvoluted with the assistance of online servers K2D2/K2D3. The K2D2 online server takes as input the CD data obtained experimentally and returns a estimated value of percentage (%) of secondary structure most present in the protein. The K2D3 online server is an updated version of K2D2, thus being more refined since it takes in consideration the protein size (either using the length in amino acids or the weight in kDa). The values obtained are shown below in **Table *3.2***.

Table 3.2 – Estimated secondary structure through online deconvolution servers K2D2 and K2D3.

Method	Percentage of α-helix (%)	Percentage of β-sheet (%)
K2D2	84	1,2
K2D3	94	0,1

Analyzing the percentage obtained from each server, the K2D2 server predicted that the protein was 84% α-helix and 1,2% β-sheet, being the remaining 14,8% associated to a disordered region/random coil. In the meanwhile, the K2D3 server predicted that the ZIKV C protein was 94% α-helix and 0,1% β-sheet, being the remaining 5.9% associated to a disordered region/random coil.

Looking for the insights that are accessible from the literature, it is expected that the ZIKV C protein is constituted mostly by α-helix structures and that it possesses a disordered region (random coil)

in the N-terminal region, that acquires structure upon proximity with lipid systems. Thus, the online server whose results better represent the reality and fits the current insights that is the K2D3.

3.2.2 Protein stability

To study the stability of the purified ZIKV C protein, a temperature denaturation study was performed using circular dichroism spectroscopy.

The spectrum of ZIKV C at 15 µM was acquired with increasing temperature from 6 ºC to 96 ºC and a denaturation profile using the mean residues molar ellipticity signal at 222 nm was made (**Figure *3.13***). With the increase of temperature, a loss of structure is observed due to the increase of the minimal peaks characteristic of α-helix, although the protein does not completely unfold since a distinctive spectrum for disordered/random coil structure is not seen (**Figure *3.13A***). This is confirmed by the denaturation profile (**Figure *3.13B***), since a plateau is not reached for high temperatures.

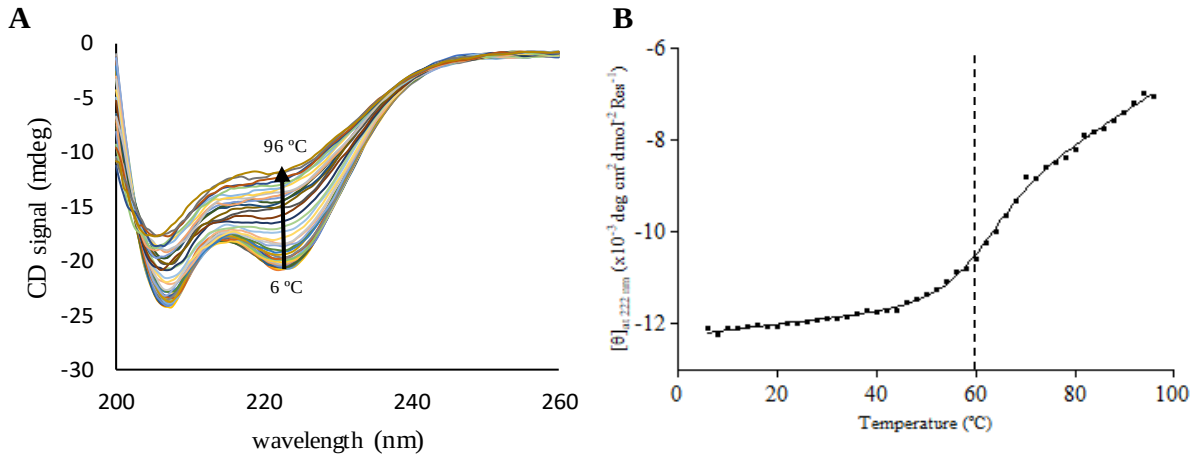


Figure 3.13 – Temperature denaturation study followed by CD for ZIKV C protein.
(A) Temperature denaturation study was performed followed by CD spectroscopy, where ZIKV C spectra were obtained with increasing temperature from 6 ºC to 96ºC. **(B)** Denaturation profiles of the mean residues molar ellipticity at 222 nm as a function of temperature. The line between dots represents the melting temperature, T_m, 59.9 ± 4.7 ºC.

From this, is possible to calculate an approximate value for temperature melting point (T_m), which is the temperature where half of the protein is completely folded, and the other half is completely unfolded. Since a state of complete unfolding is not reached, the value calculated is approximation, being this way, a temperature melting point apparent (T_m^{app}). The value obtained was about 59,9 ± 4,7 ºC, using the simplest assumption (a two-state mechanism, although no plateau was observed).

To inquire if the process of unfolding is reversible, after the ZIKV C protein reached the 96 ºC the system was set back to 25 ºC and let to equilibrate for 15 min. Afterwards a spectrum was acquired and superimposed with the spectrum acquired before the denaturation (**Figure *3.14***). From this it is observable that the protein refolds almost completely which indicates its high stability. In the context of the predicted structure of ZIKV C protein, the data can be also interpreted as a denaturation of the N-terminal region and the α1 region of the dimer. Since the protein does not totally unfold, presumably the main bulk of the dimer (α2-α4 region) remains intact and largely stable and folded. Upon decrease of temperature to 25 ºC, the N-terminal and α1 regions returns to its original fold.

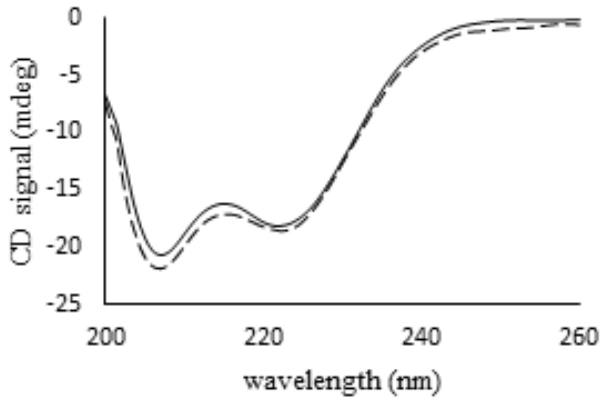


Figure 3.14 – CD spectrum for ZIKV C protein at 25 ºC and renatured.
CD spectrum for ZIKV C protein at 25 ºC (black line) and CD spectrum obtained after denaturation (dashed line). This indicates that the protein refolds almost completely demonstrating its high stability.

3.3 ZIKV C protein interaction with host lipid systems

After establishing that ZIKV C recombinant protein expressed and purified was functional, stable and with an α-helical conformation, the binding of ZIKV C to host lipid systems, such as lipid droplets and lipoproteins, was investigated. In previous work, the binding of DENV C to host lipid droplets was studied[48,50], with a peptide based on the N-terminal region conserved motif (residues 14th to 23rd) of 16 mosquito-borne flavivirus was designed, characterized and patented as pep14-23 (WO Patent Nr 2012159187), an inhibitor of DENV C-LDs binding. [48,114]

DENV C interaction with other host lipid systems, namely blood-circulating lipoproteins was also accessed in the host lab since such interactions could prompt the formation of highly infective lipoviroparticles.[67,70] DENV C protein binds to VLDL, but not to LDL. Moreover, there is a remarkable similarity between VLDL and LDs binding to DENV C, with a striking structural resemblance between proteins present at the surface of both lipid systems. Similar proteins of each lipid system and potassium ions were shown to be essential for both interactions. In addition, pep14-23 inhibits not only DENV C binding to LDs but also to VLDL. Given that and ZIKV C high sequence homology with DENV C, it was hypothesized that ZIKV C may also be able to bind to the same host lipid systems. That hypothesis was tested with the results obtained described ahead.

3.3.1 ZIKV C interacts with LDs

Following previous approaches[48,50], ζ-potential measurements were performed to understand if ZIKV C protein interacts with LDs. In the absence of ZIKV C, LDs in K^+ buffer have a negative ζ-potential value of -19,7 ± 2,7 mV, as expected.[48] Adding ZIKV C to LDs induces a progressive increase in the scattering particle charge (ζ-potential values increase), stabilizing at positive values of 10,7 ± 1,5 mV (**Figure 3.15*A and B*** where in ***B*** the same data are displayed in logarithmic scale, to facilitate analysis, page 43) for 0,5 μM ZIKV C or higher. A correlation between the increase of ζ-potential values and the concentration of ZIKV C added to LDs is observed. This concentration dependent increase in Δζ-potential, indicates that ZIKV C binds LDs as expected.

To inquire whether the interaction requires proteins attached to the surface of LDs (described as essential in DENV C studies), a similar approach was performed with trypsinized LDs, showing an initial value of -10,3 ± 2,6 mV that increases and stabilizes at 2,5 μM ZIKV C, reaching a plateau around 6,8 ± 0,8 mV (**Figure 3.15*C and D*** where in ***D*** the same data are displayed in logarithmic scale, to facilitate analysis). The increase in Δζ-potential is much higher in the non-trypsinized LDs than the increase observed for trypsinized LDs. This indicates that this interaction is stronger in non-trypsinized LDs (where the surface proteins are preserved) than in trypsinized LDs (where the protein cloth was removed by the action of trypsin).

This data indicates that the LDs surface proteins are required for the interaction between ZIKV C and lipid droplets. The experimental data was fitted to two different equations (eq. 3 and eq. 4 of Material & Methodologies section) with good R^2 values for both fits. The maximum amplitude of variation of ζ-potential promoted by the interaction of ZIKV C ($\Delta\zeta_{max}$) and the ZIKV C protein concentration at which half the lipid droplets binding sites are occupied ($C_{1/2}$) for both fits were determined, shown in **Table *3.3***.

For the non-cooperative fit, defined by equation (3), non-trypsinized $C_{1/2}$ was 178 ± 35 nM and for trypsinized LDs, $C_{1/2}$ was 1050 ± 336 nM (5.9-fold higher), indicating that LDs surface proteins are essential to the interaction (**Figure 3.15*E***).

In the cooperative fit defined by equation (4) (**Figure 3.15*F***), for non-trypsinized LDs the value obtained of $C_{1/2}$ was 38,5 ± 26,4 nM and for trypsinized LDs, the value obtained of $C_{1/2}$ was 455 ± 315

nM (11.8-fold higher). The value of n (cooperativity), when higher than 1, indicates cooperative binding of ZIKV C proteins to LDs. The values were 1,81 and 1,64 for non-trypsinized and trypsinized LDs, respectively. In the presence of LDs surface proteins, binding is not only stronger but also the binding of one ZIKV C dimer molecule seems to promote further binding, suggesting that it is specific, not solely electrostatic driven and involving LDs surface proteins.

Table 3.3 – Parameters obtained by ζ-potential analysis of LDs under different conditions. The values of $C_{1/2}$ and $\Delta\zeta_{max}$ were obtained through the fitting of ζ-potential experimental data with equation 2 and 3.

