Electronic Supplementary Information for:

Predicting Biomolecule Adsorption on MoS_2 Nanosheets with High Structural Fidelity

Le Nhan Pham[*a] and Tiffany Walsh[*a]

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

*E-mail: tiffany.walsh@deakin.edu.au

Computational Details

For the production DFT calculations, energy convergence testing suggested a kinetic cutoff of 50 Ry and charge density cutoff of 300 Ry in combination with a uniform k-point mesh of 5x5x1. To improve the accuracy, these parameters were increased to 75 Ry and 500 Ry respectively, and 7x7x1 to calculate single point energies of optimised configurations. All further details of the calculation settings are provided in the SI Section "First Principles Calculations and Adsorption Energies" and were based on previous work.[1]

First Principles Calculations and Adsorption Energies

The computational process used here was similar to that published in previous work[1], but was extended to 21 adsorbates as listed in Table 1 of the main text. Briefly, all production quantum chemical calculations were conducted at the vdW-DF2 level of theory. Two supercell sizes (4x4 and 6x6) of MoS_2 were built with the initial experimental information of the MoS_2 unit cell[2] to reduce any interaction between adsorbates in two consecutive supercells. A 22 Å thickness was used to separate two slabs of MoS_2 in the z dimension. The dimensions of the two supercell sizes are provided in Figure S1.

The gas-phase interaction was determined using the following formula:

$$E_{ads} = E_{MoS_2 + adsorbate} - (E_{MoS_2} + E_{adsorbate}) \quad \text{(S1)}$$

where E_{ads} is the adsorption energy, and the three terms on the right-hand side are the energies of adsorbate-surface, surface, and adsorbate, respectively.

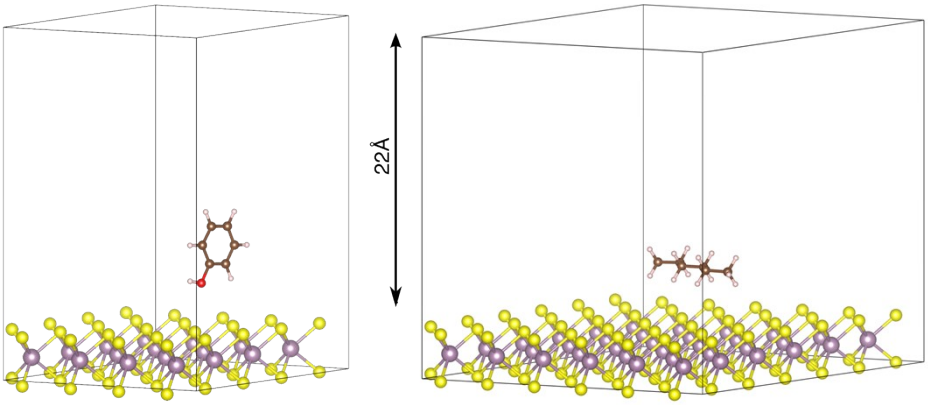

Figure S1. Two supercell sizes (4x4 and 6x6) of the MoS_2 surface constructed to calculate adsorption energy at the vdW-DF2 level of theory.

Table S1. DFT benchmarking data of thiophene and butadiene adsorbed on the basal plane of mono MoS_2 layer against experimental data.[3,4] CT1 to CT6 and CB1 to CB7 are thiophene- and butadiene-surface configurations, respectively.

