**Nina Elisabeth Karguth**

BSc. Biomedicine

# Study of Dengue virus capsid protein in the context of its interaction with nucleic acids

Dissertation to obtain the Master of Science Degree in
Molecular Genetics and Biomedicine

Supervisor: Dr. Ivo C. Martins, Principal Investigator,
Instituto de Medicina Molecular João Lobo Antunes (iMM),
Lisbon, Portugal

Co-Supervisor: Prof. Dr. Nuno C. Santos, Head of Unit,
Instituto de Medicina Molecular João Lobo Antunes (iMM),
Lisbon, Portugal
Associate Professor with Habilitation, Faculdade de Medicina,
Universidade de Lisboa, Lisbon, Portugal

Júri:

Presidente: Prof. Dr. Paula Gonçalves

Orientador: Dr. Ivo C. Martins

Arguente: Prof. Dr. Francisco Caldeira

**September 2019**

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

*Preface*

This work is the result of a one-year research on dengue virus, a member of the *Flavivirus* genus, *Flaviviridae* family, a taxon with many important human pathogens such as Zika and West Nile viruses. Concerning dengue, this is an increasing global threat, not only restricted to developing countries. It is becoming more present also in Europe, due to the expansion of its mosquito vector. This thesis is focused on dengue virus capsid protein interaction with nucleic acids. The capsid protein is crucial to the viral life cycle, regulating important steps, such as viral assembly and encapsidation, with the interaction with host lipid droplets being essential for viral replication. Understanding the capsid protein mode of action can help to develop peptide inhibitors of specific interactions, a possible way to treat dengue and other *Flavivirus* infections.

To investigate this important protein, biophysical techniques, namely circular dichroism and intrinsic fluorescence spectroscopy were employed. Moreover, computational approaches helped to explore proteins structural data and obtain new insights into its mode of action.

My contribution to the outcome of this work included the production of recombinant protein, an intensive optimization of circular dichroism experiments and their implementation, alongside intrinsic fluorescence spectroscopy studies, complemented by a careful computational analysis of protein structure. As a major accomplishment, I was able to show interaction between the capsid protein and single-strand DNA, that might lead to a structural change and, furthermore, supports the current hypothesis that viral RNA binds to the positively charged C-terminal part of the capsid protein.

This thesis is integrated in the Masters in Molecular Genetics and Biomedicine from the *Faculdade de Ciências e Tecnologia da Universidade Nova* (FCT, Lisbon, Portugal) and was developed at the Biomembranes and Nanomedicine Unit of the *Instituto de Medicina Molecular João Lobo Antunes* (iMM*), Faculdade de Medicina da Universidade de Lisboa* (FMUL), *Universidade de Lisboa,* Lisbon, Portugal, under the supervision of Dr. Ivo C. Martins and Prof. Dr. Nuno C. Santos.

The work is incorporated in the Biomembranes and Nanomedicine Unit's research line, already ongoing before my arrival, and is compilated in this thesis, constituted by four main chapters. First, the reader will find state-of-the-art information regarding the dengue virus infection, virion structure and the viral life cycle, with a special focus on dengue virus capsid protein. The methodology section contains the theoretical and practical basis of the techniques and procedures employed. The next chapter represents the results achieved, before being discussed in the context of the unit's research and current knowledge in this field. Finally, the conclusions are integrated in a model hypothesizing a new mechanism of interaction between the capsid protein and nucleic acids.

I had the chance to participate in a national meeting, the 4th Chemistry: Shaping the future & 2019 Summer School (4ECQUL), Lisbon, Portugal in which I presented and discussed the results in poster format (please consult *Appendix II: Current outcomes of the work presented* – Poster communication).

- **Nina Karguth**, Ana S. Martins, Nuno C. Santos, Ivo C. Martins. Biophysical studies of dengue virus capsid protein interaction with nucleic acids. 4th Chemistry: Shaping the future & 2019 Summer School (4ECQUL), July 16 – 18 2019, Lisbon, Portugal.

Furthermore, I had the opportunity to present and discuss my work in another national meeting, NOVA Biophysica, September 4-6, 2019 in Lisbon, Portugal. I got selected for the presentation of a flash talk.

Besides that, the novel findings of my research work during this Master's project have been incorporated into one publication (please consult *Appendix II: published articles*).

- André F. Faustino, Ana S. Martins, **Nina Karguth**, Vanessa Artilheiro, Francisco J. Enguita, Joana C. Ricardo, Nuno C. Santos and Ivo C. Martins. Structural and functional properties of the capsid protein of Dengue and related *Flavivirus*. *Int J Mol Sci.* 2019 Aug 8;20(16).

*Acknowledgments*

First of all, I want to thank Prof. Nuno C. Santos for accepting me as a member of his lab and giving me the opportunity to perform my master thesis at the *Instituto de Medicina Molecular* in Lisbon. Furthermore, I wish to express my gratitude to every one of the Biomembranes and Nanomedicine Unit at *iMM* who helped me during this work.
A special thanks goes to my supervisor Dr. Ivo Martins. Thank you very much for the everyday support not only in the scientific area, all the help during every stage of the thesis and always being open to talk and discuss ideas.
I also would like to thank especially Ana, Marco and Mário from the Nuno Santos Lab for their help, knowledge and support during the whole time. Ana, thank you for teaching me various techniques and always taking time to answer my questions and help me. Obrigada!

Overall, I want to thank my family – my mom, my dad and my sister – for all their support and trust in me during my studies, especially during my time in Portugal. My family gave me all the conditions to focus on my graduation. Thank you for always encouraging me to follow my goals and dreams and giving me the opportunity to be brave and try new things.

Last but not least, I would like to give a huge thanks to my friends, which are now spread all over Europe. Every one of you is really important to me and I hope to see you as often as I can. Additionally, a special thanks goes to my Lisbon climbing crew, for giving me the right balance.

*Abstract*

Dengue virus (DENV) is an increasing global threat due to expansion of its mosquito vector. New effective therapies require full understanding of the viral life cycle at the level of molecular mechanisms, since this knowledge can trigger progresses in basic and applied research. The viral capsid (C) protein is of special importance since, due to its interaction with different ligands, it mediates key steps of the viral life cycle. DENV C is a structural protein involved in the viral assembly, binding and encapsidation of the viral genome. However, molecular details of DENV C binding to nucleic acids are not yet clear. DENV C contains a disordered N-terminal region, an intermediate flexible fold section and a very stable conserved fold region. The interaction of DENV C with intracellular lipid droplets was previously studied by the host lab. This led to understanding structural and functional roles of DENV C N-terminal disordered domain and to the design of pep14-23, a novel patented viral inhibitor peptide. Furthermore, sections within the viral genome involved in the binding to DENV C were identified. Now, the focus was to produce recombinant DENV C protein and to characterize DENV C binding to specific nucleic acid sequences via biophysical approaches. For that, single-strand (ss) DNA analogous to ssRNA sequences, identified as relevant to DENV C binding, were used as proxies of the viral genome. Different interaction profiles with ssDNA sequences were observed. The data show specific binding to one particular structured ssDNA. This causes a minor structural change upon DENV C binding to nucleic acids, possibly via its C-terminal $\alpha$4-$\alpha$4' region. This knowledge may enable, later on, to target the most relevant sequences to inhibit these interactions. Such crucial knowledge regarding DENV C interaction with nucleic acids may yield promising advances in therapies against dengue and closely related flaviviruses.

**Keywords:** Dengue virus; capsid protein; protein nucleic acids interaction; circular dichroism; fluorescence spectrometry

*Resumo*

O vírus da Dengue (DENV) é uma ameaça global crescente, devido à expansão dos seus vetores – os mosquitos *Aedes* spp. O DENV é responsável pela maioria dos casos de febre hemorrágica a nível mundial e pertence ao género *Flavivirus*. Novas terapias eficazes requerem a compreensão do ciclo da vida viral ao nível dos mecanismos moleculares. A proteína da cápside (C) é de especial importância, regulando vários passos cruciais do ciclo da vida viral. É uma proteína estrutural envolvida na montagem viral, ligando-se e ajudando a encapsular o genoma. A proteína forma homodímeros, sendo que os dois monómeros constituintes têm estruturas terciárias idênticas: apresentam uma região N-terminal não-estruturada, seguida de quatro hélices α. Pensa-se que a região hidrofóbica ($\alpha$2-$\alpha$2') irá interage com lípidos e a região positivamente carregada ($\alpha$4-$\alpha$4') interage com o RNA viral. Foi descoberto que a proteína C do vírus da Dengue se liga a corpúsculos lipídicos intracelulares. Esta interação foi estudada e, com base nisto, foi desenvolvido um péptido (pep14-23) capaz de inibir a ligação da proteína C aos corpúsculos lipídicos. É importante comprender também a interação da proteína C com o RNA viral. O primeiro objetivo deste trabalho consistiu em produzir proteína C recombinante do Dengue. Depois, o objetivo foi caracterizar a ligação da DENV C a sequências específicas de ácidos nucleicos. Para tal foram usadas abordagens biofísicas. Sequências de cadeia única de DNA análogas ao RNA viral relevante foram usadas como *proxies* do genoma viral. Os dados mostram ligações a sequências de ssDNA específicas que poderão originar uma pequena alteração estrutural na DENV C, possivelmente através da região C-terminal. Juntamente com a compreensão da ligação da DENV C a sistemas lipídicos, este conhecimento pode levar ao desenho de inibidores, um passo crucial no desenvolvimento de terapias contra Dengue e flavivírus relacionados.

**Palavras chaves:** vírus da dengue; proteína da cápside; dicroísmo circular; espectroscopia de fluorescência

**INDEX**

*List of Figures and Tables*

*List of Abbreviations*

The acronyms used are defined on first usage. For reasons of text economy very common acronyms, scientific or not (such as “RNA” or “USA”), are not described. Amino acid residues and sequences are indicated using the one-letter code.

| | |
|---|---|
| **APOE** | apolipoprotein E |
| **C** | capsid protein |
| **CD** | circular dichroism spectroscopy |
| **cryo EM** | cryo-electron microscopy |
| **DC** | dendritic cell |
| **DENV** | dengue virus |
| **DF** | dengue fever |
| **DHF** | dengue haemorrhagic fever |
| **DMSO** | dimethyl sulfoxide |
| **DSS** | dengue shock syndrome |
| **E** | envelope protein |
| **EDTA** | ethylenediamine-tetraacetic acid |
| **ER** | endoplasmatic reticulum |
| **G-rich** | Guanin-rich |
| **HCV** | hepatitis C virus |
| **ICAM3** | intercellular adhesion molecule 3 |
| **IDP** | intrinsically disordered protein |
| **Ig** | immunoglobulin |
| **IFN** | interferon |
| **IL** | interleukin |
| **IPTG** | isopropy-β-D-1-thiogalactopyranoside |
| **kDa** | kilo dalton |
| **LB** | lysogeny broth |
| **LDs** | lipid droplets |
| **M** | membrane protein |
| **MALDI-TOF MS** | matrix-assisted laser desorption/ionization, time-of-flight mass spectrum |
| **min** | minutes |
| **NMR** | nuclear magnetic resonance spectroscopy |
| **NS** | non-structural protein |
| **$OD_{600}$** | optical density at 600nm |
| **ORF** | open reading frame |
| **PDB (ID)** | protein data bank (identification), www.pdb.org |

| | |
|---|---|
| **pep14-23** | a rationally designed peptide with sequence H-NMLKRARNRV-$NH_2$ that was found to inhibit the DENV C-LDs interaction (International Patent Office, WO2012/159187) |
| **PLIN** | perilipin |
| **prM** | pre-membrane protein |
| **RT** | room temperature |
| **SDS-PAGE** | sodium dodecyl sulfate – polyacrylamide-based gel electrophoresis |
| **ssDNA** | single-strand DNA |
| **ss(+)RNA** | single-stranded positive-sense RNA |
| **SPMR** | signal per million reads |
| **TB** | Tris-Borate |
| **TNF** | tumor necrosis factor |
| **Trp** | tryptophane |
| **Tyr** | tyrosine |
| **UV** | ultra violet |
| **VLDL** | very low-density lipoprotein |
| **vRNA** | viral RNA |
| **WHO** | World Health Organization |
| **WNV** | West Nile virus |
| **YFW** | Yellow fever virus |
| **ZIKV** | Zika virus |
| | |
| **1R6R** | NMR-structure of dengue virus capsid protein, PDB ID |
| **2ZKO** | X-ray structure of NS1 protein of human influenza virus A, PDB ID |

# 1 State of the art

This chapter first details dengue epidemics and the global impact of the disease, as well as the basic aspects of dengue virus, namely its life cycle and virion structure. Then, the main object of study of this thesis, the dengue virus capsid protein that is involved in key steps of viral assembly and encapsidation, is presented. Finally, the working hypothesis and the main aim of this work are described.

Dengue is a serious viral infection, generating a considerable economic burden for healthcare systems in tropical and sub-tropical regions of the world. The disease causes a febrile illness, for which there is no antiviral treatment. Concerning dengue infection, there are four serotypes of dengue virus (DENV): DENV-1, DENV-2, DENV-3 and DENV-4. The cross infection of different serotypes is believed to increase the risk of dengue haemorrhagic fever.

## 1.1 Dengue disease – epidemics and the global impact

Dengue is a mosquito-borne viral infection, causing a severe illness with flu-like symptoms, sometimes leading to a potentially lethal complication called severe dengue. Severe dengue, previously known as dengue haemorrhagic fever (DHF), was first recognized in the 1950s during a dengue epidemics in the Philippines and Thailand[1]. Dengue is widespread in tropical areas, with risk factors influenced by local spatial variations of rainfall, temperature, humidity, degree of urbanization and quality of vector control services in urban areas. According to the World Health Organization (WHO), there are annually 50 – 100 million infections occurring in over 100 endemic countries, putting almost half of the world's population at risk, with recent estimates suggesting even higher numbers. Dengue has become a leading cause of hospitalization and death among children and adults in the most affected regions like Asia and Latin America.

The incidence of dengue has grown dramatically over the last decades. Exact numbers are unknown, since many cases, especially in less developed health systems, are unreported and/or are likely misclassified. Recent estimations indicate 390 million dengue infections per year, of which 96 million manifest clinically[1]. A study of prevalence of dengue estimates that 3.9 billion people in 128 countries are at a serious risk of infection by dengue virus[2]. Before 1970, only 9 countries had experienced severe dengue epidemics. Today, the WHO estimates that the disease is endemic in more than 100 countries in regions of Africa, the Americas, the Eastern Mediterranean, South-East Asia and the Western Pacific. Furthermore, explosive outbreaks of dengue still occur and the number of cases is increasing. The threat of a possible outbreak in Europe also exists, as local transmission was reported in France and Croatia in 2010. In 2012 there was a dengue outbreak on Madeira island, Portugal, which led to over 2000 cases in the island, with additional imported cases being detected in mainland Portugal[1]. Globally, a decline of 28 % in case fatality has been observed between 2010 and 2016 due to case management capacity building at country level[1]. Among travellers returning from low- and middle-income countries, dengue is the second most diagnosed cause of fever, after malaria.

## 1.2 Mode of transmission

Dengue is an arthropode-borne viral disease. The vectors are females of *Aedes* spp. mosquitoes[3,4]. The most effective vector is *Aedes aegypti*, because it feeds mostly of human blood. Thus, in dense urban areas where this mosquito is established, DENV life cycle can go on between human and vector organisms[5]. *Aedes albopictus* is less effective, but still a potential vector. It feeds not only of human blood but also of domestic and wild animals[6]. These mosquito vectors are endemic in many countries of Asia, the Americas and Africa, and have now spread to temperate, non-tropical regions, due to globalization of trade and travel, accompanied by climate change[7]. The mosquito salivary proteins released during the bite might promote dengue infection[8]. When a mosquito bites a DENV-infected human, the viral particles enter the mosquito dietary tract and infects its epithelial cells[9]. The replication of the virus takes place and consequently the salivary glands get infected after approximately 10 – 20 days[9]. When the infected mosquito bites another human, it injects its saliva, which is fully loaded with DENV infectious particles, directly into the bloodstream, causing a new human infection and closing the transmission cycle. In cities with a high population density and where mosquitoes live and thrive well, this process repeats frequently over time (if mosquito numbers are uncontrolled), resulting in the observed seasonal dengue outbreaks[10].

## 1.3 Dengue pathogenesis

Dengue infection in humans is often asymptomatic but, if not, it can originate a wide range of clinical manifestations. Following transmission through the mosquito bite, the virus infects first immature dendritic cells (DC) through the dendritic cell-specific intercellular adhesion molecule 3 (ICAM3)-grabbing non-integrin receptor[11,12]. After maturation, infected DC migrate to local or regional lymph nodes and present the viral antigens to T cells, which initiates the cellular and humoral responses[11]. Furthermore, there is evidence of DENV replication in parenchymal cells of the liver, in macrophages of lymph nodes, spleen, liver and in peripheral blood monocytes[13].
After infection, the virus circulates in the blood for 2 – 7 days before the person shows any symptoms. The clinical features of dengue fever can vary widely according to the patients' age and immunological status, besides other factors. Initial dengue infections are frequently unreported and/or undiagnosed, as symptoms can be very mild or absent. Most children show only mild unspecific febrile syndromes, while other persons develop severe flu-like symptoms with high fever, accompanied by severe headache, pain behind the eyes, nausea, vomiting swollen glands, muscle and joint pains and rash[14,15]. During secondary infection, the pathophysiology of dengue can change dramatically, particularly in sequential infections where an infection with DENV-1 is followed by infection with DENV-2 or DENV-3, or infection with DENV-3 followed by DENV-2[16,17]. Infections like this can result in potentially deadly complication with an acute vascular permeability syndrome, known as dengue shock syndrome (DSS). DSS can be

accompanied by plasma leaking, fluid accumulation, respiratory distress, severe bleeding or organ impairment[18]. The severity of DSS is also age-dependent, but vascular leakage is the most severe syndrome in young children. This phenomenon is thought to be related to the intrinsic integrity of the capillaries[19,20]. A primary infection in adults, independent of the DENV serotype, can result in dengue fever (DF) and is mostly accompanied by a tendency for bleeding, leading to severe haemorrhages[7]. There are some host factors that can either increase or reduce the risk of severe dengue disease. Host factors for increased risk include being of the female sex, specific human leukocyte antigen class I alleles, the occurrence of a single-nucleotide polymorphism in the tumour necrosis factor (TNF), the presence of a particular promoter variant of the dendritic cell-specific ICAM3-grabbing non-integrin receptor gene and having the AB blood group[21–24]. During secondary infection, host factors that reduce risk of a severe disease include ethnic background, second or third degree malnutrition, vitamin D receptor genes and polymorphisms in the Fc$\gamma$ receptor[25–27]. Also, the secondary dengue infection severity is higher, if a longer time interval between the first and the second infection ocurred[28].

### 1.3.1 Clinical signs and therapy

Severe forms of the disease can be successfully managed by careful monitoring patients' symptoms and the initiation of intravenous rehydration therapy at an early stage (if needed). Early febrile stages include symptoms like fever, headache, body pains and rash. Later, during defervescence (reduction of the fever), symptoms like bleeding, thrombocytopenia (<100,000 platelets/mm$^{-3}$), ascites or pleural effusion can occur. Besides that, severe and continuous abdominal pain, restlessness and somnolence, persistent vomiting and a sudden reduction in temperature associated with profuse perspiration can indicate plasma extravasation and the likely occurrence of dengue shock syndrome. To avoid haemodynamic instability and hypotension, patients must receive fluid replacements for rehydration. Treatment of uncomplicated dengue cases is only supportive to alleviate symptoms. This includes oral fluids and paracetamol intake during the febrile period.[7]

### 1.3.2 Immunological response

When DENV enters the human organism, the acquired immune response to the infection includes antibodies, that are primarily directed against the virus envelope proteins. Thus the response varies depending on if it is a primary or secondary infection[29,30]. Individuals, who are not immune to dengue show a primary antibody response, whereas patients with a previous dengue infection present a secondary immune response. A primary infection is defined by a slow and low-titre antibody response[7]. The first isotype to appear are immunoglobulin (Ig) M antibodies, with peak levels approximately two weeks after the onset of fever[31,32]. At the end of the first week of illness, dengue-specific IgG is detectable at low titre and slowly increases. A secondary infection is

characterized by high levels of IgG, even in the acute phase, and rising intensely over the following two weeks[32]. Compared to primary infection, IgM levels are significantly lower[33]. After dengue infection, IgG can remain lifelong circulating in the bloodstream, complicating the serodiagnosis of past, recent and current infections[32,34]. Immunological responses of IgA and IgE have been also documented, but the utility of detecting these immunoglobulins as markers for dengue serodiagnosis is not yet proven[35].

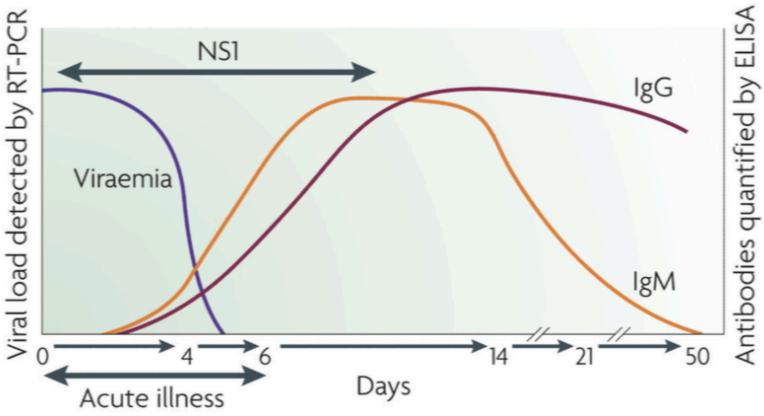


**Figure 1.1: Immunological response after primary dengue infection.** Viral load is elevated during acute illness but declining after the immune response starts. While the viraemia is decreasing, levels of IgM and IgG are rising. After acute dengue infection, IgG antibodies can remain in the bloodstream for a long period, possibly interfering with the diagnosis of new infections. Adapted from Guzman, 2010[7].

### 1.3.3 Treatment and vaccination

Until now, there is no specific treatment or effective therapy for dengue disease, but early detection and proper medical care can lower fatality rates below 1 %[1,36]. A working vaccine, Dengvaxia®, was registered in Mexico in December 2015, approved for official use in some endemic regions of Latin America and Asia and, by the end of October 2018, also in Europe[30-32]. However, this vaccine is not 100 % effective against all dengue serotypes. Thus, research into new prophylactics is still ongoing, with a new vaccine proposed, recently being in Phase 3 clinical trials[40]. Fully effective prophylactics and therapeutic approaches are lacking. This is partly due to a poor understanding of key steps in the viral life cycle, discussed ahead.

## 1.4 Dengue virus

Dengue virus is a member of the *Flavivirus* genus and part of the *Flaviviridae* family. This taxon includes other relevant human pathogen species such as West Nile (WNV), Zika (ZIKV), yellow fever (YFV) or hepatitis C (HCV) viruses[41,42]. Flavivirus are enveloped viruses with a single-strand positive sense RNA genome (ss(+)RNA). The ~10.8 kb genome of DENV contains one open reading frame (ORF), which is translated into a single polyprotein with the help of the host cell translation machinery[43]. The polyprotein is co- and post-translationally cleaved by viral and host proteases into ten proteins – three structural and seven non-structural (NS) proteins. The structural proteins have a critical role in the viral life cycle and are present in the mature virion structure[43,44]. They are encoded in the amino terminus of the genome and are namely the capsid protein (C), the pre-membrane protein (prM) and the envelope protein (E)[43]. The non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) are essential for the viral replication and are encoded in the remaining part of the genome. Each DENV shares around 65% of the genome. Despite some differences, each serotype causes nearly identical syndromes in humans and circulate in the same ecological niche[45]. Important biological properties of DENV include receptor binding, haemagglutination of erythrocytes and the induction of neutralizing antibodies and the protective immune response[7].

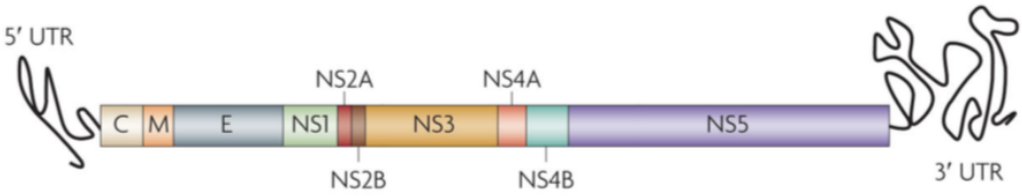


**Figure 1.2: The dengue virus genome.** The single open reading frame encodes the three structural proteins (C, M and E glycoproteins) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and N55). Adapted from Guzman, 2010[7].

### 1.4.1 The viral life cycle

The life cycle of *Flaviviridae* members can be subdivided into seven sequential steps. Virions attach to the surface of the host cell and then enter through receptor-mediated endocytosis. The acidic environment of the endosome triggers conformational changes in the virion, especially the E protein undergoes an irreversible trimerization, that results in fusion of the viral and cell membranes[46,47]. Subsequently after fusion, the ss(+)RNA is released into the cytosol, the polyprotein is translated and processed into the ten viral proteins[43]. Following, replication takes place: NS proteins transcribe molecules of ss(-)RNA to serve as a template for the synthesis of new ss(+)RNA copies[48]. The viral genome copies are conjugated with the C protein and

encapsidated in an assembly process mediated by unclear mechanisms, involving the structural proteins and the endoplasmatic reticulum (ER) membranous system[49,50]. This process results in infectious virions and subviral particles, of which are both formed by the assembly of prM and E proteins and are believed to undergo the same maturation process[51]. The immature particles are transported through the trans-Golgi network, where they are cleaved by the host protease furin, resulting in infectious and mature virions, that are released from the host cell by exocytosis[52,53]. The new built virions are now prone to infect other cells. Flavivirus use both host and viral proteins to complete their life cycle, as well as the host cellular structures. This brief overview of the process includes the main steps, which are important for the understanding of this work. In vitro studies show an increased virus replication when human monocytes and mature dendritic cells got infected, due to the suppression of the interferon (IFN) host protection system[54]. Also, Type I IFN-associated genes are less abundantly activated in peripheral blood mononuclear cells taken from patients with severe dengue cases compared with milder disease patients[55]. Consequently, the increased number of infected cells present targets for CD4+ and CD8+ T cells, which results in large quantities of interleukin (IL)-10, IL-2, IFN$\gamma$ and TNF, which might contribute to endothelial damage and changes in the hemostasis[7]. Damage of endothelial cells might be directly from virions released from infected cells[56]. Furthermore, the uptake of dengue virus non-structural protein NS1 by hepatocytes might promote the viral infection of the liver[57].

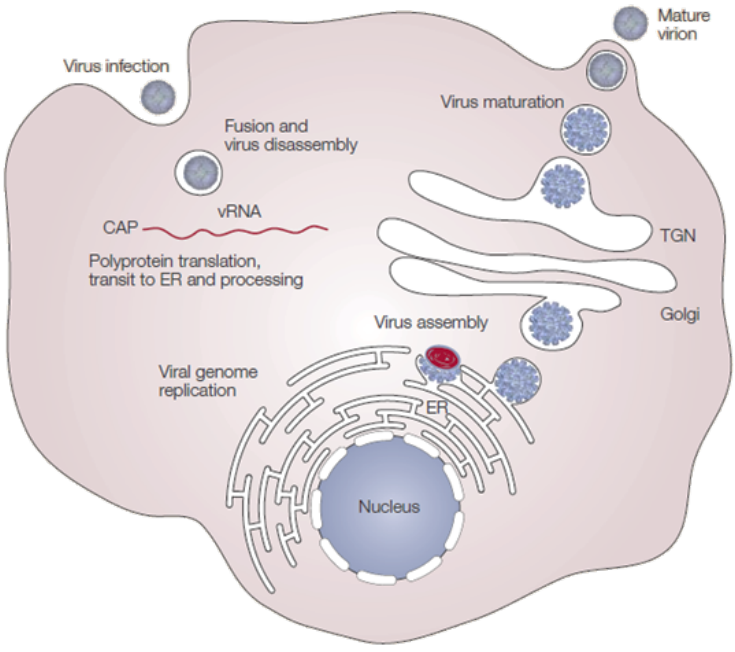


**Figure 1.3: The *Flaviviridae* life cycle.** The virus enters the host cell and the genome is released after the fusion of the viral and endocytic vesicle membranes. The polyprotein is translated and processed in the ER membranes. Later on, non-structural proteins promote the replication, while the structural proteins assemble with RNA. Host proteases are responsible for the maturation of the virion, following the secretory pathway, until the virus is released to the extracellular space. Mukhopadhyay *et al.*, 2005[43].

### 1.4.2 The viral structure

The Flavivirus structure has been solved by cryo-electron microscopy (cryo EM) for DENV, YFV and WNV[44,58–60]. This technique made it possible to identify the structure of the viral particles in a high resolution at molecular level. The virion presents a 40-50 nm diameter[7]. It displays an internal nucleocapsid, enveloped by a lipopolysaccharide envelope, the lipid bilayer[44]. The outer layer is composed by the envelope (E) and membrane (M) proteins. DENV E protein is arranged in anti-parallel dimers, showing an icosahedral symmetry[44]. During the Flavivirus life cycle, the protein is proposed to promote the fusion between viral and host cell membranes[43]. The prM protein is also embedded into the lipid bilayer, exposing the prM domain in the immature virion to protect the E proteins fusion region. Later, the prM domain is cleaved by the cellular protease furin to promote maturation[60,61]. The C protein is present in the nucleocapsid, where it is complexed with viral RNA.

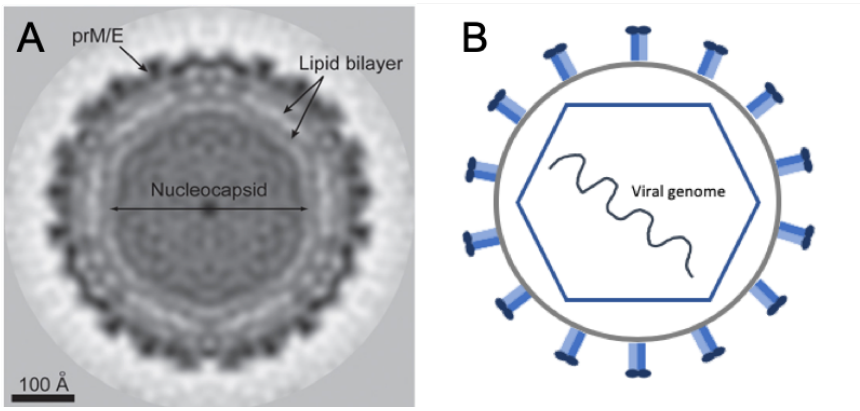


**Figure 1.4: The viral structure. A**) Cryo-EM image of the DENV virion, showing the internal structure. The nucleocapsid is composed by viral RNA, conjugated with C protein copies. Yu *et al.*, 2008[58]. **B**) Schematic drawing of DENV. The internal capsid with the viral genome is enveloped by a lipid bilayer. The outer layer is composed by the membrane proteins, covered by envelope proteins[62].

## 1.5 The capsid protein

The capsid protein of DENV is a highly basic protein, composed of 100 amino acids, that form a homodimer with an intrinsically disordered protein (IDP) region in the N-terminal, followed by 4 α-helices (α1 - α4) per monomer[63]. If not otherwise indicated, DENV refers to DENV-2 in the following sections. The main structural region is composed by the disordered N-terminal, a short flexible intermediate fold and a large conserved fold region, which stabilizes the proteins homodimer structure[64–68]. The dimer is presenting an asymmetric charge distribution, due to the antiparallel helical interfaces of α2-α2' and α4-α4'. One side presents a hydrophobic pocket (α2-α2'), responsible for the binding to host lipid droplets (LDs), while the C-terminal side (α4-α4') is positively charged and proposed to mediate the C protein binding to the viral RNA[63]. In addition,

the C protein has not only a structural role. It also provides a protection to the viral genome during transition between environments, shielding the RNA against exposure to chemical hazards[69,70]. The interactions of the C protein with host lipid systems and viral RNA are important for viral replication[71,72]. A key function of DENV C is genome encapsidation during viral assembly. However, the exact mode of action of the C protein is still unclear.

