



Article

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

André F. Faustino <sup>1,†</sup> , Ana S. Martins <sup>1</sup>, Nina Karguth <sup>1</sup>, Vanessa Artilheiro <sup>1</sup>,  
Francisco J. Enguita <sup>1</sup>, Joana C. Ricardo <sup>2,‡</sup> , Nuno C. Santos <sup>1,\*</sup> and Ivo C. Martins <sup>1,\*</sup>

<sup>1</sup> Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisbon, Portugal

<sup>2</sup> 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  $\alpha$ -helices,  $\alpha 1$  to  $\alpha 4$ , 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 ( $\alpha 2$ – $\alpha 2'$  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 ( $\alpha 4$ – $\alpha 4'$  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].

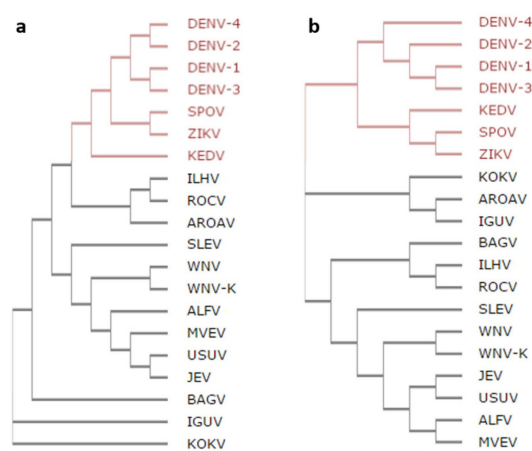
DENV infection affects the host lipid metabolism, increasing host intracellular LDs and unbalancing plasma lipoprotein levels and composition [17–19]. Importantly, DENV C binds LDs, an interaction essential for viral replication [18,20]. DENV C-LDs binding requires potassium ions, the LDs surface protein perilipin 3 (PLIN3) and involves specific amino acid residues of DENV C  $\alpha 2$ – $\alpha 2'$  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  $\alpha$ -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

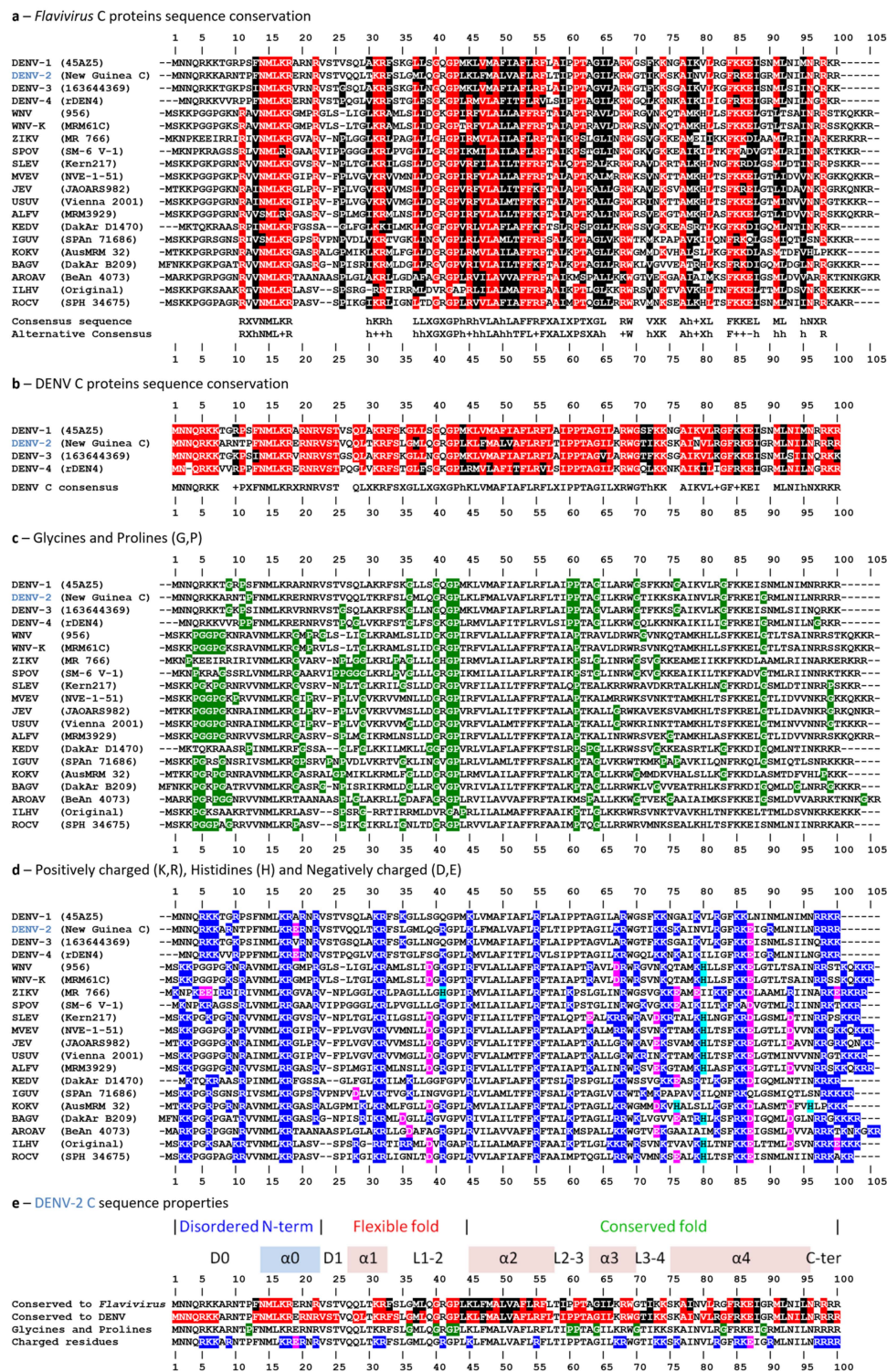
### 2.1. Analysis of Amino Acid Sequence Conservation Among Flavivirus C proteins