vdW-DF	adsorption energy (kJ mol^{-1})												
	CT1	**CT2**	**CT3**	**CT4**	**CT5**	**CT6**	**CB1**	**CB2**	**CB3**	**CB4**	**CB5**	**CB6**	**CB7**
df-c09	51.71	45.72	42.60	21.66	25.97	28.10	41.91	41.75	42.26	41.58	25.28	25.28	38.77
df-cx	47.41	42.49	40.02	21.42	24.75	26.47	39.14	39.04	38.55	39.00	25.20	38.75	36.61
df-ob86	51.85	46.35	43.35	22.62	26.86	28.30	42.47	41.92	42.15	26.23	42.37	39.59	13.63
df-obk8	52.01	46.58	43.48	22.63	26.95	28.16	42.37	42.42	41.91	42.03	25.96	42.44	39.58
df2-b86r	41.60	36.20	33.27	16.11	19.79	21.43	32.59	32.55	32.03	32.21	18.57	32.54	29.80
df2-c09	33.39	28.47	28.86	11.62	14.35	28.10	25.11	25.00	24.47	24.76	13.50	24.83	22.54
df	45.24	42.15	39.98	25.81	25.81	25.46	38.35	38.69	38.11	38.12	26.30	38.28	36.75
df2	43.55	38.82	36.01	18.79	22.44	22.87	35.11	35.32	34.54	34.20	24.05	35.04	32.63
expt.	39.75							35.56					

Visualizations of the most stable MoS_2-adsorbate configurations are given in Figure S2. All crystallographic information files (CIFs) and energies (associated Excel spreadsheet) of the remaining structures used for the fitting and validation processes are provided in the Supporting Information bundle. These data can be accessed through the github repository https://github.com/lenhanpham/MoS2-FF-paper.