DENV infection affects the host lipid metabolism, increasing host intracellular LDs and unbalancing plasma lipoprotein levels and composition[71,73,74]. DENV C-LDs binding is essential for viral replication, requiring $K^+$, the LDs surface protein perilipin (PLIN3) and specific amino acid residues of DENV C α2-α2' helical hydrophobic core and of the N-terminal[65,75]. With the help of that knowledge, pep14-23, a peptide based on a flavivirus C protein conserved N-terminal motif, was designed and patented by the host lab[76]. Furthermore, it was established that pep14-23 inhibits DENV C-LDs binding, acquiring α-helical structure in the presence of anionic phospholipids[65,66]. It was shown, that DENV C binds specifically to very low-density lipoproteins (VLDL), requiring $K^+$ and apolipoprotein E (APOE), a specific VLDL surface protein, being also inhibited by pep14-23[77]. This analogous interaction and similarities between APOE and PLIN3 in the context of DENV C binding reinforces the suggestion of a common binding mechanism[78]. The important role of LDs in flavivirus infection is well known and has been recently reviewed[71,79,80]. However, further research seeks to contribute to a better understanding of the biological function of the C protein of flaviviruses in general, providing ground for future drug development.

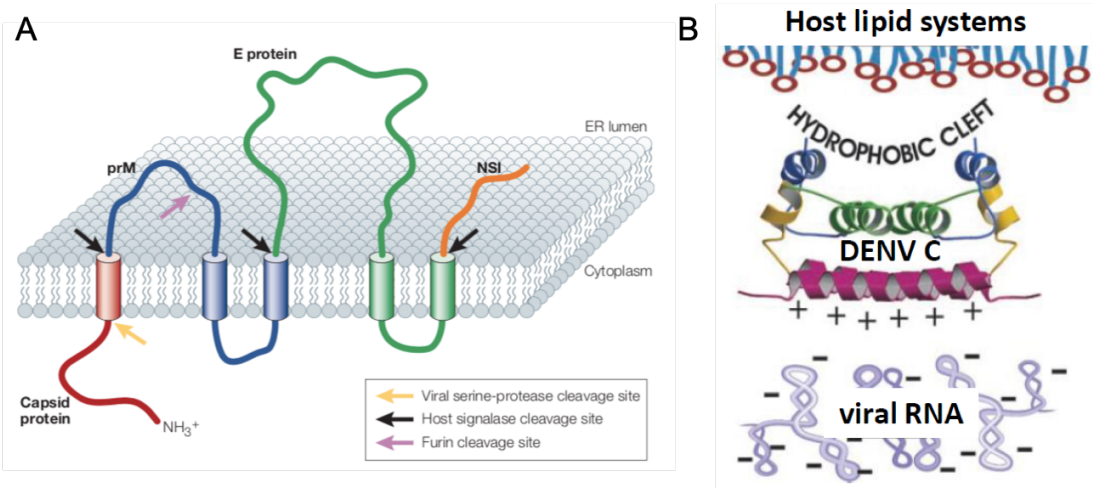


**Figure 1.5: Capsid protein membrane topology and structure. A**) The viral genome encodes a polyprotein that is cleaved into ten viral proteins. The capsid protein is the first protein of the polyprotein. the N-terminal end **B**) DENV C protein is an α-helical homodimer with an asymmetric charge distribution. The disordered N-terminal (not shown) and the hydrophobic cleft, specially the α2-α2' interface (green), are involved in the interaction with host lipid droplets. The positively charged C-terminal α4-α4' interface (purple) has been proposed to interact with viral RNA. Mukhopadhyay *et al.*, 2005[43], Martins *et al.*, 2012[65] and adapted from Ma *et al.*, 2004[63].

## 1.6 Objective of the work

Studies at the host laboratory of DENV C protein interaction with host lipoproteins and intracellular LDs led to the design of pep14-23, a peptide that inhibits DENV C binding to these host lipid systems. Now, to complement the understanding of DENV C biological activity, studies on DENV C binding to nucleic acids should be conducted. DENV C structure suggests that its α4-α4' region may bind the viral RNA[63]. It was recently shown that ZIKV C protein possesses a broad binding capability to different nucleotide types, including single-stranded (ss) and double-stranded (ds) RNA or DNA[81]. The complete understanding of DENV C interactions can give rise to a model of flavivirus C protein function, providing ground for future development of DENV C-targeted treatments. Thus, this work aims to understand molecular details of DENV C interaction with nucleic acids, especially ssDNA. To achieve this, a combined *in vitro* and *in silico* study of the C protein-DNA interaction was performed. First, recombinant DENV C protein was produced. Second, adequate nucleotide sequences were selected to test the interaction with DENV C. After an intensive optimization of the experimental workflow, DENV C binding to nucleic acids was tested via biophysical approaches, namely circular dichroism spectroscopy and fluorescence spectroscopy. To complement the molecular information gathered, a computer-aided analysis was performed to compare the structure of DENV C with the structure of other RNA-binding viral proteins. Finally, a model describing the interaction is proposed.

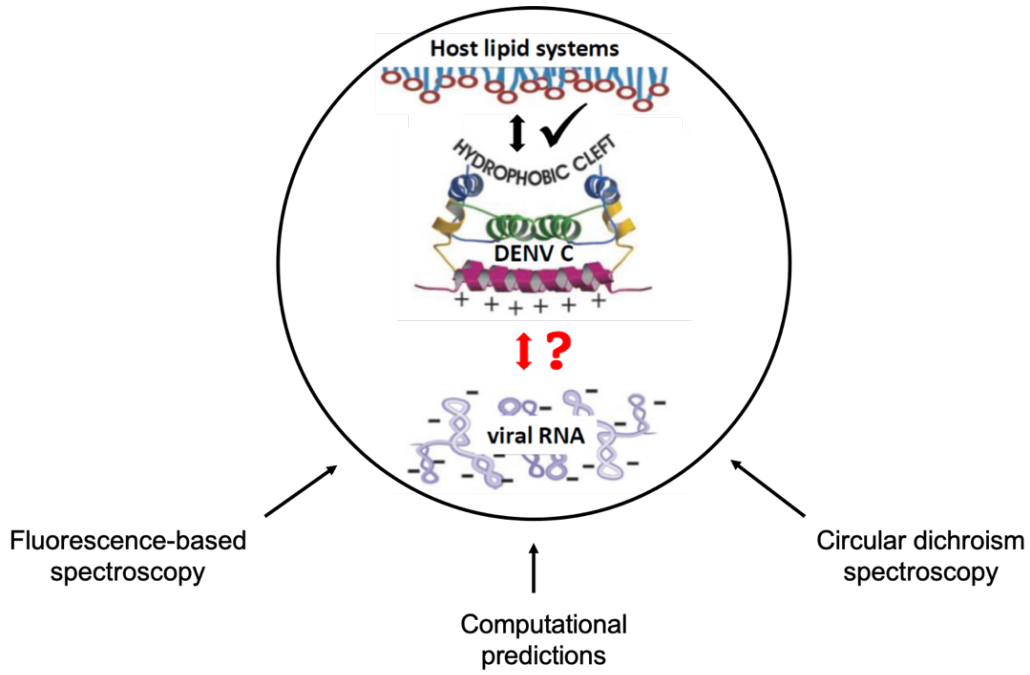


**Figure 1.6: Study of dengue virus capsid protein interaction with nucleic acids.** It was hypothesized that DENV C interacts with the viral RNA through its positively charged face of the dimer. Biophysical and computational techniques were used to access this interaction. Adapted from *Ma et al.*, 2004[63].

# 2 Methodology

Experimental procedures and sources of materials are presented in this chapter. In brief, the protocol for protein expression and purification is detailed. Then, biophysical techniques employed to analyze the interaction of DENV C with nucleic acids are described.

## 2.1 Materials

Recombinant DENV C protein was expressed in *Escherichia coli* (*E. coli*) cells, transformed with the DENV C serotype 2 C protein gene (encoding residues 1 to 100), cloned into the plasmid pET21a, containing resistance for ampicillin. Competent cells were treated in Tris-Borate (TB) buffer, composed by 10 mM HEPES, pH 6.7, 250 mM KCl, 15 mM $CaCl_2$, 55 mM $MnCl_2$. During DENV C expression, a NaCl buffer was used, consisting of 25 mM HEPES, pH 7.4, 1 mM EDTA, 5% Glycerol and 0.2 M / 2 M NaCl. The buffer used for the purification of DENV C consisted of 550 mM KCl and 55 mM $KH_2PO_4$, pH 6.0. For circular dichroism measurements, different buffers were used. The first experiments were done in 10 mM $KH_2PO_4$, pH 6.0 and 100 mM KCl and, while the experiments later on were performed in 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl or 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl and 20 mM $MgO_4S \bullet 7\ H_2O^-$. Buffers were filtered, using a 0.1-μm filter. Unless stated otherwise, the chemicals were purchased from Merck (Darmstadt, Germany). DNA oligonucleotides were obtained from Eurofins Genomics (Ebersberg, Germany), supplied by NZYTech, Lda. (Lisbon, Portugal).

## 2.2 Production of recombinant DENV C protein

The expression and purification of recombinant DENV C protein was based on previous work[72]. The full protocol is detailed below.

### 2.2.1 Preparation of competent *E. coli* cells

A colony of *E. coli* strain C41, previously grown in solid media of lysogeny broth (LB) agar, was inoculated into 3 mL of liquid LB medium, overnight, at 37 °C and agitation at 200 rpm. Then, 500 μL of the overnight culture were added to 10 mL LB liquid medium (1:50 dilution) and incubated at 30 °C, until the optical density at 600 nm reached approximately 0.5. After this, the culture was incubated on ice for 10 min. Following, the cell suspension was centrifuged at 4 °C for 10 min, 4600 g (Heraeus Multifuge 1 L-R, Thermo Fisher Scientific, Waltham, USA). After discarding the supernatant, the pellet was gently resuspended in 3 mL ice-cold TB buffer. The culture was incubated on ice for 20 min and then again centrifuged at 4 °C, for 7 min, 4600 g (Heraeus Multifuge 1 L-R, Thermo Fisher Scientific, Waltham, USA). The pellet was resuspended in 930 μL ice-cold TB buffer and 70 μL dimethyl sulfoxide (DMSO), followed by an incubation on ice for 10 min. The cell suspension was aliquoted (200 μL per tube), before shock-freezing in liquid nitrogen and storing at -80 °C, for later use.

### 2.2.2 Transformation of competent *E. coli* cells

A 200 μL aliquot of competent *E. coli*, strain C41 was thawed at room temperature (RT). Then, 1 μL of pET21a-DENV C 1-100 plasmid (containing DENV C gene sequence) was added to 100 μL of competent cells in a tube. The remaining 100 μL of competent cells were used as a negative control, adding only water (no plasmid). Both tubes were incubated on ice for 30 min. After this, both samples were heat shocked for 45 sec at 42 °C and immediately placed on ice for 2 min. Next, 900 μL of LB medium was added to both tubes. The samples were incubated for 1 h at 37 °C and 220 rpm. After, 100 μL of these cells were plated onto LB agar plates with 100 μg/mL of the antibiotic ampicillin (to select for the competent cells that are transformed with the plasmid). Plates were incubated overnight at 37 °C and stored at 4 °C after successful transformation. Competent cells are also plated, in LB medium without antibiotic (diluted 1:10.000) in order to know if competent cells are viable. If no growth of transformed cells occurs and the non-transformed cells do grow, the problem would be due to the plasmid or the transformation assay, but not due to the competent cells viability.

### 2.2.3 DENV C expression and purification

The protein was expressed in *E. coli* strain C41 cells transformed with the DENV serotype 2 C protein gene (encoding residues 1 to 100) cloned into plasmid pET21a. The cells were grown overnight at 37 °C and 220 rpm in 6 mL LB medium in the presence of 100 μg/mL ampicillin. 4 mL of the culture grown overnight was transferred into a freshly prepared LB medium in the presence of the same antibiotics. The protein expression was induced at an optical density at 600 nm ($OD_{600}$) of 0.8 with 125 μL of 1 mM isopropy-β-D-1-thiogalactopyranoside (IPTG), being grown overnight at RT and 220 rpm. Next, the cell culture was transferred to 50 mL tubes and centrifuged at 6100 g (Heraeus Multifuge 1 L-R, Thermo Fisher Scientific, Waltham, USA) for 30 min at 4 °C, the supernatant was discarded. The pellets were resuspended in 20 mL of NaCl buffer and 10 μM protease inhibitor mix (c0mplete, EDTA-free, Roche Diagnostics, Mannheim, Germany). Following, the cells were lysed by 5 cycles of sonication on ice, 2 min turned on at the maximum power of the apparatus (Sonicator Soniprep 150, MSE, East Sussex, UK) and 2 min turned off, cooling in ice. DENV C protein was eluted with an increasing NaCl concentration gradient. To reach a concentration of 2 M, NaCl was added to the 20 mL cell solution, and then incubated for 1 h on ice, while continuously stirred. The cell lysate was centrifuged at 6100 g (Heraeus Multifuge 1 L-R, Thermo Fisher Scientific) for 30 min at 4 °C. The supernatant was again centrifuged for 15 min at 16100 g at 4 °C. Next, 60 mL of 0.2 M NaCl were added to the supernatant, to get a final salt concentration of 0.5 M NaCl. Soluble protein extracts containing C protein were subjected to cation-exchange chromatography, followed by size-exclusion chromatography. The 1.5 mL fractions were confirmed to contain DENV C protein via 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and mass spectrometry analysis.

### 2.2.4 Quality control

To test the quality of the expressed DENV C protein, first, a 15% SDS-PAGE was performed, alongside evaluating the protein secondary structure via circular dichroism spectroscopy and mass spectrometry analysis (please consult the next chapter, sections 3.1 and 3.2, for more details).

## 2.3 Interaction of DENV C with nucleic acids

DENV C interaction with nucleic acids (ssDNA) was analyzed by biophysical approaches, namely circular dichroism and fluorescence spectroscopy, via a titration experiment. Briefly, 1 μM of DNA was dissolved in the appropriate buffer (detailed ahead) and DENV C protein was titrated until reaching 2 μM (monomer concentration), via successive 5.4 μL additions of protein from at a stock concentration of 180 μM, in a final volume of 2227 μL.

### 2.3.1 Circular dichroism spectroscopy

Circular dichroism (CD) is a spectroscopic method that is based on the differential absorption of left- and right-circularly polarised light by optically active molecules[82]. Proteins are optically very active compounds that, due to their peptide bonds, have the property of chirality. Two isomers are chiral if they have the same chemical composition, functional groups and similar physical properties, only differing in their spatial orientation, being the exact mirror counterpart, but not superimposing with each other. CD measures the difference in the absorbance between left- and right-circularly polarised light, reporting this difference in terms of ellipticity (Θ) in degrees[83]. Briefly, the fraction of light that is absorbed by a sample follows the Beer-Lambert law:

$$A_\lambda = log_{10}\left(\frac{I_0}{I}\right) = \varepsilon l C \qquad (1)$$

where $A_\lambda$ is absorbance at a given wavelength, $I_0$ the intensity of light entering the sample, $I$ the intensity of light leaving the sample. If pure protein samples are used, the signal is proportional to the concentration of protein in the sample, $C$ (that can be expressed in M), the pathlength, $l$, (that can be expressed in cm) and to $\varepsilon$, which is the molar extinction coefficient of a given protein at a given wavelength (that can be expressed in M.cm)$^{-1}$. In CD spectroscopy, this also applies for right- and left-circularly polarized light, so CD is the difference in absorption, such that:

$$CD\ signal = \Delta A = A_L - A_R = \varepsilon_L l C - \varepsilon_R l C = \Delta\varepsilon l C \qquad (2)$$

where the direction of rotation of light is denoted by the subscripts. For historical reasons, CD instruments display the data from the CD signal in ellipticity ($\theta$) units (mdeg), converted from the difference in absorbance via the equation below.

$$[\theta] = 3298.2\,\Delta\varepsilon \qquad (3)$$

To make comparison simpler and take sample concentration and pathlength into account, CD spectra are presented in molar ellipticity ($\theta$) units (degrees-cm$^2$ dmol$^{-1}$), defined as:

$$[\Theta] = 100\theta Cl \qquad (4)$$

where *C* is the molar concentration and *l* is the pathlength in cm.[82]

CD spectra of α-helical proteins are defined by three peaks: two negative peaks at, roughly, 222 nm and 208 nm and a stronger positive peak at 192 nm[82]. Spectra of β-sheets are characterized by a negative peak at 217 nm and a positive peak at 195 nm, with only half the intensity of the α-helix peak in this area[84]. Random coil or unordered structures often show spectra similar to the polyproline helix spectrum, which is characterized by a small peak around 215-230 nm and a negative peak at 200 nm[82,85]. Changes in CD spectra provide information on interactions of proteins and ligands and can monitor protein folding and unfolding. Furthermore, they can be used to study protein stability[82].

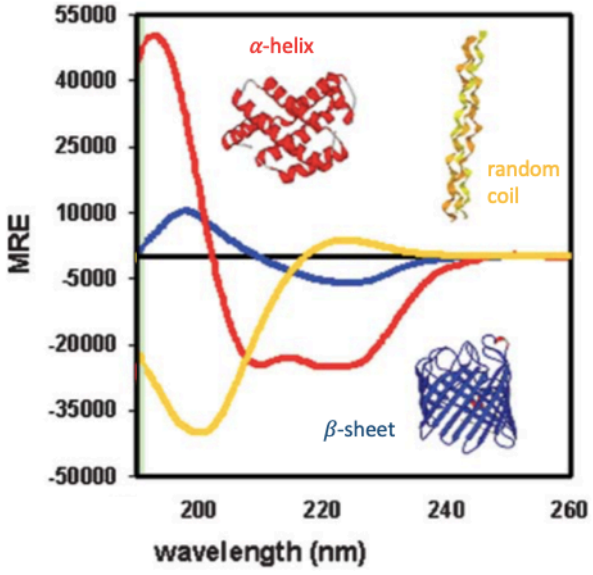


**Figure 2.1: Circular dichroism spectra of proteins.** CD spectroscopy is based on light absorption. When absorption occurs in chiral chromophores, it gives characteristic CD spectra. If a molecule is not chiral or absorbs equally the polarized light, there is no CD signal. Different protein secondary structures have each their own typical CD spectrum: α-helix, red spectrum; β-sheet, blue spectrum and random coil, yellow spectrum. Adapted from Miles *et al.*, 2006[82].

CD measurements were carried out in a JASCO J-815 CD spectrometer (JASCO, Tokyo, Japan), using 0.1 or 1 cm path length quartz cuvettes with 220 μL or 2200 μL of total volume, data pitch of 0.5 nm, velocity of 200 nm/min with a data integration time of 2 sec and performing 3 accumulations with an automatic baseline (buffer spectra) correction. Spectra were acquired in the far UV region, between 200 and 260 nm, with 1.0 nm bandwidth. Temperature was controlled by a JASCO PTC-423S/15 Peltier equipment, at 25 °C. C protein concentration ranged from 1 μM to 20 μM (monomer) in different buffer conditions. A series of optimization was performed to obtain clear spectra of the interaction of DENV C with ssDNA. Finally, spectra were smoothed through the means-movement method (using 5 points) and normalized to mean residue molar ellipticity, [θ] (in deg $cm^2$ $dmol^{-1}$ $Res^{-1}$).

### 2.3.2 Fluorescence Spectroscopy

Fluorescence spectroscopy is a crucial technique in biochemical research, due to its robustness, high sensitivity and non-invasiveness[86]. Fluorophores can absorb light of a specific wavelength ($\lambda_{EX}$) and emit the energy after a brief interval, namely the fluorescence lifetime ($\tau$), at a longer and specific wavelength ($\lambda_{EM}$)[86]. The intrinsic fluorescence of proteins, originates mostly from the aromatic residues of amino acids, mainly tryptophan. This property has been explored to study protein dynamics and conformational changes, gathering information from fluorescence intensity, absorption and emission maximum, band shape, anisotropy and fluorescence lifetimes of the residues in native proteins[87]. A molecule, that has been excited with, for example, ultraviolet (UV) / visible light, can loose the excess energy, as it returns to its ground state – in fluorescence this is achieved by the emission of light[88]. The fluorescence quantum yield is described as the ratio of the number of photons of light emitted as fluorescence to the number of photons initially absorbed[89]. Studying protein-nucleic acid interactions is simplified by the fact that all detectable fluorescence arises from the protein, since naturally occurring bases in RNA and DNA are essentially non-fluorescent[88]. Almost all fluorescence found in proteins is ascribed to the tyrosine (Tyr) and, above all, tryptophan (Trp) residues. When both amino acid residues are present, the emission spectrum will be dominated by Trp, unless the ratio of Tyr to Trp is very high[89]. Due to this, fluorescence spectroscopy can be used as a sensitive non-destructive technique for protein-DNA interactions studies. The intrinsic emission spectra obtained for a free protein and a protein-DNA complex can be compared and give information about the environment of Trp and Tyr residues in these two states. Quenching of the fluorescence intensity of an intrinsic emission spectrum and/or a shift in the wavelength maximum often occurs upon protein-DNA binding. Changes in the fluorescence emission can be used to determine the stoichiometry of binding and equilibrium-binding constants, in some cases the data can also give an indication of the location of particular residues within the protein[88]. Besides the fact that the experiments are quick and easy to perform[88], the technique is very sensitive, allowing low concentrations (typically in μM range), required for estimation of binding constants for many protein-DNA interactions[89].

Concerning tryptophan fluorescence, it is widely used as a tool to monitor changes in proteins[91]. Trp has the strongest fluorescence quantum yield of natural amino acids[92]. For this reason, intrinsic protein fluorescence normally refers to the fluorescence emission of Trp, as in the case of DENV C, where only Trp is present (no Tyr). The indole group of Trp is the dominant source of UV absorbance at excitation ~280 nm ($\lambda_{EX}$) and emission ~350 nm ($\lambda_{EM}$) in proteins[87]. The photophysical properties of Trp are highly sensitive and influenced by its local environment[93,94]. The Trp fluorescence maximum ($\lambda_{EM}$) and intensity are highly influenced by the polarity of its micro-environment, hydrogen bonding and other non-covalent interactions[87,95].

Due to this sensitivity to the polarity of the surrounding solvent, $\lambda_{EM}$ for Trp can range from 330 nm in a hydrophobic environment to 355 nm in water[88,96]. Protein fluorescence is quite weak and can be non-existent, depending not only on the environment, but also upon protein structure. The emission of Trp can be quenched by other amino acids in adjacent positions in the protein, by the solvent, ligand binding or by impurities[96].

The experiments of intrinsic protein fluorescence studies described here were performed in a VARIAN CARY Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, USA), using a 5 mm cuvette. Data were collected in the fluorescence emission mode. Samples excitation was performed at 280 nm, measurements were collected between 300 and 450 nm, excitation slit of 10 nm, emission slit at 20 nm with a slow scan control and a medium detector voltage.

## 2.4 Protein structure visualization and comparison

Protein structures coordinates were extracted from the Protein Data Bank (PDB). PDB identification codes are specified ahead after each protein name. PDB files of DENV C (1R6R)[63] and influenza NS1 (2ZKO)[97] were used as templates for comparisons. Protein structures were superimposed through UCSF Chimera 1.13.1 software MatchMaker tool. Then, we carefully analyzed the superposition visually. Protein structure figures were obtained using UCSF Chimera 1.13.1.

# 3 Results

The results of the experimental work are shown in this chapter. First, the expression of the purified DENV C protein is evaluated, alongside the characterization of the recombinant protein. Next, the optimization of the protocol for the measurement of DENV C interaction with nucleic acids at low concentrations is illustrated. The process of selection of the nucleic acid sequences is also described, before showing the outcome of the biophysical experimental evaluation of the interaction of DENV C with nucleic acids.

In this thesis a combination of *in vitro* and *in silico* studies was used to gain insights into DENV C biological activity and structure, in the context of its interactions with nucleic acids. To achieve these aims, it was first necessary to express and purify the DENV C protein and, afterwards, to characterize it. Following, after optimizing parameters and conditions for CD measurements at low protein concentrations (necessary for these goals), experiments were performed to test DENV C binding to nucleic acids. Sequences of ssDNA were used as proxies of relevant viral ssRNA genome regions, with the rational for selecting them explained below.

## 3.1 Expression and evaluation of the purification of DENV C

Recombinant DENV C protein was expressed and purified in the host laboratory to study the interaction with nucleic acids, following previous work[72]. Briefly, a pET-21a plasmid was used, containing DENV serotype 2 capsid protein gene. Two independent recombinant protein expression and purification assays were concluded. To confirm the success of the process, an intermediate quality control step was performed, via 15% sodium dodecyl sulfate – polyacrylamide-based gel electrophoresis (SDS-PAGE) (*Figure 3.1)*. IPTG induced DENV C expression in *E. coli* C41 cells, when compared to the basal expression, as expected. Some protein is found in the pellet of sonicated cells (*Figure 3.1 A, column 5*). That was unexpected, as, in previous work, DENV C was mostly all in the supernatant of the cells. The supernatant of sonicated cells was clear and, as most of the protein was found there in previous work by others, the supernatant was used in the ensuing purification chromatography steps. DENV C is about 11.5 kDa (monomer size). A band roughly corresponding to it is visible in the SDS-PAGE of size exclusion chromatography (described ahead), in the expected region (*Figure 3.1 B, column 3)*.

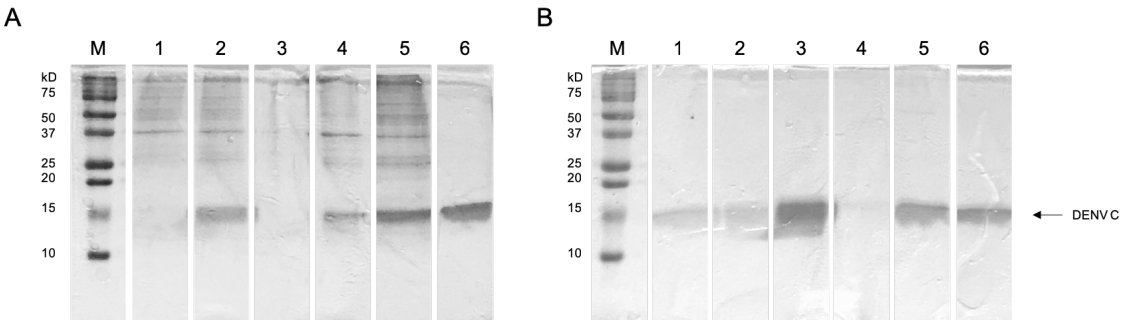


**Figure 3.1: SDS-PAGE (15%) of recombinant protein production in *E. coli* C41 cells. A) Expression steps.** (**M**) Protein marker. (**1**) Overnight basal protein expression (not induced). (**2**) Cells induced overnight with IPTG. (**3**) Supernatant of induced cells. (**4**) Pellet of induced cells. (**5**) Pellet of induced cells after being sonicated. (**6**) Recombinant DENV C previously purified (positive control). **B) Size exclusion chromatography fractions. (M)** Protein marker. **(1)** Fraction 3. **(2)** Fraction 4, resuspended pellet. **(3)** Fraction 4. **(4)** Fraction 5, resuspended pellet. **(5)** Fraction 5. **(6)** Recombinant DENV C from other purifications (positive control).

Purification of DENV C was first performed by affinity chromatography, using a heparin column. The protein is highly positively charged and, as heparin is negatively charged, it binds to it, eluting later on when high salt concentrations are passed through the column. Fractions corresponding to the elution profile of the protein were collected and, then, submitted to size exclusion chromatography. A S75 column was used for this final purification step. Fractions A4 and A5 were collected (from here onwards referred to as F4 and F5, respectively). Aliquots of each fraction were saved to perform a gel electrophoresis (*Appendix I, SF 1).* A band of 15 kDa is seen, in line with previous work at the host lab and collaborators, as DENV C protein gives a slightly larger size than expected in SDS-PAGE.

A second DENV C protein expression was done in exactly the same manner. A final gel electrophoresis was performed, to compare the result of the two purifications (*Figure 3.2)*. It can be seen that DENV C from the first purification batch had a higher concentration. This is due to the fact that only half of the volume after protein expression was used in the second purification. The reason for using less amount of protein was to avoid any protein precipitation at high concentrations, as it had occurred in the past with this and similar capsid proteins.

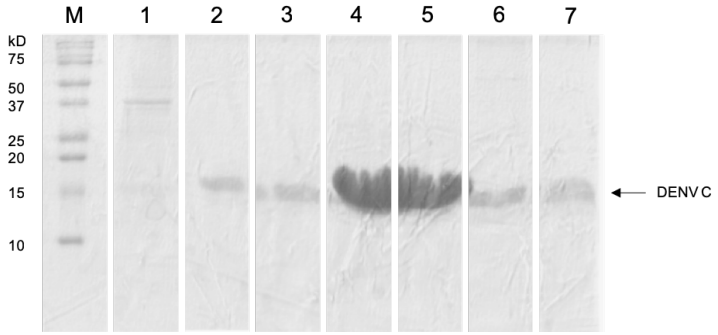


**Figure 3.2: SDS-PAGE (15%) comparing two recombinant DENV C protein production assays.** (**M**) Protein marker. (**1**) Second purification, after sonication. (**2**) Second purification, size exclusion chromatography, fraction 5. (**3**) Second purification, size exclusion chromatography, fraction 6. (**4**) First purification, size exclusion chromatography, fraction 4. (**5**) First purification, size exclusion chromatography, fraction 4 as a duplicate. (**6**) First purification, size exclusion chromatography, fraction 5. (**7**) Recombinant DENV C from other purifications, used as a control.

## 3.2 Characterization of purified DENV C

Protein samples quality was further assessed via a commercial analysis mass spectrometry service at i3S (*Instituto de Investigação e Inovação da Universidade do Porto*), Porto, Portugal, using a MALDI-TOF MS (matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry) apparatus, via LC-MS Orbitrap (separation by a C4 column) and mass spectra acquisition in the positive mode. Fractions F4 and F5 from the first purification and fractions F5 and F6 from the second purification were tested, as shown in Appendix I, SI 2 – 5.

The highest peak is consistent with the expected mass of the protein monomer (11.5 kDa). Fraction 4 of the first purification and the fractions 5 of the first and second purifications show this exact result. Fraction 6 from the second purification shows a different result and it was not used for the *in vitro* experiments. Then, to further test the recombinant DENV C protein obtained, a quality control of secondary structure content was performed. Briefly, a CD spectrum of DENV C protein at 50 μM, 25° C was acquired (*Figure 3.3*). The spectrum of fraction 4 of the first purification shows two minimal peaks at 208 nm and 222 nm, characteristic of an α-helical secondary structure, as expected.

