DOI:
10.1016/j.cplett.2018.05.068

Document Version
Peer reviewed version

On the hydration structure of the pro-drug GPG-NH_2 and its derivatives

Paul Smith[a], Nicola Steinke[b], John F. C. Turner[c], Sylvia E. McLain[b,*], Christian D. Lorenz[a,*]

[a]*Biological Physics & Soft Matter Group, Department of Physics, King's College London, London WC2R 2LS, United Kingdom*
[b]*Department of Biochemistry, University of Oxford OX1 3QU, United Kingdom*
[c]*Department of Chemistry, University of Sussex, Brighton BN1 9QJ, United Kingdom*

Abstract

Molecular dynamics simulations were used to investigate the hydration and structure of the tripeptide GPG-NH_2, and the effect of substituting a fluorine or hydroxyl gorup onto one of the C_α positions in the glycinamide portion of the molecule. The fluorinated and hydroxylated peptides both display a slight dehydration of the proline and glycinamide residues and a different conformation of the glycinamide residue backbone than the GPG peptide. These two effects result in a significant decrease in the water-mediated interactions between the Gly1 and glycinamide residues, which had previously been shown to nucleate beta turns in GPG-NH_2.

Keywords: molecular dynamics simulations, peptide hydration, fluorinated peptide, hydroxylated peptide, water-mediated folding

1. Introduction

How a protein folds from a given linear sequence into its three-dimensional biologically active structure in solution has yet to be determined. That this phenomenon occurs in aqueous solutions, while not in other liquids, necessitates that water must contribute in some way to the folding process. There are a variety of theories as to the role of water in the folding process, some which favor the 'hydrophobic effect' [1, 2] and some of which favor the dominance of electrostatic interactions as the driving force behind the process of folding [3, 4]. The hydration of a peptide is linked to the amino acid sequence, yet the details of the hydrophobic/hydrophilic balance within a protein and how this affects the hydration structure (or vice versa) in order to allow for folding remains unclear.

For many proteins, slight changes in their amino acid sequence leads to a collapse of protein structure and thereby function, and can lead to a variety of diseases [5, 6]. For instance, in the protein haemoglobin, misfolding is caused by the mutation of a single amino acid from glutamic

*Corresponding author
Email addresses: sylvia.mclain@bioch.ox.ac.uk (Sylvia E. McLain), chris.lorenz@kcl.ac.uk (Christian D. Lorenz)
URL: http://www2.bioch.ox.ac.uk/mclaingroup/ (Sylvia E. McLain), https://nms.kcl.ac.uk/lorenz.lab/wp/ (Christian D. Lorenz)

acid to valine [7] that leads to the collapse of the protein structure in sickle-cell anemia. Cystic fibrosis is another disease that results from only a slight change to the structure of a protein, as the deletion of a single phenyalanine from the CFTR protien leads to an inability for the proper cytosis of Cl^- ions [8]. How these minor changes lead to a disruption of the native structure and subsequent loss of function of the protein in solution is not well understood, particularly with respect to how the altered hydration changes the folding process.

Previous investigations on small peptides in solution have yielded insights into the role water might have on the folding process [4, 9], suggesting it may act as an aide in the formation of β-turns in the peptide GPG (glycyl-L-prolyl-glycine-NH2) [4]. It is of great interest to determine how a small chemical change to this structure might lead to the disruption of the macromolecular protein structure in solution. In this study, the structure of GPG-NH_2 in aqueous solution has been simulated using molecular dynamics (MD). Previous MD simulations on this peptide in more concentrated solutions have been shown to relatively accurately reproduce the hydration structure, as measured by neutron diffraction[4, 10]. In addition, two derivatives of GPG-NH_2 were also simulated using MD. Specifically, GPG-NH_3^+-F and GPG-NH_3^+-OH, where a -F atom and an -OH group have been substituted for one of the H-C_α positions on the glycinamide portion of the GPG-NH_2 molecule (Fig. 1).

Of the two modifications that are investigated, fluorination of peptides has been studied much more thoroughly in the past. The addition of fluorinated amino acids to peptides and proteins has been shown to confer stability to their structures [11–21]. A particular focus has been placed on the increased stability of coiled-coil and helix-bundle proteins when using fluorinated proteins [15–21], which has been attributed to the hyper-hydrophobic and fluorophilic character of fluorinated amino acid side chains. While there has been a significant amount of work on the effect of fluorinated amino acids on the structure, stability and dynamics of a protein, little work has been done in investigating the effect that fluorinating amino acids has on the hydration of a peptide/protein. Kwon *et al.* investigated the hydration dynamics at fluorinated protein surfaces with time-dependent fluorescence Stokes shifts, and their results show that the hydrating water molecules move slower when around fluorinated residues than around non-fluorinated residues. It was suggested that these slow water dynamics were caused by the strong dipole induced at the protein surface by the introduction of the fluorinated residues [22].

The majority of the research on the effect of hydroxylation of peptides is focused on the hydration and structure of two common extracellular matrix proteins: collagen and elastin. In collagen, post-translational modifications result in the formation of 3− and 4−hydroxproline (HyP) as well as 5−hydroxylysine (HyK), which are thought to each play key structural roles in stabilizing the triple-helix structure of collagen [23–30]. In elastin, the function of the HyP has yet to be identified. However, the hydroxylation of prolines within elastin has been shown to increase the temperature at which coacervation of elastin fibres occurs [31, 32] and the presence of HyP has also been shown to significantly change the supramolecular structures formed by elastin proteins [31]. It is presumed that the presence of HyP in the polypeptide chains of elastin alters the clathrate formation of water molecules around the hydrophobic residues, by enhancing the hydrophilic hydration which results in less aligned water molecules around the hydrophobic residues within the peptide. This modification of the hydration around proteins results in a reduced propensity of coacervation of the proteins and results in a higher temperature being required to induce coacervation because more energy is required to disrupt the hydrophilic hydration shell [32]. This disruption of the calthrate-like water shells around the hydrophobic residues also seems to play a role in the more polymorphic supramolecular assemblies observed in elastin proteins with HyP as the fibrils seem to be stabilized by the long-range order of clathrate-like

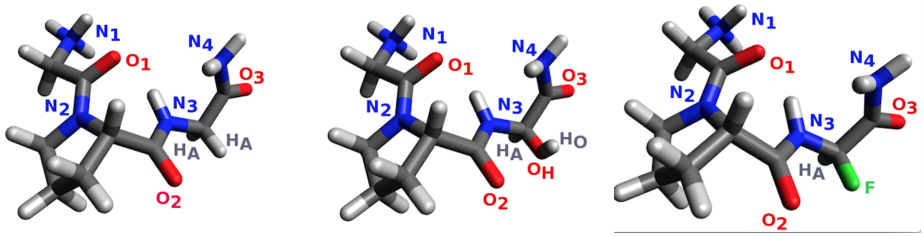


Figure 1: GPG-NH_2 (left) and its derivatives, GPG-NH_2-OH (middle) and GPG-NH_2-F (right).

water molecules [31].

