Supplementary Information

Graphitic nanoflakes modulate the structure and binding of human amylin

Alexa Kamboukos[1], Billy J Williams-Noonan[2], Patrick Charchar[1], Irene Yarovsky[1], Nevena Todorova[1a)]

[1] School of Engineering, RMIT University, Melbourne, Victoria, Australia, 3001

[2] School of Science, RMIT University, Melbourne, Victoria, Australia, 3001

[a)] Author to whom correspondence should be addressed: nevena.todorova@rmit.edu.au

1. Amylin and nanoflake (NF) models

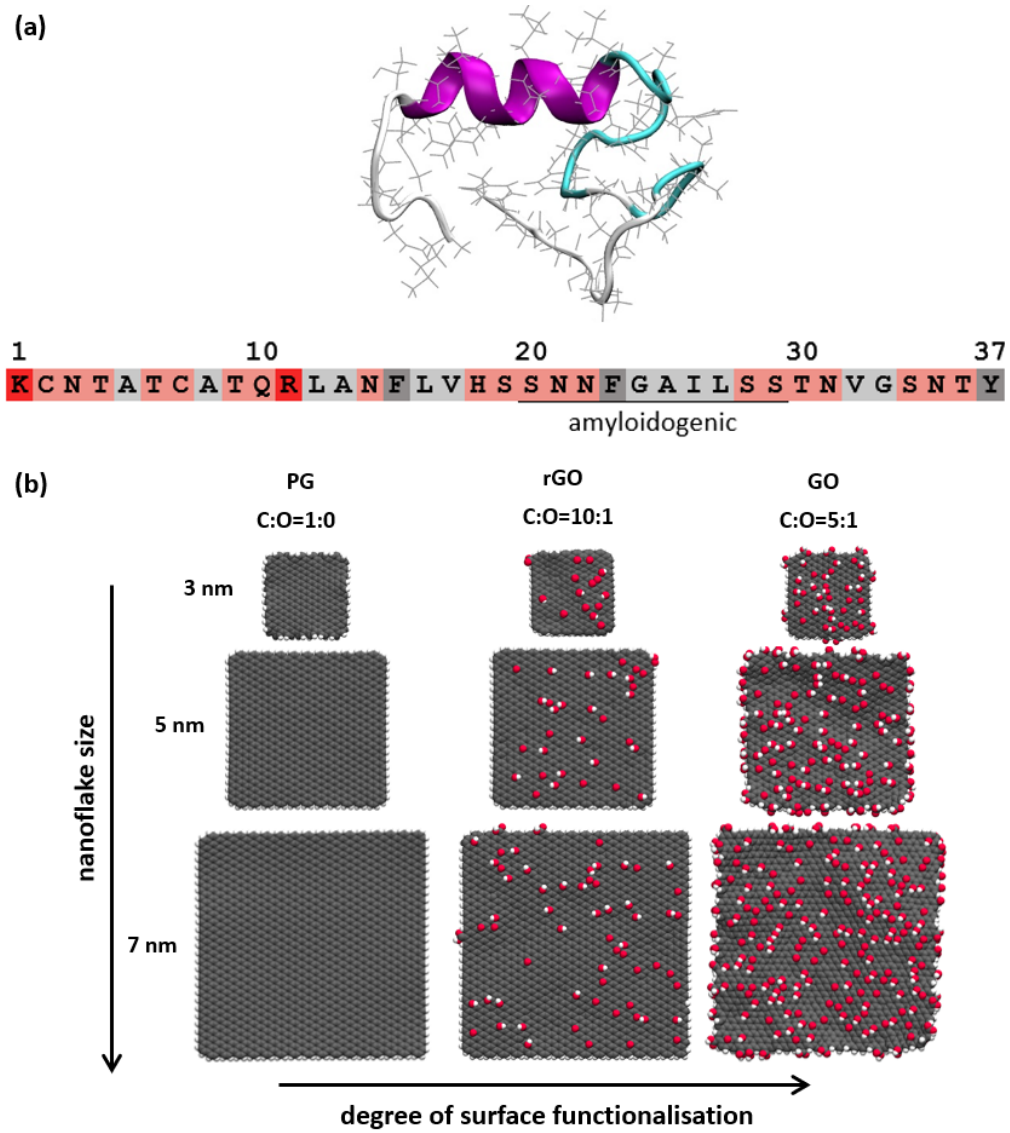

Fig. S1 Amylin and nanoflake models. (a) Representative aqueous native structure of human amylin with secondary structure displayed in cartoon representation (white coils, cyan turns, and magenta α-helix) and light grey lines showing the atomistic peptide sidechains. Primary sequence of amylin is shown and residue numbers are coloured by sidechain physicochemical properties: dark red - charged (Lys and Arg); light red - hydrophilic (Asn, Thr, Gln, His, Ser and Cys); light grey - hydrophobic (Ala, Leu, Val, Gly and Ile); and dark grey - aromatic (Phe and Tyr). **(b)** NF size (3, 5 and 7 nm) and degree of surface oxidation (PG, rGO and GO) varies between each model. Atoms are drawn with a space-filling representation and coloured: carbon (grey), hydrogen (white), and oxygen (red). Carboxyl groups are distributed on the NF edge, hydroxyl and epoxy groups are scattered on the planar NF region. Water and counter ions are hidden for visual clarity.

2. Additional computational details

2.1 Periodic box sizes

Periodic box sizes (nm) of 6.2 x 6.2 x 6.2 (GO3), 8.0 x 8.0 x 8.0 (GO5), 11.1 x 11.1 x 11.1 (GO7), 6.0 x 6.0 x 6.0 (rGO3), 7.8 x 7.8 x 7.8 (rGO5), 11.3 x 11.3 x 11.3 (rGO7), 5.4 x 5.5 x 6.0 (PG3), 7.8 x 7.8 x 7.8 (PG5), and 9.6 x 9.6 x 9.6 (PG7) were used.

2.2 BE-MetaD parameters for the collective variables CV1–3

Table S1 BE-MetaD parameters for CV1–3

Collective Variable	Gaussian width (σ)
CV1: Amylin gyration	0.1 Å
CV2: Amylin–NF COM distance	0.1 Å
CV3: Amylin–NF orientation angle	0.05 radians

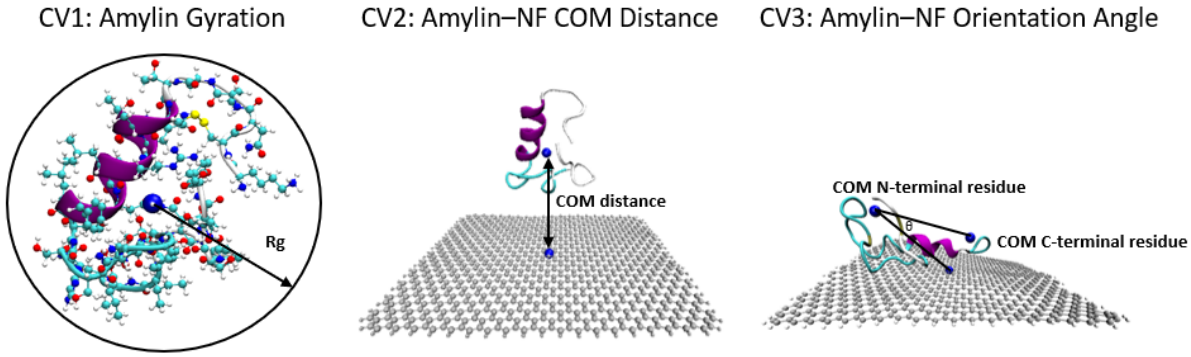

Fig. S2 Schematic illustration of collective variables CV1–3 used to bias the amylin–NF simulations. Molecule centre of mass (COM) is shown as a blue sphere.

2.3 Convergence of BE-MetaD simulations

Convergence indication of the BE-MetaD simulations was first monitored by the evolution of gaussian height for CV1–3 (Fig. S3–S5). Spikes in gaussian height indicate a new region of phase space is being explored, with gaussian height remaining stable close to zero for a sufficient period indicating convergence. Simulations were considered converged when gaussian height over time was below 0.03 kJ/mol for at least 75% of the simulation time between the 450-650 ns time range. Gaussian height for amylin–NF COM distance (CV2) fluctuates more than for CV1 and CV3, which could be associated with the inherently disordered nature of amylin resulting in a fluctuating structure.

Convergence was further investigated by examining the change in the free energy surface (FES) for all structures below an energy threshold of 3 kT for CV1–3 (Fig. S6). Convergence was considered reached when the RMSD in free energy, relative to the final FES, for coordinates with a free energy less than 3 kT from the minimum, remained below 10 kJ/mol for approximately 50 ns for each CV. Although this criterion was satisfied for most systems by 650 ns of the simulation (Fig. S6), the FES RMSD for CV1 in the amylin–rGO5 and amylin–GO7 simulations showed FES fluctuations between 600-650 ns (Fig. S6e and S6i), with greater changes observed in the amylin–rGO5 simulation. Examination of the 1D free energy profile of CV1 in the amylin-rGO5 simulation at 600, 650 and 700 ns provided assurance that the FES is changing minimally between 650 and 700 ns, especially for the low energy minima (Fig. S7). Additionally, given that gaussian height over time was below 0.03 kJ/mol

between 450-650 ns (Fig. S4 and S5), we decided to end these simulations due to research time constraints. Analysis was conducted on the neutral (non-biased) replica.