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  $\alpha 2$  helix;  $^{68}RW^{69}$  of  $\alpha 3$  helix; and, finally, the  $^{84}F++-h^{88}$  motif from  $\alpha 4$  (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  $\alpha$ -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.

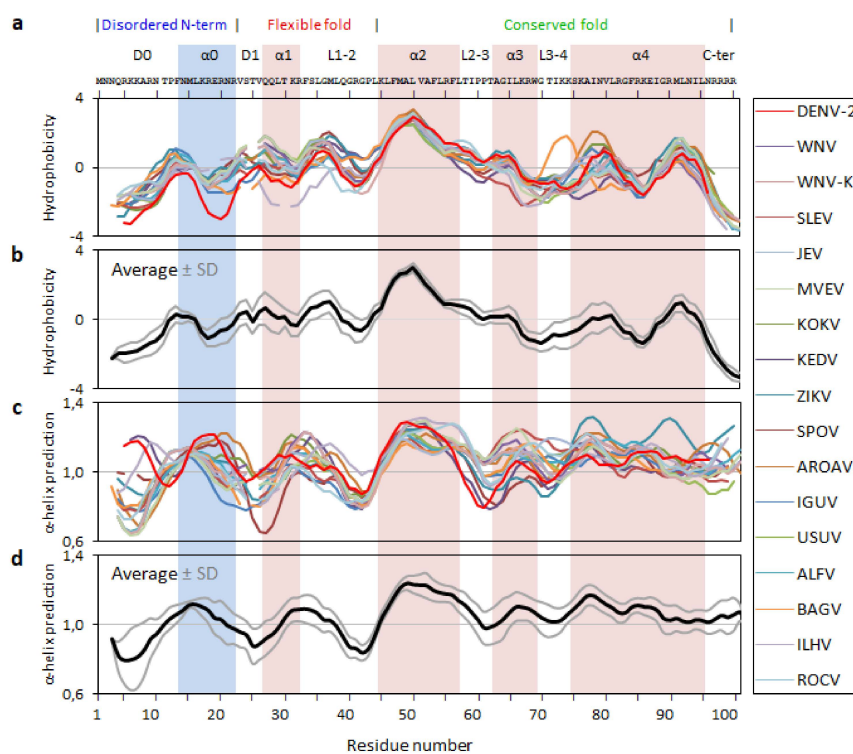


**Figure 2.** *Flavivirus* C proteins amino acid residues sequence conservation. (a) Mosquito-borne *Flavivirus* C protein are 55% conserved, with residues being alternative conserved if, in a given position, more than 15 are equal (red) or stereochemically similar (black). (b) Conservation between DENV 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 *Flavivirus* C Protein Sequences Hydrophobicity and Secondary Structure Propensity

Hydrophobicity and  $\alpha$ -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  $\alpha 0$  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  $\alpha 3$  and  $\alpha 4$  domains, with the most hydrophobic domain being  $\alpha 2$ , as expected from the sequence analysis (Figure 2) and from the literature [12,14,18].