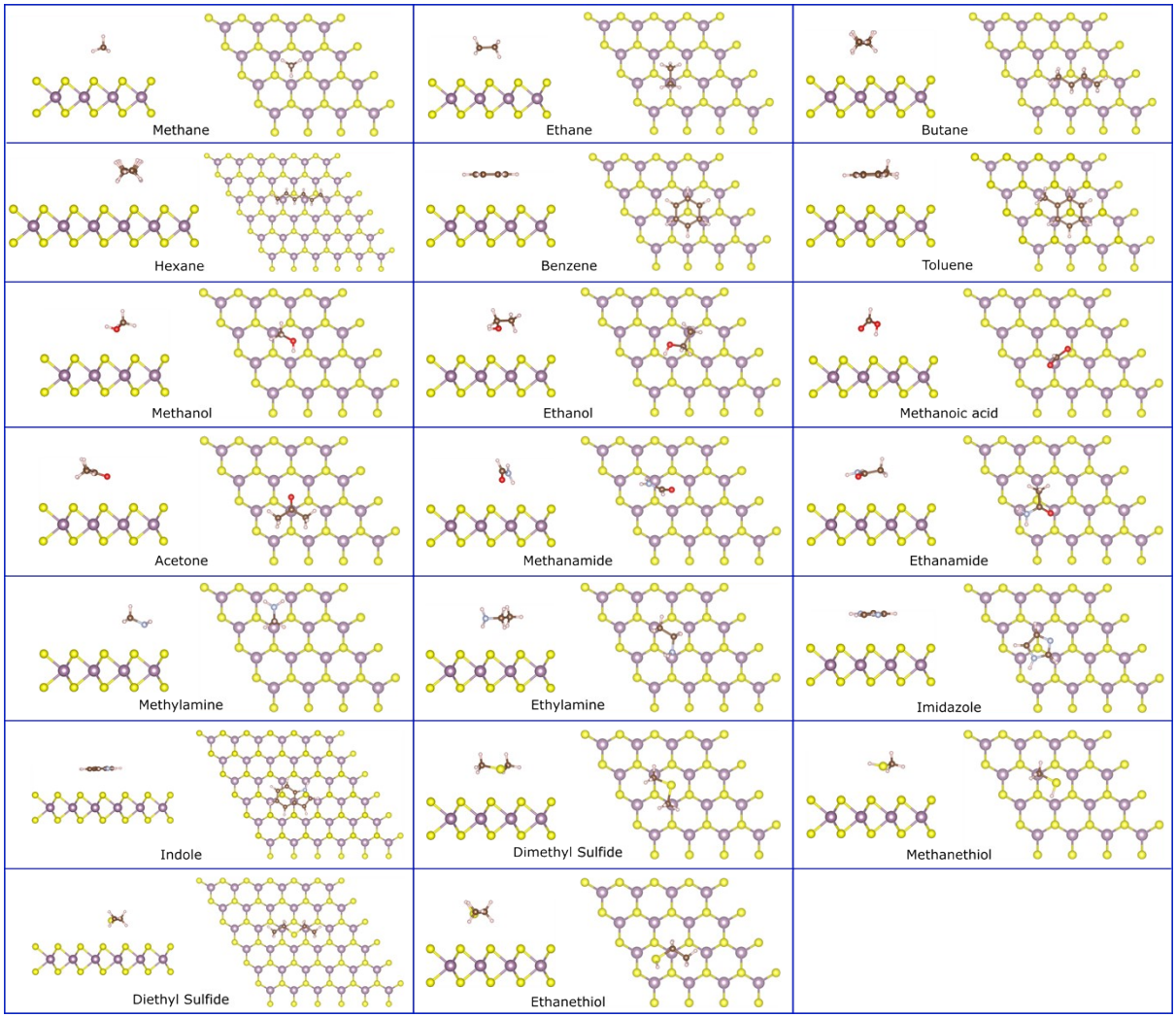

Figure S2. Geometrical structures of the strongest MoS_2-adsorbate interactions determined at the vdW-DF2 level of theory. The corresponding adsorption energies are listed in Table 1 of the main text.

Force Field Fitting to vdW-DF2 Energies

The homo-atomic parameters ($\varepsilon_{Mo\text{-}Mo}$, $\varepsilon_{S\text{-}S}$, $\sigma_{Mo\text{-}Mo}$ and $\sigma_{S\text{-}S}$) were fitted first to reproduce the adsorption energies of alkanes, based on mixing rules. Hetero-atomic van der Waals parameters were then derived for chemically-specific interactions. Specifically, hetero-atomic parameters for aromatic carbon were next fitted after alkane carbon, since this type of carbon will serve to fit other bio-based functional groups (phenol and indole) later on as alkane did. After this point, bespoke parameters for the hydroxyl, thiol, and amine groups were then fitted. Once the parameters for description of the hydroxyl group were available, the carboxylic and amide groups were treated next. Separate parameters were also required to be fitted specifically for indole/imidazole.

All MD simulations for the force field fitting process used a 20x20 supercell of the MoS_2 substrate as depicted in Figure S2. Note that the substrate is kept frozen during all fitting MD simulations. The adsorption energy obtained from the force field level is provided in Eq. S2. The non-interaction system contains a molecule of adsorbates in the centre of the periodic cell.

$$E_{ads} = E_{in} - E_{non} \tag{S2}$$

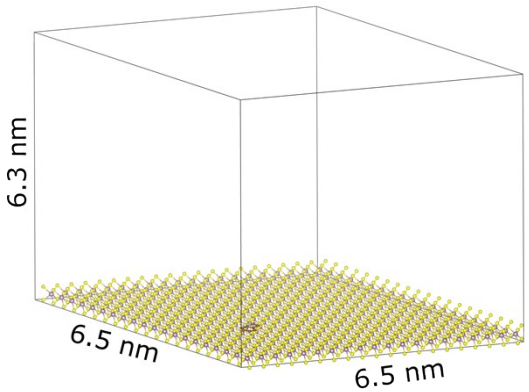

Figure S3. The 20x20 supercell used for the fitting process. Dimensions of the periodic cell are indicated.

The Gromacs 2020 package was used during the fitting process. Convergence of force was controlled by a threshold of 10 kJ mol^{-1}. The Lennard-Jones and electrostatic interactions were considered within a cutoff of 11.0 Å, smoothly switching from a distance of 10.0 Å for the Lennard-

Jones potential. The Particle-Mesh Ewald (PME)[5,6] technique was used for evaluation of electrostatics.

The optimized geometries of the adsorbates on the basal plane of MoS_2 and the adsorption energy data were used for the fitting process. The vdW-DF2 adsorbate-MoS_2 geometrical configurations were kept unchanged during the fitting process, except for reduction of the vertical distances from the adsorbates to MoS_2 surface. As has been done previously, because the vdW-DF functional was found to overestimate the distance from adsorbates to surfaces,[7–9] the molecule-surface distance was systematically reduced by 0.2 Å.