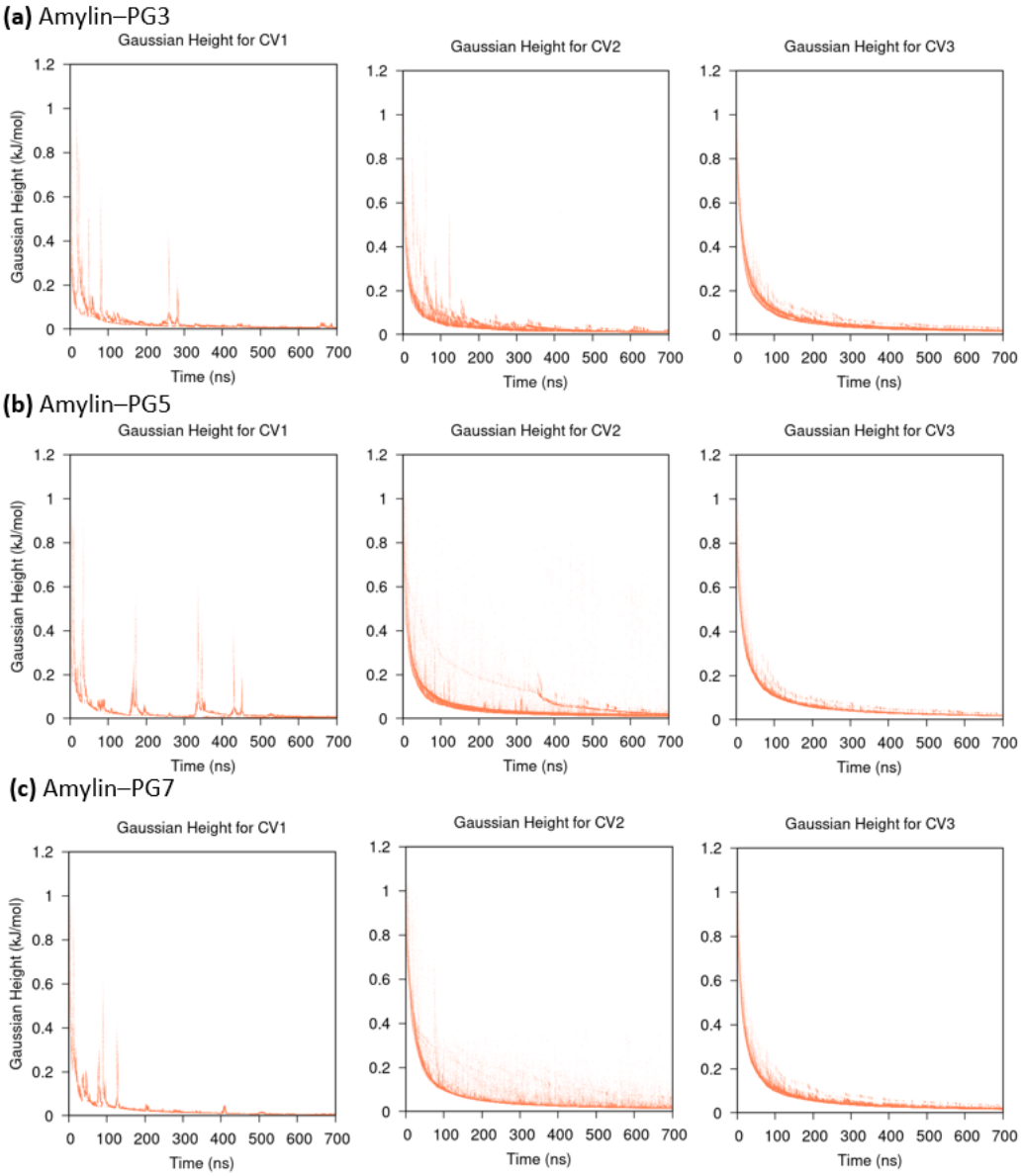

Fig. S3 Gaussian height as a function of time for CV1–3 in the amylin–PG BE-MetaD simulations.

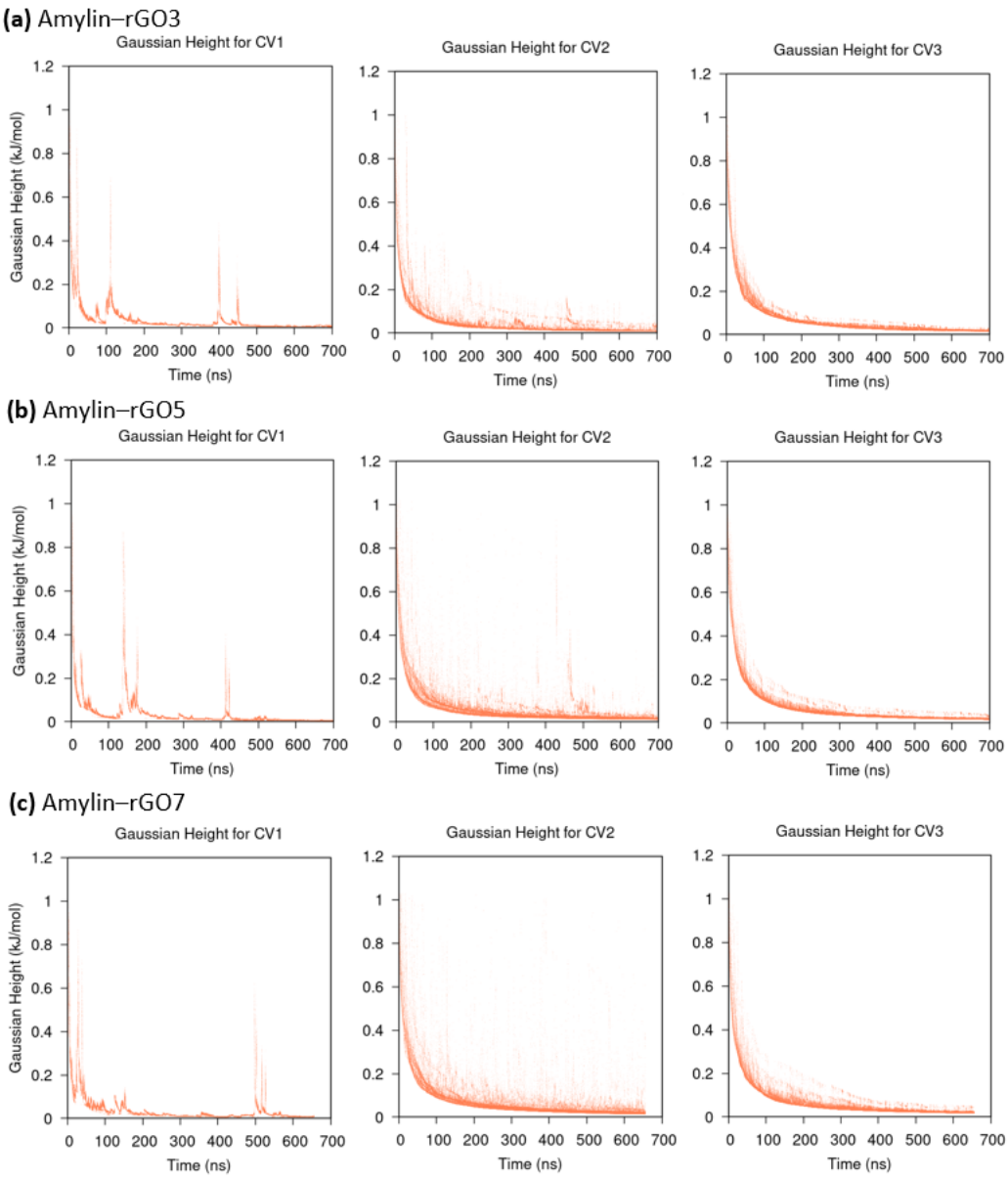

Fig. S4 Gaussian height as a function of time for CV1–3 in the amylin–rGO BE-MetaD simulations.

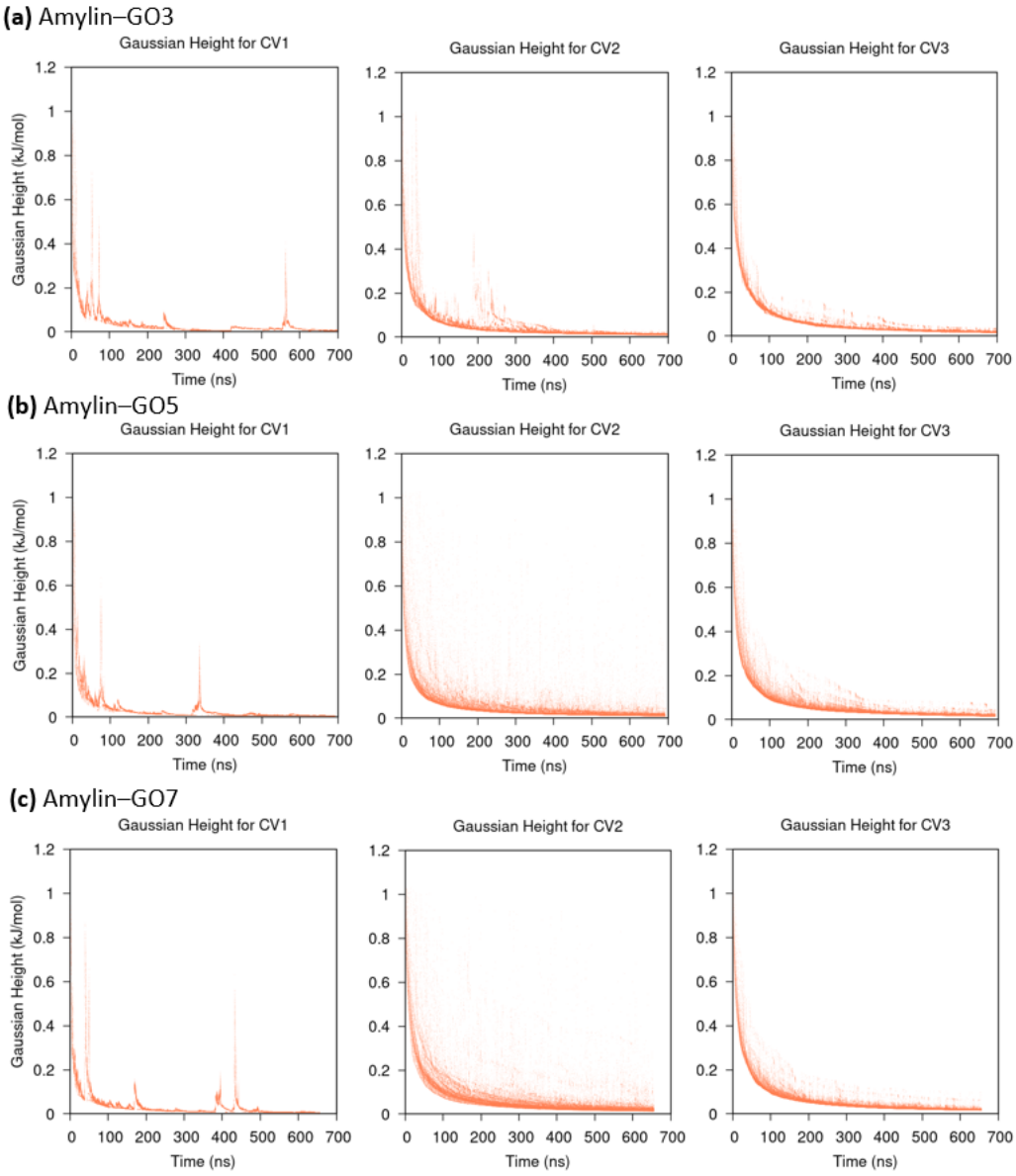

Fig. S5 Gaussian height as a function of time for CV1–3 in the amylin–GO BE-MetaD simulations.

