Electronic Supplementary Information for

Selective Manipulation of Peptide Orientation on Hexagonal Boron Nitride Nanosheets

Nermina Brljak[1], Ruitao Jin[2], Tiffany R. Walsh,[2,,‡] Marc R. Knecht[1,3,*,‡]*

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

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

3. Dr. J.T. Macdonald Foundation Biomedical Nanotechnology Institute, University of Miami, 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

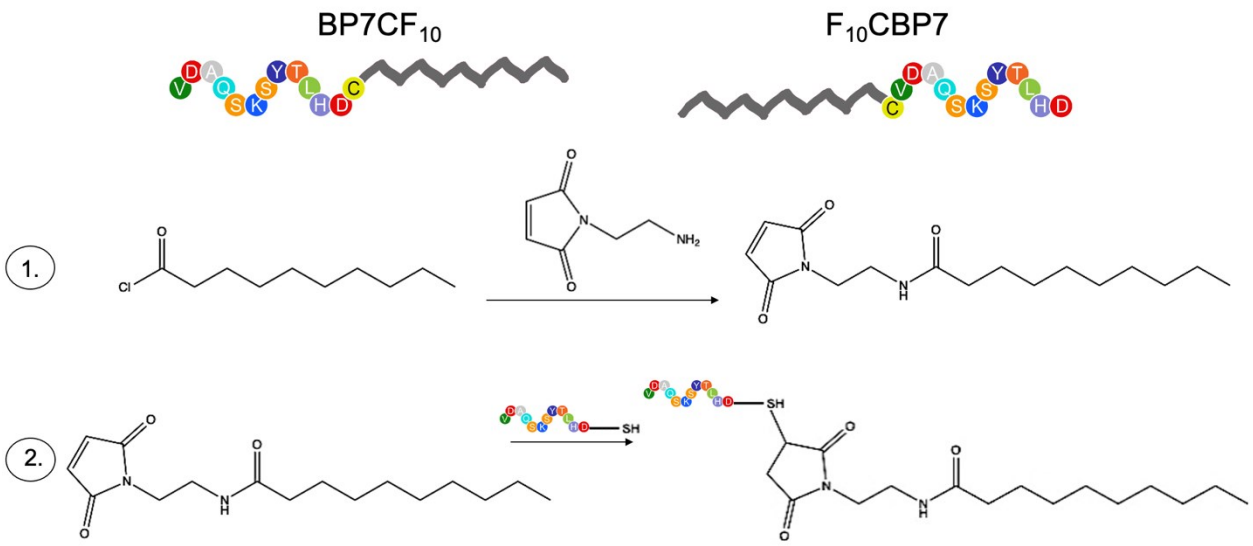

Figure S1. Synthesis scheme to (1) prepare the maleimide-containing fatty acid and (2) couple the fatty acid into the cysteine-appended BP7 peptide.