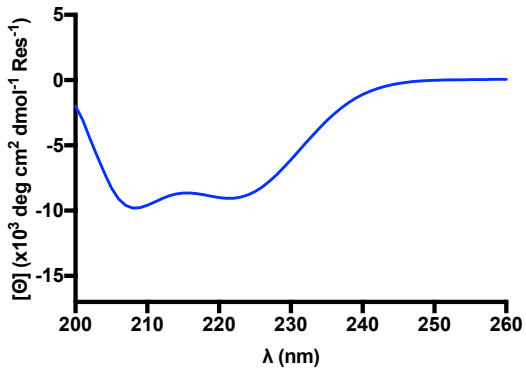


**Figure 3.3: CD spectrum of recombinant DENV C protein.** Fraction 4 of the first purification, at a concentration of 50 μM (protein monomer). in 55 mM $KH_2PO_4$, pH 6.0, 550 mM KCl. A typical spectrum for an α-helical secondary structure is seen, with minimal peaks at 208 and 222 nm.

## 3.3 Optimization of CD measurements at very low DENV C concentrations

The goal here was to obtain purified DENV C and establish conditions to test its interaction with nucleic acids. For such purpose, CD spectroscopy was used. Given the low availability of nucleic acids and the risk of protein aggregation upon interaction with them, there is a need to work at low protein concentrations. Thus, it is necessary to optimize conditions to measure an adequate CD spectrum of DENV C at very low protein concentration. Otherwise, at high concentrations, protein aggregation promoted by the nucleic acids would impair the biophysical analysis (or at least make it very difficult).

First, CD spectra of DENV C at a concentration of 1 μM (monomer) dissolved in water was acquired (*Figure 3.4 A*). Pure water is not an adequate solvent (as expected), since the protein shows higher random coil (which could eventually, however, facilitate the binding to ssDNA). Moreover, this makes very clear that buffer conditions must be carefully controlled. Unlike some globular proteins, DENV C (that has a crucial IDP section) seems to be very easily affected by the buffer environment. Next, a buffer with 10 mM $KH_2PO_4$, pH 6.0, 100 mM KCl was tested. This composition was chosen because it is compatible with further analysis of DENV C-ssDNA interactions, via, for example, nuclear magnetic resonance (NMR). Besides, pH 6.0 is still a physiological pH and was employed in other experiments with this protein[64]. However, it was not possible to acquire a stable CD spectrum at low DENV C concentrations using that buffer (*Figure 3.4 B*).

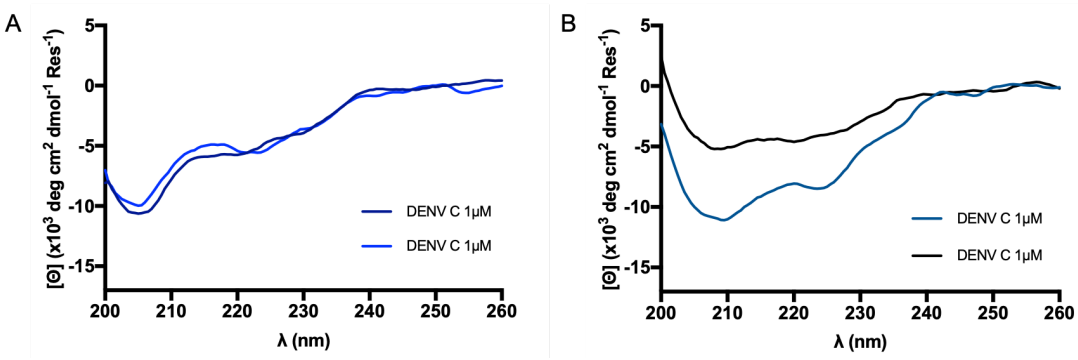


**Figure 3.4: CD spectra of DENV C in different solutions in a 0.1 cm cuvette. A**) DENV C in water. **B**) DENV C in buffer 10 mM $KH_2PO_4$, pH 6.0, 100 mM KCl. Both water and this buffer are inadequate, as both give rise to some level of random coil secondary structure (which is not the typical DENV C CD spectra). Moreover, three experiments, in different days, were performed, with poor reproducibility and the loss of α-helical structure, when the above buffer was used.

Then, DENV C was tested in 10 mM $KH_2PO_4$, pH 6.0, 100 mM KCl buffer, at 2 μM (*Figure 3.5*) and 10 μM (*Figure 3.5 B*). Even at this higher protein concentration, it was not possible to acquire stable DENV C spectra. This is clear from measurements of the same sample giving different spectra through time (*Figure 3.5 A*), even when the sample remains in the quartz cuvette

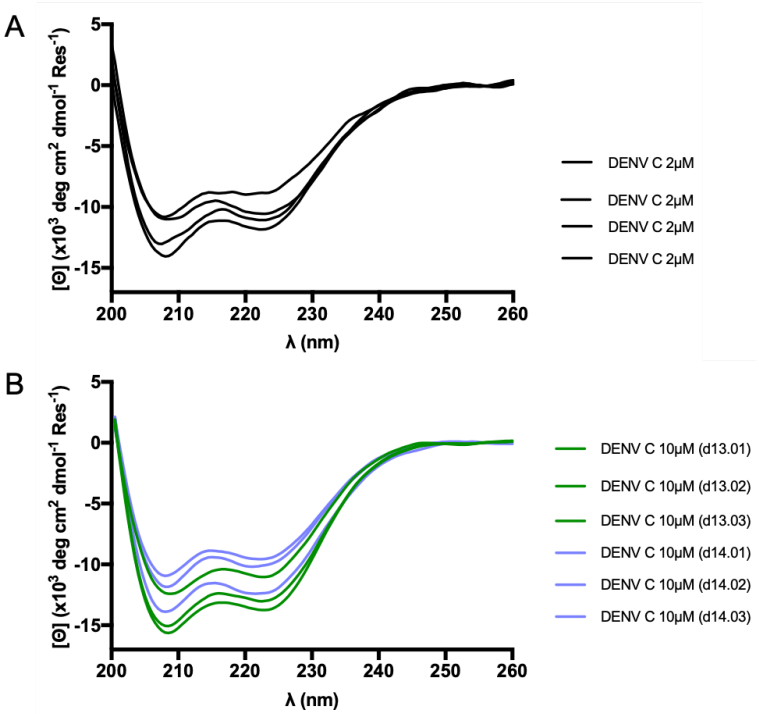


**Figure 3.5: CD spectra of DENV C in 10mM $KH_2PO_4$, pH 6.0, 100 mM KCl. A**) Unstable spectra of DENV C in the concentration of 2 μM, measured in a cuvette of 0.1 cm. **B**) Unstable spectra of DENV C in the concentration of 10 μM, measured in a cuvette of 0.1 cm. Measurements of 2 different samples on 2 days. All measurements were conducted in a 0.1 cm path length cuvette.

Finally, a buffer with the composition of 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl was used, in line with previous work[64]. The higher salt concentration would help to stabilize the spectra of the protein, even at low DENV C concentrations (*Figure 3.6 A*). Protein concentrations of 10 μM and lower did not give reproducible results. Measurements with a DENV C concentration of 20 μM showed the best result, including the expected and well-documented α-helical spectrum. Since it is not feasible to use 20 μM in the experiments for interaction with nucleic acids, it was necessary to change the cuvette path length. CD signal is directly proportional to the path length and protein concentration. Thus, with a 10-fold larger path length is possible to use one tenth of the protein concentration and, theoretically, obtain the same signal (assuming that protein structure and/or aggregation state is not affected by protein concentration, within that range). According to this, a 1 cm path length cuvette was used to measure DENV C at 2 μM. The results (*Figure 3.6 B*) show clear and stable spectra, indicative of α-helical secondary structure. Furthermore, these spectra superimpose with those of DENV C at 20 μM, measured with a path length of 0.1 cm, when the CD signal is expressed in mean residue molar ellipticity. Therefore, these were viable conditions. Consequently, cuvettes with a path length of 1 cm were employed to measure 2 μM DENV C in the presence of ssDNA.

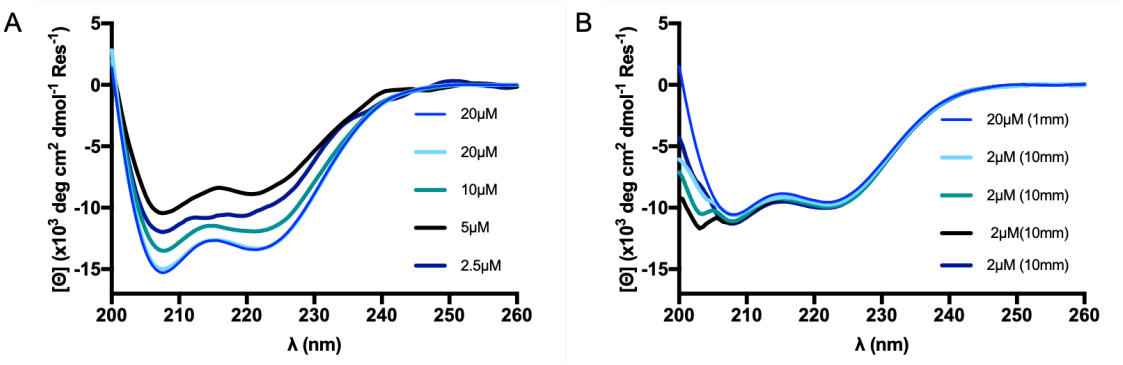


**Figure 3.6: CD spectra of DENV C in 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl. A**) DENV C at different concentrations, 0.1 cm cuvette. The best and most stable spectra are obtained with 20 μM protein, where two samples overlap. **B**) Superimposing DENV C spectra in a concentration of 2 μM, measured in a cuvette of 1 cm path length. As a reference, DENV C at 20 μM in a cuvette of 0.1 cm path length is shown.

## 3.4 Selection of nucleic acid sequences

The selection of nucleic acid sequences was done based on results from a collaboration with the Bioinformatics Institute A*STAR in Singapore. Distinct sections of the viral genome to which DENV C specifically binds were identified[98]. From the resulting data, sequences according to the following criteria were chosen: first, sections that showed a high affinity binding between DENV C and the viral genome were selected (binding profiles to the viral genome can be consulted in Appendix I, SI 6). The viral genome secondary structure can affect the binding. We thus selected nucleotide sequences with high affinity and either a linear or a structured secondary structure. Secondary structure predictions of 23ssDNA and 51ssDNA are available in Appendix I, SI 7. Last, we excluded from those sequences the ones that were either too long or very short, when compared to those that remained. Very long ones cannot be easily obtained by commercial synthesis. Very short sequences may not interact in a specific manner. Finally, two sequences from the remaining lot were selected to perform experiments. These were conducted using analogous ssDNA sequences as proxies of the viral genome.

**Table 1: Selected sequences for the interaction with DENV C.**

| Name | RNA Sequence | DNA sequence | criteria |
|---|---|---|---|
| 23 | CAGAAGGCUCA UGAAGAAGGCA UUUGUGGAAUC CGCUCAGUAAC AAGACUGGAG | CAGAAGGCTCA TGAAGAAGGCA TTTGTGGAATC CGCTCAGTAAC AAGACTGGAG | • High binding profile<br>• Structured<br>• Long sequence |
| 51 | UAGAGAUGGGUGAAG | TAGAGATGGGTGAAG | • High binding profile<br>• Linear<br>• Short sequence |

## 3.5 Interaction of DENV C with ssDNA

DENV C was titrated, up to 2 μM (monomer), into 1 μM of ssDNA in a 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl buffer (*Figure 3.7*). DENV C and 23ssDNA seem to interact causing a structural change in the protein (*Figure 3.7 A*). No interaction is seen between 51ssDNA and DENV C as there is no visible change and the CD spectra of DENV C alone (negative control) superimposes with it. The experiment was also performed with a buffer with $Mg^{2+}$ (composition: 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl, 20 mM $MgO_4S$). The interaction between 23ssDNA and DENV C is still seen, but to a lower degree. On the contrary, $Mg^{2+}$ does not affect the interaction (or lack of it) of 51ssDNA with DENV C.

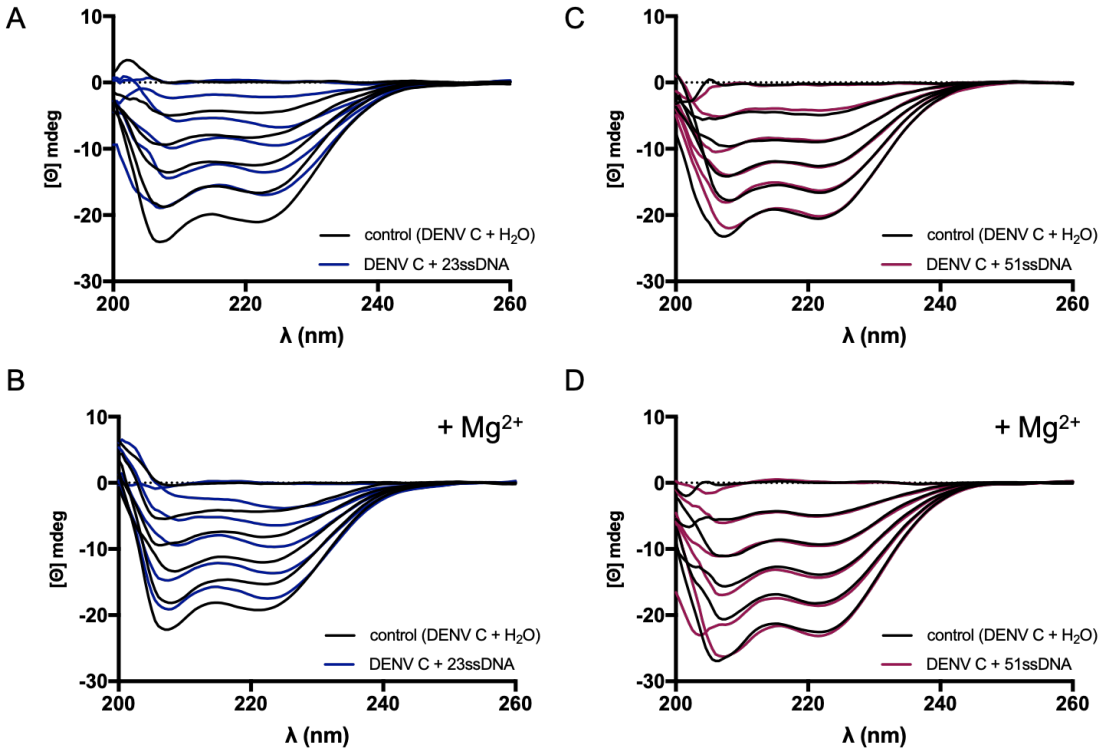


**Figure 3.7: DENV C-ssDNA interaction – CD spectra.** Titration of ssDNA with up to 2 μM DENV C in a 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl buffer, with and without $Mg^{2+}$ ions present (20 mM $MgO_4S$). **A**) DENV C-23ssDNA interaction, without $Mg^{2+}$. **B**) DENV C-23ssDNA interaction, with $Mg^{2+}$. **C**) DENV C-51ssDNA interaction, without $Mg^{2+}$. **D**) DENV C-51ssDNA interaction, with $Mg^{2+}$. Only 23ssDNA seems to interact with DENV C.

In Figure 3.8 information about DENV C interaction with 23ssDNA is obtained, by plotting the ellipticity ([θ]) at 222 nm (a local minimum for α-helices) as a function of DENV C concentration. A difference is seen between control (DENV C + $H_2O$) and the 23ssDNA sample, possibly due to structural changes after ssDNA binding. In Figure 3.8 B the signal of ssDNA after binding of DENV C is shown. The rising graph indicates an interaction between 23ssDNA and DENV C. In buffer with $Mg^{2+}$, 23ssDNA-DENV C interaction is weaker. Again, DENV C shows no interaction

with 51ssDNA. Also, the signal of the 51ssDNA stays relatively flat, which indicates that there is no interaction. Again, in buffer with $Mg^{2+}$, no interaction could be seen between DENV C and the 51ssDNA.

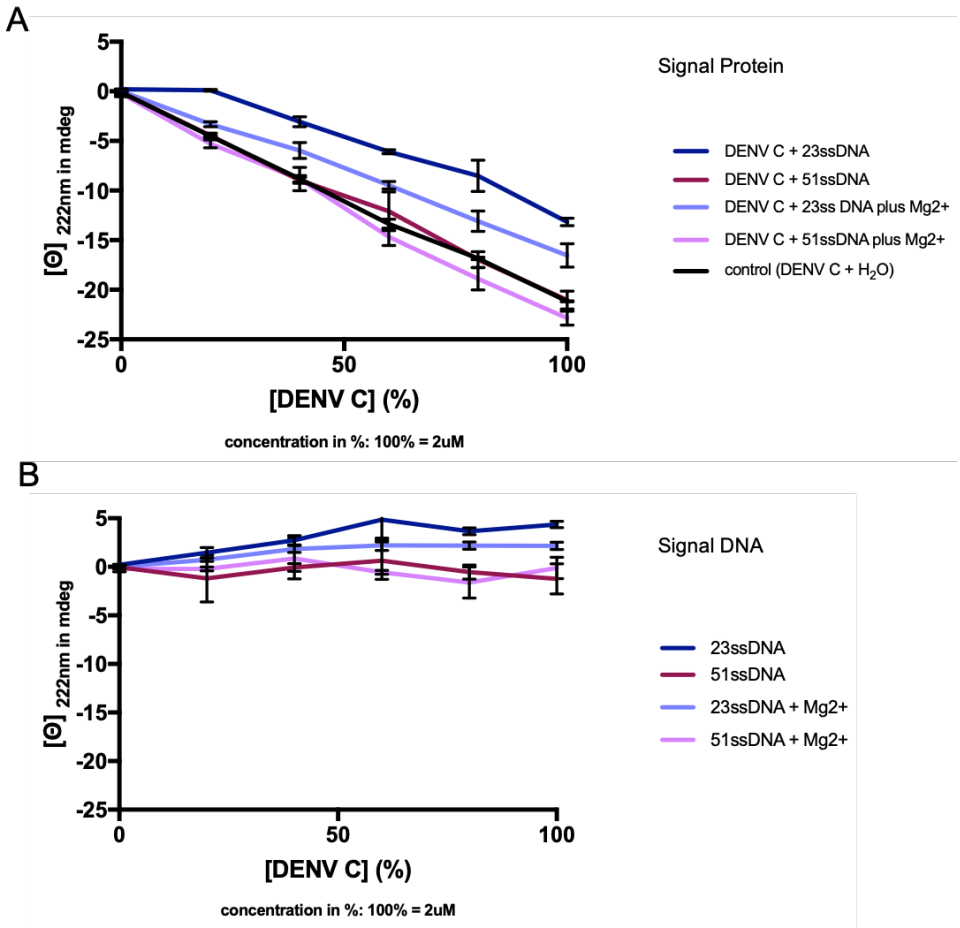


**Figure 3.8: Interaction of DENV C with two different ssDNA sequences. A)** Protein signal at 222 nm after subtraction of the DNA signal. **B**) DNA signal at 222 nm after subtraction of the protein signal. Both data indicate that 23ssDNA interacts with DENV C, but not 51ssDNA. Magnesium ions partially decrease the interaction, perhaps by making DNA more positively charged. However, if only electrostatics attraction played a role, 51ssDNA would be expected to bind DENV C at least to a small degree, which is not observed.

To complement the structural analysis of the interaction between DENV C and DNA, the same samples were tested via fluorescence spectroscopy, namely intrinsic protein fluorescence. The signal of the intrinsic protein fluorescence might be not strong enough to show the interaction between the C protein and DNA (*Figure 3.9*). There is no change visible in the spectra, neither in experiments with 23ssDNA, nor with 51ssDNA. After normalization to the maximum peak of the data, the spectra were superimposing, meaning that the DENV C protein always presents the same tryptophan environment and structural shape. Results of the normalization can be consulted in Appendix I, SI 8.

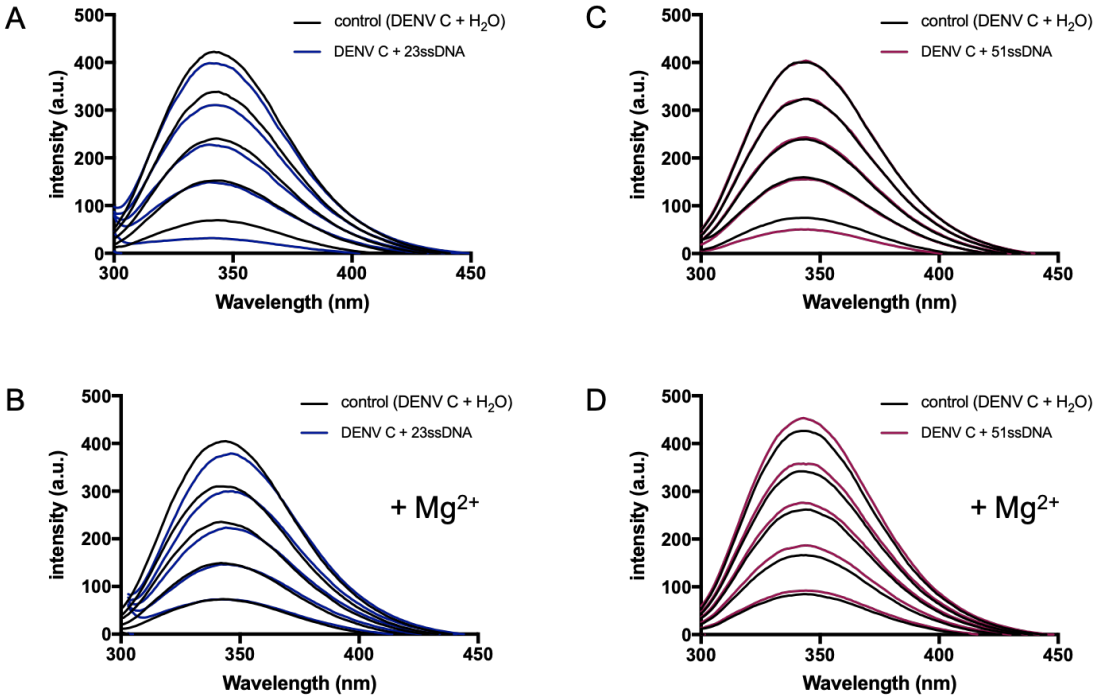


**Figure 3.9: DENV C-ssDNA interaction – intrinsic fluorescence spectroscopy analysis.** Titration of DENV C (up to 2 μM) to two ssDNA sequences, in a buffer of 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl, with and without $Mg^{2+}$ ions present (20 mM $MgO_4S$). **A**) DENV C-23ssDNA interaction, without $Mg^{2+}$. **B**) DENV C-23ssDNA interaction, with $Mg^{2+}$. **C**) DENV C-51ssDNA interaction, without $Mg^{2+}$. **D**) DENV C-51ssDNA interaction, with $Mg^{2+}$. No interaction is clear, although a small loss of signal is seen, possibly indicating some small level of protein aggregation and/or precipitation. Nevertheless, the peak maximum and the spectral shape suffer minimal changes.

# 4 Discussion

In this chapter, the results of the experimental work of the thesis are discussed within the frame of the current knowledge and recent publications. There will be a deeper insight into the interaction of DENV C with its ligands. A mode of action for the binding of DENV C to the viral genome and possible structural changes are proposed.

## 4.1 DENV C protein expression and characterization

DENV C protein was successfully expressed and purified. No major contaminants were seen in the SDS-PAGE performed after size exclusion chromatography. DENV C quality and purity was further analyzed by mass spectrometry, showing the expected mass in fraction 4 and 5 of the first purification and in fraction 5 of the second purification (*Appendix I, SI 2 – 5).* Fraction 6 of the second purification does not show the expected mass spectrum, thus it was not used for the experiments.

CD spectra of $\alpha$-helical proteins display negative peaks at ~222 nm and ~208 nm[82]. Spectra from $\beta$-sheet proteins show a negative peak at ~217 nm[82,84]. According to this, we could obtain a CD spectrum (*Figure 3.3*) and assert the $\alpha$-helical secondary structure for the DENV C, as expected.

## 4.2 Optimization of CD measurements at very low DENV C concentration

The optimization was done to test the best conditions possible to measure DENV C protein in very low concentrations, making the work with ssDNA more feasible and to prevent protein aggregation. First of all, buffers used for CD spectroscopy should not contain any optically active materials and should be as transparent as possible[99]. Samples dissolved only in water show the highest transparency. However, some proteins denature in the absence of salt. Going in line with this, first tests were performed in water, but did not give any stable results and indicated some level of random coil (*Figure 3.4*), so buffers with different salt concentrations were tested. Furthermore, buffers for CD analysis should be free of particulate matter, so filtering the buffers before use is crucial. High transparency quartz cuvettes (cells) were used, with path length between 0.1 and 1 cm. Cells of very short pathlengths (> 0.1 cm) have a small total sample volume and a small surface area facing the light beam of the CD apparatus. In those cases, it is important that the light beam is tightly focused, since large artifacts may be produced otherwise[99]. Using larger path length cells is only advisable to facilitate the production of stable spectra, but at the cost of larger volumes. When using high protein concentrations in very short path length cells (or low concentrations in long path way cells), a high-tension voltage artifact can occur (*Figure 3.6 B*). This happens when strong absorption of the light beam by the sample does not allow a sufficient amount of light to reach the detector and in turn creates artifacts[100]. Taking all of the above into account, 1 cm cuvettes were the most adequate, showing reproducible results (*Figure 3.6 B*) and were used henceforth.

## 4.3 Selection of nucleic acid sequences

As described in the results, a previous study by the host laboratory[98] helped to identify specific sections of the viral genome important for DENV C binding to it. Briefly, the architecture of the nucleocapsid of DENV serotype 2 and ZIKV was elucidated by using a nuclease digestion assay and next generation sequencing to map the respective viral RNA (vRNA) genome wide association with the C protein *in vitro*[98]. Despite an irregular organization of the nucleocapsid, it was shown that the C protein binds in a non-random manner along the vRNA with distinct patterns of association. Furthermore, it shows a binding preference for Guanin (G)-rich sequences, with the potential to form G-quadruplexes[98]. As already established, G-quadruplexes have the potential to affect viral life cycles[101]. According to this, the chosen nucleic acid sequences used in this thesis are both G-rich.

In addition, it is known that ZIKV C protein binds single-stranded and double-stranded RNA, as well as DNA[81]. Whether this is specific binding or not, is not yet clear. In any case, this broad binding feature of ZIKV C protein suggests that it might involve a nucleotide-associated host response, in addition to its viral ssRNA binding for virus assembly[81]. Since ZIKV and DENV are closely related viruses, the same may occur with DENV C. In consonance with all of the above, ssDNA oligonucleotides were used as proxies of ssRNA viral genome sequences. As the dengue genome is ssRNA, ssDNA is an adequate substitute, and much easier to work with than ssRNA, that is prone to being quickly degraded. The ssDNA sequences were selected from the most promising sites identified by the host laboratory and collaborators[98], according to the rational detailed before in the results section. Further details can be found in Appendix I, SI6. Overall, studying these ssDNA further will help identify the precise binding sites of DENV C protein within the viral ssRNA genome.

## 4.4 Interaction of DENV C with ssDNA

To further investigate the mechanism of action of DENV C, we focused on the interaction of the C protein with nucleic acids. It is important to characterize DENV C binding to the viral ssRNA, since it is a crucial step in the viral life cycle. Analogous ssDNA sequences were used as proxies of the viral genome. As already described before, the C proteins positively charged $\alpha$4-$\alpha$4' region may govern its binding to nucleic acids, especially to double helical structures. The 23ssDNA, a long and structured sequence, showed interaction, as well as a change in structure of the C protein after ssDNA binding. However, the 51ssDNA (short, linear) showed no interaction. Next, it was important to discriminate whether this interaction is specific or not. The results indicate that the length, nucleic acid structure and also the ssDNA sequence may be crucial for the specific DENV C-ssDNA binding. Addition of $Mg^{2+}$ does not seem to promote the interaction, as it lowers

it, decreasing the spectra's intensity. It can be hypothesized that in presence of $Mg^{2+}$ the structure of the ssDNA might change and affect DENV C binding. Alternatively, the $Mg^{2+}$ may bind the ssDNA, making it a more positively charged particle, and thus decreasing electrostatics. This could force attraction of the nucleic acid particle to the C protein. To evaluate this, more structured ssDNA sequences need to be tested. Additional investigations may be performed, replacing $Mg^{2+}$ with another divalent ion, such as $Ca^{2+}$ or $Mn^{2+}$, to help determine if DENV C interaction with 23ssDNA is specific or due to indiscriminate electrostatic forces. Nevertheless, if DENV C binding to ssDNA was purely electrostatics, it would bind to some degree to 51ssDNA, as it is negatively charged as well. Since it was shown that only the 23ssDNA interacted, this suggests that the protein is at least partially selective, binding to specific sequences.

Analysing the data of the intrinsic fluorescence measurements (*Figure 3.9*) reveals that, although there is no difference in the maximum peak of the spectra, the intensity of DENV C with 23ssDNA is slightly lower than the control. This can mean a small decrease in protein concentration. Since this effect is stronger with the 23ssDNA than the 51ssDNA, it can be an effect of binding associated with some degree of protein aggregation and deposition, causing the concentration of DENV C in solution to decrease. Such interpretation goes also in line with results of CD analysis. A closer look at the acquired spectra reveals a change upon interaction with 23ssDNA, which might be due, to a small extent, to protein precipitation from solution. However, this does not explain all changes, since the shape of the spectra changes. There is a shift in DENV C secondary structure towards possibly more random coil or $\beta$-sheet. Analysis through deconvolution approaches, such as K2D2[102] (*Appendix I, SF 10*) suggest that the interaction with 23ssDNA causes a decrease in DENV C $\alpha$-helical content. Overall, if spectra are normalized, no isodichroic point is seen, suggesting a mechanism with more than one species present in solution and not a simple change from $\alpha$-helix to another structure of DENV C, such as random coil. The ssDNA signal is residual, if compared to the protein signal, and cannot account for all the changes.

Looking at the protein structure, the CD and fluorescence data suggest that the main structure remains unchanged. The tryptophan environment, forming part of the protein inner "conserved fold" core stable region, remains the same. It is likely that the C-terminal region is prone to interact with negatively charged nucleic acids, due to its positive net charge (+7). Each C-terminal of the DENV C monomer provides a net charge of +8. Since the protein forms homodimers in solution, this net charge is in fact the double (+16), making it more susceptible to interact with ssDNA and thus viral RNA. Still, it cannot be excluded that the N-terminal might interact as well, as each N-terminal monomer also shows a positive net charge of almost the same degree. These hypotheses demand further enquiry, using other techniques, such as nuclear magnetic resonance, to precisely map the DENV C-ssDNA binding sites.

Few studies have tried to evaluate the interaction of the capsid protein with RNA. A possible role of DENV C as an RNA chaperone was studied and it was shown that the C protein facilitates the formation of the typical hammerhead structure of the RNA[103]. However, the N-terminal regions of DENV C are intrinsically disordered, exhibiting high flexibility in folding and also possess many

positively charged amino acid residues. They have been shown to be essential for DENV propagation in human and mosquito cells[104]. Besides this, it was shown that truncated C protein variants of other Flaviviruses do not impair viral fitness to the same range[105,106]. Hence, and supporting the observations concerning the N-terminal charge, even if the interaction of the C-protein with the viral RNA occurs mostly through the α4-α4' interface, the N-terminal IDP regions may still play a role[98]. It is important to notice, that intrinsically disordered protein domains are a natural way for increasing the functional activity of viruses and other organisms[107,108]. Another crucial hypothesis is that the α4-α4' mediated binding to nucleic acids is dependent upon the *in vivo* interaction with host lipid systems, via the N-terminal and the hydrophobic α2-α2' interface[98]. This interaction may promote structural arrangements that allow C protein binding to the vRNA. Thus, the role of host lipid systems in the binding of DENV C to nucleic acids should also be considered.

