Electronic Supplementary Information for 'On the hydration structure of the pro-drug GPG-NH_2 and its deriviatives'

Paul Smith,[†] Nicola Steinke,[‡] John F. C. Turner,[¶] Sylvia E. McLain,[*,‡] and Christian D. Lorenz[*,†]

†*Biological Physics & Soft Matter Group, Department of Physics, Strand Campus, Kings College London, Strand, London WC2R 2LS*

‡*Department of Biochemistry, University of Oxford OX1 3QU, United Kingdom*

¶*Department of Chemistry, University of Sussex, Brighton BN1 9QJ, United Kingdom*

E-mail: sylvia.mclain@bioch.ox.ac.uk; chris.lorenz@kcl.ac.uk

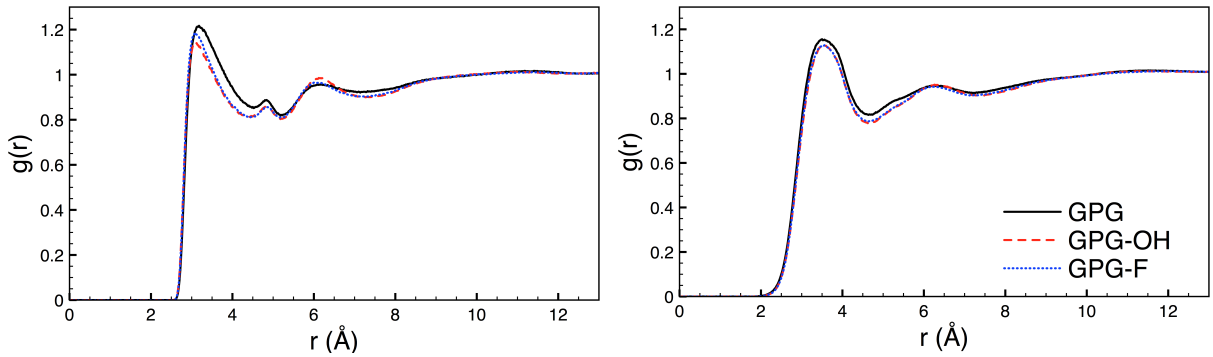
First neighbor distances and coordination numbers of backbone atoms in GPG, GPG-F & GPG-OH peptides

Radial distribution functions ($g(r)$s; RDFs) have been determined for various parts of the GPG, GPG-F and GPG-OH peptides in order to quantify the hydration properties of the three different peptides. Figure 1 shows the $g(r)$s for the C-terminal Ncap N_4 atoms and the O_w and H_w atoms in the surrounding water molecules. The $g(r)$s for the 2PRO-3GLY O_2 and 3GLY-Ncap O_3 are shown in Fig. 2. Finally, the $g(r)$s for the atoms attached to C_α in the 3GLY residue, where the modification of the derivative peptides has occurred, are shown in Fig. 3.

The first neighbor distances $r_{\max}$ and coordination numbers n_i^j are determined from the radial distribution functions for the various interactions of interest. The first neighbor distance is taken from the distance at which the first minimum after the first peak occurs. While the coordination numbers are determined by integrating the radial distribution function from a distance of 0 to the distance $r_{\max}$. In Tables 1 & 2, the first neighbor distances and coordination numbers for the backbone nitrogens and oxygens, respectively, have been tabulated. The values of $r_{\max}$ are then used in other analysis where a distance criteria is applied for determining bound molecules or first neighbor molecules (e.g. the water mediated chains).

