Supporting Information for:

Aqueous Peptide-TiO₂ Interfaces: Iso-energetic Binding *via* either Entropically- or Enthalpically-driven Mechanisms

Anas M. Sultan,[1,#] Zayd C. Westcott,[2,#] Zak E. Hughes,[1] J. Pablo Palafox-Hernandez,[1] Tristan Giesa,[3] Valeria Puddu,[2] Markus J. Buehler,[3] Carole C. Perry,[2,] and Tiffany R. Walsh [1,*]*

[1]Institute for Frontier Materials, Deakin University, Geelong, Victoria 3216, Australia, [2]Interdisciplinary Biomedical Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, U.K., [3]Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA.

[#]These authors contributed equally.

*To whom correspondence should be addressed:

TRW: tiffany.walsh@deakin.edu.au; CCP: carole.perry@ntu.ac.uk

Contents:

Section S1: Overlayer Thickness Estimates from QCM-D data

Section S2: Additional Computational Methodology

Section S3: Implicit Solvent Simulations

Table S1: Analysis of peptide adsorption/desorption from QCM-D.

Table S2: The estimated amount of peptide required for a complete monolayer coverage in two idealized peptide-surface orientations.

Table S3: Cluster populations for the top-ten most populated clusters for Ti-1 and Ti-2, determined from the 'in-solution' and 'surface-adsorbed' REST-only simulations.

Table S4: Residue-surface contact percentages for Ti-1 and Ti-2 adsorbed at the aqueous titania interface, including 'direct' and 'indirect' (solvent mediated) contact modes.

Table S5: Cross-cluster comparisons indicating 'matched' structures between the in-solution and surface-adsorbed conformational ensembles for Ti-1.

Table S6: Cross-cluster comparisons indicating 'matched' structures between the in-solution and surface-adsorbed conformational ensembles for Ti-2.

Table S7: Definition of sites on each residue side-chain used to calculate residue—surface separation, and the distance cut-off used to determine 'direct' surface contact.

Figure S1: Schematic of two idealized peptide-surface orientations, horizontal and vertical, for determining estimates of peptide surface packing.

Figure S2: Histograms of the relative frequency with which the REST+metaD CV (peptide *com*—titania distance) was sampled as a function of simulation time-step.

Figure S3: Evolution of the free energy profile corresponding to the surface adsorption of Ti-1 and Ti-2, as a function of REST+MetaD simulation time.

Figure S4: Evolution of the adsorption free energy of Ti-1 and Ti-2, as a function of REST+MetaD simulation time.

Figure S5: Titania surface model used in the molecular simulations.

Figure S6: Representative structures of titania-adsorbed structures predicted from REST-only simulations.

Figure S7: Ramachandran analysis of REST-only trajectories of Ti-1 and Ti-2, in both the in-solution and surface-adsorbed states.

Figure S8: Histograms of residue-surface contact distances for both Ti-1 and Ti-2, calculated from the REST-only trajectories.

Figure S9: Structures of the three most populated clusters for Ti-1 and Ti-2 in the absence of the titania surface, predicted from 'in solution' REST-only simulations.

Figure S10: Mobility of replicas through "effective temperature" space for the REST-only and REST+MetaD simulations.

Figure S11: Number of clusters as a function of REST-only MD simulation timestep, for Ti-1 and Ti-2, for both the 'in-solution' and 'surface-adsorbed' states.

Figure S12: Symmetrized final free energy profiles of Ti-1 and Ti-2 adsorbed at the aqueous titania interface.

Figure S13: Most populated clusters for Ti-1 and Ti-2 determined from the implicit solvent REMD simulations.

Figure S14: Most populated clusters of Ti-1 determined from the post-implicit-solvent (water) standard MD simulation in explicit solvent in the presence of 150 mM sodium chloride.

Figure S15: Most populated cluster of Ti-2 determined from the post-implicit-solvent annealing MD simulation in explicit solvent (150 mM sodium chloride in water).

Overlayer Thickness Estimates from QCM-D data:

A thickness estimate of a single, rigid absorbed peptide layer can be achieved using

$$d = \frac{\Delta m}{\rho}$$

where Δm is the mass of peptide bound to the QCM sensor surface and ρ is the effective density of the adhering layer. ρ can be approximated using the method from Fischer *et al.*[1] The average protein density is a molecular-weight-dependent function:[1]

$$\rho(M) = \left[1.410(6) + 0.145(28). \exp\left(-\frac{M(kDa)}{13(4)}\right)\right] \, g/cm^3$$

However, this assumes formation of a complete layer of peptide with no defects or patterning. This yielded the following data, which were used in the expression for d, given above.

Peptide	Molecular Weight g/mol	Molecular Weight kDa	Estimated Peptide Density* (g/mL)
Ti-1	1395.62	1.39562	1.540239929
Ti-2	1407.51	1.40751	1.571580544

We can also use our QCM-D data in respect of the mass of peptide adsorbed to the sensor surface to infer the average sparseness of the idealized adsorbed peptide overlayer. To do this, we considered two extremes of peptide surface area in the adsorbed state – an ideal binding area for the horizontally-oriented state (HBA) and an ideal binding area for the vertically-oriented state (VBA), as shown in Figure S1. Other assumptions included that the total surface area of the sensor was available for binding, that no solvent molecules or ions present on the sensor surface, and, neglect of inter-peptide interactions. These estimates, summarized in Table S2, reveal the likelihood of an extremely sparse monolayer coverage. An assessment of the HBA and VBA modes indicated that Ti-1 could adsorb in approximate isolation within an area $\approx 23 \times$ greater than its ideal surface area in the HBA case, or $\approx 120 \times$ greater than its ideal surface area in the VBA case.

Additional Computational Methodology:

REST Simulations:

Structural Details of the Titania Surface. We modelled the negatively-charged hydroxylated titania interface using a five-layer rutile TiO_2 slab of dimensions $60.7 \times 57.5 \times 14.5$ Å^3, modeled using the force-field developed by Predota *et al.* (described in their paper as their negatively-charged non-hydroxylated surface model) with 12.5% hydroxyl coverage[2]. The slab featured 25 hydroxyl groups randomly placed on one facet of the slab and another 25 groups with the same coordinates in the lateral plane on the opposite facet of the slab. A structural model of the slab is provided in Figure S5. The slab atoms were fixed in space throughout the duration of simulations, however surface hydroxyls were allowed to move within the constraints of the force-field bonded terms. The surface charge density was fixed at $\sigma = -0.104$ C m^{-2}), such that each facet of the slab carried $-25e$ charge ($-50e$ in total).