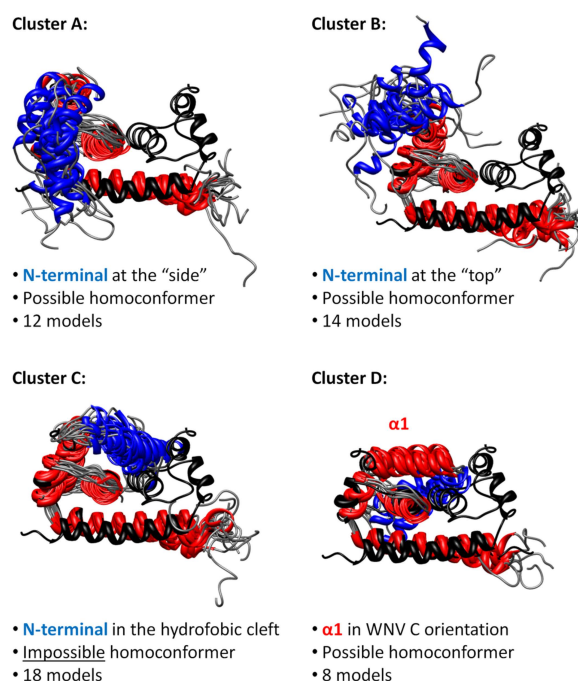
**Figure 3.** *Flavivirus* C proteins hydrophobicity and secondary structure predictions. (a) Hydrophobicity predictions and (b) respective average (black line)  $\pm$  standard deviation, SD (gray lines). (c)  $\alpha$ -helical secondary structure predictions and (d) respective average (black line)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

For  $\alpha$ -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  $\alpha 0$  occurring for these proteins, as hypothesized earlier [15]. Roughly, between positions 12 to 20, occurs a disordered region

with high tendency to acquire  $\alpha$ -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  $\alpha$ -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  $\alpha$ -helix conformation are in blue, the other  $\alpha$ -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  $\alpha 1$  helix in the DENV C experimentally determined conformation (Protein Data Bank (PDB) ID: 1R6R [12]). In cluster D the  $\alpha 1$  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  $\alpha 4$ – $\alpha 4'$  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  $\alpha 1$ – $\alpha 2$ – $\alpha 2'$ – $\alpha 1'$  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  $\alpha 1$  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  $\alpha 1$  domain, and partially blocks the  $\alpha 2$ – $\alpha 2'$  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
<b>Total</b>	<b>12</b>	<b>14</b>	<b>18</b>	<b>8</b>	<b>28</b>

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.

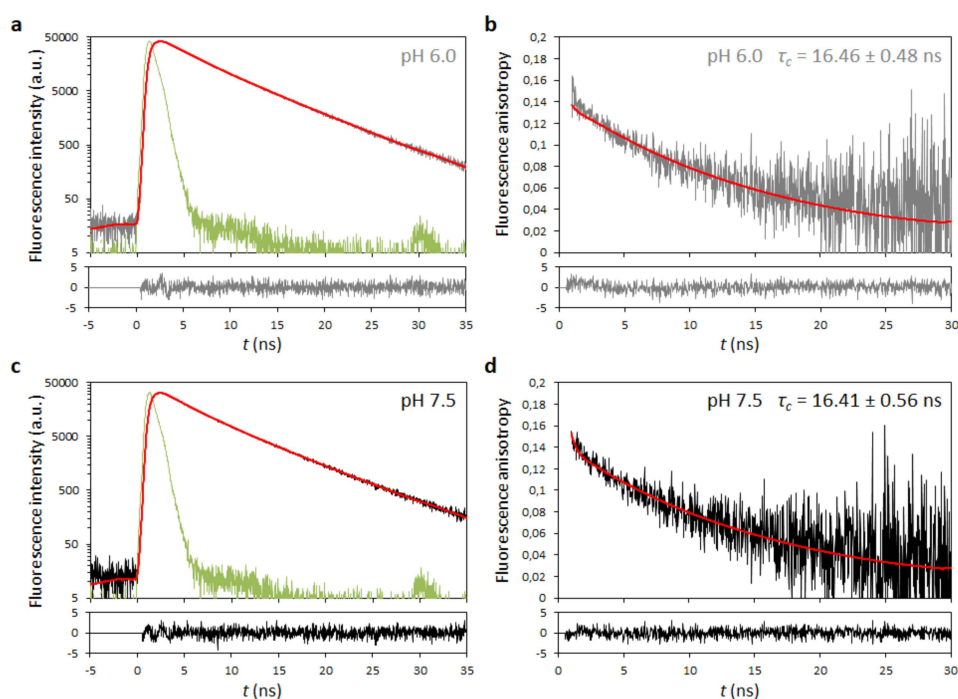
#### 2.4. Analysis of Dengue Virus (DENV) C Protein Rotational Correlation Time