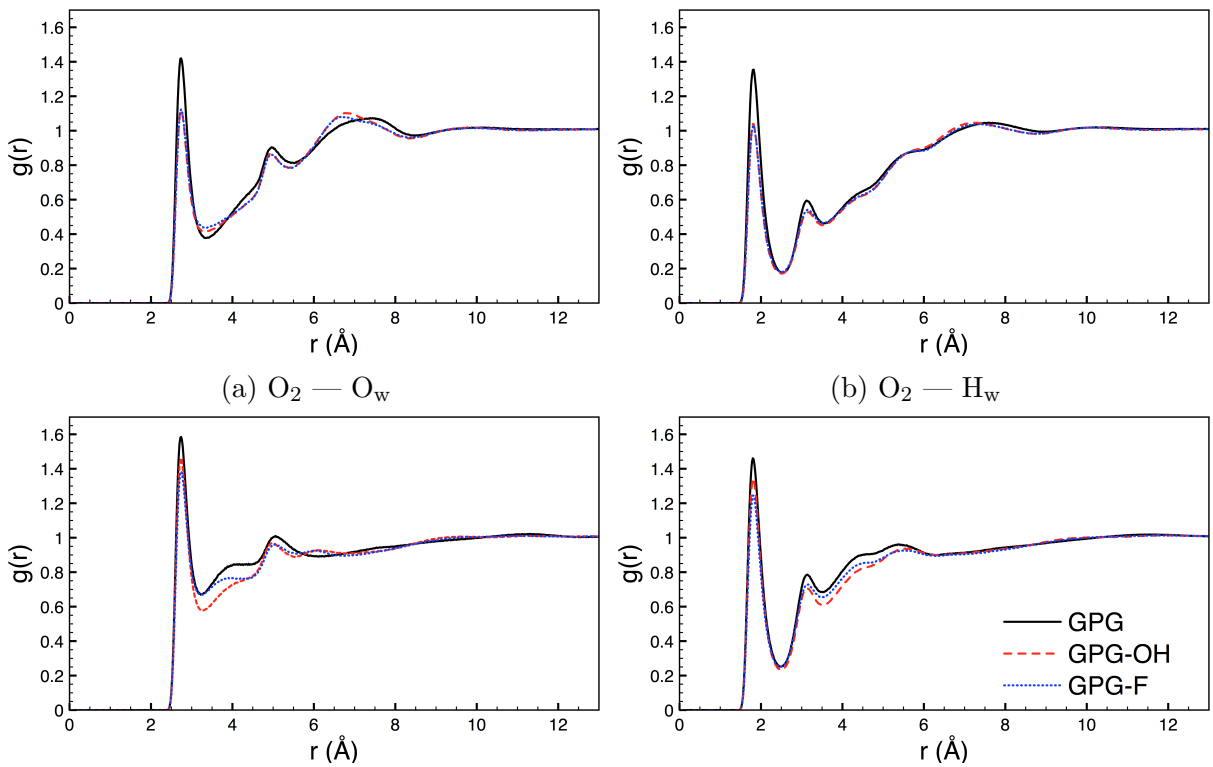
Table 2: Coordination numbers n_i^j and nearest neighbour distances $r_{\max}$ of the solvent atoms around the O_x atoms in the GPG, GPG-F and GPG-OH peptides.

		GPG		GPG-OH		GPG-F	
Type i	Type j	n_i^j	$r_{\max}$ (Å)	n_i^j	$r_{\max}$ (Å)	n_i^j	$r_{\max}$ (Å)
O_1	O_w	0.94	3.00	0.94	3.00	0.92	3.00
	H_w	0.97	2.44	0.97	2.44	0.95	2.44
O_2	O_w	1.9	3.38	1.7	3.38	1.7	3.38
	H_w	1.6	2.54	1.3	2.54	1.3	2.54
O_3	O_w	2.2	3.28	2.0	3.28	2.0	3.28
	H_w	1.8	2.50	1.7	2.50	1.6	2.50



(a) N_4 — O_w (b) N_4 — H_w

Figure 1: Radial distribution functions for (a) O_w and (b) H_w atoms in the water molecules around N_4 atoms of the GPG, GPG-F and GPG-OH peptides.



