The disordered N-terminal region of dengue virus capsid protein contains a lipid-droplet-binding motif

Ivo C. MARTINS*,†, Francisco GOMES-NETO*,‡, André F. FAUSTINO†, Filomena A. CARVALHO†, Fabiana A. CARNEIRO§, Patricia T. BOZZA¶, Ronaldo MOHANA-BORGES‖, Miguel A. R. B. CASTANHO†, Fábio C. L. ALMEIDA*‡, Nuno C. SANTOS† and Andrea T. DA POIAN*†

*Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil, †Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisbon, 1649-028, Portugal, ¶Centro Nacional de Ressonância Magnética Nuclear, Universidade Federal do Rio de Janeiro and National Institute of Structural Biology and Bioimaging, Rio de Janeiro, RJ, 21941-902, Brazil, ‡Polo Avanço de Xerém, Universidade Federal do Rio de Janeiro and National Institute of Structural Biology and Bioimaging, Rio de Janeiro, RJ, 21941-902, Brazil, §Centro Nacional de Ressonância Magnética Nuclear, Universidade Federal do Rio de Janeiro and National Institute of Structural Biology and Bioimaging, Rio de Janeiro, RJ, 21941-902, Brazil, ‖Laboratório de Genomica Estrutural, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil

Dengue is the major arthropod-borne human viral disease, for which no vaccine or specific treatment is available. We used NMR, zeta potential measurements and atomic force microscopy to study the structural features of the interaction between dengue virus C (capsid) protein and LDs (lipid droplets), organelles crucial for infectious particle formation. C protein-binding sites to LD were mapped, revealing a new function for a conserved segment in the N-terminal disordered region and indicating that conformational selection is involved in recognition. The results suggest that the positively charged N-terminal region of C protein prompts the interaction with negatively charged LDs, after which a conformational rearrangement enables the access of the central hydrophobic patch to the LD surface. Taken together, the results allowed the design of a peptide with inhibitory activity of C protein–LD binding, paving the way for new drug development approaches against dengue.

Key words: atomic force microscopy (AFM), capsid protein, conformational selection, dengue virus (DENV), intrinsically disordered protein, lipid droplet, NMR.

INTRODUCTION

Viruses have evolved several strategies to manipulate cellular functions during infection, allowing their efficient replication and spread. This is possible even with very small viral genomes, as viral proteins usually accumulate distinct activities and properties, being remarkable examples of multifunctional proteins. It can be hypothesized that the diversity of functions of viral proteins could be achieved through a high structural flexibility, conferred by the presence of intrinsically disordered regions, which are interconverting dynamic structures that lack a defined folding, yet are able to carry out specific functions [1,2]. Indeed, a comparative structural analysis of viral and cellular proteins revealed that viral proteins show a high occurrence of short disordered regions and loosely packed cores, suggesting that they have evolved to be more adapted to changes in the environment than to thermodynamic stability [3]. An analysis of the functions carried out by the intrinsically disordered regions of viral proteins showed that they are usually involved in recognition, regulation, signalling and interaction with nucleic acids and proteins, functions for which structural flexibility is required [4].

The capsid (or core) proteins from viruses of the family Flaviviridae are good examples in which the diversity of functions can be attributed to their intrinsically disordered nature [5]. Flaviviruses are positive-sense single-stranded RNA viruses that comprise several human pathogens, including DENV (dengue virus), WNV (West Nile virus), TBEV (tick-borne encephalitis virus), JEV (Japanese encephalitis virus) and YFV (yellow fever virus). Their genomic RNA is translated from a single open reading frame, generating a polyprotein that is cleaved by cellular and viral proteases into three structural proteins [C (capsid), prM (pre-membrane) and E (envelope)] and seven NS (non-structural) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [6]. Flavivirus capsid proteins can be considered multifunctional proteins, since they bind to viral RNA in nucleocapsid assembly [7], act as nucleic acid chaperones [8] and interact with a variety of cellular proteins, modulating transcription, cellular metabolism, apoptosis and immune response [9–11]. The flexibility conferred by their intrinsically disordered regions is expected to facilitate these functions [5].

The present study focused on the structural features of DENV C protein necessary for its interaction with LDs (lipid droplets), endoplasmic reticulum-derived organelles crucial for virus replication and infectious particle formation [12], whose biogenesis is stimulated by infection [13]. DENV causes the most important arthropod-borne human viral disease, with three billion people at risk, 50–100 million infections and more than 20 000 deaths annually [14]. DENV infection can evolve to a severe and life-threatening disease, for which no specific treatment is currently available. Additionally, vaccine development against DENV is still a challenge [14]. Therefore a detailed structural understanding of the DENV life cycle is of utmost importance to tackle this problem.
DENV particles are composed of a lipid bilayer surrounding a nucleocapsid, in which genomic RNA is located, complexed with multiple copies of the DENV C protein [15]. This protein forms symmetric homodimers in solution, with each monomer containing four α-helices (α1–α4) connected by short loops [16,17]. The segment containing the first 20 amino acid residues in DENV C protein was found to be disordered in solution [17], in agreement with the disorder prediction for the same region of WNV and YFV C proteins [5]. The protein structure and surface charge distribution suggest that capsid assembly occurs via C protein interaction with lipid membranes through a central hydrophobic region located in the α2/α2′ interface, which orients the positively charged α4/α4′ C-terminus area, allowing its interaction with the viral RNA [17]. In agreement with this proposal, mutation of two residues in the α2/α2′ interface disrupted LD targeting and abolished virus particle formation [12].