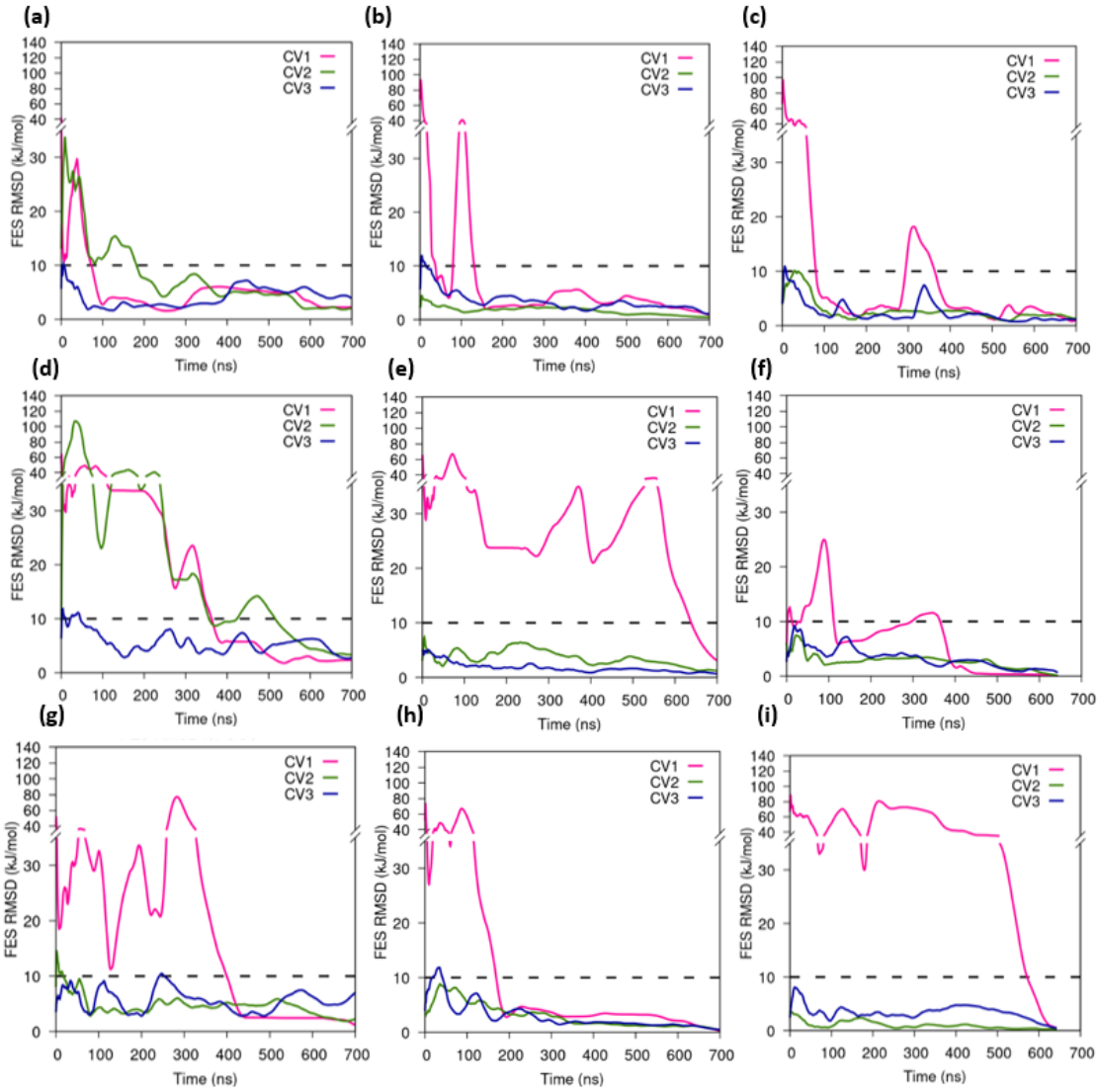

Fig. S6 Free energy surface (FES) root-mean-square-deviation (RMSD) for all structures below an energy threshold of 3 kT for CV1–3 in the BE-MetaD simulations. A dashed line at 10 kJ/mol represents the FES RMSD target selected for justifying convergence. **(a)** Amylin–PG3, **(b)** Amylin–PG5, **(c)** Amylin–PG7, **(d)** Amylin–rGO3, **(e)** Amylin–rGO5, **(f)** Amylin–rGO7, **(g)** Amylin–GO3, **(h)** Amylin–GO5, **(i)** Amylin–GO7.

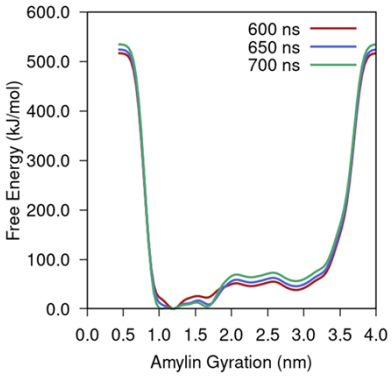

Fig. S7 FES as a function of CV1 at different simulation times for the Amylin-rGO5 simulation.

2.4 SPON-MD simulation details

Equivalent simulation parameters and preparation steps were used in the SPON-MD and BE-MetaD simulations (refer to main manuscript), with the exception of a smaller cut-off value of 1.0 nm being used for calculating van der Waals and short-range electrostatics. We acknowledge that this shorter cut-off, while increasing computational performance, deviates from the recommended 1.2 nm cut-off that the CHARMM force field was parameterised with and could potentially introduce simulation artefacts. However, our primary focus here is on surface-induced protein (un)folding, and previous studies have demonstrated that "the free energy of folding is relatively insensitive to the choice of cut-off beyond 9 Å, [provided that] an Ewald method is used to account for long-range electrostatic interactions."[1] Moreover, our selection of this specific cut-off was aimed at ensuring consistency with the SPON-MD parameters employed in our preceding studies.[2, 3]

SPON-MD was performed for 1 µs for each of the NF systems (two orientations of amylin per graphitic NF), totalling 18 µs of SPON-MD simulation data collected across all systems. RMSD of the amylin backbone was calculated and plotted against simulation time to monitor for equilibrium (Fig. S8). Most simulations show an initial spike in RMSD followed by a plateau, signifying amylin reached a stable equilibrium conformation. Variations in RMSD backbone values were observed in more than half of the amylin–NF systems (amylin–PG5, amylin–PG7, amylin–rGO5, amylin–GO3 and amylin–GO5), suggesting some degree of orientation bias and/or conformational trapping may be present in those specific SPON-MD simulations.

For each system, once amylin is identified to be adsorbed to the NF surface (minimum distance < 0.4 nm), the collection of adsorbed frames in the two independent SPON-MD simulations were combined to produce an ensemble trajectory. All analysis techniques conducted on the BE-MetaD simulations were also performed on the SPON-MD ensemble for consistency and the results are shown in Fig. S9–S14.

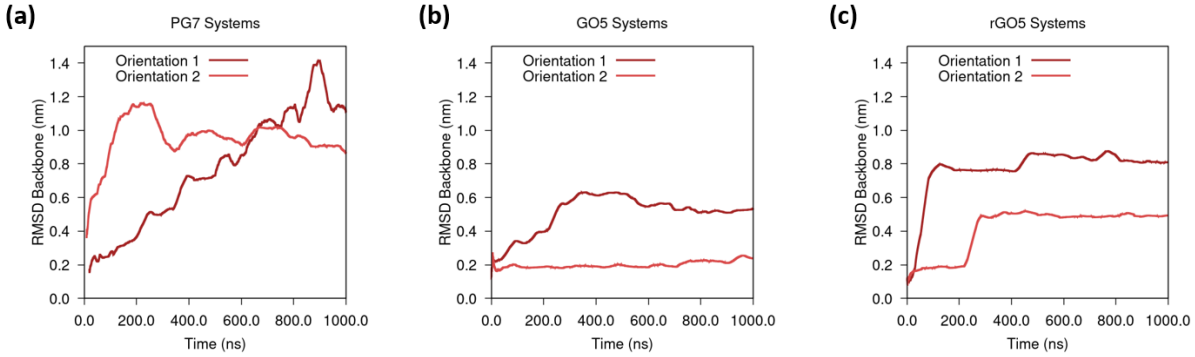

Fig. S8 Representative RMSD backbone plots for the amylin–NF SPON-MD simulations involving two different starting amylin orientations. (a) Amylin–PG7, **(b)** Amylin–GO5, and **(c)** Amylin–rGO5 systems.

2.5 Analysis details

2.5.1 Analysis on adsorbed states

To understand the conformational variance in surface-adsorbed amylin, RMSD cluster analysis was performed. An optimal cut-off value of 0.12 nm (backbone only) was identified to produce cluster population between 30–50% for the top three most populated clusters. Contact maps were constructed using VMD to identify the probability of each NF atom being in contact with amylin. A contact was defined as formed for any NF atom within 0.4 nm from any amylin atom. To identify the most stable surface-bound amylin conformations sampled in the unbiased BE-MetaD replicate and the SPON-MD ensemble, 2D free energy maps were constructed based on the amylin radius of gyration and amylin–NF contact area. These maps provide insights into the evolution of the protein compactness and extent of protein adsorption upon interaction with each NF. The contact area between amylin and the NF surface was determined using GROMACS by performing a series of solvent accessible surface area (SASA) calculations. A probe radius of 1.4 nm was employed, and three SASA

calculations were conducted to determine the Connolly surface area for: the amylin–NF complex ($A_{complex}$), amylin alone (A_{amylin}), and the NF alone (A_{NF}). The contact surface area between the solutes was calculated for each analysis frame as half the difference between the SASA of the amylin–NF complex and the sum of the individual SASA of each solute, i.e. $0.5 \times (A_{complex} - (A_{amylin} + A_{NF}))$.

2.5.2 Analysis on the frequently sampled (low energy) adsorbed states of amylin

In each free energy map, hotspots were defined as groups of highly sampled regions (low energy bound conformations) at a particular correlated value of amylin gyration and amylin–NF contact area, and analysis was conducted on the structures in the most sampled (visited) bin**.** RMSD cluster analysis was performed using the same procedure discussed above to extract representative structures of the low energy states. The secondary structure of amylin in adsorbed state was calculated using the STRIDE[4] algorithm on every 2nd structure in each hotspot pool (for computational efficiency) to categorise each residue as either α-helix, 3_{10} helix, π-helix, coil, turn, extended or isolated bridge. The average percentage of amylin adopting a specific secondary structure (coil, turn, helix and β-rich) with the two types of helical components (α-helices and 3_{10} helices) and two types of β-rich components (extended and isolated bridge) combined. The minimum distance of individual protein residues to the NF atoms was also calculated on the low energy states, as well as hydrogen bond formation between amylin and rGO/GO NFs.

