

# SCIENTIFIC REPORTS



OPEN

## Fast NMR method to probe solvent accessibility and disordered regions in proteins

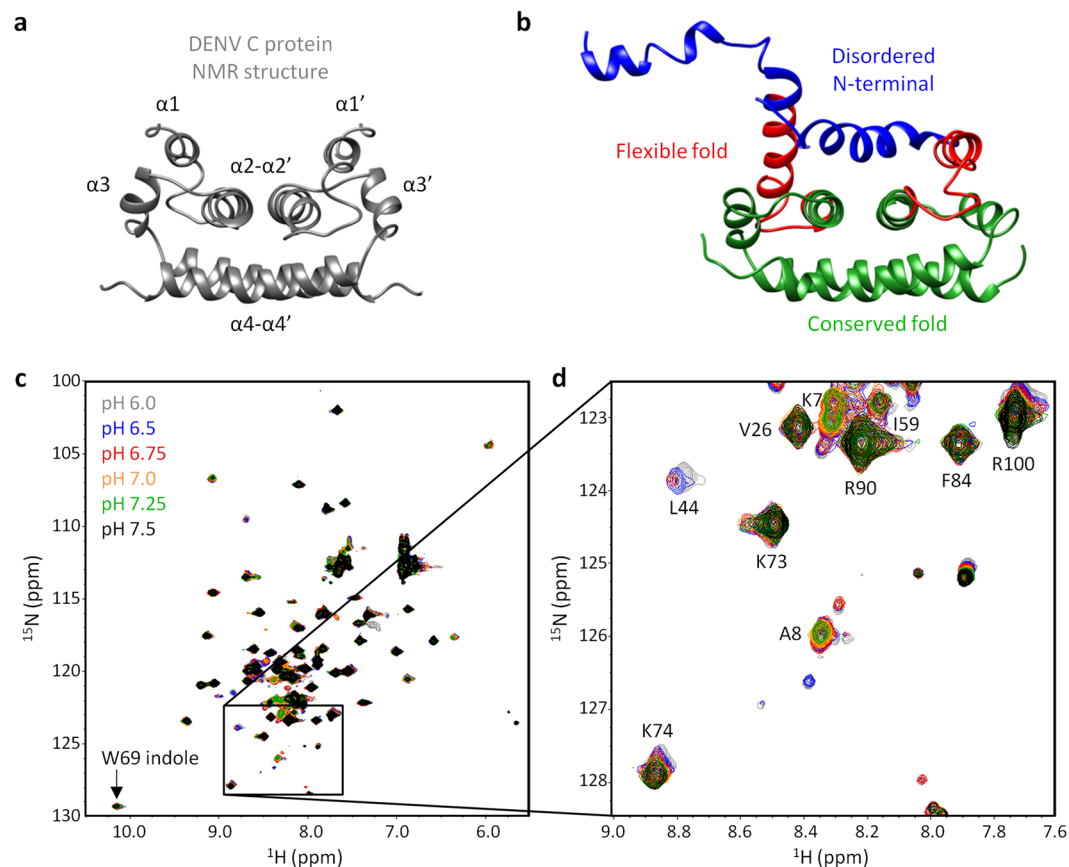
André F. Faustino<sup>1,5</sup>, Glauce M. Barbosa<sup>2</sup>, Micael Silva<sup>4</sup>, Miguel A. R. B. Castanho<sup>1</sup>, Andrea T. Da Poian<sup>2</sup>, Eurico J. Cabrita<sup>4</sup>, Nuno C. Santos<sup>1</sup>, Fabio C. L. Almeida<sup>2,3</sup> & Ivo C. Martins<sup>1</sup>

Understanding protein structure and dynamics, which govern key cellular processes, is crucial for basic and applied research. Intrinsically disordered protein (IDP) regions display multifunctionality via alternative transient conformations, being key players in disease mechanisms. IDP regions are abundant, namely in small viruses, allowing a large number of functions out of a small proteome. The relation between protein function and structure is thus now seen from a different perspective: as IDP regions enable transient structural arrangements, each conformer can play different roles within the cell. However, as IDP regions are hard and time-consuming to study via classical techniques (optimized for globular proteins with unique conformations), new methods are required. Here, employing the dengue virus (DENV) capsid (C) protein and the immunoglobulin-binding domain of streptococcal protein G, we describe a straightforward NMR method to differentiate the solvent accessibility of single amino acid N-H groups in structured and IDP regions. We also gain insights into DENV C flexible fold region biological activity. The method, based on minimal pH changes, uses the well-established <sup>1</sup>H-<sup>15</sup>N HSQC pulse sequence and is easily implementable in current protein NMR routines. The data generated are simple to interpret, with this rapid approach being an useful first-choice IDPs characterization method.

Nuclear magnetic resonance (NMR) spectroscopy is the technique of excellence to obtain structural and dynamics atomic resolution information of macromolecules, especially proteins<sup>1</sup>. NMR is compatible with room temperature solution measurements, a major advantage over other high-resolution structural techniques (X-ray crystallography and cryo-electron microscopy). Via NMR, protein structures can now be determined with chemical shifts data alone (employing the CS-Rosetta package)<sup>2-6</sup>, which is extremely important when merely sparse data is available<sup>3,4</sup>. Noteworthy, NMR provides protein dynamics information in physiological conditions, via the probing of different macromolecular motion timescales<sup>7-9</sup>. Backbone amide hydrogen exchange experiments are particularly informative of protein N-H solvent accessibility, being related to both structure and dynamics (sensitive to the millisecond timescale)<sup>9-13</sup>. Nevertheless, complementary data is frequently still necessary<sup>14-18</sup>, especially to study intrinsically disordered protein (IDP) regions<sup>17,19</sup>. Therefore, there is a major unmet demand for fast straightforward methodologies to analyze IDP regions, which play key roles in health and disease mechanisms.

Here, based on the amide hydrogen exchange process<sup>10-12,20-24</sup>, we report a simple and fast method to gain information on solvent accessibility of each amino acid residue N-H group of a protein. Benefiting from our previous work<sup>25-30</sup>, we used dengue virus (DENV) capsid (C) protein as a model (Fig. 1a, PDB ID 1R6R)<sup>26,31,32</sup> since it possesses three distinct structural regions (Fig. 1b and Table S1): the disordered N-terminal, the flexible fold and the conserved fold. The nomenclature “conserved fold” refers to a structurally persistent fold that was

<sup>1</sup>Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028, Lisboa, Portugal. <sup>2</sup>Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-902, RJ, Brazil. <sup>3</sup>Centro Nacional de Ressonância Magnética Nuclear, Universidade Federal do Rio de Janeiro and National Institute of Structural Biology and Bioimage (CENABIO), Rio de Janeiro, 21941-902, RJ, Brazil. <sup>4</sup>UCIBIO, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516, Monte de Caparica, Portugal. <sup>5</sup>Present address: iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2780-901, Oeiras, Portugal. Correspondence and requests for materials should be addressed to N.C.S. (email: [nsantos@fm.ul.pt](mailto:nsantos@fm.ul.pt)) or F.C.L.A. (email: [falmeida@bioqmed.ufrj.br](mailto:falmeida@bioqmed.ufrj.br)) or I.C.M. (email: [ivomartins@medicina.ulisboa.pt](mailto:ivomartins@medicina.ulisboa.pt))



**Figure 1.** DENV C structure, overall dynamics and  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra from pH 6.0 to 7.5. **(a)** DENV C homodimer experimental structure. This protein is positively charged, with 26 cationic and 2 anionic out of 100 residues *per* monomer. From amino acid residue 21 to 100, it contains four  $\alpha$ -helices named  $\alpha 1$  to  $\alpha 4$  (PDB ID 1R6R<sup>32</sup>). The first 20 residues are not shown since they are intrinsically disordered in solution<sup>32</sup>. **(b)** Molecular dynamics simulation structure of DENV C<sup>29</sup>, which highlights the three main structure/dynamics regions: disordered N-terminal (blue, residues 1–22)<sup>31,32</sup>; flexible fold (red, residues 23–44)<sup>26</sup>; and conserved fold (green, residues 45–100)<sup>26,32–34</sup>. **(c)** Superimposed DENV C  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra at pH 6.0 (gray), 6.5 (blue), 6.75 (red), 7.0 (yellow), 7.25 (green) and 7.5 (black). Our approach requires only  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities data. **(d)** Zoom on a spectral region where all types of response to the pH variation are observed: peaks from L44, I59 and K74 vary in intensity and chemical shift; K7, A8 and K73 just vary in intensity; R100 just varies in chemical shift; and, V26, F84 and R90 neither vary in intensity nor in chemical shift. UCSF Chimera v1.9 software<sup>54</sup> was used for protein structure visualization. The data agrees with the current understanding of flaviviruses C proteins structure/dynamics and biological activity<sup>25–34,55</sup>.

found in the flaviviruses capsid protein structures of Dengue, West Nile and, recently, Zika viruses (with PDB IDs, respectively, 1R6R, 1SFK and 5YGH)<sup>26,32–34</sup>. This approach employs a small pH variation, preserving DENV C overall structure and dynamics, which easily allows probing the backbone N-H groups' solvent accessibility (using only  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities). The three DENV C structure/dynamics regions can be clearly distinguished via this method, which supports not only the current understanding of DENV C structure/dynamics properties, but also the use of this technique to access and differentiate solvent-exposed N-H groups in IDP regions. Moreover, since the solvent accessibility of each N-H group is related to its intramolecular H-bond pattern, our approach also informs on the secondary structure content. Importantly, this methodology is readily applicable to study other proteins structure and dynamics. To demonstrate this, we also tested the effect of varying the pH, in a similar 1.5 pH units interval, on the NMR  $^1\text{H}$ - $^{15}\text{N}$  HSQC intensities using another model system, the B1 immunoglobulin binding domain of streptococcal protein G (GB1)<sup>35–38</sup>. GB1 is particularly interesting as it possesses  $\alpha$ -helical and  $\beta$ -sheet regions, complementing the approach. As described ahead, the method is readily applicable to this protein as well, supporting its use.

