Supporting Information for

Achieving Regioselective Materials Binding using Multidomain Peptides

*Ruitao Jin,[a, †] Nermina Brljak,[b, †] Robert Sangrigoli,[b] Tiffany R. Walsh,[a, †, *] and Marc R. Knecht[b,c, †, *]*

a. Institute for Frontier Materials, Deakin University, Geelong, 3216 VIC, Australia.

b. Department of Chemistry, University of Miami, 1301 Memorial Drive, Coral Gables, Florida 33146, United States.

c. Dr. J.T. Macdonald Foundation Biomedical Nanotechnology Institute, University of Miami, UM Life Science Technology Building, 1951 NW 7th Ave, Suite 475, Miami, Florida 33136, United States.

† These authors contributed equally.

* To whom correspondence should be addressed: TRW - tiffany.walsh@deakin.edu.au; MRK – knecht@miami.edu

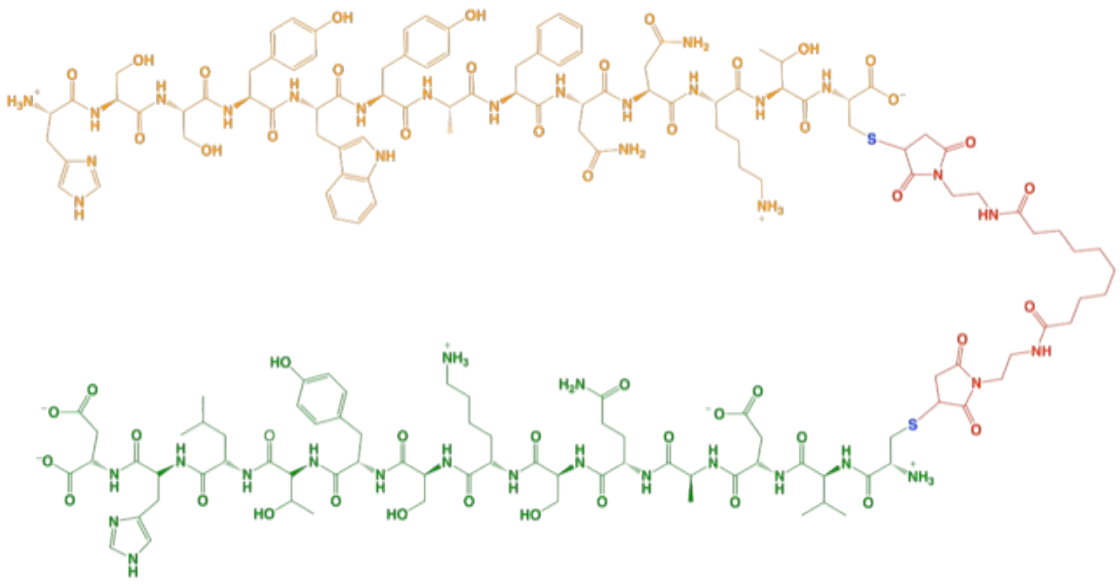

Figure S1. Chemical structure of the BEAM with the P1 domain in orange, the $F_{10}F$ domain in red, and the BP7 domain in green.