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 ( $\pm$ % 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 ( $\pm$ 3.9%)	0.520 ( $\pm$ 4.0%)
$\tau_2$ (ns)	3.106 ( $\pm$ 0.4%)	3.108 ( $\pm$ 0.9%)
$\tau_3$ (ns) *	6.328 ( $\pm$ 0.4%)	6.506 ( $\pm$ 0.4%)
$\alpha_1$ *	0.275 ( $\pm$ 0.7%)	0.178 ( $\pm$ 3.4%)
$\alpha_2$ *	0.315 ( $\pm$ 0.9%)	0.385 ( $\pm$ 0.4%)
$\alpha_3$ *	0.410 ( $\pm$ 0.4%)	0.437 ( $\pm$ 0.4%)
$\tau_c$ (ns) *	16.46 ( $\pm$ 2.9%)	16.41 ( $\pm$ 3.4%)
$r_0$	0.130 ( $\pm$ 0.8%)	0.131 ( $\pm$ 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  $\sim 6.4$  ns (with a significant weight  $\alpha_3$  of  $\sim 0.42$ ), this means that, at both pH values, we could measure  $\tau_c$  values up to a limit of  $\sim 19$  ns. In both pH conditions, the  $\tau_c$  measured was



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

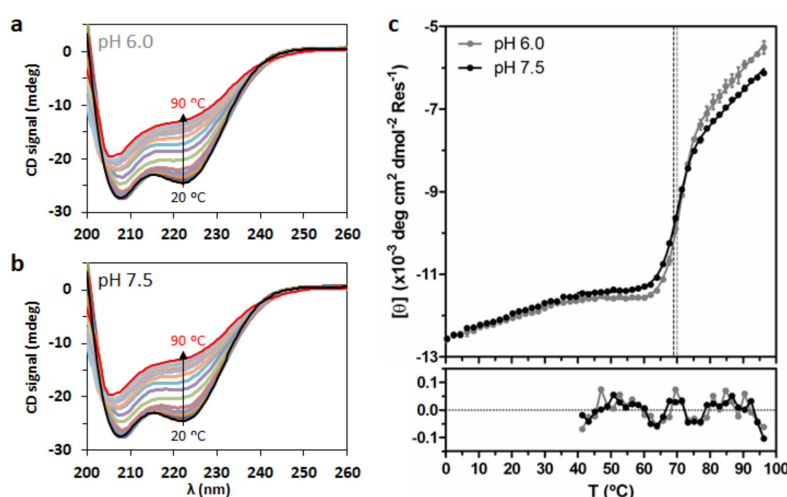
Rossi et al. [36] correlated the  $\tau_c$  of 16 globular proteins at 20 °C with their molecular weight (MW in kDa), based on NMR data, leading to the relation:  $\tau_c \approx 0.6$  MW. Assuming DENV C as a 23.5 kDa fully globular homodimer and correcting for the temperature ( $T$ ) and viscosity ( $\eta$ ) [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  $15.2 \pm 0.5$  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<sub>2</sub>O were calculated using Equation (10)).