## Results

**Suitability of the protein to the pH-based variation approach used.** Our approach is based on small variations of pH within an interval that does not lead to major protein conformational changes. DENV C (Fig. 1a,b) was used as a model to study the relationship between the protein structure/dynamics and backbone N-H solvent accessibility (*i.e.*, the ability of the N-H hydrogen to exchange with water hydrogens). The interval

used is between pH 6.0 and 7.5, a physiological range that is suitable for most proteins, including DENV C, as described ahead. Taking advantage of the pH dependent amide N-H hydrogen exchange process<sup>10–12,20–24</sup>, <sup>1</sup>H-<sup>15</sup>N HSQC spectra of DENV C were acquired at pH 6.0 and 7.5 (Fig. 1c,d, gray and black, respectively). Only specific peaks show decreased intensity and/or a chemical shift variation (Fig. 1c,d). The variations in intensity affect more peaks and are more pronounced than chemical shift variations, suggesting no major conformational change triggered by pH. To ascertain that, we acquired <sup>1</sup>H-<sup>15</sup>N HSQC spectra of DENV C at several pH values (6.0, 6.5, 6.75, 7.0, 7.25 and 7.5), assessing the spectral evolution as a function of pH (Fig. 1c,d). The spectral region represented in Fig. 1d shows the four amino acid peaks that present the most pronounced variation of chemical shift, namely L44, K73, K74 and R100. Even for these residues, the changes are minimal, implying the conservation of DENV C overall architecture.

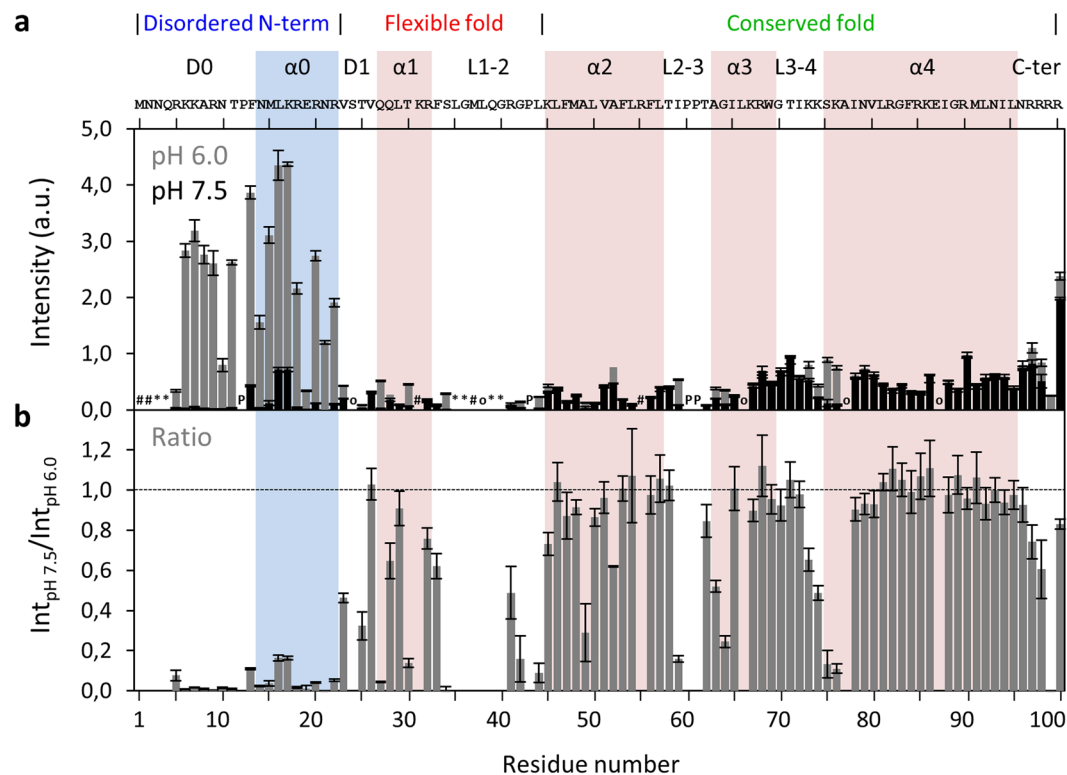
We then consulted the pK<sub>a</sub> values of titratable residues (Fig. S1)<sup>39</sup>, since acid-base equilibrium could cause conformational changes that would difficult the interpretation of the results. DENV C theoretical isoelectric point is at pH 12.6 and, importantly, its sequence does not contain amino acid residues titratable within the pH range studied (Fig. S1, orange bar). We also measured the NMR transverse <sup>15</sup>N amide relaxation rates (R<sub>2</sub>) in both pH conditions (Fig. S2), which demonstrated that there are no conformational transitions triggered by pH. This parameter is sensitive to alterations on the size/shape of a protein (since, in globular proteins, it generally increases with the protein hydrodynamic diameter), as well as to local fluctuations in the flexibility of particular amino acid residues<sup>7,9,40–42</sup>. It is clear from Fig. S2 that the R<sub>2</sub> values obtained for DENV C are overall invariant in this pH range. Therefore, taking all of the above into account (namely Figs. 1c,d, S1 and S2), the overall DENV C structural arrangement is maintained. Thus, the pH-induced <sup>1</sup>H-<sup>15</sup>N HSQC spectral differences are solely due to amide hydrogen exchange with water<sup>10–12,20–24</sup>, which reports on solvent accessibility.

**Probing solvent accessibility to the protein backbone.** The spectral changes observed are consistent with an amide hydrogen exchange process (*i.e.*, where N-H groups exchange their hydrogens with water hydrogens)<sup>10–12,20–24</sup>. Such process only occurs if N-H groups are exposed to the solvent and not in an intramolecular hydrogen bond. Therefore, these changes directly report on N-H groups' solvent accessibility. At constant temperature, this exchange process occurs at a rate that increases 10 fold *per* pH unit<sup>10–12,20–24</sup>. Thus, here, by increasing the pH from 6.0 to 7.5, the hydrogen exchange rate constants increase 31.6 fold (*i.e.*, 10<sup>(7.5–6.0)</sup>). This causes a decrease of the NMR peak intensity, since the fact that the amide proton starts to jump more frequently back and forth between the water and the amide sites leads to an enhanced decay of the transverse magnetization during acquisition. Spectral changes are therefore dependent on the extent of the increase of the N-H exchange rate constant. For the most solvent accessible N-H groups, peaks may even disappear from the spectrum at pH 7.5 (*e.g.*, residues K7, A8 and L44 in Fig. 1d). These spectral changes can be highly informative if properly explored, reporting on structural and dynamic properties of proteins. As such, we studied them here, to develop a method that provides insights into protein structure and function, both at the individual amino acid and domain level.

To establish this new methodology, we first compared the maximum variation of pH values tested, by plotting the intensities at pH 6.0 and 7.5 as a function of protein sequence (Fig. 2a, gray and black, respectively). At pH 6.0, the N-H groups of N-terminal region residues display higher peak intensities, consistent with their disordered nature<sup>26,31,32</sup>. To simplify the analysis and compensate for differences in initial intensity (Int<sub>pH 6.0</sub>), results were normalized by the ratio between the intensities at pH 7.5 and 6.0 (Int<sub>pH 7.5</sub>/Int<sub>pH 6.0</sub>; Fig. 2b). The whole N-terminal region and specific residues located in the α1 and near loop regions decrease their intensity as the pH increases to 7.5. These findings are worth considering in the context of DENV C three main structural regions (Fig. 1b and Table S1). Briefly, at pH 7.5, the peak intensity of some residues is less than half of their initial values (Fig. 2b), namely: R5 to R22 (except P12) in the N-terminal region; V23 and T25 in the D1 domain; Q27 and T30 in α1; S34, R41, G42 and L44 in L1-2; A49 in α2; I59 and G64 near L2-3; K74, S75 and K76 near L3-4; and, R99 in the C-terminal domain. As such, the three main structure/dynamics regions of DENV C are clearly distinguishable, as further detailed ahead.

**Average solvent accessibility of protein regions.** Given the above, we then analyzed the intensity changes in the context of the protein structure and dynamics of the main regions of the protein. For such purpose and although each N-H group of an individual amino acid behaves differently in response to pH<sup>10–12,20–24</sup>, we considered that structural factors are more determinant and we averaged the backbone N-H group response to pH across regions. Those that are protected, either by being buried within the structure or within an intramolecular hydrogen bond, will not be affected by pH. The amino acids that are not protected will be responsive to pH within the pH range tested here. The ratios determined in Fig. 2b, when averaged across a protein region or domain, provide a single parameter to distinguish between structural and dynamics sections.