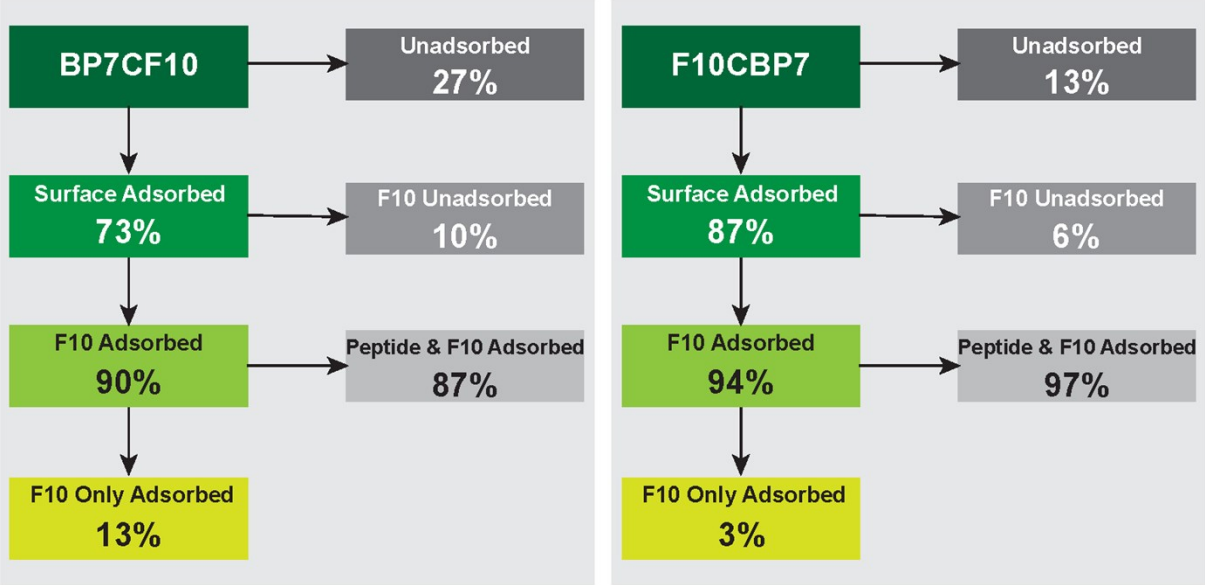

Figure S2. Breakdown of peptide/fatty acid contact with the *h*-BN surface, as predicted from the REST-MD simulations. Percentages reflect the number of frames in each category, with respect (ultimately) to the total number of frames in the REST-MD trajectory.

Peptide Orientation Analysis

Each of the REST-MD simulation trajectories for $BP7CF_{10}$ and $F_{10}CBP7$ was analyzed to capture the number of upright surface-adsorbed states, as detailed in 'Computational Details' (provided herein). This analysis used a distance-based cut-off approach, identifying the furthest atom from the surface (in terms of vertical distance), for each frame in the REST-MD reference trajectory. This analysis revealed that 43% of the $BP7CF_{10}$ surface-adsorbed states were in an upright configuration, compared with 13% of the corresponding $F_{10}CBP7$ surface-adsorbed states.

Once these upright states were identified via this cut-off and counted, their structures were characterized. From visual inspection, many of the upright states appeared to be mediated by fatty acid contact with the surface, with the peptide mostly detached from the surface. To probe this further, a breakdown of each type of surface-adsorbed state, in terms of the contact between each of the peptide and fatty acid regions and the *h*-BN surface was performed, provided in Figure S1. These data, along with the distance-based analysis described above, along with visual inspection, confirm that the F_{10}-only adsorbed states (where only the fatty acid makes *exclusive* contact with the surface, comprising 13% for $BP7CF_{10}$, and 3% for $F_{10}CBP7$) were entirely categorized as upright conformations. However, these F_{10}-only adsorbed states comprise approximately one-quarter to one-third of all upright states as determined from the distance-based analysis. Further analysis revealed the remainder of the F_{10}-adsorbed states to arise from chiefly F_{10}-mediated contact, assisted by one or two additional peptide residue contacts, as summarized in Figure S3.