3. Spontaneous MD (SPON-MD) Simulation Results

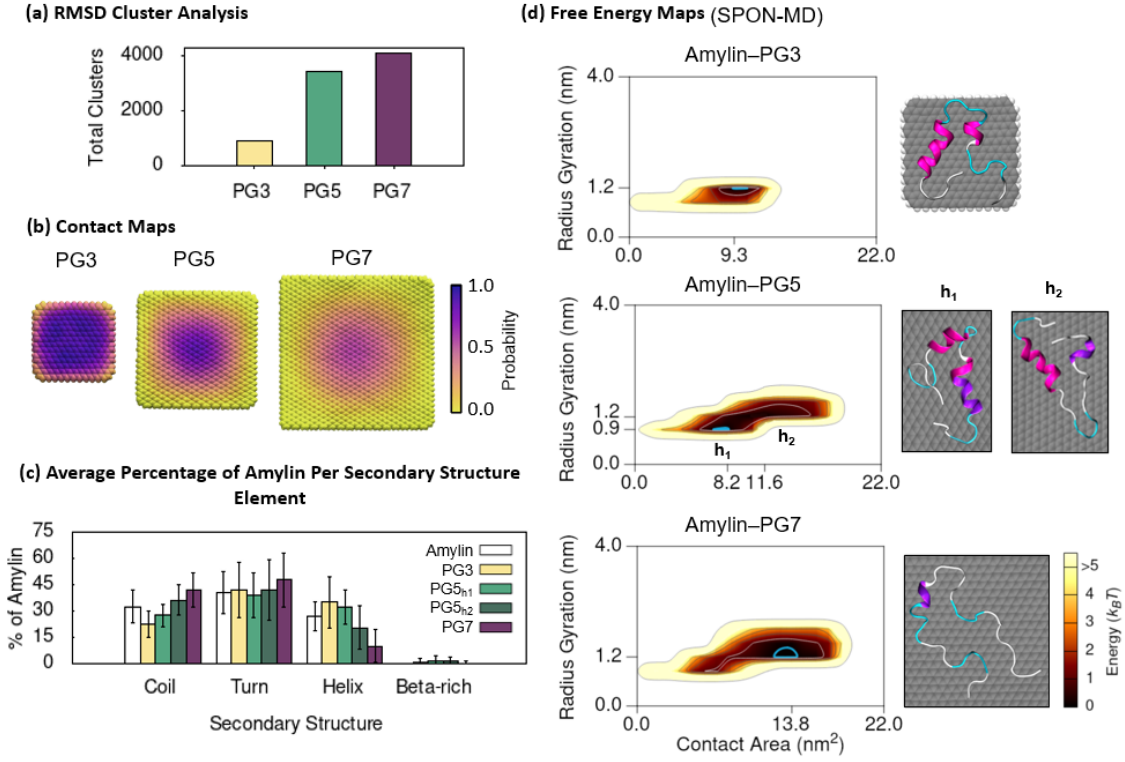

Fig. S9 SPON-MD simulation results of amylin conformational preferences and contacts with pristine graphene. (a) Total number of unique surface-bound amylin conformations identified by RMSD backbone cluster analysis from the amylin–PG simulations. **(b)** Contact probability maps showing the relative proportion of simulation time each PG atom maintains close association (less than 0.4 nm) with amylin. The probability values are coloured with a perceptually linear, sequential colour scale to identify the dynamics of amylin adsorption and persistent binding locations. **(c)** Average percentage of amylin adopting different type of secondary structure within the hotspot regions with errorbars representing standard deviation. The average secondary

structure of amylin in-solution[2] is provided in white bars for reference. **(d)** Amylin–PG free energy maps showing the relationship between amylin radius gyration and amylin–NF contact area. Hotspots (h) are sequentially labelled numerically based on their free energy and the blue highlight corresponds to the lowest energy hotspot region. Inset images illustrate representative structures from the given hotspots, as determined by RMSD backbone cluster analysis, and are coloured based on their typical secondary structure features: coils (white), turns (cyan), α-helix (magenta), 3_{10}-helix (violet), extended conformation (yellow), and isolated bridge (orange). Solvent hidden for clarity.

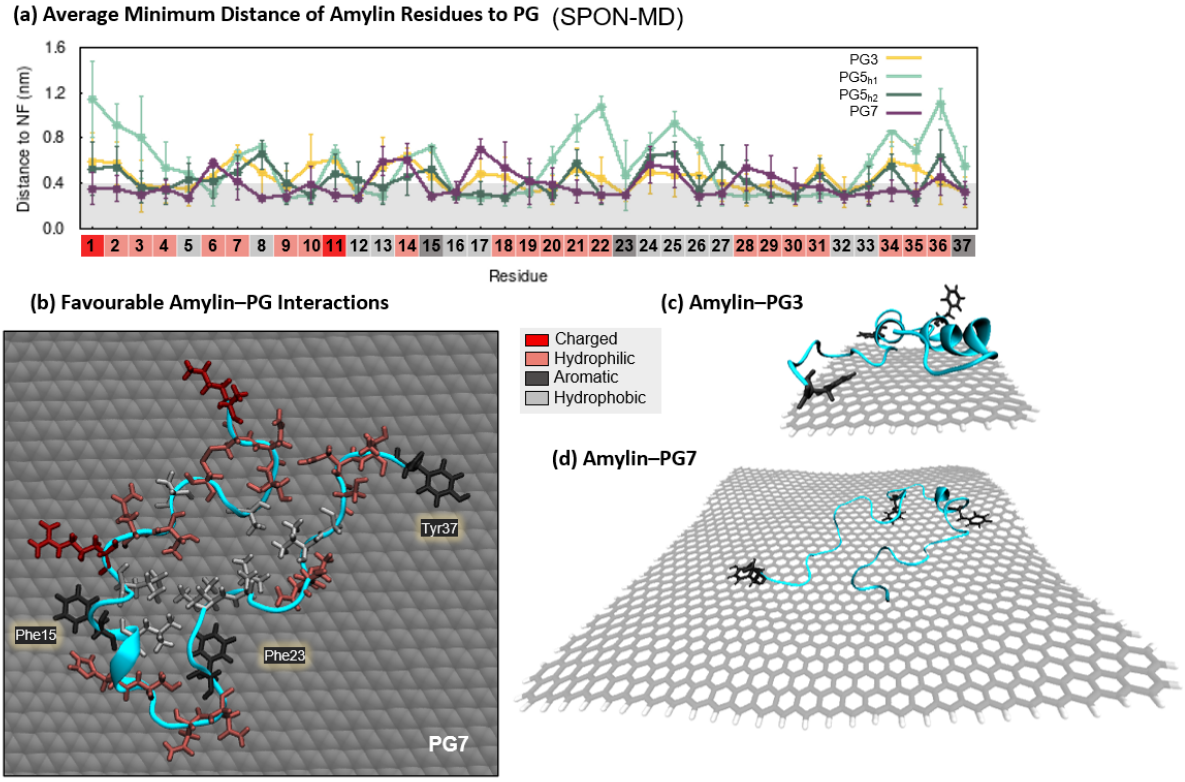

Fig. S10 SPON-MD simulation results of specific interactions between amylin and pristine graphene. (a) Average minimum distances measured between individual amylin residues and the PG3, PG5, and PG7 surfaces. Calculated on the low energy states of each system (shown in Fig. S9), with errorbars representing standard deviation. Shaded grey on the plot is used to indicate residue–NF contacts, defined as distances less than 0.4 nm. Residue numbers are coloured by sidechain physicochemical properties: dark red - charged (Lys and Arg); light red - hydrophilic (Asn, Thr, Gln, His, Ser and Cys); light grey - hydrophobic (Ala, Leu, Val, Gly and Ile); and dark grey - aromatic (Phe and Tyr). **(b)–(d)** Representative snapshots that highlight favourable amylin–PG interactions. The amylin backbone is shown in cyan cartoon representation, residue sidechains are drawn in licorice atomic detail and coloured based on their properties (i.e., charged, hydrophilic, hydrophobic, and aromatic), and graphene carbon atoms are shown in grey with either space-filling (B) or licorice (C, D) representations. Water and ions are hidden for clarity. **(b)** Top-view of amylin–PG7 showing residues within 0.4 nm of the NF. **(c), (d)** Side-view of amylin interacting with PG3 and PG7, emphasising the orientation of aromatic residues relative to the substrates.

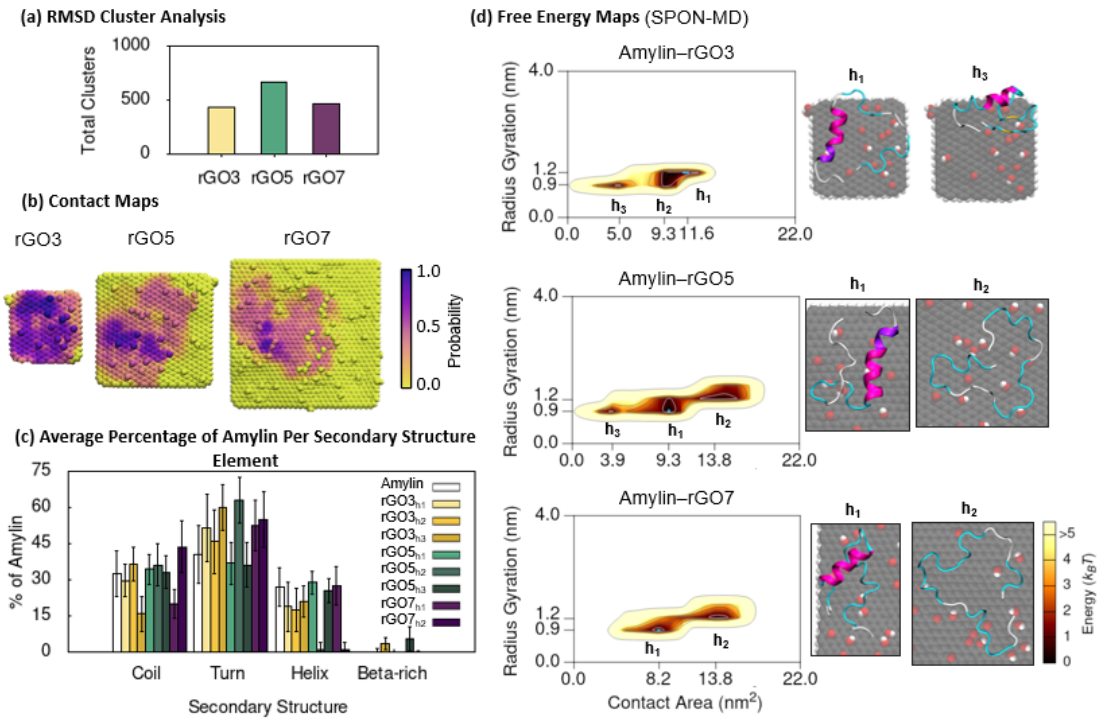

Fig. S11 SPON-MD simulation results of amylin conformational preferences and contacts with reduced graphene oxide. (a) Total number of unique surface-bound amylin conformations identified by RMSD backbone cluster analysis from the amylin–rGO simulations. **(b)** Contact probability maps showing the relative proportion of simulation time each rGO atom maintains close association (less than 0.4 nm) with amylin. **(c)** Average percentage of amylin adopting different type of secondary structure within the hotspot regions with errorbars representing standard deviation. The average secondary structure of amylin in-solution[2] is provided in white bars for reference. **(d)** Amylin–rGO free energy maps showing the relationship between amylin radius gyration and amylin–NF contact area. Hotspots (h) are sequentially labelled numerically based on their free energy and the blue highlight corresponds to the lowest energy hotspot region.