Condition (Equation fitted)	**Mean $C_{1/2}$ ± S.E. (nM)**	**Mean $\Delta\zeta_{max}$ ± S.E. (mV)**	**n (cooperativity)**	**R^2**
Non-Trypsinized (3)	178 ± 35	33.4 ± 1.8	-	0.91
Trypsinized (3)	1050 ± 336	21.8 ± 2.3	-	0.90
Non-Trypsinized (4)	38.5 ± 26.4	31.3 ± 1.4	1.81	0.93
Trypsinized (4)	455 ± 315	18.0 ± 1.8	1.64	0.91

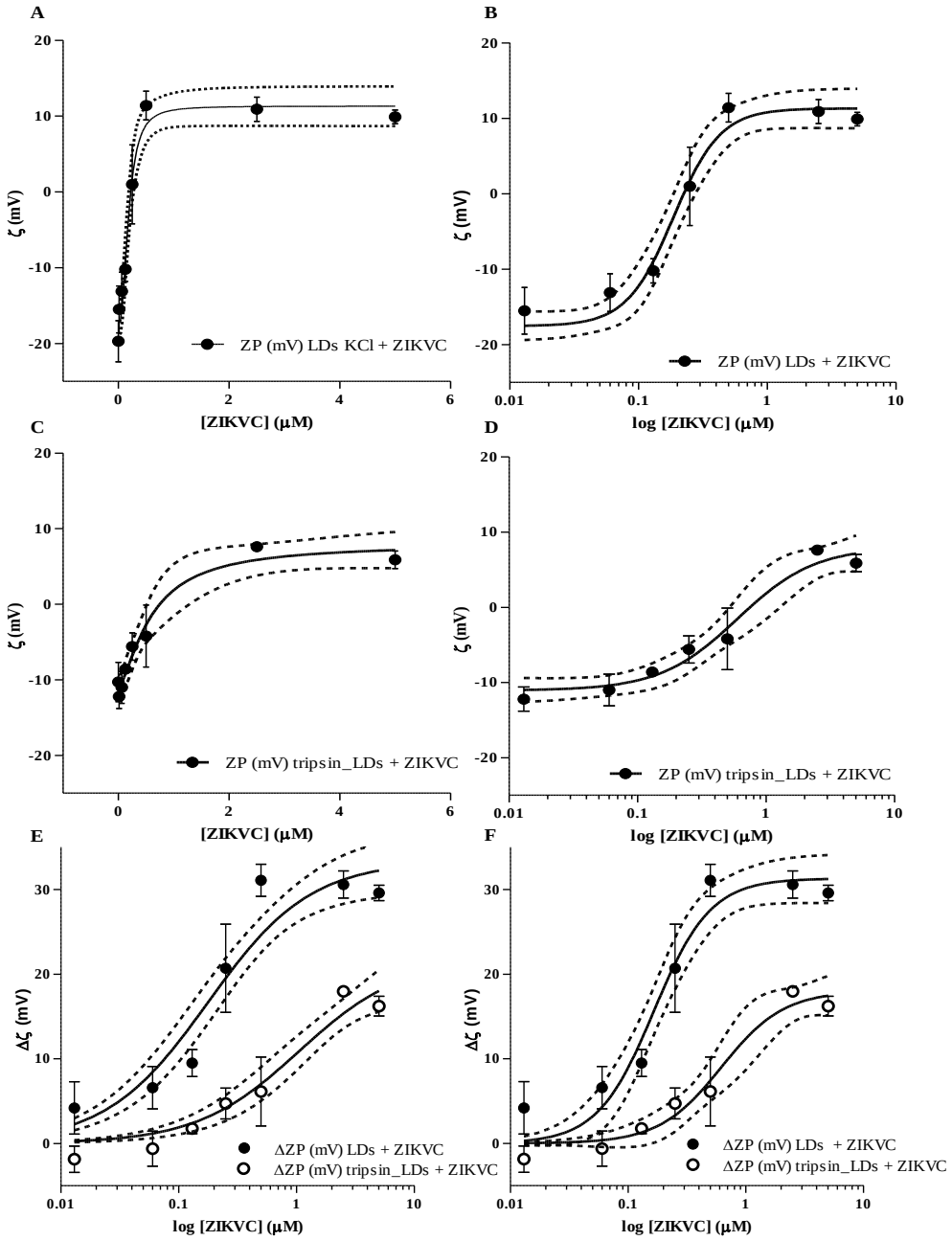


Figure 3.15 – Lipid droplets ζ-potential determination and analysis with different ZIKV C concentrations.
ζ-potential for the isolated LDs was accessed in the absence and in the presence of ZIKV C protein, ranging from 0,013 μM to 5 μM. The values were plotted in a linear scale **(A)** and in a logarithmic scale **(B)**, where the black dots represent the obtained values, the black line represents the fit used and the dashed lines represent the 95% confidence interval of the fit used. Results are represented as mean ± standard error of 3 independent repeats.
To inquire if the surface proteins have some role in the interaction of ZIKV C and LDs, LDs were trypsinized and ζ-potential was measure in the absence and in the presence of ZIKV C protein, ranging from 0,013 μM to 5 μM. Values were plotted in a linear scale (**C**) and in logarithmic scale (**D**) where the black circles represent the obtained values, the black line represents the fit used and the dashed lines represent the 95 % confidence interval of the fit used. Results are represented as mean ± standard error of 3 independent repeats.
To evaluate the effect of cooperativity two fits were made to the experimental data of LDs/LDs trypsinized. Values are expressed in variation of zeta potential (Δζ) and were accessed without trypsin preincubation (filled circles) and with LDs preincubated with trypsin (empty circles). Black line was obtained through the fitting of experimental data with equation 3 and dashed lines represent the 95% confidence interval **(E)**. Another fit was made taking in consideration the effect of cooperativity where the black line is the fit of the experimental data with equation 4 and dashed lines represent the 95% confidence interval **(F)**. Results are represented as mean ± standard error and both graphics are in logarithmic scale.

3.3.2 ZIKV C interaction with LDL

To test ZIKV C interaction with LDL in a similar manner as DENV C protein (since they are homologous proteins and may have a homologous function as mentioned in section 3.3), DLS was employed to observe whether there would be an increase of the lipoproteins' size. Such increase would be reflected in a shift in the correlograms, demonstrating an interaction and binding.

First, the interaction with LDL was determined with increasing concentrations of ZIKV C up to 5 μM (above which strong flocculation and aggregation was observed which made the measurement noisy). As a control, the interaction with DENV C, which in previous reports was shown not to interact with LDL, was also evaluated. The area under the curve of the correlograms was also calculated to quantify the shift observed in each experiment (**Figure 3.16**).

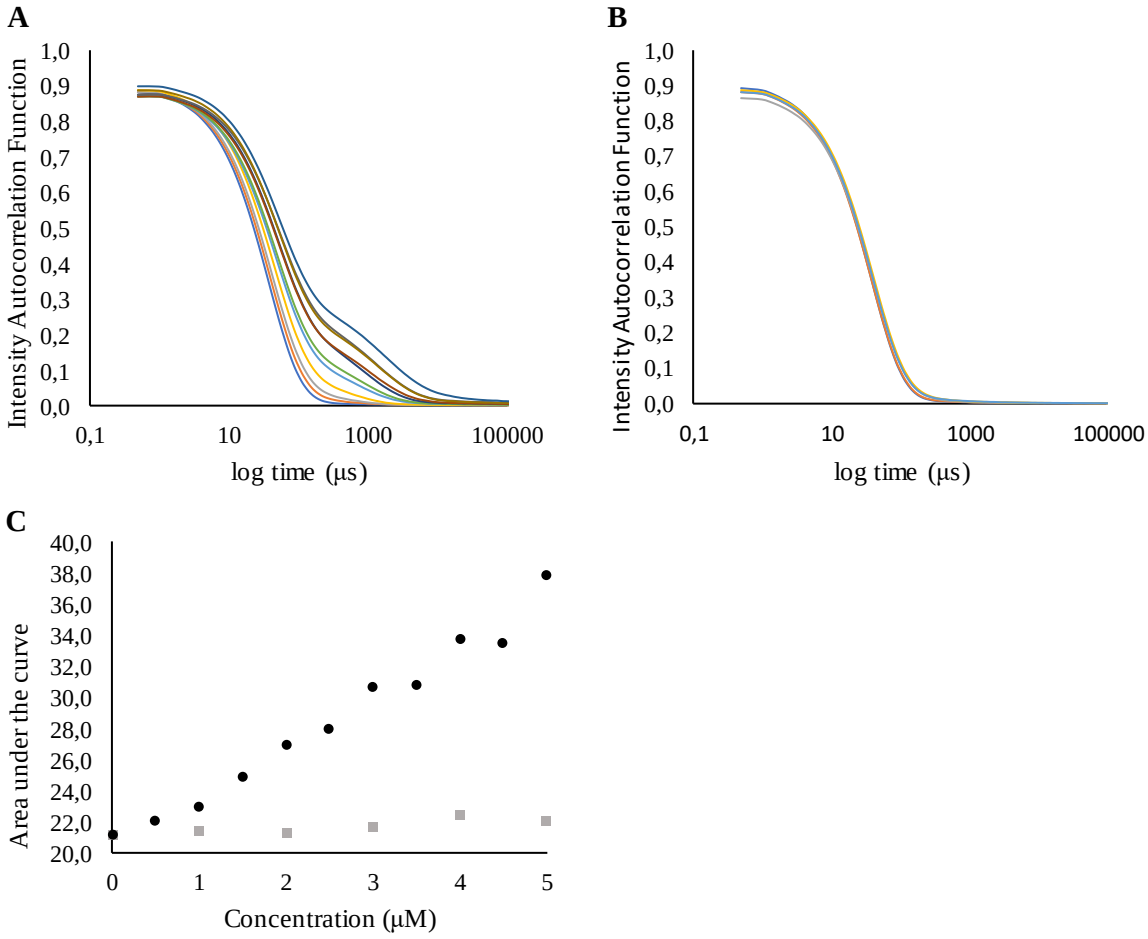


Figure 3.16 – Correlograms obtained through DLS for LDL with increasing concentrations of C proteins.
(A) Assessment of the interaction between LDL and ZIKV C protein through dynamic light scattering, resulting in a shift observed in the correlograms. ZIKV C concentration is ranging from 0,5 μM to 5 μM in 0,5 μM of interval. **(B)** Assessment of DENV C and LDL interaction which showed no shift in the correlogram superimposing with the correlogram with LDL in the absence of DENV C. DENV C protein concentration ranged from 0 μM to 5 μM in 1 μM of interval. **(C)** Area under the curve of the correlograms for ZIKV C (black dots) and DENV C (gray squares), which demonstrates an increase of the shift of ZIKV C comparing with DENV C.

Analyzing the correlograms we can observe there is a shift upon titration of LDL with increasing ZIKV C concentrations (***Figure 3.16 A***) indicating an interaction. In contrast, with DENV C protein, a superimposition is observed of all correlograms of increasing DENV C concentration with the correlogram obtained with LDL in the absence of DENV C (***Figure 3.16 B***). This was expected as this behavior was previously reported.[67] Calculating the area under the curve of each correlogram (**Figure**

3.16 *C*) these distinct interactions between ZIKV C or DENV C with LDL can be better observed showing that ZIKV C may bind to LDL, contrarily to DENV C.

Measuring the scattered light intensity fluctuations that occurs due to the Brownian motion of lipoproteins through DLS, is possible to calculate the hydrodynamic diameter (D_H) of particles in solution, through the Stokes-Einstein equation. Applying this technique, the average particle size of LDL was measured in the absence and in the presence of ZIKV C protein, demonstrating an increase of the LDL size upon titration with ZIKV C (**Figure 3.17**).

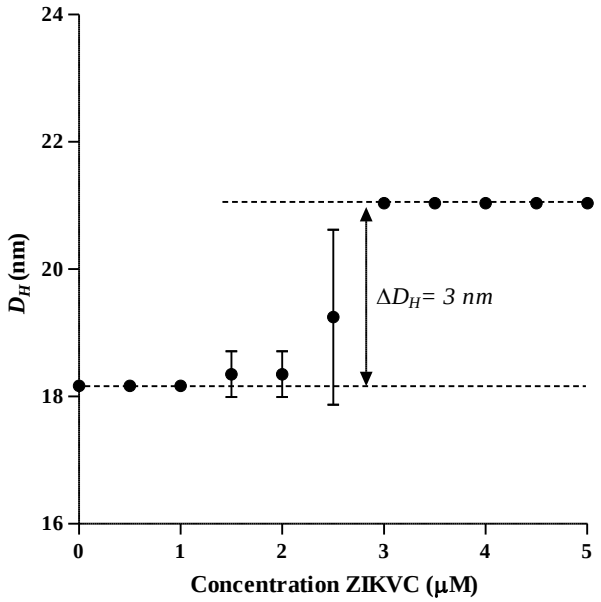


Figure 3.17 – Dynamic light scattering (DLS) analysis of lipoproteins in potassium buffer. Lipoproteins hydrodynamic diameters (D_H) were determined in the absence and in the presence of increasing ZIKV C concentration, ranging from 0,5 μM to 5 μM. The maximum LDL size increase due to ZIKV C reaches about 3 nm in diameter. Each result is represented by mean ± standard error.

The experimental data shows that the LDL average size is 18,17 nm D_H, which is in accordance with the literature. Upon ZIKV C addition, there was a statistically significant increase in the average LDL size, up to 21,04 nm D_H ($p < 0,005$), at the higher ZIKV C protein concentrations (>2,5 μM). The hydrodynamic diameter (D_H) of the conjugate LDL – ZIKV C particle is thus 2,87-nm larger than LDL particles. Although some aggregations may there occur at the higher concentrations of ZIKV C, observed by the distribution of the measurements of $I(D_H)$, and that may slightly affect the size estimations, this only happens in a small fraction of the total particles and the estimations are *per* particle, it is thus feasible to correlate the increase of the average LDL size to its interaction with ZIKV C.