The MD simulation supercell sizes and dimensions of the periodic cells are provided in Figure S3. For the sake of practicality, all bespoke hetero-atomic parameters (σ_{ij} and ε_{ij}) were fitted within a space of values ranging from 0 to 8, as determined from preliminary fitting evaluations. The fitting space was then narrowed gradually to locate as many as possible wells where the optimal values of parameters can be obtained; each well corresponds to a region where the difference in energies between the DFT fitting dataset and force-field outputs is smaller. Further fitting processes were conducted for each well to refine and locate the best possible parameters.

Helical Peptide Simulations

Four peptide α-helical structures, namely the wild type peptide (X) (KWKLFKKIGIGAVLKVLTTGLPALIS), mutant A (KAKLAKKIGIGAVLKVLTTGLPALIS), mutant B (SWSLFSSIGIGAVLKVLTTGLPALIS), and mutant C (KWKFFKKIGIGAVLKVLTTGLPALIS), were constructed using Tinker.[10] Each peptide was placed in the simulation cell (dimensions provided in Figure S4) filled with 18995 TIPS3P water molecules, such that the water density in the centre of the cell was its bulk value at 300 K and 1 atm pressure. The whole simulation system was ensured to be charge-neutral by adding counter ions (Na^+ or Cl^-). 10 ns NVT simulations were carried out at 300 K and Newton's equations of motion were integrated with a time step of 1 fs. The Nosé-Hoover thermostat[11,12] and the PME technique[5,6] were used to maintain the system temperature and evaluate long-ranged electrostatic contributions, respectively. MD fitting and simulations were conducted using

the GROMACS 2020.4 package.[13] All MD simulations were conducted in the NVT ensemble at 300K, and the system temperature was maintained with the Nosé–Hoover thermostat.[11,12] The Lennard-Jones and electrostatic interactions were considered within a cutoff of 11.0 Å, smoothly switching from a distance of 10.0 Å for the Lennard-Jones potential. The Particle-Mesh Ewald (PME)[5,6] technique was used for evaluation of electrostatics.

To build the simulation systems for each peptide mentioned above, a multi-step procedure was designed and followed. First, each individual peptide chain was constructed in a helical geometry using the Tinker force field explorer[10] software package, and geometrically relaxed in an aqueous solution by running a 20 ps *NVT* simulation at 300 K. Following this, the individual peptide chain was placed on the frozen MoS_2 surface vertically with one terminus positioned close to the surface and the other protruding into solution. For each peptide, a total of eight initial (upright) configurations of peptides on the surface were explored, of which four configurations were constructed with the C-terminus close to the MoS_2 surface and the other four with the N-terminus close to the surface. Since experimental data found that mutant C does not produce tilting behaviours and lies down flat on the surface, after the first eight simulations, an additional eight different initial (upright) configurations were prepared with the same procedure, to further probe this system. These eight additional simulations were performed to further consolidate the performance of MoSu-CHARMM in terms of reproducing experimental data. Note that prior to each 10 ns NVT production simulation, the temperature of the whole system was gradually increased from 0 to 300 K through 11 steps (100 ps) with an average increase of 27 K per step.