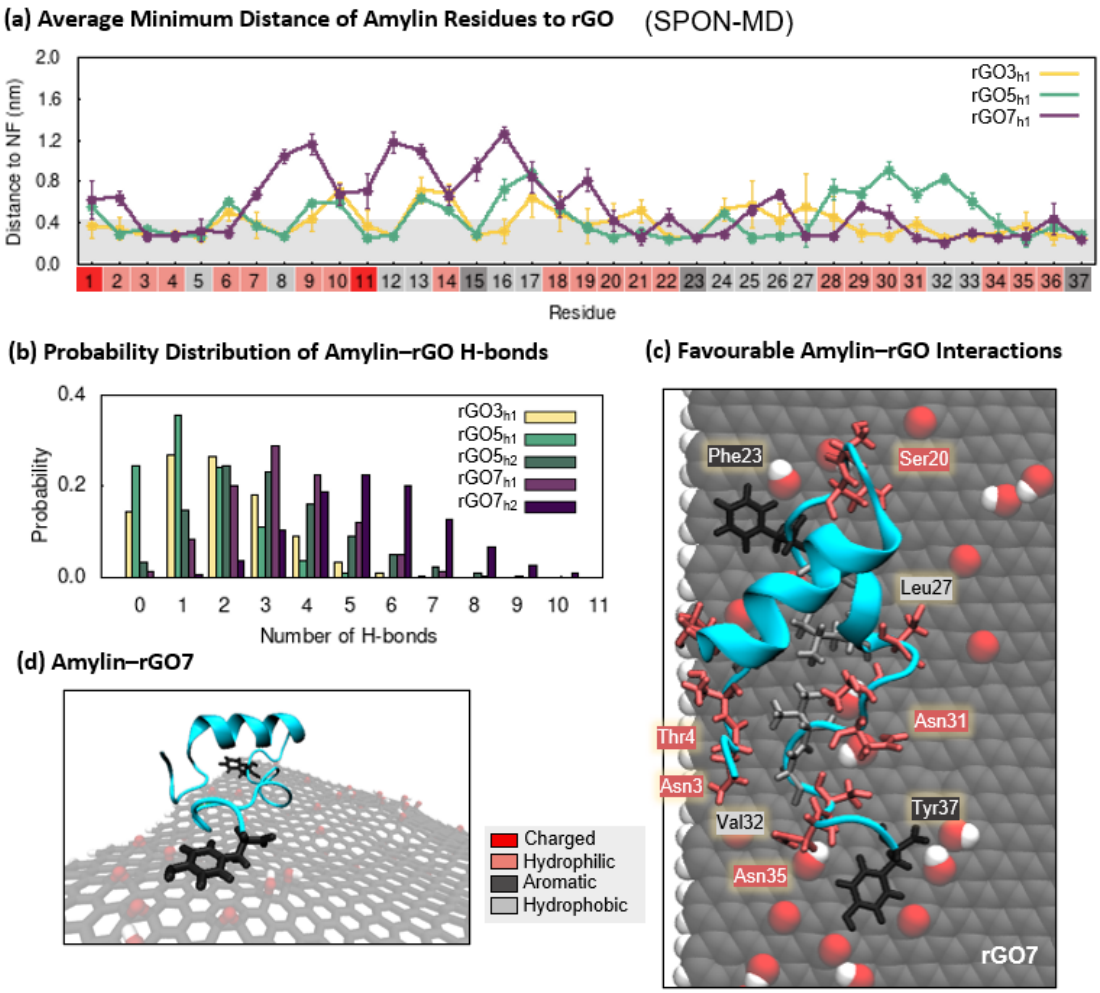

Fig. S12 SPON-MD simulation results of specific interactions between amylin and reduced graphene oxide. (a) Average minimum distances measured between individual amylin residues and the rGO3, rGO5, and rGO7 surfaces, with errorbars representing standard deviation. Shaded grey on the plot is used to indicate residue–NF contacts, defined as distances less than 0.4 nm. Residue numbers are coloured by sidechain physicochemical properties. **(b)** Probability distribution of hydrogen bonds formed between amylin and rGO in the amylin–rGO hotspots. **(c)–(d)** Representative snapshots that highlight favourable amylin–rGO interactions. The amylin residue sidechains are drawn in licorice atomic detail and coloured based on their properties. **(c)** Top-view of amylin–rGO7 showing residues within 0.4 nm of the NF. **(d)** Side-view of amylin interacting with rGO7, emphasising the orientation of aromatic residues relative to the substrate.

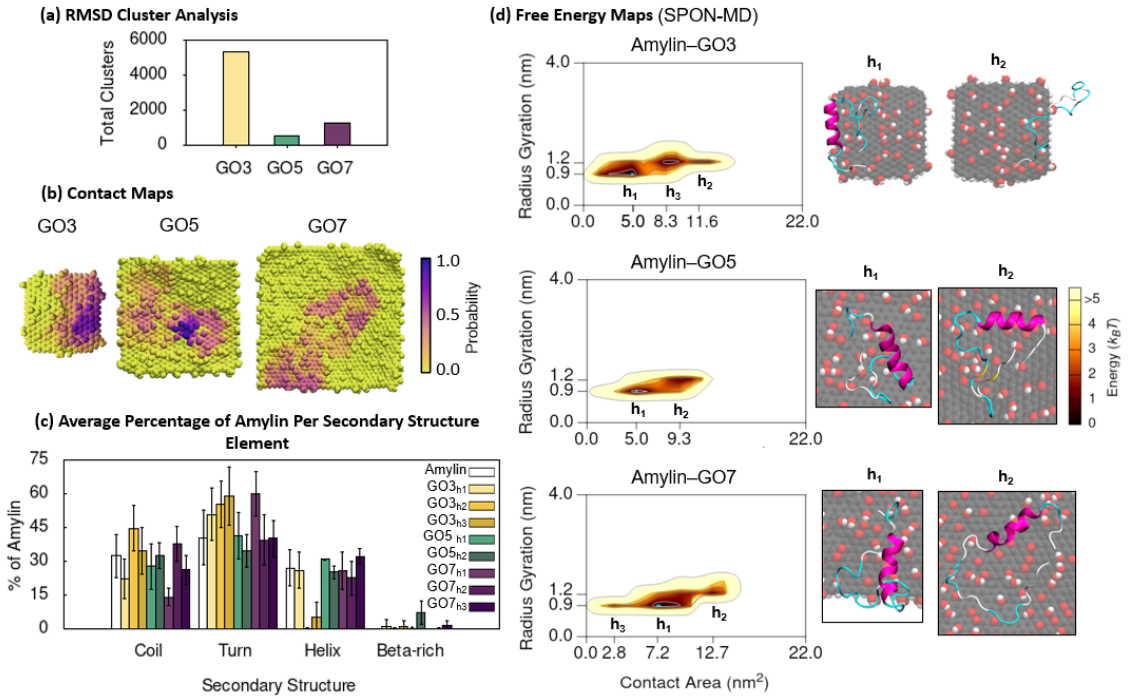

Fig. S13 SPON-MD simulation results of amylin conformational preferences and contacts with graphene oxide. **(a)** Total number of unique surface-bound amylin conformations identified by RMSD backbone cluster analysis from the amylin–GO simulations. **(b)** Contact probability maps showing the relative proportion of simulation time each GO atom maintains close association (less than 0.4 nm) with amylin. **(c)** Average percentage of amylin adopting different type of secondary structure within the hotspot regions with errorbars representing standard deviation. The average secondary structure of amylin in-solution[2] is provided in white bars for reference. **(d)** Amylin–GO free energy maps showing the relationship between amylin radius gyration and amylin–NF contact area. Hotspots (h) are sequentially labelled numerically based on their free energy and the blue highlight corresponds to the lowest energy hotspot region.