Following previous approaches[48,50], pep14-23 inhibition of LDL binding to ZIKV C was tested. LDL were pre-incubated with 100 μM of pep14-23 for 15 min and then 1 μM, 3 μM and 5 μM of ZIKV C protein were added to the system. DLS measurements were acquired as described above.

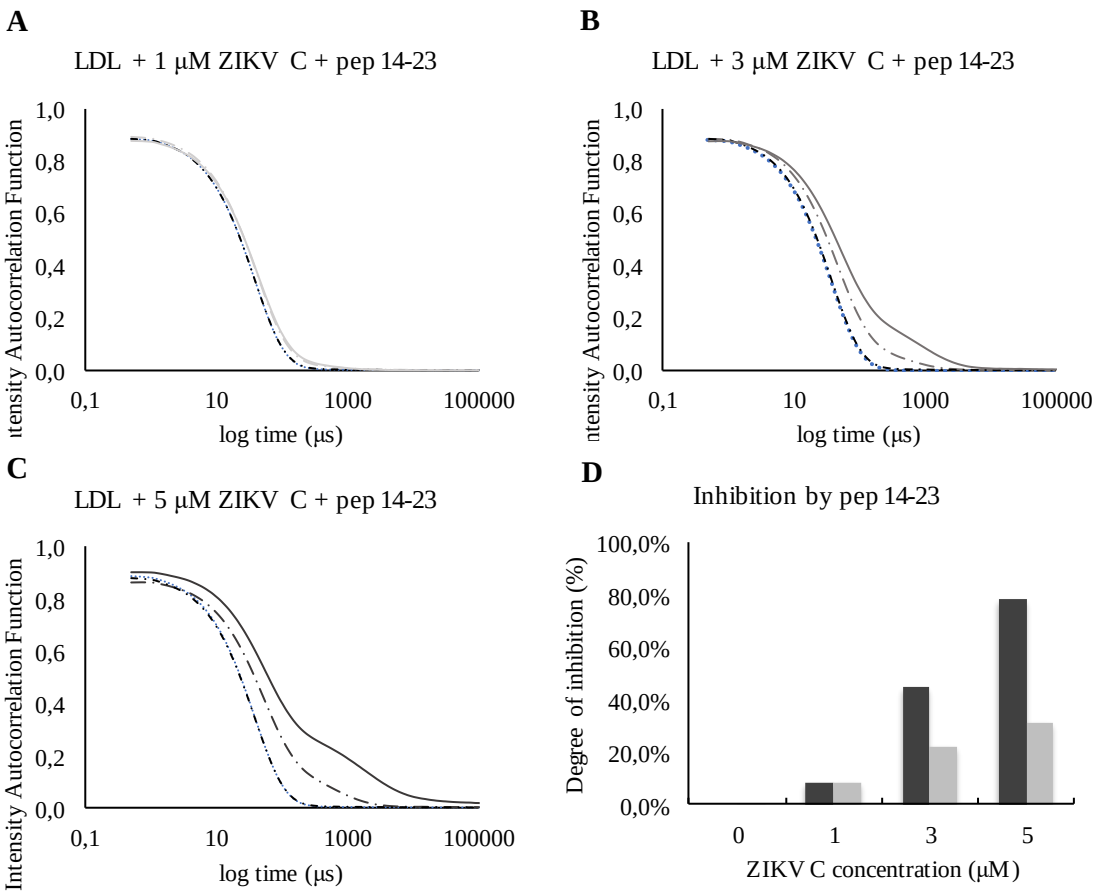


Figure 3.18 – Correlograms obtained through DLS for the inhibition of the interaction LDL – ZIKV C with pep14-23. (A) Correlogram obtained for 1 µM of ZIKV C and 100 µM pep14-23, where dark dashed line represents LDL, blue dots represents LDL + pep14-23 and the dashed and the solid gray line represents the system LDL + 1 µM ZIKV C + pep14-23 and LDL + 1 µM ZIKV C respectively; **(B)** Correlogram obtained for 3 µM of ZIKV C and 100 µM where dark dashed line represents LDL, blue dots represents LDL + pep14-23 and the dashed and the solid gray line represents the system LDL + 3 µM ZIKV C + pep14-23 and LDL + 3 µM ZIKV C respectively; **(C)** Correlogram obtained for 5 µM of ZIKV C and 100 µM where dark dashed line represents LDL, blue dots represents LDL + pep14-23 and the dashed and the solid gray line represents the system LDL + 5 µM ZIKV C + pep14-23 and LDL + 5 µM ZIKV C respectively. **(D)** Correlogram area increase (compared to LDL without ZIKV C) observed for LDL in the presence of 0 µM, 1 µM, 3 µM and 5 µM ZIKV C (black columns) and for the same systems but pre-incubated with 100 µM pep14-23 (grey columns).

Analyzing the correlograms obtained from ***Figure 3.18***, LDL with and without 100 µM pep14-23, in the absence of ZIKV C protein, display a similar correlogram. For 1 µM ZIKV C protein there is an overlay in the correlograms in the absence and in the presence of pep14-23 (***Figure 3.18 A***).

However, if we analyze the correlograms for higher concentrations of ZIKV C (3 µM ***Figure 3.18 B*** and 5 µM ***Figure 3.18 C***), when the LDL were pre-incubated with pep14-23 there was a smaller shift in the correlogram indicating inhibition. Calculating the area under the curve of those correlograms we can determine the degree of inhibition by pep14-23 (***Figure 3.18D*** and **Appendix *III***)

3.3.3 ZIKV C interaction with VLDL

The interaction between VLDL and ZIKV C was studied via DLS. Preliminary tests with this VLDL stock showed that there was some disparity among sizes, which could be observed by the intensity peak. Fresh VLDL should have a single peak around 40 nm to 50 nm of diameter (**Figure *3.21***).

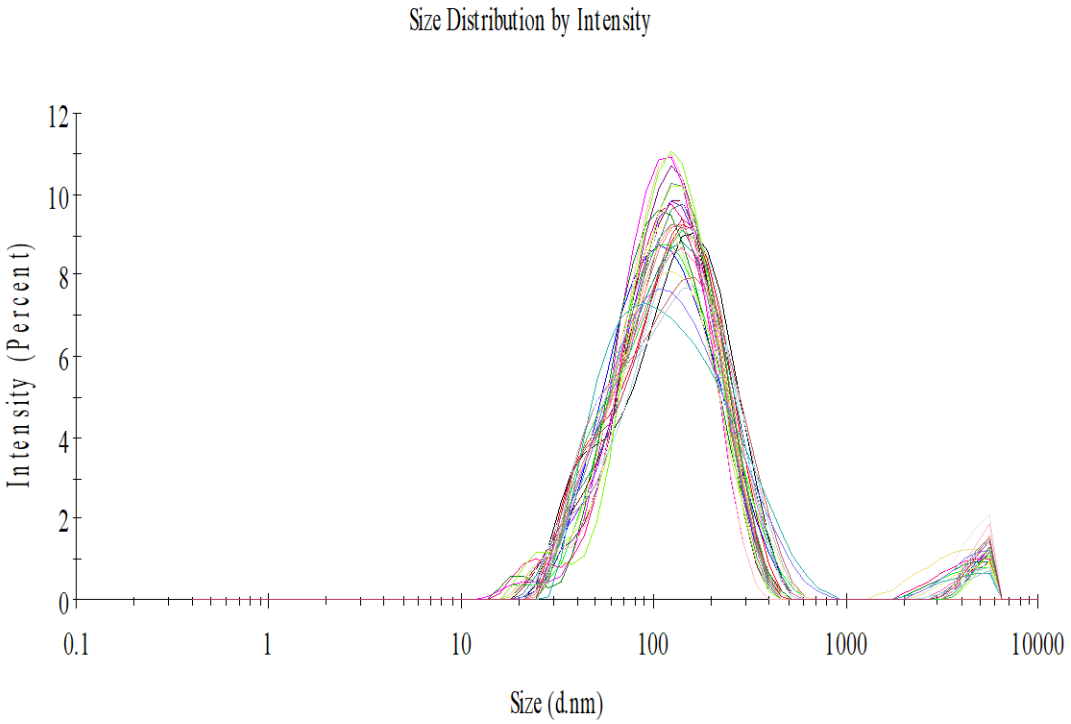


Figure 3.19 – Size distribution by intensity measured by DLS showing a disform variability of VLDL lipoproteins' size (batch 1). Variety of peaks in the intensity of VLDL, showing disparity in size of particles. Besides, the peaks are above 100 nm, which in normal situations should be around 40-50 nm, maybe some signs of coalescence of VLDL are present. Also, aggregation is observable (see signal above 1 µm).

Analyzing the intensity peak measured with the VLDL available in the lab, it is observable that there is a wide variety of sizes of VLDL (peak around 100 nm and some fluctuations in the intensity, showing some high particles suggesting aggregation) and that somehow, they coalesced. This coalescence can be explained by the durability of the vial bought as it was near the expiration date. Preliminary tests with ZIKV C protein and the above mention VLDL (which from now on will be called "VLDL1") were made, via DLS. Upon titration of this lipoproteins with ZIKV C and DENV C there was a shift in the correlograms, indicating a possible interaction between VLDL lipoproteins and ZIKV C (VLDL and DENV C were already studied and reported to interact).

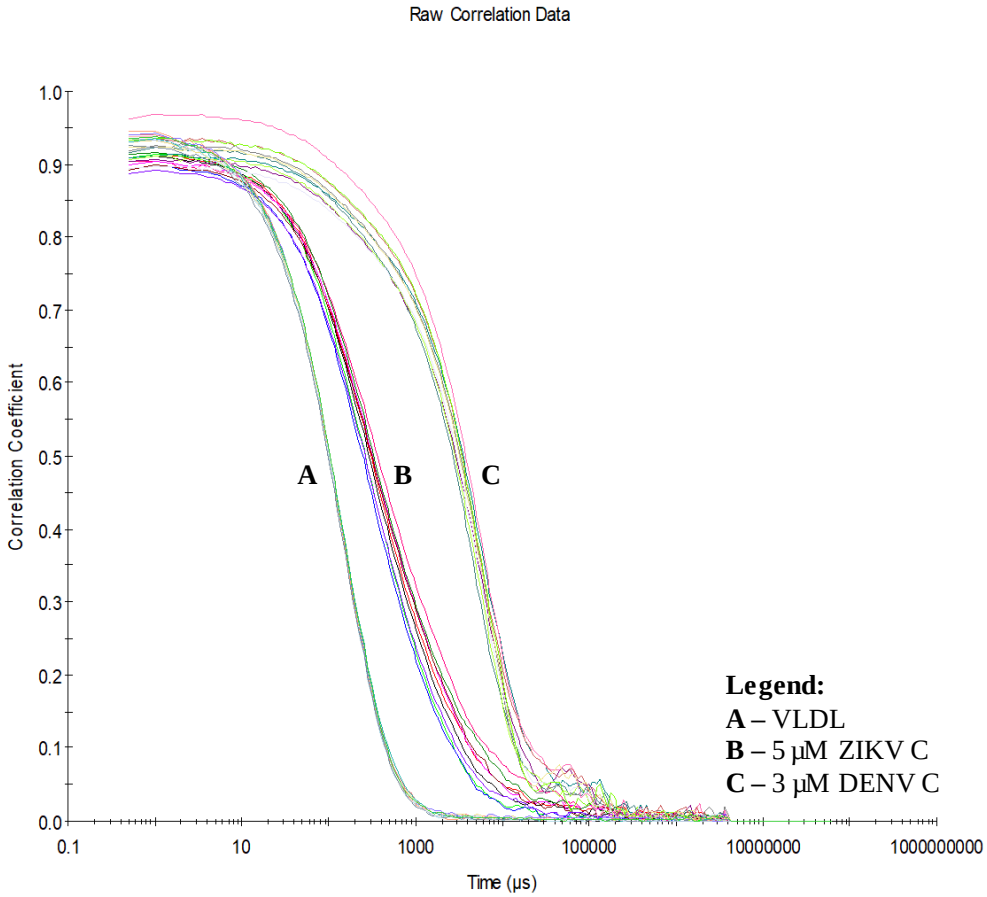


Figure 3.20 – DLS correlograms for VLDL (batch 1) with ZIKV C or DENV C proteins. With this batch of VLDL1, an interaction is observable for ZIKV C-VLDL due to the shift observable in the correlograms. Controls with DENV C were also made, which in previous reports shown that this protein interacts with VLDL at the concentration studied.