Importantly, it was recently shown that mutations in the dimer interfaces of the DENV C protein affect the structural stability and impair the RNA-capsid interaction[109]. In 2007, *Patkar et. al*.[106] predicted the structure for the YFV capsid proteins and mutations at the residues 81 and 88 lead to an impaired viral assembly[106]. Based on those results, Figueira-Mansur *et. al.* (2019)[109] generated single mutants of DENV C at L50 and L54 of the α-helix, which was shown to interfere with the integration of the capsid into LDs and at residues L88 and I88 located in the α-helix, which was shown to affect viral assembly. All mutations promoted a reduction in the proteins stability and furthermore they affected the ability of DENV C to interact with RNA, especially pronounced for mutations in the α4-helix[109]. Keeping all this in mind, the α4-α4' interface could be a potential target for antiviral drug design in the future.

## 4.5 Computational analysis

The findings of this thesis should be considered in light of DENV C biologically relevant interactions with LDs[78] and RNA (*Figure 4.1*). As a basis for this analysis, the structure of DENV C (PDB ID: 1R6R [63]) was used. The experimental structure shows three distinct structural regions: the disordered N-terminal (starting at the N-terminal up to residue R22), the flexible fold (residues V23 to L44, where α-helix 1 is located) and the conserved fold that includes helices α2, α3 and α4, containing the residues R68 and W69 that are highly conserved among *Flavivirus*[63]. R68 is located in the α3 helix, with its side chain pointing to the protein interior[63]. W69 locates at DENV C α3 as well, having a crucial role in structural stability of the dimer[63]. Along with dimer structural stability, these interactions enable allosteric communication and movements between DENV C more hydrophobic section (α2-α2´dimer interface) and its remaining sections, namely the α4-α4' region. Figure 4.1 displays this, in the context of the C protein biologically relevant interactions, as they are understood on the basis of recent studies[65,66,71,72,77–80]. Looking further, it is important to consider that the binding of DENV C to host LDs is mediated by both the N-terminal and the

α2 region[65]. V51 of α2 is affected by the interaction with LDs and stabilizes the dimer by contact with α3 (I65). Another interaction via salt bridges from α2 (K45 and R55') to α4 (E87) stabilizes the homodimer (*Figure 4.1 A*). The C protein binding to host LDs, which affects α2-α2', can lead to changes in the α4-α4' structural arrangement (*Figure 4.1 B*).

With the above in mind, we did an extensive search in PDB for similar proteins. Influenza A non-structural protein 1 (influenza NS1) (PDB ID: 2ZKO[97]), a RNA-binding viral protein, with a two α-helix interface domain similar to DENV C α4-α4', was identified (*Figure 4.1 C*). To take a closer look to the similarity with the α4-α4' motif, please consult Appendix I, SI 9. Influenza NS1 has interesting features: it accumulates in the nuclei of host cells after being translocated by importin α and β and works as a viral immune-suppressor by weakening the host cell gene expression[110]. DENV C was also reported to have an importin α-like motif in the N-terminal [66,111].

DENV C may function as an immunosuppressor, similarly to influenza NS1, by interacting with importins α and/or β. Importantly, ivermectin, a specific inhibitor of importin α/β-mediated nuclear import, is also able to inhibit the replication of HIV-1 and DENV [112]. The mechanism might involve the C protein, specifically the N-terminal IDP, which is similar to the N-terminal IDP of importin α[66]. Moreover, influenza NS1 can counteract the RNA-activated protein kinase (PKR)-mediated antiviral response due to direct interaction with PKR[113]. Besides that, influenza NS1 blocks the activation of the interferon (IFN) regulatory factor 3, which in turn prevents the induction of IFN-related genes [114]. DENV inhibits the IFN signalling pathway in a similar manner[115]. By its N-terminal dsRNA-binding ability, influenza NS1 inhibits the nuclear export of mRNAs and modulates pre-mRNA splicing, also to suppress antiviral response[97]. Similarities between DENV C and influenza NS1 also extend to the later ability to bind RNA (*Figure 4.1 C*). Recognition of dsRNA is made by influenza NS1 RNA-binding domain, that forms a homodimer[97]. Following, a slight change in R38-R38' orientation leads to anchoring the dsRNA by a hydrogen bond network to the protein[97].

Taking all this into account, it can be proposed, as with influenza NS1, a small conformational change in DENV C α4-α4' interface occurs after contact of its α2-α2' interface with LDs, modulated by transitions between N-terminal alternative "open" and "closed" conformations. Binding to LDs requires an open conformation (*Figure 4.1 D*), decreasing the conformational variability and entropy of DENV C protein, which would trigger the allosteric movements affecting the C-terminal α4-α4'. As with influenza NS1, the *Flavivirus* C protein would remain in the same overall fold, but a small opening of α4-α4' would facilitate its binding to RNA. Further analysis will help to understand the interaction between DENV C and its ligands, as well as the development of C protein-targeted drugs.

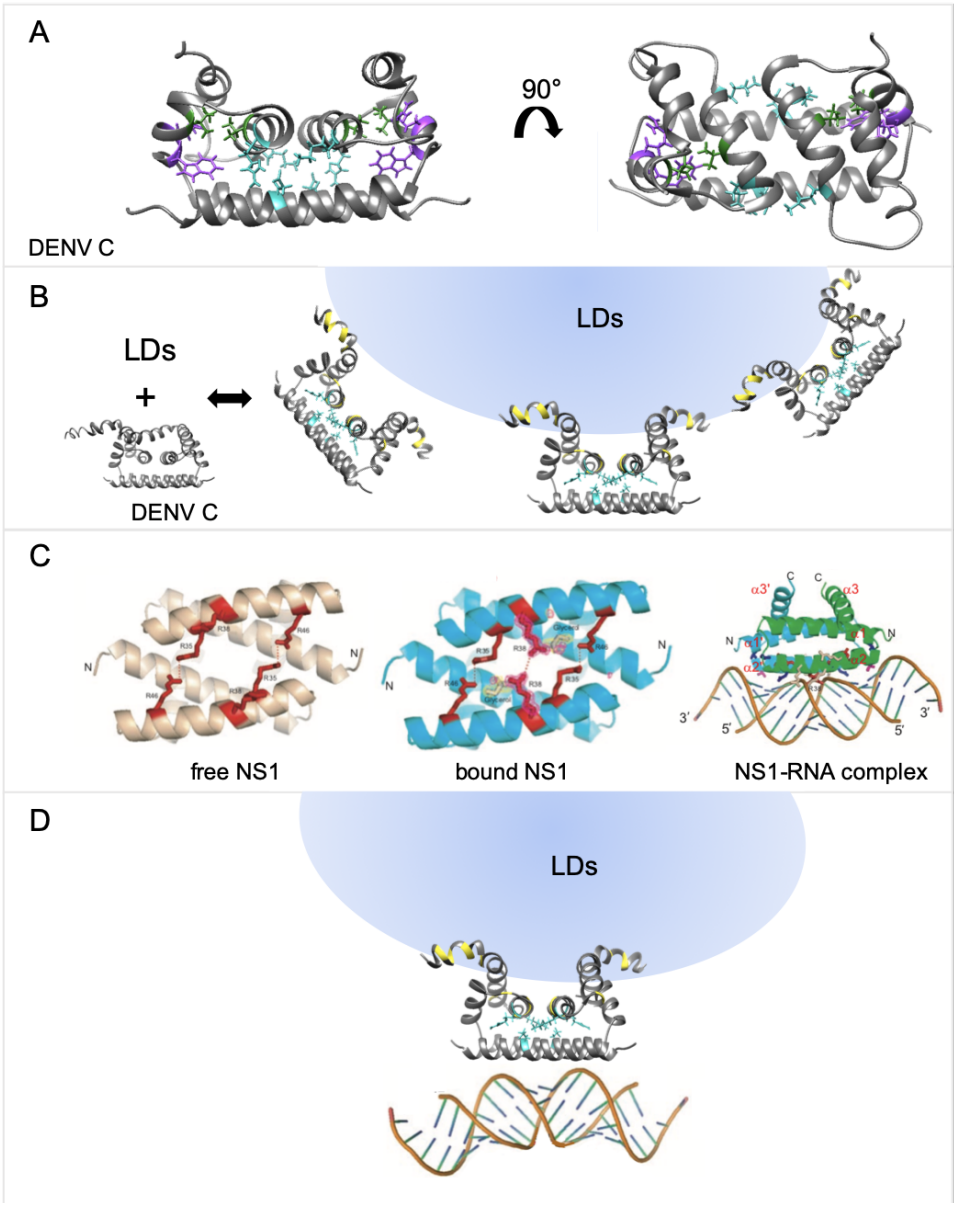


**Figure 4.1: Protein structures of DENV C and influenza NS1.** (**A**) DENV C structure from two different angles with the conserved residues R68 and W69 (purple), and the interface stabilizing residues V51 and I65 (green), as well as E87, R55 and K45, forming the salt bridge (cyan). (**B**) DENV C structure in a N-terminal closed formation and next in an open conformation with schematic binding of LDs and the affected amino acids (yellow). (**C**) The RNA-binding domain of NS1 protein from Influenza A in a RNA-free (left) and RNA-bound state (middle and right), showing an organization similar to DENV C α4-α4' region (adapted from Cheng *et al.*, 2009). (**D**) DENV C structure with schematic bound LDs and RNA. DENV C amino acid residues affected by the binding to are colored yellow, while a key internal salt bridge is shown in cyan. DENV C binding to host LDs may enable allosteric rearrangements (eventually involving the salt bridge), that may lead to a small conformational change in α4 side chains, namely the positively charged residues that stabilize the RNA-bound form of DENV C and, similarly, of the other *Flavivirus* C protein.

## 4.6 Future work

The mechanism of binding of DENV C to nucleic acids and their role in the *Flavivirus* life cycle is still poorly understood. Recent studies elucidated that DENV C may interact with specific regions of the viral genome[98] and this thesis could partly prove this interaction. Keeping this in mind and that the binding to host LDs affects DENV C $\alpha$2-$\alpha$2' region, as well as its N-terminal IDP region, it can be hypothesized that DENV C-LDs binding enables some allosteric movements. Such movements can result in a change on the $\alpha$4-$\alpha$4' structural arrangement and in turn facilitates the binding to the viral RNA. However, more ssDNA sequences need to be tested. NMR experiments can also help to map the specific binding sites of nucleic acids on DENV C. It is also of great interest to test the interaction of DENV C with viral RNA, both in the absence and presence of host lipid systems, to further elucidate how the assembly of the nucleocapsid is performed in *Flavivirus* in general.

In the future, DENV C binding to RNA can be studied, using biophysical approaches like CD and light scattering spectroscopies, to determine in detail binding parameters and affinities, as well as any structural changes and estimate the size of the RNA-C protein complexes. Calorimetry assays upon the binding of DENV C to RNA can complement the thermodynamic binding data. Later on, this knowledge may enable targeting of the most relevant sequences, seeking to inhibit key interactions and develop new therapeutic approaches against DENV and closely related flaviviruses.

# 5 References


1. WHO. *Dengue and severe dengue*. *Fact sheets: Dengue and severe dengue* (2018).
2. 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).
3. Lupi, O. Mosquito-borne hemorrhagic fevers. *Dermatol. Clin.* **29**, 33–38 (2011).
4. Weaver, S. C. & Reisen, W. K. Present and future arboviral threats. *Antiviral Res.* **85**, 328–345 (2010).
5. Weissenbock, H., Hubalek, Z., Bakonyi, T. & Nowotny, N. Zoonotic mosquito-borne flaviviruses: worldwide presence of agents with proven pathogenicity and potential candidates of future emerging diseases. *Vet. Microbiol.* **140**, 271–280 (2010).
6. Diallo, M. *et al.* Potential role of sylvatic and domestic African mosquito species in dengue emergence. *Am. J. Trop. Med. Hyg.* **73**, 445–449 (2005).
7. Guzman, M. G. *et al.* Dengue: a continuing global threat. *Nat. Rev. Microbiol.* **8**, S7-16 (2010).
8. Schneider, B. S., Soong, L., Zeidner, N. S. & Higgs, S. Aedes aegypti salivary gland extracts modulate anti-viral and TH1/TH2 cytokine responses to sindbis virus infection. *Viral Immunol.* **17**, 565–573 (2004).
9. Salazar, M. I., Richardson, J. H., Sanchez-Vargas, I., Olson, K. E. & Beaty, B. J. Dengue virus type 2: replication and tropisms in orally infected Aedes aegypti mosquitoes. *BMC Microbiol.* **7**, 9 (2007).
10. Chan, E. H., Sahai, V., Conrad, C. & Brownstein, J. S. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl. Trop. Dis.* **5**, e1206 (2011).
11. Rathore, A. P. S. & St John, A. L. Immune responses to dengue virus in the skin. *Open Biol.* **8**, (2018).
12. Wu, S. J. *et al.* Human skin Langerhans cells are targets of dengue virus infection. *Nat. Med.* **6**, 816–820 (2000).
13. Jessie, K., Fong, M. Y., Devi, S., Lam, S. K. & Wong, K. T. Localization of dengue virus in naturally infected human tissues, by immunohistochemistry and in situ hybridization. *J. Infect. Dis.* **189**, 1411–1418 (2004).
14. Juneja, D. *et al.* Clinical profile, intensive care unit course, and outcome of patients admitted in intensive care unit with dengue. *J. Crit. Care* **26**, 449–452 (2011).
15. Esler, D. Dengue - Clinical and public health ramifications. *Aust. Fam. Physician* **38**, 876–879 (2009).
16. Guzman, M. G. *et al.* Dengue hemorrhagic fever in Cuba, 1981: a retrospective seroepidemiologic study. *Am. J. Trop. Med. Hyg.* **42**, 179–184 (1990).
17. Alvarez, M. *et al.* Dengue hemorrhagic Fever caused by sequential dengue 1-3 virus infections over a long time interval: Havana epidemic, 2001-2002. *Am. J. Trop. Med. Hyg.* **75**, 1113–1117 (2006).
18. Endy, T. P. *et al.* Determinants of inapparent and symptomatic dengue infection in a prospective study of primary school children in Kamphaeng Phet, Thailand. *PLoS Negl. Trop. Dis.* **5**, e975 (2011).
19. Gamble, J. *et al.* Age-related changes in microvascular permeability: a significant factor in the susceptibility of children to shock? *Clin. Sci. (Lond).* **98**, 211–216 (2000).
20. Guzman, M. G. *et al.* Effect of age on outcome of secondary dengue 2 infections. *Int. J. Infect. Dis.* **6**, 118–124 (2002).
21. Stephens, H. A. F. *et al.* HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. *Tissue Antigens* **60**, 309–318 (2002).
22. Kalayanarooj, S. *et al.* Blood group AB is associated with increased risk for severe dengue disease in secondary infections. *J. Infect. Dis.* **195**, 1014–1017 (2007).
23. Sakuntabhai, A. *et al.* A variant in the CD209 promoter is associated with severity of dengue disease. *Nat. Genet.* **37**, 507–513 (2005).
24. Fernandez-Mestre, M. T., Gendzekhadze, K., Rivas-Vetencourt, P. & Layrisse, Z. TNF-alpha-308A allele, a possible severity risk factor of hemorrhagic manifestation in dengue fever patients. *Tissue Antigens* **64**, 469–472 (2004).
25. Thisyakorn, U. & Nimmannitya, S. Nutritional status of children with dengue hemorrhagic

fever. *Clin. Infect. Dis.* **16**, 295–297 (1993).
26. Loke, H. *et al.* Susceptibility to dengue hemorrhagic fever in vietnam: evidence of an association with variation in the vitamin d receptor and Fc gamma receptor IIa genes. *Am. J. Trop. Med. Hyg.* **67**, 102–106 (2002).
27. de la C Sierra, B., Kouri, G. & Guzman, M. G. Race: a risk factor for dengue hemorrhagic fever. *Arch. Virol.* **152**, 533–542 (2007).
28. Guzman, M. G. *et al.* Enhanced severity of secondary dengue-2 infections: death rates in 1981 and 1997 Cuban outbreaks. *Rev. Panam. Salud Publica* **11**, 223–227 (2002).
29. Gubler, D. J. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* **11**, 480–496 (1998).
30. Guzman, M. G. & Kouri, G. Dengue diagnosis, advances and challenges. *Int. J. Infect. Dis.* **8**, 69–80 (2004).
31. TDR/WHO. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control.* (2009).
32. Innis, B. L. *et al.* An enzyme-linked immunosorbent assay to characterize dengue infections where dengue and Japanese encephalitis co-circulate. *Am. J. Trop. Med. Hyg.* **40**, 418–427 (1989).
33. Chanama, S. *et al.* Analysis of specific IgM responses in secondary dengue virus infections: levels and positive rates in comparison with primary infections. *J. Clin. Virol.* **31**, 185–189 (2004).
34. Gubler, D. J. Serological diagnosis of dengue/dengue haemorrhagic fever. *Dengue Bull.* **20**, 20–23 (1996).
35. Vazquez, S. *et al.* Kinetics of antibodies in sera, saliva, and urine samples from adult patients with primary or secondary dengue 3 virus infections. *Int. J. Infect. Dis.* **11**, 256–262 (2007).
36. Gibbons, R. V & Vaughn, D. W. Dengue: an escalating problem. *BMJ* **324**, 1563–1566 (2002).
37. Durbin, A. P. A Dengue Vaccine. *Cell* **166**, 1 (2016).
38. Sanofi. Dengvaxia® vaccine approved for prevention of dengue in Europe. *19.12.2018* (2018).
39. Villar, L. *et al.* Efficacy of a tetravalent dengue vaccine in children in Latin America. *N. Engl. J. Med.* **372**, 113–123 (2015).
40. Takeda. Takeda's Dengue Vaccine Candidate Meets Primary Endpoint in Pivotal Phase 3 Efficacy Trial. *29.01.2019* (2019). Available at: https://www.takeda.com/newsroom/newsreleases/2019/takedas-dengue-vaccine-candidate-meets-primary-endpoint-in-pivotal-phase-3-efficacy-trial/. (Accessed: 25th March 2019)
41. Calisher, C. H. & Gould, E. A. Taxonomy of the virus family Flaviviridae. *Adv. Virus Res.* **59**, 1–19 (2003).
42. Grard, G. *et al.* Genomics and evolution of Aedes-borne flaviviruses. *J. Gen. Virol.* **91**, 87–94 (2010).
43. Mukhopadhyay, S., Kuhn, R. J. & Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.* **3**, 13–22 (2005).
44. Kuhn, R. J. *et al.* Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717–725 (2002).
45. Halstead, S. B. Dengue virus-mosquito interactions. *Annu. Rev. Entomol.* **53**, 273–291 (2008).
46. Kaufmann, B. & Rossmann, M. G. Molecular mechanisms involved in the early steps of flavivirus cell entry. *Microbes Infect.* **13**, 1–9 (2011).
47. Allison, S. L. *et al.* Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *J. Virol.* **69**, 695–700 (1995).
48. Urcuqui-Inchima, S., Patino, C., Torres, S., Haenni, A.-L. & Diaz, F. J. Recent developments in understanding dengue virus replication. *Adv. Virus Res.* **77**, 1–39 (2010).
49. Welsch, S. *et al.* Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell Host Microbe* **5**, 365–375 (2009).
50. Brinton, M. A. The molecular biology of West Nile Virus: a new invader of the western hemisphere. *Annu. Rev. Microbiol.* **56**, 371–402 (2002).
51. Wang, P.-G. *et al.* Efficient assembly and secretion of recombinant subviral particles of the four dengue serotypes using native prM and E proteins. *PLoS One* **4**, e8325 (2009).

52. Guirakhoo, F., Heinz, F. X., Mandl, C. W., Holzmann, H. & Kunz, C. Fusion activity of flaviviruses: comparison of mature and immature (prM-containing) tick-borne encephalitis virions. *J. Gen. Virol.* **72 ( Pt 6)**, 1323–1329 (1991).
53. Elshuber, S., Allison, S. L., Heinz, F. X. & Mandl, C. W. Cleavage of protein prM is necessary for infection of BHK-21 cells by tick-borne encephalitis virus. *J. Gen. Virol.* **84**, 183–191 (2003).
54. Chareonsirisuthigul, T., Kalayanarooj, S. & Ubol, S. Dengue virus (DENV) antibody-dependent enhancement of infection upregulates the production of anti-inflammatory cytokines, but suppresses anti-DENV free radical and pro-inflammatory cytokine production, in THP-1 cells. *J. Gen. Virol.* **88**, 365–375 (2007).
55. Simmons, C. P. *et al.* Patterns of host genome-wide gene transcript abundance in the peripheral blood of patients with acute dengue hemorrhagic fever. *J. Infect. Dis.* **195**, 1097–1107 (2007).
56. Couvelard, A. *et al.* Report of a fatal case of dengue infection with hepatitis: demonstration of dengue antigens in hepatocytes and liver apoptosis. *Hum. Pathol.* **30**, 1106–1110 (1999).
57. Alcon-LePoder, S. *et al.* The secreted form of dengue virus nonstructural protein NS1 is endocytosed by hepatocytes and accumulates in late endosomes: implications for viral infectivity. *J. Virol.* **79**, 11403–11411 (2005).
58. Yu, I.-M. *et al.* Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* **319**, 1834–1837 (2008).
59. Zhang, Y. *et al.* Structures of immature flavivirus particles. *EMBO J.* **22**, 2604–2613 (2003).
60. Plevka, P. *et al.* Maturation of flaviviruses starts from one or more icosahedrally independent nucleation centres. *EMBO Rep.* **12**, 602–606 (2011).
61. Yu, I.-M. *et al.* Association of the pr peptides with dengue virus at acidic pH blocks membrane fusion. *J. Virol.* **83**, 12101–12107 (2009).
62. Martin-Acebes, M. A., Vazquez-Calvo, A. & 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).
63. Ma, L., Jones, C. T., Groesch, T. D., Kuhn, R. J. & Post, C. B. Solution structure of dengue virus capsid protein reveals another fold. **101**, 1–6 (2004).
64. Faustino, A. F. *et al.* Fast NMR method to probe solvent accessibility and disordered regions in proteins. *Sci. Rep.* 1–13 (2019). doi:10.1038/s41598-018-37599-z
65. Martins, I. C. *et al.* The disordered N-terminal region of dengue virus capsid protein contains a lipid-droplet-binding motif. *Biochem. J.* **444**, 405–415 (2012).
66. 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).
67. Jones, C. T. *et al.* Flavivirus Capsid Is a Dimeric Alpha-Helical Protein. *J. Virol.* **77**, 7143–7149 (2003).
68. 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.* **101**, 3414–3419 (2004).
69. Willows, S., Hou, S. & Hobman, T. C. RNA virus capsid proteins: more than just a shell. *Future Virol.* **8**, 435–450 (2013).
70. 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).
71. Samsa, M. M. *et al.* Dengue virus capsid protein usurps lipid droplets for viral particle formation. *PLoS Pathog.* **5**, e1000632 (2009).
72. 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**, 2096–2108 (2012).
73. van Gorp, E. C. M. *et al.* Changes in the Plasma Lipid Profile as a Potential Predictor of Clinical Outcome in Dengue Hemorrhagic Fever. *Clin. Infect. Dis.* **34**, 1150–1153 (2002).
74. Suvarna, J. C. & Rane, P. P. Serum lipid profile: a predictor of clinical outcome in dengue infection. *Trop. Med. Int. Heal.* **14**, 576–585 (2009).
75. 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).
76. Martins, I., Almeida, F., Santos, N. & Da Poian, A. DENV–derived peptides and methods for the inhibition of Flavivirus replication. (2012). International Patent Publication Nr. WO/2012/159187
77. Faustino, A. F. *et al.* Dengue virus capsid protein interacts specifically with very low-density lipoproteins. *Nanomedicine Nanotechnology, Biol. Med.* **10**, 247–255 (2014).
78. Faustino, A. F. *et al.* Understanding Dengue Virus Capsid Protein Interaction with Key Biological Targets. *Sci. Rep.* **5**, 10592 (2015).
79. Martins, A. S., Carvalho, F. A., Faustino, A. F., Martins, I. C. & Santos, N. C. West Nile Virus Capsid Protein Interacts With Biologically Relevant Host Lipid Systems. *Front. Cell. Infect. Microbiol.* **9**, 8 (2019).
80. Martins, A. S., Martins, I. C. & Santos, N. C. Methods for Lipid Droplet Biophysical Characterization in Flaviviridae Infections. *Front. Microbiol.* **9**, 1951 (2018).
81. 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).
82. 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).
83. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–139 (2005).
84. Woody, R. W. *Circular Dichroism and the Conformational Analysis of Biomolecules*. (Plenum Press, 1996).
85. Rucker, A. L. & Creamer, T. P. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci.* **11**, 980–985 (2002).
86. Ghisaidoobe, A. B. T. & Chung, S. J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *Int. J. Mol. Sci.* **15**, 22518–22538 (2014).
87. TEALE, F. W. & WEBER, G. Ultraviolet fluorescence of the aromatic amino acids. *Biochem. J.* **65**, 476–482 (1957).
88. Favicchio, R., Dragan, A. I., Kneale, G. G. & Read, C. M. Fluorescence spectroscopy and anisotropy in the analysis of DNA-protein interactions. *Methods Mol. Biol.* **543**, 589–611 (2009).
89. Lakowicz, J. *Principles of Fluorescence Spectroscopy.* (Springer, 2006).
90. Harris, D.A., Bashford, C. L. *Spectrophotometry and Spectrofluorimetry: A Practical Approach.* (IRL Press, 1987).
91. Vivian, J. T. & Callis, P. R. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.* **80**, 2093–2109 (2001).
92. Chen, R. F. Fluorescence Quantum Yields of Tryptophan and Tyrosine. *Anal. Lett.* **1**, 35–42 (1967).
93. Beechem, J. M. & Brand, L. Time-resolved fluorescence of proteins. *Annu. Rev. Biochem.* **54**, 43–71 (1985).
94. Eftink, M. R. Fluorescence techniques for studying protein structure. *Methods Biochem. Anal.* **35**, 127–205 (1991).
95. Gryczynski, I., Wiczk, W., Johnson, M. L. & Lakowicz, J. R. Lifetime distributions and anisotropy decays of indole fluorescence in cyclohexane/ethanol mixtures by frequency-domain fluorometry. *Biophys. Chem.* **32**, 173–185 (1988).
96. Munishkina, L. A. & Fink, A. L. Fluorescence as a method to reveal structures and membrane-interactions of amyloidogenic proteins. *Biochim. Biophys. Acta - Biomembr.* **1768**, 1862–1885 (2007).
97. Cheng, A., Wong, S. M. & Yuan, Y. A. Structural basis for dsRNA recognition by NS1 protein of influenza A virus. *Cell Res.* **19**, 187–195 (2009).
98. P. L.S. Boon, A. S. Martins, F. J. Enguita, X. N. Lim, N. C. Santos, P. T. Matsudaira, I. C. Martins, W. Yue, P. J. Bond, R. G. H. Genome-wide Associations of Flavivirus Capsid Proteins. *BioRxiv* (2019). doi:10.1101/606905
99. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2006).
100. Kumar, C. V. & Buranaprapuk, A. Chiral Photochemical Scissors Targeting Proteins. in 400 (World Scientific Publishing Company, 2018).

101. Métifiot, M., Amrane, S., Litvak, S. & Andreola, M.-L. G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.* **42**, 12352–12366 (2014).
102. Perez-Iratxeta, C. & Andrade-Navarro, M. A. K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol.* **8**, 25 (2008).
103. Pong, W.-L., Huang, Z.-S., Teoh, P.-G., Wang, C.-C. & Wu, H.-N. RNA binding property and RNA chaperone activity of dengue virus core protein and other viral RNA-interacting proteins. *FEBS Lett.* **585**, 2575–2581 (2011).
104. 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**, 1046–1058 (2012).
105. Byk, L. A. & Gamarnik, A. V. Properties and Functions of the Dengue Virus Capsid Protein. *Annu. Rev. Virol.* **3**, 263–281 (2016).
106. Patkar, C. G., Jones, C. T., Chang, Y., Warrier, R. & Kuhn, R. J. Functional requirements of the yellow fever virus capsid protein. *J. Virol.* **81**, 6471–6481 (2007).
107. Peng, Z. *et al.* Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **72**, 137–151 (2015).
108. Fan, X. *et al.* The intrinsic disorder status of the human hepatitis C virus proteome. *Mol. Biosyst.* **10**, 1345–1363 (2014).
109. Figueira-Mansur, J., Estefania, A. A., Stoque, R. M. & Ventura, G. T. Mutations in the dimer interfaces of the dengue virus capsid protein affect structural stability and impair RNA-capsid interaction. *Sci. Rep.* **9**, 2829 (2019).
110. Fernandez-Sesma, A. *et al.* Influenza Virus Evades Innate and Adaptive Immunity via the NS1 Protein. *J. Virol.* **80**, 6295–6304 (2006).
111. Wang, S.-H. *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).
112. Wagstaff, K. M., Sivakumaran, H., Heaton, S. M., Harrich, D. & Jans, D. A. Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. *Biochem. J.* **443**, 851–856 (2012).
113. Bergmann, M. *et al.* Influenza Virus NS1 Protein Counteracts PKR-Mediated Inhibition of Replication. *J. Virol.* **74**, 6203–6206 (2000).
114. Kochs, G., Garcia-Sastre, A. & Martinez-Sobrido, L. Multiple Anti-Interferon Actions of the Influenza A Virus NS1 Protein. *J. Virol.* **81**, 7011–7021 (2007).
115. Rodriguez-Madoz, J. R., Bernal-Rubio, D., Kaminski, D., Boyd, K. & Fernandez-Sesma, A. Dengue Virus Inhibits the Production of Type I Interferon in Primary Human Dendritic Cells. *J. Virol.* **84**, 4845–4850 (2010).
116. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

# 6 Appendix

Complementary information, not essential for understanding the main body of text, is found here. This concerns mainly results and supporting information that complement the data found in the above chapters. In addition, the outcomes of this work are shown as abstracts of the results, which were already presented through meetings and original articles, recently submitted. A list of the material found in the appendix is found ahead.

## Appendix – list of materials

Appendix I: Supplementary Information (SI)



Appendix II: Current outcomes of the work presented – Poster communication

## Appendix I: Supplementary Information (SI)

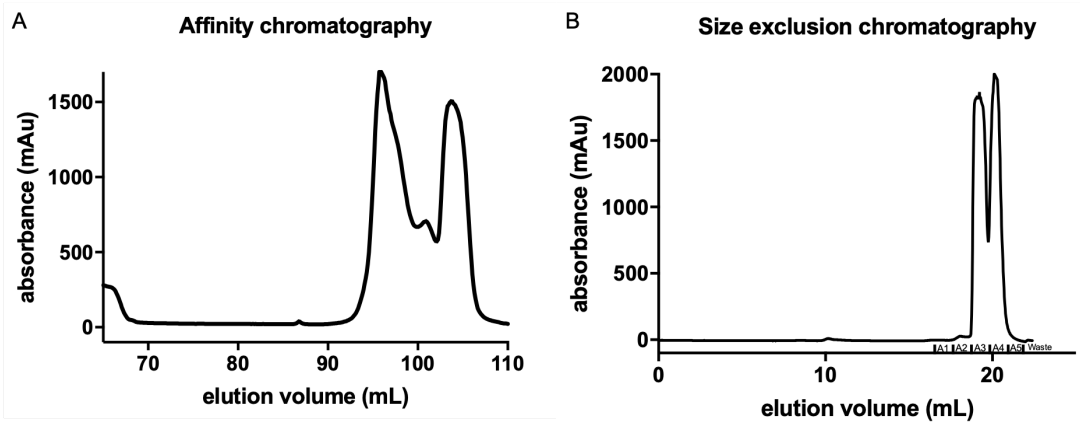


**Figure SI1: Chromatogram of the first recombinant DENV C purification. A**) Affinity chromatography (heparin column) of the sonicated cell lysate supernatant. Samples corresponding to DENV C eluted in the second peak between. **B**) Size exclusion chromatography (S75 column) of samples collected during the affinity chromatography. Samples corresponding to pure DENV C eluted in the second peak. Aliquots of 1.5 mL were collected.