In the present study, using high-resolution NMR, zeta potential measurements and AFM (atomic force microscopy)-based force spectroscopy to map DENV C protein LD-binding sites, we identified local structural changes involving specific conserved amino acids residues located in the disordered N-terminal region, in the central hydrophobic patch and in the loop between α1 and α2. On the basis of these results, it was possible to attribute a new function to the N-terminal disordered region of DENV C protein, reinforcing the concept of multiple functions for intrinsically disordered regions in proteins. Furthermore, we applied that information to develop an original and successful peptide-based approach for the specific inhibition of the C protein–LD interaction.

**EXPERIMENTAL**

**Chemicals**

The peptides pep14–23 (NMLKRARNRV) and pep5–26 (RKKTGRPSFNMLKRARNVSTV) were purchased from JPT Peptide Technologies. [15N]NH4Cl and D2O were purchased from Cambridge Isotope Laboratories.

**Recombinant DENV C protein expression and purification**

The C protein of DENV serotype 2 New Guinea strain was expressed and purified following established procedures [16,18], except that M9 minimal medium supplemented with [15N]NH4Cl was used for expression of 15N-labelled C protein.

**Production and purification of LDs**

BHK (baby-hamster kidney)-21 cells were maintained in α-MEM (minimum essential medium-α) supplemented with 10% (v/v) fetal bovine serum at 37 °C in 5% CO2. At 80% confluence, cells were treated with 10 μM oleic acid for 24 h. LDs were isolated by nitrogen cavitation (model 4639, Parr Instrument Company) and purified by sucrose-gradient ultracentrifugation, as described previously [18]. To ensure LD purity, only fractions negative for lactate dehydrogenase activity were used. LDs samples were kept at 4 °C and used within 2 weeks.

**NMR spectroscopic analysis**

NMR experiments were performed at 300 K in Bruker Avance III 800 MHz and 600 MHz instruments, equipped with triple-resonance (1H,13C,15N) probes. C protein (200 μl) in 50 mM phosphate buffer, pH 6.0, and 0.2 M NaCl, were added to 125 μl of LDs (~4×10^10/ml) in TEE buffer (25 mM Tris/HCl, pH 7.4, 1 mM EDTA and 1 mM EGTA), or to TEE buffer alone, to obtain a final concentration of uniformly labelled 15N-DENV C protein between 250 and 275 μM. Chemical shifts were referenced with respect to the H2O signal at 4.77 p.p.m. (pH 6.8, 25 °C) relative to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). NMR spectra were processed and analysed using NMRPipe [19] and NMRView [20]. Chemical shift assignments were obtained by comparison with the deposited data for C protein (Biomagnetic Resonance Data Bank 5973 [17]). Ambiguous assignments were solved using three-dimensional 15N-edited NOESY [NOE (nuclear Overhauser enhancement) spectroscopy]-HSQC (heteronuclear single-quantum coherence) and TOCSY-HSQC by looking at αN (i,i+1) and NH (i,i+1) NOEs, and 85% of the backbone amidic resonances were unambiguously assigned. NOESY-HSQC (100 ms mixing time) and TOCSY-HSQC (60 ms MLEV) were collected with and without LDs, at 800.13 MHz, 1024×48×64 complex points. Residues Met1, Asn2, Thr14, Leu40 and Arg99 were not assigned in the previously determined NMR structure [17]. The assignments of the residues Asp1, Gin4, Asn10, Gly38, Gin39, Thr62, Ala77 and Arg100 could not be localized or confirmed. Quadrature detection in indirect dimension was done using States-TPPI (time proportional phase incrementation) (1H) and echo–anti-echo (15N). Gradient selection HSQC spectra were acquired at 800.13 MHz with 1024×128 complex points (echo–anti-echo). Peptide–LD interaction experiments (using pep14–23 or pep5–26) were performed at the same conditions as for C protein, except that the final concentrations were 400 or 800 μM. Sequential backbone resonance assignment was performed for free peptides using (1H,13C)-gradient-selection-HSQC (natural abundance of 13C), NOESYS (100 and 200 ms mixing time) and TOCSY (60 ms MLEV) in 10% 2H2O, acquired at 800.13 MHz, with 4096×512 complex points for NOESY and TOCSY, and 1024×128 for HSQC. Quadrature detection in indirect dimension was done using States-TPPI (1H) for NOESY and TOCSY and echo–anti-echo (13C) for the HSQC. Water suppression was achieved using the 3-9-19 WATERGATE sequence for NOESY and TOCSY [21].

**Alignment of the sequence of the C proteins from mosquito-born flaviviruses**

DENV C protein sequence from strain New Guinea was aligned with other flavivirus C protein sequences using BLASTp [22]. The search was performed against non-redundant protein sequences and excluding the DENV group {TaxID [NCBI (National Center for Biotechnology Information) Taxonomy identifier] 11052; http://www.ncbi.nlm.nih.gov/Taxonomy/]. Only Flavivirus spp. gave a score high enough to be considered for further analysis. The best-scored sequence of each different virus strain was selected. Virus taxonomical classification was accessed by the NCBI Taxonomy browser and the International Committee on Taxonomy of Viruses 2009 list (http://ictvonline.org/virusTaxonomy.asp?version=2009). Unclassified Flavivirus spp. (TaxID 11051) were rejected. As a result, 16 polypeptide sequences, each from an independent Flavivirus spp. strain remained. In all of them, the C protein aligned sequence was in the first ~110 amino acid residues. The residues next to the NS2B-NS3 protease cleavage site were excluded from further analysis in all polypeptides, leaving the presumed C protein sequences, which were aligned by multiple alignments on the Clustal W2 tool of the European Bioinformatics Institute web server [23] against the DENV C protein sequence. In Supplementary Table
S1 (at http://www.BiochemJ.org/bj/444/bj4440405add.htm), the GenBank® accession numbers of each virus strain studied can be retrieved.