Analyzing the correlograms obtained (**Figure *3.20***) a shift in the correlograms is seen for the ZIKV C protein, indicating there might occur an interaction between VLDL and this protein. Experiments were repeated with DENV C protein as a positive control since it was previously reported and published that this interaction occurs at 3 µM of DENV C protein. Since the shift observed for ZIKV C protein is lesser than the observe for DENV C protein, this might indicate that the interaction between VLDL-ZIKV C protein is somehow weaker that the interaction between VLDL-DENV C protein.

A new vial of VLDL from a different supplier was used in the experiments. The intensity peak of the new vial of VLDL (which from now on will be called "VLDL2") was studied via DLS. This new batch of VLDL had a single peak and less fluctuations comparing to the "VLDL1" (single peak around 40 to 50 nm (for comparison see **Figure *3.19*** and **Figure *3.21***))

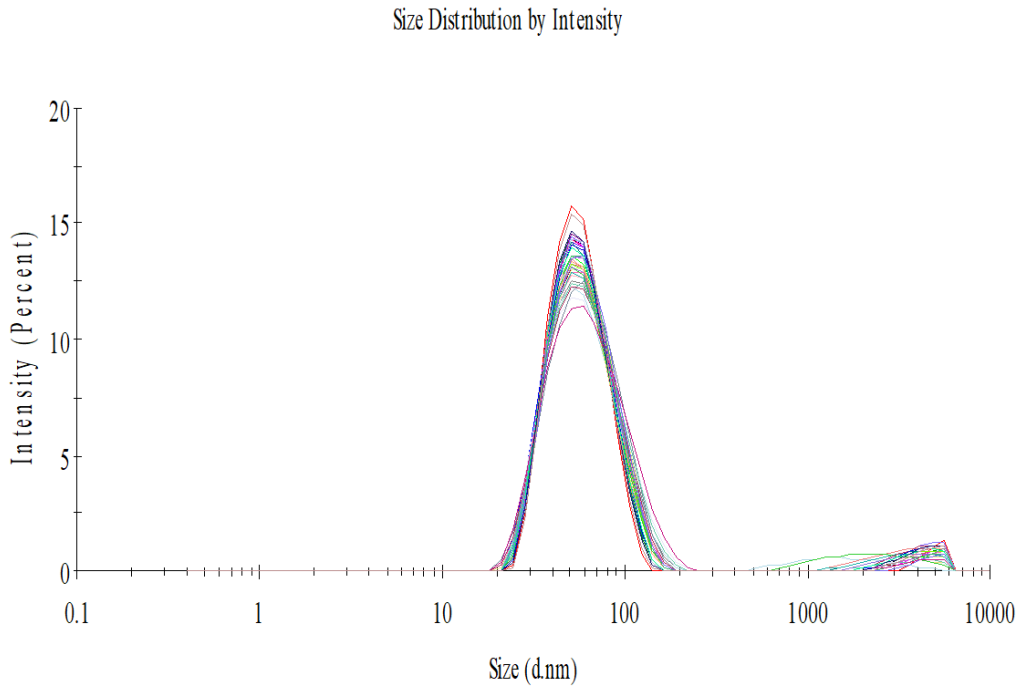


Figure 3.21 – Size distribution by intensity measured by DLS showing a single VLDL lipoproteins peak size (batch 2) around 40 – 50 nm. This new batch of VLDL showed a single peak around 50 nm, consisting with the size of these lipoproteins. Fewer signs of aggregation are observable.

The same procedures as described above were performed for the "VLDL2". The interaction between VLDL and ZIKV C was accessed with DLS technique and VLDL were titrated with increasing concentrations of ZIKV C, ranging from 1 μM to 5 μM where from 1 μM to 3 μM were made 0,5 μM increase at each time. A control was also made with DENV C protein, which already was described to interact with VLDL. The area under the curve of each correlogram obtained was also calculated to better observe the increase of the shift in the correlograms, allowing a better assessment if the interaction occurs between the VLDL lipoproteins and ZIKV C protein. Analyzing the correlograms there is no shift observable with the ZIKV C protein which can be also seen by the area under the curve of those correlograms that remains constant (**Figure 3.22 *A and* Figure 3.22 *C – black dots***). Observing the correlograms obtained with DENV C it is possible to observe a shift in the correlograms, in line with the previously reported for DENV C interaction with VLDL at 3 μM. Analyzing the area under the curve of those correlograms it is possible to observe that this data is again in concordance with the described in the literature.

Nonetheless, this batch of results enters in contradiction with the batch of results of the "VLDL1". In the "VLDL1" it was possible to observe a major shift for DENV C – VLDL interaction where with the "VLDL2" the same shift is present but in a smaller magnitude. For ZIKV C the results of "VLDL1" and "VLDL2" also contradict each other. In "VLDL1" a shift in the correlograms is present indicating that there is an interaction between VLDL and ZIKV C protein but with the "VLDL2" the shift is absent indicating that this interaction does not occur for ZIKV C protein.

No definitive conclusions can be drawn from these two batches of results relatively to the ZIKV C and VLDL interaction because with one batch of results there is an interaction (with the "VLDL1") and with the other such interaction does not exist (with the "VLDL2"). Since these results are inconclusive it is necessary to address this interaction with other technique as such atomic force microscopy (AFM), which exploits single-target molecular interactions.

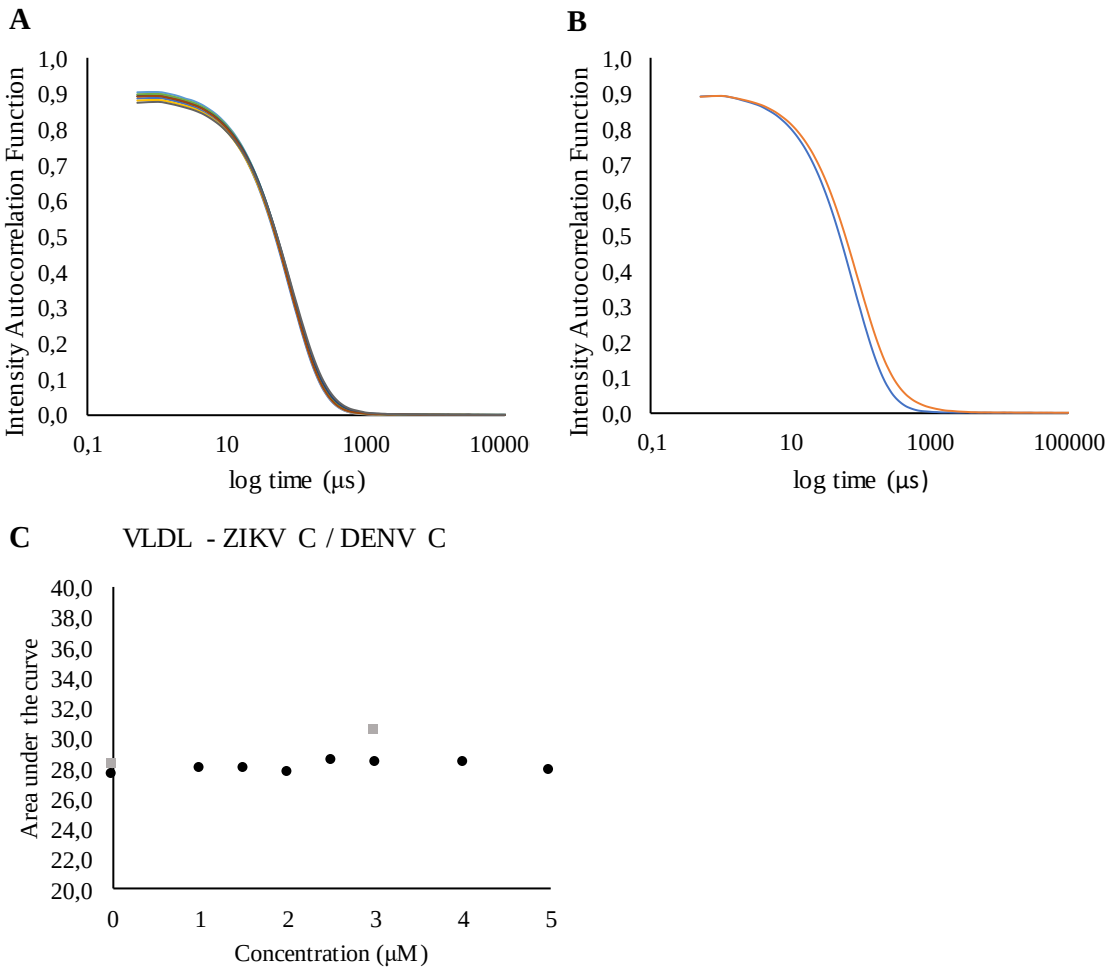


Figure 3.22 – **DLS correlograms for VLDL (batch 2) with increasing concentrations of ZIKV C and DENV C protein.** **(A)** Assessment of the interaction between VLDL and ZIKV C protein through dynamic light scattering, resulting in a no shift observed in the correlograms, where all correlograms superimpose to the correlogram obtained for VLDL. ZIKV C concentration is ranging from 1 µM to 5 µM where from 1 µM to 3 µM measurements were made in 0,5 µM of interval. **(B)** Assessment of DENV C and VLDL interaction which showed a small shift in the correlogram indicating that there is an interaction, which was previously reported. DENV C protein concentration utilized was 3 µM since it was the concentration necessary to promote the interaction between DENV C protein and VLDL. **(C)** Area under the curve of the correlograms versus ZIKV C (black dots) protein concentration and DENV C concentration (gray squares), which demonstrates that there is no increase of the shift of ZIKV C comparing with DENV C.

.

4 Discussion & Main Conclusions

4.1 Similarities between homologous *Flavivirus* C proteins

Flavivirus' C proteins have high similarity of sequence and structure. In terms of secondary structure, it was reported that these proteins have three distinct conserved motifs: a "conserved fold" between amino acids 33rd to 100th covering the α2, α3 and α4 regions of the capsid protein, a "flexible fold" between amino acids 24th and 32nd covering the α1 region, and a disordered region corresponding to the N-terminal region. Regarding the "conserved fold" that exists among 16 mosquito-borne Flavivirus[48] , there is a high probability of this region promoting dimerization, leading to the formation of a homodimer, including in ZIKV C. This has been described in the literature recently (March 30, 2018), where a truncated ZIKV C protein (without the N-terminal) forms a homodimer, revealed by X-Ray crystallography[45]. The analysis of predicted structures shows that the N-terminal and the α1 region may adopt alternative conformations. For DENV C, for example, several biological implications have been found for the location of the N-terminal, since it can be in a "open conformation" allowing interactions with host lipid systems, but also in a "closed conformation" responsible for a autoinhibitory mechanism which prevents such interactions.

Concerning the mosquito-borne flavivirus C proteins analyzed, understanding which of those conformations are more likely to occur in the context of a homodimer might be useful to develop new drugs. Previous work at host lab analyzing the predicted structures by visually examining the position of the N-terminal region allowed to classify them in 4 major clusters: 3 of those clusters corresponds to a "open conformation" (A, B, D, in which D is more closed that the others) and one corresponding to a "closed conformation" (C) (**Figure *3.2***). From that, it is possible to conclude that the autoinhibitory conformation is likely to occur and that might be a property of Flavivirus C proteins.

Based on this, a new analytical method to redistribute the structures into clusters was developed. This method agrees with almost all the previous predictions, except for: three of the structures previously belonging to cluster D, two of those shifted to cluster B (green) and the other to cluster C (red); and one structure belonging to cluster A (blue), which was assigned to cluster D (black). This difficulty in determining the cluster of the structures belonging to cluster D is due to the odd torsion that is observed among those structures in the α2 and α1 regions alongside the N-terminal position. This probably led to a misclassification by the visual analysis. Since this analytical method rests on a theoretical base and follows less biased criteria, it might be a better approach than the visual method (one relevant criteria is, for example, the determination of the probability of one specific structure belonging to the assigned cluster and not to the others by comparing the normal distribution of the data intra-cluster).