In this manuscript, the effect of the fluorination and hydroxylation of the 3GLY residue in the GPG-NH_2 peptide on the structure and hydration of the peptide has been investigated in detail. The results show that the hydration and peptide structure are quite similar for the fluorinated and hydroxlated peptides, and highlights some interesting differences in these properties as compared to those of the GPG-NH_2 peptide.

In this study, we used fixed point charge classical forcefields (TIP3P [33] for water and CHARMM [34, 35] for the peptides and ions) to investigate these systems, which we have used in several previous studies [4, 9, 10, 36] and shown good agreement with experimental results. However, the continual development of polarisable forcefields (e.g. [37–40]) provides further ability to investigate these systems using forcefields which may capture physics that is not captured in these simulations. We have used the AMOEBA polarisable forcefield [39] to investigate the hydration properties of highly concentrated solutions of proline and seen that it provides a better representation of the experimental data than classical fixed charged forcefields [41].

2. Methods

Molecular dynamics simulations

Molecular dynamics simulations have been performed on GPG-NH_2 (GPG) in aqueous solution as well as for two analogs of the GPG-NH_2 molecule, GPG-NH_2-OH (GPG-OH) and GPG-NH_2-F (GPG-F) which are shown in Fig. 1. Each system contains 125 GPG/GPG-OH/GPG-F molecules, 125 Cl^- counter ions, and 7250 water molecules. The GPG molecules, their derivatives and the Cl^- ions were modelled using the CHARMM forcefield [34, 35], which is the same forcefield that we have previously used to investigate GPG and GPG-containing peptides and their interactions with water, ions and urea [4, 9, 10, 36]. The water molecules were modeled using TIP3P [33] modified for the CHARMM forcefield and all of the bonds and angles of the water molecules were constrained using the SHAKE algorithm.

All of the simulations were conducted using GROMACS 4 [42]. For each system, position restraints were placed on the backbone of the peptides in the simulated systems while an energy minimisation simulation was conducted in order to eliminate any clashes between atoms that resulted from the building of the initial configurations without artifically disturbing the initial structure of the peptides. Then the temperature was equilibrated by using the NVT ensemble to run a 2 ns simulation at 300 K. Finally, the pressure was allowed to equilibrate within the system by carrying out a simulation using the NPT ensemble at a temperature of 300 K and a pressure of 1 atm, which is 2 ns in duration. Finally, the production simulations were carried out using

the NPT ensemble, where the temperature was 300 K and the pressure was 1 atm, which were run for 40 ns using a 2 fs timestep. In all simulations, the temperature was controlled using the Nose-Hoover thermostat [43, 44] and the pressure was controlled using the Martyna-Tuckerman-Tobias-Klein (MTTK) barostat [45]. The van der Waals interaction cutoff was at 14 Å, while the particle mesh Ewald (PME) algorithm [46, 47] was used to compute the long-range Coulomb interactions.

ANGULA

The ANGULA analysis suite was used for analyzing the dihedral angles, and identifying the different representative conformations of the peptides in solution as well as the whole molecule analysis spatial density maps (WMAs), which give the most probable location of water molecules around in the peptide molecules in three dimensional space [9, 48, 49]. To generate the WMAs, the peptide molecules were first separated into the different conformers based on selection of the dihedral angle in question (*vide infra*) and then subsequently the first nearest neighbor water molecules (within a distance range up to 2.2 Å) to each atomic site were identified for each peptide in the solution. This information was used to create a cloud of most probable density of water molecules around the peptides in question. For each derivative 5000 frames of the trajectory were used in the analysis where each frame contained 125 molecules. A more detailed description of this analysis is proved in the SI.

3. Results

Interaction of water molecules with the peptide backbone

The interaction of water molecules with the peptide backbone has been characterised by first determining the radial distribution functions ($g(r)$; RDFs) for nitrogens and oxygens in the GPG, GPG-OH and GPG-F peptides and the O_w and H_w atoms in the surrounding water molecules. The $g(r)$s for the C-terminal Ncap N_4 atoms and 2PRO-3GLY O_2 and 3GLY-Ncap O_3 are shown in the Supplementary Information, while the $g(r)$s for the other oxygen and nitrogen atoms in the peptide backbone are not shown as there is no appreciable difference for the three different peptides. The coordination numbers for all of these $g(r)$s are also provided in the Supplementary Information.

Fig. 2a shows the probability distribution of a N_4 atom being hydrated by a given number of water molecules, which was generated by determining the number of water molecules whose O_w atom is within r_{max} for the N_4-O_w (see SI) of the N_4 atom of a peptide in each configuration of the trajectory. The probability histograms show that the most probable number of water molecules around the N_4 is 5, 4 and 4 for the GPG, GPG-OH and GPG-F peptides, respectively, signifying that the N_4 in GPG is relatively more hydrated than the GPG-OH and GPG-F peptides. It appears that the different functional groups on the C_α of the 3GLY residue has only a small affect on the location of the water molecules around the N_4 atoms.

The interaction of water molecules with the oxygen atoms in the peptide backbone also play a significant role in the hydration structure of these peptides. The coordination numbers for all of the peptide oxygen $g(r)$s are included in the Supplementary Information. Figures 2b and 2c show the probability distributions of finding a certain number of hydrating water molecules around the O_2 and O_3 atoms within the backbones of the GPG, GPG-OH and GPG-F peptides, respectively.

Both the O_2 and O_3 atoms are dehydrated in the GPG-OH and GPG-F peptides as compared to the GPG peptide. The $g(r)$s in the SI show that for the O_2 atoms the magnitude of the first

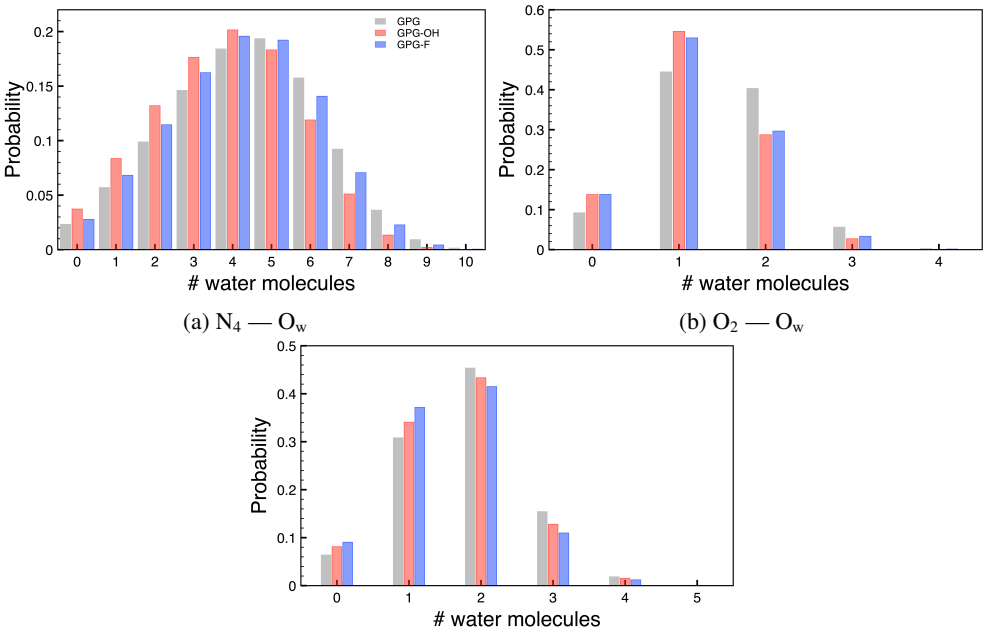


Figure 2: Histograms of the number of hydrating water molecules around the (a) N_4, (b) O_2 and (c) O_3 atoms of the GPG, GPG-OH and GPG-F peptides.