**Structural comparison of DENV and WNV-K (WNV strain Kunjin) C proteins**

The UCSF Chimera software package [24] was used to superimpose DENV C protein model number 21 (PDB code 1R6R) [17], which is the energy-minimized averaged structure of the 53 NMR models obtained, with WNV-K (WNV strain Kunjin) C protein model number 2 (PDB code 1SFK) [25], which presented the lowest RMSDs (root mean square deviations) of the α-carbons of the amino acids when superimposed with the selected DENV C protein structure and analysed using the Swiss-PdbViewer software package [26]. For superimposition analysis, only amino acids 43–96 of DENV C protein were used, since immediate visual inspection showed that only the regions encompassing α-helices 2, 3 and 4 could superimpose with the available WNV-K C protein structure, with a final RMSD of 2.19 Å (1 Å = 0.1 nm).

**Zeta potential analysis**

Zeta potential (ζ) measurements were performed using a Malvern Zetasizer Nano ZS system, equipped with a helium–neon laser (λ = 632.8 nm). LD interaction with C protein, pep14–23 or pep5–26, in the buffer 20 mM Tris/HCl, pH 7.4, 100 mM KCl, 1 mM EDTA and 1 mM EGTA was analysed at 25 °C (average of 15 measurements, 100 runs each), by PALS (phase analysis light scattering), using disposable zeta cells with platinum gold-coated electrodes. Values of viscosity and refractive index were 0.8872 cP and 1.330 respectively. The electrophoretic mobility obtained was used for the zeta potential calculation through the Smoluchowski equation [27]:

\[
\zeta = \frac{4\pi \eta u}{\varepsilon}
\]

where \( u \) is the electrophoretic mobility, \( \eta \) is the viscosity of the solvent and \( \varepsilon \) is its dielectric constant. For the experiments in which proteolysis of LDs was performed, LD samples were previously incubated with 10 μM trypsin for 15 min at room temperature (23°C), as described previously [18]. The reaction was stopped by the addition of 1 mM PMSF to the mixture. After 5 min of incubation with PMSF, pep14–23 was added and the zeta experiments were carried out.

**AFM-based force spectroscopy measurements**

A NanoWizard II atomic force microscope (JPK Instruments) mounted on an Axiovert 200 inverted optical microscope (Carl Zeiss), equipped with a 15 μm z-range linearized piezoelectric scanner and an infrared laser, was used following standard procedures [28], recently adapted to the probing of DENV C protein–LD interaction [18]. Briefly, 50 μl of LD suspension were allowed to deposit for 30 min at room temperature on a mica surface. Non-adherent LDs were removed by five washes with TEE buffer. Force spectroscopy measurements were performed using OMCL TR-400-type silicon nitride tips (Olympus) functionalized with C protein. Molecular recognition was searched by pressing the DENV C protein-functionalized tips on different points of the LDs in the presence of different pep14–23 concentrations. Force spectroscopy curves were analysed using the JPK image processing program v.3. Each experiment was performed at least three times, using different samples and different functionalized tips, with approximately 5000 force–distance curves collected, analysed and fitted to the WLC (worm-like chain) model [18]. Histograms of the (un)binding forces of each dataset were constructed choosing the ideal bin size to achieve the best fit (6 pN). Force rupture values below 10 pN were considered to represent noise, artefacts or unspecified interactions. From each histogram, the most likely single DENV C protein–LD binding rupture force was determined fitting the distributions of the rupture forces with the Gaussian model. The maximum values of the Gaussian peaks represent a single-molecule-based statistical measure of the strength of the molecular bond.

**RESULTS**

The N-terminal region of the DENV C protein interacts with LDs

The interaction between the DENV C protein and LDs was studied by NMR, taking advantage of the previously determined structure of the protein in solution [17]. The comparison of the \((^{15}N,^1H)\)-HSQC spectra of DENV C protein in the absence and in the presence of LDs revealed specific local interactions that could be mapped by CSPs (chemical shift perturbations) and changes in peak intensity for specific residues, with maintenance of the overall protein structure (Supplementary Figure S1 at http://www.BiochemJ.org/bj/444/bj4440405add.htm, and spectra sections shown in Figures 1a–1d).

Chemical shifts are highly sensitive to small changes in the chemical environment, which may occur as a result of the direct participation of the atom in the interaction, or by an indirect effect of the binding (e.g. conformational changes induced by the molecular recognition) [29]. Analysis of the \((^{15}N,^1H)\)-HSQC spectra revealed that the larger CSPs occurred for residues located in the N-terminal region (Arg5, Lys6, Arg9, Arg19, Arg22, Ser41 and Thr42), in the L1–2 loop (Gly40) and in the α2–α2’ dimer interface (Ala49, Val51, Phe53 and Leu54) (Figure 1e). In contrast, residues located elsewhere were almost not affected by the presence of LDs, as exemplified for residues within helices α1 (Gln28), α3 (Arg68), α4 (Asn79 and Val80) or in the C-terminus (Arg98).

Peak intensity (I) change upon binding is also a good reporter for measuring protein recognition [30]. The intensities ratios for LD-bound to free C protein residues (I_{LD}/I_{0}) are shown in Figure 1(f). For most of the residues, I_{LD}/I_{0} was slightly less than 1, as expected for non-interacting regions of a protein complex. Remarkably, a significant increase in I_{LD}/I_{0} was observed for several residues in the N-terminal region (Arg5, Lys7, Ala8, Arg9, Asn14, Leu16, Arg18 and Arg22), for Leu36, located in the L1–2 loop, and for Ala52 and Phe53, located in the α2–α2′ central hydrophobic patch (Figure 1f). In addition, Ile56, located in the second loop (L2–3), and Glu57, located in α4, were also affected by the interaction. The increase in intensity may be explained by the conformational selection as the mechanism for molecular recognition. This mechanism of interaction presumes that the protein in the free state contains regions in conformational exchange. Residues in conformational exchange show low peak intensities due to the line-broadening caused by the exchange contribution to transverse relaxation rate. Binding stabilizes one conformation, thus leading to an increase in peak intensity. The results showed that a significant increase in I_{LD}/I_{0} was observed for regions that are prone to be in conformational exchange, such as the N-terminal region and the α2–α2′ dimer interface.