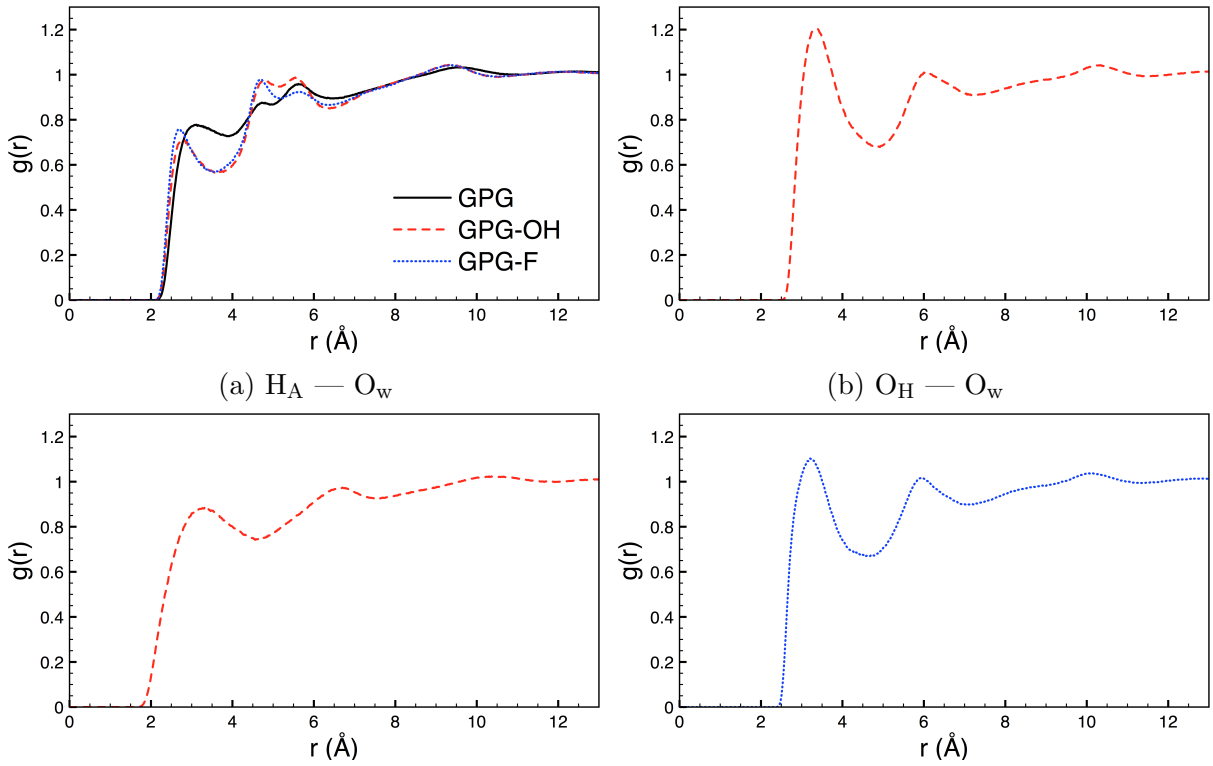
(a) O_2 — O_w (b) O_2 — H_w

(c) O_3 — O_w (d) O_3 — H_w

Figure 2: Radial distribution functions for (a) O_w and (b) H_w atoms in the water molecules around O_2 atoms and for the (c) O_w and (d) H_w atoms around O_3 atoms of the GPG, GPG-F and GPG-OH peptides.

Distributions of the configurations of GPG, GPG-F & GPG-OH backbones

The dihedrals ϕ_6 & ϕ_7 (as shown in Fig. 6(a) of manuscript) are the only dihedrals which shown any significant difference in their conformations in the three different peptides that



(a) H_A — O_w (b) O_H — O_w

(c) H_O — O_w (d) F — O_w

Figure 3: Radial distribution functions of the interactions of the different derivative groups on the GPG, GPG-F and GPG-OH with the O_w a atoms of the water molecules.

Table 1: Coordination numbers n_i^j and nearest neighbour distances r_{max} of the solvent atoms around the N_x atoms in the GPG, GPG-F and GPG-OH peptides.

		GPG		GPG-OH		GPG-F	
Type i	Type j	n_i^j	r_{max} (Å)	n_i^j	r_{max} (Å)	n_i^j	r_{max} (Å)
N_1	O_w	4.2	3.58	4.1	3.50	4.2	3.58
	H_w	15.5	4.38	15.5	4.38	15.4	4.38
	Cl^-	0.28	3.92	0.28	3.92	0.28	3.92
N_2	O_w	28.4	6.60	28.0	6.60	28.0	6.60
	H_w	68.5	7.02	67.5	7.02	67.6	7.02
	Cl^-	0.71	7.10	0.71	7.10	0.72	7.10
N_3	O_w	0.99	3.36	0.99	3.36	1.0	3.36
	H_w	9.6	4.34	9.1	4.34	9.2	4.34
	Cl^-	0.03	4.10	0.03	4.10	0.03	4.10
N_4	O_w	8.4	4.52	7.8	4.52	7.9	4.52
	H_w	18.7	4.66	17.8	4.66	17.9	4.66
	Cl^-	0.10	4.36	0.09	4.36	0.10	4.36

have been studied. In Fig. 4, the distributions of these two dihedrals for the three different peptides have been plotted. The behavior of the peptide backbones has been described in the main manuscript.