REST Simulations – General Details. For all REST simulations, the system temperature of 300 K was maintained using the Nose-Hoover thermostat[3-4] with a relaxation time of 0.2 ps. Three-dimensional periodic boundary conditions were used throughout. The Leapfrog algorithm was used to solve Newton's equations of motion, with an integration time step of 1 fs and coordinates saved every 1000 steps. Long-range electrostatics were described using the particle mesh Ewald (PME)[5] summation, with a 12 Å real-space cutoff. The same cutoff was used for Lennard-Jones interactions. In both cases (surface-adsorbed and in-solution), REST-only simulations of each peptide were carried out for 15 ns. Each simulation comprised 16 replicas of the system, spanning an "effective temperature" range of 300-433 K, with each replica featuring a different initial peptide conformation. Note that the "effective temperatures" are not thermal temperatures – the thermal temperature of each replica was set to 300 K. All simulations were implemented according to the method outlined in Terakawa *et al.*[6] and Wright *et al.*[7] Briefly, for the system used here, the potential energy of replica *j* was scaled according to:

$$V_j(X) = \frac{\beta_j}{\beta} V_{pp}(X) + \sqrt{\frac{\beta_j}{\beta}} V_{ps}(X) + V_{ss}(X)$$

where V_{pp} is the intra-peptide, V_{ps} the peptide–water and peptide–surface, and V_{ss} the water–water, water–surface and surface–surface potential energies of the system X, respectively. β and β_j are the inverse of the system and 'effective' temperatures of replica j and are related *via* λ_j:

$$\beta_j = \beta(1 - \lambda_j) + \beta_H \lambda_j$$

where $0 < \lambda_j < 1$. The lambda values of the 16 replicas were (0.000, 0.067, 0.133, 0.200, 0.267, 0.333, 0.400, 0.467, 0.533, 0.600, 0.667, 0.733, 0.800, 0.867, 0.933, 1.000). The highest 'effective' temperature (β_H) was chosen to be 433 K.

Before initiating the REST simulations, starting configurations were equilibrated for 0.5 ns at their target potentials with no exchange moves attempted. Similar to previous work,[8-9] the solute group comprised the peptide and counterions. An exchange between two neighboring replicas was attempted every 1 ps. Evidence of sampling efficacy and approach of equilibrium are provided in Figures S10 and S11.

REST-only Simulation Details. For the "surface-adsorbed" REST-only simulations, one chain of either the Ti-1 or Ti-2 peptide was placed in an orthorhombic periodic cell, along with the titania slab, and 6951 TIPS3P[10-11] water molecules. The cell dimension perpendicular to the slab surface was adjusted to ensure bulk water density (at ambient pressure and a temperature of 300 K) in the center of the inter-slab space of the simulation cell. To neutralize the −50e charge of the slab, and to provide a reasonable salt concentration in the bulk solution, 54 Na^+ and 4 Cl^- ions were added to the solvent. For Ti-2, 2 Na^+ ions were replaced with 2 water molecules due to the +2e charge of the peptide. The cell dimensions were $60.7 \times 57.5 \times 61.9$ $Å^3$. Because 34 Na^+ ions adsorbed strongly to the slab facets, the remaining 20 Na^+ ions contributed to a bulk (*i.e* non surface-adsorbed) Na^+ concentration of 0.15 M. We also performed REST-only simulations of each of these peptides in aqueous solution in the absence of the titania slab. For these "in solution" REST-only simulations, we modeled one chain of either the Ti-1 or Ti-2 peptide in a cubic simulation cell along with 6773 TIPS3P water molecules with 19 Na^+ and Cl^- ions. The positive charge of Ti2 was balanced by replacing two water molecules with 2 Cl^- ions. The periodic cell dimensions were determined *via* an initial simulation in the isothermal-isobaric ensemble (*NPT* ensemble), at 1 atm pressure and 300 K.

REST-only Cluster Analysis. The Boltzmann-weighted ensemble of conformations was obtained from the reference trajectory of the REST-only simulations at the interface, corresponding to the effective temperature of 300 K (the unscaled Hamiltonian). The trajectories were categorised into groups of similar structures based on their backbone conformation, determined using the Daura clustering method[12]. The entire 15 ns trajectory of each simulation was used for the analysis, with a 2 Å root mean squared deviation (RMSD) cutoff between backbone atom positions.

REST-only Residue-Surface Contact Analysis. Residue-surface contact was evaluated for two contact modes: 'direct' contact, where a residue side chain has displaced water molecules in the first interfacial water layer; and 'solvent-mediated' contact, where a residue made contact with the interfacial water layer rather than directly with the surface itself. To determine if a residue was in contact (either directly or *via* solvent mediation) each residue was assigned a reference site on its side chain, and the separation distance (in the *z*-dimension) between the side chain reference site and the surface was calculated. A residue was considered to be in direct contact or solvent-mediated with the surface if the separation distance was within a given cutoff distance determined for each contact mode. The contact propensity of each residue was classified according to the percentage of frames that a given residue was located within the range of cutoff distances, calculated from the entire reference trajectory. Details of cutoff distances and reference sites for each residue are provided in Table S7 of the Supporting Information. Cutoff distances were determined by histograms of the residue-surface distance calculated over the REST-only trajectories; see Figure S8 for these data.

Details of Histidine Protonation: There are grand challenges associated with the treatment of His residues in materials-binding peptides. In the amino acid form, the side-chain of histidine is likely to have both protonated and deprotonated forms present in equilibrium in aqueous solution at a pH value of 7, because the relevant pK_a value is ~6.2 for the imidazole side-chain. However, the pK_a corresponding to protonation of the side-chain of a His residue in a peptide sequence in solution may differ substantially from that of the corresponding amino acid.[13] This is due to the modification in the immediate dielectric environment of the relevant His in a protein/peptide, which can depend on the conformational state (or ensemble of states) of the protein/peptide.

S7

Moreover, the relevant pK_a of a His residue in the surface-adsorbed state at the aqueous titania interface might also differ from that of a free (*i.e.* un-adsorbed) peptide in solution. At present, the prediction of pK_a values for residues such as His for peptides/proteins in aqueous solution, which can account for environment/conformation effects, is extremely challenging.[14] Going further, to the authors' knowledge, information regarding the influence exerted by the presence of the aqueous titania interface on the pK_a of His residues in surface-bound peptide is also scarce. We have therefore chosen one possible configuration for the protonation state of each His residue, corresponding to the approx. 1-in-3 probability of the imidazole group being protonated. To investigate this phenomenon in more detail, a more in-depth study to probe the dependence of His protonation for His-containing materials-binding peptides would be required, but is beyond the scope of this current study.