peak for the GPG-OH and GPG-F peptides is smaller than that for the GPG peptide. The RDFs of the water interactions with the O_3 atoms of the different peptides show that once again the first neighbor peak of the O_3-O_w $g(r)$ has a slightly smaller magnitude for the GPG-OH and GPG-F peptides than the GPG peptide. However, the more significant difference is that the corresponding minimum in the $g(r)$ occurs at smaller magnitude in the GPG-OH peptides. The $g(r)$s for the O_3-O_w and the O_3-H_w interactions show that the second neighbor shell in the GPG-OH and GPG-F peptides is significantly dehydrated as compared to the GPG peptides.

The distributions of hydrating waters around the O_2 atoms show that the most probable number of water molecules in the first hydration shell is 1 for all three peptides. However, the probability of having 1 or less water molecules hydrating the O_2 in the GPG-OH and GPG-F peptides is greater than that for the GPG peptides and the probability of 2 or more hydrating water molceules is larger for the GPG peptides than that for the GPG-OH and GPG-F peptides.

Meanwhile, the most probable number of water molecules in the first hydration shell around the O_3 atoms is 2 for all three peptides, as shown in Fig. 2c. However the probability of having 1 or fewer water molecules in the first hydration shell of an O_3 atom in the GPG-OH and GPG-F peptides is greater than that for the GPG peptides. While the GPG peptides are more likely to have 2 or more water molecules in the first hydration shell around an O_3 atom than for GPG-OH and GPG-F peptides.

Interaction of water molecules with 3GLY C_α

The C_α of the 3GLY residue in GPG is the site of modification in the derivative peptides where a -H was replaced by a -OH in the GPG-OH peptide and a -F in the GPG-F peptide. As

a result, it is of interest to understand how the hydration of this group is modified by the change to a functional group. The salient $g(r)$s for these functions are shown in the Supplementary Information and the coordination numbers for these functions are in Table 1. The H_A-O_w RDF for the three GPG derivatives, perhaps unsurprisingly, show significant differences. For the GPG-OH and GPG-F peptides, there are well defined peaks at r = 2.78 Å and 2.72 Å, respectively, which indicate that there are strongly bound water molecules within the first hydration shell. Whereas, the H_A-O_w RDF for the GPG peptide has a first peak which is similar in magnitude to that for GPG-OH and GPG-F but the peak is comparatively more broad signifying a larger radius for the first hydration shell (r = 3.92 Å) than either the GPG-OH (3.68 Å) or GPG-F (3.56 Å) peptides. Further the RDFs for both GPG-OH and GPG-F (see SI) also show more prominent second nearest neighbor peaks for the hydration around this groups. This is suggestive of a radial charge effect where introducing the -OH or -F atom would change the dipole around this portion of the peptide which leads to a closer interaction with the water molecules hydrating 3GLY and a more structured second coordination sphere.

In Table 1, the coordination numbers of O_w around H_A for the GPG-OH and the GPG-F peptides are shown to be significantly lower than for the GPG peptide. The $g(r)$s of the interaction between the O_w atoms in water molecules and the O_H and H_O atoms in the GPG-OH and the F atoms in the GPG-F peptides are shown in the SI. In all cases, there is a first neighbor peak indicating that the water molecules form an ordered hydration shell around the -OH and -F groups in the GPG-OH and GPG-F peptides, respectively. From the values of the coordination numbers in Table 1, the GPG-OH and GPG-F peptides have approximately 4 more water molecules in their hydration shell around the 3GLY C_α than the GPG peptides, as both the H_O and F atoms have coordination numbers of approximately 8 water molecules, whereas the coordination number of water around a H_A atom is approximately 4.

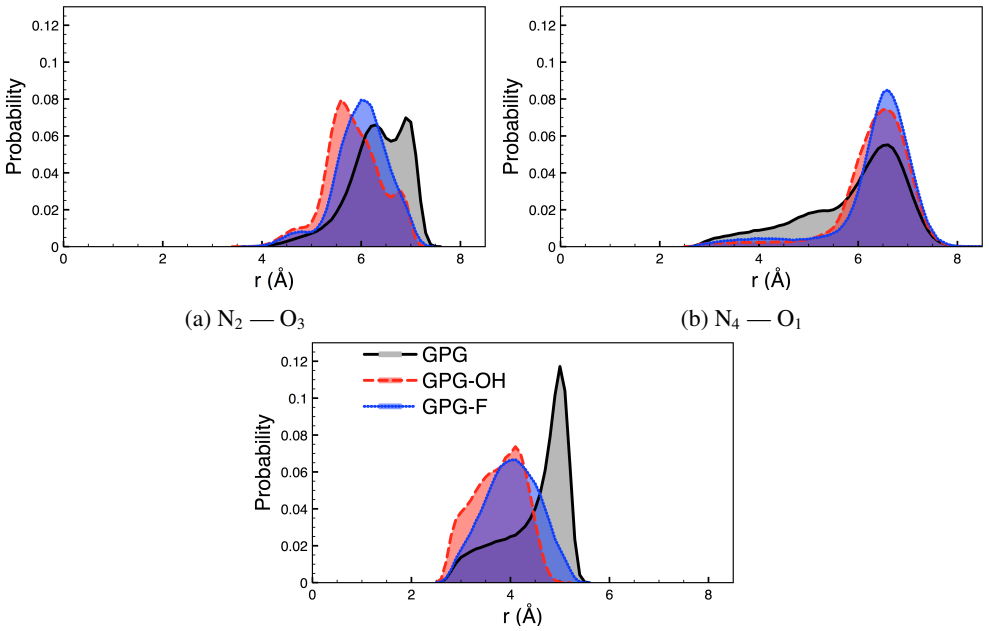
The second nearest neighbor peaks observed in the RDFs for the H_A-O_w interaction are also different for the three different GPG derivatives. While the locations of the peak and corresponding minimum occur at nearly identical values of r, the magnitudes of the peaks for the GPG-OH and GPG-F peptides are nearly equal and both are larger than that for the GPG peptides.

Table 1: Coordination numbers n_i^j and nearest neighbor distances r_{max} (Å) of the solvent atoms around the atoms attached to the C_α of the 3GLY residue in the GPG, GPG-OH and GPG-F peptides.