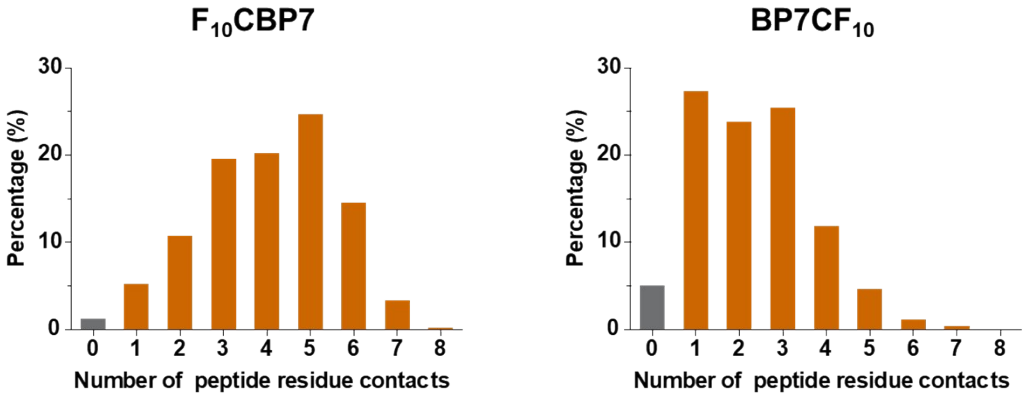

Figure S3. Distribution of additional peptide residue surface contacts for all F_{10}-adsorbed contact states, predicted for both $F_{10}CBP7$ and $BP7CF_{10}$.

These data reveal the origins of the difference in the relative proportion of upright and horizontal adsorbed states in each case. For $BP7CF_{10}$, the number of additional residues assisting the fatty acid in terms of surface contact is skewed to low numbers (1-3 residues), whereas the opposite is true for $F_{10}CBP7$. Further characterization regarding the identities of the specific residues that are responsible for this assisted-upright contact is presented in Figure S4, and provide deeper understanding of the origins of the adsorption orientation control.

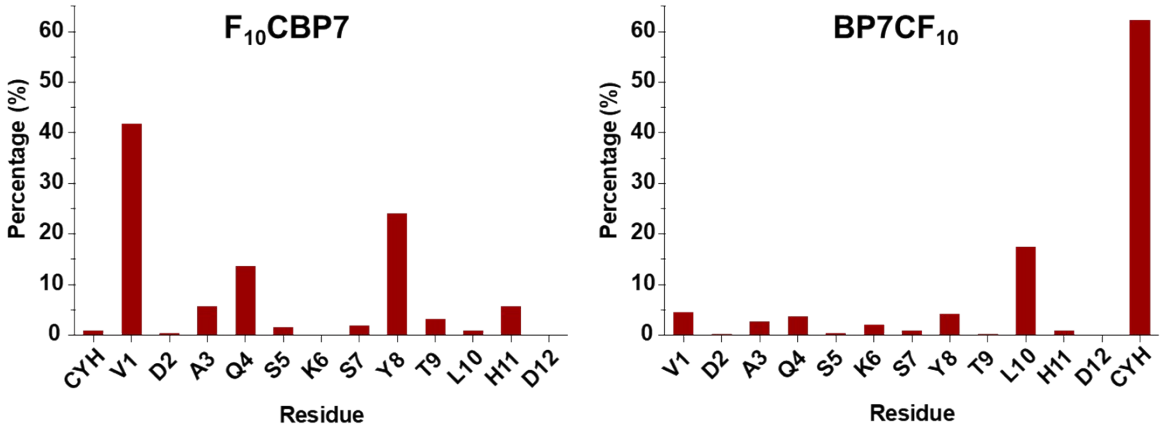

Figure S4. Breakdown of the degree of contact (expressed as a percentage of trajectory frames) for additional peptide residues assisting the F_{10}-adsorbed states, determined for the F_{10}-adsorbed states of F_{10}CBP7 and BP7CF_{10}.

Specifically, Figure S4 indicates that the assisting peptide residues for BP7CF_{10} are dominated by residues that are closest to the F_{10} moiety attachment point (Cys13, Leu10). The contact for Tyr8 (which is an anchor in the parent BP7 case) is substantially down-modulated, leading to a large number of non-adsorbed states for the segment of the peptide covered by residues 1-9. This facilitates the higher proportion of upright states observed for BP7CF_{10}. In contrast for F_{10}CBP7, not only do more residues provide greater assistance, but crucially the Tyr8 contact is enhanced relative to BP7CF_{10}. As a consequence, the peptide segment comprising residues 1-9 experiences a greater degree of pinning to the surface, leading to a greater proportion of horizontal adsorbed states.