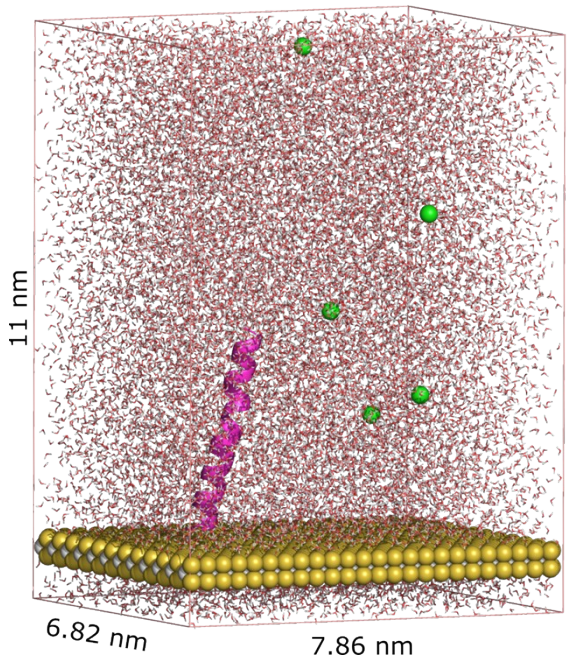

Figure S4. The periodic cell and its dimensions used for simulations of the four peptides.

Umbrella Sampling MD Simulations

The adsorption free energy between each of the 20 amino acids and the aqueous MoS_2 interface was calculating using umbrella sampling MD simulations. To construct the initial configurations for each window, the amino acids were pulled from the starting point defined by a centre-of-mass (COM) vertical distance (in the *z*-coordinate) 0.05 nm close to the MoS_2 surface and stopped at 2.25 nm (near the central vertical plane of the simulation cell); the pulling force constant used was 3000 kJ mol^{-1} nm^{-2}. Each umbrella sampling MD simulation was done with 45 windows with a window spacing of 0.05 nm. A 100 ns simulation was conducted for each window. All umbrella sampling MD simulations were conducted by using the NVT ensemble at 300K maintained with the Nosé-Hoover thermostat. All simulation control settings used for the Lennard-Jones and electrostatic interactions were the same as those used in the fitting process. The umbrella sampling MD simulation cell contained 4070 TIPS3P[14,15] water molecules such that the density of TIPS3P water was bulk-like at 300 K and 1 atm pressure and two frozen MoS_2 sheets, and its dimensions are provided in Figure S5. After 45x 100 ns simulations for each amino acid, the 45 pull force output files were used as inputs in the Weighted Histogram Analysis Method (WHAM) to construct the potential of mean force (PMF) curve. The temperature (300 K) of

the umbrella sampling simulations was passed to WHAM for its construction of PMF curve. An example for tyrosine is provided in Figure S6.

The binding free energy of the amino acids at the aqueous MoS_2 interface were calculated from the resultant potential of mean force (Figure S6) obtained after the umbrella sampling MD simulations. The Boltzmann average was used to calculate final binding free energy as formulated in the following equation.

$$\langle A \rangle = \frac{\int A e^{-E_i/k_B T}}{\int e^{-E_i/k_B T}} \quad \text{(S3)}$$

where A is the quantity or observable (energy), to be calculated, of system; k_B and T are the Boltzmann constant and temperature, respectively. In this case, "A" is the binding energy, and integration in Eq. S3 was performed over the reaction coordinate as defined above.

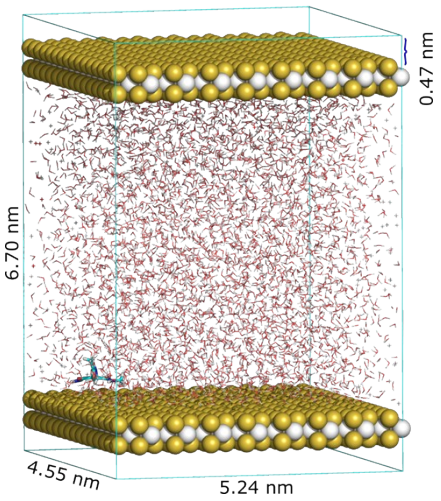

Figure S5. The system cell and its dimensions used in the umbrella sampling simulations.