		GPG		GPG-OH		GPG-F	
Type i	Type j	n_i^j	r_{max}	n_i^j	r_{max}	n_i^j	r_{max}
H_A	O_w	4.0	3.92	2.7	3.68	2.4	3.56
O_H	O_w	—	—	10.0	4.92	—	—
H_O	O_w	—	—	8.6	4.66	—	—
F	O_w	—	—	—	—	8.1	4.64

Intrapeptide interactions

From the RDFs and the coordination numbers, three different parts of the peptide backbone have showed significantly different hydration when comparing GPG, GPG-OH and GPG-F. These differences may be the result of changes to the intra-peptide structure within the backbone of a given peptide. In order to investigate this, the distances between the various N_x and O_x atoms on the peptides were calculated for each peptide in each configuration of the production trajectories and the probability distributions of a given N and O atom being a certain distance



(a) N_2 — O_3 (b) N_4 — O_1

(c) N_4 — O_2

Figure 3: Probability distributions of the (a) N_2 and O_3, (b) N_4 and O_1 and (c) N_4 and O_2 atoms on the same peptide being within a given distance from one another.

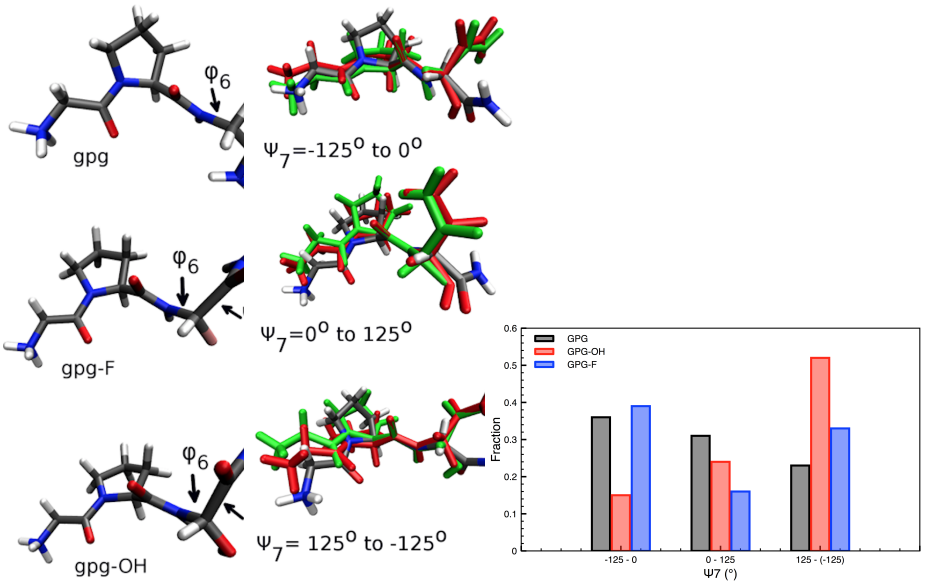
apart were generated for the GPG, GPG-OH and GPG-F peptides. The three interactions that were significantly different (N_2-O_3, N_4-O_1 & N_4-O_2) include the three atoms which showed significant differences in the hydration properties discussed previously (O_2, O_3 & N_4), and have been plotted in Figure 3.

The coordination numbers of water molecules around the O_2, O_3 and N_4 atoms (which are shown in SI) show a slight yet significant dehydration of these atoms in the GPG-OH and GPG-F peptides when compared to the GPG peptides. The distribution for the N_2-O_3 interactions in the GPG-OH peptides displayed in Fig. 3 shows a shoulder from r = 4 Å to 4.95 Å and then the distribution increases until it reaches a peak at r = 5.6 Å, which represents the most probable distance between the two atoms, as well as a further peak which is significantly smaller in magnitude at r = 6.8 Å. The N_2-O_3 distribution for the GPG-F peptides shows a similar shoulder as observed with the GPG-OH peptides and a similar in the distribution until reaching a peak at r = 6 Å, and the GPG-F distribution decreases to zero without going through any other maximums. Conversely, the distribution for the GPG peptides is significantly different, it does not have the shoulder at small distances and instead constantly increases until reaching a first peak at r = 6.3 Å and a second, slightly larger peak at r = 6.9 Å. From these distributions, it appears that the shorter range interactions between N_2 and O_3, which are more probable in the GPG-OH and GPG-F peptides than in the GPG peptides, are correlated with the dehydration of the O_3 atoms in the peptide backbone in the two derivatives of GPG.

The distributions of the distances between the N_4 and O_1 atoms in the peptide backbone show that the most probable distance (r = 6.6 Å) is the same for the GPG, GPG-OH and GPG-F

peptides. The only significant difference observed in these distributions is that the GPG peptides are most likely to be found in configurations such that the distance between the N_4 and O_1 is less than 5.8 Å. However, as the most probable distance is the same for the three peptides and there is no significant difference in the hydration of the O_1 atoms in the three different peptides, it appears that this interaction does not contribute to the dehydration of the N_4 atoms in the GPG-OH and GPG-F peptides.

The interactions within the GPG-OH and GPG-F peptides results in the distribution of distances between the N_4 and O_2 atoms in the peptide backbones being shifted to smaller distances as compared to the GPG peptides. The most probable configurations of the GPG-OH and GPG-F peptides are such that the distance between the N_4 and O_2 atoms is $\sim$ 4.1 Å while within the GPG peptides it is $\sim$ 5 Å. There is a slight difference in the distribution of N_4-O_2 distances for the GPG-OH and GPG-F peptides, in that the GPG-OH peptides are more likely to have N_4-O_2 distances of less than $\sim$ 4.3 Å. N_4 and O_2 are both dehydrated in the derivatives, unlike the N_2-O_3 and N_4-O_1 distances, where only one of the atoms is dehydrated upon OH/F substitution. This might explain why the change in intramolecular distance is most pronounced here.



(a) (b) (c)

Figure 4: Peptide backbone dihedrals. (a) Depiction of the two dihedrals of interest, ϕ_6 and ψ_7, in the GPG, GPG-F and GPG-OH peptides. (b) Representative structures of the GPG (grey & blue), GPG-F (green) and GPG-OH (red) at the different ranges of values of the ψ_7 dihedral. (c) Probability distribution of the ψ_7 dihedral being within a certain range of values for the three different GPG peptides.

The dihedral angles along the GPG, GPG-OH and GPG-F peptide backbones have also been calculated. There are only two backbone dihedral angles that show any significant difference between the three peptides, ϕ_6 and ψ_7, which are defined in Fig. 4a. The distributions of these two dihedrals that were observed during the simulations is shown in the SI. In the case of ϕ_6, the distribution for the GPG peptide show a broad range in values from $-180°$ to $-45°$ and then also from $45°$ to $180°$, with three peaks within the distribution, the largest at $180°$ ($-180°$) and

two smaller peaks at $-90°$ and $90°$. The distributions of ϕ_6 for the GPG-F and GPG-OH peptides show a much more narrow range from $-175°$ to $-45°$ for GPG-F and from $-105°$ to $-45°$ for GPG-OH, where each has a peak at approximately $-85°$.

The distributions of the ψ_7 dihedral are unique for each of the three different peptides. This ψ_7-distribution for the GPG peptide shows three significant peaks, where one is found at $\psi_7 \sim -75°$, another at $\sim 55°$ and the last at $\sim 180°$. Meanwhile, the distribution of the ψ_7 angles for the GPG-OH peptides shows peaks at almost identical values of the dihedral, with largest peak at $\sim 180°$, and the distribution for the GPG-F peptides shows peaks only at $\psi_7 \sim 60°$ and $\sim 180°$.