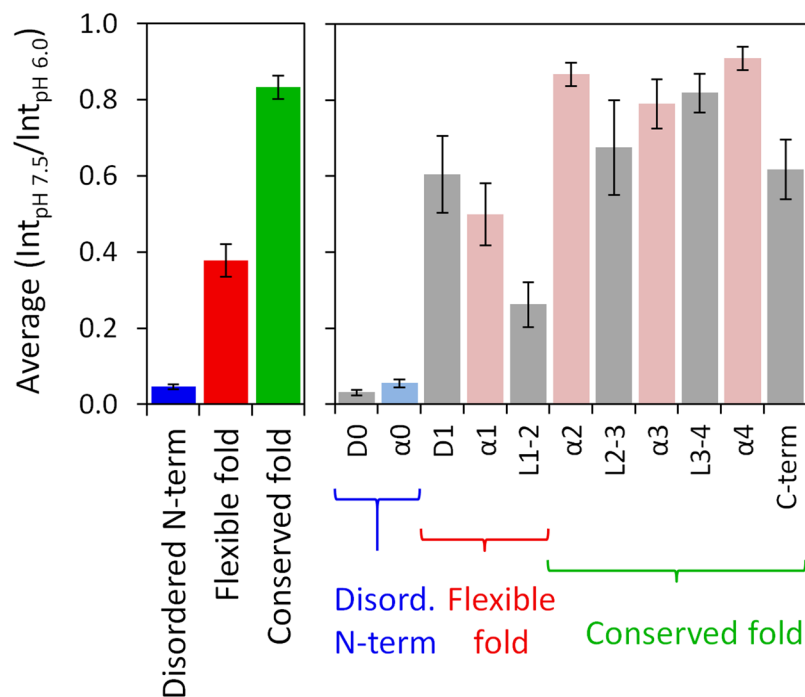
Fig. 3 depicts the average Int<sub>pH 7.5</sub>/Int<sub>pH 6.0</sub> of the three main structure/dynamics regions (left panel) and of the secondary structure domains (right panel). The main structural regions are distinguished by their average backbone solvent accessibility (Fig. 3, left panel): the disordered N-terminal backbone is highly exposed, the flexible fold is partially accessible, and the conserved fold is mostly inaccessible to the solvent. Looking at the secondary structure domains (Fig. 3, right panel), the average values show that the backbone of the α0 domain, which is disordered and may transiently adopt an α-helical secondary structure<sup>29</sup>, is highly exposed to the solvent. Among the α-helices, the backbone of α1 presents values that are in between the obtained for α0 and those of the α2, α3 and α4 backbones, suggesting an intermediate exposure of α1 backbone to the solvent, implying a certain degree of flexibility. Therefore, DENV C α1 has more freedom to interact with the solvent, in line with our previous studies<sup>26,29</sup>. Moreover, the average backbone values for loop regions L2-3 and L3-4 are, in general, similar to those of nearby α-helices (Fig. 3, right panel). Thus, this approach probes the differences in backbone N-H groups solvent accessibility for the main structure/dynamics regions (Fig. 3, left panel), as well as for the secondary structure domains (Fig. 3, right panel).



**Figure 2.** DENV C  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities at pH 6.0 and pH 7.5, and their ratio. **(a)** DENV C  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities at pH 6.0 (gray bars) and pH 7.5 (black bars), and **(b)** ratio of HSQC peak intensities at pH 7.5 and pH 6.0 ( $\text{Int}_{\text{pH } 7.5} / \text{Int}_{\text{pH } 6.0}$ ). Error bars represent standard error (SE). The symbols in **a** encode the reason why the respective residues could not be analyzed by NMR: ‘#’ for residues that are not assigned, ‘o’ for overlaps, ‘\*’ for absent resonances due to line broadening, and ‘P’ for prolines. The horizontal line on **b** marks the ratio equal to 1. The main structural features are indicated on the top of the figure: the three structure/dynamics regions<sup>26,31,32</sup>, the secondary structure domains<sup>29,31,32</sup>, and the protein primary sequence. Colored columns are a guide for the data corresponding to each secondary structure domain (pink columns represent the experimentally determined  $\alpha$ -helices<sup>31,32</sup>, while the cyan column corresponds to the transient  $\alpha$ -helix suggested by our previous work<sup>29</sup>).

**$^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities in function of pH.** The changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum for each pH tested (6.0, 6.5, 6.75, 7.0, 7.25 and 7.5) were assigned to the respective individual N-H groups of the protein (Fig. 1c,d), to give a complete picture of the peak evolution with pH. Fig. 4a shows the evolution of  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities as a function of pH for three residues (M15, T30 and R97) representative of the three main structural regions. Importantly, the solvent accessibility probed via the approach presented reports the interaction of each specific amide group with water. It can be used to distinguish the solvent accessibility of backbone and side-chain N-H groups within the same residue, as shown for W69 N-H groups (Fig. S3), where the backbone amide is not affected by pH, while the indole N-H group value varies significantly. Therefore, each N-H group reports its own microenvironment. The highly localized probing sensitivity illustrates the methodology great resolution level, a property that can be exploited to gain vital structural and dynamics information.

Besides this, we can analyze the normalized intensity of the backbone N-H groups, as a function of pH, for each amino acid, with striking differences between amino acids from different regions of the protein (Fig. 4a). An average of the normalized intensity of all the amino acids for each main region can then be obtained (Fig. 4b). The three key structural regions of DENV C are clearly distinguished (Fig. 4b): the conserved fold suffers no major changes (green), the N-terminal suffers the greatest change (blue), while the flexible fold shows an intermediate regime (red). The flexible fold also has larger error bars (Fig. 4b, red), indicative of higher heterogeneity among the constituting residues solvent accessibility. For each secondary structure domain, the average intensities as a function of pH are available in Fig. S4. Since the  $\alpha 0$  domain<sup>29</sup> is mostly disordered in solution<sup>26,31,32</sup>, its backbone average solvent accessibility is higher than for other  $\alpha$ -helical domains, as expected. Importantly,  $\alpha 1$  displays an intermediate accessibility and the other  $\alpha$ -helical domains backbones are generally not exchanging the amide hydrogen with the solvent, in agreement with Fig. 3 data. Regarding loop domains, D0 and L1-2 have their backbone N-H groups mostly interacting with the solvent, while other loops are roughly unable to perform amide hydrogen exchange, in accordance with the analysis of Fig. 3. Therefore, we can obtain a single parameter that describes individual and regional exposure to the solvent, as described hereafter.



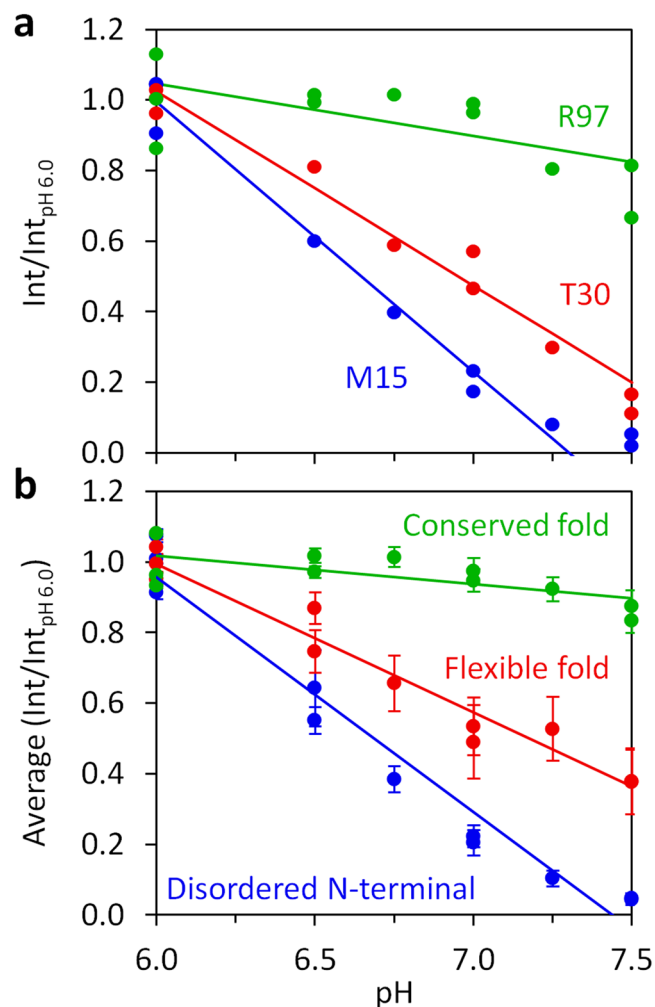
**Figure 3.** Average of the NMR peak intensities ratio for the three structure/dynamics regions and secondary structure domains of DENV C. The NMR peak intensities ratio between pH 7.5 and pH 6.0 for each residue (data from Fig. 2b) were averaged across the residues that comprise each of the three structure/dynamics regions (left panel) and secondary structure domains (right panel) of DENV C. Error bars are SE.