These were picked by eye to show a cluster of different conformations that the peptides adopt in solution with repsect to the variant dihedrals phi6 and psi 7 as discussed in the main draft. Figure 4 in the ESI shows the dihedral curves for these variant conformations between the peptides. The peptides were chosen to be representative for the minimum and maximum variation as well as the average position of the dihedral, shown as a molecular structure in Fig 6 of the main draft. This allows for the range of conformations to be depicted, while not cluttering up the figure so that the conformations could not be observable.

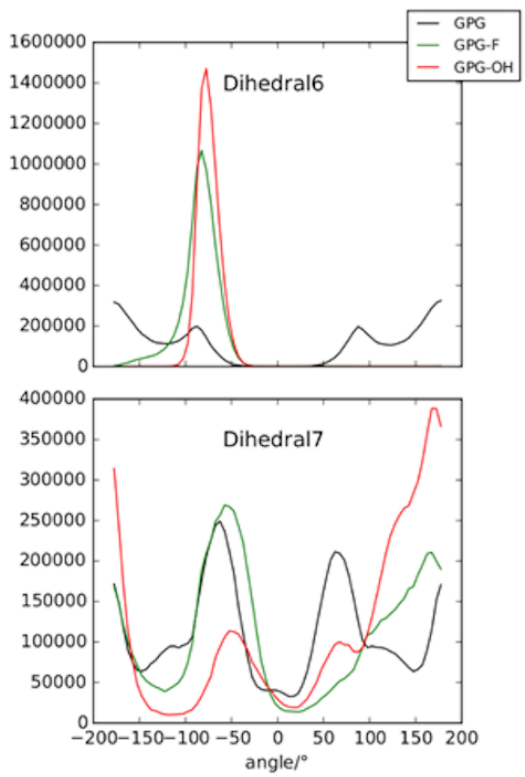


Figure 4: Probability distributions of the ϕ_6 (top) and ψ_7 (bottom) dihedrals of the GPG, GPG-F and GPG-OH peptides.

Also, a principal component analysis (PCA) has been carried out of the three different peptides. Figure 5 shows a two-dimensional projection of the two major eigenvectors that are found from the PCA for each of the peptides. The results show that the structures of the GPG-F and GPG-OH peptides generally fall into two different clusters of conformations, while the GPG peptides are found in one more diffuse cluster of conformations. Also, the GPG-F and GPG-OH clusters seem to generally be quite similar. The two eigenvectors in each system generally represent the two dihedrals identified in the manuscript in Fig. 4.

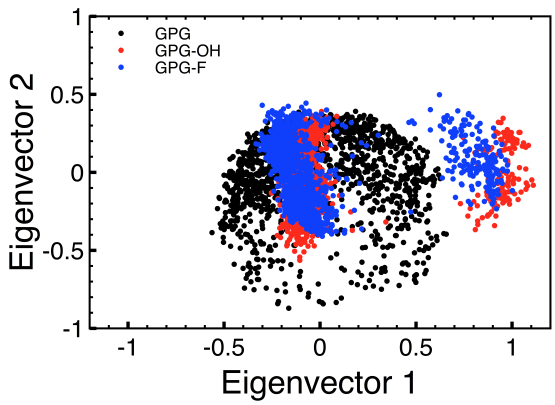


Figure 5: Principal component analysis for GPG (black), GPG-OH (red) and GPG-F (blue).

Description of ANGULA analysis

ANGULA is a program which can be used to analyze the three-dimensional atomic-scale information from computational models of liquids and amorphous systems. It allows for the relative positions and orientations of molecules and atoms to each other to be obtained. ANGULA can also be used to analyze and plot frequency distributions of molecular properties and their possible correlation (e.g. inter- or intramolecular distances, (dihedral) angles) and to select molecules with any of these criteria. Further, these analysis options can be combined with each other to provide a comprehensive range of analytical techniques.

In this study, ANGULA was used to determine the distribution of dihedral angles along the peptide backbone. Simultaneously, the three-dimensional distribution of neighboring

water oxygens around different conformational groups of the whole peptide was determined.

Five thousand coordinate files from each of the MD trajectories were used and orthonormal Cartesian coordinate systems were assigned to the peptide and water molecules. Each coordinate system is defined *via* the addition of four pseudo-atoms: the first in the position of the origin ⓞ and the others (ⓧ, ⓨ and ⓩ) located on the intended x-, y- and z-axes.