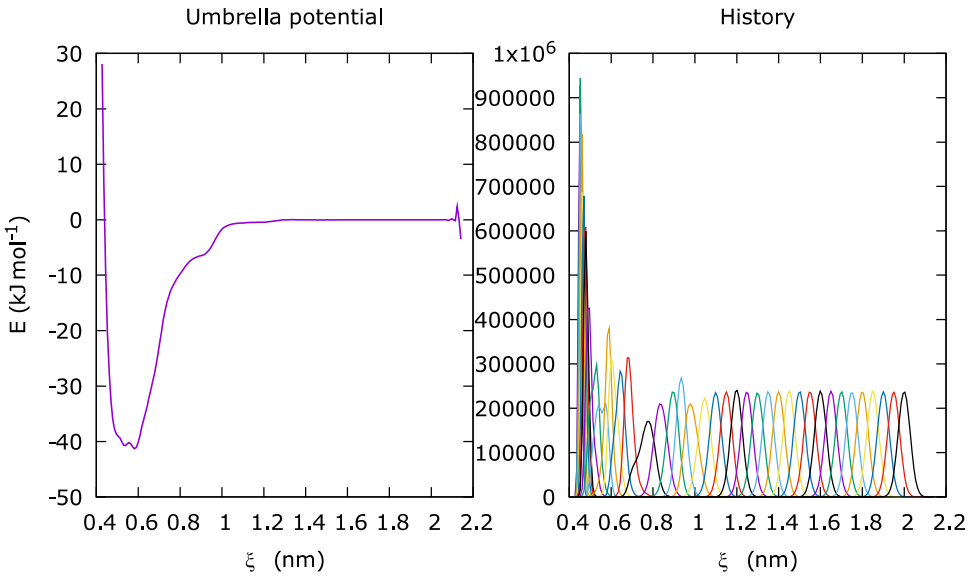

Figure S6. An example of umbrella potential of mean force and sampling history as obtained for tyrosine at the aqueous MoS_2 interface.

Table S2. Adsorption free energies of the twenty amino acids at the aqueous interface of the MoS_2 basal plane. Histidine C is the protonated state of Histidine.

Amino acid	Energy (kJ mol^{-1})	Amino acid	Energy (kJ mol^{-1})
Tryptophan	-52.7 ± 1.3	Asparagine	-30.4 ± 1.2
Arginine	-50.5 ± 1.3	Proline	-30.3 ± 1.1
Tyrosine	-42.4 ± 1.2	Cysteine	-30.2 ± 1.2
Histidine	-38.7 ± 1.1	Threonine	-30.2 ± 1.1
Lysine	-36.2 ± 1.2	Isoleucine	-29.1 ± 1.1
Histidine C	-36.2 ± 0.6	Glutamate	-28.3 ± 1.2
Serine	-36.2 ± 1.2	Leucine	-27.7 ± 1.2
Methionine	-34.8 ± 1.1	Alanine	-27.0 ± 1.1
Glutamine	-34.5 ± 1.2	Valine	-24.5 ± 1.0
Glycine	-32.1 ± 1.0	Aspartate	-24.4 ± 1.1
Phenylalanine	-31.0 ± 0.6		