**Linearity of intensities versus pH.** The backbone N-H peak intensities of individual amino acid residues follow a roughly linear decrease with pH (Fig. 4a), which is also observed for the average of the main regions (Fig. 4b) and domains (Fig. S4). As such, an approximation was used by fitting the following empirical linear equation to the data:

$$\frac{\text{Int}}{\text{Int}_{\text{pH } 6.0}} \approx \text{Slope} \times (\text{pH} - 6.0) + 1 \quad (1)$$

where Int is the  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensity at a given pH and  $\text{Int}_{\text{pH } 6.0}$  is the average intensity from 3 independent measurements at pH 6.0. The fitting of this equation retrieves the slope, which is a parameter that describes the average value of the derivative  $d(\text{Int}/\text{Int}_{\text{pH } 6.0})/d\text{pH}$  throughout the pH interval probed. A formal approach was also devised based on the literature<sup>10–12,20–24</sup>, which can be found on the Supplementary Note (of the Supplementary Information file), leading to the pH dependencies of both  $\text{Int}/\text{Int}_{\text{pH } 6.0}$  and  $d(\text{Int}/\text{Int}_{\text{pH } 6.0})/d\text{pH}$ . Importantly, this simpler slope-based (linear) approach retrieves a single fitting parameter that entirely describes the trend, being independent of external parameters estimation that sometimes are difficult to determine (*i.e.*,  $k_{\text{c}}$  or  $t$  values; for details, please consult the Supplementary Note). In practice, the more negative is the slope, the more susceptible to exchange is the corresponding N-H group. Slopes and  $\text{Int}_{\text{pH } 7.5}/\text{Int}_{\text{pH } 6.0}$  values are comparable, as explained hereafter. Since slopes are originated from measurements at several pH values, they are a better parameter to represent each N-H group solvent accessibility, being of use to more advanced applications, and were employed henceforth.

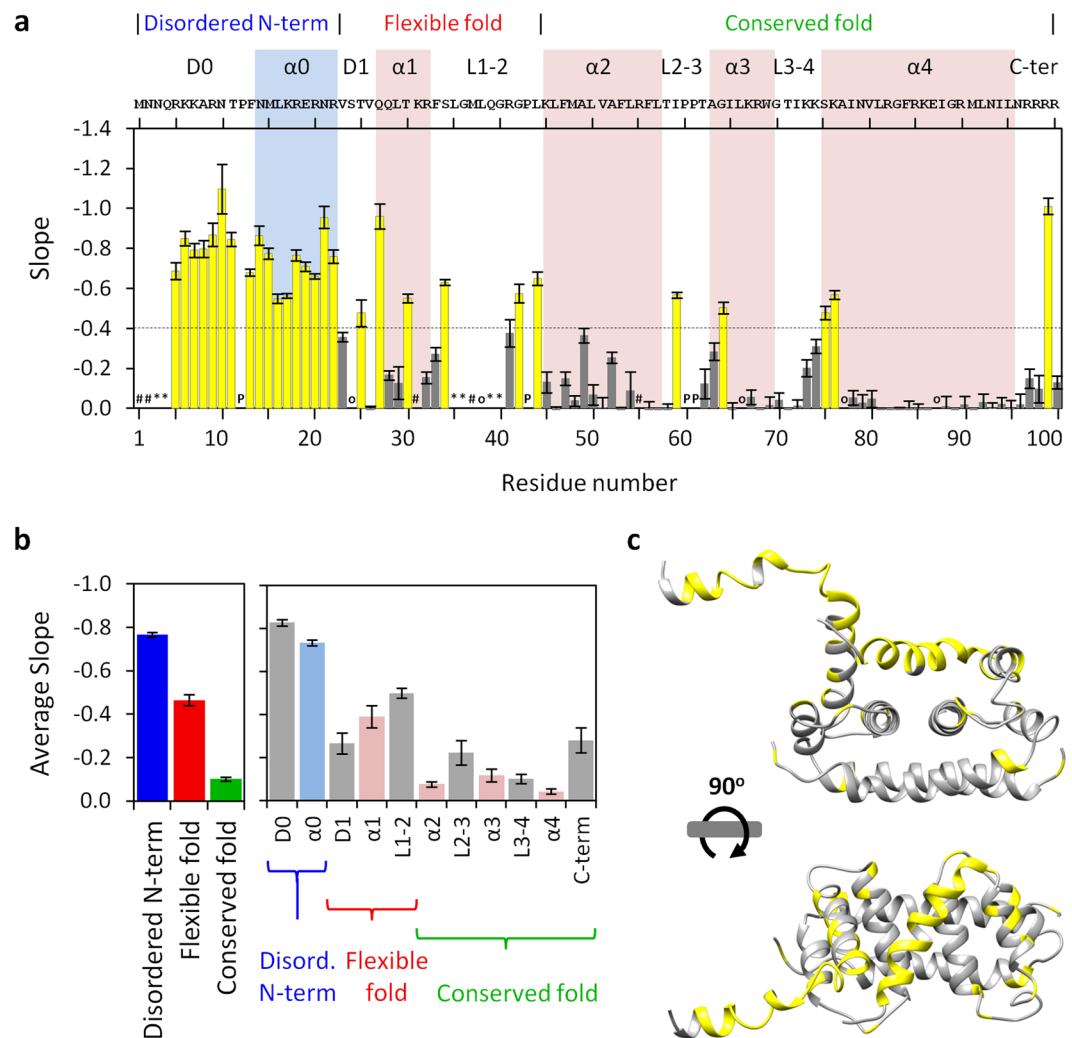
**DENV C structure/dynamics and the slope information.** Slope values were calculated via equation 1 for each analyzable DENV C backbone N-H group (Fig. 5a). The average slope values of the three major regions and of the secondary structure domains (Fig. S4) were then computed (Fig. 5b). The information obtained is similar to the one derived from the  $\text{Int}_{\text{pH } 7.5}/\text{Int}_{\text{pH } 6.0}$  values (compare Figs. 5a and 2b, and also Figs. 5b and 3). In Fig. 5a, it is easy to distinguish the individual N-H groups that are fully exposed, intermediately exposed or buried away from the solvent. This is also clear in Fig. 5b for the three main regions and the several secondary structure domains. Noteworthy, within a given region, interconnecting loops seem to be more dynamic and exposed than adjacent  $\alpha$ -helical domains, in consonance with the protein structure. Slope values of 0 (Fig. 5a) are from N-H groups of residues that cannot change their hydrogen with the solvent (corresponding to  $\text{Int}_{\text{pH } 7.5}/\text{Int}_{\text{pH } 6.0}$  values of 1, in Fig. 2b). Slopes with absolute value higher than 0.7 (Fig. 5a) arise from N-H groups which are performing H-bonds with the water (corresponding to  $\text{Int}_{\text{pH } 7.5}/\text{Int}_{\text{pH } 6.0}$  values of 0, in Fig. 2b). A slope threshold of  $-0.4$  distinguishes the more solvent exposed N-H groups (Fig. 5a, yellow) from those less exposed (Fig. 5a, gray). A detailed analysis of Fig. 5a using this threshold reveals that the most solvent accessible N-H groups are from residues R5 to R22, T25, Q27, T30, S34, G42, L44, I59, G64, S75, K76 and R99. These residues are accordingly



**Figure 4.** Normalized peak intensities of individual N-H groups and the average among DENV C structure/dynamics regions, as a function of pH. **(a)**  $\text{Int}/\text{Int}_{\text{pH}6.0}$  variation with pH for the backbone N-H groups of M15, T30 and R97, representing residues in the disordered N-terminal, flexible fold and conserved fold regions, respectively. **(b)** Average intensities ratio for the three major structure/dynamics regions of DENV C. In all graphs, lines are fits of equation 1 to the data, from which slopes were extracted.

depicted in the protein structure (Fig. 5c, matching yellow and gray residues), providing direct information on both IDP and ordered regions of DENV C protein, which are immediately distinguishable. Moreover, many of these residues are located at the beginning of all the protein  $\alpha$ -helices, which gives information on protein structure.

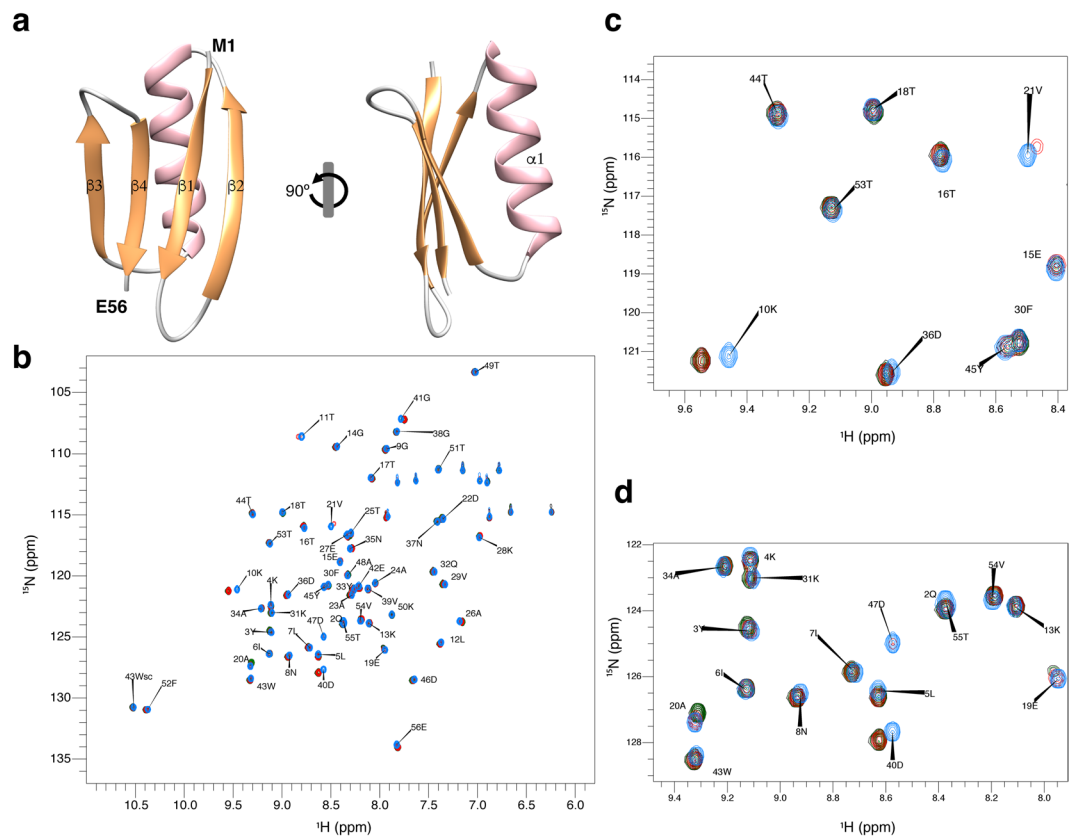
The fact that many of the first  $\alpha$ -helical residues have N-H groups exposed to the solvent is a direct insight into the nature of  $\alpha$ -helices. In an  $\alpha$ -helix, the first residue is establishing H-bond via its C=O group with the N-H group of the fourth residue, leaving its own N-H group free of H-bonds involved in the  $\alpha$ -helix stabilization. This means that the N-H groups of the first three residues of  $\alpha$ -helices are free to establish H-bonds with other nucleophile groups (that serve as hydrogen bond acceptors) either from the protein, becoming unavailable to the solvent, or from the solvent. If they are exposed to the water, their amide hydrogen can exchange with those from the solvent. This is exactly what we observe in DENV C  $\alpha$ -helices, by analyzing the backbone N-H groups performing intramolecular H-bonds within the DENV C structure (Fig. S5a). We then compared the slopes information with the normalized frequency of intramolecular H-bonds *per* N-H group (Fig. S5b), finding a clear correlation of the slopes with DENV C structure. Interestingly, residues that have low frequency of N-H intramolecular H-bonds ( $<0.5$ ) and low slope values (between  $-0.4$  and  $0$ ) are thus free to perform hydrogen exchange, but are unable to do so. This suggests that they are not facing the solvent because they are buried within the protein. In summary, an N-H group from a specific residue needs to be both free of intramolecular H-bonds and exposed to the solvent in order to exchange its hydrogen with the water. Overall, our findings suggest that the probing of the N-H groups' solvent accessibility of a protein, via minor pH changes, may be used as an additional structure and dynamics restraint to help on the calculation of protein structures.