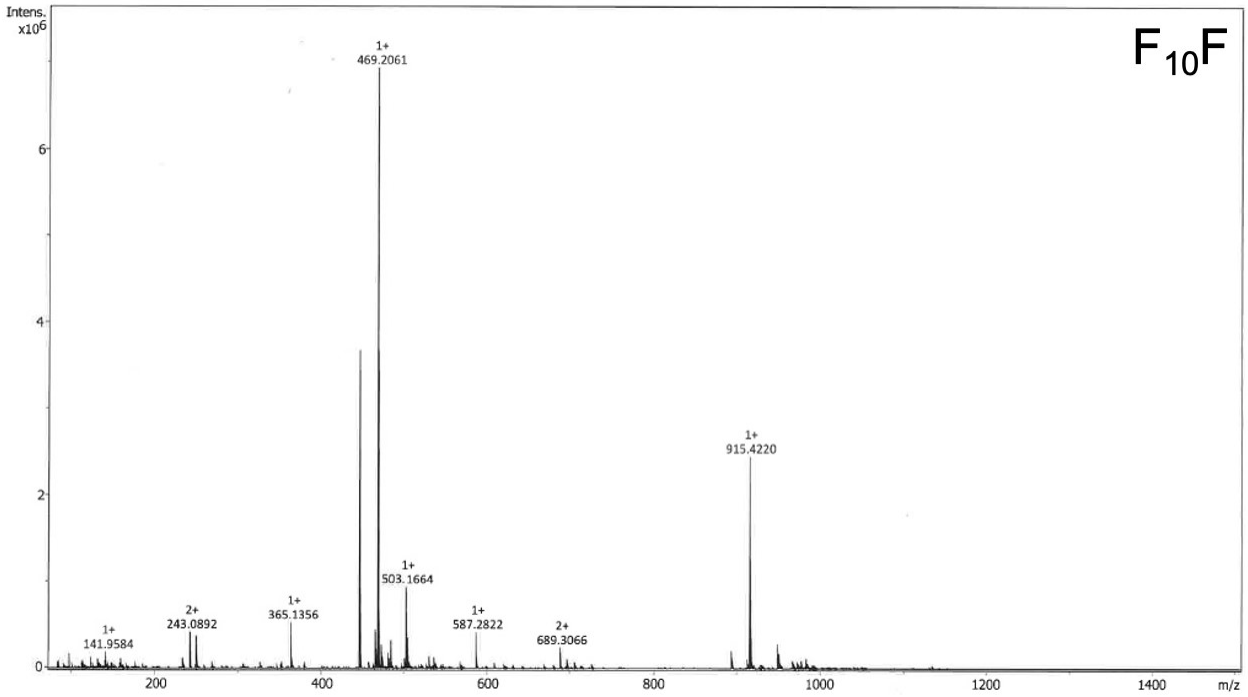

Figure S2. ESI-TOF mass spectrum of synthesized $F_{10}F$ fatty acid spacer.