However, it is also important to consider some other aspects. The first one is that the most variable amino acid might not reflect the overall position and direction of the N-terminal if some torsions in those regions exist. The other one is that the algorithm is in constant update which might reflect different assignments for each run performed. In this case a manual curation is needed to address the validity of these newly assigned clusters. Despite this, the new method, based on theoretical criteria, is less biased than visually analyzing each structure and manually curating it to a specific cluster. Thus, this method might be applicable, for example, for analyzing IDPs (intrinsically disordered proteins, or regions within the protein that lack well defined structures), by studying homologous proteins distributing them into clusters of predicted conformers, determine the most likely conformations and then identify key properties of such proteins.

Regarding ZIKV, only two conformations of the capsid protein are viable in the context of a homodimer. From this, three viable homodimers were generated, one with the interface α2- α2' totally exposed to the solvent, which allows the interaction to host lipid systems; one completely closed in a autoinhibitory conformation; and one partially exposed. This suggests that these conformations are interchangeable from a autoinhibitory conformation to a transient state and then to a fully open conformation in which the capsid protein interacts/binds to host lipid systems. Also, computational analysis proved that ZIKV C is more compact than DENV C protein and that a superimposition of the pep14-23 with the N-terminal region exists.

From this it is hypothesized that the ZIKV C might have a functional similarity comparatively to DENV C protein, interacting with host lipid systems and that the interaction might be inhibited by pep14-23.

4.2 ZIKV C protein characterization

ZIKV C protein was expressed and purified with a high degree of purity. No contaminations were seen in the SDS page performed after the size exclusion column. To corroborate this, a single peak with the expected mass of the ZIKV C protein is observable in the mass spectrum acquired (see **Appendix I**, **Appendix II**). It is of note that when expressing this protein, some levels of toxicity were observed causing turbidity in culture probably due to cell lysis.

The protein easily aggregates in solution if it passes a certain threshold of concentration (> 115 µM in present conditions). This aggregation is not exclusively electrostatic-driven, since upon change of the conditions in the storing buffer to a high ionic strength aggregation it is still observable. Thus, some hydrophobic forces might be behind the aggregation observed. When a buffer containing 5% glycerol was added to a small fraction of protein that aggregated, all aggregation disappeared, corroborating the previous statement that hydrophobic forces might be the driving force behind the aggregation. Further studies are needed to acquire more data in this aspect. Despite this, enough pure soluble protein was obtained.

CD spectroscopy revealed that ZIKV C protein has an α-helix secondary structure, which was expected based on computational predictions. Temperature denaturation studies demonstrated that the protein is highly stable and does not fully denature in the range of temperatures studied (6 ºC to 96ºC). Also, when allowed to cool to room temperature the protein fully regains the secondary structure, showing an almost complete reversible process. From this data it is hypothesized that a highly stable structure is formed, probably through a dimer, as recently published.[45] When at high temperatures, ZIKV C protein partial denaturation suggests that only the N-terminal and the α1 region suffers denaturation, while the protein hydrophobic core keeps the protein intact and stable. This is also in line with the analysis of the computational structure.

4.3 ZIKV C interaction with host lipid systems

4.3.1 Interaction with LDs

The interaction of DENV C with LDs was shown to be essential for viral replication cycle. [51] In this thesis, the interaction between recombinant ZIKV C protein and purified LDs from BHK-21 was studied by ζ-potential measurements (chapter 3.3.1). LDs ζ-potential revealed a negative surface charge at pH 7,4 and 100 mM KCl, with an average of - 19,7 ± 2,7 mV, in line with the described in the literature[48,50,114]. Upon interaction with the ZIKV C protein, this negative surface charge becomes positive, with a maximum plateau at 10,7 ± 1,5 mV, at the highest ZIKV C concentration tested (5 μM).

For DENV C, a model was proposed[46], in which the α4-α4' region (positively charged) interacts with viral RNA (negatively charged) and the apolar α2-α2' region interacts with the viral membrane. DENV C shares high similarity with ZIKV C, thus this model could be translated to ZIKV C biological activity. ZIKV C protein dimer net charge at neutral pH is +46 (assuming solution-state physiologic pK_a values). The region with the highest density of positive charge is the α4-α4' helical domain. ZIKV C-LDs binding is likely to occur via the N-terminal and hydrophobic interface, leaving the polycationic domain (the α4-α4' region) exposed to the aqueous environment. This exposure can be an important step for RNA recognition since cationic charges are available to interact with the anionic charge of viral RNA, binding to it, for assembling and forming new virions. The interaction of ZIKV C with LDs might promote the protein binding to RNA and/or prevent it from aggregating *in vivo*.

Also, it was shown that LDs-DENV C interaction is highly dependent on potassium ions at physiologic concentrations that cannot be substituted by sodium ions. As proposed before, this K^+-dependent interactions between DENV C and LDs may be a mean for regulating DENV C protein release in the cytosol. K^+ is the main intracellular cation, while its concentration is low in the extracellular medium, where NaCl is responsible for osmolarity. Several viruses modulate infected cells' biochemistry and physiology. This fine-tuning of the ideal cellular conditions for viral replication often occurs at the specific intracellular ion concentrations level. In this thesis, all interactions were performed in potassium buffer, so ZIKV C follows that line of thought. Experiments in sodium buffer were also performed, showing that ZIKV C can equally interact in sodium buffer (see **Appendix IV**), contrarily to the reported for DENV C. From this, it is possible to begin to observe some differences relatively to DENV C, despite the high similarity among them.

On LDs' surface there are several proteins, such as proteins belonging to PAT family, namely perilipin (perilipin 1); ADRP (perilipin 2), TIP47 (perilipin 3), as well other proteins in smaller quantities.[129] This family of proteins covers at least 15-20 % of the surface of LDs[60] and are negatively charged at pH 7.4, which contributes to the negative charge of the ζ-potential measured for LDs before the addition of ZIKV C protein. To inquire if those proteins were responsible for the interaction of LDs with ZIKV C protein, the LDs were treated with trypsin which removes the protein coat attached to the LDs' surface. From the results after the removal by trypsin, obtained through ζ-potential measurements, it is possible to observe that a direct influence in the interaction occurs. This suggests that ZIKV C binds to a protein component on the surface of LDs, which is yet to be identified. Additionally, cooperativity among the ZIKV C proteins to interact with LDs might occur, but the fit without considering it is still reasonable and acceptable.

In summary, based on the findings in this thesis, it is proposed that the ZIKV C binds (possiblyin a cooperative way) to proteins on the surface of intracellular LDs and that this interaction is not solely electrostatic-driven. From the preliminary data it is possible to conclude that it might interact not only in the presence of K^+ ions, as observed for DENV C-LDs binding, but also in the presence of Na^+ ions.

Overall, this interaction may regulate the availability of ZIKV C protein in the cytosol (or other cellular compartment), thus modulating its role on the virus replication cycle.

4.3.2 Interaction with lipoproteins (VLDL and LDL)

The interaction between ZIKV C and lipoproteins was demonstrated by DLS spectroscopy, which monitored the particle size. In the absence of ZIKV C, VLDL and LDL had an average hydrodynamic diameter (D_H) of 28 nm and 18,2 nm, respectively. These values are in agreement with those reported in the literature.[63] Upon titration of these lipoproteins with ZIKV C, ZIKV C – LDL complex is formed, originating an increase in the diameter of 2,9 nm. This small increase in size reflects an interaction with LDL, suggesting either a protein with smaller size or that the protein is more inserted in the surface of the LDL. This is in accordance with the proposed structures of ZIKV C in the computational chapter, since they are more compact than the ones of DENV C protein.

Before proceeding, it is important to briefly discuss the recent publication of ZIKV C partial structure (February 2018), revealing a much compact structure, supporting the experimental and computational findings reported in this thesis. Briefly, ZIKV C protein experimental homodimer is very similar to WNV and DENV C experimental protein structures[46,47]. This is especially true regarding the "conserved fold" of these two proteins, identified by the host lab before, after comparing DENV and WNV C protein structures.[48] The host lab also identified a "flexible fold" in WNV and DENV, including the first α-helix (α1) and the N-terminal IDP region.[48] Comparing the residues of the same region of ZIKV C for which there is experimental information, with the homologous residues of WNV and DENV C shows that, overall, ZIKV is more compact and similar to WNV C than DENV in that region.[45] However, contrarily to DENV C, which structure was obtained via nuclear magnetic resonance (NMR) spectroscopy using a full-length protein, WNV and ZIKV C structures were obtained via X-Ray crystallography using truncated proteins, without, roughly the first 26 N-terminal amino acids. That corresponds to the IDP region that has a core role in DENV C biological activity, such as its interaction with host lipid systems, namely host LDs, an interaction essential for viral replication. [48,50,51,114] Thus, it is expected that WNV and ZIKV C structures are more compact and similar since they are both truncated and exclude a highly flexible IDP region. Also, they were both obtained via X-Ray, which promotes packing into compact structures to have adequate crystals. Thus, such structures are incomplete, and their size must be forcibly larger than reported in order to accommodate the IDP region, as described for the alternative predicted structures of DENV C N-terminal.[114] Despite this, the work described here is still of great relevance since there are yet no proposed structures that include the N-terminal IDP region.

The predicted structures of ZIKV C monomer, obtained out of its sequence, resulted in five conformers being possible. Since any model of the full length ZIKV C protein structure should account for dimerization and taking into consideration the similarities with DENV and WNV (similar C proteins), conformers that did not fit in a dimer structure were rejected and ended up with two monomer conformers and three likely dimer conformations, which make sense in light of the recently published structure of ZIKV C homodimer, as described below. A computational approach allowed to measure ZIKV C structure both in "fully open conformation" and "fully closed conformation". Two planes were drawn, one corresponding to the α4-α4' region and another to the N-terminal region. The distance between centroids is a measure of the size of the C protein in that conformation (**Figure *4.1***). From this analysis it was determined that ZIKV C in the open conformation measures 25,9 Å (2,6 nm) and in the closed conformation 19,5 Å (2,0 nm). This is roughly in accordance with the size increase of LDL upon interaction with ZIKV C protein, described previously.

No significant changes were observed in the "VLDL2" size. This suggests that ZIKV C only interacts with LDL and does not interact with VLDL, which was surprising comparing to the previously reported for DENV C.[67] Even though the interaction with VLDL could not be conclusive (due to their quality), the data seems to suggest that ZIKV C interacts with LDL but not VLDL. LDL derive from VLDL and present small but significant differences. VLDL in the bloodstream loses triacylglycerol and enriches in cholesterol esters, loses some of the intrinsic proteins and becomes smaller in size becoming LDL. Mature LDL have at their surface apoB100 (being the most important apolipoprotein in LDL) and apoE is not present at their surface. In previous reports[50,67,130], DENV C interaction with LDs and VLDL occurs through the binding to perilipin 3 (also known as TIP 47) and to apoE respectively, which have some degree of similarity. However, since ZIKV C interacts with LDs and with LDL and not with VLDL, a common protein responsible for the interaction/binding among LDL and LDs is yet to be identified. However, it is possible to affirm that the ZIKV C binds to these lipid systems in a mechanism that is dependent of K^+ and possibly Na^+, with the involvement of the ZIKV C' N-terminal and α2-α2' hydrophobic regions. Regarding the formation of a layer with ZIKV C dimer bound to LDL surface, this leads to the hypothesis of the formation of lipoviroparticles, a shared property with other *Flavivirus* (as such HCV, DENV and WNV) increasing dramatically the infectivity of the virus, possibly evading the host immune system and facilitating the crossing of blood-brain barrier, causing severe systemic and neurologic diseases. Considering the inhibition studied by pep14-23, it is possible to observe that it prevents the interaction between ZIKV C and LDL. According to previous reports, pep14-23 can inhibit the interaction between DENV C and LDs/VLDL. Since this peptide is designed based on the conserved motif of the N-terminal region of all 16 mosquito-borne flavivirus, this might be a powerful breakthrough for the design of a treatment that can prevent infection caused by those viruses. No treatment is available for ZIKV and, since this peptide may block a step of viral replication, it might lead to future pharmacological developments against this and related *Flavivirus*.