In closing, the cause(s) for the down-modulation of the Tyr8 contact for the F10-adsorbed states are currently proposed to be due to a proximity effect. Tyr8 has been previously identified as a surface anchor for the BP7 peptide. The separation between Tyr8 and the F10 attachment point appears to be a factor in the Tyr8 binding modulation. When attached at the C-terminus, F10 is closer to Tyr8 (6 residue spacing) compared with the case when attached at the N-terminus (9 residue spacing). It is conceivable that 9 residues is too distant for the down-modulation influence of F10 to be sufficiently exerted. Further investigations will probe this hypothesis in future.

Clustering Analysis

A clustering analysis uses a measure of structural similarity to classify the conformational ensemble into groups of "like structures" (denoted clusters). Clustering yields two key outcomes: the set of unique types (or classes) of structures (clusters) in the conformational ensemble, and, the relative population of each cluster in the ensemble. The total number of clusters signifies the number of thermally-accessible conformational states, and a high total number of clusters can therefore be indicative of a substantial degree of conformational disorder, provided that the most populated clusters are not dominant (e.g. more than ~50% of the population).

The clustering procedure is described in the Computational Details section. In brief, the clustering was based on the positions of the backbone atoms of the peptide, plus the "backbone" of the fatty acid chain (54 atoms in total). The populations of the top five most populated clusters are

summarized along with the total number of clusters for both bioconjugates in the surface-adsorbed state.

Table S1. Populations of the top-5 most populated clusters in the surface adsorbed state, determined over the entire molecular backbone. The total number of clusters is given in parentheses.

Cluster Rank	$BP7CF_{10}$	$F_{10}CBP7$
1	4	7
2	3	5
3	2	5
4	2	5
5	2	4
	(518)	(266)

The difference in the total number of clusters (Table S1) is remarkable, and signifies a very substantial difference in the degree of conformational disorder in terms of the number of thermally-accessible conformational states for $BP7CF_{10}$. This high total number of clusters, 518, is consistent with the prevalent upright binding configuration, in which many different F10-adsorbed (assisted with mostly 1-2 F10-proximal peptide residues) configurations are possible, relative to the more surface-pinned $F_{10}CBP7$ molecule. The percentage populations of the top-five most populated clusters are also uncharacterisitically low, compared with previous studies of fatty-acid conjugated materials binding peptides. [1]

Computational Details

Set-Up: Replica Exchange with Solute Tempering Molecular Dynamics (REST-MD) simulations [2-3] were used to predict the Boltzmann-weighted conformational ensemble for the three Ala mutants (BP7 V1A, BP7 D2A and BP7 Y8A), and the two fatty acid bioconjugates $BP7CF_{10}$ and $F_{10}CBP7$, under aqueous conditions in the presence of the *h*-BN surface. Two periodic *h*-BN surfaces were placed in an orthorhombic periodic simulation cell of lateral dimensions ~7.5 × ~6.5 nm, with a cell dimension vertical to the *h*-BN plane of 12 nm, with a vertical spacing (between the parallel h-BN sheets) of 8 nm. This 8 nm vertical space between the two sheets was filled with liquid water, along with one chain of the relevant biomolecule. Periodic boundary conditions were applied in all three principal directions. All simulations were performed in the Canonical (NVT) ensemble, at a thermal temperature of 300K, maintained using the Nosé-Hoover thermostat,[4,5] with a coupling constant of τ = 0.2 ps. Newton's equations of motion were solved using an integration time-step of 1fs. Coordinates were saved every 1ps. Long-ranged electrostatic interactions were treated using Particle-mesh Ewald (PME),[6] with a cut-off at 11 Å, whereas a force-switched cut- off, starting at 9 Å and ending at 10 Å was used for the Lennard-Jones non-bonded interactions.