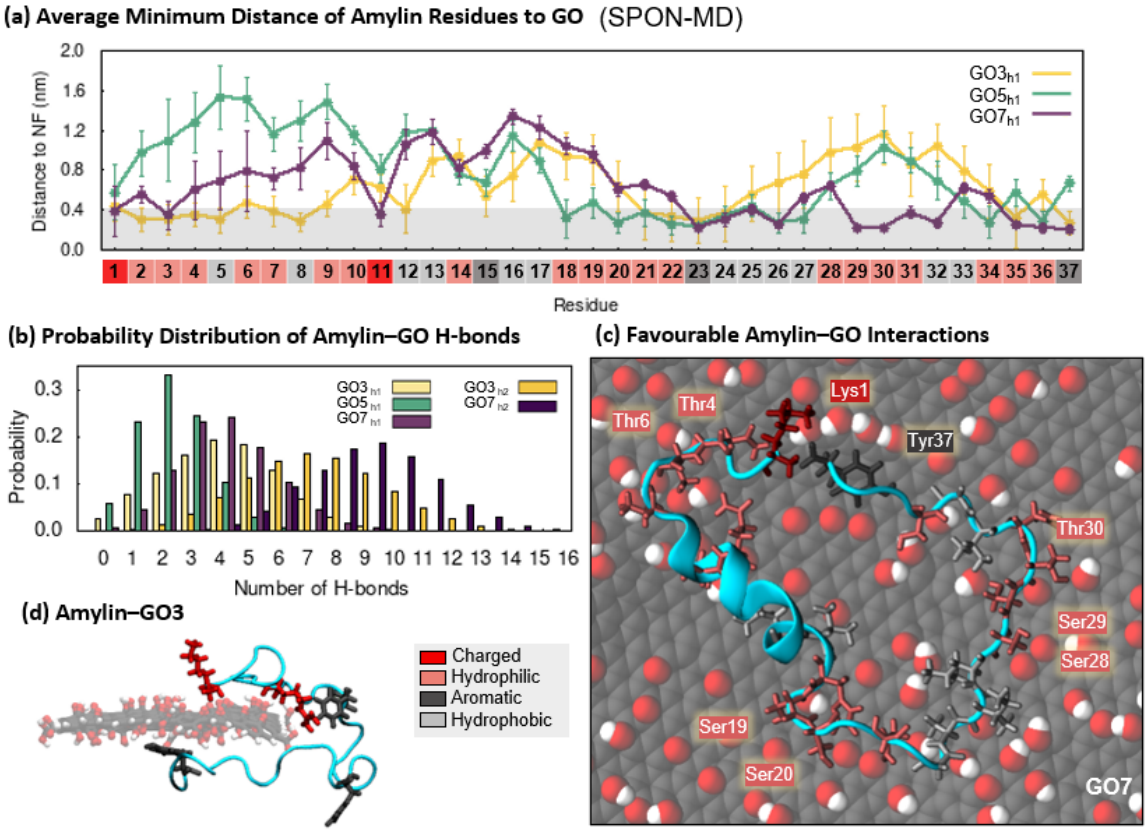

Fig. S14 SPON-MD simulation results of specific interactions between amylin and graphene oxide. (a) Average minimum distances measured between individual amylin residues and the GO3, GO5, and GO7 surfaces, with errorbars representing standard deviation. Shaded grey on the plot is used to indicate residue–NF contacts, defined as distances less than 0.4 nm. Residue numbers are coloured by sidechain physicochemical properties. **(b)** Probability distribution of hydrogen bonds formed between amylin and GO in the amylin–GO hotspots. **(c)–(d)** Representative snapshots that highlight favourable amylin–GO interactions. The amylin residue sidechains are drawn in licorice atomic detail and coloured based on their properties. **(c)** Top-view of amylin–GO7 showing residues within 0.4 nm of the NF. **(d)** Side-view of amylin interacting with GO3, emphasising the orientation of aromatic and charged residues relative to the substrate.

4. Additional BE-MetaD and SPON-MD Simulations Results

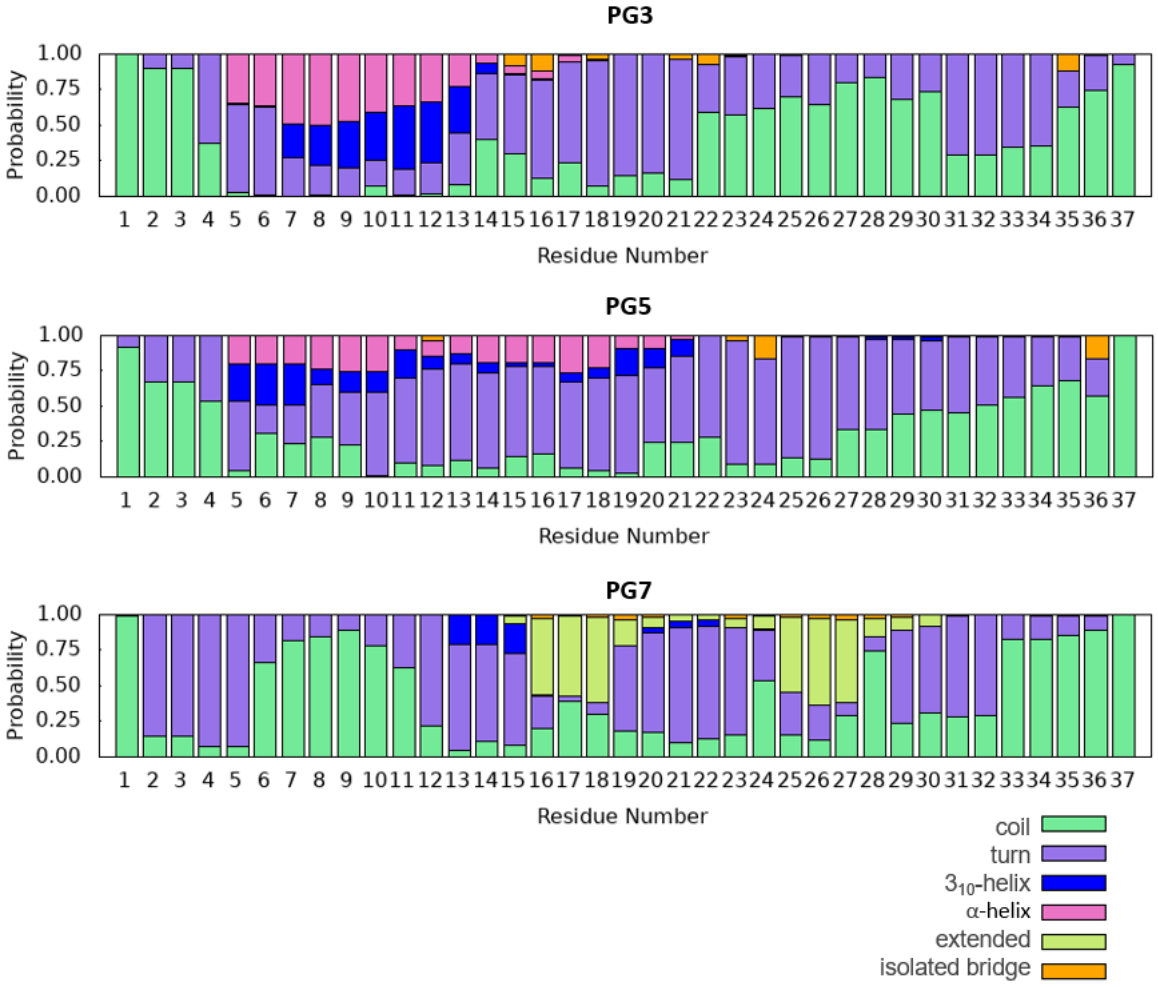

Fig. S15 Per-residue secondary structure probability of amylin determined on the BE-MetaD amylin–PG low energy states.

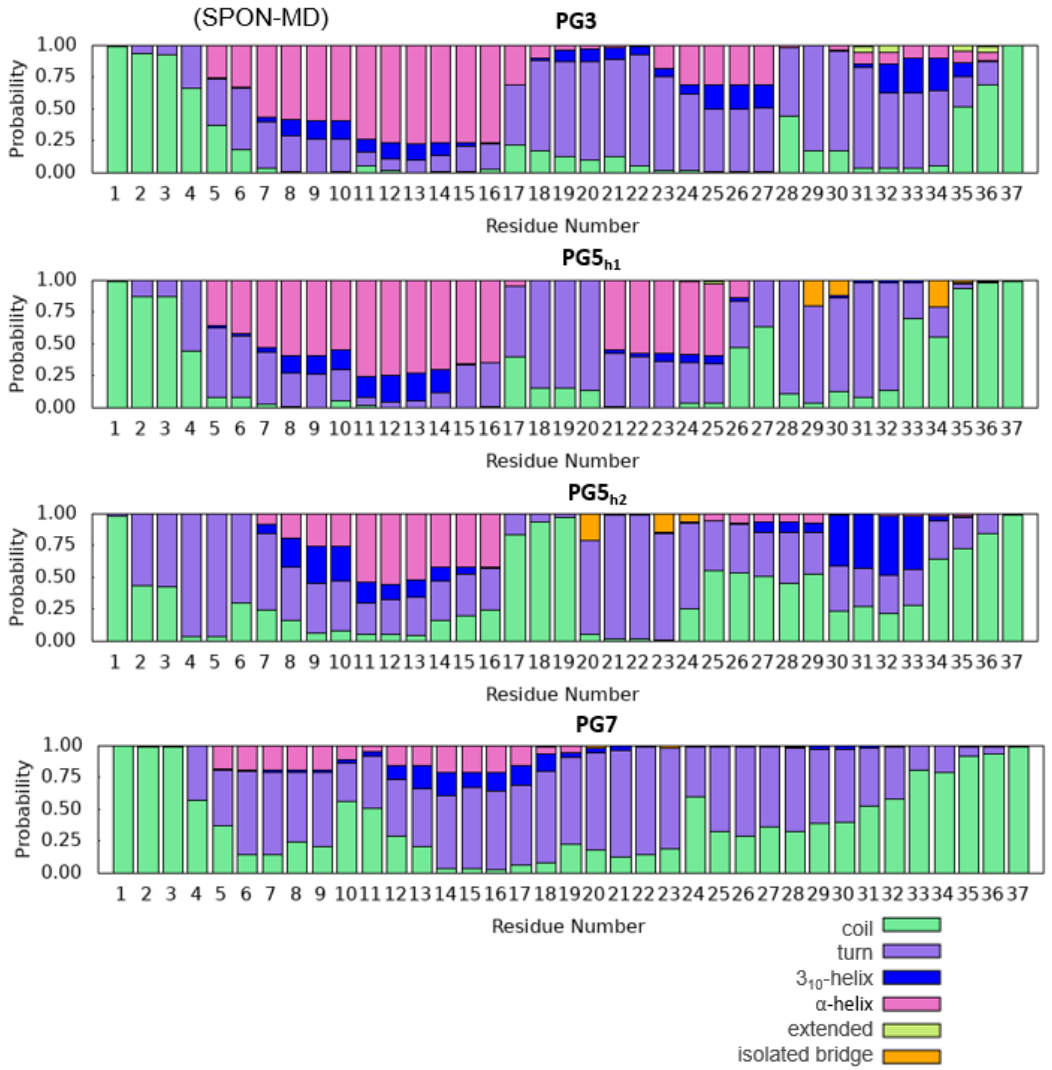

Fig. S16 Per-residue secondary structure probability of amylin determined on the SPON-MD amylin–PG low energy states.