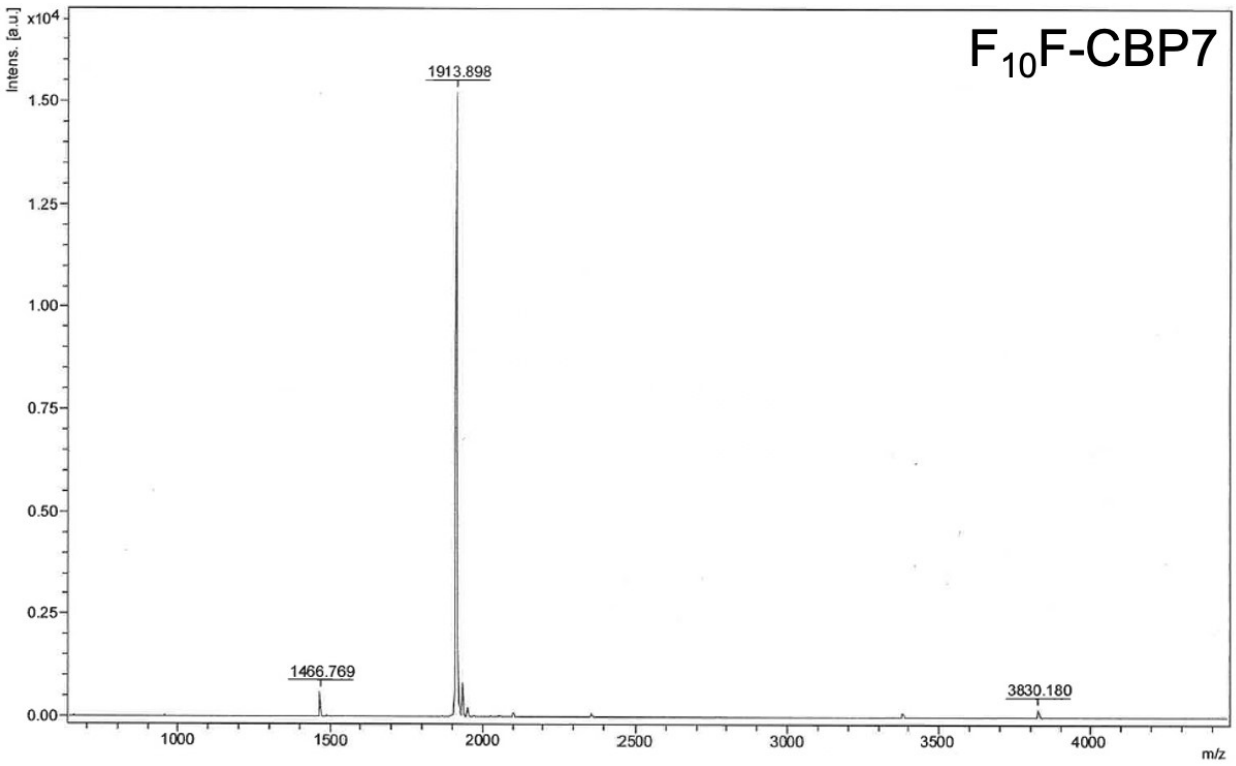

Figure S3. MALDI-TOF mass spectrum of F_{10}F-CBP7.

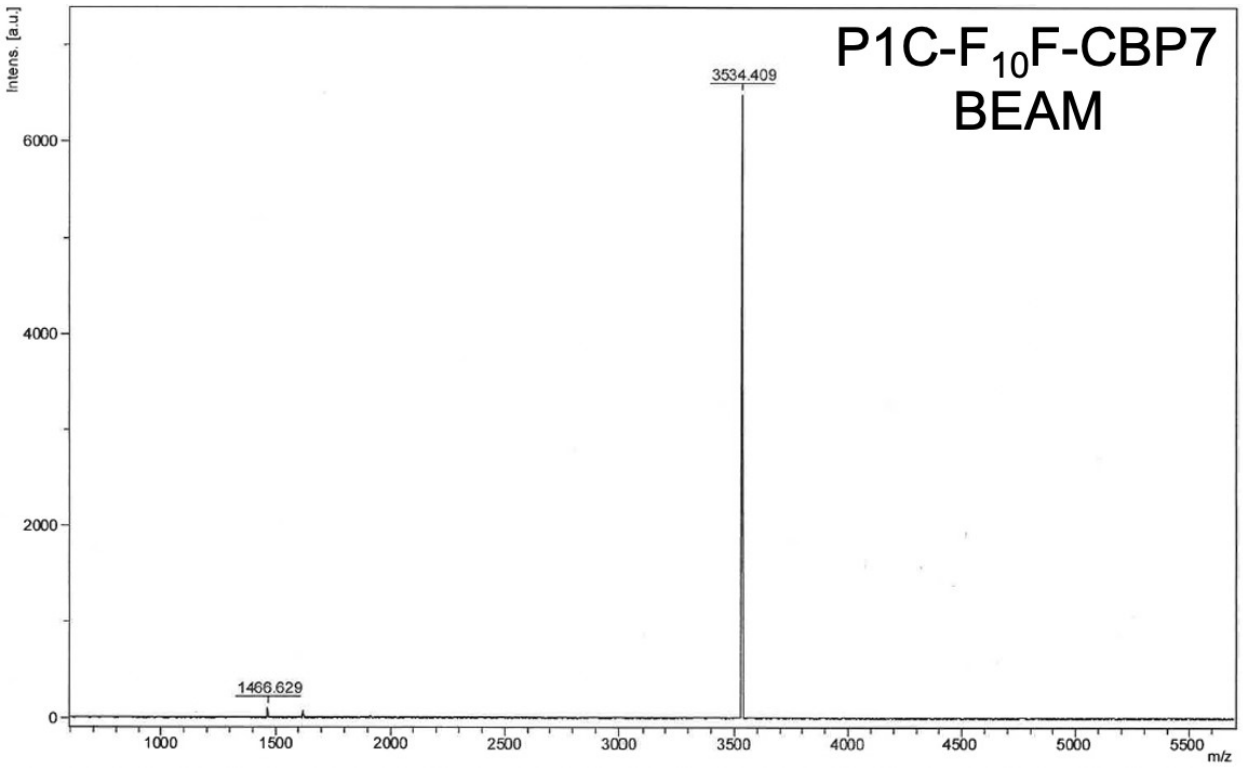

Figure S4. MALDI-TOF mass spectrum of P1C-F_{10}F-CBP7 (BEAM).