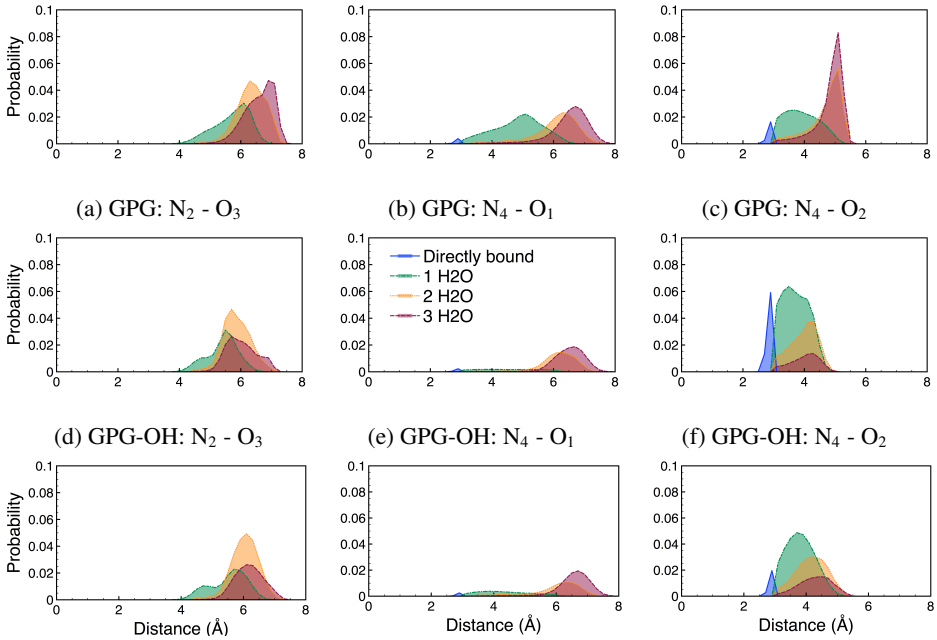
In order to quantify the different conformations of these peptides, we have divided the conformational space into three different ranges of ψ_7 which are each centered around one of the peaks observed: $\psi_7 = -125°$ - $0°$; $0°$ - $125°$; and $-125°$ - $125°$. In Fig. 4b, conformations of the GPG, GPG-F and GPG-OH peptides with values of ψ_7 which represent the minimum, maximum and average values of the distributions of this dihedral for each peptide (as shown in the SI) are shown. These conformations show that in each range of ψ_7 the GPG-OH and GPG-F peptides have quite similar structures, while the GPG structure is quite similar until reaching the location of the ϕ_6 dihedral along the backbone chain which results in the NH_2 (N_4) capping group on the C-terminal end of the peptide being rotated into a different region of space than the other two peptides. This is because the ϕ_6 angle is $\sim 180°$ for GPG and $\sim -85°$ for the derivatives. As a result of this difference, the GPG-OH and GPG-F peptides take conformations where the N_4 atom is near the O_2 atom in the peptide backbones when $\phi_7 = -125°$ - $0°$ and $\phi_7 = 0°$ - $125°$. A principal component analysis of these peptides has also shown that the GPG-F and GPG-OH show two distinct clusters of configurations, which are quite similar (see ESI). Meanwhile, when $\phi_7 = 125°$ - $-125°$, all three peptides take significantly extended conformations. In Fig. 4c, the probability distributions for each of the three peptides existing in a conformation with a certain ψ_7 angle are reported. From this plot, it is observed that the GPG and GPG-F peptides are most commonly found in conformations with ψ_7 between $-125°$ and $0°$, although as discussed already this does mean they have very similar structures. The GPG-OH peptides are most commonly found with ψ_7 between $125°$ and $-125°$.

Therefore, the stronger interactions between N_4 and O_2 observed in both the intramolecular distance distributions and the distributions of the ϕ_6 and ψ_7 dihedral angles in the GPG-OH and GPG-F peptides seemingly result in the dehydration of the two atoms in the peptide backbone (as shown in Fig. 2) as compared to that found in the GPG peptides.

Effect of water-mediated interactions

Water-mediated interactions within peptides can be measured from the trajectories of the MD simulations. The first neighbor distances of water molecules with the various oxygen and nitrogen atoms found from the $g(r)$s (shown in the SI) are used to determine the water molecules that are bound to each atom on the peptide. Then the first neigbor distance between a pair of water molecules is found from the radial distribution function of water molecules. Then in a given configuration, two atoms on the peptide are said to have a water-mediated interaction via a single water molecule if the same water molecule is found to be bound to each of the two atoms. Likewise, if water molecule x is bound to one atom i on the peptide and another water molecule y is bound to another atom j on the peptide, and the two water molecules x and y are first neighbors to one another then atoms i and j of the peptide are said to have water-mediated interactions via a two water molecule chain. Finally, a similar search is done for chains of three water molecules.

Figure 5 shows the probability that the N_2 & O_3, N_4 & O_1, and N_4 & O_2 atoms within the same peptide are either directly bonded, or interacting via a hydrogen bonded chain of 1, 2 or



(a) GPG: N_2 - O_3 (b) GPG: N_4 - O_1 (c) GPG: N_4 - O_2

(d) GPG-OH: N_2 - O_3 (e) GPG-OH: N_4 - O_1 (f) GPG-OH: N_4 - O_2

(g) GPG-F: N_2 - O_3 (h) GPG-F: N_4 - O_1 (i) GPG-F: N_4 - O_2

Figure 5: Probability distributions of the (a) N_2 and O_3, (b) N_4 and O_1 and (c) N_4 and O_2 atoms on the same peptide being directly bound (blue), or having 1 (green), 2 (orange) or 3 (maroon) water molecules form a hydrogen bonded chain between the two atoms.

3 water molecules as a function of distance between the two atoms. The distributions for the interactions between the N_2 & O_3 atoms on the GPG, GPG-OH and GPG-F (Figs. 5a, 5d, 5g, respectively) are very similar.

In the case of the interactions between the N_4 & O_1 atoms, the most significant difference when comparing the behavior of the GPG peptides (Fig. 5b) with that of the GPG-OH (Fig. 5e) and GPG-F (Fig. 5h) peptides is that there is a significant population of peptides in which the two atoms are seperated by 3 - 6 Å, whose interactions are mediated by a single water molecule. While, in the GPG-OH and GPG-F peptides, there are almost no peptides in which the configurations are mediated by a single water molecule. There are small populations of peptides for all three of the derivatives which are directly bound, and larger populations which are bound by two and three water molecules within the same distance ranges. However, in all cases the water-mediated populations of the GPG-OH and GPG-F peptides are smaller than in the GPG peptides.