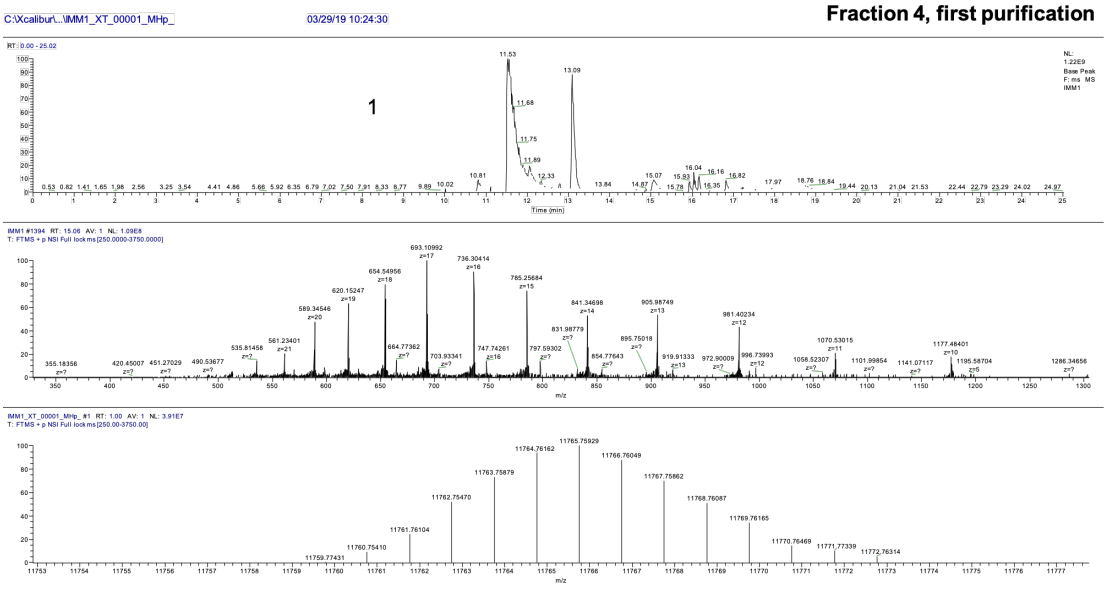


**Figure SI2: Result of the mass spectrometry analysis of fraction 4, first purification.** This fraction showed the expected result of the mass spectrometry analysis and was used to perfom the experiments with DNA.

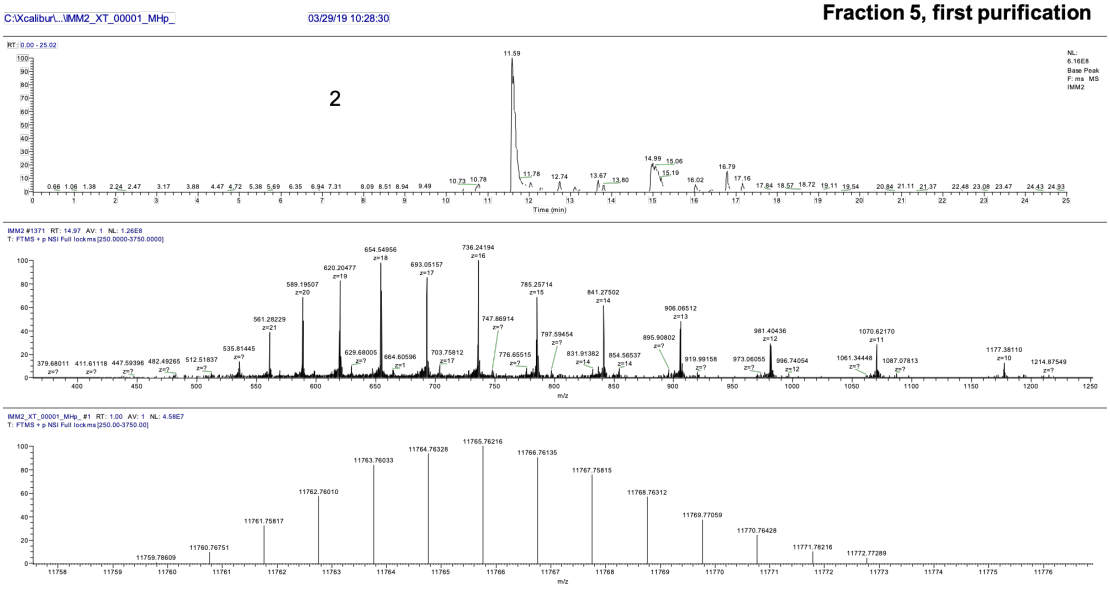


**Figure SI3: Result of the mass spectrometry analysis of fraction 5, first purification.** This fraction showed the expected result of the mass spectrometry analysis, but was only used for preliminary tests.

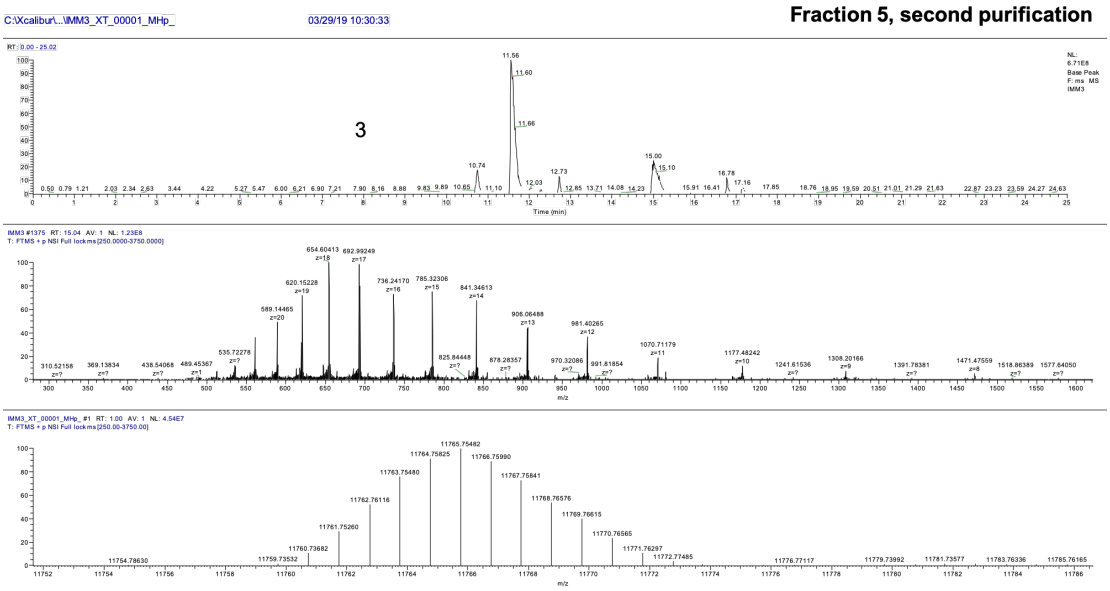


**Figure SI4: Result of the mass spectrometry analysis of fraction 5, second purification.** This fraction showed the expected result of the mass spectrometry analysis, but was only used for preliminary tests.

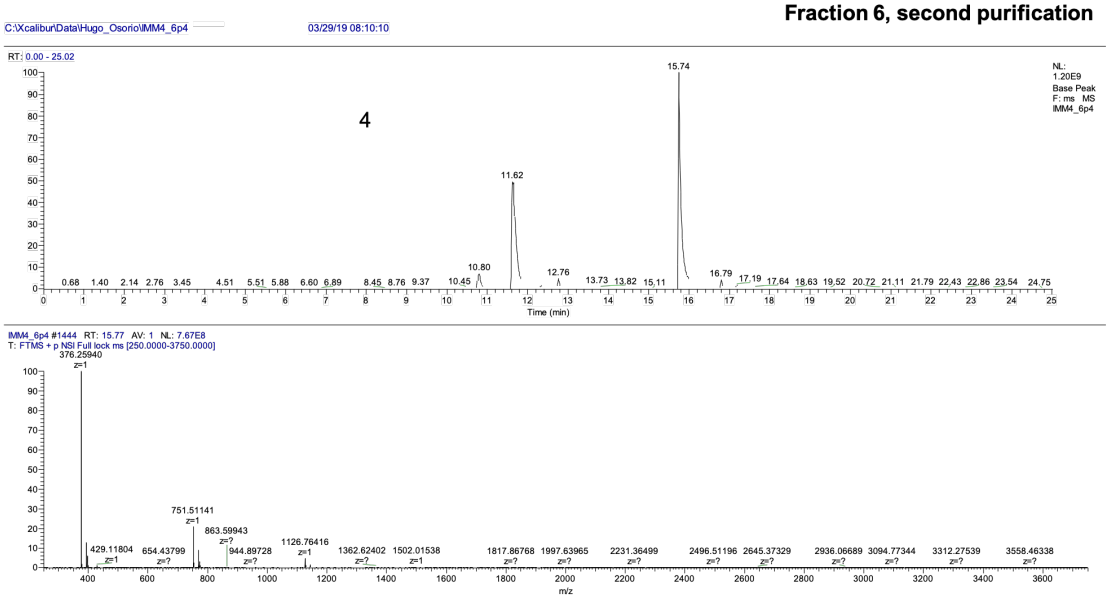


**Figure SI5: Result of the mass spectrometry analysis of fraction 6, second purification.** This fraction does not show the expected mass of DENV C, thus it was not used for experiments.

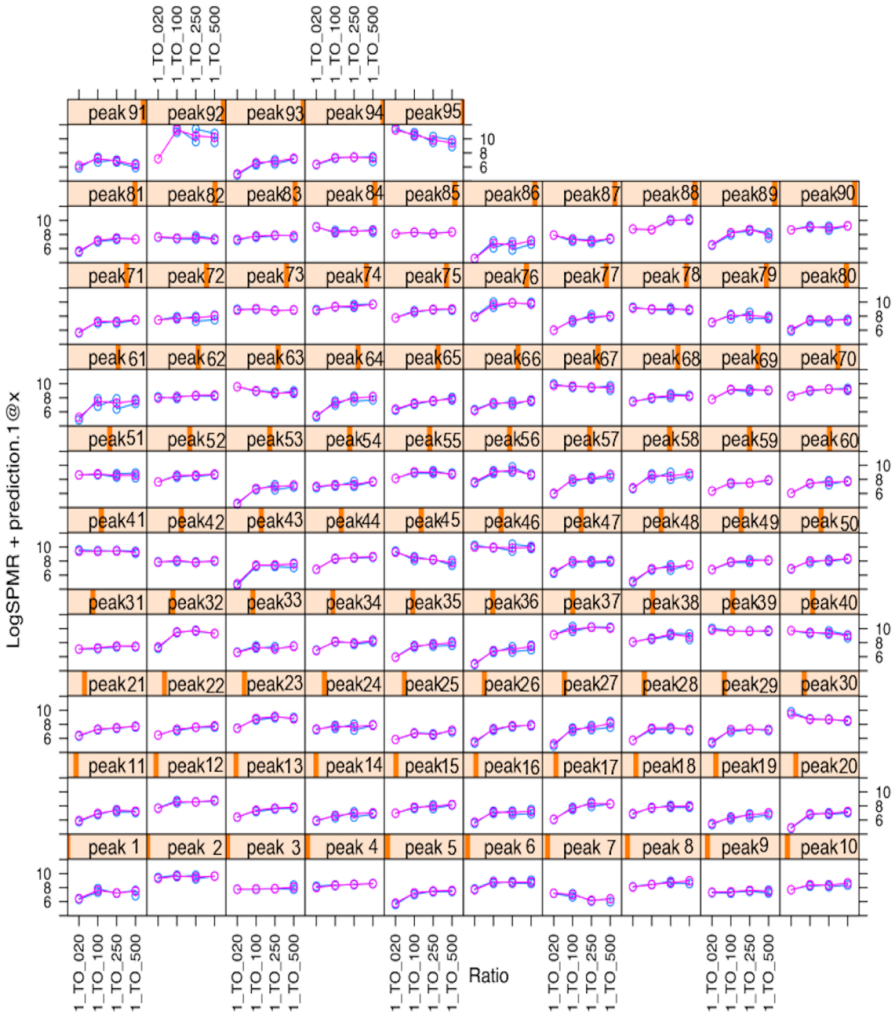


**Figure SI6: Binding profiles of tested sequences of the viral genome binding to DENV C.** The figure shows the 95 peaks previously identified by the host laboratory, through a collaboration with the Bioinformatics Institute A*STAR, Singapore, and used to select the ssDNA sequences.

The peaks can be analysed with the help of the following information:

Binding affinity: High – all four points are relatively flat
Low – the four points show a discernible positive slope

Signal strength: High – most of the points lie between 9 – 10 log signal per million reads (SPMR)
Medium – most of the points lie around 8 log SPMR
Low – most of the points lie below 7 log SPMR

*Recap for the figure. The blue dots/line represents the log (normalized peak height) of the raw data. The pink dots/line represents the model obtained using the generalized linear mixed effect model that was used to fit the data and obtain p-values. The model and the raw data fit well.

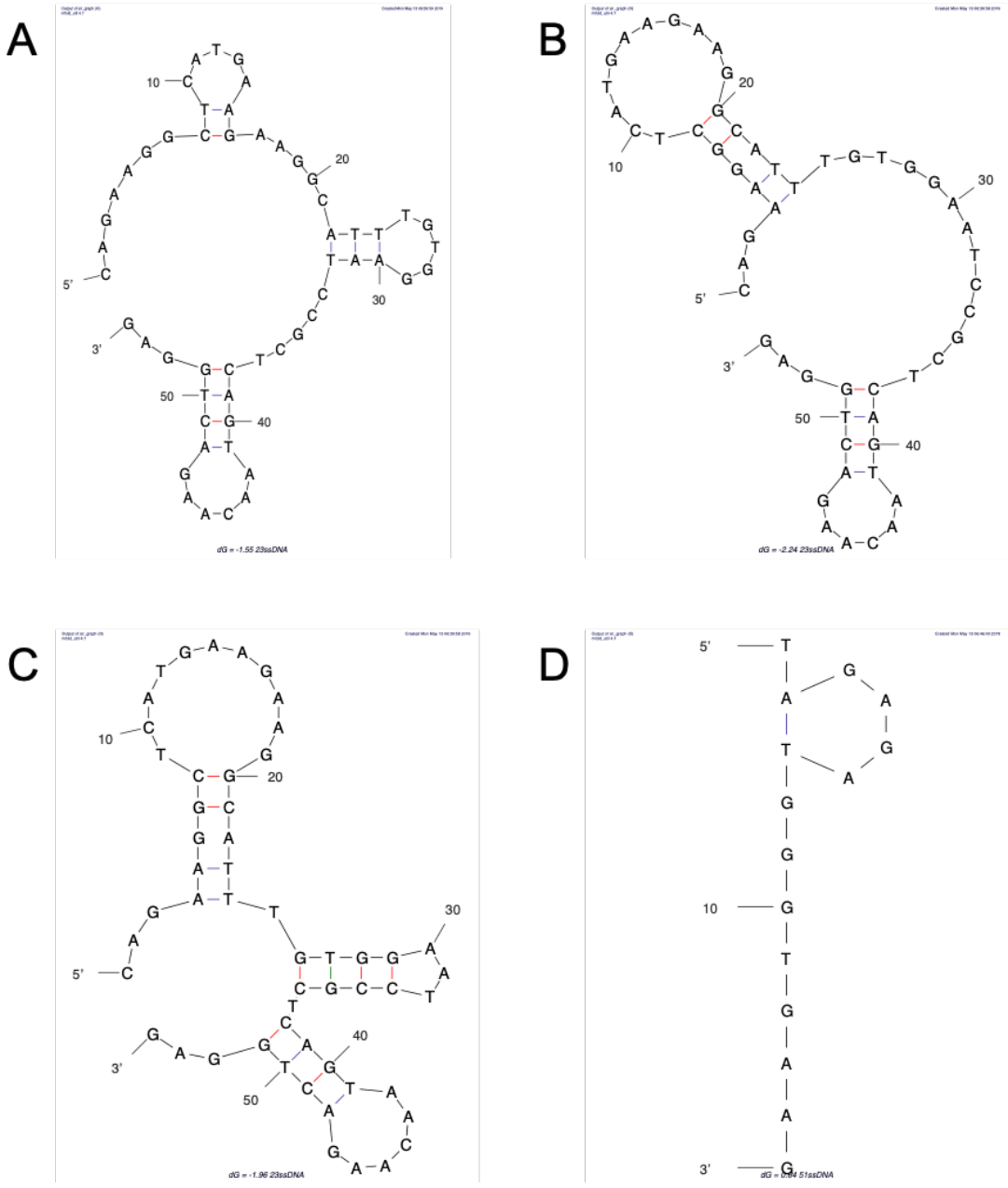


**Figure SI7: Secondary structure prediction of ssDNA. A**) Structure prediction of 23ssDNA, alternative A. **B**) Structure prediction of 23ssDNA, alternative B. **C**) Structure prediction of 23ssDNA, alternative C. **D**) Structure prediction of 51ssDNA. Structure predictions were done by using the mfold web server[116].

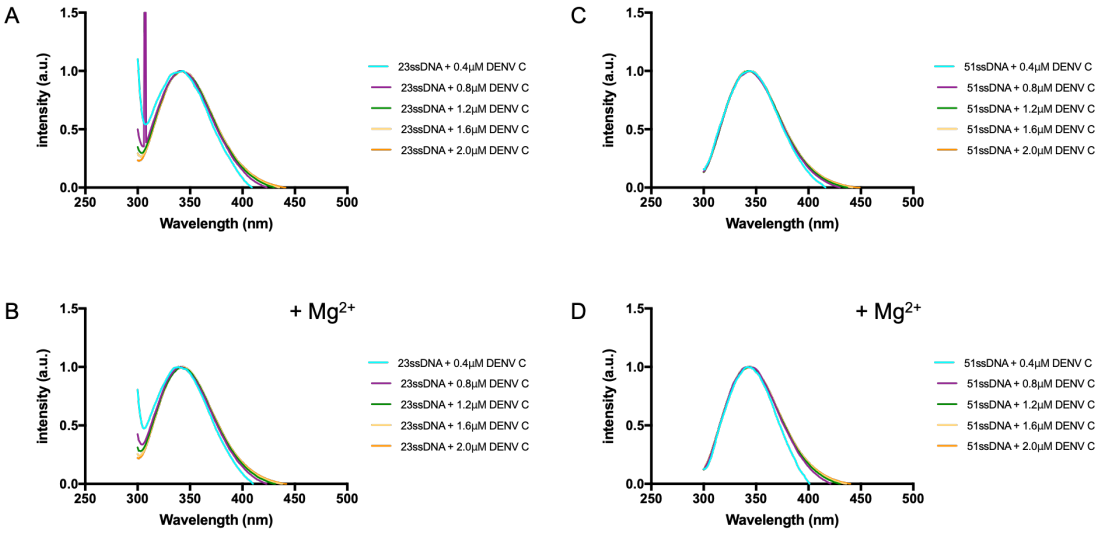


**Figure SI8: DENV C-DNA interaction – fluorescence spectroscopy, normalized data.** Titration of DENV C (up to 2 μM) into two ssDNA samples, in a buffer of 50 mM $KH_2PO_4$, pH 6.0, 200 mM KCl, with and without $Mg^{2+}$ ions present (20 mM $MgO_4S$). **A**) DENV C-23ssDNA interaction, without $Mg^{2+}$. **B**) DENV C-23ssDNA interaction, with $Mg^{2+}$. **C**) DENV C-51ssDNA interaction, without $Mg^{2+}$. **D**) DENV C-51ssDNA interaction, with $Mg^{2+}$. No major changes are seen.

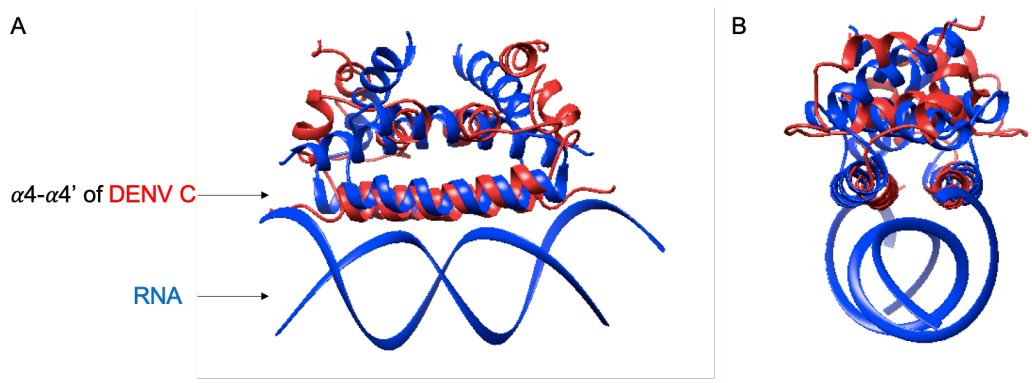


**Figure SI9: Comparison of DENV C and influenza NS1 protein structures.** DENV C (red) and influenza NS1 (blue) superimposed. The figure is shown with influenza NS1 in the RNA-bound form. **A**) Superimposing of the α4-α4' part of DENV C with the similar motif of influenza NS1. **B**) Different angle of the structure comparison. It can be seen that the α4-α4'-like structure of influenza NS1 is a somewhat more open, when bound to nucleic acids.

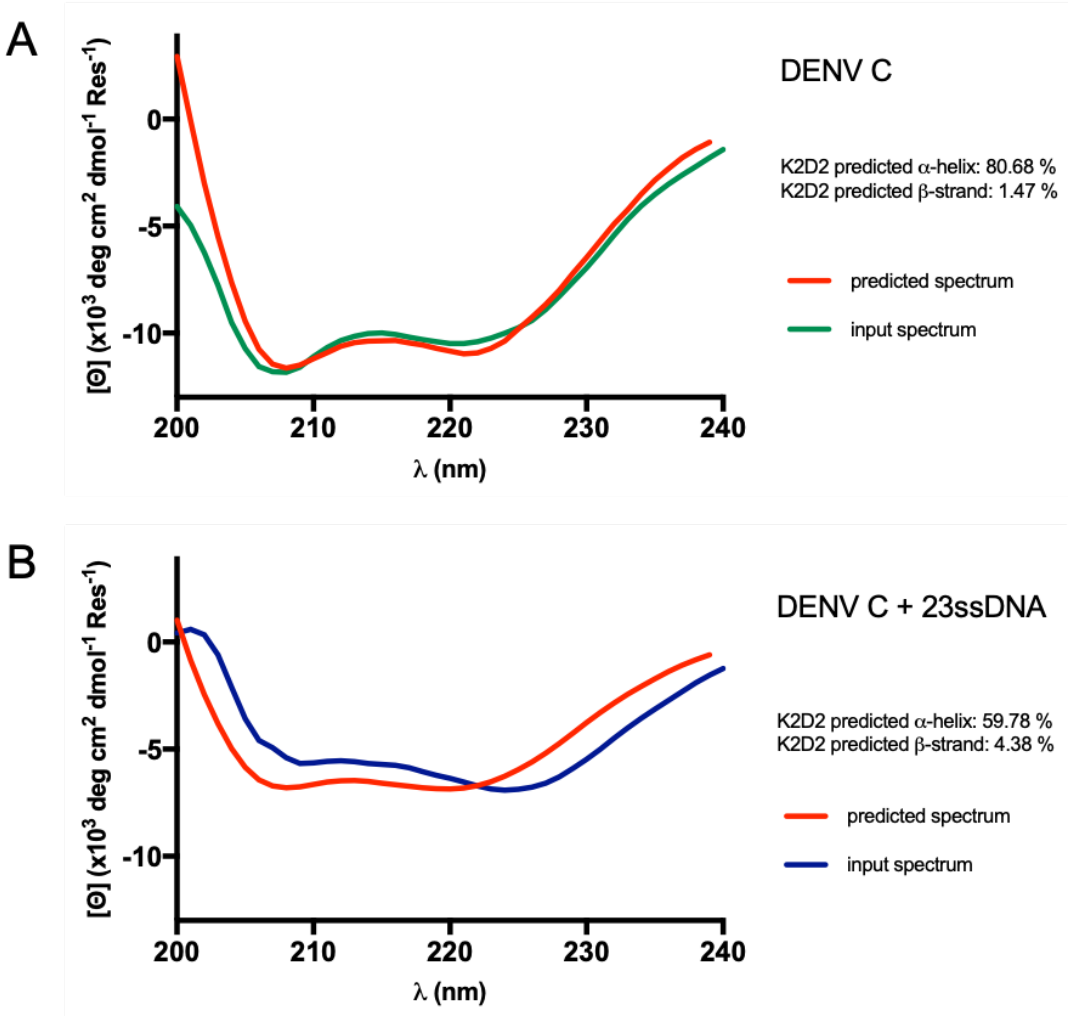


**Figure SI10: K2D2 estimation of DENV C secondary structure content from CD spectra.** Predicted spectrum of 0.8 μM DENV C (red) using K2D2 algorithm and CD data between 200 and 240 nm, without (**A**) and with 1 μM 23ssDNA (**B**). The original input spectra are also shown. The original input spectra are also shown. The input and predicted spectra superimpose well for **A**. In **B** the input and predicted K2D2 spectra do not superimpose perfectly, but the amount of α-helical structure could still be estimated, showing a decrease upon interaction. 51ssDNA gives a similar profile as the protein alone (not shown).

## Appendix II: Current outcomes of the work presented

Poster communication: National colloquium – Chemistry: Shaping the future & 2019 Summer School (4ECQUL), Lisbon, Portugal

**Date and venue:** July 16 – 18 2019, Lisbon, Portugal

**Title:** **Biophysical studies of dengue virus capsid protein interaction with nucleic acids**

**Authors:** **Nina Karguth**, Ana S. Martins, Nuno C. Santos, Ivo C. Martins

**Affiliation:** Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisbon, Portugal

**Abstract:** Dengue virus (DENV) is an increasing global threat due to the expansion of its mosquito vector. New effective therapies require full understanding of the viral life cycle, at the level of the molecular mechanisms governing protein-ligand interactions, since this knowledge can trigger progresses in basic and applied research. The viral capsid (C) protein is of special importance, due to its interaction with different ligands, mediating key steps of the viral life cycle. DENV C is also a structural protein involved in the viral assembly, binding and helping to encapsulate the viral genome. However, the molecular details of DENV C binding to nucleic acids are not yet clear. DENV C contains a disordered N-terminal region, an intermediate flexible fold section and a very stable conserved fold region. We previously studied the C protein interaction with host lipoproteins and intracellular lipid droplets. This led to understanding of the structural and functional roles of DENV C N-terminal disordered domain and, furthermore, to the design of pep14-23, a novel patented viral inhibitor peptide. We also identified specific sections of the viral genome to which DENV C specifically binds. To complement this knowledge, here we characterized DENV C binding to these sections via biophysical approaches, using analogous DNA sequences as proxies of the viral genome. We identified different binding interactions with different DNA sequences. This may enable, later on, the targeting of the most relevant sequences, seeking to inhibit these interactions. Together with the understanding of DENV C binding to host lipid systems, such crucial knowledge regarding DENV C interaction with nucleic acids may yield promising advances in therapies against dengue and closely related flaviviruses.

Oral communication - Flashtalk: NOVA Biophysica, Lisbon, Portugal

**Date and venue:** September 4 – 6, 2019, Lisbon, Portugal

**Title:** **Dengue virus capsid protein and its interaction with nucleic acids**

**Authors:** **Nina Karguth**, Ana S. Martins, Nuno C. Santos, Ivo C. Martins

**Affiliation:** Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisbon, Portugal

**Abstract:** Dengue virus (DENV) is an increasing global threat due to the expansion of its mosquito vector. New effective therapies require full understanding of the viral life cycle, at the level of the molecular mechanisms governing protein-ligand interactions, since this knowledge can trigger progresses in basic and applied research. The viral capsid (C) protein is of special importance, due to its interaction with different ligands, mediating key steps of the viral life cycle. DENV C is also a structural protein involved in the viral assembly, binding and helping to encapsulate the viral genome. However, the molecular details of DENV C binding to nucleic acids are not yet clear. DENV C contains a disordered N-terminal region, an intermediate flexible fold section and a very stable conserved fold region. We previously studied the C protein interaction with host lipoproteins and intracellular lipid droplets. This led to a better understanding of key structural and functional roles of DENV C N-terminal disordered domain. It also led to the design of pep14-23, a novel patented viral inhibitor peptide. Recently, we identified sections of the viral genome to which DENV C specifically binds. To complement this, using analogous single-strand DNA sequences as proxies of the viral genome, here we test DENV C binding to nucleic acid sequences. Biophysical analysis via circular dichroism and intrinsic fluorescence spectroscopy studies supports the binding to specific structured single-strand sequences. This may enable to, later on, target relevant nucleotide sequences, seeking to inhibit DENV C binding to the viral RNA. Together with understanding DENV C binding to host lipid systems, such crucial knowledge regarding DENV C interaction with nucleic acids may thus yield promising advances in therapies against dengue and closely related flaviviruses.

Published articles:

**Title:** **Structural and functional properties of the capsid protein of Dengue and related *Flavivirus***

**Authors:** André F. Faustino[1,#], Ana S. Martins[1], **Nina Karguth**[1], Vanessa Artilheiro[1], Francisco J. Enguita[1], Joana C. Ricardo[2,§], Nuno C. Santos[1,*] and Ivo C. Martins[1,*]

**Affiliation:** [1] Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisbon, Portugal
[2] Centro de Química-Física Molecular, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[#]Present address: Instituto de Biologia Experimental e Tecnológica (iBET), Apartado 12, 2780-901 Oeiras, Portugal
[§]Present address: Department of Biophysical Chemistry, J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Dolejskova 3, 182 23 Prague 8, Czech Republic
*Correspondence: ivomartins@fm.ul.pt; Tel.: +351217999476; nsantos@fm.ul.pt; Tel.: +351217999480;

**Current state:** published: Int J Mol Sci. 2019 Aug 8;20(16).

**Abstract:** Dengue, West Nile and Zika, closely related viruses of the *Flaviviridae* family, are an increasing global threat, due to the expansion of their mosquito vectors. They present a very similar viral particle with an outer lipid bilayer containing two viral proteins and, within it, the nucleocapsid core. This core is composed by the viral RNA complexed with multiple copies of the capsid protein, a crucial structural protein that mediates not only viral assembly, but also encapsidation, by interacting with host lipid systems. The capsid is a homodimeric protein that contains a disordered N-terminal, an intermediate flexible fold section and a very stable conserved fold region. Since a better understanding of its structure can give light into its biological activity, here, first, we compared and analyzed relevant mosquito-borne *Flavivirus* capsid protein sequences and their predicted structures. Then, we studied the alternative conformations enabled by the N-terminal region. Finally, using Dengue virus capsid protein as main model, we correlated the protein size, thermal stability and function with its structure/dynamics features. The findings suggest that the capsid protein interaction with host lipid systems leads to minor allosteric changes that may modulate the specific binding of the protein to the viral RNA. Such mechanism can be targeted in future drug development strategies, namely by using improved versions of pep14-23, a Dengue virus capsid protein peptide inhibitor, previously developed by us. Such knowledge can yield promising advances against Zika, Dengue and closely related *Flavivirus*.