REST+Metadynamics Free Energy Extraction. In the limit of an infinite metadynamics simulation ($t \rightarrow \infty$), the bias added during a metadynamics simulation approaches the negative of the free energy of the system, $V(X, t) \rightarrow -G(X, t)$, where V, G and X are the metadynamics bias added, the free energy of the system and its co-ordinates, respectively. The symmetrical nature of our simulation set-up means that two estimates of the binding affinity for the TiO_2 rutile (110) surface could be generated per run; adsorption to the top face of the titania slab ($\Delta G_{ads,t}$) and adsorption to the underside of the periodic neighboring slab ($\Delta G_{ads,b}$) respectively (referred to herein as the bottom face). Using the definition of Schneider and Colombi Ciacchi[15], $\Delta G_{ads,t}$ and $\Delta G_{ads,b}$ were estimated using:

$$\Delta G_{ads,t} = -k_B T ln \left(\frac{c_{ads,t}}{c_{bulk}}\right) \quad (1)$$

$$\Delta G_{ads,b} = -k_B T ln \left(\frac{c_{ads,b}}{c_{bulk}}\right) \quad (2)$$

where $c_{ads,t}$ is the concentration of adsorbed peptide at the top face, $c_{ads,b}$ is the concentration of the adsorbed peptide at the bottom face, and c_{bulk} is the peptide concentration in the bulk. These concentrations are given by:

$$c_{ads,t} = \frac{1}{z_0 - z_{min}} \int_{z_{min}}^{z_0} exp\left[{-G(X, t_f)}\middle/{k_B T}\right] dX \quad (3)$$

$$c_{ads,b} = \frac{1}{z_{max}-z_1} \int_{z_1}^{z_{max}} exp\left[-G(X,t_f)\Big/k_B T\right] dX \quad (4)$$

$$c_{bulk} = \frac{1}{z_1-z_0} \int_{z_0}^{z_1} exp\left[-G(X,t_f)\Big/k_B T\right] dX \quad (5)$$

where z_0 and z_1 indicate the values of the CV for which the peptide is considered to be in the 'bulk' solution (*i.e.* the peptide was defined as not adsorbed for $z_0 < z < z_1$, and was defined as adsorbed for all other values of z; see Figure S12). z_{min} is the z coordinate of the top (upper side) of the titania slab, and correspondingly z_{max} is the z coordinate bottom surface of the underside of slab as its periodic image. T is the temperature and $t_f = 120$ ns, the duration of the REST+Metadynamics simulations performed in this work. Specifically, z_0 was defined using the final symmetrized free energy profiles of each system (see Figure S12) to be the minimum value of the CV for which $G(X, t_f) > -4$kJmol^{-1}; z_1 was then assigned the same distance from the bottom surface of the periodic image of the slab. Herein, we quote our calculated binding free energies as ΔG_{ads}, the mean of $\Delta G_{ads,t}$ and $\Delta G_{ads,b}$. The associated error was defined as half the difference between $\Delta G_{ads,t}$ and $\Delta G_{ads,b}$.

Regarding the peptide-surface interaction at distances such as 20 Å (around z = 30 Å) from the optimal binding position and their impact on the free energy profile, we are alert to the fact that the contour length of the fully extended peptides is ~40 Å, which might give rise to counter-example instances where the center of mass of the peptide may be distant from the surface while still allowing the peptide to interact with the surface. This somewhat misleading identification of an incorrectly labelled "unadsorbed" state could be satisfied by the scenario where the peptide is interacting with the surface only *via* the N-terminal or C-terminal regions (such that the peptide is fully outstretched and roughly vertically oriented with respect to the surface plane). We have inspected the reference replica trajectory of our REST+metadynamics simulations in each case to determine if such configurations have made a substantial contribution to the ensemble of configurations corresponding to this value range of the collective variable. We found only a trivial number of instances for which this was the case; ~8% of frames defined by our cut-off to be "unadsorbed" that were inappropriately classified as "unadsorbed" by our cut-off criterion for Ti-1 and ~10% of such "unadsorbed" frames for Ti-2. We have also checked the robustness of our results with respect to our choice of the dividing surface. By moving the

S9

"adsorbed"/"unadsorbed" cut-off 2 Å further out, the adsorption energies changed only slightly, with the error bars overlapping (-14.4±0.2 and -18.2±3.9 kJ mol^{-1} for Ti-1 and Ti-2 respectively). Our count of frames that were inappropriately classified as "unadsorbed" by our cut-off criterion reduced to ~3% for Ti-1 and ~1% for Ti-2. We therefore argue that our division of unadsorbed and surface-adsorbed states is reasonable in light of these facts.

Moreover, we argue that the periodic cell is sufficiently large in the dimension perpendicular to the surface plane. Previous simulations (Ref 23 in the main text) of adsorption at the aqueous titania interface using a different free energy technique (potential of mean constraint force simulations) indicated that the forces on adsorbates in the center of the inter-slab space were sufficiently small to indicate negligible interaction between the adsorbate and the surface. Therefore, if the simulation cell was increased along the dimension perpendicular to the slab surface, we would expect that the central portion of the free energy profile would remain approximately flat. In the current manuscript, one of the main findings of our work is that the free energies of adsorption of two peptides are very close in value (both in experiment and simulation). While a larger simulation cell might modify the absolute value of the predicted adsorption free energies slightly, we emphasize here that both peptides were modelled under the same conditions and would be subject to the same effect.

Implicit Solvent Simulations:

Methodology: The Ti-1 and Ti-2 peptide structures were built using Tripos Sybyl. N- and C-terminal charges were then introduced using the VMD topology tool. In these REMD simulations, 16 replicas were simulated at temperatures between 300 and 800 K. The simulations were carried out using the CHARMM27 force field implemented in NAMD, using the Generalized Born implicit solvent model (α-cutoff 12.0, ion concentration 0.3). Solvent friction was added *via* a Langevin friction term (10 ps^{-1}) that allowed for high mobility and conformational sampling. A cutoff of 15 Å was used for long-range interactions and a simulation timestep of 2 fs was used throughout, in conjunction with LINCS constraints.