When considering the interactions between the N_4 & O_2 atoms on the three peptides, there is a significant difference in the nature of the interactions observed. While 1.9% and 2.2% of the GPG (Fig. 5c) and GPG-F (Fig. 5i) peptides, respectively, are configured such that the two atoms are directly bound to one another, the GPG-OH peptides (Fig. 5f) have significantly larger populations of their configurations in which the two atoms are directly bound (7.3%). The GPG-OH and GPG-F peptides have significantly larger populations of their configurations (40.1% & 31.0%, respectively) in which the N_4 - O_2 interactions are mediated by a single water molecule than are found within the configurations of the GPG peptide (20.1%). Meanwhile the GPG peptide shows that 45% of the peptide configurations are mediated by 2 or 3 water molecules at distances larger than 4.5 Å, whereas for the GPG-F and GPG-OH peptides only 15% and 8%, respectively, are found to have such large N_4 - O_2 distances. Therefore, the close (directly bound and single water-mediated) interactions of the N_4 and O_2 atoms in the GPG-OH and GPG-F peptides appear to lead to the relative dehydration of the N_4 and O_2 atoms that is observed in the hydration numbers of these two atoms within the peptide as compared to that for the GPG peptides.

While the investigation of the role that water plays in the interactions that seemingly differentiate the structures that the three peptides take in solution provides useful insight, it is also interesting to see how the overall hydration of the peptides are affected by such slight changes in their chemistry. Therefore, spatial distribution functions for the whole molecule (WMA[48]) have been calculated around each atom of the peptide and these WMAs have been used to provide a three dimensional image of where the nearest (and therefore most tightly bound) water molecules between 0 and 2.2 Å are in relation to the various atoms within the peptide (Fig. 6). In Fig. 6a, the WMAs have been plotted for each of the three peptides when ψ_7 is between 125° and −125°. From these three figures, it is clear that the water molecules which are tightly bound are found in the same relative positions in each of the three peptides, with the only significant difference being observed in the size of the population of water molecules around the N_3 atom in which there seems to be similar amounts in the GPG-F and GPG-OH peptides and significantly less in the GPG peptide. Fig. 6b once again shows the hydration of the three different peptides, but this time a set of characteristic structures of GPG with ψ_7 between 0° and 125° are shown, while the structures for GPG-F and GPG-OH are when ψ_7 is between −125° and 0°. These figures show very similar hydration patterns for the three different peptides, where there is significant hydration around the NH_3 and NH_2 groups on the terminal ends of the peptides and then a large population of water molecules which seemingly could mediate the interaction between N_2 and N_3.

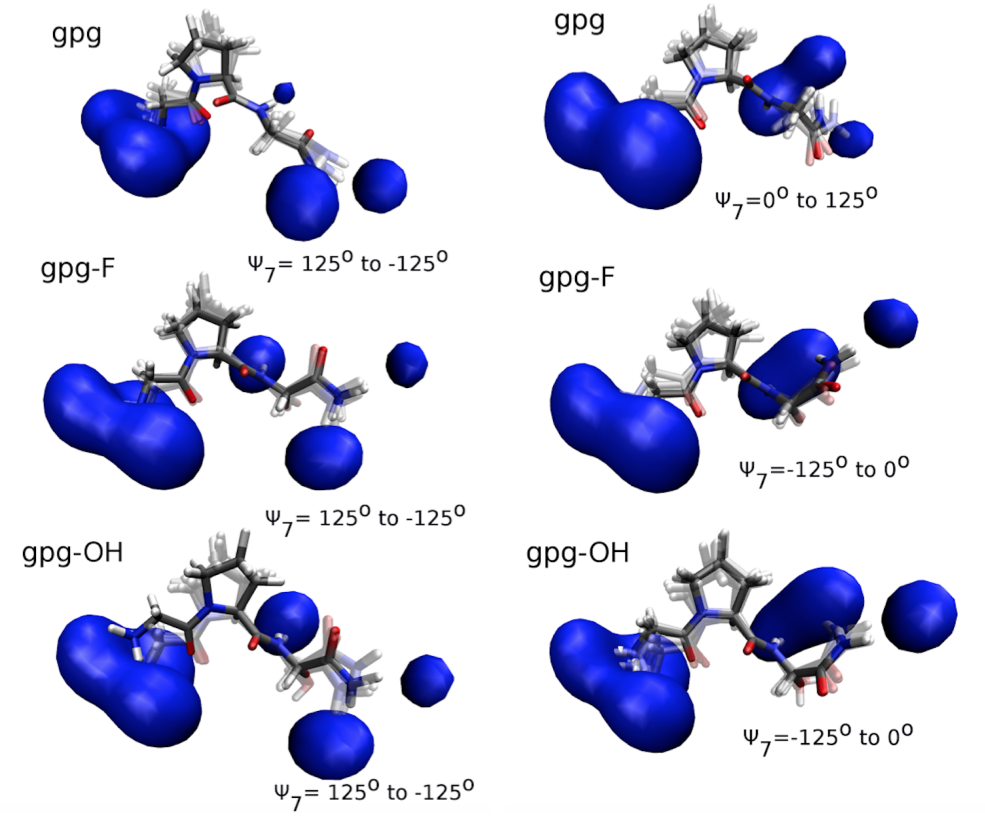


Figure 6: Whole molecule analysis for water between 0-2.2 Å around the peptides (a) where ψ_7 is between 125°and −125°and (b) where the hydration patterns are similar but the value of ψ_7 is different. The isocountour levels shown in each figure represent the top 40% of the surrounding water molecules.

4. Discussion & Conclusions

The analysis that has been presented in this manuscript shows that the average conformation and hydration of the GPG-F and GPG-OH peptides are different from that of the GPG peptide. However, the conformation and hydration structures observed in the GPG-F and GPG-OH peptides are quite similar to one another. Interestingly, while most of the dihedrals along the backbone of the GPG molecules remain unchanged, the dihedrals around 3GLY are significantly different in GPG-F and GPG-OH compared with GPG, and lead to different intramolecular structures for these two derivatives compared with GPG. This likely results in a change of the electronic structure around the C-terminus of the peptide, which leads to a change in not only the hydration, but also the mechanism of conformational change and possibly folding nucleation in the GPG derivatives. Both the location and distribution of water molecules around these tripeptides are affected by the modifications to the 3GLY residue. The oxygen atoms in the peptide backbone of both 2PRO and 3GLY as well as the nitrogen in the -NH_2 cap on the C-termini are somewhat dehydrated for GPG-F and GPG-OH as compared to that observed in GPG, where the degree of dehydration around these three atoms is quite similar for GPG-F and GPG-OH. Further, the dehydration that is observed around these appears to encourage the direct and water-mediated intramolecular binding between the peptide bond oxygen in the 2PRO residue and the N_4 in the NH_2 cap for the GPG-F and GPG-OH, which is not nearly as pronounced in GPG.

This change in conformation and hydration of the two derivatives suggests a change in their possible folding initiation mechanisms compared with GPG. The similarity between the results for GPG-F and GPG-OH suggest a similar folding mechanism via two different routes. In the case of the GPG-OH, the backbone oxygen in the 3GLY residue (O_3) is dehydrated by the -OH group on the C_α which is likely the result of a competition between these to groups for nearest neighbor water molecules. This possibly instigates a fold within the peptide which occurs such that the backbone oxygen in the 2PRO residue (O_2) and the nitrogen atom in the NH_2 cap (N_4) are placed within closer proximity of one another. Whereas in GPG-F, the polarity of F, perhaps, acts to reduce the electronegativity of the backbone oxygen in the 3GLY residue (O_3) atom, which therefore reduces its ability to form hydrogen bonds with neighboring water molecules, leading to a reduction in nearest neighbor hydration around it. Subsequently, the GPG-F molecule would be able to fold in a similar way to GPG-OH in solution.