Interestingly, the three major C protein regions found to be affected by LD interactions are in the same side of the protein, probably facing the LD upon binding (Figure 1g).
Figure 1  DENV C protein residues involved in the interaction with LDs are in the disordered N-terminal region and in α-helix 2

(a–d) Sections of the HSQC spectra of 15N-labelled DENV C protein in the presence (red) and absence (black) of LDs, highlighting Arg5, Arg18 and Arg22, which belong to the disordered N-terminal region; Gly40, located in the L1–2 loop; Phe53 and Leu54, located in α2; and Gly28 of α1, Arg68 of α3, Val79 and Arg98 of α4. (e) CSP and (f) peak intensity ratios (I_{b} / I_{0}) of bound/free protein as a function of protein residue. Bars are coloured to indicate the residues showing CSP values higher than 0.05 in (e), and I_{b} / I_{0} values higher than 3 in (f): green, residues within the N-terminal region for which no structure is available; dark blue, residues within the disordered N-terminal region for which a structure is available; cyan, the residue in the L1–2 loop; and orange, residues in α2. (g) Ribbon structure of DENV C protein (PDB code 1R6R) in two alternative views, with the residues that showed significant CSP highlighted with the same colour code used in (e) and (f). Note that the segment containing residues 1–20 is missing, since its structure could not be determined by NMR [17]. Residues Met1, Asn2, Thr11, Leu35 and Arg55 were not assigned in the previously determined NMR structure [17]. The assignments of Asp3, Gin4, Asn10, Gly30, Gin37, Thr62, Ala79 and Arg100 could not be localized or confirmed. Pro22, Pro31, Pro40 and Pro51 are not represented in the graphs.
DENV C protein displays intrinsically disordered behaviour for its N-terminal region

The central hydrophobic patch at α-helix 2 has previously been assigned to participate in C protein interaction with membranes [17,31] and LDs [12]. In the present study we show that loop L1–2 and the previously unstudied N-terminal region also play a role in C protein binding to LDs. Thus, to get information regarding the conformational flexibility of these regions, two parameters for each residue in the protein sequence were analysed and compared: the peak intensities in the [(15)N,1H]-HSQC of the free protein (Figure 2a) and the prediction of the order parameter (S2) of the N–H bond based on the random coil index obtained from the deposited chemical shift values [32] (Figure 2b). Most of the residues in the N-terminal region presented lower intensities and lower S2 values when compared with the residues in the globular region. Additionally, within the N-terminal region, values of peak intensities inversely correlate with S2, as clearly observed for residues Ala8, Met15, and Glu19, which display the highest intensities and the lowest S2 values within this region. This behaviour is expected, since regions of low S2, and probably high thermal flexibility, generally show sharper resonance lines. A remarkable observation was the presence of high order (S2 ~0.8) for Pro12 and Phe13 within this region, which also display low peak intensities. This segment is a hydrophobic triad (Thr11, Pro12, Phe13) in which Thr11 was not assigned in the previous determined structure [17]. For Asn10, only 13C′ and the amidoic 1H and 15N could be found, and Pro12 was found only in the trans-conformation. The lack of those assignments could be a consequence of conformational exchange in the milli- to microsecond timescale, leading to line-broadening and, consequently, to the decrease in peak intensity. Conformational exchange results from an equilibrium among two or more discrete conformational states, indicating some degree of conformational order in the flexible N-terminal region, possibly due to the hydrophobic character of the triad and the presence of a proline. This is a typical behaviour of intrinsically disordered proteins, which display flexibility, but yet some degree of order [1,2]. Taken together, these observations agree with the proposed conformational selection mechanism for C protein binding to LDs suggested by the increased peak intensity ratio (Figure 1f).

Both S2 values and peak intensities agreed on the presence of higher ordering in the C protein globular region, which extends from Val21 to the C-terminus. All secondary structure elements (residues contained in α1–α4) presented an S2 value greater than 0.8. On the other hand, a decrease in the order parameter on the L1–2 loop could be observed, indicating thermal motion within this long loop that extends from Arg32 to Pro41. The most-reduced order parameters were observed for the segments extending from Arg32 to Leu35 (RFSL) and from Gln39 to Pro43 (QGRGP) (Figure 2b). The reduction in S2 was followed by a significant increase in peak intensity for residues Ser44 and Arg45 (Figure 2a). TALOS+ prediction of secondary structure [33] suggests the presence of one turn of a helix within these segments, comprising residues Gly36, Met37 and Leu38 (Figure 2c). Analysis of the DENV C protein structure in solution (PDB code 1R6R) corroborates the presence of the one-turn helix involving these residues. Indeed, for this segment, an S2 value greater than 0.8 was found (Figure 2b). Additionally, reduced order parameters, but not as pronounced as for the L1–2 loop, were observed for the L2–3 and L3–4 loops (Figure 2b). It is important to mention that the assignments of Leu35 and Arg35 were missing in the previously determined C protein structure [17], possibly due to conformational exchange, which is interesting since Arg35 is within the central hydrophobic patch that is also involved in the binding to LDs.