REMD was run for 10 ns (an aggregate of 160 ns per each T-REMD simulation) and the temperatures exchanged every 0.2 ps. Statistical analysis was performed using the MMTSB toolbox, with a k-means clustering algorithm (cluster distance 2 Å). Final ensemble structures were chosen from the lowest temperature replica (300 K). Representative structures of Ti-1 and Ti-2 that were generated by this clustering analysis were then exported for subsequent MD simulation in explicit solvent. Regular MD simulations of the most populated cluster in explicit solvent were found to yield very few clusters in either case, chiefly because the helical structures were very persistent. Therefore, we subjected each peptide to a different post-implicit-solvent MD simulation procedure to explore alternative strategies. For Ti-1, we chose two clusters, Ti-1_t13 (the most populated cluster at 29% of the population), and Ti-1_t65 (an arbitrarily-chosen low-population cluster at ~6% of the population). These were both subjected to subsequent standard MD simulation at 300 K for 50 ns. For Ti-2, we tried a different strategy, in which we selected the most populated cluster, denoted Ti-2_t59, and subjected this system to thermal annealing. The annealing was run for 40 ns, by heating up the structure by 50K and cooling down in 5 ns cycles and then equilibrating at 300 K again for 10 ns. All of these explicit simulations were performed using a cubic periodic cell containing TIP3P water and 0.15 M NaCl, using the CHARMM27 force-field within Gromacs. To prevent image interactions, the cell was larger than the peptide by at least 10 Å in all dimensions. Equilibration was performed with Langevin dynamics at 300 K. The Particle Mesh Ewald (PME) approach was used for the long-range electrostatic interactions.

S11

Results: The results of the clustering analysis for the implicit solvent T-REMD simulations revealed that the most likely structures had a considerable helical character (see Figure S13). The results of the subsequent explicit solvent simulations varied according to the simulation strategy employed.

The top two most populated clusters (~35% and 22% respectively) resulting from the Ti-1_t13 regular MD simulation are shown in Figures S14 a) and b); these revealed a substantial helical structural character, particularly for the most populated cluster (Figures S14 a)). We also show the top two most populated clusters (~12% and 11%) resulting from an arbitrarily-chosen non-helically-structured cluster, Ti-1-t65 (Figures S14 c) and d)). As might be expected, these structures exhibit a more random-coil character. At first glance it may appear that choice of a low-population cluster from an implicit-solvent REMD simulation may be a promising strategy. However, the arbitrary choice of this low-population cluster as the initial structure for the explicit-solvent MD simulation is problematic; there is no systematic way to select this cluster. Moreover, by making a systematic cross-cluster comparison between the ensemble of structures generated from the Ti-1-t65 initial structure, and those generated using the explicit-solvent REST simulations, we found no significant structural similarity between these two ensembles. In summary, this strategy is not able to deliver reliable results; it is clear that the conformational ensemble generated from Ti-1_t13 and Ti-1_t65, even when combined, does not approach anything close to the REST-generated Boltzmann-weighted ensemble of states.

For Ti-2, the trajectory resulting from the thermal annealing MD simulation was clustered. The most populated cluster is shown in Figure S15, and revealed a persistent helical conformation, which comprised ~46% of the entire conformational ensemble. Again, we performed a comprehensive cross-cluster analysis between the ensemble generated from thermal annealing and that generated from the explicit-solvent REST simulations, again, we found no significant structural similarity between these two ensembles. These results suggest that thermal annealing is also not a viable strategy.

Previous studies suggest that the outcomes from explicit-solvent REST simulations of similar-sized peptides in solution should provide reasonably reliable benchmarks[8-9, 16]. Therefore, our results discussed above reinforce the conclusion that simulation data generated using implicit-solvent models, even when followed up with explicit solvation simulations, are unable

to provide a reliable characterization of the conformational ensemble of materials-binding peptides.

Table S1: Analysis of peptide adsorption/desorption from QCM-D.

Experimental Stage	Comparison	NaCl (0.15 M)		Water at pH 7.4	
		Ti-1	**Ti-2**	**Ti-1**	**Ti-2**
Addition of Peptide	Peptide Adsorption (ng/cm^2)	10.39	21.21	65.16	46.92
	Rate of Peptide Adsorption $(ng\ cm^{-2}\ min^{-1})$	0.88	5.90	42.87	13.42
Buffer re-added	Peptide Desorption (ng/cm^2)	3.91	7.89	64.39	20.10
	Rate of Peptide desorption $(ng\ cm^{-2}\ min^{-1})$	0.57	0.66	37.92	3.68

Table S2: The estimated amount of peptide required for a complete monolayer coverage in both HBA and VBA orientations (see Figure S1 for definitinons of HBA and VBA). The mass bound using QCM-D is shown for comparison.

Peptide (Buffer)	QCM-D Adsorption (ng/cm^2)	Peptide required for a complete HBA monolayer coverage (ng/cm^2)	Peptide required for a complete VBA monolayer coverage (ng/cm^2)
Ti-1 NaCl (0.15 M)	10.39	231	1181
Ti-2 NaCl (0.15 M)	21.21	233	1187

Table S3: Cluster populations (given as a percentage) for the top ten most populated clusters for Ti-1 and Ti-2, as determined from both the in-solution and surface-adsorbed REST-only simulations. Values in parentheses give the total number of clusters in each case.

Cluster Rank	Ti-1		Ti-2	
	In Solution (270)	Adsorbed (182)	In Solution (200)	Adsorbed (113)
1	5.3	11.7	8.7	23.4
2	5.0	10.2	6.4	12.9
3	4.4	5.6	4.9	9.6
4	3.8	4.1	4.6	6.3
5	2.5	4.0	3.5	6.0
6	2.5	3.7	3.5	5.1
7	2.4	3.2	3.4	4.1
8	2.3	3.2	3.2	3.4
9	2.2	3.1	3.0	2.5
10	2.2	2.6	2.8	2.5

Table S4: Residue-surface contact percentages for Ti-1 and Ti-2 adsorbed at the aqueous titania interface, for both 'direct' and 'indirect' (solvent mediated) contact. Values are rounded to the nearest whole number. Shaded entries indicate residues where the total (=direct+indirect) contact is less than 10%.