$\tau_c$ (ns) at $T$	$T$ (°C)	$\tau_c$ (ns) at 25 °C in H <sub>2</sub> O	Method	Source
$16.4 \pm 0.5$	22	$15.2 \pm 0.5$	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) $\approx 0.6 \times$ 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  $\alpha$ -helical structure is partially lost upon increasing temperature (Figure 6a,b). However, even at 96 °C, the protein does not 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,  $[\theta]$ , as a function of temperature,  $T$ , reveals a transition at  $\sim 70$  °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,  $[\theta]$ , 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 ( $T_m^\circ$ ), the enthalpy variation at  $T_m^\circ$  ( $\Delta H_{T_m^\circ}^\circ$ ) and the entropy variation at  $T_m^\circ$  ( $\Delta S_{T_m^\circ}^\circ$ ), with all parameters at standard thermodynamics conditions (symbolized by  $^\circ$ ). 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  $\pm$  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 $\pm$ 0.63	69.03 $\pm$ 0.65
$T_m^\circ$ (°C)	88.26 $\pm$ 0.80	88.80 $\pm$ 0.83
$\Delta H_{T_m^\circ}^\circ$ (kJ mol <sup>-1</sup> )	612 $\pm$ 26	564 $\pm$ 23
$\Delta S_{T_m^\circ}^\circ$ (kJ mol <sup>-1</sup> K <sup>-1</sup> )	1.693 $\pm$ 0.073	1.557 $\pm$ 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 <sup>40</sup>GXGP<sup>43</sup> motif at loop L1-2, that marks the transition from the flexible to the conserved fold region [14]; the <sup>68</sup>RW<sup>69</sup> at  $\alpha 3$  forms a hydrophobic pocket that accommodates the W69 side chain involving residues from  $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$  [12,32]; and, the <sup>44</sup>h+hhLAhhAFF+F<sup>56</sup> and <sup>84</sup>F++-h<sup>88</sup> motifs, respectively from  $\alpha 2$  and  $\alpha 4$  helices, maintain the homodimer structure both via the  $\alpha 2$ - $\alpha 2'$  hydrophobic interaction and via the salt bridges of residues [RK]<sup>45</sup> and [RK]<sup>55'</sup> with [ED]<sup>87</sup> [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  $\alpha$ -helix propensities (Figure 3). The conserved motif  $^{13}\text{hNML}+\text{R}^{18}$ , at the N-terminal region, and the  $\alpha 2-\alpha 2'$  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  $\alpha 2-\alpha 2'$  and  $\alpha 4-\alpha 4'$  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  $\alpha$ -helical propensity and  $\alpha 2-\alpha 2'$  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  $\alpha 1$  aligns with WNV C [14,31] and ZIKV C [25]. Moreover, if two monomers are in a D conformation (D–D' homoconformer), the dimer  $\alpha 2-\alpha 2'$  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  $\alpha 4-\alpha 4'$  Cluster A (Figure 4), the most likely RNA binding site [12,39,40]. Furthermore, binding to RNA via the C-terminal  $\alpha 4-\alpha 4'$  interface may be favored by a previous or simultaneous interaction of the protein with host LDs via the N-terminal region and  $\alpha 2-\alpha 2'$  interface. Access to  $\alpha 2-\alpha 2'$  (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  $\tau_c$ ,  $15.2 \pm 0.5$  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  $\sim 70$  °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.