The water-mediated interactions between the oxygen (O_1) in the 1GLY and the nitrogen atom in the NH_2 capping group (N_4) that are observed in the GPG-NH_2 peptide, and previously have been suggested to be a mechanism by which β-turns nucleate[4], are disrupted by the modifications made in the GPG-F and GPG-OH derivatives. This disruption serves to shunt the folding frame for the peptides and induce a comparatively tighter turn in the GPG derivatives, namely between O_2 and the N_4 atoms where the surrounding water molecules still mediate this interaction. This suggests that it is the interplay between both the electron density distribution of the peptides and its hydrophobic/hydrophilic balance and the surrounding water solvent which nucleates the folding process. The results here suggest that it is the charge distribution of the peptide that guides water into the appropriate location to initiate the folding process and that very minor disruptions to the electron density structure of peptides, can result in large changes to a folding site. These observations also provide further support for the assumption that the slower hydration dynamics observed near fluorinated residues within proteins will change the nature of the water-mediated interactions within the protein [22]. Likewise, the hydroxylation of elastin and elastin-like peptides has shown that there is a reduction of coacervation and the aggregates that do form have a reduction in fibril structure, which is probably due to difference in the struc-

ture that is taken by the individual peptides [31, 32]. Therefore this work provides further insight into how modifying a peptide surface can change its hydration and therefore structure, which is a key step in exploiting these type of modifications of peptides/proteins in biological systems.

5. Acknowledgements

The MD simulations were performed via our membership of the U. K.'s HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202); this work used the ARCHER U. K. National Supercomputing Service (http:// www.archer.ac.uk). Additionally, C. D. L. and P. S. thank the Faculty of Natural and Mathematical Sciences and the Department of Physics at King's College London for funding summer internships for P. S. S. E. M. and N. S. thank the ISIS Facility (Rutherford Appleton Laboratories, STFC, U.K.) for the allocation of neutron beam time and the U.K. Engineering and Physical Sciences Research Council (EP/ J002615/1) and the Leverhulme Trust (RPG-2015-135) for funding.

References

[1] W. Kauzmann, Some factors in the interpretation of protein denaturation11the preparation of this article has been assisted by a grant from the national science foundation., Adv. Protein Chem. 14 (Supplement C) (1959) 1–63. doi:https://doi.org/10.1016/S0065-3233(08)60608-7.
URL http://www.sciencedirect.com/science/article/pii/S0065323308606087

[2] D. Chandler, Interfaces and the driving force of hydrophobic assembly, Nature 437 (7059) (2005) 640–647.

[3] S. R. Durell, A. Ben-Naim, Hydrophobic-hydrophilic forces in protein folding, Biopolymers 107 (8) (2017) e23020–n/a, e23020. doi:10.1002/bip.23020.
URL http://dx.doi.org/10.1002/bip.23020

[4] S. Busch, C. D. Bruce, C. Redfield, C. D. Lorenz, S. E. McLain, Water mediation is essential to nucleation of β-turn formation in peptide folding motifs, Angew. Chem. Int. Ed. 52 (49) (2013) 13091–13095.

[5] J. S. Valastyan, S. Lindquist, Mechanisms of protein-folding diseases at a glance, Disease Models & Mechanisms 7 (1) (2014) 9–14. arXiv:http://dmm.biologists.org/content/7/1/9.full.pdf, doi:10.1242/dmm.013474.
URL http://dmm.biologists.org/content/7/1/9

[6] C. L. McCafferty, Y. V. Sergeev, In silico mapping of protein unfolding mutations for inherited disease, Scientific Reports 6.

[7] A. Cole-Strauss, K. Yoon, Y. Xiang, B. C. Byrne, M. C. Rice, J. Gryn, W. K. Holloman, E. B. Kmiec, et al., Correction of the mutation responsible for sickle cell anemia by an rna-dna oligonucleotide, Science (1996) 1386–1388.

[8] B. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox, A. Chakravarti, M. Buchwald, L.-C. Tsui, Identification of the cystic fibrosis gene: genetic analysis, Trends in Genetics 5 (1989) 363–363.

[9] N. Steinke, R. J. Gillams, L. C. Pardo, C. D. Lorenz, S. E. McLain, Atomic scale insights into urea-peptide interactions in solution, Phys. Chem. Chem. Phys. 18 (2016) 3862–3870. doi:10.1039/C5CP06646H.

[10] S. Busch, L. C. Pardo, W. B. O'Dell, C. D. Bruce, C. D. Lorenz, S. E. McLain, On the structure of water and chloride ion interactions with a peptide backbone in solution, Phys. Chem. Chem. Phys. 15 (2013) 21023–21033. doi:10.1039/C3CP53831A.
URL http://dx.doi.org/10.1039/C3CP53831A

[11] P. J. Baker, J. K. Montclare, Enhanced refoldability and thermoactivity of fluorinated phosphotriesterase, ChemBioChem 12 (12) (2011) 1845–1848.

[12] C. Yuvienco, H. T. More, J. S. Haghpanah, R. S. Yu, J. K. Montclare, Modulating supramolecular assemblies and mechanical properties of engineered protein materials by fluorinated amino acids, Biomacromolecules 13 (8) (2012) 2273–2278.

[13] B. C. Buer, R. de la Salud-Bea, H. M. Al Hashimi, E. N. G. Marsh, Engineering protein stability and specificity using fluorous amino acids: The importance of packing effects, Biochemistry 48 (45) (2009) 10810–10817.

[14] H. T. More, N. Zhang, Kevin S. Srivastava, J. A. Frezzo, J. K. Montclare, Influence of fluorination on protein-engineered coiled-coil fibrils, Biomacromolecules 16 (2015) 1210–1217.

[15] C. J *a*ckel, M. Salwiczek, B. Koksch, Fluorine in a native protein environmenthow the spatial demand and polarity of fluoroalkyl groups affect protein folding, Angewandte Chemie International Edition 45 (25) (2006) 4198–4203. doi:10.1002/anie.200504387.
URL http://dx.doi.org/10.1002/anie.200504387

[16] Y. Tang, D. A. Tirrell, Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host, Journal of the American Chemical Society 123 (44) (2001) 11089–11090, pMID: 11686725. arXiv:http://dx.doi.org/10.1021/ja016652k, doi:10.1021/ja016652k.
URL http://dx.doi.org/10.1021/ja016652k

[17] J. K. Montclare, S. Son, G. A. Clark, K. Kumar, D. A. Tirrell, Biosynthesis and stability of coiled-coil peptides containing (2*S*,4*R*)-5,5,5-trifluoroleucine and (2*S*,4*S*)-5,5,5-trifluoroleucine, ChemBioChem 10 (1) (2009) 84–86.

[18] B. Bilgiçer, A. FIchera, K. Kumar, A coiled coil with a fluorous core, J. Am. Chem. Soc. 123 (19) (2001) 4393–4399.