Ti-1	Resid	Direct contact %	Indirect contact %	Total contact %
	N-term	28	14	42
	GLN1	7	15	22
	PRO2	0	21	21
	TYR3	1	23	24
	LEU4	0	4	4
	PHE5	0	3	3
	ALA6	0	10	10
	THR7	9	5	14
	ASP8	5	8	13
	SER9	1	9	10
	LEU10	0	5	5
	ILE11	1	17	18
	K12	40	10	51
	C-term	27	12	39
Ti-2	**Resid**	**Direct contact %**	**Indirect contact %**	**Total contact %**
	N-term	82	7	89
	GLY1	80	9	89
	HIS2	11	21	32
	THR3	0	27	27
	HIS4	32	30	62
	TYR5	24	17	40
	HIS6	0	18	18
	ALA7	0	9	9
	VAL8	0	3	3
	ARG9	92	2	94
	THR10	0	2	2
	GLN11	0	26	26
	THR12	1	10	11
	C-term	44	8	52

Table S5: Cross-cluster comparisons indicating 'matched' structures between the in-solution and surface-adsorbed conformational ensembles for Ti-1. Entries highlighted in yellow indicate a match between the top ten most populated clusters. Cluster ranks beyond 10 have negligible population in either ensemble.

Cluster rank (adsorbed)	Cluster rank (in solution)	RMSD (nm)
3	214	0.1915
3	223	0.1950
4	3	0.1916
4	45	0.1734
5	65	0.1460
5	87	0.1951
6	12	0.1667
6	138	0.1948
8	5	0.1642
8	11	0.1793
8	70	0.1884
8	73	0.1895
10	58	0.1978
10	87	0.1613
23	6	0.1653
25	8	0.1520
34	8	0.1968
48	9	0.1952
57	8	0.1994
60	5	0.1884
63	3	0.1942
73	7	0.1954
89	3	0.1860
100	5	0.1828
128	6	0.1907
158	4	0.1930
161	10	0.1876

Table S6: Cross-cluster comparisons indicating 'matched' structures between the in-solution and surface-adsorbed conformational ensembles for Ti-2. Entries highlighted in yellow indicate a match between the top ten most populated clusters. Cluster ranks beyond 10 have negligible population in either ensemble.

Cluster rank (adsorbed)	Cluster rank (in solution)	RMSD (nm)
1	15	0.1987
1	30	0.1996
1	64	0.1827
1	157	0.1823
3	1	0.1852
3	3	0.1741
3	38	0.1857
4	3	0.1688
4	122	0.1973
5	2	0.1953
5	77	0.1745
7	18	0.1899
8	5	0.1848
8	19	0.1877
8	67	0.1840
9	1	0.1799
10	26	0.1702
10	49	0.1220
10	64	0.1862
11	1	0.1766
11	6	0.1462
14	1	0.1766
21	1	0.1979
22	3	0.1761
23	5	0.1743
24	2	0.1907
24	9	0.1829
48	5	0.1846
51	10	0.1860
81	1	0.1963
81	2	0.1989
91	7	0.1998

Table S7: Definition of sites on each residue side-chain used to calculate residue—surface separation, and the distance cut-off used to determine 'direct' surface contact. Cut-offs for solvent-mediated contact involved a uniform addition of 2.0 Å to the 'direct' cut-offs.

Residue	Reference site	Cutoff distance (Å)
Ala	Beta carbon	5.5
Arg	Cz in the guanidinium group	5.0
Asp	Gamma carbon	5.0
Gln	Side chain oxygen	4.5
Gly	Alpha carbon	5.5
His	Center-of-mass of imidazole ring	5.5
Ile	Beta carbon	7.0
Leu	Gamma carbon	6.5
Lys	Side chain nitrogen	5.0
Pro	Center-of-mass of ring	5.5
Phe	Center-of-mass of benzyl ring	5.5
Ser	Side chain oxygen	5.0
Thr	Side chain oxygen	5.0
Tyr	Phenol oxygen	5.0
Val	Beta carbon	6.5
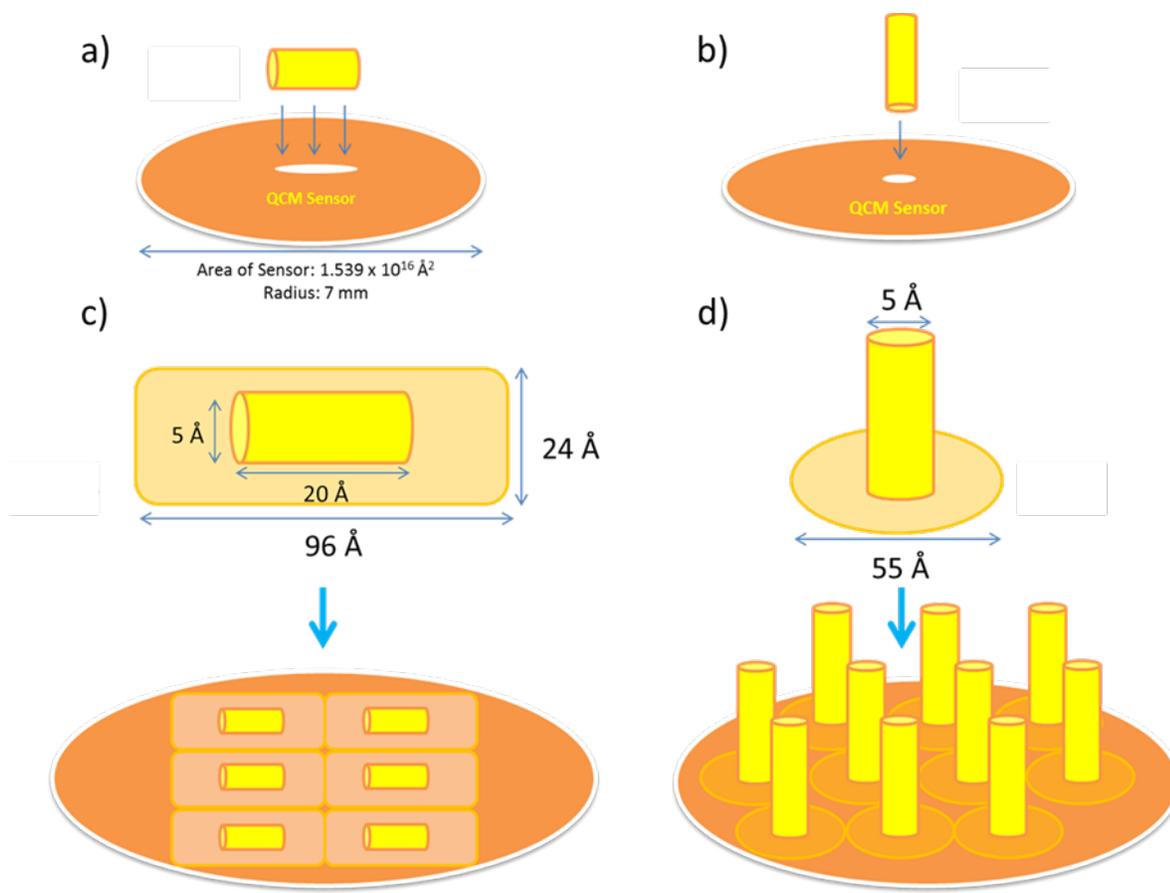
Figure S1: a) A representation of the Horizontal Binding Area (HBA) and b) Vertical Binding Area (VBA) of Ti-1. Using the estimated dimensions of each peptide*, a prediction of the area each peptide would occupy on the sensor surface was carried out. Estimations into the nature of the peptide layer reveal the likelihood of an extremely sparse monolayer coverage. c) An assessment of the HBA and d) VBA showed Ti-1 could comfortably sit in isolation within an area ≈23 (HBA) or ≈120 (VBA) times larger than its surface area. Similar results were obtained for Ti-2.