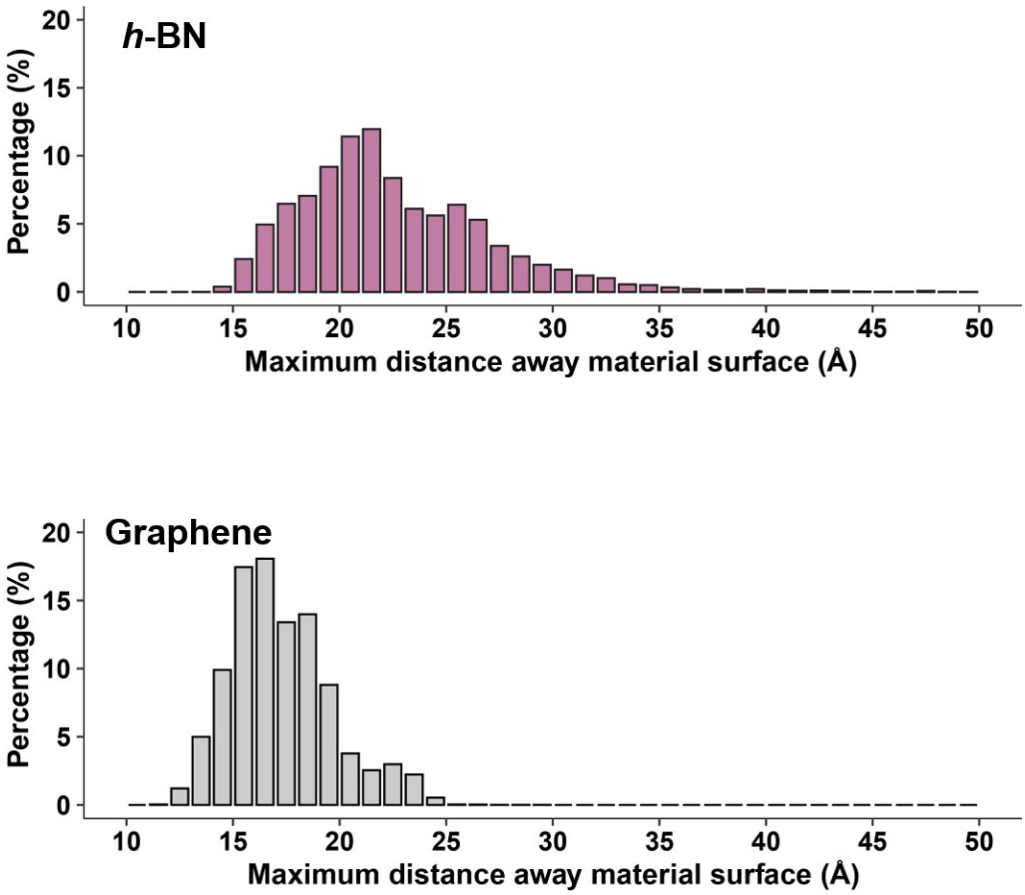

Figure S5. Histogram of the maximal vertical molecule-surface distance of the BEAM in the surface adsorbed state at the aqueous graphene and *h*-BN interfaces.