All atoms in the *h*-BN sheet were held fixed in space during these simulations. Tests indicate that there is very little difference between binding obtained using a rigid substrate, vs. using a slab where all atoms can move.[7] Similar to our procedure reported previously for interfacial REST-

MD simulations of peptide bio-conjugates,[1,8] we used 16 replicas and produced trajectories of 20 ns duration (amounting to 16 × 20 ns = 0.32 μs of nominal total simulation time). The initial structures of the 16 replicas included a wide range of secondary structure motifs, including α-helices, β-turns, polyproline II, and 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 time. During the REST simulations, the interval between exchange attempts was set to every 1 ps.

The Gromacs software package, v.2019 was used. Full technical details of the Terakawa implementation[2] of REST have been given by us previously.[3] In these REST simulations, an 'effective temperature' window of 300-430K was used, spanned by 16 replicas. The adsorbate structure for each replica was initially placed within ~5Å distance from the top surface of the *h*-BN sheet. The 16 values of lambda used to scale our force-field were:
λ_j = 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.

A previously-tested force-field combination of CHARMM22*[9,10] was used for the peptides (with parameter modifications to describe the maleimide-mediated fatty acid linkage, as reported previously) and the BoNi-CHARMM force-field for *h*-BN,[11] along with the modified TIP3P force-field for water.[12,13] The CHARMM22* force-field (as opposed to the much older CHARMM22 parametrization) has a demonstrated track record of providing peptide conformational data that are consistent with experimental findings[14,15] for materials-specific dodecapeptides such as those studied here.

Analysis: To characterize the conformational ensemble produced by the REST-MD trajectories, a clustering analysis over the positions of the peptide backbone atoms was used. The outcomes of this analysis identify the key peptide backbone conformations present in the ensemble and their relative prevalence in the ensemble. Clustering over the peptide backbone atoms of the conjugated molecules was chosen to ensure a like-for-like comparison with the parent BP7 peptide, and also to specifically probe how the F_{10} attachment point affected the peptide conformation. Additional clustering analyses were carried out over the entire molecule "backbone" (i.e. peptide backbone and all additional N/C atoms in the F_{10} chain) for the N- and C-conjugated moieties. This latter analysis provides insights into the most populated structures in the conformational ensemble of the entire molecule, including the F_{10} chain. The Daura algorithm[16] was used for the clustering, with a cutoff in the root mean-squared deviation (RMSD) in backbone atom positions of 2.5 Å for entire "backbone" clustering of the bioconjugate, as determined in previous studies. [1]

Similar to previous studies,[1,17] 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 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].

Two types of contact analysis were performed. In the first type, a distance based cut-off was applied to determine if a molecule was deemed in "surface contact" state or an "unadsorbed" state. This was accomplished by calculating the vertical distance between the surface and each atom in the molecule in each frame of the trajectory. If any atom was located within 5 Å of the surface, the frame was counted as "surface adsorbed". The total number of "in contact" frames was divided by the total number of frames, multiplied by 100 to give the "in contact" percentage. The breakdown provided in Figure S2 was similarly obtained, except in this instance the range of atoms under consideration was restricted to either the peptide or the fatty acid regions of the molecule.

The second analysis was a residue-surface contact analyses that counted the number of frames in each REST-MD reference trajectory that a specific residue (or F10 chain) was determined to be in "contact" with *h*-BN, based on a vertical distance cut-off criterion. Reference sites and cut- off values for each residue can be found in previous work.[18] 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 *h*-BN for the entire trajectory, whereas a contact value of 0% would mean the residue was never in contact with *h*-BN sheet for the entire trajectory.

Finally, a peptide orientation analysis was performed. This analysis was done to count the number of surface-adsorbed states in a trajectory that corresponded to an upright binding configuration. Guided by (and confirmed with) visual inspection, a distance-based cut-off categorization was implemented. In this process, the vertical distance between each peptide atom and the *h*-BN surface was calculated and the furthest atom-surface distance was identified, for each frame that corresponded to an "in contact" binding state. If the furthest atom-surface distance reached or exceeded 2.0 nm, this frame was counted as an upright binding state.