Dimensions of peptides estimated using MM2 energy minimisation calculations on ChemDraw
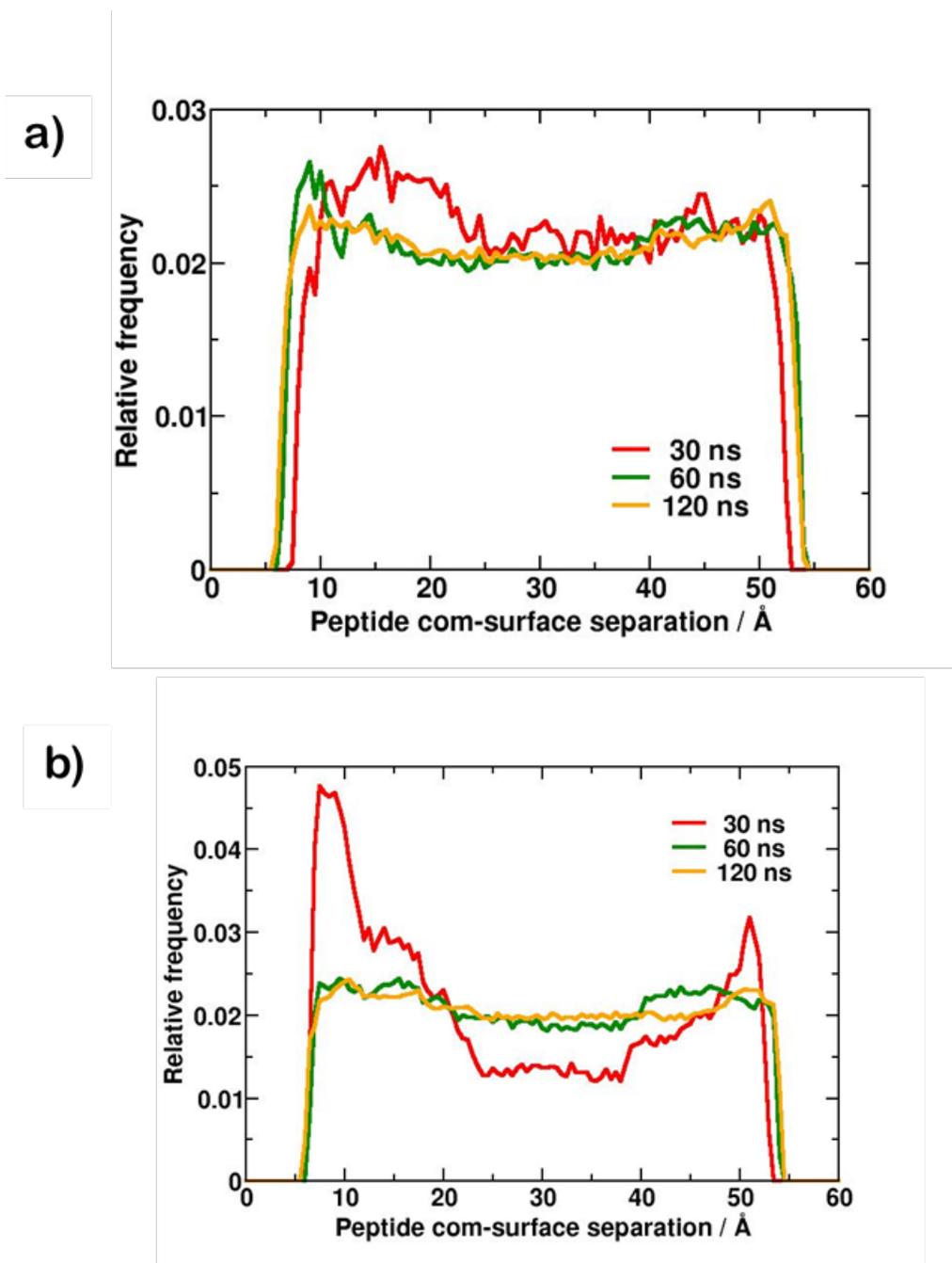
Figure S2: Histograms of the relative frequency with which the REST+MetaD CV (peptide *com*—titania distance) was sampled after 30, 60 and 120 ns for **a)** Ti-1 and **b)** Ti-2.
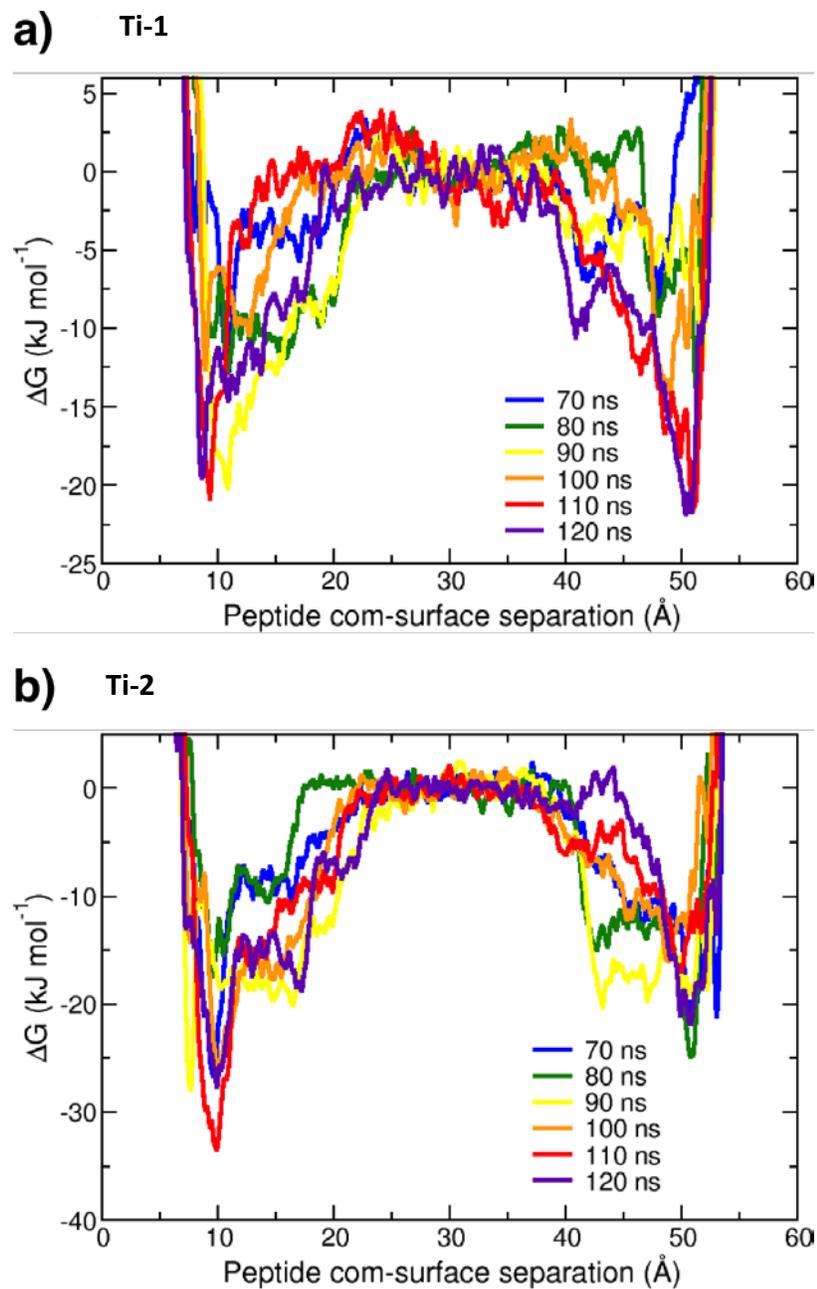
Figure S3: Evolution of the free energy profile corresponding to the surface adsorption of **a)** Ti-1 and **b)** Ti-2, as a function of REST+MetaD simulation time.
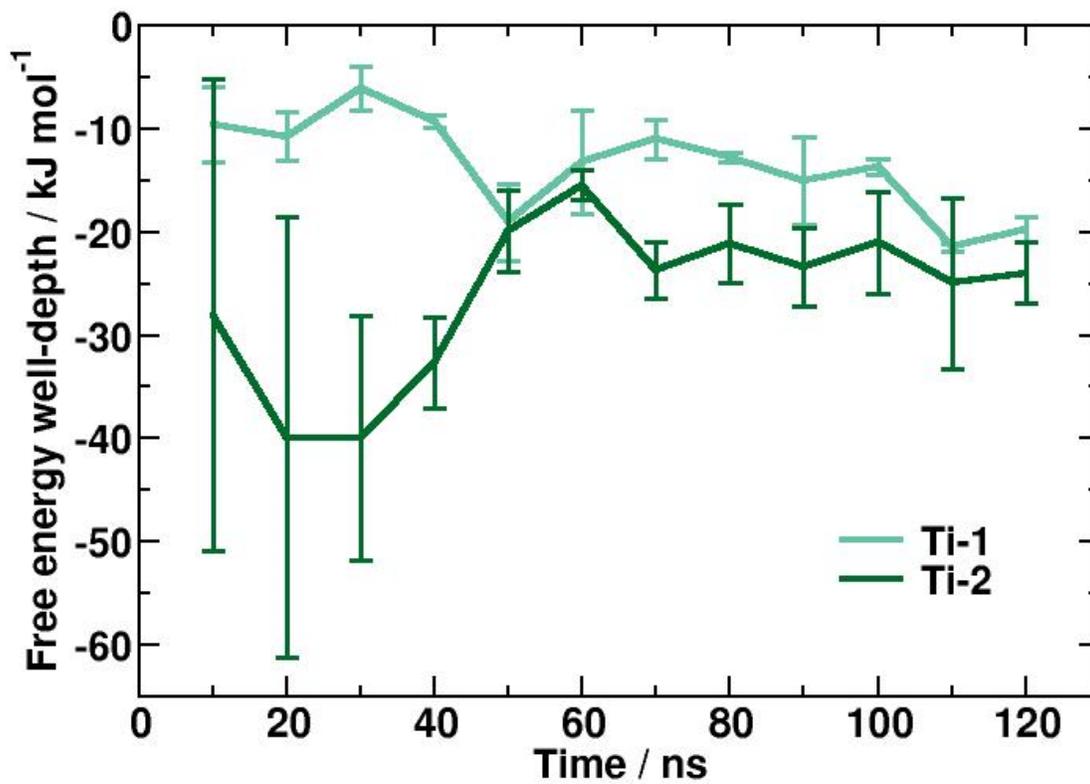