A peptide corresponding to the conserved N-terminal motif of C protein binds to LDs

To further evaluate the structural features involved in the interaction between the disordered N-terminal region and LDs, two peptides were designed, corresponding to residues 5–26 (pep5–26), and to residues 14–23 (pep14–23). In both peptides, Glu19 was replaced with an alanine residue, in order to make their...
sequences more similar to that of C proteins of all other DENV strains, as well as of all other flaviviruses, in which a neutral amino acid is present at that position. pep5–26 includes all N-terminal residues affected by the interaction with LDs, starting at residue 5 and ending just before the beginning of $\alpha_1$, at residue 26. pep14–23 is a shorter version of pep5–26 that excludes the N-terminal residues and the hydrophobic segment that showed a high order parameter. pep14–23 includes the conserved motif NML + R, beginning in the first residue of the conserved motif (Asn$^{14}$) and extending to the last residue affected by LD interaction, just before the beginning of $\alpha_1$.

To quantify peptide binding to LDs, zeta potential measurements were performed. This parameter is dependent on the charge of a given scattering particle in solution [27]. Upon the addition of DENV C protein to an LD suspension, there was a concentration-dependent increase in the zeta potential values, from the initial negative value characteristic of LDs (approximately $-20$ mV) to a limit positive value, as expected for the binding of the positively charged C protein (Figure 4). pep14–23 (net charge +4) was also able to bind to LDs, leading to an increase in the zeta potential with a magnitude similar to that observed when LDs interact with DENV C protein. In

Figure 3  C protein residues involved in the LD interaction are conserved among Flavivirus spp.

(a) Alignment of the DENV C protein sequence with C protein sequences from other flaviviruses, highlighting fully conserved segments (red) and partially conserved residues (black). The conserved segments and the DENV C protein sequence are depicted, with the residues shown to interact with LDs underlined and coloured according to the colour code used in Figure 1(b). Both the N-terminal NML + R motif and the $\alpha_2$ HAPL/F motif are conserved and contain the residues indicated by the NMR studies to interact with LDs. (b) Comparison of C protein structures of DENV (grey) and WNV-K (red). Two regions can be depicted: a conserved fold (from residues 44 to the C-terminus), and a particularly flexible region with alternative folds, containing $\alpha_1$ and the disordered section (residues 1–26). DENV C protein residues that showed a significant CSP are highlighted in blue, cyan and orange (following the colour code used in Figure 1b).
contrast, pep5–26 (net charge +8) did not change the measured zeta potential values, indicating an absence of significant binding to LDs. In a previous study, we found that C protein binds to a protein component in the surface of LDs [18]. To evaluate whether binding of pep14–23 to LDs was also dependent on a surface protein, LD limited proteolysis was performed before zeta potential measurements. As previously observed for C protein, LD trypsinization eliminated the ability of pep14–23 to interact with LDs (Figure 4).

NML+ is the LD-binding motif of pep14–23

To map the interaction of pep14–23 with LDs, free pep14–23 was firstly assigned using a sequential assignment strategy [34], with a combination of TOCSY and NOESY spectra. A (1H,13C)-HSQC spectrum was also used to solve ambiguities and to assign the 13C resonances. It was possible to unambiguously and fully assign Met2, Leu3, Lys4, Ala6 and Val10. The amide resonances of Met2 and the N-terminal Asn1 were not found in the spectrum, probably due to line broadening. Both Asn4 and Asn8 were assigned, although they could not be unambiguously differentiated. All arginine residues were superimposed in the spectrum, showing degenerated chemical shift for the amide and side-chain resonances. The only exception was for the (1H,13C)-chemical shift, for which two sets of peaks could be differentiated. The 80 % assignment of the (1H,13C)-HSQC spectrum enabled the measurement of the CSP for each residue after addition of LDs to the peptide sample. Met2 and Leu3 Hα/Cα cross-peak showed the largest CSP, along with one of the asparagine residue, which has been attributed to Asn1 due to the proximity in the sequence to Met2 and Leu3 (Figure 5). It should be noted that Hα/Cα cross-peaks for the arginine residue did not change with the addition of LDs. Smaller changes in CSP for Lys4, Ala6 and Val10 were also observed. It is important to remark that all chemical shift changes were double-checked in the HSQC and TOCSY spectra. Figure 5(b) shows a significant perturbation in the Hβ/Cβ cross-peak for the same asparagine residue, attributed as Asn1. CSP was not observed for any other side-chain resonance. The Hα chemical shift changes promoted by the presence of LDs were quantified, showing that the highest changes occurred for the resonances of Asn1, Met2 and Leu3 (Figure 5c). The chemical shift of the basic arginine residues did not change significantly, showing that the positive charges of the N-terminal residues and Lys4 were dominant for the interaction. To complement the findings, the interaction of pep5–26 with LDs was evaluated employing the same methodology. As previously observed using zeta potential, no interaction of this peptide with LDs was detected by NMR (Supplementary Figure S2 at http://www.BiochemJ.org/bj/444/bj4440405add.htm).

pep14–23 inhibits the binding of DENV C protein to LDs

Since pep14–23 was shown to specifically interact with LDs, a relevant question is whether the peptide is able to compete with C protein for its LD-binding site. In order to tackle this issue, the effects of pep14–23 on the strength of the C protein–LD interaction was evaluated by AFM-based force spectroscopy, following procedures established previously [18,28]. Tapping on LDs with DENV C protein-derivatized AFM tips allowed measurement of the frequency of C protein–LD-binding events (by the subsequent unbinding), as well as measurement of the force necessary to break the bond between a single C protein dimer and an LD. The (un)binding frequencies and forces for C protein–LD interaction were determined in the absence and in the presence of different concentrations of pep14–23 (Figures 6a–6f). It was found that C protein binding to LD was inhibited by the addition of the peptide, in a concentration-dependent manner (Figure 6f). The force necessary to break the
binding decreases from 33 pN, in the absence of the peptide, to 19 pN in the presence of 100 μM pep14–23. Most striking was the abrupt decrease in the number of (un)binding events, from 54% in the absence of peptide to only 23% at the highest peptide concentration tested.