Overlayer simulations

Standard MD simulations were used to probe multi-chain effects on the molecule-surface orientation of the two bioconjugates. To this end, an overlayer comprising eight chains of either $BP7CF_{10}$ or $F_{10}CBP7$ were place in the same simulation cell as described earlier (containing two *h*-BN surfaces in a wide slit-pore configuration, vertically separated by 8 nm). Each conformation of the eight chains was chosen from likely states identified from the REST-MD simulations. The initial positions and spatial arrangement of the eight chains was chosen to ensure a near-uniform (e.g. non-aggregated) coverage. Liquid water was added as per the process described earlier. Two independent samples were created for each system (either the $BP7CF_{10}$ overlayer or the $F_{10}CBP7$ overlayer) with different arrangements of the eight chains, amounting to four simulations in total. Each system was subjected to 100ns of *NVT* simulation at 300 K, with the same settings as mentioned above. The final 50ns of the simulation was used for analysis in each case.

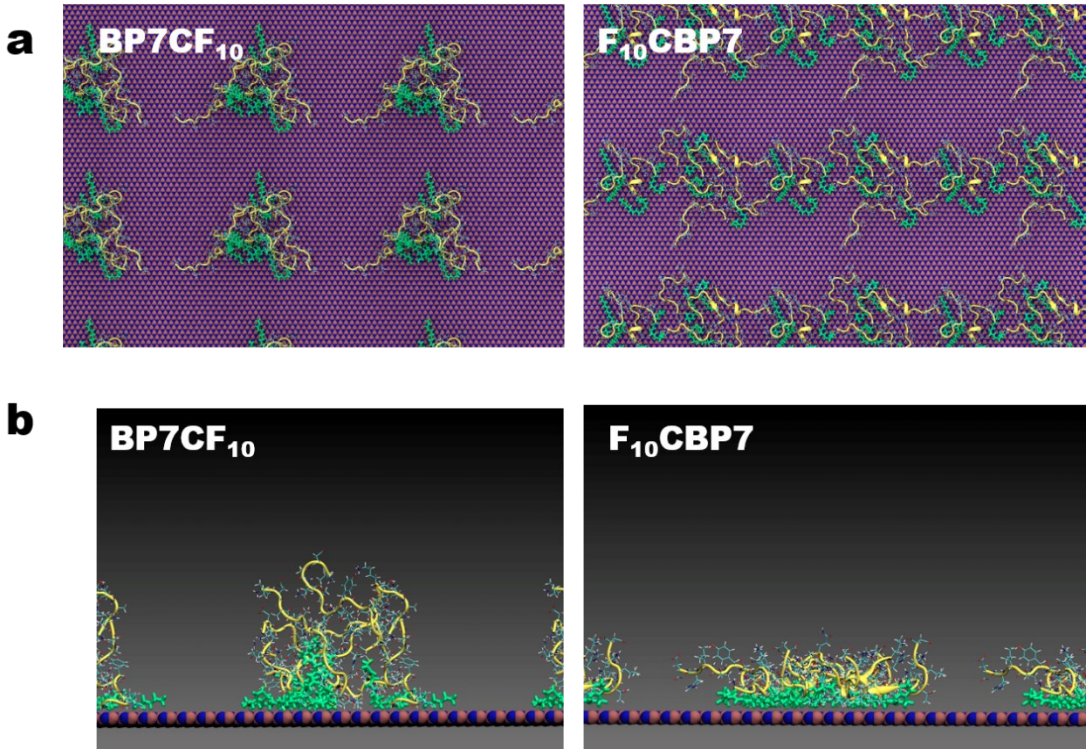

Figure S5. Example views of the resultant eight-chain overlayer following 100ns of standard MD simulation. a) Plan view (shown as a 3x3 supercell), and b) side view. Water not shown for clarity. Green indicates the F10 moiety, yellow indicates the peptide backbone.