Figure S4: Evolution of the surface adsorption free energy of Ti-1 and Ti-2, as a function of REST+MetaD simulation time-step.
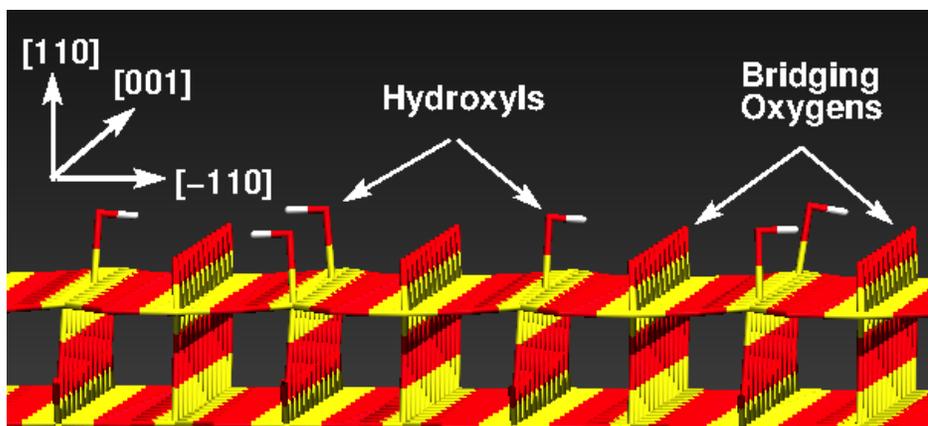
Figure S5: Summary of the structural model used to describe the negatively-charged, hydroxylated rutile TiO$_2$ (110) surface.
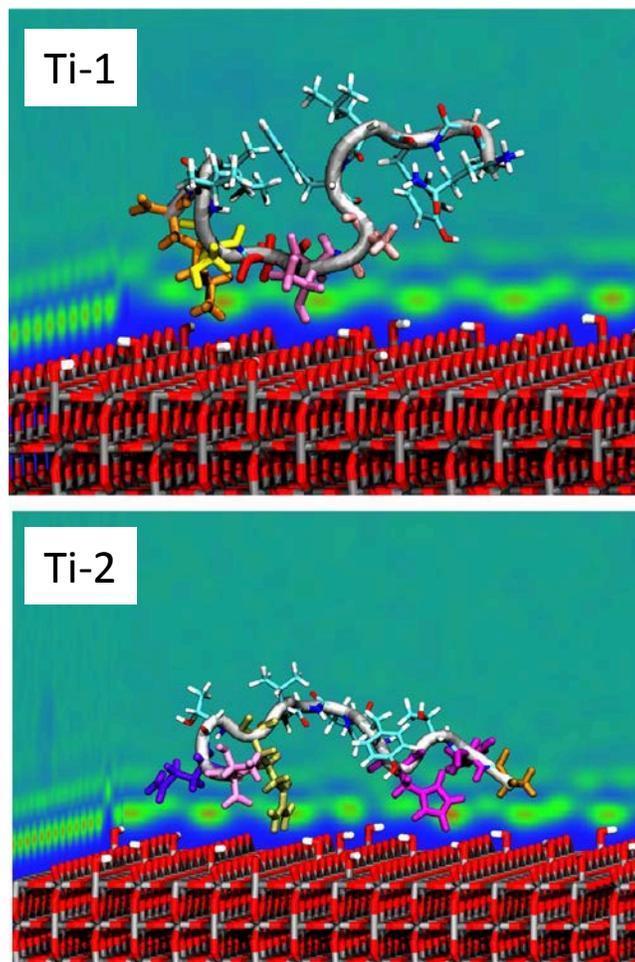
Figure S6: Alternative views of representative structures of Ti-1 and Ti-2 adsorbed at the aqueous rutile titania [110] interface, predicted using REST-only simulations. Color scheme