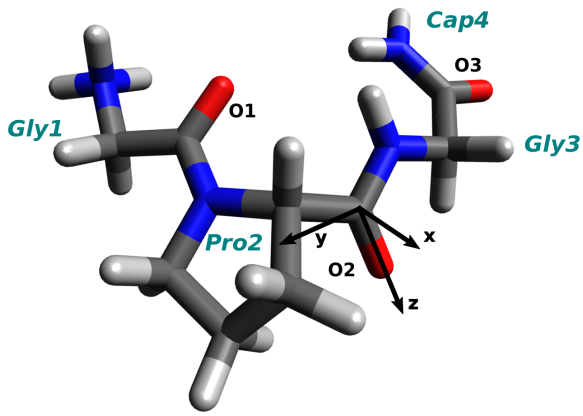


Figure 6: Peptide coordinate system used within ANGULA analysis.

The origin of the peptide coordinate system (Fig. 6) was placed at the proline carbonyl carbon and the pseudo-atom ⓞ was superimposed onto this position and the z-axis pseudo-atom ⓩ is superimposed on the proline carbonyl oxygen (O2). The positions of the pseudo-atoms (ⓧ and ⓨ) need to be defined via the vector product. For the x-axis, the z-axis from origin to oxygen and the vector from origin to the next glycine nitrogen (N3) map out a plane to which the cross product resulting vector is perpendicular, according to the right-hand-rule. As the vector product only defines a direction and not a position, the distance to the origin (1 Å) needs to be defined as well for the positioning of ⓧ. Subsequently, the y-axis is defined by placing ⓨ at a distance of 1 Å to the origin on the vector product of the x- and y-axis. Similarly, a coordinate system can be assigned to water (Fig. 7) with the origin at the water oxygen, the z-axis along one of the OH-bonds and the x- and y-axes defined *via* cross products.

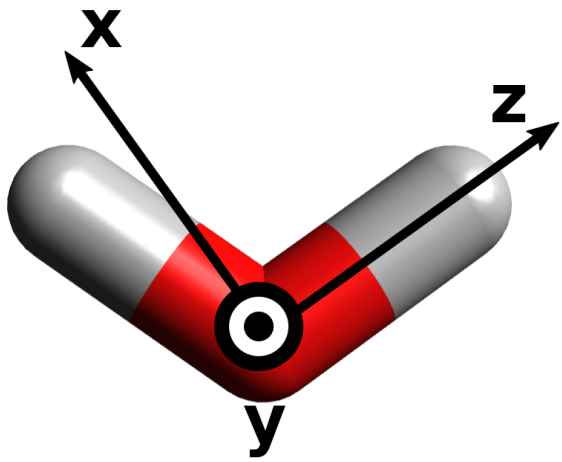


Figure 7: Coordinate system used for water molecules within ANGULA analysis

The relative position of molecules to each other can be obtained by calculating a spatial distribution. A central and a surrounding molecular species with coordinate systems can be chosen, e.g. water around the peptide. A Whole Molecule Analysis (WMA) is a specific type of spatial distribution that displays neighboring molecule atoms (in this case water oxygens) within a specific distance range (0-2.2 Å) from any of the central peptide molecule atoms. ANGULA calculates the position of surrounding atoms as spherical polar coordinates (r, θ_{CM}, ϕ_{CM}) from the central molecule origin. The calculation of density clouds relies on the conversion of spherical polar-coordinates calculated by ANGULA to Cartesian xyz-coordinates. Each surrounding atom within the defined distance range is represented by a density point. In most cases it is not practical to display 100 % of the data points and only the top 40 % within the density distribution are shown. Consequently, the displayed data point clouds make up a surface of equal density (i.e. isopycnic).

For small central molecules this WMA method can easily be applied without further steps. Larger and more flexible molecules, like the peptides in this study, need to be analyzed for their conformational properties as well. As the position of the neighbor atom is only determined by one coordinate system, the further this neighbor atom is away from the central origin, the more the position will be influenced by the central molecule conformation. There are multiple criteria to determine and select for a specific conformation, such as in-

tramolecular distances, bond angles or dihedrals. For the peptides in this work the backbone dihedral angle distributions were determined in parallel with the WMA calculations as additional variables. The ϕ_6 and ψ_7 angles on the second glycine (as shown in Fig. 6(a) of the manuscript) are the only dihedral angles which have shown any significant difference in their conformation in the three peptides that have been studied. In Fig. 4 the distributions of these two dihedrals for the three different peptides have been plotted. The WMA data was subsequently split into three different conformational groups depending on ψ_7 either being between 0° to 125°, -125° to 0° or 125° to -125°.