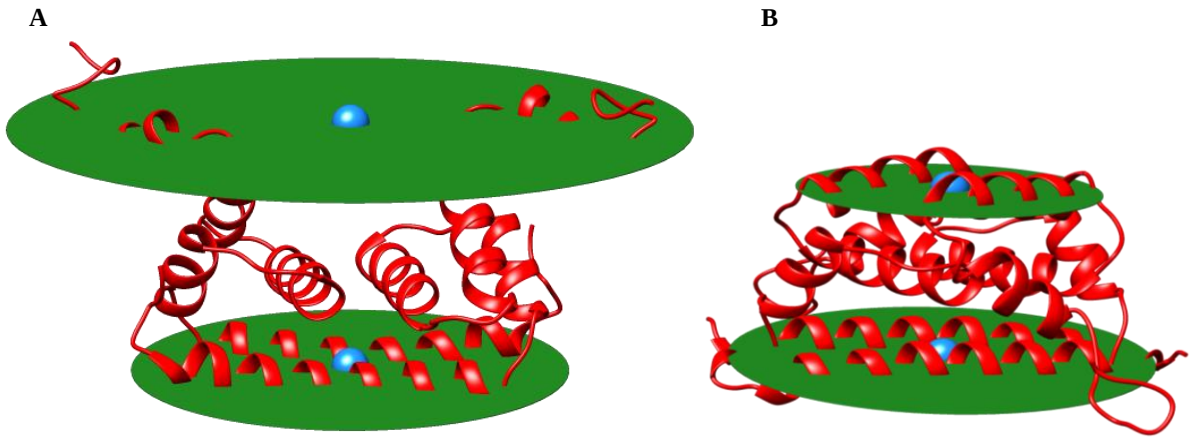


Figure 4.1 – Computational approach to determine the size of the capsid protein. Capsid protein in fully open conformation structure (A) in which planes were drawn across the N-terminal and α4-α4' region (green) and distance between centroids were measured (blue). The same procedure was applied to the fully closed conformational arrangement (B). The respective distances are 25,9 Å and 19,5 Å.

4.4 Future work

The underlying mechanism of the nucleocapsid formation in the Flavivirus cycle is still poorly understood. From this thesis it is elucidated that ZIKV C interacts with LDs and LDL, a surprise against the previous viruses studied in the host lab (DENV and WNV) which interacted with LDs and VLDL. However, ZIKV C interaction with VLDL needs to be repeated with new samples of VLDL to have conclusive data. A computational approach will also be used to inquire a common protein present in both LDs and LDL surface and then the data acquired can be integrated in molecular mechanistic models. Furthermore, experiments will be conducted using atomic force spectroscopy (AFM), a high-resolution technique that allows to evaluate the frequency of single molecule interactions between ZIKV C and host lipid systems as well as the (un)binding forces involved in those interactions. From that, interactions with LDs, VLDL and LDL will be accessed. After a common protein has been identified as a potential molecular target on LDs and lipoproteins, this technique can be further applied. Using antibodies that block the accessibility of ZIKV C to surface proteins of host lipid systems, it will be possible to ascertain that such molecular targets are indeed required for the binding.

Considering the inhibition promoted by pep14-23, further studies must be performed, whether by AFM (or other techniques), since it potentially impairs viral replication. Also, several versions derived from pep14-23 should be tested to obtain greater inhibition or better relation between the inhibition observed and the concentration of peptide used. This will clarify the mechanism of action and might provide grounds for improvements, in its anti-ZIKV activity, which may be then translated into the development of a drug inhibitor for related *Flavivirus* infections.

It is of great interest to study the interaction of ZIKV C with viral RNA, both in the absence or in the presence of lipid systems, to further elucidate how the assembly of the nucleocapsid is performed in the flavivirus in general. The study of peptide inhibitors should also be analyzed in this context.

With this multidisciplinary approach is thus expected to succeed in the future development of viral assembly-targeted inhibitors for ZIKV and in a large scope for all 16 mosquito-borne *Flavivirus.*

5 References

1. Mukhopadhyay, S., Kuhn, R. J. & Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.* **3,** 13–22 (2005).

2. Kuno, G., Chang, G. J., Tsuchiya, K. R., Karabatsos, N. & Cropp, C. B. Phylogeny of the genus Flavivirus. *J. Virol.* **72,** 73–83 (1998).

3. Lindenbach, B. D., Thiel, H.-J. & Rice, C. M. Flaviviridae: The Viruses and Their Replication. *Fields Virol.* (2001).

4. Weaver, S. C. & Barrett, A. D. T. Transmission cycles, host range, evolution and emergence of arboviral disease. *Nat. Rev. Microbiol.* **2,** 789–801 (2004).

5. Wilder-Smith, A. *et al.* Epidemic arboviral diseases: priorities for research and public health. *Lancet Infect. Dis.* **17,** e101–e106 (2017).

6. Pierson, T. C. & Kielian, M. Flaviviruses: braking the entering. *Curr. Opin. Virol.* **3,** 3–12 (2013).

7. Fernandez-Garcia, M.-D., Mazzon, M., Jacobs, M. & Amara, A. Pathogenesis of Flavivirus Infections: Using and Abusing the Host Cell. *Cell Host Microbe* **5,** 318–328 (2009).

8. WHO | Yellow fever. *WHO* (2018). Available at: http://www.who.int/mediacentre/factsheets/fs100/en/. (Accessed: 29th January 2018)

9. WHO | Dengue and severe dengue. *WHO* (2018). Available at: http://www.who.int/mediacentre/factsheets/fs117/en/. (Accessed: 26th February 2018)

10. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).

11. Brady, O. J. *et al.* Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus. *PLoS Negl. Trop. Dis.* **6,** e1760 (2012).

12. Lessler, J. *et al.* Assessing the global threat from Zika virus. *Science (80-.).* **353,** aaf8160-aaf8160 (2016).

13. Mlakar, J. *et al.* Zika Virus Associated with Microcephaly. *N. Engl. J. Med.* **374,** 951–958 (2016).

14. Cao-Lormeau, V.-M. *et al.* Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet* **387,** 1531–1539 (2016).

15. Perera-Lecoin, M., Meertens, L., Carnec, X. & Amara, A. Flavivirus entry receptors: an update. *Viruses* **6,** 69–88 (2013).

16. Bressanelli, S. *et al.* Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO J.* **23,** 728–738 (2004).

17. Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427,** 313–319 (2004).

18. Harak, C. & Lohmann, V. Ultrastructure of the replication sites of positive-strand RNA viruses. *Virology* **479–480,** 418–433 (2015).

19. Apte-Sengupta, S., Sirohi, D. & Kuhn, R. J. Coupling of replication and assembly in flaviviruses. *Curr. Opin. Virol.* **9,** 134–142 (2014).

20. Neufeldt, C. J., Cortese, M., Acosta, E. G. & Bartenschlager, R. Rewiring cellular networks by members of the Flaviviridae family. *Nat. Rev. Microbiol.* **16,** 125–142 (2018).

21. Murray, C. L., Jones, C. T. & Rice, C. M. Architects of assembly: roles of Flaviviridae non-structural proteins in virion morphogenesis. *Nat. Rev. Microbiol.* **6,** 699–708 (2008).

22. Martín-Acebes, M. A., Vázquez-Calvo, Á. & Saiz, J.-C. Lipids and flaviviruses, present and future perspectives for the control of dengue, Zika, and West Nile viruses. *Prog. Lipid Res.* **64,** 123–137 (2016).

23. Mukhopadhyay, S., Kim, B.-S., Chipman, P. R., Rossmann, M. G. & Kuhn, R. J. Structure of West Nile Virus. *Science (80-.).* **302,** 248–248 (2003).

24. Sirohi, D. *et al.* The 3.8 A resolution cryo-EM structure of Zika virus. *Science (80-.).* **352,** 467–470 (2016).

25. Freire, J. M., Santos, N. C., Veiga, A. S., Da Poian, A. T. & Castanho, M. A. R. B. Rethinking the capsid proteins of enveloped viruses: multifunctionality from genome packaging to genome transfection. *FEBS J.* **282,** 2267–2278 (2015).

26. Amberg, S. M. & Rice, C. M. Mutagenesis of the NS2B-NS3-mediated cleavage site in the flavivirus capsid protein demonstrates a requirement for coordinated processing. *J. Virol.* **73,** 8083–94 (1999).

27. Driggers, R. W. *et al.* Zika Virus Infection with Prolonged Maternal Viremia and Fetal Brain Abnormalities. *N. Engl. J. Med.* **374,** 2142–2151 (2016).

28. Oliveira, E. R. A., Mohana-Borges, R., de Alencastro, R. B. & Horta, B. A. C. The flavivirus capsid protein: Structure, function and perspectives towards drug design. *Virus Res.* **227,** 115–123 (2017).

29. Tadano, M., Makino, Y., Fukunaga, T., Okuno, Y. & Fukai, K. Detection of Dengue 4 Virus Core Protein in the Nucleus. I. A Monoclonal Antibody to Dengue 4 Virus Reacts with the Antigen in the Nucleus and Cytoplasm. *J. Gen. Virol.* **70,** 1409–1415 (1989).

30. Makino, Y. *et al.* Detection of Dengue 4 Virus Core Protein in the Nucleus. II. Antibody against Dengue 4 Core Protein Produced by a Recombinant Baculovirus Reacts with the Antigen in the Nucleus. *J. Gen. Virol.* **70,** 1417–1425 (1989).

31. Bulich, R. & Aaskov, J. G. Nuclear localization of dengue 2 virus core protein detected with monoclonal antibodies. *J. Gen. Virol.* **73,** 2999–3003 (1992).

32. Westaway, E. G., Khromykh, A. A., Kenney, M. T., Mackenzie, J. M. & Jones, M. K. Proteins C and NS4B of the Flavivirus Kunjin Translocate Independently into the Nucleus. *Virology* **234,** 31–41 (1997).

33. Hu, S.-T. *et al.* Intracellular localization and determination of a nuclear localization signal of the core protein of dengue virus. *J. Gen. Virol.* **83,** 3093–3102 (2002).

34. Mori, Y. *et al.* Nuclear localization of Japanese encephalitis virus core protein enhances viral replication. *J. Virol.* **79,** 3448–58 (2005).

35. Tsuda, Y. *et al.* Nucleolar Protein B23 Interacts with Japanese Encephalitis Virus Core Protein and Participates in Viral Replication. *Microbiol. Immunol.* **50,** 225–234 (2006).

36. Bhuvanakantham, R., Chong, M.-K. & Ng, M.-L. Specific interaction of capsid protein and importin-α/β influences West Nile virus production. *Biochem. Biophys. Res. Commun.* **389,** 63–69 (2009).

37. Colpitts, T. M., Barthel, S., Wang, P. & Fikrig, E. Dengue Virus Capsid Protein Binds Core Histones and Inhibits Nucleosome Formation in Human Liver Cells. *PLoS One* **6,** e24365 (2011).

38. Freire, J. M. *et al.* Peptides as models for the structure and function of viral capsid proteins: Insights on dengue virus capsid. *Biopolymers* **100,** 325–336 (2013).

39. Kiermayr, S., Kofler, R. M., Mandl, C. W., Messner, P. & Heinz, F. X. Isolation of capsid protein dimers from the tick-borne encephalitis flavivirus and in vitro assembly of capsid-like particles. *J. Virol.* **78,** 8078–84 (2004).

40. Schlick, P. *et al.* Characterization of West Nile virus live vaccine candidates attenuated by capsid deletion mutations. *Vaccine* **28,** 5903–5909 (2010).

41. Kofler, R. M., Heinz, F. X. & Mandl, C. W. Capsid Protein C of Tick-Borne Encephalitis Virus Tolerates Large Internal Deletions and Is a Favorable Target for Attenuation of Virulence. *J. Virol.* **76,** 3534–3543 (2002).

42. Samsa, M. M., Mondotte, J. A., Caramelo, J. J. & Gamarnik, A. V. Uncoupling cis-Acting RNA elements from coding sequences revealed a requirement of the N-terminal region of dengue virus capsid protein in virus particle formation. *J. Virol.* **86,** 1046–58 (2012).

43. Yamshchikov, V. F. & Compans, R. W. Processing of the intracellular form of the west Nile virus capsid protein by the viral NS2B-NS3 protease: an in vitro study. *J. Virol.* **68,** 5765–71 (1994).