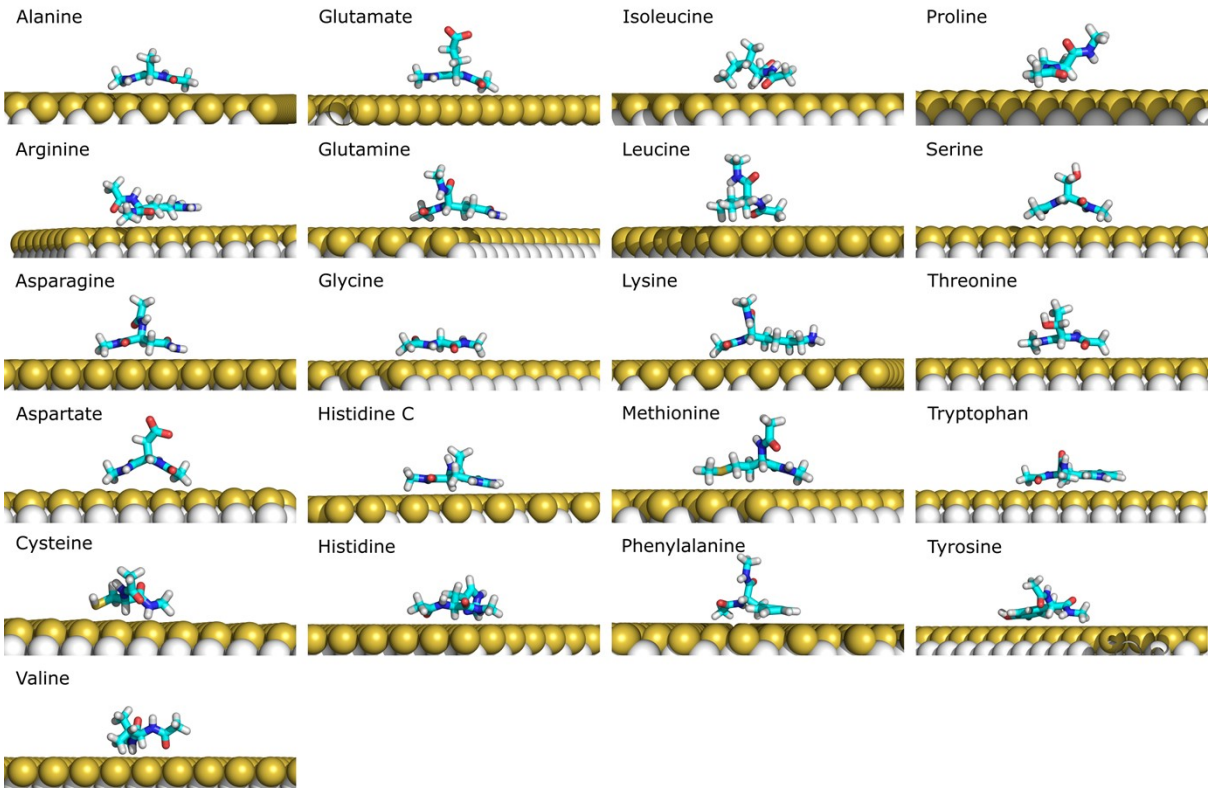

Figure S7. Side views of the most populated configurations of the twenty amino acids obtained from the umbrella sampling simulations. Water not shown for clarity.

References

(1) Pham, L. N.; Walsh, T. R. Force Fields for Water–Surface Interaction: Is Reproduction of the Experimental Water Contact Angle Enough? *Chem. Commun.* **2021**, *57*, 3355–3358.

(2) Schönfeld, B.; Huang, J. J.; Moss, S. C. Anisotropic Mean-Square Displacements (MSD) in Single-Crystals of 2 *H* - and 3 *R* -MoS_2. *Acta Crystallogr., Sect. B: Struct. Sci.* **1983**, *39*, 404–407.

(3) Salmeron, M.; Somorjai, G. A.; Wold, A.; Chianelli, R.; Liang, K. S. The Adsorption and Binding of Thiophene, Butene and H2S on the Basal Plane of MoS_2 Single Crystals. *Chem. Phys. Lett.* **1982**, *90*, 105–107.

(4) Peterson, S. L.; Schulz, K. H. Ethanethiol Decomposition Pathways on MoS_2 (0001). *Langmuir* **1996**, *12*, 941–945.

(5) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(6) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(7) Budi, A.; Walsh, T. R. A Bespoke Force Field to Describe Biomolecule Adsorption at the Aqueous Boron Nitride Interface. *Langmuir* **2019**, *35*, 16234–16243.

(8) Hughes, Z. E.; Tomásio, S. M.; Walsh, T. R. Efficient Simulations of the Aqueous Bio-Interface of Graphitic Nanostructures with a Polarisable Model. *Nanoscale* **2014**, *6*, 5438–5448.

(9) Hughes, Z. E.; Wright, L. B.; Walsh, T. R. Biomolecular Adsorption at Aqueous Silver Interfaces: First-Principles Calculations, Polarizable Force-Field Simulations, and Comparisons with Gold. *Langmuir* **2013**, *29*, 13217–13229.

(10) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.

(11) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(12) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(13) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(14) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(15) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902–1921.