**DISCUSSION**

DENV C protein is known to interact with LDs and this association was shown to be mandatory for virus particle formation [12]. Taking into account the biological relevance of that interaction, we investigated the determinant structural features of C protein for the binding to LDs. Local structural changes involving specific amino acid residues could be mapped in the disordered N-terminal region, in the central hydrophobic patch and in the L1–2 loop.

The finding that four consecutive residues in helix α2 (residues 51–54) are affected by LD interaction agrees with the suggestion that the interaction with membranes would occur through the central hydrophobic patch formed by the α2–α2′ dimer helical interface [17,31]. Additionally, mutational analysis showing that Leu50 and Leu54 are essential for targeting C protein to LDs [12] also supports the NMR data of the present study. Finally, it is clear that the residues within the hydrophobic α2–α2′ core involved in LD interaction are conserved among flaviviruses, both in terms of sequence (are part of an hAFL/F conserved sequence) and structural alignment, reinforcing their importance for C protein functions.

Furthermore, our findings expanded the previous understanding of DENV C protein interaction with LDs by pointing to a role of specific residues in the disordered N-terminal region and in the L1–2 loop that have been overlooked so far. Our hypothesis is that loop residues do not participate directly in the interaction, but provide structural flexibility, as observed for other viral proteins [35,36]. Most intriguing is the involvement of the N-terminal residues located in the intrinsically disordered region, which had not been studied before. The intrinsically disordered nature of this region points to its involvement in molecular recognition events, such as the novel function of LD binding described in the present paper. Intrinsically disordered proteins or regions undergo equilibrium among several conformational states [1,2]. Thus, the pre-existent conformational states allow recognition of different targets. Binding to each of the targets, in turn, leads to conformational selection [37–39]. Indeed, conformational selection as a recognition mechanism in C protein–LD interaction is supported by the increase in peak intensity ratios. Conformational selection may offer an evolutionary advantage for multifunctional proteins, such as the C proteins of flaviviruses. In fact, despite their low sequence similarity, these proteins share similar activities attributed to their intrinsically disordered regions, such as RNA binding [40,41] or RNA chaperoning [8]. In the case of DENV C protein, a nuclear localization signal has already been identified in the disordered N-terminal region, but mutations in this region had partial effects on nuclear migration [42]. Interestingly, antibodies produced against DENV C protein presented specific reactivity to the
disordered N-terminal region, suggesting that it corresponds to the predominant immunogenic segment [43]. The high immunoreactivity of C protein N-terminal region may be related to the intrinsically disordered nature of this segment, enabling its interaction with multiple targets, including LDs, as shown in the present study.

As the interaction with LD did not significantly change the globular core of C protein structure, we hypothesized that LD binding may alter the relative orientation of the N-terminal region and α1 with respect to the globular core. This hypothesis was supported by the dynamics of the C protein in the absence of LDs, which clearly identified flexible segments corresponding to the N-terminal region and the L1–2 loop, contrasting with a rigid core, extending from α2 to the C-terminus. The comparison of the C protein structures of DENV [17] and WNV-K [25] also revealed that the segment from the N-terminus to residue 43 behaves as a plastic region, with alternate folds, corroborating the analysis of DENV C protein dynamics and reinforcing our hypothesis. Furthermore, the fact that the same N-terminal residues that interact with LDs are contained in a conserved flavivirus C protein segment (NML + R, residues 14–18) highlights the importance of these residues for the C protein functions. Interestingly, NS5A from HCV (hepatitis C virus, also a member of the Flaviviridae), a protein that is found associated with surface of LDs in infected cells [44], contains a relatively similar sequence (Y209MLP KK R210), further supporting the role of the NML + R sequence as a LD-binding motif.

Taking into account the findings described above, we hypothesized that the disordered region, rich in positively charged residues, containing the NML + motif, might prompt an initial interaction with the negatively charged LDs, after which a conformational rearrangement facilitated by the conformational freedom provided by the flexible L1–2 loop enables the access of LDs to the later specific interactions with the central α2–α2′ hydrophobic dimer interface. Indeed, the increase in peak intensity observed for α2-helix residues Ala35 and Phe39 corroborates the stabilization of binding through this region. The results of the present study suggest that the α2–α2′ interface is in conformation exchange in the free state, possibly interconverting between an open and a closed state. Indeed, Arg33 is missing in the previous assignment of DENV C protein structure [17], and residues in helices α1 and α2 showed consistently lower peak intensities than those of α3 and α4, possibly indicating line-broadening due to conformational exchange. This would facilitate the DENV C protein interaction with LDs and may have special importance in the case of WNV-K, in which deletions in large parts of its inner core lead to still-viable viruses [45,46]. In those studies, viruses containing large deletions in the genome that resulted in the complete removal of C protein helices α2 and α3 generated viable viral progeny. In those shortened C protein versions, the disordered N-terminus, helix α1, part of the L1–2 loop and part of helix α4 remained intact and seemed to be sufficient for viral replication to proceed.