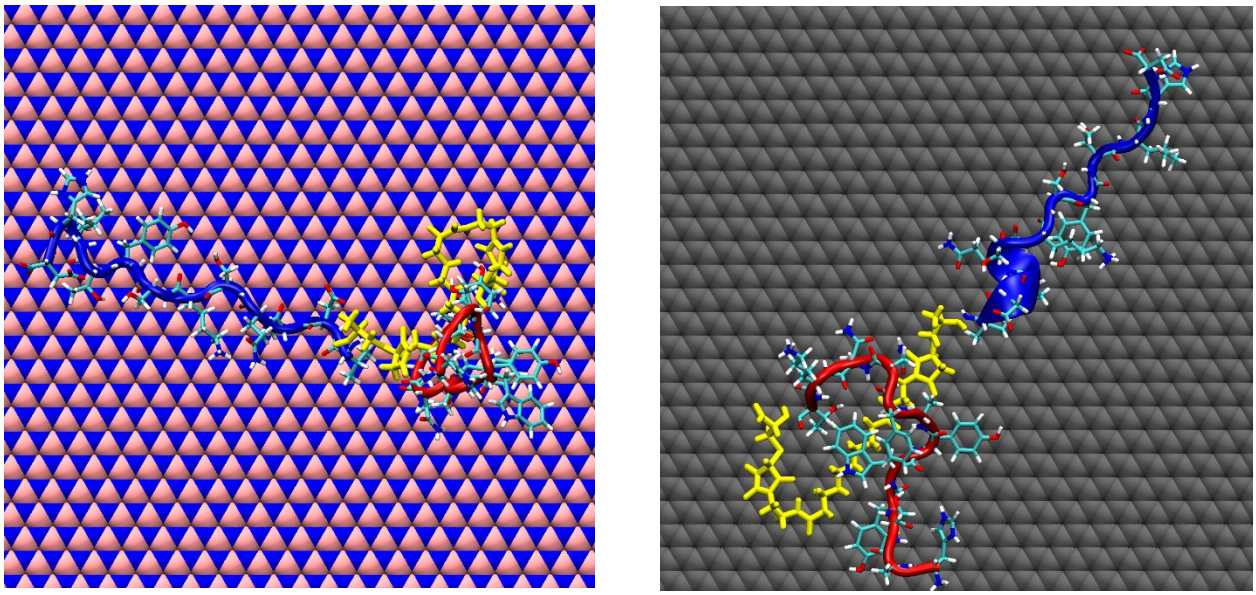

Figure S6. Snapshots of the BEAM molecule adsorbed at the aqueous *h*-BN (left) and graphene (right) interfaces, shown in plan view. Water not shown for clarity.

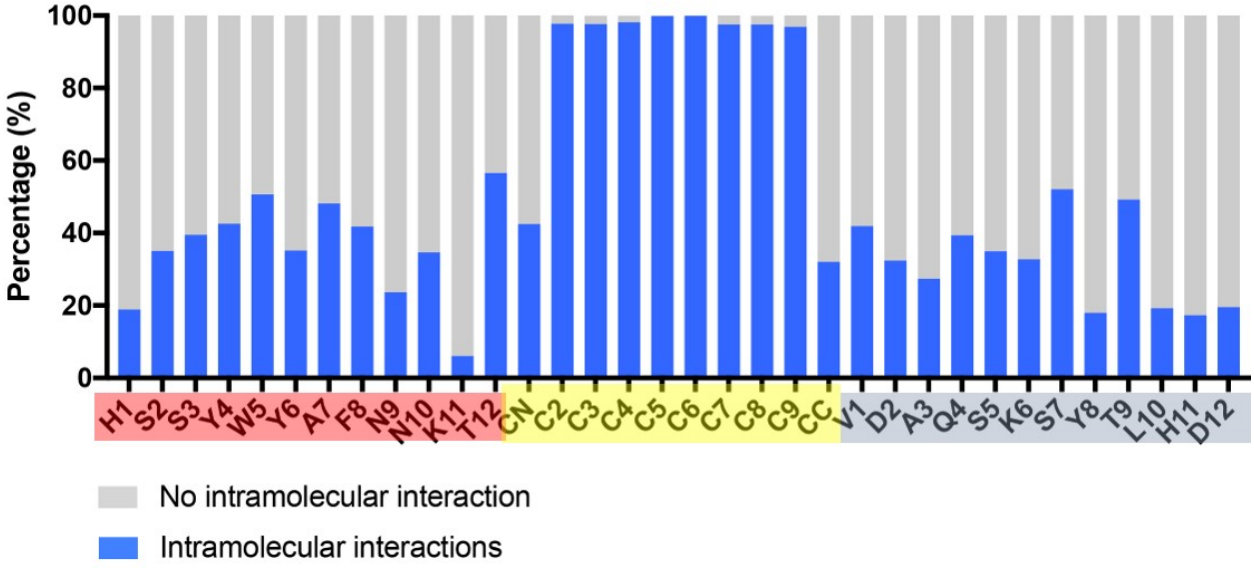

Figure S7. The proportion of the trajectory for which a given residue has any interaction with the rest of the BEAM molecule, indicated by blue bars, in the *h*-BN adsorbed state. The residues indicated on the ordinate are colored by domain (red for the P1 domain, yellow for $F_{10}F$, blue for the BP7 domain).