44. Yamshchikov, V. F. & Compans, R. W. Formation of the flavivirus envelope: role of the viral NS2B-NS3 protease. *J. Virol.* **69,** 1995–2003 (1995).

45. Shang, Z., Song, H., Shi, Y., Qi, J. & Gao, G. F. Crystal Structure of the Capsid Protein from Zika Virus. *J. Mol. Biol.* **430,** 948–962 (2018).

46. Ma, L., Jones, C. T., Groesch, T. D., Kuhn, R. J. & Post, C. B. Solution structure of dengue virus capsid protein reveals another fold. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3414–9 (2004).

47. Dokland, T. *et al.* West Nile Virus Core Protein. *Structure* **12,** 1157–1163 (2004).

48. Martins, I. C. *et al.* The disordered N-terminal region of dengue virus capsid protein contains a lipid-droplet-binding motif. *Biochem. J.* **444,** 405–15 (2012).

49. Markoff, L., Falgout, B. & Chang, A. A Conserved Internal Hydrophobic Domain Mediates the Stable Membrane Integration of the Dengue Virus Capsid Protein. *Virology* **233,** 105–117 (1997).

50. Carvalho, F. A. *et al.* Dengue virus capsid protein binding to hepatic lipid droplets (LD) is potassium ion dependent and is mediated by LD surface proteins. *J. Virol.* **86,** 2096–108 (2012).

51. Samsa, M. M. *et al.* Dengue Virus Capsid Protein Usurps Lipid Droplets for Viral Particle Formation. *PLoS Pathog.* **5,** e1000632 (2009).

52. Tauchi-Sato, K., Ozeki, S., Houjou, T., Taguchi, R. & Fujimoto, T. The surface of lipid droplets is a phospholipid monolayer with a unique Fatty Acid composition. *J. Biol. Chem.* **277,** 44507–12 (2002).

53. Ohsaki, Y., Cheng, J., Fujita, A., Tokumoto, T. & Fujimoto, T. Cytoplasmic lipid droplets are sites of convergence of proteasomal and autophagic degradation of apolipoprotein B. *Mol. Biol. Cell* **17,** 2674–83 (2006).

54. Buhman, K. K., Chen, H. C. & Farese, R. V. The enzymes of neutral lipid synthesis. *J. Biol. Chem.* **276,** 40369–72 (2001).

55. Pol, A., Gross, S. P. & Parton, R. G. Review: biogenesis of the multifunctional lipid droplet:

lipids, proteins, and sites. *J. Cell Biol.* **204,** 635–46 (2014).

56. Gao, Q. & Goodman, J. M. The lipid droplet—a well-connected organelle. *Front. Cell Dev. Biol.* **3,** 49 (2015).

57. Lohmann, D. *et al.* Monoubiquitination of Ancient Ubiquitous Protein 1 Promotes Lipid Droplet Clustering. *PLoS One* **8,** e72453 (2013).

58. Velázquez, A. P., Tatsuta, T., Ghillebert, R., Drescher, I. & Graef, M. Lipid droplet-mediated ER homeostasis regulates autophagy and cell survival during starvation. *J. Cell Biol.* **212,** 621–31 (2016).

59. Romero-Brey, I. *et al.* Three-Dimensional Architecture and Biogenesis of Membrane Structures Associated with Hepatitis C Virus Replication. *PLoS Pathog.* **8,** e1003056 (2012).

60. Walther, T. C. & Farese, R. V. The life of lipid droplets. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1791,** 459–466 (2009).

61. Heaton, N. S. & Randall, G. Dengue virus-induced autophagy regulates lipid metabolism. *Cell Host Microbe* **8,** 422–32 (2010).

62. Zhang, J., Lan, Y. & Sanyal, S. Modulation of Lipid Droplet Metabolism—A Potential Target for Therapeutic Intervention in Flaviviridae Infections. *Front. Microbiol.* **8,** 2286 (2017).

63. Cushley, R. J. & Okon, M. NMR Studies of Lipoprotein Structure. *Annu. Rev. Biophys. Biomol. Struct.* **31,** 177–206 (2002).

64. Boyle, J. *Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. Biochemistry and Molecular Biology Education* **33,** (John Wiley & Sons Inc., 2005).

65. Welsch, S. *et al.* Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites. *Cell Host Microbe* **5,** 365–375 (2009).

66. André et al. *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy - Google Livros*. (2002).

67. Faustino, A. F. *et al.* Dengue virus capsid protein interacts specifically with very low-density lipoproteins. *Nanomedicine-Nanotechnology Biol. Med.* **10,** 247–255 (2014).

68. Bartenschlager, R., Penin, F., Lohmann, V. & André, P. Assembly of infectious hepatitis C virus particles. *Trends Microbiol.* **19,** 95–103 (2011).

69. Agnello, V., Abel, G., Elfahal, M., Knight, G. B. & Zhang, Q. X. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 12766–71 (1999).

70. Faustino, A. F. *et al.* Understanding Dengue Virus Capsid Protein Interaction with Key Biological Targets. *Sci. Rep.* **5,** (2015).

71. Haddow, A. D. *et al.* Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage. *PLoS Negl. Trop. Dis.* **6,** e1477 (2012).

72. Dick, G. W. ., Kitchen, S. . & Haddow, A. . Zika Virus (I). Isolations and serological specificity. *Trans. R. Soc. Trop. Med. Hyg.* **46,** 509–520 (1952).

73. Saiz, J.-C. *et al.* Zika Virus: the Latest Newcomer. *Front. Microbiol.* **7,** 496 (2016).

74. SMITHBURN, K. C. Neutralizing antibodies against certain recently isolated viruses in the sera of human beings residing in East Africa. *J. Immunol.* **69,** 223–34 (1952).

75. Faye, O. *et al.* Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. *PLoS Negl. Trop. Dis.* **8,** e2636 (2014).

76. MacNamara, F. . Zika virus : A report on three cases of human infection during an epidemic of jaundice in Nigeria. *Trans. R. Soc. Trop. Med. Hyg.* **48,** 139–145 (1954).

77. Duffy, M. R. *et al.* Zika Virus Outbreak on Yap Island, Federated States of Micronesia. *N. Engl. J. Med.* **360,** 2536–2543 (2009).

78. Rocklöv, J. *et al.* Assessing Seasonal Risks for the Introduction and Mosquito-borne Spread of Zika Virus in Europe. *EBioMedicine* **9,** 250–256 (2016).

79. Musso, D., Nilles, E. J. & Cao-Lormeau, V.-M. Rapid spread of emerging Zika virus in the Pacific area. *Clin. Microbiol. Infect.* **20,** O595–O596 (2014).

80. Oehler, E. *et al.* Zika virus infection complicated by Guillain-Barre syndrome--case report, French Polynesia, December 2013. *Euro Surveill.* **19,** (2014).

81. Roth, A. *et al.* Concurrent outbreaks of dengue, chikungunya and Zika virus infections – an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012–2014. *Eurosurveillance* **19,** 20929 (2014).

82. Zanluca, C. *et al.* First report of autochthonous transmission of Zika virus in Brazil. *Mem. Inst. Oswaldo Cruz* **110,** 569–572 (2015).

83. Foy, B. D. *et al.* Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerg. Infect. Dis.* **17,** 880–2 (2011).

84. Cao-Lormeau, V.-M. *et al.* Zika Virus, French Polynesia, South Pacific, 2013. *Emerg. Infect. Dis.* **20,** 1084–1086 (2014).

85. Musso, D. & Gubler, D. J. Zika Virus. *Clin. Microbiol. Rev.* **29,** 487–524 (2016).

86. Schuler-Faccini, L. *et al.* Possible Association Between Zika Virus Infection and Microcephaly — Brazil, 2015. *MMWR. Morb. Mortal. Wkly. Rep.* **65,** 59–62 (2016).

87. de Paula Freitas, B. *et al.* Ocular Findings in Infants With Microcephaly Associated With Presumed Zika Virus Congenital Infection in Salvador, Brazil. *JAMA Ophthalmol.* **134,** 529 (2016).

88. Martín-Acebes, M. A. & Saiz, J.-C. West Nile virus: A re-emerging pathogen revisited. *World J. Virol.* **1,** 51 (2012).

89. Medlock, J. M. *et al.* A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector Borne Zoonotic Dis.* **12,** 435–47 (2012).

90. Rezza, G. *et al.* Infection with chikungunya virus in Italy: an outbreak in a temperate region. *Lancet* **370,** 1840–1846 (2007).

91. Reiter, P. Yellow fever and dengue: a threat to Europe? *Eurosurveillance* **15,** 19509 (2010).

92. Gjenero-Margan, I. *et al.* Autochthonous dengue fever in Croatia, August-September 2010. *Euro Surveill.* **16,** (2011).

93. Diagne, C. T. *et al.* Potential of selected Senegalese Aedes spp. mosquitoes (Diptera: Culicidae) to transmit Zika virus. *BMC Infect. Dis.* **15,** 492 (2015).

94. Anez, G., Chancey, C., Grinev, A. & Rios, M. Dengue virus and other arboviruses: a global view of risks. *ISBT Sci. Ser.* **7,** 274–282 (2012).

95. Marchette, N. J., Garcia, R. & Rudnick, A. Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia. *Am. J. Trop. Med. Hyg.* **18,** 411–5 (1969).

96. Grard, G. *et al.* Zika Virus in Gabon (Central Africa) – 2007: A New Threat from Aedes albopictus? *PLoS Negl. Trop. Dis.* **8,** e2681 (2014).

97. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *Elife* **4,** e08347 (2015).

98. Paupy, C., Delatte, H., Bagny, L., Corbel, V. & Fontenille, D. Aedes albopictus, an arbovirus vector: From the darkness to the light. *Microbes Infect.* **11,** 1177–1185 (2009).

99. Dyer, O. Sixty seconds on . . . Zika virus. *BMJ* **352,** i467 (2016).

100. Thomas, S., Obermayr, U., Fischer, D., Kreyling, J. & Beierkuhnlein, C. Low-temperature threshold for egg survival of a post-diapause and non-diapause European aedine strain, Aedes albopictus (Diptera: Culicidae). *Parasit. Vectors* **5,** 100 (2012).

101. Higgs, S. Zika Virus: Emergence and Emergency. *Vector-Borne Zoonotic Dis.* **16,** 75–76 (2016).

102. Scott, T. W. & Takken, W. Feeding strategies of anthropophilic mosquitoes result in increased risk of pathogen transmission. *Trends Parasitol.* **28,** 114–121 (2012).

103. Ladhani, S. N., O'Connor, C., Kirkbride, H., Brooks, T. & Morgan, D. Outbreak of Zika virus disease in the Americas and the association with microcephaly, congenital malformations and Guillain-Barré syndrome. *Arch. Dis. Child.* **101,** 600–2 (2016).

104. Besnard, M., Lastère, S., Teissier, A., Cao-Lormeau, V. & Musso, D. Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. *Eurosurveillance* **19,** 20751 (2014).

105. Musso, D. *et al.* Potential Sexual Transmission of Zika Virus. *Emerg. Infect. Dis.* **21,** 359–361 (2015).

106. Domínguez-Moreno, R. *et al.* [Mortality associated with a diagnosis of Guillain-Barré syndrome in adults of Mexican health institutions]. *Rev. Neurol.* **58,** 4–10 (2014).

107. Butler, D. First Zika-linked birth defects detected in Colombia. *Nature* **531,** 153–153 (2016).

108. Gouw, S. C. *et al.* Intensity of factor VIII treatment and inhibitor development in children with severe hemophilia A: the RODIN study. *Blood* **121,** 4046–4055 (2013).

109. Puccioni-Sohler, M., Orsini, M. & Soares, C. N. Dengue: a new challenge for neurology. *Neurol. Int.* **4,** 15 (2012).

110. Calvet, G. *et al.* Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect. Dis.* **16,** 653–660 (2016).

111. Rubin, E. J., Greene, M. F. & Baden, L. R. Zika Virus and Microcephaly. *N. Engl. J. Med.* **374,** 984–985 (2016).

112. Musso, D. *et al.* Detection of Zika virus in saliva. *J. Clin. Virol.* **68,** 53–55 (2015).

113. Balm, M. N. D. *et al.* A diagnostic polymerase chain reaction assay for Zika virus. *J. Med. Virol.* **84,** 1501–1505 (2012).