**Figure 5.** Slopes of  $\text{Int}/\text{Int}_{\text{pH } 6.0}$  versus pH in the context of the DENV C structure. **(a)** Slope of the intensities ratio versus pH along the DENV C sequence (an inverse scale is shown since the more efficient is the hydrogen exchange process, the more negative is the slope). The threshold of  $-0.4$  (dashed line) was defined to identify the DENV C backbone N-H groups that are highly exposed to the solvent (yellow bars). For details on the protein structural information and symbols, on top and within the graph (respectively), please refer to the legend of Fig. 2. **(b)** Average slopes of the three structure/dynamics regions (b, left panel) and of the secondary structure domains (b, right panel) of DENV C. Error bars in **a** and **b** represent SE. **(c)** DENV C residues in which the backbone N-H is highly exposed to the water were highlighted within the protein structure (yellow regions). Clearly, from **a** and **c**, all the residues of the disordered N-terminal region, some specific residues on the flexible fold region and residues in the beginning of the  $\alpha$ -helices are able to exchange their backbone amide hydrogen with the water.

**Applying the method to GB1 protein.** Having established the method applicability with DENV C, we proceeded to test it with the B1 immunoglobulin binding domain of streptococcal protein G (GB1), which contains 56 amino acid residues and a structure of four stranded  $\beta$ -sheets with one long  $\alpha$ -helix on top (Fig. 6a), as shown by X-Ray diffraction crystallography as well as by NMR (PDB ID: 2GB1 and 5JXV)<sup>35,36,43</sup>. GB1 has been extensively studied by different biophysical methods and is one of the smallest stable folded globular domains known. A pH interval of 1.5 was assayed as well, but now changing the pH from 6.5 to 8.0. No major conformational changes were seen (Fig. 6b), only minor local switches (Fig. 6c,d). As the overall protein structure remains highly stable within that pH range<sup>37</sup>, we went further and tested if the intensities of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC peaks revealed any changes (Fig. 7). As for DENV C (Fig. 2), by directly comparing peak intensities at pH 6.5 and 8.0 on GB1 (Fig. 7a) or the ratio between these intensities (Fig. 7b), the major regions of the protein with exposed backbone amide nitrogen atoms can be readily identified, namely the loops, the outer strands of the four-stranded  $\beta$ -sheet (*i.e.*,  $\beta 2$  and  $\beta 3$ ) and the beginning of the  $\alpha$ -helix, which are free of backbone intramolecular H-bonds and accessible to the solvent.

Then, with the above in mind, we tested the use of the slope to map the protein regions most accessible to the solvent (Figs. 8 and S6), using the same cut-off as for DENV C. The information fits well with the known pattern



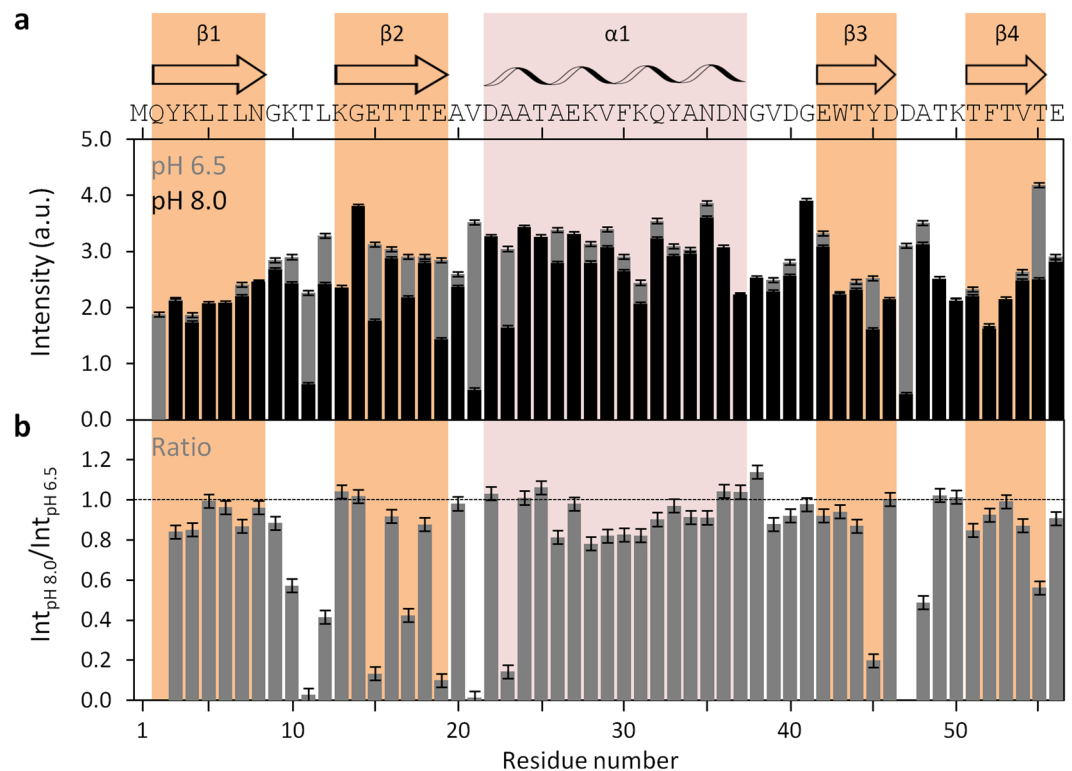
**Figure 6.** GB1 structure and  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra from pH 6.5 to 8.0. (a) GB1 experimental structure. GB1 is a domain of immunoglobulin G binding protein that is negatively charged, with 6 cationic and 10 anionic residues out of 56. It consists of four  $\beta$ -sheets, named  $\beta 1$  to  $\beta 4$  (colored in orange), plus one  $\alpha$ -helix, named  $\alpha 1$  (colored in pink), connected by short loops (PDB ID 5JXV<sup>43</sup>). (b) Superimposed GB1  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra at pH 6.5 (blue), 7.3 (red), 7.6 (green) and 8.0 (black). Major chemical shifts corresponding to large conformational rearrangements are not observed. (c and d) Zoom on spectral regions where all types of responses to the pH variation are observed: peaks that vary in intensity and/or display a small chemical shift perturbation. Most peaks neither vary in intensity nor in chemical shift. UCSF Chimera v1.9 software<sup>54</sup> was used for protein structure visualization<sup>43</sup>.

of GB1 solvent-accessible surface area (SASA) along the sequence<sup>44</sup>, supporting the methodology employed. Moreover, our results are in accordance with H/D exchange studies, as the same regions identified as being more accessible to exchange with the solvent are the most sensitive to pH<sup>38</sup>. More protected residues that exchange through a global unfolding mechanism (e.g., residues K4, L5, A26, F30 and T44) or a local high energy unfolding mechanism (e.g., residues K28, Y33, N35 and T55) display minimal changes with pH, while the regions that correspond to fast exchanging non H-bonded N-H groups are clearly visible (e.g., residues T17, E19 and V21). This is also supported by the intramolecular H-bonds frequency analysis (Fig. S6), similarly to DENV C (Fig. S5). All this information further validates the methodology employed and suggests its applicability in other studies, as discussed ahead.