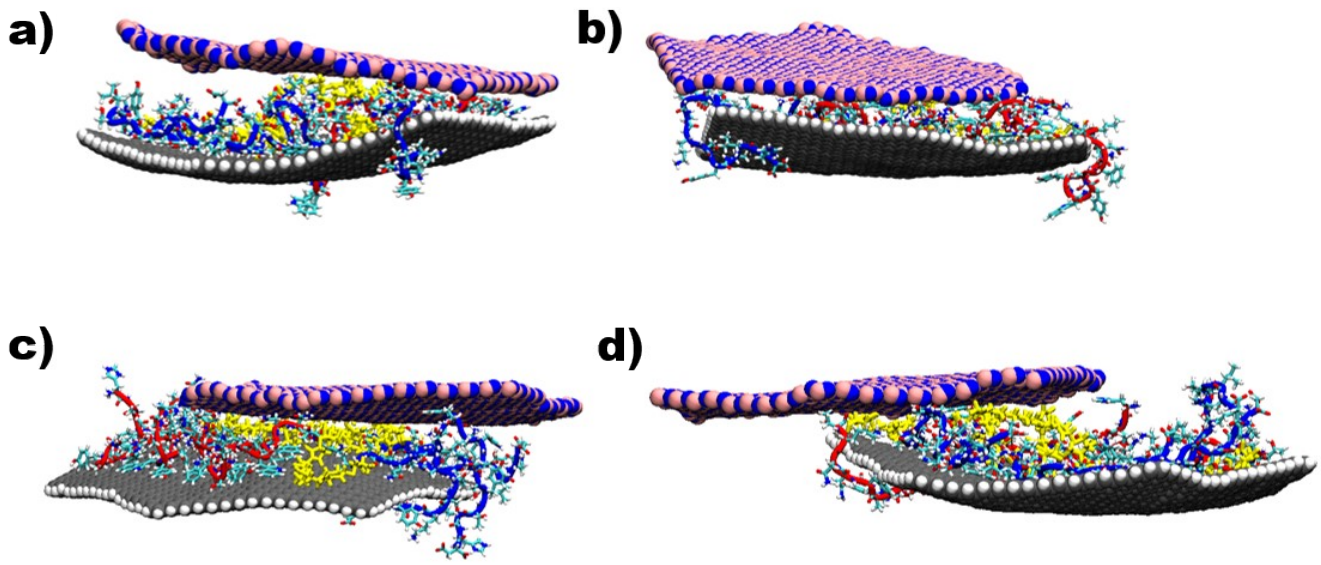

Figure S8. Additional views of the BEAM-mediated graphene/h-BN bilayer in aqueous solution. a) and b): additional views of the snapshots provided in Figure 8 of the main text. c) and d): snapshots from two of the additional independent simulations. Water not shown for clarity.

Computational Details

Replica Exchange with Solute Tempering Molecular Dynamics (REST-MD) Simulations.