International Journal of
*Molecular Sciences*

MDPI

*Article*

# Structural and Functional Properties of the Capsid Protein of Dengue and Related *Flavivirus*

**André F. Faustino [1,†], Ana S. Martins [1], Nina Karguth [1], Vanessa Artilheiro [1], Francisco J. Enguita [1], Joana C. Ricardo [2,‡], Nuno C. Santos [1,*] and Ivo C. Martins [1,*]**

1 Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisbon, Portugal
2 Centro de Química-Física Molecular, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal
* Correspondence: nsantos@fm.ul.pt (N.C.S.); ivomartins@fm.ul.pt (I.C.M.); Tel.: +351-217-999-480 (N.C.S.); +351-217-999-476 (I.C.M.)
† Present address: Instituto de Biologia Experimental e Tecnológica (iBET), Apartado 12, 2780-901 Oeiras, Portugal.
‡ Present address: Department of Biophysical Chemistry, J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Dolejškova 3, 182 23 Prague 8, Czech Republic.

Received: 21 June 2019; Accepted: 6 August 2019; Published: 8 August 2019


**Abstract:** Dengue, West Nile and Zika, closely related viruses of the Flaviviridae family, are an increasing global threat, due to the expansion of their mosquito vectors. They present a very similar viral particle with an outer lipid bilayer containing two viral proteins and, within it, the nucleocapsid core. This core is composed by the viral RNA complexed with multiple copies of the capsid protein, a crucial structural protein that mediates not only viral assembly, but also encapsidation, by interacting with host lipid systems. The capsid is a homodimeric protein that contains a disordered N-terminal region, an intermediate flexible fold section and a very stable conserved fold region. Since a better understanding of its structure can give light into its biological activity, here, first, we compared and analyzed relevant mosquito-borne *Flavivirus* capsid protein sequences and their predicted structures. Then, we studied the alternative conformations enabled by the N-terminal region. Finally, using dengue virus capsid protein as main model, we correlated the protein size, thermal stability and function with its structure/dynamics features. The findings suggest that the capsid protein interaction with host lipid systems leads to minor allosteric changes that may modulate the specific binding of the protein to the viral RNA. Such mechanism can be targeted in future drug development strategies, namely by using improved versions of pep14-23, a dengue virus capsid protein peptide inhibitor, previously developed by us. Such knowledge can yield promising advances against Zika, dengue and closely related *Flavivirus*.

**Keywords:** Dengue virus (DENV); capsid protein (C protein); *Flavivirus*; intrinsically disordered protein (IDP); protein–RNA interactions; protein–host lipid systems interaction; circular dichroism; time-resolved fluorescence anisotropy

## 1. Introduction

Viral hemorrhagic fever is a global problem, with most cases due to dengue virus (DENV), which originates over 390 million infections per year worldwide, being a major socio-economic burden, mainly for tropical and subtropical developing countries [1]. A working vaccine was registered in Mexico in December 2015, approved for official use in some endemic regions of Latin America and Asia and, as of October 2018, also in Europe [2–4]. However, this vaccine is not 100% effective against all

DENV serotypes. Thus, research into new prophylactics is still ongoing, with a new vaccine proposed recently being now in phase 3 clinical trials [5]. In spite of these recent developments, fully effective prophylactics approaches are lacking and there are no effective therapies. This is in part, due to a poor understanding of key steps of the viral life cycle.

There are four dengue serotypes occurring: DENV-1, DENV-2, DENV-3 and DENV-4 [6]. Here, if not otherwise indicated, DENV refers to DENV-2. DENV is a member of the *Flavivirus* genus, part of the Flaviviridae family, a genus which comprises 53 viral species [6]. Many of these are important human pathogens as well, such as hepatitis C (HCV), tick-borne encephalitis (TBEV), yellow fever (YFV), West Nile (WNV) and Zika (ZIKV) viruses [6–9]. Flaviviridae are single-stranded positive-sense RNA viruses with approximately 11 kb, containing a single open reading frame [10]. Using the host cell translation machinery, the *Flavivirus* RNA genome is translated into a polyprotein that is co- and post-translationally cleaved by cellular and viral proteases into three structural proteins and seven non-structural proteins [10]. Structural proteins are named as such since they are present in the mature virion structure [11]. Nevertheless, they may also have non-structural roles, such as the capsid (C) protein. This is a structural protein that also mediates viral assembly and encapsidation, crucial steps of the viral life cycle. Given the C protein key roles, it is the focus of this work and will be described in detail below.

DENV C contains 100 amino acid residues, which form an homodimer with an intrinsically disordered protein (IDP) region in the N-terminal followed by four α-helices, α1 to α4, per monomer [12]. Overall, the main structural/dynamics regions consist of the disordered N-terminal, a short flexible intermediate fold and, finally, a large conserved fold region, which greatly stabilizes the protein homodimer structure [12–16]. The C protein has an asymmetric charge distribution: one side of the dimer contains a hydrophobic pocket (α2–α2′ interface), responsible for, alongside the disordered N-terminal, the binding to host lipid droplets (LDs) [12–16]. The other is the positively charged C-terminal side (α4–α4′ interface), proposed to mediate the C protein binding to the viral RNA [12]. It is noteworthy that several transient conformations for DENV C N-terminal were proposed, which may help modulate DENV C interaction with host lipid systems, via an autoinhibition mechanism [15].

DENV infection affects the host lipid metabolism, increasing host intracellular LDs and unbalancing plasma lipoprotein levels and composition [17–19]. Importantly, DENV C binds LDs, an interaction essential for viral replication [18,20]. DENV C-LDs binding requires potassium ions, the LDs surface protein perilipin 3 (PLIN3) and involves specific amino acid residues of DENV C α2–α2′ helical hydrophobic core and of the N-terminal [14,20]. This knowledge led us to design pep14-23, a patented peptide, based on a *Flavivirus* C protein conserved N-terminal motif. We then established that pep14-23 inhibits DENV C-LDs binding [14], acquiring α-helical structure in the presence of anionic phospholipids [15]. Moreover, we also found that DENV C binds specifically to very low-density lipoproteins (VLDL), requiring $K^+$ ions and a specific VLDL surface protein, apolipoprotein E (APOE), being also inhibited by pep14-23 [21]. This is analogous to DENV C-LDs interaction. The similarities between APOE and PLIN3 further reinforce this, suggesting a common mechanism [22]. The role of LDs in *Flavivirus* infection is well known and has been recently reviewed [14,18,20,23–25]. Given that, pep14-23 is an excellent drug development lead. Further developments require a better understanding of the function of the C protein of dengue and of *Flavivirus* in general.

Therefore, here, we seek to contribute to understand the C proteins biological activity, with a special focus on DENV C. Briefly, we studied DENV C structure-activity relationship in the context of similar and highly homologous mosquito-borne *Flavivirus* C proteins. Our findings shed light into the structure-function relationship behind the C protein biological roles, which may contribute to future therapeutic approaches against DENV and closely related *Flavivirus*.

## 2. Results

### *2.1. Analysis of Amino Acid Sequence Conservation Among Flavivirus C proteins*

A phylogenetic analysis of the *Flavivirus* C protein and the polyprotein amino acid residue sequences reveals if the C protein is an indicator of phylogenetic similarity (Figure 1). C proteins of Spondweni group viruses, i.e., ZIKV, Spondweni virus (SPOV) and Kedougou virus (KEDV), cluster together, being the most similar to DENV (Figure 1a). Another cluster corresponds to mosquito-borne encephalitis-causing *Flavivirus*: Saint Louis encephalitis (SLEV), WNV, WNV serotype Kunjin (WNV-K), Alfuy (ALFV), Murray Valley encephalitis (MVEV), Usutu (USUV) and Japanese encephalitis (JEV) viruses. The *Flavivirus* polyproteins sequences show similar clusters (Figure 1b). As such, the C protein is a good indicator of viral genetic similarity. Thus, we investigated the C protein amino acid sequences, seeking common patterns relevant to biological activity.

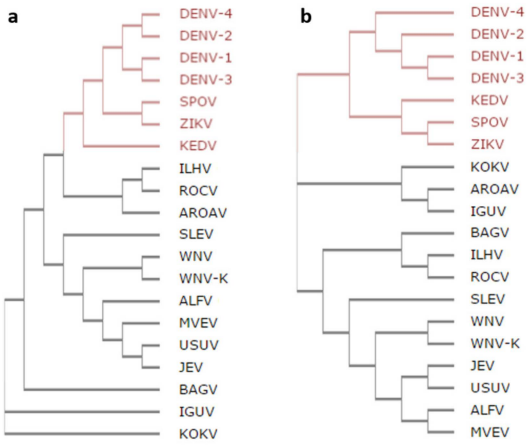


**Figure 1.** *Flavivirus* phylogenetic trees. Phylogenetic trees of (**a**) *Flavivirus* C proteins, highlighting in red the viruses with the C protein most similar to dengue virus (DENV) C (Spondweni group viruses (ZIKV), Spondweni virus (SPOV) and Kedougou virus (KEDV)) and of the (**b**) entire viral polyproteins of the same *Flavivirus*. Overall, despite some differences, the same general clusters are seen regardless of the clustering being based on the polyprotein or the capsid protein.

The amino acid residues sequences of the *Flavivirus* C proteins identified above were analyzed in the context of the three main regions identified in DENV C sequence, *i.e.*, the conserved fold region, the flexible fold region and the N-terminal IDP region (Figure 2). This was done for all mosquito-borne *Flavivirus* relevant for human diseases (Figure 2a), as well as for the four main DENV C serotypes (Figure 2b). For this, the 16 mosquito-borne *Flavivirus* and the 4 DENV serotypes amino acid sequence of the C protein are jointly aligned. In agreement with previous work [12,14], five conserved motifs are found in the mosquito-borne *Flavivirus* C proteins and deserve attention, namely: the N-terminal conserved $^{13}$hNML+R$^{18}$; $^{40}$GXGP$^{43}$ in loop L1-2; $^{44}$h+hhLAhhAFF+F$^{56}$ in α2 helix; $^{68}$RW$^{69}$ of α3 helix; and, finally, the $^{84}$F++–h$^{88}$ motif from α4 (with 'h', '+' and '–' representing hydrophobic, positively charged and negatively charged residues, respectively). Between residues 70–100, other motifs, not previously reported and containing hydrophobic and positively charged residues, are visible. Moreover, amino acid residues G and P, that can break the continuity of α-helices, are conserved in specific positions of the protein, especially in the disordered N-terminal and the flexible fold regions (Figure 2c). Charged residues are also conserved in specific locations. They are mostly in the conserved fold region, especially after position 95 (Figure 2d). Overall, the disordered N-terminal and the flexible fold regions, when compared with the conserved fold region, have an average of, respectively, 10 versus 4 G and P residues (Figure 2c), green, 10 versus 15 K and R residues (Figure 2d), blue, and 1 versus 2 D and E residues (Figure 2d), magenta.

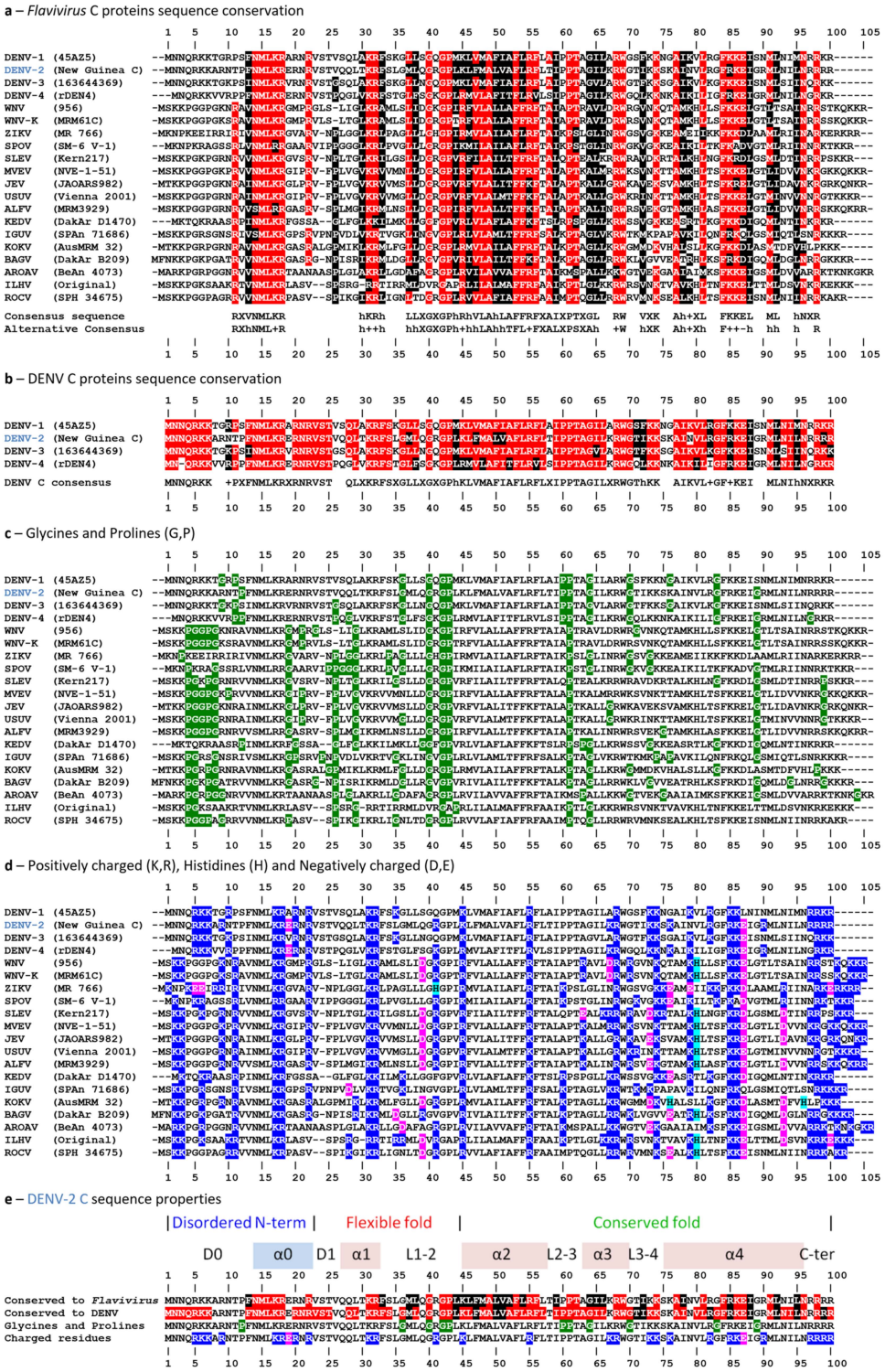


**Figure 2.** *Flavivirus* C proteins amino acid residues sequence conservation. (**a**) Mosquito-borne *Flavivirus* C protein are 55% conserved, with residues being considered conserved if, in a given position, more than 15 are equal (red) or stereochemically similar (black). (**b**) Conservation between DENV serotypes is 80%, with the same criteria as in (**a**). (**c**) Structure-breaking residues G and P (green). (**d**) Charged residues: dark blue for positively charged residues (K and R), light blue for H, and magenta for negatively charged residues (D or E). (**e**) Overall conserved regions of *Flavivirus* C proteins: the disordered N-terminal and the conserved fold are clearly conserved in terms of charged and G/P amino acids. In contrast, the flexible fold region allows higher variability. Thus, its main role seems to be to connect the disordered N-terminal and the conserved fold regions, and to enable alternative conformations. DENV C serotype 2 is highlighted in blue, with amino acid residues numbered according to its sequence. Amino acid residues are numbered according to the consensus, coinciding with DENV-2 residues numbers. The viruses' full designation is found in the abbreviations section.

Several motifs in the *Flavivirus* C protein sequences can be identified. These represent the main sections of the protein, conserved during evolution as these must be crucial to protein function (Figure 2e). The N-terminal region, although disordered, is highly conserved, in terms of charged amino acid and G/P residues. The flexible fold section allows greater variability, in line with previous reports by us and others, suggesting that it can adopt several conformations [15].

### 2.2. Analysis of the Flavirus C Protein Sequences Hydrophobicity and Secondary Structure Propensity

Hydrophobicity and α-helical propensity predictions were performed as previously reported [15], using the Kite-Doolittle [26] and the Deleage-Roux [27] scales on ProtScale server, respectively, for the 16 mosquito-borne *Flavivirus* C proteins analyzed (Figure 3). The hydrophobicity scale ranges from −4.5, for highly polar amino acids (hydrophilic), to 4.5, for highly hydrophobic amino acid residues [26]. Therefore, when plotting the average values for each amino acid residue of the *Flavivirus* C sequences, negative local minima and positive local maxima indicate, respectively, hydrophilic and hydrophobic regions (Figure 3a,b). All proteins display a similar profile even in the N-terminal and flexible fold regions despite the slightly higher amino acid residues variability (Figure 2). The α0 domain, homologous to pep14-23, is amphipathic, with average values near 0. In the flexible fold region, which is mostly amphipathic too, there is a peak of hydrophobicity between residues 30 and 40, possibly explaining its intermediate structure/dynamics behavior [13,14]. Some peaks of hydrophobicity are observed in the α3 and α4 domains, with the most hydrophobic domain being α2, as expected from the sequence analysis (Figure 2) and from the literature [12,14,18].

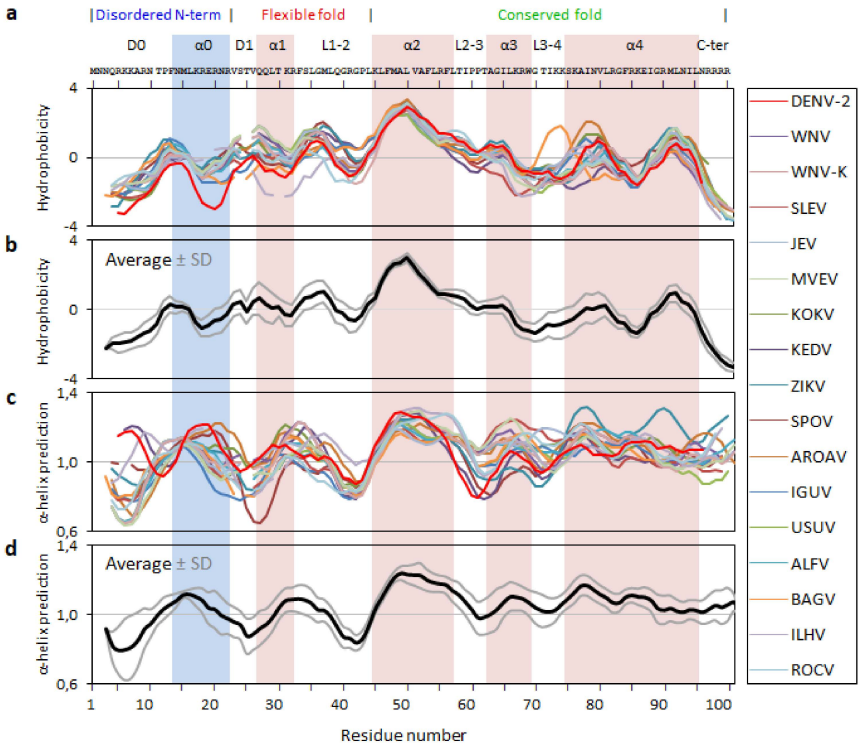


**Figure 3.** *Flavivirus* C proteins hydrophobicity and secondary structure predictions. (**a**) Hydrophobicity predictions and (**b**) respective average (black line) ± standard deviation, SD (gray lines). (**c**) α-helical secondary structure predictions and (**d**) respective average (black line) ± SD (gray lines). Amino acid residues are numbered according to the consensus, coinciding with DENV 2 residues numbers.

For α-helical predictions secondary structure is highly probable above a threshold of 1.0 [27]. *Flavivirus* C proteins secondary structure predictions correlate well with the known secondary structure of DENV C (Figure 2e) [12]. Such agreement supports the concept of a transient α0 occurring for these proteins, as hypothesized earlier [15]. Roughly, between positions 12 to 20, occurs a disordered region

with high tendency to acquire α-helical secondary structure. Importantly, the values of the predictions are similar and the same tendencies are found in all proteins, with peaks and valleys co-localizing (Figure 3). Along with data from the last subsection, these results strengthen the idea that *Flavivirus* C proteins have similar structure and dynamics properties.

### 2.3. Analysis of the Flavivirus C Protein Tertiary Structure Propensity

*Flavivirus* C proteins tertiary structure was then investigated, complementing the α-helical predictions, to help understanding the disordered N-terminal region role(s). Following previous work [15], I-TASSER [28–30] was used to predict tertiary structures for the 16 closely related mosquito-borne *Flavivirus* C proteins (Figure 4). Eighty monomer conformations were obtained (several for each sequence) and superimposed with the DENV C homodimer partial structure deposited at the Protein Data Bank (PDB) and obtained via nuclear magnetic resonance (NMR) spectroscopy (PDB ID: 1R6R). Noteworthy, DENV [12,16], WNV [31] and ZIKV [25] C proteins form homodimers, stabilized by hydrophobic and electrostatic interactions involving their conserved fold region [12–14,25,31–33]. Since this is the most conserved region of *Flavivirus* C proteins sequences (Figure 2), a homodimer is thus not only a stable conformational arrangement, but also likely to occur. Thus, as 28 conformers had more than 5 backbone clashes with the other monomer when superimposed in a homodimer structure (not allowing a viable homodimer), those conformers were discarded Table 1. The remaining 52 *Flavivirus* C proteins conformational models were analyzed, while superimposed with DENV C homodimer (PDB ID: 1R6R, model 21 [12]). These were then grouped into four clusters by visual inspection of their similarity (Figure 4).

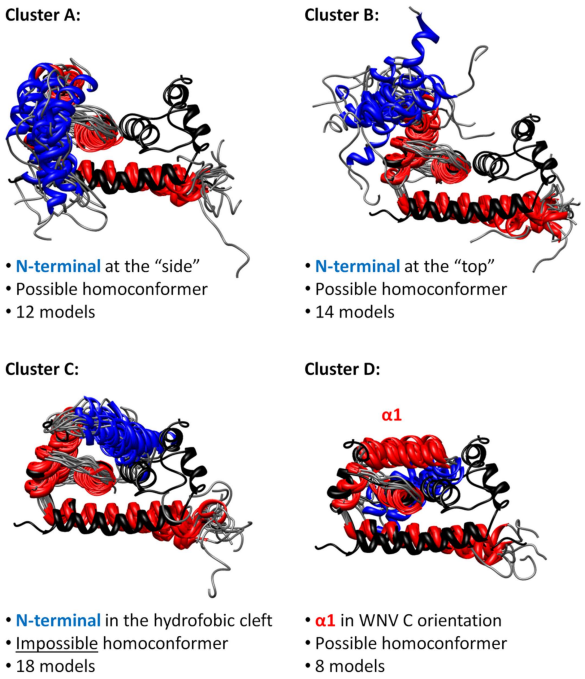


**Figure 4.** *Flavivirus* C proteins tertiary structure predictions, organized into four conformational clusters. The *Flavivirus* C proteins conformations predicted by I-TASSER are superimposed with DENV C experimental homodimer structure (black). Amino acid residues of the N-terminal region in α-helix conformation are in blue, the other α-helices in red and the loops in gray. From the 80 conformers, 52 can be clustered by similarity of conformations, from cluster A to D. Clusters A, B and C have the α1 helix in the DENV C experimentally determined conformation (Protein Data Bank (PDB) ID: 1R6R [12]). In cluster D the α1 is in West Nile Virus (WNV) C and ZIKV C conformation (PDB IDs: 1SFK [31] and 5YGH [25], respectively). The closed autoinhibitory conformation of cluster C seems the most probable, having the highest number of models. Although unlikely given their transient unstable nature, N-terminal IDP regions may interact with each other. Table 1 specifies each cluster composition.

Most sequences have a conformer in each cluster (Figure 1 and Table 1). In cluster A, some N-terminal amino acid residues are close to α4–α4′ and may interact with RNA, namely the positively charged residues. Cluster B has the most scattered conformers, with the N-terminal region at the "top", not interacting with other protein regions, resembling a transition between more ordered states. In cluster C, the N-terminal region is in an autoinhibitory conformation, blocking the access to the α1–α2–α2′–α1′ region, as previously suggested by us for DENV C [15]. 18 conformer models are predicted in this closed conformation with, at least, one model from most of the C proteins tested (except JEV C and ZIKV C; see Table 1). Therefore, it can occur in most *Flavivirus* C proteins. As for cluster D conformation, the α1 helix is in the conformation of WNV [14,31] and ZIKV [25] C experimental structures, an arrangement not previously reported for DENV C [15]. This closed conformation also involves the N-terminal region and α1 domain, and partially blocks the α2–α2′ hydrophobic cleft (or totally blocks it, when both monomers are in the same conformation). Importantly, both cluster C and D are closed conformations, supporting the autoinhibition hypothesis.

**Table 1.** Distribution of the I-TASSER predicted models through the four clusters.

| Protein | Cluster A | Cluster B | Cluster C | Cluster D | Excluded |
|---|---|---|---|---|---|
| ALFV C | 1 | 0 | 1 | 0 | 3 |
| AROAV C | 1 | 1 | 1 | 0 | 2 |
| BAGV C | 1 | 0 | 2 | 1 | 1 |
| DENV C | 1 | 2 | 1 | 0 | 1 |
| IGUV C | 1 | 2 | 1 | 0 | 1 |
| ILHV C | 0 | 2 | 2 | 0 | 1 |
| JEV C | 1 | 1 | 0 | 1 | 2 |
| KEDV C | 0 | 1 | 1 | 1 | 2 |
| KOKV C | 1 | 0 | 1 | 1 | 2 |
| MVEV C | 0 | 2 | 1 | 0 | 2 |
| ROCV C | 1 | 0 | 1 | 2 | 1 |
| SLEV C | 1 | 0 | 2 | 0 | 2 |
| SPOV C | 1 | 2 | 1 | 0 | 1 |
| USUV C | 1 | 0 | 2 | 0 | 2 |
| WNV C | 1 | 0 | 1 | 1 | 2 |
| ZIKV C | 0 | 1 | 0 | 1 | 3 |
| **Total** | **12** | **14** | **18** | **8** | **28** |

Dimers with A or B conformers in one monomer enable the simultaneous co-existence of all other conformers (A to D) on the other monomer. The C conformer neither permits the existence of C-C′ homoconformers (i.e., both monomers in the same conformation) nor the heteroconformers of C-D′ and D-C′. Despite that, D-D′ homoconformers are allowed, similarly to the conformation that WNV C adopts in the crystal form [31]. Moreover, to go from cluster A to cluster C or D, the N-terminal region should pass by cluster B. These constraints suggest a path for transitions between conformations, discussed ahead. Overall, the autoinhibition hypothesis proposed for DENV C [15] is supported and such conformation can occur in other *Flavivirus* C proteins.

### 2.4. Analysis of Dengue Virus (DENV) C Protein Rotational Correlation Time

Given the close similarities between *Flavivirus* C proteins (Figures 1–4), DENV C can be used as a general model for them. Hence, we proceeded to determine DENV C overall rotational correlation time ($\tau_c$), taking advantage of the tryptophan residue in position 69 (W69) intrinsic fluorescence. Our computational data support three main structure/dynamics regions, including a disordered N-terminal region, which would increase its expected apparent size (as it would not be globular and folded), a property detectable by such an approach. Upon testing molecules in aqueous solution and at room temperature, fluorescence lifetimes are usually in the ns timescale, and the fluorescence decays are sensitive to the anisotropy of the fluorophore, which depends on its $\tau_c$ (vd. Equations (1)–(8), describing

these relations, in the Methods section [34,35]). Thus, the time-resolved fluorescence decay of DENV C W69 and the corresponding anisotropy decay were determined, both at pH 6.0 and 7.5 (Figure 5).

**Table 2.** Fitting parameters of DENV C time-resolved fluorescence anisotropy data analysis. Parameters obtained from fitting Equations (5) and (8) to the data of Figure 5. Values are average (±% standard error, SE). * Statistically significant differences ($p < 0.05$) between the values obtained for the two pH values tested.

| Parameter | pH 6.0 | pH 7.5 |
|---|---|---|
| $\tau_1$ (ns) * | 0.209 (± 3.9%) | 0.520 (± 4.0%) |
| $\tau_2$ (ns) | 3.106 (± 0.4%) | 3.108 (± 0.9%) |
| $\tau_3$ (ns) * | 6.328 (± 0.4%) | 6.506 (± 0.4%) |
| $\alpha_1$ * | 0.275 (± 0.7%) | 0.178 (± 3.4%) |
| $\alpha_2$ * | 0.315 (± 0.9%) | 0.385 (± 0.4%) |
| $\alpha_3$ * | 0.410 (± 0.4%) | 0.437 (± 0.4%) |
| $\tau_c$ (ns) * | 16.46 (± 2.9%) | 16.41 (± 3.4%) |
| $r_0$ | 0.130 (± 0.8%) | 0.131 (± 1.1%) |

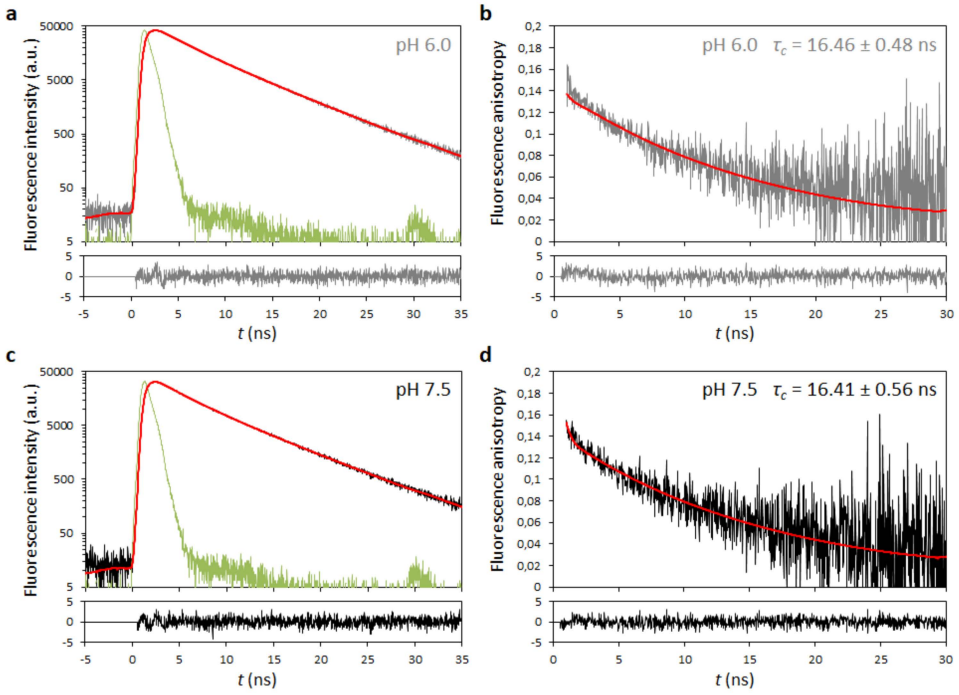


**Figure 5.** DENV C time-resolved fluorescence anisotropy. Time-resolved fluorescence decay at pH (**a**) 6.0 and (**c**) 7.5, with the corresponding anisotropy decays at pH (**b**) 6.0 and (**d**) 7.5. Fluorescence and anisotropy decays at both pH values are similar (gray and black decays, respectively). Fitting of experimental data (red) took into account the instrument response function (IRF; in green) and the corresponding residuals distribution, displayed below each graph. The equations used for fitting are presented on the Methods Equations (5) and (8). The parameters obtained are shown in Table 2.