Figure S5 provides snapshots of the resultant overlayer structure in plan view and side view. The plan view images indicate the relatively greater aggregation of the overlayer morphology for $BP7CF_{10}$, with greater exposure of the underlying *h*-BN surface. The corresponding view for $F_{10}CBP7$ indicates a less tight aggregate that has greater aerial coverage of the substrate. The side views reveal the clear difference in the degree of extension from the surface. $BP7CF_{10}$ clearly features almost exclusively the upright states observed in the counterpart REST-MD simulations, whereas $F_{10}CBP7$ features mostly prostrate conformations.

The degree of upright extension from the surface was quantified by measuring the most maximal vertical distance between any atom in each chain and the *h*-BN surface, at each time point (every frame, which was saved every 10ps). These eight maximal distances were then averaged over all frames for the analysis portion of the trajectory, and this average was then averaged over the eight chains to produce one value, defined as the extension distance. The extension distance was determined for each of the two independent runs for each bioconjugate and averaged. The sample-averaged extension distance was 2.43nm with a standard deviation of 0.2 nm for $BP7CF_{10}$. The corresponding value for $F_{10}CBP7$ was 1.81nm with a standard deviation of 0.1 nm. These data are consistent with both the single-chain REST-MD data and the dissipation data from the experimental QCM observations.