General: All simulations were performed using Gromacs software package (version 2021.1)[1]. The simulation system comprised two sheets placed in an orthorhombic periodic simulation cell in a parallel orientation (both perpendicular to the z-direction), at values of z=11nm. For the graphene system the periodic cell had approximate dimensions 8.9nm x 8.9nm x 12.5 nm, and for the *h*-BN system the cell dimensions were 9.0 × 9.5 x 12 nm. This vertical inter-sheet gap (along the z-dimension) was filled with liquid water and counter-ions (where appropriate). All simulations were performed in the Canonical (NVT) ensemble at 300K, using the Nose-Hoover thermostat[2]. The leap-frog algorithm[3] was used to integrate the equations of motion using a timestep of 1 fs. The Verlet grid cut-off scheme[4] was applied for neighbor searching. The PME algorithm[5] was used for evaluating long-ranged electrostatic interactions. A tested force field combination was used, comprising the CHARMM22* force field for peptides and the fatty acid linkage, as reported previously[6–8], the TIPS3P water model[9], the polarizable GRAPPA force-field[10] for graphene, and the BoNi-CHARMM force field[11] for *h*-BN. During the simulations, all atoms in the 2D nanosheets were held fixed in space during these simulations, except those dipoles belonging to the graphene substrate, which were able to freely rotate. Each simulation comprised two 2D periodic substrates, liquid water (approx. 27000 or 30000 water molecules for graphene and *h*-BN), counter-ions, and one BEAM molecule.

REST-MD Simulation Details: Full technical details of the Terakawa implementation[12] of the REST approach have been given previously[13]. In total 16 replicas were used with an 'effective temperature' window of 300-430K. The REST-MD simulation trajectories were of 50 ns duration (this yields 16× 50 ns = 0.8 μs of aggregate simulation time per simulation – amounting

to 1.6 μs of total simulation time for the two REST-MD simulations). The 16 initial replica structures included a wide range of secondary structure motifs, including α-helices, β-turns, and polyproline II conformations, as well as random coil structures. Prior to initiation of each REST-MD simulation, the 16 initial configurations were equilibrated at their target potential for 0.5 ns, with no exchange moves attempted during this period. After this point, the time interval between exchange attempts was 1 ps. The 16 values of lambda used to scale our force-field were: $\lambda_i = 0.000$, 0.057, 0.114, 0.177, 0.240, 0.310, 0.382, 0.458, 0.528, 0.597, 0.692, 0.750, 0.803, 0.855, 0.930, 1.000. Frames were saved every 1 ps.

Analysis: The clustering analysis was done using the positions of the entire molecule "backbone" (i.e. all peptide backbone atoms and all additional N/C atoms in the $F_{10}F$ linker). This analysis provides insights into the most populated structures in the conformational ensemble of the entire molecule, including the fatty acid chain. The Daura algorithm[14] was used for the clustering, with a cutoff in the root mean-squared deviation (RMSD) of atomic positions of 4.0 Å for the entire "backbone" clustering, as determined from test simulations.

Similar to previous studies, the conformational entropic contribution, S_{conf}, was estimated for each of the conjugated molecules in the surface-adsorbed state. This estimate was based on the calculation of the discrete entropy of the distribution of cluster populations. The larger the value of S_{conf}, the greater the conformational entropic contribution to binding. These values can be considered together with a measure of the enthalpic contribution to binding (described below), to provide an overall binding assessment that can be considered alongside the experimentally-determined binding free energies. The discrete entropy, set here to be equivalent to S_{conf}, was calculated using:

$$S_{conf} = -\sum_{i=1}^{N_c} p_i \log(p_i)$$

where N_c is the total number of clusters, and p_i is the population (relative fraction) of the i^{th} cluster, expressed as a value on the interval [0,1].

Table S1. Summary of reference sites and cutoff distances used for determining residue-surface contact.

Residue	Reference Site and Cutoff Distance (Å) on Graphene	Reference Site and Cutoff Distance (Å) on *h*-BN
Ala	CB, 5Å	CB, 4Å
Asp	CG, 5Å	CG, 5Å
Asn	ND2, 4.5Å	ND2, 4.5Å
Cys	SG, 4.5Å	SG, 4.5Å
Gln	NE2, 4.5Å	NE2, 4.5Å
His	Center of mass of ring, 4.5Å	Center of mass of ring, 5Å
Leu	CG, 5Å	CG, 5Å
Lys	NZ, 6Å	NZ, 5Å
Phe	Center of mass of ring, 4Å	Center of mass of ring, 4Å
Ser	OG, 4.25Å	OG, 4Å
Thr	OG1, 4.25Å	CG2, 4Å
Trp	Mid of central bond, 3.7Å	Mid of central bond, 4Å
Tyr	Center of mass of ring, 4Å	Center of mass of ring, 4Å
Val	CB, 5Å	CB, 5Å