[19] Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado, D. A. Tirrell, Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability, Angewandte Chemie International Edition 40 (8) (2001) 1494–1496. doi:10.1002/1521-3773(20010417)40:8¡1494::AID-ANIE1494¿3.0.CO;2-X.
URL http://dx.doi.org/10.1002/1521-3773(20010417)40:8<1494::AID-ANIE1494>3.0.CO;2-X

[20] S. Son, I. C. Tanrikulu, D. A. Tirrell, Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues, ChemBioChem 7 (2006) 1251–1257.

[21] M. Salwiczek, S. Samsonov, T. Vagt, E. Nyakatura, E. Fleige, J. Numata, H. C *o*lfen, M. Pisabarro, B. Koksch, Position-dependent effects of fluorinated amino acids on the hydrophobic core formation of a heterodimeric coiled coil, Chemistry A European Journal 15 (31) (2009) 7628–7636. doi:10.1002/chem.200802136.
URL http://dx.doi.org/10.1002/chem.200802136

[22] O.-H. Kwon, T. H. Yoo, C. M. Othon, J. A. Van Deventer, D. A. Tirrell, A. H. Zewail, Hydration dynamics at fluorinated protein surfaces, Proceedings of the National Academy of Sciences 107 (40) (2010) 17101–17106. arXiv:http://www.pnas.org/content/107/40/17101.full.pdf, doi:10.1073/pnas.1011569107.
URL http://www.pnas.org/content/107/40/17101.abstract

[23] K. M. Reiser, R. J. McCormick, R. B. R *u*cker, Enzymatic and nonenzymatic cross-linking of collagen and elastin, FASEB J. 6 (1992) 2439–2449.

[24] J. Bella, B. Brodsky, H. M. Berman, Hydration structure of a collagen peptide, Structure 3 (9) (1995) 893–906.

[25] K. Takaluoma, J. Lantto, J. Myllyharju, Lysyl hydroxylase 2 is a specific telopeptide hydroxylase, while all three isoenzymes hydroxylate collagenous sequences, Matrix Biol. 26 (2007) 396–403.

[26] M. A. Weis, D. M. Hudson, L. Kim, M. Scott, J.-J. Wu, D. R. Eyre, Location of 3-hydroxyproline residues in collagen types i, ii, iii and v/xi implies a role in fibril supramolecular assembly, J. Biol. Chem. 285 (2010) 2580–2590.

[27] R. A. Berg, D. J. Prockop, Thermal transition of a non-hydroxylated form of collagen – evidence for a role for hydroxyproline in stabilizing triple-helix of collagen, Biochem. Biophys. Res. Commun. 52 (1973) 115–120.

[28] K. L. Gorres, R. T. Raines, Prolyl 4-hydroxylase, Crit. Rev. Biochem. Mol. Biol. 45 (2010) 106–124.

[29] K. Terao, K. Mizuno, M. Murashima, Y. Kita, C. Hongo, K. Okuyama, T. Norisuye, H. P. B *a*chinger, Chain dimensions and hydration behavior of collagen model peptides in aqueous solution: [glycyl-4(*R*)-hydroxyprolyl-4(*R*)-hydroxyproline]$_n$, [glycylprolyl-4(*R*)-hydroxyproline]$_n$, and some related model peptides, Macromolecules 41 (2008) 7203–7210.

[30] L. Vitagliano, R. Berisio, L. Mazzarella, A. Zagari, Structural bases of collagen stabilization induced by proline hydroxylation, Biopolymers 58 (2001) 459–464.

[31] B. Bochicchio, A. Laurita, A. Heinz, C. E. H. Schmelzer, A. Pepe, Investigating the role of (2*S*,4*R*-4-hydroxyproline in elastin model peptides, Biomacromolecules 14 (2013) 4278–4288.

[32] J. Dandurand, V. Samouillan, C. Lacabanne, A. Pepe, B. Bochicchio, Water structure and elastin-like peptide aggregation, J. Therm. Anal. Calorim. 120 (2015) 419–426.

[33] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (2) (1983) 926–935.

[34] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. M. Jr., Charmm general force field (cgenff): A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields, J. Comput. Chem. 31 (2010) 671–690.

[35] J. A. D. MacKerell, M. Feig, I. C. L. Brooks, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, J. Comput. Chem. 25 (2004) 1400–1415.

[36] N. Steinke, A. Genina, C. D. Lorenz, S. E. McLain, Salt interactions in solution prevent direct association of urea with a peptide backbone, J. Phys. Chem. B 121 (2017) 1866–1876. doi:10.1021/acs.jpcb.6b12542.

[37] S. Patel, C. L. Brooks, Charmm fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations, Journal of Computational Chemistry 25 (1) 1–16.

arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.10355, doi:10.1002/jcc.10355.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.10355

[38] S. Patel, A. D. Mackerell, C. L. Brooks, Charmm fluctuating charge force field for proteins: Ii protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model, Journal of Computational Chemistry 25 (12) 1504–1514. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20077, doi:10.1002/jcc.20077.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20077

[39] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, P. Ren, Polarizable atomic multipole-based amoeba force field for proteins, Journal of Chemical Theory and Computation 9 (9) (2013) 4046–4063, pMID: 24163642. arXiv:https://doi.org/10.1021/ct4003702, doi:10.1021/ct4003702.
URL https://doi.org/10.1021/ct4003702

[40] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, R. A. Friesner, Opls3: A force field providing broad coverage of drug-like small molecules and proteins, Journal of Chemical Theory and Computation 12 (1) (2016) 281–296, pMID: 26584231. arXiv:https://doi.org/10.1021/acs.jctc.5b00864, doi:10.1021/acs.jctc.5b00864.
URL https://doi.org/10.1021/acs.jctc.5b00864

[41] S. Busch, C. D. Lorenz, J. Taylor, L. C. Pardo, S. E. McLain, Short-range interactions of concentrated proline in aqueous solution, The Journal of Physical Chemistry B 118 (49) (2014) 14267–14277, pMID: 25383939. arXiv:https://doi.org/10.1021/jp508779d, doi:10.1021/jp508779d.
URL https://doi.org/10.1021/jp508779d

[42] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, Gromacs 4: Algorithms for highly efficient, load-balaced and scalable molecular dimulation, J. Chem. Theory Comput. 4 (2008) 435–447.

[43] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Mol. Phys. 52 (1984) 255–268.

[44] W. G. Hoover, Canonical dynamcs: equilibrium phase-space distributions, Phys. Rev. A 31 (1985) 1695–1697.

[45] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, M. L. Klein, Explicit reversible integrators for extended systems dynamics, Mol. Phys. 87 (1996) 1117–1157.

[46] T. Darden, D. York, L. Pedersen, Particle mesh ewald: An n*log(n) method for ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.

[47] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh ewald potential, J. Chem. Phys. 103 (1995) 8577–8592.

[48] A. J. Johnston, S. Busch, L. C. Pardo, S. K. Callear, P. C. Biggin, S. E. McLain, On the Atomic Structure of Cocaine in Solution, Phys. Chem. Chem. Phys. 18 (2) (2016) 991–999. doi:10.1039/C5CP06090G.

[49] [link].
URL https://gcm.upc.edu/en/members/luis-carlos/angula/ANGULA