Following this line of reasoning, it can be speculated that peptides mimicking the interacting region of the C protein, especially if the NML + R motif is included, might be targeted to the LD–C protein binding site. Thus we rationally designed two peptides based on this region, and tested their ability to interact with LDs by zeta potential measurements. The advantage of employing zeta potential analysis is that it allows monitoring the changes in the surface charge density of particles, such as LDs, as a result of an interaction with small ligands, without any further modification of the ligand [27]. The results showed that pep14–23 is able to interact with LDs to a similar extent as the full protein. The extensive zeta potential variation (approximately 34 mV), changing from negative to positive values, fits well with a model of interaction in which the positively charged α4 is exposed, conferring a positive charge to the outer surface of the C protein–LD complex, and supporting the previously described conclusions based on NMR data [17]. Additionally, the interaction between pep14–23 and LDs was dependent on a protein component in the surface of LD, exactly as previously found for C protein–LD interaction [18]. The detailed analysis of the Hr chemical shift changes in the (1H,13C)-HSQC spectra of pep14–23 in the presence and in the absence of LDs revealed that the higher changes occurred exactly for the conserved NML + motif, fitting adequately with the NMR analysis of the full-length protein. As pep14–23 is rather small and highly basic, with the five positive charges distributed along positions 1, 4, 5, 7 and 9, one can hypothesize that electrostatic interactions contribute to a first stage of the peptide binding to the negatively charged surface of the LDs. Then, the NML + R residues may become more tightly bound, due to the contribution of hydrophobic interactions, which may also occur in the full-length protein and, in this way, anchor the C protein to the LDs, allowing further interaction of the hydrophobic α2–α2′ core. On the other hand, both zeta potential and NMR data showed that pep5–26 does not interact with LDs. This result is similar to those found for other peptide–ligand interactions, in which shorter sequences were more effective in eliciting a response than longer versions [47]. The larger size of this peptide may render the NML + R sequence inaccessible to LD binding. The inaccessibility of the NML + R sequence in pep5–26 may result from presence of the hydrophobic triad (Thr4–Pro5–Phe6) that displays high order parameter, which may impose some conformational constraint. Remarkably, pep14–23 comprises only the low order parameter segment. This binding behaviour is another evidence of the intrinsically disordered nature of the C protein N-terminal region.

Taken together, the results of the present study point to the ability of using pep14–23 to inhibit C protein binding to LDs. Indeed, the inhibitor potential of the peptide was demonstrated by single-molecule AFM-based force spectroscopy data. The peptide blocked C protein binding to LDs in a concentration-dependent manner, with a marked decrease in the probability of the occurrence of the binding to one fifth of the initial value at the highest peptide concentration tested. It is important to bear in mind that the histograms (obtained by thousands of measurements for each peptide concentration) clearly show that even the small fraction of binding events occurring at high pep14–23 concentrations corresponds mostly to low (un)binding forces, typical of unspecific interactions. To conclude, we propose that sequences encompassing or within pep14–23 actively contribute to the C protein–LD interaction. Moreover, therapeutic approaches using peptides and peptide analogues were effective against other viruses, such as HIV [48,49] and HCV [50], a strategy that may be followed also for DENV and similar flaviviruses.

In summary, the results of the present study revealed that the disordered N-terminal region of DENV C protein is crucial for its interaction with LDs, and that a peptide designed on the basis of a conserved segment within this region was able to inhibit this interaction, which is a key event of the replication of DENV (and possibly of other flaviviruses). Therefore the novel advances reported in the present paper pave the way to drug development approaches, in which peptide inhibitory efficiency may be further improved through lead optimization strategies.
AUTHOR CONTRIBUTION
Ivo Martins, Fabio Almeida, Nuno Santos and Andrea Da Poian conceived and designed the experiments. Ivo Martins, André Faustino, Filomena Carvalho and Fabiana Carmeno performed the experiments. Ivo Martins, Francisco Gomes-Neto, André Faustino, Filomena Carvalho, Fabio Almeida, Nuno Santos and Andrea Da Poian analysed the data. Ronaldo Mohana-Borges, Patrícia Bozza, Fabio Almeida, Nuno Santos and Andrea Da Poian contributed reagents/materials/analysis tools. Ivo Martins, Miguel Castanho, Fabio Almeida, Nuno Santos and Andrea Da Poian wrote the paper.

ACKNOWLEDGEMENTS
We thank Teresa Freitas (IMM, FUMUL, Lisbon, Portugal) for technical assistance and Marco Domingues (IMM) for help with the zeta potential measurements.

FUNDING
This work was supported by the European Union 7th Framework Programme PEOPLE IRSES (International Research Staff Exchange Scheme) project MEMPEACROSS, the National Institute of Science and Technology in Dengue (INCTD, Brazil), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ, Brazil), the Fundação para a Ciência e a Tecnologia (FCT)–Ministério da Educação e Ciência (MEC) (Portugal) [grant number PTDC/QUI-BIO/112929/2009], the Fundação Calouste Gulbenkian (Portugal), and FCT–Fundacão Coordenacão de Aperfeicãoamento do Pessoal de Nível Superior (CAPES) Portugal–Brazil joint co-operation projects. I.C.M. is the recipient of consecutive postdoctoral funding from a Marie Curie International Outgoing Fellowship [grant number MIC-IOF237373] and FCT postdoctoral fellowships [grant numbers SFRH/BDP/46324/2008 and SFRH/BDP/74287/2010].

REFERENCES

© The Authors. Journal compilation © 2012 Biochemical Society
Dengue virus capsid protein binding to lipid droplets


Received 22 December 2011/6 March 2012; accepted 19 March 2012
Published as BJ Immediate Publication 19 March 2012, doi:10.1042/BJ20112219
SUPPLEMENTARY ONLINE DATA

The disordered N-terminal region of dengue virus capsid protein contains a lipid-droplet-binding motif

Ivo C. MARTINS*†, Francisco GOMES-NETO*‡, André F. FAUSTINO†, Filomena A. CARVALHO†, Fabiana A. CARNEIRO§, Patricia T. BOZZA∥, Ronaldo MOHANA-BORGES¶, Miguel A. R. B. CASTANHO⊥, Fábio C. L. ALMEIDA*‡, Nuno C. SANTOS† and Andrea T. DA POIAN*1

*Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil, †Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisbon, 1649-028, Portugal, ‡Centro Nacional de Ressonância Magnética Nuclear, Universidade Federal do Rio de Janeiro and National Institute of Structural Biology and Bioimage, Rio de Janeiro, RJ, 21941-902, Brazil, §Polo Avançado de Xerém, Universidade Federal do Rio de Janeiro, Duque de Caxias, RJ, 25245-360, Brazil, ¶Laboratório de Genomica Estrutural, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil

Figure S1 (13C,1H)-HSQC spectra of DENV C protein in the presence (red) and absence (black) of LDs, showing that the overall protein structure is maintained

Figure S2 (13C,1H)-HSQC spectra of pep5–26 in the presence (red) and absence (black) of LDs

(a) Cα region. (b) Cβ region. (c) Methyl region of the spectra. Hα/Cα paired resonances in three different regions of the (13C,1H)-HSQC spectra do not show significant chemical shift changes, reinforcing the previous conclusion of the absence of pep5–26–LD binding.