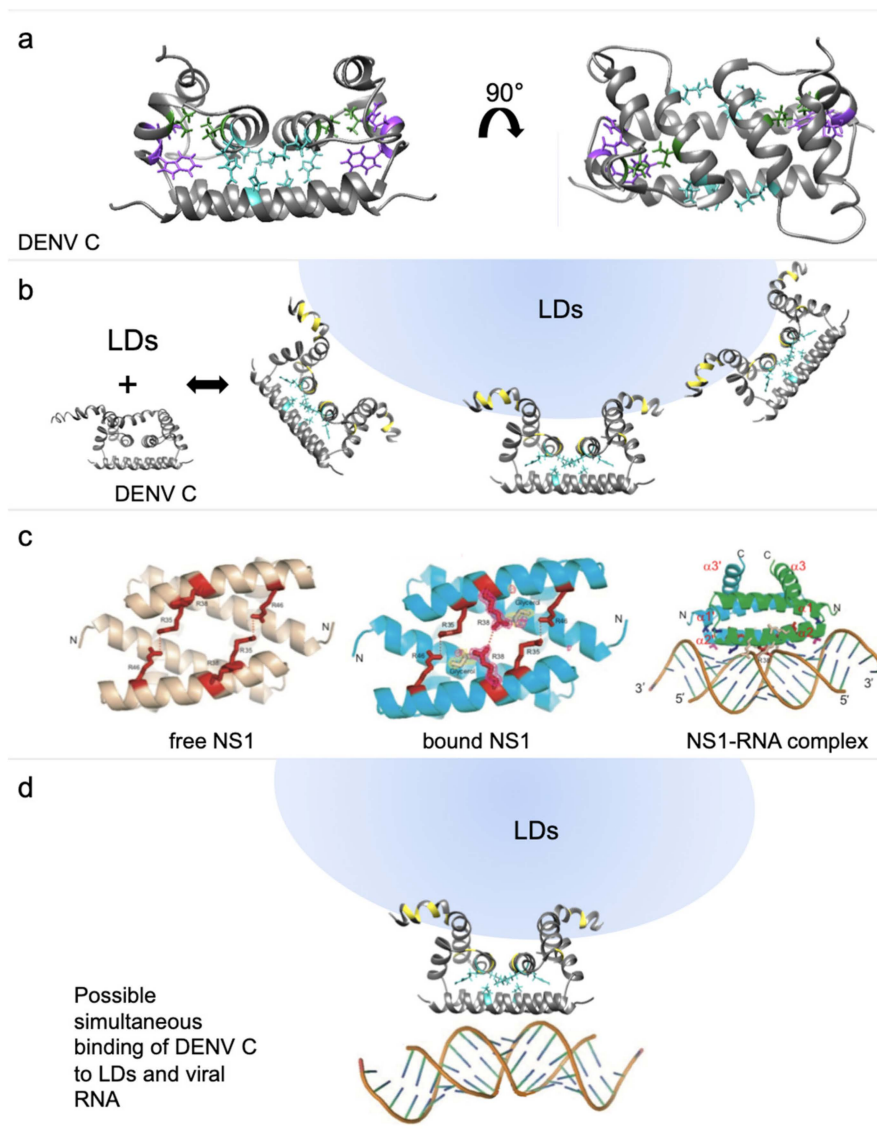
These findings must also be considered in light of DENV C biologically relevant interactions with LDs [22] and RNA (Figure 7). DENV C experimental structure [12] contains three distinct structural regions [13]: a disordered N-terminal region (from the N-terminal up to residue R22), a flexible fold (residues V23 to L44, where  $\alpha$ -helix 1 is located) and a conserved fold with helices  $\alpha 2$ ,  $\alpha 3$  and

$\alpha 4$ , containing the R68 and W69 amino acid residues, highly conserved among *Flavivirus* [12]. R68 terminates  $\alpha 3$  helix, with its side chain pointing to the protein interior [12]. W69 locates at DENV C  $\alpha 4$ – $\alpha 4'$  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 ( $\alpha 2$ – $\alpha 2'$  dimer interface) and its remaining sections, namely the  $\alpha 4$ – $\alpha 4'$  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  $\alpha 2$ – $\alpha 2'$  interface [14]. V51 of  $\alpha 2$  is affected by the interaction with LDs and stabilizes the dimer by contacting with  $\alpha 3$  (I65). Another interaction via salt bridges, between  $\alpha 2$  (K45 and R55') and  $\alpha 4$  (E87), stabilizes the homodimer (Figure 7a). The C protein binding to host LDs, which affects the  $\alpha 2$ – $\alpha 2'$ , can lead to changes in the  $\alpha 4$ – $\alpha 4'$  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  $\alpha 4$ – $\alpha 4'$  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  $\alpha$  and  $\beta$  and works as a viral immuno-suppressor by weakening the host cell gene expression [45]. DENV C was also reported to have an importin  $\alpha$ -like motif in the N-terminal [15,46]. Regarding the targets that may interact with importin  $\alpha$  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  $\alpha 1$  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  $\alpha$ , 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  $\beta$ , given the similarities between the N-terminal region of DENV C and importin  $\alpha$  [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  $\alpha$  and/or importin  $\beta$ . Ivermectin, a specific inhibitor of importin  $\alpha/\beta$ -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  $\alpha$  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  $\alpha 4$ – $\alpha 4'$  interface occurs after the contact