## Discussion

Taking advantage of the amide hydrogen exchange with water<sup>10–12,20–24</sup>, we established a new NMR approach to determine protein backbone solvent accessibility. Solvent accessibility correlates with the general structure/dynamics regions of a protein, with the explanation being straightforward: in more dynamic regions the N-H groups are more susceptible to hydrogen exchange with the solvent, as they are not involved in stable secondary structure elements. We may thus probe solvent accessibility by slightly varying the pH of a protein solution, as demonstrated by the study with DENV C, where a 1.5 pH units interval was employed (pH 6.0 to 7.5). The main protein regions (Fig. 1b) can be clearly distinguished via their individual amino acids (Figs. 2b and 5a) and their averaged backbone solvent exposure (Figs. 3, 4b and 5b). We can discriminate between the exposed disordered N-terminal, the partially accessible flexible fold and the mostly inaccessible conserved fold, in accordance with the current understanding of DENV C properties<sup>26,29,31,32</sup>. Noteworthy, the flexible fold intermediate behavior, previously predicted<sup>26</sup>, is confirmed here (Figs. 3, 4b and 5b, red). Furthermore, at single residue resolution, the solvent accessibility of each N-H group clearly correlates with its normalized frequency of intramolecular H-bonds in DENV C structure (Fig. S5). Such correlation is specific for each secondary structure element (such as  $\alpha$ -helices and  $\beta$ -sheets), depending on the relative position in the protein structure. The same approach was then applied to a different system, GB1, a well structured protein domain, the B1 domain of immunoglobulin protein



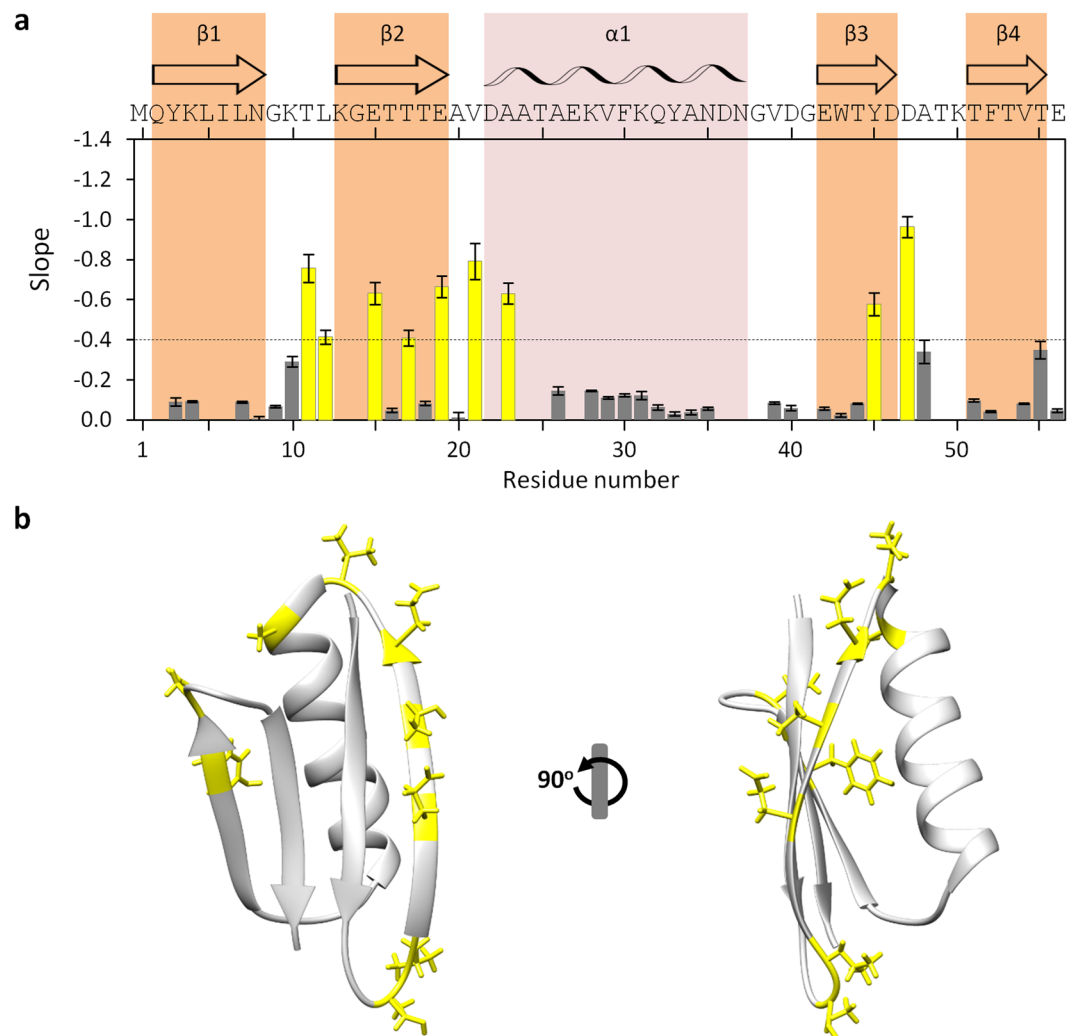


**Figure 7.** GB1  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities at pH 6.5 and pH 8.0, and ratio between peaks. (a) GB1  $^1\text{H}$ - $^{15}\text{N}$  HSQC peak intensities at pH 6.5 (gray bars) and pH 8.0 (black bars), and (b) ratio of HSQC peak intensities between pH 8.0 and pH 6.5 ( $\text{Int}_{\text{pH } 8.0}/\text{Int}_{\text{pH } 6.5}$ ). Error bars represent standard error (SE). The M1 residue is not assigned, while Q2 was not analyzed due to absent resonance, as a result of line broadening. All other amino acid residue peaks were assigned and analyzed. The horizontal line on **b** marks the ratio equal to 1. The main structural features are indicated on the top of the figure: the secondary structure domains<sup>36</sup>, and the protein primary sequence. Colored columns are a guide for the data corresponding to each secondary structure domain (orange and pink columns correspond to experimentally determined  $\beta$ -sheets and  $\alpha$ -helix, respectively).

G. We employed another physiological interval of 1.5 pH units, from 6.5 to 8.0 (Fig. 6). The data shows that also with this protein there are no major conformational changes. Moreover, the key regions of the protein are also readily identified (Fig. 7), namely at the single amino acid level (Fig. 8).

Therefore, the pH range tested (6.0–8.0) is particularly suitable since most proteins isoelectric point is not close to the physiological pH (since they would not be functional as they could precipitate) and, apart from histidines, titratable amino acids are usually not affected at this pH range, as inferred from Fig. S1. Even if minor conformational changes occur, these would be readily visible in the NMR spectra through chemical shift perturbation (CSP) studies, allowing it to be taken into account in the analysis. So, the methodology can be easily applied to other proteins, in pH ranges where their general structure and dynamics properties are maintained. After NMR assignment at low pH (e.g., pH 6.0), one can perform a pH increase on the same sample (at least one pH unit is recommended) and acquire one more  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum to determine immediately which N-H groups and protein regions interact with the solvent. A single protein preparation (at concentrations around 5–10 mg/mL) can be used and two measurements at different pH values readily provide key information. Even if there is a mild conformational change triggered by pH and/or titration of some specific amino acid residue(s), a pH variation can be performed by acquiring  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra at several slightly spaced pH values, to follow the NMR peaks evolution (CSP analysis). In particular, the pH range used here (between 6.0 and 8.0) has the advantage of being physiologically relevant and compatible with the timescale of usual  $^1\text{H}$ - $^{15}\text{N}$  HSQC measurements.

Moreover, the approach is particularly useful to study IDP regions or when NMR data is sparse due to time or other constraints. As mentioned, the intensity of a signal depends on the line width, which is mainly influenced by the protein correlation time and the chemical exchange regime: sharp for IDPs (due to correlation time below 1 ns, but potentially broadened due to accessibility for water exchange) while amides in the folded part of a protein will experience less line broadening due to slower water exchange. Overall, if one changes the pH by 1.5 units, the base-catalyzed hydrogen exchange increases by  $10^{1.5}$ , which has large effects on the intensity (the reciprocal line width) of the IDP signals and a smaller effect on the folded parts, as readily observed here. Other methodologies have been described to improve protein structure determination, namely residual dipolar couplings (RDCs)<sup>14,17</sup>, diffusion tensor parameters<sup>15</sup>, relaxation parameters<sup>16</sup> and paramagnetic relaxation enhancement (PRE) probes<sup>17</sup>, among others. Nevertheless, since the method described here is much simpler and easier to interpret, we believe that it will be widely adopted for protein structural and dynamics studies.



**Figure 8.** Slopes of  $\text{Int}/\text{Int}_{\text{pH } 6.5}$  versus pH in the context of the GB1 structure. **(a)** Slope of the intensities ratio versus pH along GB1 sequence (an inverse scale is shown since the more efficient is the hydrogen exchange process, the more negative is the slope). As in Fig. 5, the threshold of  $-0.4$  (dashed line) was defined to identify GB1 backbone N-H groups that are highly exposed to the solvent (yellow bars). Error bars represent standard error (SE). Residues with positive slopes are not displayed (these are all close to zero). The protein structural information and symbols are colored as described in Fig. 7 legend. **(b)** GB1 residues in which the backbone N-H is exposed to the water were highlighted (yellow) within the protein structure, showing that only the amino acids exposed to the solvent (and not protected by intramolecular H-bonds) exchange their backbone amide hydrogen with the water.