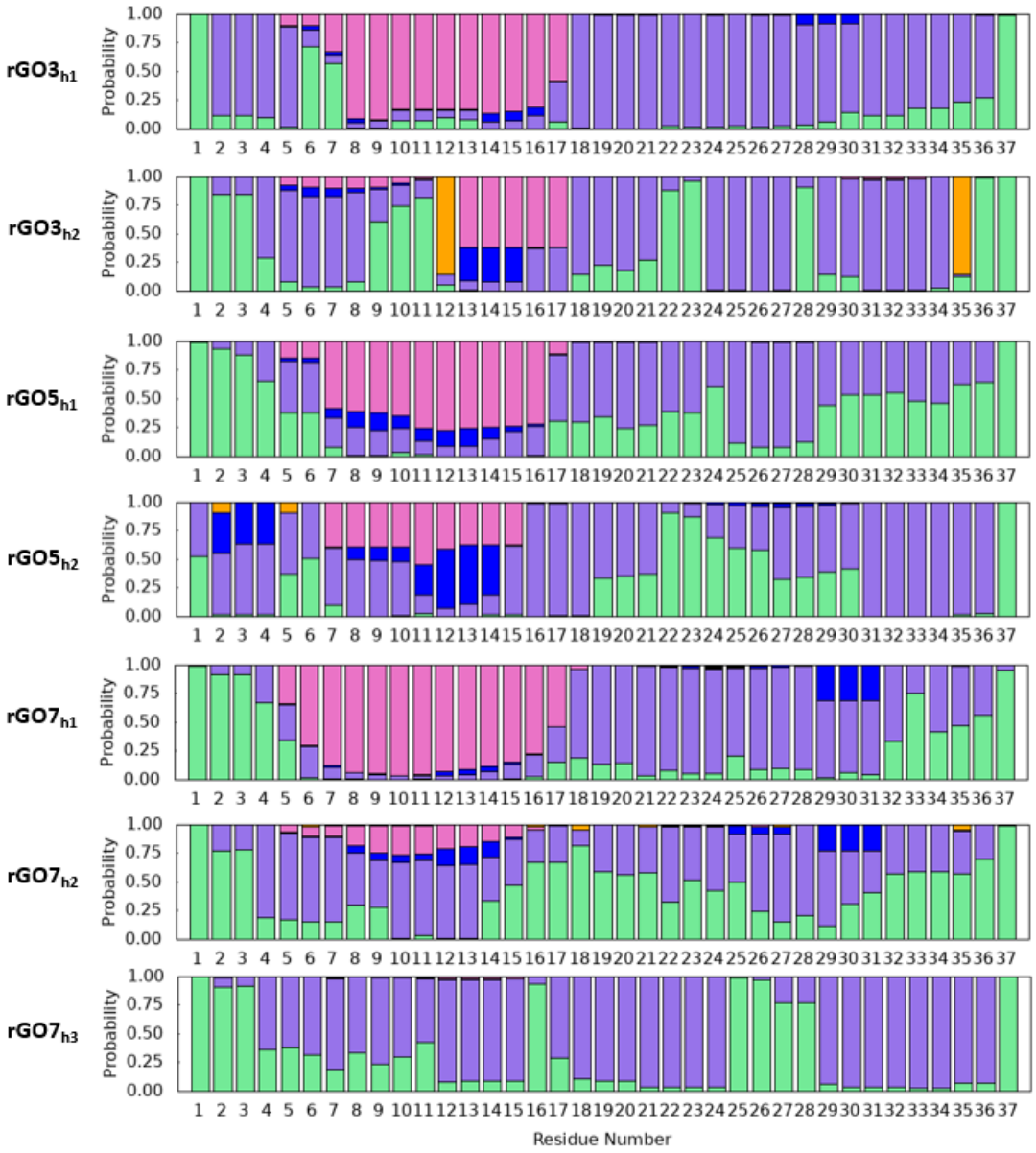

Fig. S17 Per-residue secondary structure probability of amylin determined on the BE-MetaD amylin–rGO low energy states.

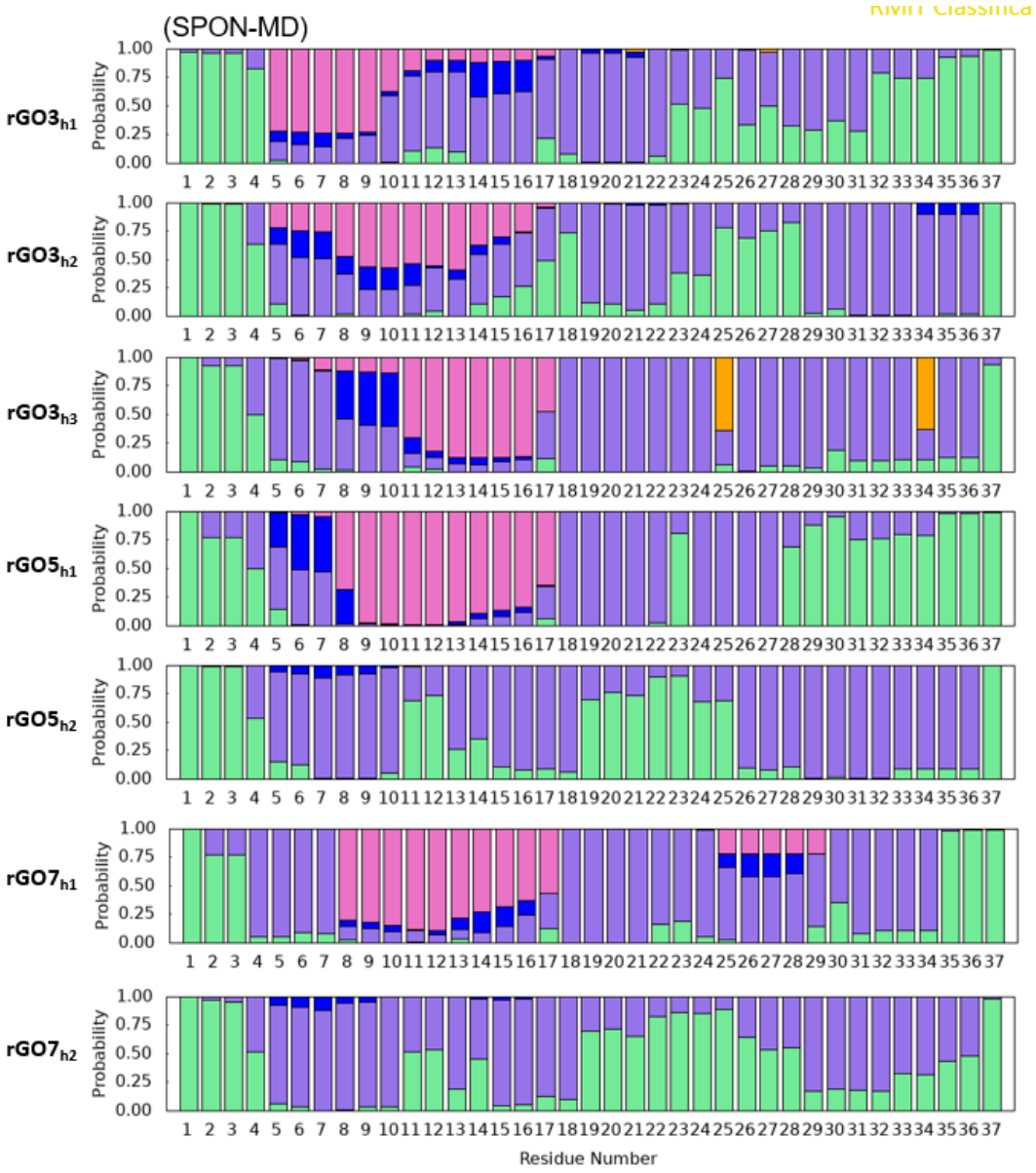

Fig. S18 Per-residue secondary structure probability of amylin determined on the SPON-MD amylin–rGO low energy states.

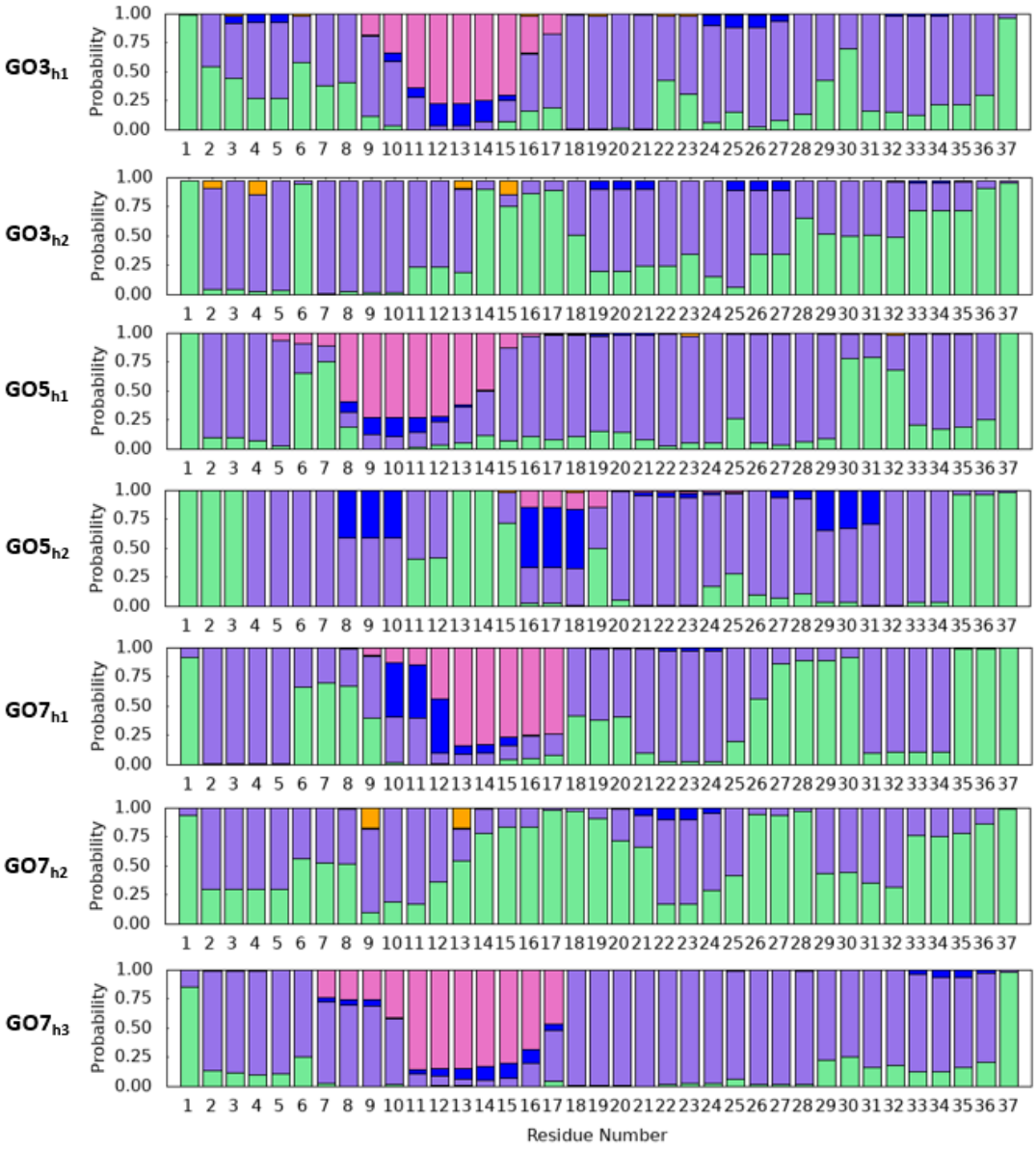

Fig. S19 Per-residue secondary structure probability of amylin determined on the BE-MetaD amylin–GO low energy states.