114. Faustino, A. F. *et al.* Understanding Dengue Virus Capsid Protein Disordered N-Terminus and pep14-23-Based Inhibition. *ACS Chem. Biol.* **10,** 517–526 (2015).

115. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751,** 119–139 (2005).

116. Miles, A. J. & Wallace, B. A. Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. *Chem. Soc. Rev.* **35,** 39–51 (2006).

117. Koepf, E. K., Petrassi, H. M., Sudol, M. & Kelly, J. W. WW: An isolated three-stranded antiparallel β-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* **8,** 841–853 (2008).

118. Panda, A., Martindale, J. & Gorospe, M. Polysome Fractionation to Analyze mRNA Distribution Profiles. *BIO-PROTOCOL* **7,** (2017).

119. Freire, J. M. *et al.* Using zeta-potential measurements to quantify peptide partition to lipid membranes. *Eur. Biophys. J.* **40,** 481–487 (2011).

120. Santos, N. C. & Castanho, M. A. Teaching light scattering spectroscopy: the dimension and shape of tobacco mosaic virus. *Biophys. J.* **71,** 1641–50 (1996).

121. Berne, B. J. & Pecora, R. *Dynamic light scattering : with applications to chemistry, biology, and physics*. (R.E. Krieger Pub. Co, 1990).

122. Zimm, B. H. The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *J. Chem. Phys.* **16,** 1093–1099 (1948).

123. Koppel, D. E. Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. *J. Chem. Phys.* **57,** 4814–4820 (1972).

124. Frisken, B. J. Revisiting the method of cumulants for the analysis of dynamic light-scattering data.

125. Provencher, S. W. CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.* **27,** 229–242 (1982).

126. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39,** 1–38 (1977).

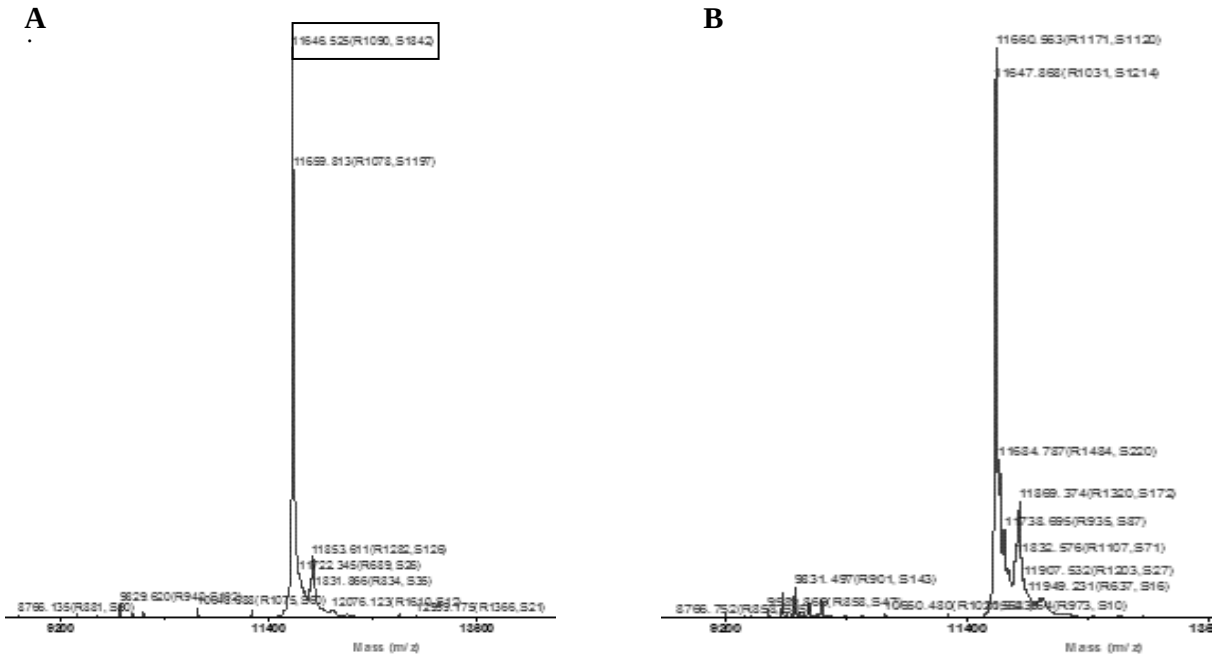
127. Snyder, J. P. *Flattening the earth : two thousand years of map projections*. (University of Chicago Press, 1993).

128. Dhinakaran, S., Afonso, A. M., Alves, M. A. & Pinho, F. T. Steady viscoelastic fluid flow between parallel plates under electro-osmotic forces: Phan-Thien–Tanner model. *J. Colloid Interface Sci.* **344,** 513–520 (2010).

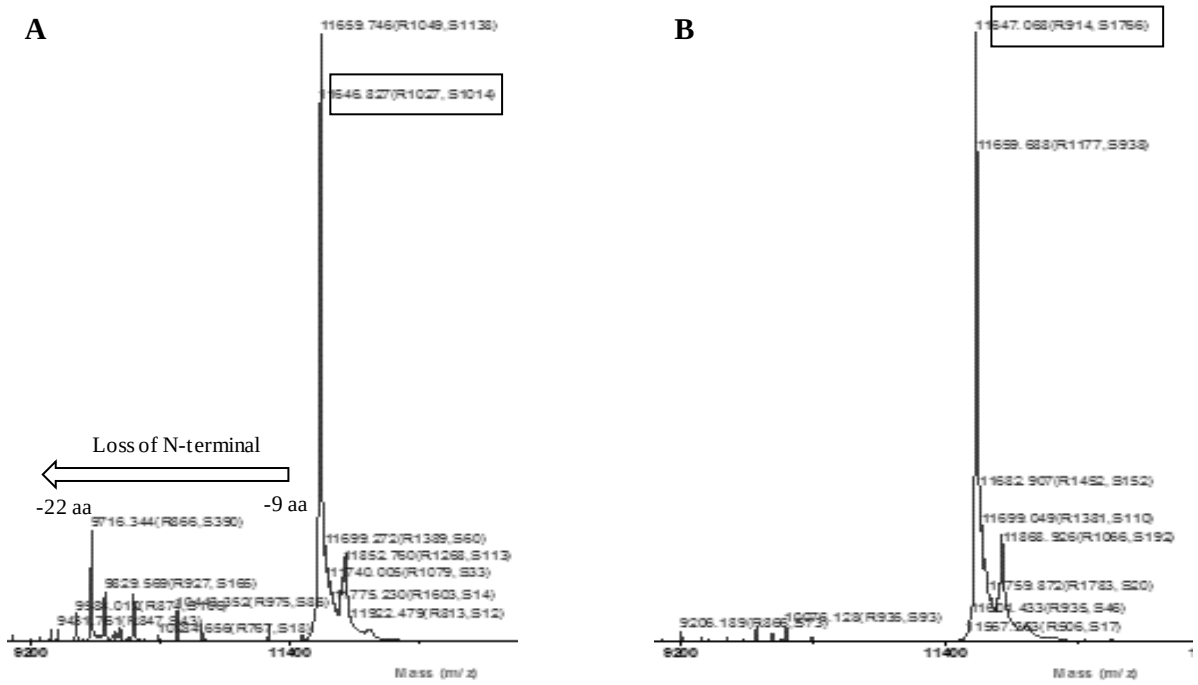
129. Olofsson, S.-O. *et al.* Lipid droplets as dynamic organelles connecting storage and efflux of lipids. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1791,** 448–458 (2009).

130. Hickenbottom, S. J., Kimmel, A. R., Londos, C. & Hurley, J. H. Structure of a Lipid Droplet Protein: The PAT Family Member TIP47. *Structure* **12,** 1199–1207 (2004).

6 Appendix



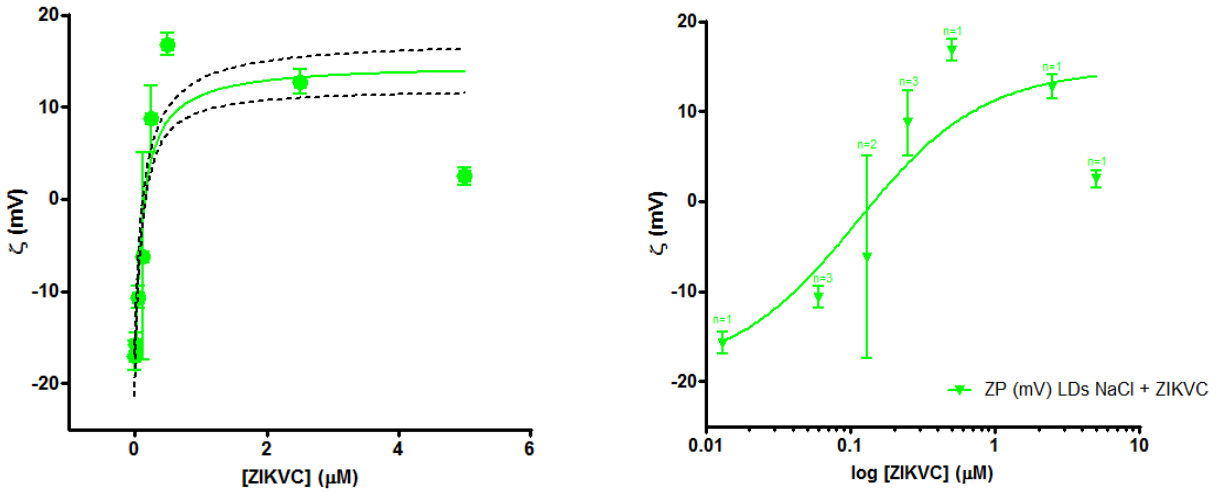
Appendix I – Mass spectrometry analysis of purified ZIKV C protein with 0,2 M KCl buffer. Quality control for the first expression/purification of ZIKV C, in which the storing buffer had 0,2 M KCl. Spectrum for fraction A5 eluted from the size-exclusion S200 column (shown to have highest protein concentration). Fraction A5 aggregated nearly after elution from the column, thus a centrifugation was performed (5 min, *16110* × *g*, 4 ºC) where **(A)** and **(B)** represent the spectrum for resultant supernatant and pellet, respectively. Boxed values represent the obtained mass for the ZIKV C protein (approximately 11647 Da), indicating that ZIKV C protein can be used in further studies.



Appendix II – Mass spectrometry analysis of purified ZIKV C protein with 2 M KCl buffer. Quality control for the second expression/purification of ZIKV C, in which the storing buffer had 2 M KCl. Spectrum for fraction A5 eluted from the size-exclusion S200 column (shown to have highest protein concentration). Fraction A5 aggregated nearly after elution from the column, thus a centrifugation was performed (5 min, *16110* × *g*, 4 ºC) where (A) and (B) represent the spectrum for resultant supernatant and pellet, respectively. Boxed values represent the obtained mass for the ZIKV C protein (approximately 11647 Da). However, the spectrum obtained for the ZIKV C protein in the supernatant (A) shows signs of degradation. This indicates that the protein has some degree of degradation in the N-terminal region (region of interest in the further studies), thus being not acceptable for use.

Appendix III – Summary table with the calculations for the determination of the degree of inhibition by pep14-23 in LDL-ZIKV C interaction.

	LDL	**LDL + pep14-23**	**LDL + 1 µM ZIKV C**	**LDL + 1 µM ZIKV C + pep14-23**	**LDL + 3 µM ZIKV C**	**LDL + 3 µM ZIKV C + pep14-23**	**LDL + 5 µM ZIKV C**	**LDL + 5 µM ZIKV C + pep14-23**
Area under the curve	21.1	21.0	22.9	22.8	30.6	25.7	37.7	27.7
Subtracted to VLD alone	0	-0.1	1.8	1.7	9.5	4.6	16.6	6.6
Increase upon adding ZIKV C (%)	N.A.	-0.5	8.5	8.1	45.0	21.8	78.7	31.3



Appendix IV – Preliminary data obtained for LDs-ZIKV C interaction in sodium buffer.
ζ-potential for the isolated LDs was accessed in the absence and in the presence of ZIKV C protein, ranging from 0,013 µM to 5 µM. The values were plotted in a linear scale **(A)** and in a logarithmic scale **(B)**, where the green dots represent the obtained values, the green line represents the fit used and the dashed lines represent the 95% confidence interval of the fit used. Results are represented as mean ± standard error.