1 To whom correspondence should be addressed (email dapoian@bioqmed.ufrj.br).
Table S1  Virus strains and sequences used to construct the alignment with different Flavivirus spp. capsid proteins in Figure 3(a) in the main paper

N.A., not applicable (query sequence).

<table>
<thead>
<tr>
<th>Virus name</th>
<th>Abbreviation</th>
<th>Strain</th>
<th>NCBI accession number</th>
<th>Total score</th>
<th>Query coverage</th>
<th>Taxonomy (Group, species)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dengue virus serotype 2</td>
<td>DENV-2</td>
<td>PR-159S1</td>
<td>1R6R_A</td>
<td>N.A.</td>
<td>N.A.</td>
<td>Dengue virus group, Dengue virus</td>
</tr>
<tr>
<td>West Nile virus</td>
<td>WNV</td>
<td>K6453</td>
<td>ADM8862.1</td>
<td>76.6</td>
<td>N.A.</td>
<td>Japanese enc. virus group, West Nile virus</td>
</tr>
<tr>
<td>WNV serotype Kunjin</td>
<td>WNV-K</td>
<td>MRM62C</td>
<td>AAP78941.1</td>
<td>75.5</td>
<td>91</td>
<td>Japanese enc. virus group, West Nile virus</td>
</tr>
<tr>
<td>Japanese encephalitis virus</td>
<td>JEV</td>
<td>YL</td>
<td>AAL33186.1</td>
<td>75.5</td>
<td>93</td>
<td>Japanese enc. virus group, Japanese enc. virus</td>
</tr>
<tr>
<td>St. Louis encephalitis virus</td>
<td>SLEV</td>
<td>GML</td>
<td>903797</td>
<td>79.3</td>
<td>93</td>
<td>Japanese enc. virus group, St. Louis enc. virus</td>
</tr>
<tr>
<td>Murray Valley encephalitis virus</td>
<td>MVEV</td>
<td>MVE-1-51</td>
<td>NP_051124.1</td>
<td>73.2</td>
<td>83</td>
<td>Japanese enc. virus group, Murray Valley enc. virus</td>
</tr>
<tr>
<td>Kokobera virus</td>
<td>KOKV</td>
<td>AusMRM 32</td>
<td>GP_001040007</td>
<td>85.9</td>
<td>91</td>
<td>Kokobera virus group, Kokobera virus</td>
</tr>
<tr>
<td>Kedougou virus</td>
<td>KEDV</td>
<td>DakAR D1470</td>
<td>ABI54477.1</td>
<td>85.5</td>
<td>96</td>
<td>Dengue virus group, Kedougou virus</td>
</tr>
<tr>
<td>Zika virus</td>
<td>ZIKV</td>
<td>MR766</td>
<td>ABI54475.1</td>
<td>84.0</td>
<td>98</td>
<td>Spondweni virus group, Zika virus</td>
</tr>
<tr>
<td>Spondweni virus</td>
<td>SPDV</td>
<td>SM-6 V1</td>
<td>AB154480.1</td>
<td>87.8</td>
<td>99</td>
<td>Spondweni virus group, Zika virus</td>
</tr>
<tr>
<td>Aroa virus</td>
<td>AROAV</td>
<td>BEAN 4673</td>
<td>GP_001040004</td>
<td>83.6</td>
<td>95</td>
<td>Aroa virus group, Aroa virus</td>
</tr>
<tr>
<td>Iguape virus</td>
<td>IGV</td>
<td>SPAe 71666</td>
<td>AV24154.1</td>
<td>80.1</td>
<td>87</td>
<td>Aroa virus group, Aroa virus</td>
</tr>
<tr>
<td>Usutu virus</td>
<td>USUV</td>
<td>SAAR-1776</td>
<td>AA55401.1</td>
<td>79.7</td>
<td>91</td>
<td>Japanese enc. virus group, Usutu virus</td>
</tr>
<tr>
<td>Alfuy virus</td>
<td>ALFV</td>
<td>MRM3929</td>
<td>AA42481.1</td>
<td>77.0</td>
<td>93</td>
<td>Japanese enc. virus group, Murray Valley enc. virus</td>
</tr>
<tr>
<td>Bagaza virus</td>
<td>BAGV</td>
<td>DakAR B209</td>
<td>GP_002790883</td>
<td>76.6</td>
<td>85</td>
<td>Ntaya virus group, Bagaza virus</td>
</tr>
<tr>
<td>Ilheus virus</td>
<td>ILHV</td>
<td>Original</td>
<td>GP_001040006</td>
<td>65.9</td>
<td>94</td>
<td>Ntaya virus group, Ilheus virus</td>
</tr>
<tr>
<td>Rocio virus</td>
<td>ROCV</td>
<td>SPM 34675</td>
<td>AA434158.1</td>
<td>64.7</td>
<td>86</td>
<td>Ntaya virus group, Ilheus virus</td>
</tr>
</tbody>
</table>