Beyond the methodological development, the analysis employed in this study also gives important information on the DENV C structure/dynamics properties and their functional implications. From this analysis it is clear that the three major regions are directly observable and that the  $\alpha$ -helices identified have different dynamic properties. Briefly,  $\alpha 0$  region is more malleable than the other secondary structure domains and  $\alpha 1$  is partially flexible when compared to  $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$ . The N-terminal IDP region as well as the first amino acids of the  $\alpha$ -helices and the loop domains are accessible to the solvent, in agreement with the available DENV C structure (PDB ID 1R6R<sup>32</sup>). Overall, the results are consistent with the current understanding of DENV C structure and function<sup>25,26,28–32</sup>, supporting the use of the proposed methodology to investigate protein architecture.

Importantly, the findings concerning the flexible fold region are particularly new and shed a new light on the protein biological activity. The term “flexible fold”, to refer to that section of the protein, was first used by us upon comparing DENV C NMR structure with the capsid protein X-Ray crystallographic structure of the closely related West Nile virus<sup>26,32,33</sup>. At that time, it only meant that this section, which contains an  $\alpha$ -helix in both C proteins, was not folded in a similar manner in the context of the homodimer. Very recent studies showed the same when comparing with the Zika virus C protein structure<sup>34</sup>. Nevertheless, only now did it become evident that in DENV C this section is much more free to interact with the solvent. This, alongside with the N-terminal IDP region, may modulate an auto-inhibitory mechanism that regulates the solvent access to the hydrophobic interface region previously proposed by us<sup>29</sup>. This hydrophobic pocket is essential for DENV C binding to lipid

droplets and, subsequently, for viral replication. Therefore, by targeting this region, such understanding may help to improve and design new drugs against dengue virus and related flaviviruses.

In addition, our studies with GB1 (Figs. 6, 7 and 8) fully demonstrate that those regions and amino acid residues most exposed to the solvent are easily distinguished. The data also correlate well with the intramolecular H-bond pattern (Fig. S7). In short, the method is applicable to different proteins and can complement, in a faster way, present NMR experimental routines.

To conclude, while gaining important information on a key viral protein, we also report a fast NMR-based method to determine the protein backbone solvent accessibility at single amino acid N-H group resolution, which is applicable to other proteins besides DENV C, as demonstrated here with GB1. The approach is of special interest to study IDP regions, where classical techniques are difficult to employ. Moreover, the method is applied in a physiologically relevant pH range, providing valuable insights into protein structure and function. This work provides the basis to further studies, where the application of small physiological pH changes to interrogate protein structure, dynamics and solvent accessibility is conducted. These can be performed in conditions suited for the analysis of IDP and structured regions, complementing the array of methodologies available for protein studies, particularly of IDP regions that are difficult to assay via current techniques.

## Methods

**Materials.** Isotopically enriched chemical compounds, namely [ $^{15}\text{N}$ ]- $\text{NH}_4\text{Cl}$ , [ $^{13}\text{C}$ ]-glucose and  $^2\text{H}_2\text{O}$ , were purchased from Cambridge Isotope Laboratories (Tewksbury, MA, USA). Chromatography columns HiTrap Heparin, HiTrap Q HP and Superdex 75 10/300 GL, the Peristaltic Pump P-1 and AKTA Start were from GE Healthcare (Little Chalfont, UK). Reagents for SDS-PAGE were from BioRad (Hercules, CA, USA). Dialysis membrane tube Spectra/Por of 3.5 kDa nominal cut-off was from Spectrum Laboratories (Los Angeles, CA, USA) and Amicon Ultra-4 or -15 centrifugal filters of 10 or 3 kDa nominal cut-off were from Millipore (Billerica, MA, USA). Unless otherwise described, all other chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA).

**DENV C – heterologous protein expression and purification.** DENV C protein purification protocol was optimized from previous studies<sup>25,26,28,31,32,45</sup>. The protein was expressed in *E. coli* BL21-CodonPlus transformed with a pET-21a plasmid with a gene encoding the capsid protein of DENV serotype 2, strain New Guinea C (NCBI ID AAC59275, corresponding to amino acids 1–100 of the polyprotein)<sup>25,26,28,46</sup>. Further details are available as Supplementary Information, in the Supplementary Methods section.

**GB1 – heterologous protein expression and purification.** The pET11a plasmid containing the gene encoding T2Q B1 immunoglobulin G binding domain of streptococcal protein G (GB1) was kindly provided by Professor Gary Pielak, from University of North Carolina at Chapel Hill. The T2Q mutation prevents N-terminal deamidation. This form is mentioned here as wild type (WT) or only “GB1”. The isolation and purification of  $^{15}\text{N}$ ,  $^{13}\text{C}$  enriched GB1 was optimized from previous studies<sup>35–37,44</sup>. Further details are available as Supplementary Information, in the Supplementary Methods section.

**DENV C – pH variation, sample preparation and NMR experiments.** Prior to the measurements, DENV C stock solution was diluted to 2/5 with a 55 mM  $\text{KH}_2\text{PO}_4$ , 13.75 mM  $\text{NaN}_3$ , pH 6.0 solution and the protein concentration was adjusted to 550  $\mu\text{M}$  (monomer), either by diluting or concentrating (with Amicon Ultra-4 Centrifugal Filters of 10 kDa cut-off) with a solution of 55 mM  $\text{KH}_2\text{PO}_4$ , 220 mM KCl and 5.5 mM  $\text{NaN}_3$ , pH 6.0. In order to change the pH in a controlled way in small volumes, we performed a dialysis of 500  $\mu\text{L}$  of protein solution, containing 550  $\mu\text{M}$  of DENV C, 55 mM  $\text{KH}_2\text{PO}_4$ , 220 mM KCl and 5.5 mM  $\text{NaN}_3$ , pH 6.0, in an uncapped eppendorf tube sealed with a Spectra/Por dialysis membrane of 3.5 kDa cut-off, which was then tightly sealed with Parafilm. Each eppendorf tube was put upside down in a different flask containing 100 mL (to be 200  $\times$  500  $\mu\text{L}$ ) of the same buffer solution in which the protein is dissolved in, but with different pH value. These dialyses were performed at room temperature at least for 1 h before the experiment. After dialysis and a quick spin-down centrifugation of the eppendorf tubes,  $^2\text{H}_2\text{O}$  was added to a 10% (v/v) final concentration and the solution was transferred to a standard NMR tube. In every measurement, we checked the final pH within the NMR tube (after  $^2\text{H}_2\text{O}$  addition), which matched the expected pH. The final solutions contained 500  $\mu\text{M}$  of DENV C monomer, 50 mM  $\text{KH}_2\text{PO}_4$ , 200 mM KCl, 5 mM  $\text{NaN}_3$  and 10% (v/v)  $^2\text{H}_2\text{O}$ , at different pH values: 6.0, 6.5, 6.75, 7.0, 7.25 or 7.5. NMR peak intensities are normalized to the average intensity at pH 6.0 ( $\text{Int}/\text{Int}_{\text{pH } 6.0}$ ).

We performed  $^{15}\text{N}$  transverse relaxation ( $R_2$ ) NMR experiments were performed at 298.15 K in a Bruker Avance III 800 MHz equipped with a triple resonance ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ) probe. Spectra were processed using NMRPipe<sup>47</sup> and analyzed with NMRView<sup>48</sup>. These spectra were acquired as pseudo-3D, with 2D  $^1\text{H}$ -detected,  $^{15}\text{N}$ -edited HSQC experiments, implementing standard pulse sequences<sup>40–42,49</sup>.  $R_2$  spectra were recorded with spectral widths of 1024  $\times$  256 complex points in the  $^1\text{H}$  and  $^{15}\text{N}$  dimensions, respectively. The field strength of the CPMG refocusing train was 500 Hz and a 1.2 ms delay was used between the refocusing pulses<sup>50,51</sup>. The effects of cross relaxation between  $^1\text{H}$ - $^{15}\text{N}$  dipolar and  $^{15}\text{N}$  chemical shift anisotropy were removed by applying  $^1\text{H}$  180° pulses during relaxation delays<sup>52</sup>. Further details are available in the Supplementary Information, Supplementary Methods section.

**GB1 – pH variation, sample preparation and NMR experiments.**  $^{13}\text{C}$ - $^{15}\text{N}$  GB1 was dissolved in 100 mM KCl buffer in 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$  containing 100  $\mu\text{M}$  DSS (used as internal reference) to a protein concentration of 1 mM. The pH was measured using a Docu-pH meter (Sartorius) calibrated with standard solutions. The initial pH was 6.50. For each pH step, the pH was adjusted with microliter additions of 0.15 M NaOH solution. The concentration of added salt amounted to less than 1 mM. Spectra were collected in steps of  $\sim 0.4$  pH units from pH 6.5 to 8.0. Data were processed using Bruker TopSpin<sup>TM</sup> 4.0 and analyzed with CCPNMR<sup>53</sup> for

cross-peak assignment and height extraction. Further details are available in the Supplementary Information, Supplementary Methods section.

**Further experimental details.** Further details are available as Supplementary Information, in the Supplementary Note and the Supplementary Methods sections.

## Data Availability

All data generated or analyzed during this study are included in this article and Supplementary Information file.