Time-resolved fluorescence anisotropy decays at both pH values are similar (Figure 5b,d). Fluorescence lifetime components ($\tau_1$, $\tau_2$ and $\tau_3$) were obtained from the intensity decays Equations (2)–(6) [34,35], with a triple-exponential retrieving the best fit (Figure 5a,c). Fitting the data retrieves similar values Table 2 for $\tau_1$, $\tau_2$ and $\tau_3$, and corresponding weights ($\alpha_1$, $\alpha_2$ and $\alpha_3$ pre-exponential factors, respectively). For accurate calculation of $\tau_c$, the condition $\tau_c < 3 \times \tau_3$ must occur [34,35]. Since $\tau_3$ values were ~6.4 ns (with a significant weight $\alpha_3$ of ~0.42), this means that, at both pH values, we could measure $\tau_c$ values up to a limit of ~19 ns. In both pH conditions, the $\tau_c$ measured was

16.4 ± 0.5 ns at 22 °C, within the limit and higher than expected for a purely globular protein of DENV C size, as predicted [13].

Rossi et al. [36] correlated the $\tau_c$ of 16 globular proteins at 20 °C with their molecular weight (MW in kDa), based on NMR data, leading to the relation: $\tau_c \approx 0.6$ MW. Assuming DENV C as a 23.5 kDa fully globular homodimer and correcting for the temperature ($T$) and viscosity ($\eta$) [37], the $\tau_c$ predicted is 12.0 ns. However, the correlational time must be slightly higher, as the protein will be partially unfolded and disordered (in the N-terminal). Jones et al. [16] measured a $\tau_c$ of 13 ns at 27 °C, by NMR, which with the corrections from Equation (10) [37], corresponds to 13.4 ns at 25 °C. Given DENV C size, this implies that the protein is not globular, in line with current knowledge of DENV C structure and dynamics [12–16]. Fluorescence anisotropy supports an even more open and partially disordered DENV C structure, given the $\tau_c$ value of 15.2 ± 0.5 ns at 25 °C Table 3, in line with in silico data (Figures 1–4).

**Table 3.** Comparing DENV C $\tau_c$ values ($\tau_c$ at 25 °C in $H_2O$ were calculated using Equation (10)).

| $\tau_c$ (ns) at $T$ | $T$ (°C) | $\tau_c$ (ns) at 25 °C in $H_2O$ | Method | Source |
|---|---|---|---|---|
| 16.4 ± 0.5 | 22 | 15.2 ± 0.5 | Time-resolved fluorescence anisotropy | This work |
| 13.0 | 27 | 13.4 | Overall NMR relaxation analysis | Jones et al., 2003 [16] |
| 14.1 | 20 | 12.0 | $\tau_c$ (ns) ≈ 0.6×MW (kDa) | Rossi et al., 2010 [36] |

## 2.5. Analysis of DENV C Conformational Stability

Circular dichroism (CD) spectroscopy was used to study DENV C secondary structure, via its thermal denaturation in solution from 0 to 96 °C, at pH 6.0 and 7.5 (2 °C steps, Figure 6). At both pH values, the α-helical structure is partially lost upon increasing temperature (Figure 6a,b). However, even at 96 °C, the protein does no become completely random coil, as seen from the spectrum shape and its high ellipticity at 222 nm (Figure 6c). Plotting the mean residue molar ellipticity at 222 nm, [θ], as a function of temperature, $T$, reveals a transition at ~70 °C at both pH (Figure 6c).

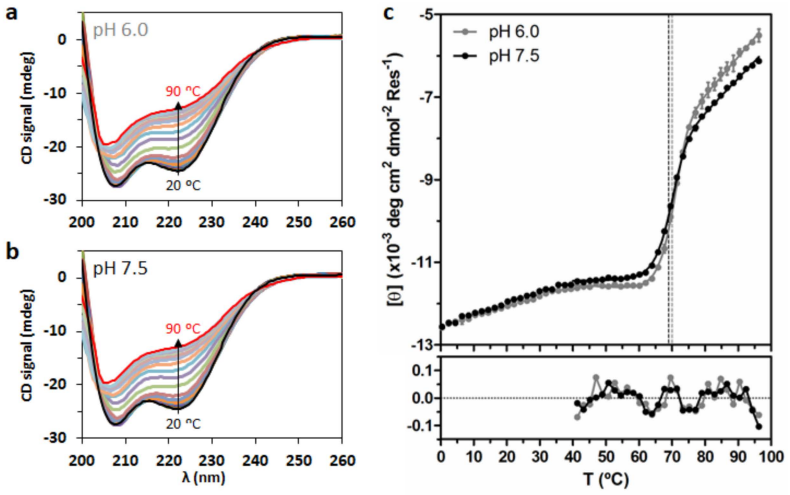


**Figure 6.** DENV C temperature denaturation followed via circular dichroism (CD) spectroscopy. CD spectra of DENV C, between 20 and 90 °C, at pH (**a**) 6.0 and (**b**) 7.5. For the sake of simplicity, the spectra from 0 to 18 °C and from 92 to 96 °C are not displayed, as they are similar to the 20 °C and the 90 °C spectra, respectively. (**c**) Mean residue molar ellipticity at 222 nm, [θ], as a function of temperature (dots) for pH 6.0 (gray) and 7.5 (black), between 0 and 96 °C. Lines correspond to the fitting of Equation (21) (combined with Equations (20), (22), (24) and (28)). Vertical dashed lines represent experimentally observed $T_m$, colored according to pH. Error bars represent SD, from three independent experiments. Residuals are shown below the graph, being lower than SD.

DENV C does not display a typical unfolding profile, as the denaturation curves do not reach a flat plateau. Still, ellipticity data were successfully fitted to a denaturation curve (Figure 6c), assuming a homodimer with one-step denaturation [32]. Briefly, Equation (21) was combined with Equations (20), (22) and (24) and fitted to the data. This allows to obtain the thermodynamic parameters of DENV C unfolding Table 4, namely the melting temperature ($T^{\circ}_{m}$), the enthalpy variation at $T^{\circ}_{m}$ ($\Delta H^{\circ}_{T^{\circ}_{m}}$) and the entropy variation at $T^{\circ}_{m}$ ($\Delta S^{\circ}_{T^{\circ}_{m}}$), with all parameters at standard thermodynamics conditions (symbolized by '°'). Equation (28) was then used to calculate the melting temperature ($T_m$) at the actual $[P_m]$ (instead of the value at $[P] = 1$ M, details in the Methods). Despite small differences, the parameters obtained are not significantly different between pH values Table 4. A small but consistent variation of the CD spectra between 0 and 40 °C is observable, implying: (i) a conformational equilibrium with temperature and/or (ii) some flexibility of the structure and/or (iii) a transition between alternative conformations. This temperature range covers the physiological conditions of both mosquitoes (20 to 40 °C, depending on the environment) and humans (36 to 40 °C). DENV C can continuously transition between conformations as temperature varies, in line with the previously hypothesized conformational equilibrium [15]. As temperature increases, the disordered conformations become more abundant but only a partial loss of structure is seen. This indicates that the C protein conserved region is thermodynamically stable. Similar observations are expected for other *Flavivirus* C proteins.

**Table 4.** Fitting parameters of DENV C temperature denaturation CD data. Parameters were estimated by fitting Equation (21) (combined with Equations (20), (22), (24) and (28)) to the data. $T_m$ is the experimentally observed melting temperature (represented by the vertical lines in Figure 6c). Estimations are average ± SE. There were no significant variations between the two pH values tested ($p < 0.05$).

| Parameter | pH 6.0 | pH 7.5 |
|---|---|---|
| $T_m$ (°C) | 70.02 ± 0.63 | 69.03 ± 0.65 |
| $T^{\circ}_{m}$ (°C) | 88.26 ± 0.80 | 88.80 ± 0.83 |
| $\Delta H^{\circ}_{T^{\circ}_{m}}$ (kJ mol$^{-1}$) | 612 ± 26 | 564 ± 23 |
| $\Delta S^{\circ}_{T^{\circ}_{m}}$ (kJ mol$^{-1}$ K$^{-1}$) | 1.693 ± 0.073 | 1.557 ± 0.065 |

## 3. Discussion

*Flavivirus* C proteins are known to have similar sequences and structure [12–16,25,31]. Here, we go further by examining common features at different structural levels, complemented with data on DENV C size and thermodynamic stability. The phylogenetic analysis of the C proteins and the polyproteins (Figure 1) shows that the former is a marker of *Flavivirus* evolution. There are several conserved motifs, highlighted in previous studies with 16 *Flavivirus* [12,14]. The work is now expanded to include the four DENV serotypes (Figure 2). When these 20 *Flavivirus* C amino acid sequences, with between 96 and 107 amino acid residues each, are jointly analyzed, it is clear that 55% of the residues are conserved or stereochemically similar (Figure 2a). About 80% of amino acid residues are equal or similar and, thus, conserved among the four DENV C serotypes (Figure 2b). From the five major conserved motifs, four are known to be involved in dimer stabilization [14]: the $^{40}$GXGP$^{43}$ motif at loop L1-2, that marks the transition from the flexible to the conserved fold region [14]; the $^{68}$RW$^{69}$ at α3 forms an hydrophobic pocket that accommodates the W69 side chain involving residues from α2, α3 and α4 [12,32]; and, the the $^{44}$h+hhLAhhAFF+F$^{56}$ and $^{84}$F++–h$^{88}$ motifs, respectively from α2 and α4 helices, maintain the homodimer structure both via the α2–α2′ hydrophobic interaction and via the salt bridges of residues [RK]$^{45}$ and [RK]$^{55'}$ with [ED]$^{87}$ [12,14,32]. *Flavivirus* C proteins must have similarly sized secondary structure domains, since G/P are in the same positions and these amino acid residues tend to break the secondary structure (Figure 2c). Charged residues are also conserved (Figure 2d), which makes sense as charges would promote the interaction of the C protein with the negatively charged host lipid systems [12,14,20–22] and the viral RNA [12]. C proteins have a common

homodimer conserved fold region (roughly, residues 45–100), as observed for DENV, WNV and ZIKV C structures [12,14,25,31]. Conserved motifs are summarized in (Figure 2e).

The above explains the C proteins similar hydrophobic and α-helix propensities (Figure 3). The conserved motif $^{13}$hNML+R$^{18}$, at the N-terminal region, and the α2–α2′ hydrophobic cleft are of particular importance for DENV C interaction with LDs and VLDL [14,20–22,38]. Mutations in specific residues of DENV C α2–α2′ and α4–α4′ also impair RNA binding. Likewise, ZIKV C also accumulates on LDs surface, with specific mutations on this protein disrupting the association [25]. ZIKV C also binds single-stranded and double-stranded RNAs [25], with, as for DENV C, the high positively charged residues density prompting the binding to LDs and RNA [12,39,40]. Given the match at the level of N-terminal α-helical propensity and α2–α2′ hydrophobicity (Figure 3), the C proteins may all be self-regulated by an autoinhibition mechanism, as proposed for DENV C [15].

The autoinhibition hypothesis is corroborated by the quaternary structure analysis (Figure 4); Table 1. Two clusters, C and D, are autoinhibited conformations. Importantly, cluster D α1 aligns with WNV C [14,31] and ZIKV C [25]. Moreover, if two monomers are in a D conformation (D–D′ homoconformer), the dimer α2–α2′ region is totally inaccessible. Cluster C does not allow a C–C′ homoconformer nor a C–D heteroconformer, imposing restrictions to the simultaneous transitions that are possible between A, B, C and D, as homodimer. The interaction between N-terminal regions within a dimer may be considered. Nonetheless, the disordered nature and high density of positively charged amino acid residues will mostly favor the repulsion between these IDP regions.

It is important to look at the clusters (Figure 4), while considering the number of positively charged residues (Figure 2) in the disordered N-terminal and flexible fold (10 K and R residues) versus those in the conserved fold (15 K and R). The charge distribution in some arrangements implies that the disordered N-terminal is at least in theory able to bind the viral RNA [39,40]. Such binding would be governed by the N-terminal region cationic amino acid residues [41,42]. Here, the structure predictions reveal that, indeed, the first 12 N-terminal residues can locate near α4–α4′ Cluster A (Figure 4), the most likely RNA binding site [12,39,40]. Furthermore, binding to RNA via the C-terminal α4–α4′ interface may be favored by a previous or simultaneous interaction of the protein with host LDs via the N-terminal region and α2–α2′ interface. Access to α2–α2′ (controlled by the N-terminal region) would modulate the interaction (Figure 4) and, thus, viral assembly. In agreement, the binding of the related hepatitis C virus core protein (homologous to DENV C) to host LDs is what enables efficient viral assembly [43]. Thus, the C protein disordered N-terminal would be critical to protein function, enabling crucial structural and functional roles.

To evaluate this, we used DENV C as a model system, measuring its $\tau_c$ value by time-resolved fluorescence anisotropy (Figure 5) and its thermal stability by CD spectroscopy (Figure 6), at pH 6.0 and 7.5 (within the usual pH range of its biological microenvironment). A similar $\tau_c$, 15.2 ± 0.5 ns, is obtained at both pH values (Figure 5; Tables 2 and 3), in line with previous work [13]. DENV C maintains its homodimer structure and dynamics behavior between pH 6.0 and 7.5. The $\tau_c$ value and respective size are higher than expected, due to the N-terminal disordered nature.

Regarding DENV C thermodynamic stability (Figure 6, Table 4), the protein $T_m$ is ~70 °C at both pH values. These denaturation parameters are in line with other authors, as a chemically synthesized DENV C 21–100 fragment (without most of the disordered N-terminal region) displays a $T_m$ = 71.6 °C [32]. DENV C high thermal stability in physiological conditions is likely due to the large hydrophobic area that is shared by the two monomers [12], but also to the W69 stabilizing interactions and, as experimentally observed [32], the formation of salt bridges (residues K45 and R55′ with E87). As structure/dynamics properties are conserved among *Flavivirus* C proteins (Figures 2–4), these observations can probably be generalized for all these proteins.

These findings must also be considered in light of DENV C biologically relevant interactions with LDs [22] and RNA (Figure 7). DENV C experimental structure [12] contains three distinct structural regions [13]: a disordered N-terminal region (from the N-terminal up to residue R22), a flexible fold (residues V23 to L44, where α-helix 1 is located) and a conserved fold with helices α2, α3 and

α4, containing the R68 and W69 amino acid residues, highly conserved among *Flavivirus* [12]. R68 terminates α3 helix, with its side chain pointing to the protein interior [12]. W69 locates at DENV C α4–α4′ interface, having a crucial role in the dimer structural stabilization [12]. Along with dimer structural stability, these interactions enable allosteric communication and movements between DENV C more hydrophobic section (α2–α2′dimer interface) and its remaining sections, namely the α4–α4′ region. Figure 7 displays this, in the context of the C protein biologically relevant interactions, as they are understood on the basis of recent studies [12–15,18,20–24].

Looking further, it is important to consider that the binding of DENV C to host LDs is mediated by both the N-terminal IDP region and the α2–α2′ interface [14]. V51 of α2 is affected by the interaction with LDs and stabilizes the dimer by contacting with α3 (I65). Another interaction via salt bridges, between α2 (K45 and R55′) and α4 (E87), stabilizes the homodimer (Figure 7a). The C protein binding to host LDs, which affects the α2–α2′, can lead to changes in the α4–α4′ structural arrangement (Figure 7b). To investigate this we searched for similar proteins. An RNA-binding protein with a two-helix domain similar to DENV C α4–α4′ was identified (Figure 7c), influenza A non-structural protein 1 (NS1, PDB ID: 2ZKO [44]). Influenza NS1 has interesting features: it accumulates in the nuclei of host cells after being translocated by importin α and β and works as a viral immuno-suppressor by weakening the host cell gene expression [45]. DENV C was also reported to have an importin α-like motif in the N-terminal [15,46]. Regarding the targets that may interact with importin α and be transported to the nucleus, they normally contain a nuclear localization sequence (NLS), consisting of a motif of at least 2 consecutive positively charged residues [47–51]. Some of these proteins contain 2 NLS motifs, with at least 8 (up to 40 or even more) residues in between, designated as a bipartite NLS motif [49–51]. Strikingly, *Flavivirus* C proteins have three motifs of two consecutive cationic residues in the N-terminal region and α1 domain, which could form a bipartite NLS. A bipartite NLS formed by the cationic residues before position 10 and at positions 17 and 18, with a spacer of 7 to 13 residues can occur. The other bipartite NLS possibility may be formed by residues at positions 17 and 18, and at positions 31 and 32, with 9 to 12 spacer residues. Possible bipartite NLS are also seen in the conserved fold region but its static nature precludes activity as NLS. If DENV C binds to importin α, it may act as a cargo protein to be transported to the nucleus. This could explain why has DENV C been found in the nucleus of DENV infected cells [46,52,53]. DENV C may directly bind importin β, given the similarities between the N-terminal region of DENV C and importin α [49]. This may allow it to disrupt the normal nuclear import/export system in DENV-infected cells. The conformational plasticity of the N-terminal and flexible fold regions is certainly compatible with interactions with importin(s). As the hypothesized bipartite NLS are conserved among *Flavivirus* C proteins, this may occur in other *Flavivirus*.

The C protein may act as an immuno-suppressor, similarly to influenza NS1, by interacting with importins α and/or importin β. Ivermectin, a specific inhibitor of importin α/β-mediated nuclear import, is able to inhibit HIV-1 and DENV replication [54]. The mechanism of DENV C inhibition might involve the C protein, specifically the intrinsically disordered N-terminal IDP region, which is similar to importin α disordered N-terminal region [15]. Moreover, influenza NS1 can counteract the RNA-activated protein kinase (PKR)-mediated antiviral response through a direct interaction with PKR [55]. Besides, influenza NS1 blocks interferon (IFN) regulatory factor 3 activation, which in turn prevents the induction of IFN-related genes [56]. DENV inhibits the IFN signaling pathway in a similar manner [57]. By its N-terminal region dsRNA-binding ability, influenza NS1 inhibits the nuclear export of mRNAs and modulates pre-mRNA splicing, suppressing antiviral response [44]. Similarities between DENV C and influenza NS1 also extend to the later ability to bind RNA (Figure 7c). Recognition of dsRNA is made by the influenza NS1 RNA-binding domain, which forms a homodimer [44]. Afterwards, a slight change in R38-R38′ orientation leads to anchoring the dsRNA to the protein by a hydrogen bond network to the protein [44]. One of the main functions of influenza NS1 binding to RNA is sequestering dsRNA from the 2′–5′ oligo(A) synthetase [58]. We propose that, as with influenza NS1, a small conformational change in DENV C α4–α4′ interface occurs after the contact

of its α2–α2′ interface with LDs, modulated by transitions between alternative N-terminal "open" and "closed" conformations. Binding to LDs requires an open conformation (Figure 7d), decreasing the conformational variability and entropy of the C protein, which trigger the allosteric movements affecting the C-terminal α4–α4′. As with influenza NS1, the *Flavivirus* C protein would remain in the same overall fold, but a small opening of α4–α4′ would facilitate its binding to RNA.

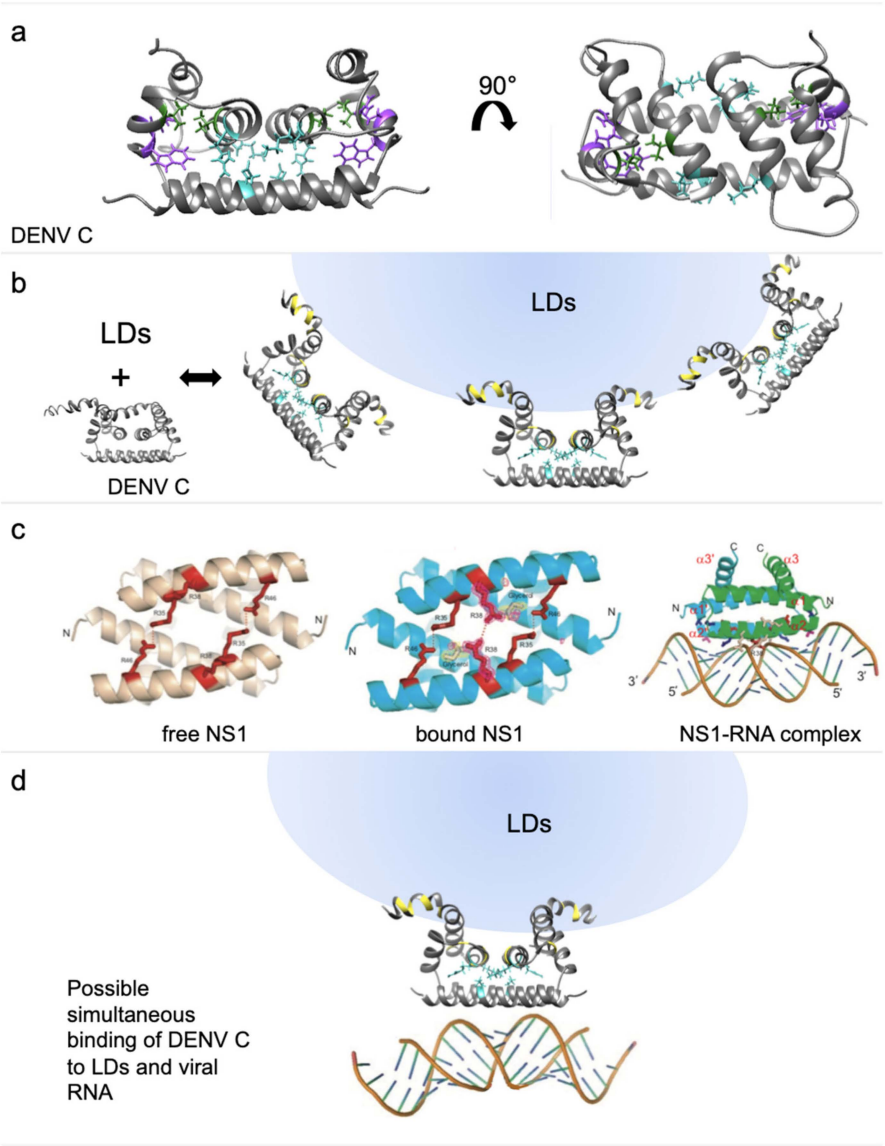


**Figure 7.** Protein structures of DENV C and influenza NS1. (**a**) DENV C structure from two different angles with the conserved residues R68 and W69 (purple) and the interface stabilizing residues V51 and I65 (green), as well as E87, R55 and K45, forming the salt bridge (cyan). (**b**) DENV C structure in a N-terminal region closed conformation and, next, in an open conformation with schematic binding of lipid droplets (LDs) and the affected amino acid residues (yellow). (**c**) The RNA-binding domain of NS1 protein from influenza A in a RNA-free (left) and RNA-bound state (middle and right), showing an organization similar to DENV C α4–α4′ region (adapted from Cheng et al., 2009 [44]). (**d**) DENV C with schematically bound to a LD and to RNA. DENV C amino acid residues affected by the binding to LDs are colored yellow, while a key internal salt bridge is shown in cyan. DENV C binding to host LDs may enable allosteric rearrangements (eventually involving the salt bridge), allowing a small conformational change in α4 side chains, namely the positively charged residues, prompting stable RNA-C protein binding.

The C-terminal is likely to be the crucial section for RNA binding given its similarity with influenza NS1 (Figure 7). Nevertheless, the N-terminal conformers must also be considered in the context of RNA binding (Figure 4). The A and D conformers allow RNA to be bound to the α4–α4′ interface and, simultaneously, to the N-terminal cationic amino acid residues. A–A′ and D–D′ conformations result in the possible binding of a single continuous portion of RNA to both the C-terminal α4–α4′ and the N-terminal IDP region, making the RNA more tightly bound. Moreover, the A–B′, B–B′ and B–C′ conformations would enable the protein to bind two distinct sections of the RNA, one bound to α4–α4′ and another to the N-terminal regions. That arrangement may allow to further compact the viral RNA. The N-terminal IDP region putative binding to RNA should not be disregarded given its positive net charge (+7). It compares very well with the C-terminal α-helical region net charge (+8 for a monomer, +16 for α4–α4′ dimer interface). Both may thus bind RNA due to, mostly, electrostatic forces. This IDP region can thus provide multi-functionality by several modes of binding and different ligands, enabled by alternative conformations. It must be stressed that this is not unlikely. Viral proteins tend to have IDP regions that increase their biological activity [59–61]. In a proteome as small as that of flaviviruses (10 proteins), IDP regions augment the number of ligands with which it can interact. Less structure often means more function. This is an increasingly hot topic of recent research, leading to design of algorithms to identify these regions [62,63]. Further analysis will help understand the interaction between DENV C and its ligands.

To conclude, the data imply a common structure and functions for mosquito-borne *Flavivirus* C proteins. Moreover, studying DENV C rotational diffusion and thermodynamics reveals a stable protein due to the conserved fold maintaining the homodimer structure. These findings apply to other *Flavivirus* C proteins, supporting a common mechanism for their biological activity. Such understanding of this key protein structure and dynamics properties may contribute to the future development of C protein-targeted drugs to impair dengue virus and other *Flavivirus* infections.

## 4. Materials and Methods

### 4.1. Materials

Chromatography columns HiTrap Heparin (1 and 5 mL), Sephadex S200 and the chromatography equipment AKTA-explorer were from GE Healthcare (Little Chalfont, UK). Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) reagents were from BioRad (Hercules, CA, USA). Unless otherwise stated, other chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA).

### 4.2. Flavivirus C Proteins Primary, Secondary and Tertiary Structural Predictions

For primary structure alignments we used the 16 non-DENV *Flavivirus* polyprotein sequences identified in reference [14], plus the four DENV reference sequences from NCBI, namely: DENV serotype 1, strain 45AZ5, NCBI ID NP_059433.1; DENV serotype 2, strain New Guinea C, NCBI ID NP_056776.2; DENV serotype 3, strain D3/H/IMTSSA-SRI/2000/1266, NCBI ID YP_001621843.1; and, DENV serotype 4, strain rDEN4, NCBI ID NP_073286.1. For the phylogenetic trees, both the entire polyproteins and the C protein regions were used. For the alignments and subsequent data analysis, the residues next to the NS2B-NS3 protease cleavage site [64,65] were excluded, leaving only the C protein sequences. Alignments and the derived phylogenetic trees were performed via Clustal Omega web tool (http://www.ebi.ac.uk/Tools/msa/clustalo/) [66,67].

Statistical comparison of the disordered N-terminal plus flexible fold regions with the conserved fold region of *Flavivirus* C proteins, for G and P content, as well as charged amino acid residues, was performed via a paired *t*-test, using GraphPad Prism v5 software. *p*-values were always lower than 0.001.

Predictions of hydrophobicity and α-helix propensity were done using ProtScale server (http://web.expasy.org/protscale/) [26,27], tertiary structure predictions were performed via I-TASSER

server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) [28–30], following previous approaches [15]. Briefly, *Flavivirus* C protein sequences from our previous work were employed [14]. DENV and WNV (serotype Kunjin) C structures were excluded, not serving as templates for the tertiary structure prediction. ZIKV C protein structure was also not included, as it was not yet determined when the modeling was conducted. This avoids a bias towards known homologous protein structures. Five I-TASSER models were obtained for each C protein sequence. These were superimposed with DENV C experimental structure (PDB ID 1R6R, model 21) [12] after root-mean-square deviation (RMSD) minimization in UCSF Chimera v1.9 software [68]. Clusters were formed based on the visual similarity between predictions. The number of N-terminal amino acid residues with backbone clashes with the other monomer backbone was calculated for each model. In our previous work [15], a DENV C predicted structure was excluded from further analysis if it had 6 clashes or more, as it would not be viable as an homodimer [15]. Here we excluded models with more than 5 clashes (28 models rejected). These would preclude homodimer formation and, thus, were not considered in the clusters analysis (Table 1 excluded models column).

### 4.3. Structure Comparison Between DENV C and Influenza NS1

Protein structures coordinates were extracted from the Protein Data Bank (PDB, www.pdb.org). PDB identification codes are specified ahead after each protein name. The protein structures were superimposed through UCSF Chimera 1.13.1 software MatchMaker tool. After that, we carefully analyzed the superposition visually. Then, using the Match-Align tool of UCSF Chimera, which returns a sequence alignment based on the regions and taking into account the structure superimposition, we identified the residues simultaneously similar in structure and sequence. Protein structure figures were obtained using UCSF Chimera 1.13.1 version [68].

### 4.4. DENV C Recombinant Protein Production and Purification

Recombinant DENV C protein expression and purification was conducted based on previous approaches [13]. We used a pET-21a plasmid containing DENV serotype 2 strain New Guinea C capsid protein gene (encoding amino acid residues 1–100) [69]. The protein was expressed in *Escherichia coli* C41 and C43 bacteria grown in lysogeny broth (LB) medium. The only differences in the purification protocol are the abolition of the ammonium sulfate precipitation step and the addition of a size exclusion chromatography step (with Sephadex S200) after the heparin affinity column chromatography, using an AKTA chromatography equipment. The C protein was purified in a 55 mM $KH_2PO_4$, pH 6.0, 550 mM KCl. DENV C protein purified fractions were concentrated with Amicon Ultra-4 Centrifugal Filters of 3 or 10 kDa nominal cut-off, from Millipore (Billerica, MA, USA). Concentrated protein samples were stored at −80 °C. Protein samples quality was assessed by SDS-PAGE and matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry (MALDI-TOF MS) analysis. Very low degradation and the highest peak consistent with the expected mass of the protein monomer (11765 Da).

### 4.5. Time-Resolved Fluorescence Anisotropy

Time-resolved fluorescence spectroscopy measurements were performed in a Life Spec II equipment with an EPLED-280 pulsed excitation light-emitting diode (LED) of 275 nm (Edinburgh Instruments, Livingston, UK), acquiring the emission at 350 nm. DENV C (monomer) concentration was 20 μM in 50 mM $KH_2PO_4$, 200 mM KCl, pH 6.0 or pH 7.5, with 550 μL total volume, in 0.5 cm × 0.5 cm quartz cuvettes. The instrument response function, IRF$(t)$, was obtained with the same settings, except emission, which was at 280 nm, with a solution of polylatex beads of 60 nm diameter diluted in Mili-Q water. Measurements were performed at 22 °C. Time-resolved fluorescence intensity measurements with picosecond-resolution were obtained by the time-correlated single-photon timing (TCSPT) methodology [35]. Measurements were performed at constant time, with 15 min per decay, acquiring 2048 time points in a 50 ns window. Four intensity decays, $I(t)$, were acquired in each condition, with excitation/emission polarizers, respectively at vertical/vertical positions, $I_{VV}(t)$, vertical/horizontal

positions, $I_{VH}(t)$, horizontal/vertical positions, $I_{HV}(t)$, and horizontal/horizontal positions, $I_{HH}(t)$. The instrumental $G$-factor was calculated as [35]:

$$G = \frac{\int_0^{50} I_{HV}(t)dt}{\int_0^{50} I_{HH}(t)dt} \quad (1)$$

The $G$-factor value obtained was 1.61. The intensity decay with emission polarizer at the magic angle (~54.7°, with respect to the vertical excitation polarizer), $I_m(t)$, avoids the effects of anisotropy. It can be calculated easily [35]:

$$I_m(t) = I_{VV}(t) + 2GI_{VH}(t) \quad (2)$$

with $I_{VV}(t)$ and $I_{VH}(t)$ depending on the time-resolved fluorescence anisotropy, $r(t)$, as:

$$I_{VV}(t) = \frac{I_m(t)}{3}(1 + 2r(t)) \quad (3)$$

$$I_{VH}(t) = \frac{I_m(t)}{3G}(1 - r(t)) \quad (4)$$

Thus, $I_m(t)$ was used to obtain the fluorescence lifetime components, $\tau_i$, and the respective amplitudes, $\alpha_i$, for the DENV C W69. $I_m(t)$ was described by a sum of three exponential terms:

$$I_m(t) = \sum_{i=1}^{3} \alpha_i e^{\left(-\frac{t}{\tau_i}\right)} \quad (5)$$

where the index $i$ represents each component of the fluorescence decay. For the fitting to the data, $\alpha_i$ and $\tau_i$ values were obtained by iteratively convoluting $I_m(t)$ with the IRF$(t)$:

$$I_m^{calc}(t) = I_m(t) \otimes \mathrm{IRF}(t) \quad (6)$$

and fitting $I_m^{calc}(t)$ to the experimental data, $I_m^{exp}(t)$, using a non-linear least squares regression method. The usual statistical criteria, namely a reduced $\chi^2$ value bellow 1.3 and a random distribution of weighted residuals, were used to evaluate the goodness of the fits [35]. Data analysis was performed using the TRFA Data Processing Package v1.4 (Scientific Software Technologies Centre, Belarusian State University, Minsk, Belarus) which allows calculating automatically the standard error (SE) for each fitted parameter [35].