Residue-surface contact analyses counted the number of frames in each REST-MD reference trajectory that a specific residue (or carbon in the $F_{10}F$ linker) was deemed to be in "contact" with graphene or *h*-BN surface, based on a vertical distance cut-off criteria. Reference sites and cut- off values for each residue are provided in Table S1. The reference site for the methylenes in the $F_{10}F$ linker used the carbon site with the same cutoff as used for Ala. The contact data are expressed as a percentage of the REST-MD trajectory for which each residue is in surface contact, such that a contact value of 100% would mean the residue was in contact with the surface for the entire trajectory, whereas a contact value of 0% would mean the residue was never in contact with surface for the entire trajectory.

The enthalpic binding score was determined as the weighted average of the product of the residue-surface contact for each residue (*i*), w_i (expressed as a fraction between 0 and 1 instead of a percentage), and the calculated free energy of binding of the corresponding amino acid, F_i, at either the aqueous graphene interface or the aqueous *h*-BN interface, as determined from previous work[11,15]. Again, the methylene units in the $F_{10}F$ linker were correspondingly treated as Ala for the purposes of this analysis.

The intra-molecule interactions were quantified in a similar fashion to the residue-surface interactions. In this case, two residues (or methylenes in the $F_{10}F$ linker) were determined to be in contact if their reference sites were found within a given distance cutoff. Here, the cutoff was chosen to be 5 Å between the reference sites selected in Table S1. Possible neighboring residues (i+1, i-1) were excluded for residue i in peptide domains if residues (i+1, i-1) belong to peptide domain, possible residues (i+1,i+2, i-1, i-2) were excluded for residues i in the $F_{10}F$ domain if residues (i+1,i+2, i-1, i-2) belonged to the $F_{10}F$ domain as well.

The BEAM orientation was also analyzed, to detect upright conformations of the BEAM in the adsorbed state. To do this, the maximal vertical distance (in the z direction, perpendicular to the plane) between the BEAM atoms and the surface plane was identified in each frame of the trajectory and histogrammed (over 50,001 frames).

References:

(1) M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *Softwarex,* 2015, **1**, 19–25.

(2) D. J. Evans, and B. L. J. Holian, *Chem. Phys.* 1985, **83**, 4069–4074.

(3) W. F. V. Gunsteren, and H. J. C. Berendsen, *Mol. Simulat.,* 1988, **1**, 173–185.

(4) S. Páll, and B. Hess, *Comput. Phys. Commun.,* 2013, **184**, 2641–2650.

(5) T. Darden, D. York, and L. J. Pedersen, *Chem. Phys.,* 1993, **98**, 10089–10092.

(6) A. D. M. Jr, D. Bashford, M. Bellott, R. L. D. Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. R. III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B,* 1998, **102**, 3586–3616.

(7) S. Piana, K. Lindorff-Larsen, and D. E. Shaw, *Biophys. J.,* 2011, **100**, L47–L49.

(8) J. P. Palafox-Hernandez, C.-K. Lim, Z. Tang, K. L. M. Drew, Z. E. Hughes, Y. Li, M. T. Swihart, P. N. Prasad, M. R. Knecht, and T. R. Walsh, *ACS Appl. Mater. Inter.,* 2016, **8**, 1050–1060.

(9) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.,* 1983, **79**, 926–935.

(10) Z. E. Hughes, S. M. Tomásio, and T. R. Walsh, *Nanoscale,* 2014, **6**, 5438–5448.

(11) A. Budi, and T. R. Walsh, *Langmuir,* 2019, **35**, 16234–16243.

(12) T. Terakawa, T. Kameda, and S. Takada, *J. Comput. Chem.,* 2011, **32**, 1228–1234.

(13) L. B. Wright, and T. R. Walsh, *Phys. Chem. Chem. Phys.,* 2013, **15**, 4715–4726.

(14) X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, *Angew. Chem. Int. Ed.,* 1999, **38**, 236–240.

(15) Z. E. Hughes, and T. R. Walsh, *J. Mater. Chem. B,* 2015, **3**, 3211–3221.