## References

- Cavanagh, J., Fairbrother, W. J., Palmer, A. G. I., Rance, M. & Skelton, N. J. *Protein NMR Spectroscopy – Principles and Practice*. Second Edition (2007).
- Raman, S. *et al.* NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014–1018 (2010).
- Korzhev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
- Lange, O. F. *et al.* Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc. Natl. Acad. Sci. USA* **109**, 10873–10878 (2012).
- van der Schot, G. *et al.* Improving 3D structure prediction from chemical shift data. *J. Biomol. NMR* **57**, 27–35 (2013).
- van der Schot, G. & Bonvin, A. M. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. *J. Biomol. NMR* **62**, 497–502 (2015).
- Mittermaier, A. K. & Kay, L. E. Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* **34**, 601–611 (2009).
- Kay, L. E. NMR studies of protein structure and dynamics. *J. Magn. Reson.* **173**, 193–207 (2005).
- Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814**, 942–968 (2011).
- Englander, S. W. & Mayne, L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Structure* **21**, 243–265 (1992).
- Krishna, M. M., Hoang, L., Lin, Y. & Englander, S. W. Hydrogen exchange methods to study protein folding. *Methods* **34**, 51–64 (2004).
- Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. Hydrogen exchange: the modern legacy of Linderstrom-Lang. *Protein Sci.* **6**, 1101–1109 (1997).
- Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *J. Am. Chem. Soc.* **125**, 15686–15687 (2003).
- Zeng, J. *et al.* High-resolution protein structure determination starting with a global fold calculated from exact solutions to the RDC equations. *J. Biomol. NMR* **45**, 265–281 (2009).
- Ryabov, Y., Suh, J. Y., Grishaev, A., Clore, G. M. & Schwieters, C. D. Using the experimentally determined components of the overall rotational diffusion tensor to restrain molecular shape and size in NMR structure determination of globular proteins and protein-protein complexes. *J. Am. Chem. Soc.* **131**, 9522–9531 (2009).
- Ryabov, Y., Schwieters, C. D. & Clore, G. M. Impact of  $^{15}\text{N}$   $R_2/R_1$  relaxation restraints on molecular size, shape, and bond vector orientation for NMR protein structure determination with sparse distance restraints. *J. Am. Chem. Soc.* **133**, 6154–6157 (2011).
- Jensen, M. R., Zweckstetter, M., Huang, J. R. & Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **114**, 6632–6660 (2014).
- Rossi, P. *et al.* A hybrid NMR/SAXS-based approach for discriminating oligomeric protein interfaces using Rosetta. *Proteins* **83**, 309–317 (2015).
- Varadi, M. *et al.* pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.* **42**, D326–335, <https://doi.org/10.1093/nar/gkt960> (2014).
- Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
- Connelly, G. P., Bai, Y., Jeng, M. F. & Englander, S. W. Isotope effects in peptide group hydrogen exchange. *Proteins* **17**, 87–92 (1993).
- Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192–197 (1995).
- Koide, S., Jahnke, W. & Wright, P. E. Measurement of intrinsic exchange rates of amide protons in a  $^{15}\text{N}$ -labeled peptide. *J. Biomol. NMR* **6**, 306–312 (1995).
- Hwang, T. L., van Zijl, P. C. & Mori, S. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **11**, 221–226 (1998).
- Carvalho, F. A. *et al.* Dengue virus capsid protein binding to hepatic lipid droplets (LD) is potassium ion dependent and is mediated by LD surface proteins. *J. Virol.* **86**, 2096–2108 (2012).
- Martins, I. C. *et al.* The disordered N-terminal region of dengue virus capsid protein contains a lipid droplet-binding motif. *Biochem. J.* **444**, 405–415 (2012).
- Martins, I. C., Almeida, F. C. L., Santos, N. C. & Da Poian, A. T. DENV-Derived Peptides and Methods for the Inhibition of the Flavivirus Replication, Universidade Federal do Rio de Janeiro (UFRJ), Universidade de Lisboa (UL), Instituto de Medicina Molecular (IMM). International Patent Publication Nr WO/2012/159187 (2012).
- Faustino, A. F. *et al.* Dengue virus capsid protein interacts specifically with very low-density lipoproteins. *Nanomedicine: NBM* **10**, 247–255 (2014).
- Faustino, A. F. *et al.* Understanding dengue virus capsid protein disordered N-terminus and pep14-23-based inhibition. *ACS Chem. Biol.* **10**, 517–526 (2015).
- Faustino, A. F. *et al.* Understanding dengue virus capsid protein interaction with key biological targets. *Sci. Rep.* **5**, 10592, <https://doi.org/10.1038/srep10592> (2015).
- Jones, C. T. *et al.* Flavivirus capsid is a dimeric alpha-helical protein. *J. Virol.* **77**, 7143–7149 (2003).
- 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* **101**, 3414–3419 (2004).
- Dokland, T. *et al.* West Nile virus core protein; tetramer structure and ribbon formation. *Structure* **12**, 1157–1163 (2004).
- Shang, Z., Song, H., Shi, Y., Qi, J. & Gao, G. F. Crystal structure of the capsid protein from Zika virus. *J. Mol. Biol.* **430**, 948–962 (2018).
- Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **33**, 4721–4729 (1994).
- Gronenborn, A. M. *et al.* A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657–661 (1991).
- Lindman, S. *et al.* Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys. J.* **90**, 2911–2921 (2006).

38. Orban, J., Alexander, P., Bryan, P. & Khare, D. Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: comparison with calorimetric data. *Biochemistry* **34**, 15291–15300 (1995).
39. Toseland, C. P., McSparrow, H., Davies, M. N. & Flower, D. R. PPDv1.0—an integrated, web-accessible database of experimentally determined protein pKa values. *Nucleic Acids Res.* **34**, D199–203, <https://doi.org/10.1093/nar/gkj035> (2006).
40. de Medeiros, L. N. *et al.* Backbone dynamics of the antifungal Psd1 pea defensin and its correlation with membrane interaction by NMR spectroscopy. *Biochim. Biophys. Acta* **1798**, 105–113 (2010).
41. de Paula, V. S. *et al.* Structural basis for the interaction of human beta-defensin 6 and its putative chemokine receptor CCR2 and breast cancer microvesicles. *J. Mol. Biol.* **425**, 4479–4495 (2013).
42. de Paula, V. S., Razzera, G., Barreto-Berger, E., Almeida, F. C. & Valente, A. P. Portrayal of complex dynamic properties of sugarcane defensin 5 by NMR: multiple motions associated with membrane interaction. *Structure* **19**, 26–36 (2011).
43. Andreas, L. B. *et al.* Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc. Natl. Acad. Sci. USA* **113**, 9187–9192 (2016).
44. Monteith, W. B. & Pielak, G. J. Residue level quantification of protein stability in living cells. *Proc. Natl. Acad. Sci. USA* **111**, 11335–11340 (2014).
45. Samsa, M. M. *et al.* Dengue virus capsid protein usurps lipid droplets for viral particle formation. *PLoS Pathog.* **5**, e1000632, <https://doi.org/10.1371/journal.ppat.1000632> (2009).
46. Irie, K., Mohan, P. M., Sasaguri, Y., Putnak, R. & Padmanabhan, R. Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene* **75**, 197–211 (1989).
47. Delaglio, F. *et al.* NMRpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
48. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
49. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
50. Carr, H. Y. & Purcell, E. M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **94**, 630–638 (1954).
51. Meiboom, S. & Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
52. Palmer, A. G., Williams, J. & McDermott, A. Nuclear magnetic resonance studies of biopolymer dynamics. *J. Phys. Chem.* **100**, 13293–13310 (1996).
53. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
54. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
55. Martins, A. S. *et al.* Methods for Lipid Droplet Biophysical Characterization in *Flaviviridae* Infections. *Front. Microbiol.* **9**, 1951, <https://doi.org/10.3389/fmicb.2018.01951> (2018).

## Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia – Ministério da Ciência, Tecnologia e Ensino Superior (FCT-MCTES, Portugal) projects PTDC/QUI-BIQ/112929/2009, PTDC/SAU-ENB/117013/2010 and PTDC/BBB-BQB/3494/2014, Calouste Gulbenkian Foundation (FCG, Portugal) project Science Frontiers Research Prize 2010, European Union Marie Skłodowska-Curie Research and Innovation Staff Exchange H2020-MSCA-RISE-2014 project INPACT (Grant 644167), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil, grant numbers 471239/2012-7 and 306669/2013-7) and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ, Brazil, grant numbers E-26/110.636/2012, E-26/110.092/2013 and E-26/201.167/2014). AFF acknowledges FCT-MCTES fellowship SFRH/BD/77609/2011. MS acknowledges FCT-MCTES fellowship PD/BD/128202/2016. ICM acknowledges consecutive funding from the FCT-MCTES fellowship SFRH/BPD/74287/2010 and the Program “Investigador FCT” (IF/00772/2013 Research Contract). This work was also supported by LISBOA01-0145-FEDER-007391 project, cofunded by FEDER, through POR Lisboa 2020 - Programa Operacional Regional de Lisboa, PORTUGAL 2020, and Fundação para a Ciência e a Tecnologia. The NMR spectrometers at FCT NOVA are part of Rede Nacional de RMN (PTNMR), supported by FCT-MCTES (ROTEIRO/0031/2013 - PINFRA/22161/2016) co-funded by FEDER through COMPETE 2020, POCI, and PORL and FCT through PIDDAC.

## Author Contributions

A.F.F., M.S., E.J.C., N.C.S., F.C.L.A. and I.C.M. conceived and designed the experiments; A.F.F., M.S., G.M.B. and F.C.L.A. performed the experiments. A.F.F., M.S., E.J.C., F.C.L.A. and I.C.M. analyzed the data. M.A.R.B.C., A.T.D.P., E.J.C., N.C.S., F.C.L.A. and I.C.M. contributed with reagents/materials/analysis tools. All authors contributed to the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-37599-z>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.