The time-resolved fluorescence anisotropy, $r(t)$, is calculated via $I_{VV}(t)$, $I_{VH}(t)$ and $G$ via_ENREF_52:

$$r(t) = \frac{I_{VV}(t) - GI_{VH}(t)}{I_{VV}(t) + 2GI_{VH}(t)} \quad (7)$$

In this case, the obtained $r(t)$ can be fitted to a single exponential decay [35]:

$$r(t) = r_0 e^{\left(-\frac{t}{\tau_c}\right)} \quad (8)$$

where $r_0$ is the anisotropy when $t \rightarrow 0$ and $\tau_c$ is the rotational correlation time. The $r(t)$ decays were globally analyzed in TRFA Data Processing Package v1.4 maintaining the previously obtained $\alpha_i$ and $\tau_i$ values constant, and convoluting Equations (3) and (4) with the respective IRF$(t)$, analogously to the analysis of $I_m(t)$, using Equation (8) to fit $r(t)$. Values obtained for both pH conditions were considered statistically different if their 95% confidence intervals (~1.96 × SE) do not overlap (corresponding to $p < 0.05$).

### *4.6. Rotational Correlation Time Corrections*

The $\tau_c$ of a molecule in solution is related with the solution viscosity, $\eta$, the molecular hydrodynamic volume, V, the Boltzmann constant, $k_B$, and the absolute temperature, $T$, as [35,70]:

$$\tau_c = \frac{\eta V}{k_B T} \quad (9)$$

Based on Equation (9), $\tau_c$ can be corrected for different temperatures, considering that the molecular volume does not change significantly in a small temperature interval (±5 °C; i.e., V and $k_B$ are constants), using [70]:

$$\frac{T_a \tau_{c,a}}{\eta_a} = \frac{T_b \tau_{c,b}}{\eta_b} \Leftrightarrow \tau_{c,b} = \tau_{c,a} \frac{\eta_b T_a}{\eta_a T_b} \quad (10)$$

where the indexes 'a' and 'b' represent a different condition of $T$ and $\eta$, taking into account the variation of $\eta$ with $T$ [37]. The $\eta$ values were assumed to be those of pure $H_2O$ or 10% $D_2O$ in the case of the corrections for the NMR-based values (those from the literature). In this way, Table 5 below shows the values employed on the calculations [37]:

**Table 5.** Values for $\eta$ employed in this work, derived from the references and Equations above.

| T (°C) | $\eta$ in $H_2O$ (cP) | $\eta$ in 10% $D_2O$ (cP) | $\frac{\eta_b T_a}{\eta_a T_b}$ in$H_2O$ | $\frac{\eta_b T_a}{\eta_a T_b}$ in10%$D_2O$ |
|---|---|---|---|---|
| 20 | 1.002 | 1.027 | 0.8736 | 0.8523 |
| 22 | 0.955 | 0.978 | 0.9231 | 0.9012 |
| 25 | 0.890 | 0.911 | 1 | 0.9770 |
| 27 | 0.851 | 0.871 | 1.0530 | 1.0293 |

### *4.7. Temperature Denaturation Measurements via Circular Dichroism (CD) Spectroscopy*

Circular dichroism spectroscopy measurements were carried out 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, data integration time (DIT) of 1 s and performing 3 accumulations. Spectra were acquired in the far UV region, between 200 and 260 nm, with 1 nm bandwidth. The temperature was controlled by a JASCO PTC-423S/15 Peltier equipment. It was varied between 0 and 96 °C, in steps of 2 °C, increasing at a rate of 8 °C/min and waiting 100 s after crossing 5 times the target temperature, $T$. Then, the system was allowed, at least, 120 s to equilibrate (sufficient time for a stable CD signal). Before and after denaturation, spectra were acquired at 25 °C, to determine the reversibility of thermal denaturation. DENV C monomer concentration was 20 µM in 50 mM $KH_2PO_4$, 200 mM KCl, pH 6.0 or pH 7.5, with 220 µL of total volume. Spectra were smoothed through the means-movement method (using 7 points) and normalized to mean residue molar ellipticity, [θ] (in deg $cm^2$ $dmol^{-1}$ $Res^{-1}$).

For the CD temperature denaturation data treatment, we assumed a dimer to monomer denaturation model [71–73] in which the folded dimer, $F_2$, separates into unfolded monomers, U, in a single step described by reaction R1:

$$F_2 \Leftrightarrow 2U \quad (R1)$$

In this system, the total protein concentration, $[P_m]$, in monomer equivalents, is described as:

$$[P_m] = 2[F_2] + [U] \quad (11)$$

Hereafter, concentrations are treated as dimensionless, being divided by the standard concentration of 1 M, in order to be at standard thermodynamic conditions. The fractions of monomer in the folded, $f_F$, and unfolded, $f_U$, states are calculated by [71,72]:

$$f_F = \frac{2[F_2]}{[P_m]} \tag{12}$$

$$f_U = \frac{[U]}{[P_m]} \tag{13}$$

$$f_F + f_U = 1 \tag{14}$$

and the concentrations of folded dimer and unfolded monomer can be written in terms of $f_U$:

$$[U] = f_U[P_m] \tag{15}$$

$$[F_2] = \frac{f_F[P_m]}{2} = \frac{(1 - f_U)[P_m]}{2} \tag{16}$$

Then, the equilibrium constant, $K_{eq}$, of R1 is defined in terms of $[U]$ and $[F_2]$, or $f_U$ and $[P_m]$:

$$K_{eq} = \frac{[U]^2}{[F_2]} = \frac{(f_U[P_m])^2}{(1 - f_U)[P_m]/2} = \frac{2[P_m] \times {f_U}^2}{(1 - f_U)} \tag{17}$$

which can be solved in order to $f_U$, with the only solution in which $f_U \in [0;1]$ being:

$$f_U = \frac{\sqrt{8[P_m]K_{eq} + {K_{eq}}^2} - K_{eq}}{4[P_m]} \tag{18}$$

The $[\theta]$ signal as a function of temperature [71,72,74], $[\theta]_T$, can be described as a linear combination of the signal of the folded, $[\theta]_{T,F}$, and unfolded states, $[\theta]_{T,U}$, weighted by $f_U$:

$$[\theta]_T = [\theta]_{T,F}(1 - f_U) + [\theta]_{T,U}f_U \tag{19}$$

where $[\theta]_{T,F}$ and $[\theta]_{T,U}$ have a variation with $T$ described here by a straight line ($i$ can be F or U) [72,74]:

$$[\theta]_{T,i} = m_i \times T + [\theta]_{0,i} \tag{20}$$

Equation (19) can be re-written to evidence $f_U$ and then substitute it by Equation (18) [71,72]:

$$[\theta]_T = [\theta]_{T,F} + \left([\theta]_{T,U} - [\theta]_{T,F}\right)\frac{\sqrt{8[P_m]K_{eq} + {K_{eq}}^2} - K_{eq}}{4[P_m]} \tag{21}$$

$K_{eq}$ can also be described by the standard Gibbs free-energy, $\Delta G^\circ$, of the reaction R1:

$$K_{eq} = e^{-\frac{\Delta G^\circ}{RT}} \tag{22}$$

where R is the rare gas constant and $T$ is the absolute temperature. The $\Delta G^\circ$ function used to fit the data contains both the enthalpic, $\Delta H^\circ$, and entropic, $\Delta S^\circ$, variations with temperature, which take into account ${\Delta H^\circ}_{T_m^\circ}$, the specific heat capacity at constant pressure, $\Delta C_p^\circ$, and the standard conditions' denaturation temperature, $T_m^\circ$, according to [74]:

$$\Delta G^\circ = {\Delta H^\circ}_{T_m^\circ}\left(1 - \frac{T}{T_m^\circ}\right) - \Delta C_p^\circ\left(T_m^\circ - T + T\ln\left(\frac{T}{T_m^\circ}\right)\right) \tag{23}$$

In our data, $\Delta C_p^\circ$ was statistically equal to 0 and, thus, Equation (23) can be simplified to:

$$\Delta G^\circ = \Delta H^\circ_{T_m^\circ}\left(1 - \frac{T}{T_m^\circ}\right) \quad (24)$$

Then, Equation (21) was combined with Equations (20), (22) and (24), and fitted to the data using GraphPad Prism v5 software, via the non-linear least squares method, to extract both the $\Delta H^\circ_{T_m^\circ}$ and $T_m^\circ$, along with the respective SE values. Afterwards, $\Delta S^\circ_{T_m^\circ}$ can be obtained, since $\Delta G^\circ = 0$ kJ mol$^{-1}$ at $T_m^\circ$, via the following Equation:

$$\Delta H^\circ_{T_m^\circ} - T_m^\circ \Delta S^\circ_{T_m^\circ} = 0 \Rightarrow \Delta S^\circ_{T_m^\circ} = \frac{\Delta H^\circ_{T_m^\circ}}{T_m^\circ} \quad (25)$$

The SE of $\Delta S^\circ_{T_m^\circ}$ was calculated based on $\Delta H^\circ_{T_m^\circ}$, $T_m^\circ$, and the respective SE values:

$$SE_{\Delta S^\circ_{T_m^\circ}} = \left|\frac{\Delta H^\circ_{T_m^\circ}}{T_m^\circ}\right| \times \sqrt{\left(\frac{SE_{\Delta H^\circ_{T_m^\circ}}}{\Delta H^\circ_{T_m^\circ}}\right)^2 + \left(\frac{SE_{T_m^\circ}}{T_m^\circ}\right)^2} \quad (26)$$

Interestingly, for a dimer to monomer denaturation, $K_{eq}$ depends on $[P_m]$ and, consequently, $\Delta G^\circ$ also depends on $[P_m]$. This implies that $\Delta G^\circ = 0$ at $T_m^\circ$ ($T_m$ value estimated if $[P_m] = 1$M), which is considerably higher than the observed $T_m$ (that occurs when $f_U = 0.5$). The dependence of $T_m$ with $[P_m]$ is [72]:

$$\Delta G^\circ_{f_U=0.5} = -RT_m \ln([P_m]) \Rightarrow T_m = \frac{\Delta G^\circ_{f_U=0.5}}{-R\ln([P_m])} \quad (27)$$

$$T_m = \frac{\Delta H^\circ_{T_m^\circ}}{\Delta S^\circ_{T_m^\circ} - R\ln([P_m])} \quad (28)$$

The SE of $T_m$ was based on the percentual SE value of $T_m^\circ$.

Values obtained for both pH conditions were statistically evaluated via F-tests to compare two possible fits, one assuming a given parameter as being different for the distinct data sets, and another assuming that parameter to be equal between data sets (while maintaining the other parameters different). No statistically significant difference ($p < 0.05$) was observed.

**Author Contributions:** Conceptualization, A.F.F., N.C.S. and I.C.M.; In silico studies, A.F.F., V.A., A.S.M., N.K. and I.C.M.; Recombinant protein production, A.F.F., A.S.M., F.J.E. and I.C.M.; Time-resolved fluorescence anisotropy studies, A.F.F. and J.C.R.; Circular dichroism studies, A.F.F. and I.C.M.; Formal analysis, A.F.F., J.C.R. and I.C.M.; Resources, I.C.M., F.J.E., N.C.S.; Writing-original draft preparation, A.F.F., N.K., and I.C.M.; Writing-review and editing, A.F.F., A.S.M., N.K., N.C.S. and I.C.M.; Supervision, N.C.S. and I.C.M.; Project administration, N.C.S. and I.C.M.; Funding acquisition, N.C.S. and I.C.M.

**Funding:** This work was supported by "Fundação para a Ciência e a Tecnologia–Ministério da Ciência, Tecnologia e Ensino Superior" (FCT-MCTES, Portugal) project PTDC/SAU-ENB/117013/2010, Calouste Gulbenkian Foundation (FCG, Portugal) project Science Frontiers Research Prize 2010. A.F.F., A.S.M. and J.C.R. also acknowledge FCT-MCTES fellowships SFRH/BD/77609/2011, PD/BD/113698/2015 and SFRH/BD/95856/2013, respectively. I.C.M. acknowledges FCT-MCTES Programs "Investigador FCT" (IF/00772/2013) and "Concurso de Estímulo ao Emprego Científico" (CEECIND/01670/2017). This work was also supported by UID/BIM/50005/2019, project funded by Fundação para a Ciência e a Tecnologia (FCT)/ Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through Fundos do Orçamento de Estado.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## Abbreviations

| | |
|---|---|
| ALFV | Alfuy virus |
| APOE | Apolipoprotein E |
| AROAV | Aroa virus |
| BAGV | Bagaza virus |
| C protein | Capsid protein |
| CD | Circular dichroism |
| DENV | Dengue virus |
| ICTV | International Committee on Taxonomy of Viruses |
| IDP | Intrinsically disordered protein |
| IFN | Interferon |
| IGUV | Iguape virus |
| ILHV | Ilheus virus |
| JEV | Japanese encephalitis virus |
| KEDV | Kedougou virus |
| KOKV | Kokobera virus |
| LDs | Lipid droplets |
| MVEV | Murray Valley encephalitis virus |
| NS1 | Non-structural protein 1 from influenza virus A |
| PDB | Protein Data Bank |
| pep14-23 | Inhibitor peptide pep14-23 (amino acid sequence NMLKRARNRV) |
| PLIN3 | Perilipin 3 |
| ROCV | Rocio virus |
| SLEV | Saint Louis encephalitis virus |
| SPOV | Spondweni virus |
| USUV | Usutu virus |
| VLDL | Very low-density lipoproteins |
| WNV | West Nile virus |
| WNV-K | WNV serotype Kunjin |
| YFV | Yellow fever virus |
| ZIKV | Zika virus |

## References

1. Bhatt, S.; Gething, P.W.; Brady, O.J.; Messina, J.P.; Farlow, A.W.; Moyes, C.L.; Drake, J.M.; Brownstein, J.S.; Hoen, A.G.; Sankoh, O.; et al. The global distribution and burden of dengue. *Nature* **2013**, *496*, 504–507. [CrossRef] [PubMed]
2. Sanofi Pasteur. Available online: https://www.sanofipasteur.com/en/media-room/press-releases/dengvaxia-vaccine-approved-for-prevention-of-dengue-in-europe (accessed on 30 January 2019).
3. Durbin, A.P. A dengue vaccine. *Cell* **2016**, *166*, 1. [CrossRef] [PubMed]
4. Villar, L.; Dayan, G.H.; Arredondo-Garcia, J.L.; Rivera, D.M.; Cunha, R.; Deseda, C.; Reynales, H.; Costa, M.S.; Morales-Ramirez, J.O.; Carrasquilla, G.; et al. Efficacy of a tetravalent dengue vaccine in children in Latin America. *N. Engl. J. Med.* **2015**, *372*, 113–123. [CrossRef] [PubMed]
5. Takeda. Available online: https://www.takeda.com/newsroom/newsreleases/2019/takedas-dengue-vaccine-candidate-meets-primary-endpoint-in-pivotal-phase-3-efficacy-trial/ (accessed on 4 February 2019).
6. ICTV Taxonomy. Available online: https://talk.ictvonline.org/taxonomy/ (accessed on 17 April 2019).
7. Grard, G.; Moureau, G.; Charrel, R.N.; Holmes, E.C.; Gould, E.A.; de Lamballerie, X. Genomics and evolution of Aedes-borne flaviviruses. *J. Gen. Virol.* **2010**, *91*, 87–94. [CrossRef] [PubMed]
8. Schubert, A.M.; Putonti, C. Infection, genetics and evolution of the sequence composition of flaviviruses. *Infect. Genet. Evol.* **2010**, *10*, 129–136. [CrossRef] [PubMed]
9. Calisher, C.H.; Gould, E.A. Taxonomy of the virus family *Flaviviridae*. *Adv. Virus Res.* **2003**, *59*, 1–19. [PubMed]
10. Mukhopadhyay, S.; Kuhn, R.J.; Rossmann, M.G. A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.* **2005**, *3*, 13–22. [CrossRef]

11. Kuhn, R.J.; Zhang, W.; Rossmann, M.G.; Pletnev, S.V.; Corver, J.; Lenches, E.; Jones, C.T.; Mukhopadhyay, S.; Chipman, P.R.; Strauss, E.G.; et al. Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion. *Cell* **2002**, *108*, 717–725. [CrossRef]
12. 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. USA* **2004**, *101*, 3414–3419. [CrossRef]
13. Faustino, A.F.; Barbosa, G.M.; Silva, M.; Castanho, M.A.R.B.; da Poian, A.T.; Cabrita, E.J.; Santos, N.C.; Almeida, F.C.L.; Martins, I.C. Fast NMR method to probe solvent accessibility and disordered regions in proteins. *Sci. Rep.* **2019**, *9*, 1647. [CrossRef]
14. Martins, I.C.; Gomes-Neto, F.; Faustino, A.F.; Carvalho, F.A.; Carneiro, F.A.; Bozza, P.T.; Mohana-Borges, R.; Castanho, M.A.R.B.; Almeida, F.C.L.; Santos, N.C.; et al. The disordered N-terminal region of dengue virus capsid protein contains a lipid-droplet-binding motif. *Biochem. J.* **2012**, *444*, 405–415. [CrossRef]
15. Faustino, A.F.; Guerra, G.M.; Huber, R.G.; Hollmann, A.; Domingues, M.M.; Barbosa, G.M.; Enguita, F.J.; Bond, P.J.; Castanho, M.A.R.B.; da Poian, A.T.; et al. Understanding Dengue virus capsid protein disordered N-terminus and pep14-23-based inhibition. *ACS Chem. Biol.* **2015**, *10*, 517–526. [CrossRef]
16. Jones, C.T.; Ma, L.; Burgner, J.W.; Groesch, T.D.; Post, C.B.; Kuhn, R.J. Flavivirus capsid is a dimeric alpha-helical protein. *J. Virol.* **2003**, *77*, 7143–7149. [CrossRef] [PubMed]
17. Van Gorp, E.C.M.; Suharti, C.; Mairuhu, A.T.A.; Dolmans, W.M.V.; van der Ven, J.; Demacker, P.N.M.; van der Meer, J.W.M. Changes in the plasma lipid profile as a potential predictor of clinical outcome in dengue hemorrhagic fever. *Clin. Infect. Dis.* **2002**, *34*, 1150–1153. [CrossRef]
18. Samsa, M.M.; Mondotte, J.A.; Iglesias, N.G.; Assuncao-Miranda, I.; Barbosa-Lima, G.; da Poian, A.T.; Bozza, P.T.; Gamarnik, A. V Dengue virus capsid protein usurps lipid droplets for viral particle formation. *PLoS Pathog.* **2009**, *5*, e1000632. [CrossRef]
19. Suvarna, J.C.; Rane, P.P. Serum lipid profile: A predictor of clinical outcome in dengue infection. *Trop. Med. Int. Heal.* **2009**, *14*, 576–585. [CrossRef] [PubMed]
20. Carvalho, F.A.; Carneiro, F.A.; Martins, I.C.; Assunção-Miranda, I.; Faustino, A.F.; Pereira, R.M.; Bozza, P.T.; Castanho, M.A.R.B.; Mohana-Borges, R.; da Poian, A.T.; 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.* **2012**, *86*, 2096–2108. [CrossRef]
21. Faustino, A.F.; Carvalho, F.A.; Martins, I.C.; Castanho, M.A.R.B.; Mohana-Borges, R.; Almeida, F.C.L.; da Poian, A.T.; Santos, N.C. Dengue virus capsid protein interacts specifically with very low-density lipoproteins. *Nanomed. Nanotechnol. Biol. Med.* **2014**, *10*, 247–255. [CrossRef]
22. Faustino, A.F.; Martins, I.C.; Carvalho, F.A.; Castanho, M.A.R.B.; Maurer-Stroh, S.; Santos, N.C. Understanding dengue virus capsid protein interaction with key biological targets. *Sci. Rep.* **2015**, *5*, 10592. [CrossRef]
23. Martins, A.S.; Carvalho, F.A.; Faustino, A.F.; Martins, I.C.; Santos, N.C. West Nile virus capsid protein interacts with biologically relevant host lipid systems. *Front. Cell. Infect. Microbiol.* **2019**, *9*, 8. [CrossRef]
24. Martins, A.S.; Martins, I.C.; Santos, N.C. Methods for lipid droplet biophysical characterization in *Flaviviridae* infections. *Front. Microbiol.* **2018**, *9*, 1951. [CrossRef] [PubMed]
25. Shang, Z.; Song, H.; Shi, Y.; Qi, J.; Gao, G.F. Crystal structure of the capsid protein from Zika virus. *J. Mol. Biol.* **2018**, *430*, 948–962. [CrossRef] [PubMed]
26. Kyte, J.; Doolittle, R.F. A Simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [CrossRef]
27. Deléage, G.; Roux, B. An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng.* **1987**, *1*, 289–294. [CrossRef] [PubMed]
28. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **2008**, *8*, 1–8. [CrossRef]
29. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Arbor, A.; Arbor, A. The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8. [CrossRef] [PubMed]
30. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2011**, *5*, 725–738. [CrossRef]
31. Dokland, T.; Walsh, M.; Mackenzie, J.M.; Khromykh, A.A.; Ee, K.-H.; Wang, S. West Nile virus core protein; tetramer structure and ribbon formation. *Structure* **2004**, *12*, 1157–1163. [CrossRef]
32. Zhan, C.; Zhao, L.; Chen, X.; Lu, W.; Lu, W. Total chemical synthesis of dengue 2 virus capsid protein via native chemical ligation: Role of the conserved salt-bridge. *Bioorg. Med. Chem.* **2013**, *21*, 3443–3449. [CrossRef]

33. Morando, M.A.; Barbosa, G.M.; Cruz-Oliveira, C.; da Poian, A.T.; Almeida, F.C.L. Dynamics of Zika virus capsid protein in solution: The properties and exposure of the hydrophobic cleft are controlled by the α-helix 1 sequence. *Biochemistry* **2019**, *58*, 2488–2498. [CrossRef]
34. Kumar, S.; Ravi, V.K.; Swaminathan, R. How do surfactants and DTT affect the size, dynamics, activity and growth of soluble lysozyme aggregates? *Biochem. J.* **2008**, *415*, 275–288. [CrossRef]
35. Lakowicz, J. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer Science, LLC: Berlin/Heidelberg, Germany, 2006; ISBN 9780387312781.
36. Rossi, P.; Yuanpeng, G.V.T.S.; James, J.H.; Anklin, C.; Conover, K.; Hamilton, K.; Xiao, R. A microscale protein NMR sample screening pipeline. *J. Biomol. NMR* **2010**, *46*, 11–22. [CrossRef] [PubMed]
37. Cho, C.H.; Urquidi, J.; Singh, S.; Robinson, G.W. Thermal offset viscosities of liquid $H_2O$, $D_2O$, and $T_2O$. *J. Phys. Chem. B* **1999**, *103*, 1991–1994. [CrossRef]
38. Martins, I.C.; Almeida, F.C.L.; Santos, N.C.; da Poian, A.T. DENV–Derived Peptides and Methods for the Inhibition of Flavivirus Replication. International Patent Publication Nr WO/2012/159187, 26 May 2011.
39. Ivanyi-Nagy, R.; Lavergne, J.; Gabus, C.; Ficheux, D.; Darlix, J.; Inserm, L.; Supe, E.N. RNA chaperoning and intrinsic disorder in the core proteins of *Flaviviridae*. *Nucleic Acids Res.* **2008**, *36*, 712–725. [CrossRef]
40. Ivanyi-Nagy, R.; Darlix, J. Core protein-mediated 5–3 annealing of the West Nile virus genomic RNA in vitro. *Virus Res.* **2012**, *167*, 226–235. [CrossRef]
41. Kumar, M.; Gromiha, M.M.; Raghava, G.P.S. SVM based prediction of RNA-binding proteins using binding residues and evolutionary information. *J. Mol. Recognit.* **2011**, *24*, 303–313. [CrossRef] [PubMed]
42. Järvelin, A.I.; Noerenberg, M.; Davis, I.; Castello, A. The new (dis)order in RNA regulation. *Cell Commun. Signal.* **2016**, *14*, 9. [CrossRef]
43. Shavinskaya, A.; Boulant, S.; Penin, F.; McLauchlan, J.; Bartenschlager, R. The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly. *J. Biol. Chem.* **2007**, *282*, 37158–37169. [CrossRef]
44. Cheng, A.; Wong, S.M.; Yuan, Y.A. Structural basis for dsRNA recognition by NS1 protein of influenza A virus. *Cell Res.* **2009**, *19*, 187–195. [CrossRef]
45. Fernandez-Sesma, A.; Marukian, S.; Ebersole, B.J.; Kaminski, D.; Park, M.S.; Yuen, T.; Sealfon, S.C.; Garcia-Sastre, A.; Moran, T.M. Influenza virus evades innate and adaptive immunity via the NS1 protein. *J. Virol.* **2006**, *80*, 6295–6304. [CrossRef]
46. Wang, S.H.; Syu, W.J.; Huang, K.J.; Lei, H.Y.; Yao, C.W.; King, C.C.; Hu, S.T. Intracellular localization and determination of a nuclear localization signal of the core protein of dengue virus. *J. Gen. Virol.* **2002**, *83*, 3093–3102. [CrossRef] [PubMed]
47. Kobe, B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α. *Nat. Struct. Biol.* **1999**, *6*, 388–397. [CrossRef] [PubMed]
48. Catimel, B.; Teh, T.; Fontes, M.R.M.; Jennings, I.G.; Jans, D.A.; Howlett, G.J.; Nice, E.C.; Kobe, B. Biophysical characterization of interactions involving importin-α during nuclear import. *J. Biol. Chem.* **2001**, *276*, 34189–34198. [CrossRef] [PubMed]
49. Marfori, M.; Mynott, A.; Ellis, J.J.; Mehdi, A.M.; Saunders, N.F.W.; Curmi, P.M.; Forwood, J.K.; Boden, M.; Kobe, B. Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim. Biophys. Acta* **2011**, *1813*, 1562–1577. [CrossRef] [PubMed]
50. Fontes, M.R.M.; Teh, T.; Kobe, B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-α. *J. Mol. Biol.* **2000**, *297*, 1183–1194. [CrossRef] [PubMed]
51. Marfori, M.; Lonhienne, T.G.; Forwood, J.K.; Kobe, B. Structural basis of high-affinity nuclear localization signal interactions with importin-alpha. *Traffic* **2012**, *13*, 532–548. [CrossRef] [PubMed]
52. 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.* **1989**, *70*, 1409–1415. [CrossRef] [PubMed]
53. Makino, Y.; Tadano, M.; Anzai, T.; Ma, S.P.; Yasuda, S.; Žagar, E. 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.* **1989**, *70*, 1417–1425. [CrossRef] [PubMed]
54. Wagstaff, K.M.; Sivakumaran, H.; Heaton, S.M.; Harrich, D.; Jans, D.A. Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. *Biochem. J.* **2012**, *443*, 851–856. [CrossRef]

55. Bergmann, M.; Garcia-Sastre, A.; Carnero, E.; Pehamberger, H.; Wolff, K.; Palese, P.; Muster, T. Influenza virus NS1 protein counteracts PKR-mediated inhibition of replication. *J. Virol.* **2000**, *74*, 6203–6206. [CrossRef]
56. Kochs, G.; Garcia-Sastre, A.; Martinez-Sobrido, L. Multiple anti-interferon actions of the influenza A virus NS1 protein. *J. Virol.* **2007**, *81*, 7011–7021. [CrossRef] [PubMed]
57. Rodriguez-Madoz, J.R.; Bernal-Rubio, D.; Kaminski, D.; Boyd, K.; Fernandez-Sesma, A. Dengue virus inhibits the production of type I interferon in primary human dendritic cells. *J. Virol.* **2010**, *84*, 4845–4850. [CrossRef] [PubMed]
58. Min, J.Y.; Krug, R.M. The primary function of RNA binding by the influenza A virus NS1 protein in infected cells: Inhibiting the 2′–5′ oligo (A) synthetase/RNase L pathway. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7100–7105. [CrossRef]
59. Uversky, V.N. Intrinsically disordered proteins and their "mysterious" (meta)physics. *Front. Phys.* **2019**, *7*, 10. [CrossRef]
60. Na, J.H.; Lee, W.K.; Yu, Y.G. How do we study the dynamic structure of unstructured proteins: A case study on nopp140 as an example of a large, intrinsically disordered protein. *Int. J. Mol. Sci.* **2018**, *19*, 381. [CrossRef]
61. Uversky, V.N. Introduction to intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6557–6560. [CrossRef] [PubMed]
62. Minde, D.P.; Halff, E.F.; Tans, S. Designing disorder: Tales of the unexpected tails. *Intrinsically Disord. Proteins* **2013**, *1*, e26790. [CrossRef]
63. Krystkowiak, I.; Manguy, J.; Davey, N.E. PSSMSearch: A server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. *Nucleic Acids Res.* **2018**, *46*, W235–W241. [CrossRef]
64. Bera, A.K.; Kuhn, R.J.; Smith, J.L. Functional characterization of cis and trans activity of the Flavivirus NS2B-NS3 protease. *J. Biol. Chem.* **2007**, *282*, 12883–12892. [CrossRef]
65. Niyomrattanakit, P.; Yahorava, S.; Mutule, I.; Mutulis, F.; Petrovska, R.; Prusis, P.; Katzenmeier, G.; Wikberg, J.E. Probing the substrate specificity of the dengue virus type 2 NS3 serine protease by using internally quenched fluorescent peptides. *Biochem. J.* **2006**, *397*, 203–211. [CrossRef]
66. Sievers, F.; Higgins, D.G. Clustal omega. *Curr. Protoc. Bioinform.* **2014**, *13*, 1–16.
67. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; Thompson, J.D.; Higgins, D.G.; Mcwilliam, H.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef]
68. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef]
69. Irie, K.; Mohan, P.; Sasaguri, Y.; Putnak, R.; Padmanabhan, R. Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene* **1989**, *75*, 197–211. [CrossRef]
70. Smith, P.; van Gunsteren, W. Translational and rotational diffusion of proteins. *J. Mol. Biol.* **1994**, *236*, 629–636. [CrossRef] [PubMed]
71. Mok, Y.; de Prat Gay, G.; Butler, P.; Bycroft, M. Equilibrium dissociation and unfolding. *Protein Sci.* **1996**, *5*, 310–319. [CrossRef]
72. Rumfeldt, J.; Galvagnion, C.; Vassall, K.; Meiering, E. Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.* **2008**, *98*, 61–84. [CrossRef]
73. Neet, K.E.; Timm, D.E. Conformational stability of dimeric proteins: Quantitative studies by equilibrium denaturation. *Protein Sci.* **1994**, *3*, 2167–2174. [CrossRef]
74. Allen, D.L.; Pielak, G.J. Baseline length and automated fitting of denaturation data. *Protein Sci.* **1998**, *7*, 1262–1263. [CrossRef]