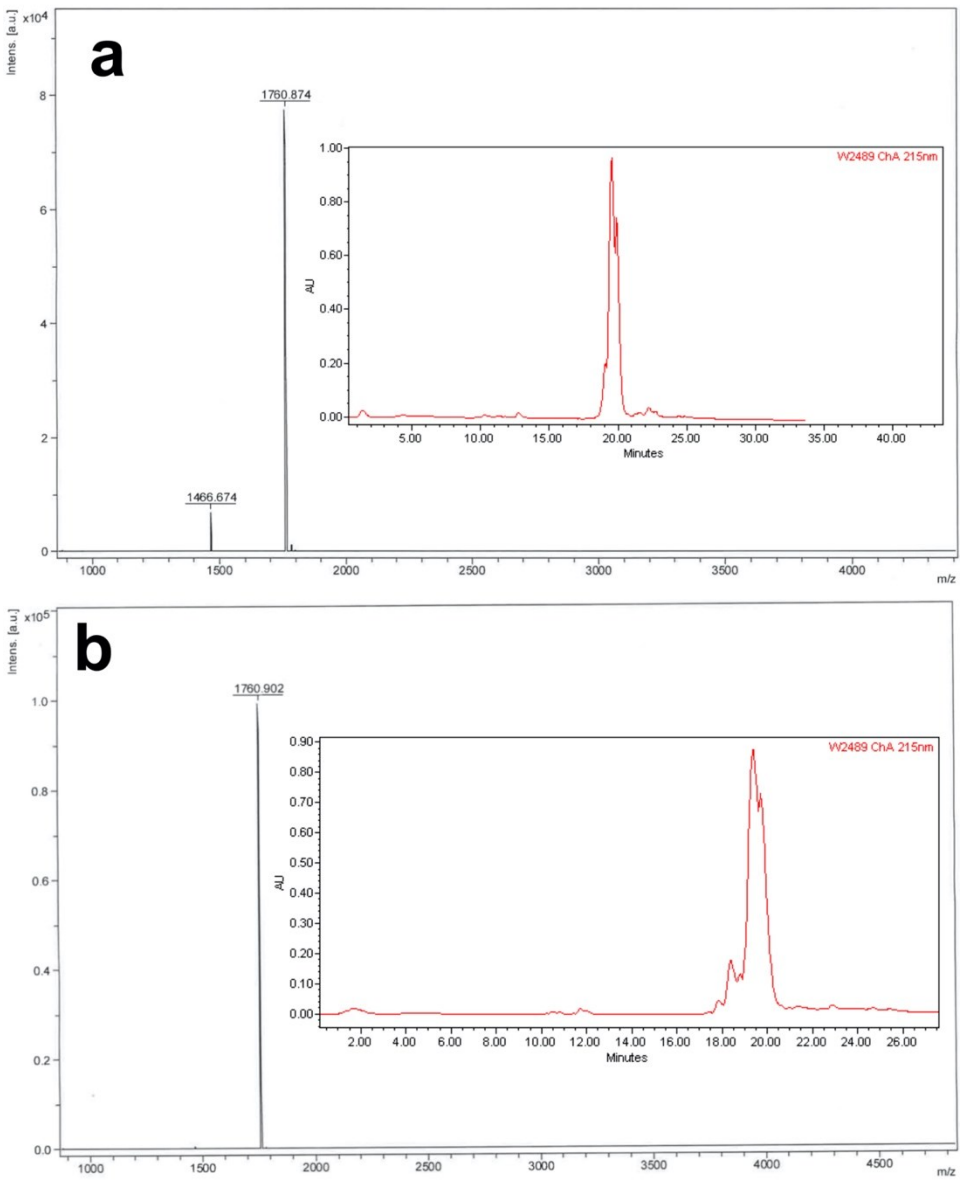

Figure S6. HPLC and MALDI-TOF analysis of (a) F_{10}CBP7 and BP7CF_{10}. The inset represents the HPLC chromatogram of the materials, while the main figure presents the mass spectrum of the sample.

References

1. A. D. Parab, A. Budi, N. Brljak, M. R. Knecht and T. R. Walsh, *Adv. Mater. Interfaces*, 2020,**n/a**, 2001659.
2. T. Terakawa, T. Kameda and S. Takada, *J. Comput. Chem.*, 2011, **32**, 1228-1234.
3. L. B. Wright and T. R. Walsh, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4715-4726.
4. S. Nosé, *Mol. Phys.,* 1984, **52**, 255-268.
5. W. G. Hoover, *Phys. Rev. A,* 1985, **31**, 1695.
6. T. Darden, D. York, D., and L. Pedersen, *J. Chem. Phys*., 1993, **98**, 10089.
7. L. B. Wright, C. L. Freeman, and T. R. Walsh, *Mol. Simul*., 2013, **39**, 1093-1102.
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. & Interfaces*, 2016, **8**, 1050-1060.
9. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, 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. Reiher, 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.
10. S. Piana, K. Lindorff-Larsen and David E. Shaw, *Biophys. J.*, 2011, **100**, L47-L49.
11. A. Budi and T. R. Walsh, *Langmuir*, 2019, **35**, 16234-16243.
12. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
13. E. Neria, S. Fischer and M. Karplus, *J. Chem. Phys.*, 1996, **105**, 1902-1921.
14. Z. Tang, J. P. Palafox-Hernandez, W.-C. Law, Z. E. Hughes, M. T. Swihart, P. N. Prasad, M. R. Knecht and T. R. Walsh, *ACS Nano*, 2013, **7**, 9632-9646.
15. Z. E. Hughes and T. R. Walsh, *J. Mater. Chem. B*, 2015, **3**, 3211-3221.
16. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, Angew. *Chem. Int. Ed*. 1999, **38**, 236-240.
17. J. P. Palafox-Hernandez, Z. Tang, Z. E. Hughes, Y. Li, M. T. Swihart, P. N. Prasad, T. R. Walsh and M. R. Knecht, *Chem. Mater.*, 2014, **26**, 4960-4969.
18. N. Brljak, A. D. Parab, R. Rao, J. M. Slocik, R. R. Naik, M. R. Knecht and T. R. Walsh, *Chem. Commun.*, 2020, **56**, 8834-8837.