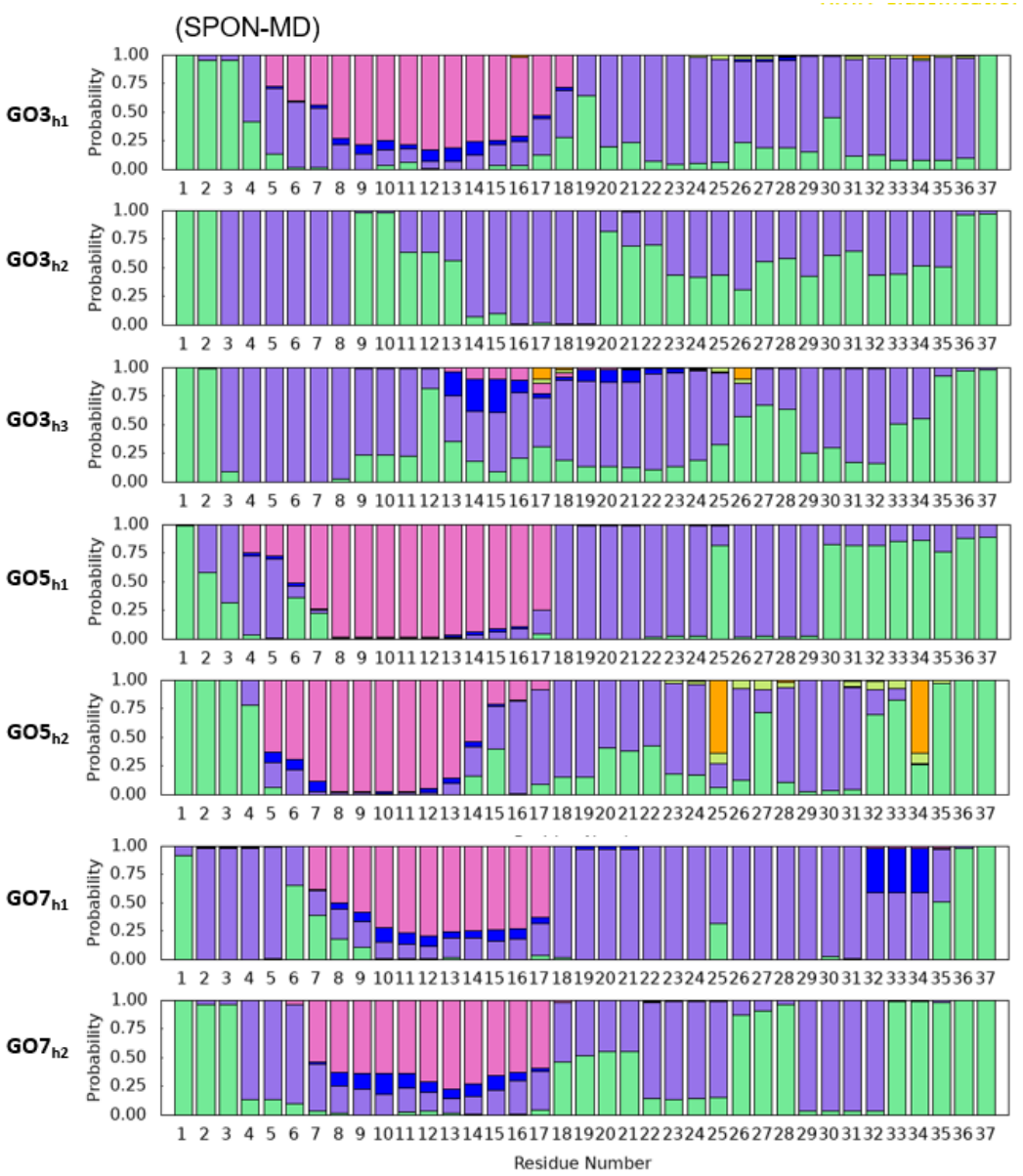

Fig. S20 Per-residue secondary structure probability of amylin determined on the SPON-MD amylin–GO low energy states.

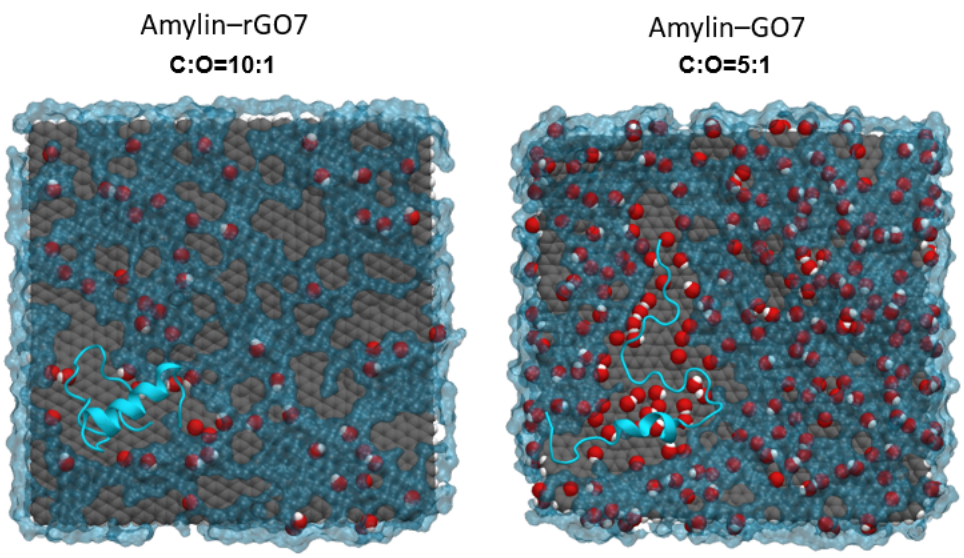

Fig. S21 Differences in nanoflake water structuring on the oxidised nanoflakes. Water atoms within a 0.35 nm distance from the NF are shown in surf representation and coloured in blue. Amylin is drawn in NewCartoon representation and coloured cyan.

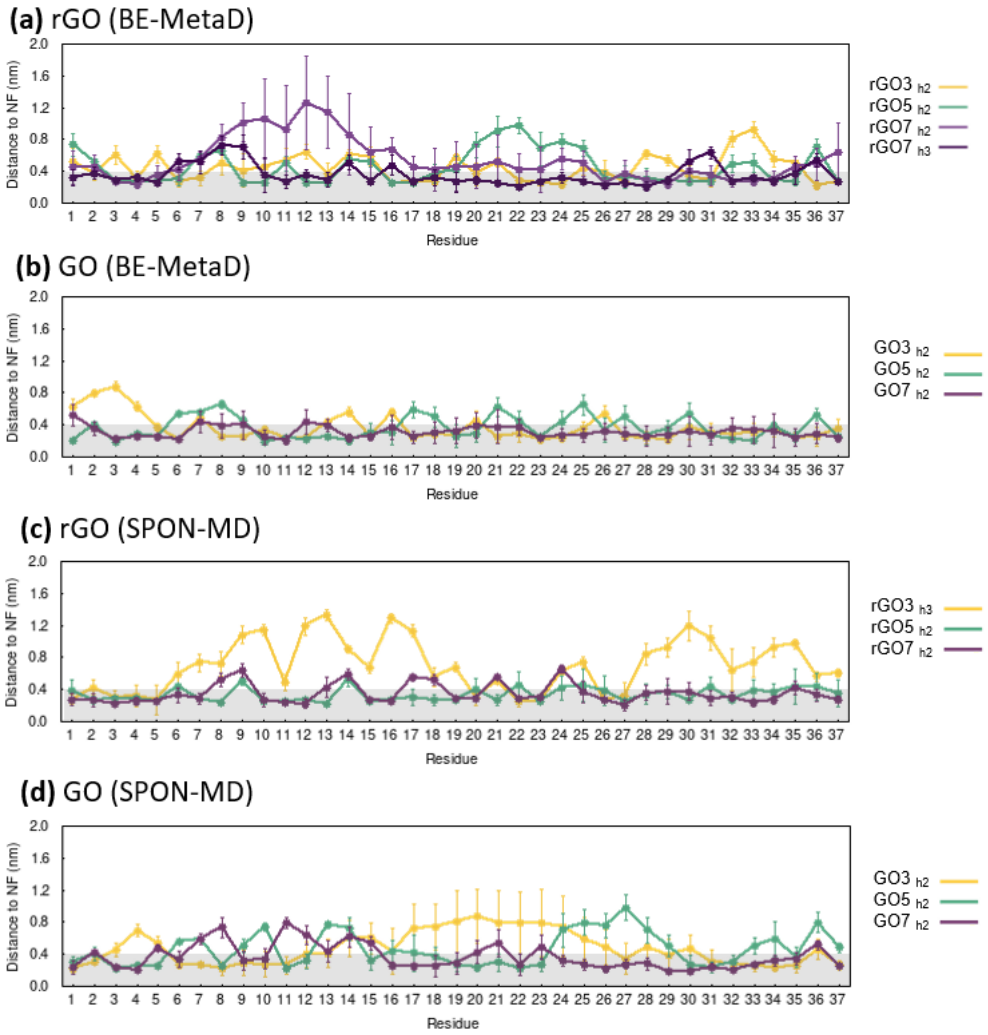

Fig. S22 Average minimum distance of individual protein residues to the NF within the additional low energy states shown in the 2D free energy maps.

References:

1. S. Piana, K. Lindorff-Larsen, R. M. Dirks, J. K. Salmon, R. O. Dror and D. E. Shaw, *PloS One*, 2012, **7**, e39918.
2. E. Peng, N. Todorova and I. Yarovsky, *PLoS One*, 2017, **12**, e0186219.
3. E. Peng, N. Todorova and I. Yarovsky, *ACS Omega*, 2018, **3**, 11497-11503.
4. D. Frishman and P. Argos, *Proteins: Struct. Funct. Genet.*, 1995, **23**, 566– 579.