of its  $\alpha 2$ – $\alpha 2'$  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  $\alpha 4$ – $\alpha 4'$ . As with influenza NS1, the *Flavivirus* C protein would remain in the same overall fold, but a small opening of  $\alpha 4$ – $\alpha 4'$  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  $\alpha 4$ – $\alpha 4'$  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  $\alpha 4$  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  $\alpha 4$ – $\alpha 4'$  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  $\alpha 4$ – $\alpha 4'$  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  $\alpha 4$ – $\alpha 4'$  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  $\alpha$ -helical region net charge (+8 for a monomer, +16 for  $\alpha 4$ – $\alpha 4'$  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  $\alpha$ -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](http://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  $\text{KH}_2\text{PO}_4$ , 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  $-80\text{ }^\circ\text{C}$ . 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  $\mu\text{M}$  in 50 mM  $\text{KH}_2\text{PO}_4$ , 200 mM KCl, pH 6.0 or pH 7.5, with 550  $\mu\text{L}$  total volume, in 0.5 cm  $\times$  0.5 cm quartz cuvettes. The instrument response function,  $\text{IRF}(t)$ , 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  $^\circ\text{C}$ . Time-resolved fluorescence intensity measurements with picosecond-resolution were obtained by the time-correlated single-photon timing (TCSP) 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 ( $\sim 54.7^\circ$ , 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:

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

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( $t$ ):

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

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 below 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]:

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

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 ( $\sim 1.96 \times \text{SE}$ ) 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 ( $\pm 5$  °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<sub>2</sub>O or 10% D<sub>2</sub>O 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 <sub>2</sub> O (cP)	$\eta$ in 10% D <sub>2</sub> O (cP)	$\frac{\eta_b T_a}{\eta_a T_b}$ in H <sub>2</sub> O	$\frac{\eta_b T_a}{\eta_a T_b}$ in 10% D <sub>2</sub> O
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  $\mu$ M in 50 mM KH<sub>2</sub>PO<sub>4</sub>, 200 mM KCl, pH 6.0 or pH 7.5, with 220  $\mu$ L of total volume. Spectra were smoothed through the means-movement method (using 7 points) and normalized to mean residue molar ellipticity,  $[\theta]$  (in deg cm<sup>2</sup> dmol<sup>-1</sup> Res<sup>-1</sup>).

For the CD temperature denaturation data treatment, we assumed a dimer to monomer denaturation model [71–73] in which the folded dimer, F<sub>2</sub>, separates into unfolded monomers, U, in a single step described by reaction 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]:

$$f_F = \frac{2[F_2]}{[P_m]} \quad (12)$$

$$f_U = \frac{[U]}{[P_m]} \quad (13)$$

$$f_F + f_U = 1 \quad (14)$$

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

$$[U] = f_U[P_m] \quad (15)$$

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

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

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

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

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

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  $f_U$ :

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

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]:

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

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

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

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

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

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  $\Delta H^\circ_{T_m}$ , the specific heat capacity at constant pressure,  $\Delta C_p^\circ$ , and the standard conditions' denaturation temperature,  $T_m^\circ$ , according to [74]:

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

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

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

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  $\Delta H_{T_m}^\circ$  and  $T_m^\circ$ , along with the respective SE values. Afterwards,  $\Delta S_{T_m}^\circ$  can be obtained, since  $\Delta G^\circ = 0 \text{ kJ mol}^{-1}$  at  $T_m^\circ$ , via the following Equation:

$$\Delta H_{T_m}^\circ - T_m^\circ \Delta S_{T_m}^\circ = 0 \Rightarrow \Delta S_{T_m}^\circ = \frac{\Delta H_{T_m}^\circ}{T_m^\circ} \quad (25)$$

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

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

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  $T_m^\circ$  ( $T_m^\circ$  value estimated if  $[P_m] = 1M$ ), 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]:

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

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

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

- 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]
- 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).
- Durbin, A.P. A dengue vaccine. *Cell* **2016**, *166*, 1. [CrossRef] [PubMed]
- 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]
- 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).
- ICTV Taxonomy. Available online: <https://talk.ictvonline.org/taxonomy/> (accessed on 17 April 2019).
- 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.* **2019**, *91*, 87–94. [CrossRef] [PubMed]
- Schubert, A.M.; Putonti, C. Infection, genetics and evolution of the sequence composition of flaviviruses. *Infect. Genet. Evol.* **2010**, *10*, 129–136. [CrossRef] [PubMed]
- Calisher, C.H.; Gould, E.A. Taxonomy of the virus family *Flaviviridae*. *Adv. Virus Res.* **2003**, *59*, 1–19. [PubMed]
- 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  $\alpha$ -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<sub>2</sub>O, D<sub>2</sub>O, and T<sub>2</sub>O. *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  $\alpha$ . *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- $\alpha$  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- $\alpha$ . *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  $\alpha/\beta$ -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](#)]