of highlighted residues: K, orange; S, yellow; D, red; T, mauve; A, pink; Q, violet; R, tan; H, purple; G, ochre. The image background indicates the averaged interfacial solvent structuring. Liquid water is not shown for clarity.
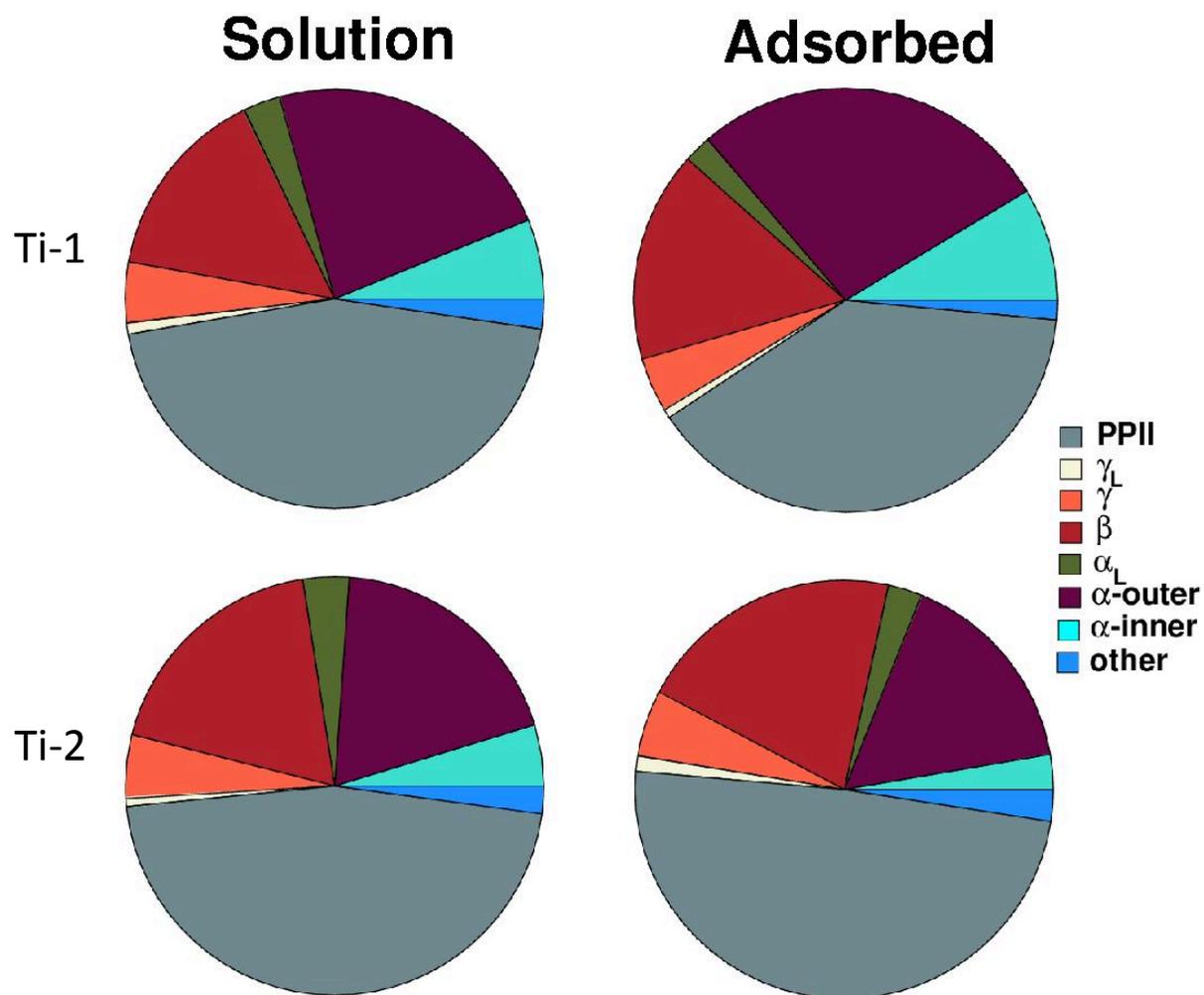
Figure S7: Population of secondary structure motifs taken from Ramachandran analysis of the reference replica generated using the REST-only trajectories of Ti-1 and Ti-2, for both the 'in-solution' and 'surface-adsorbed' states at the aqueous titania interface.
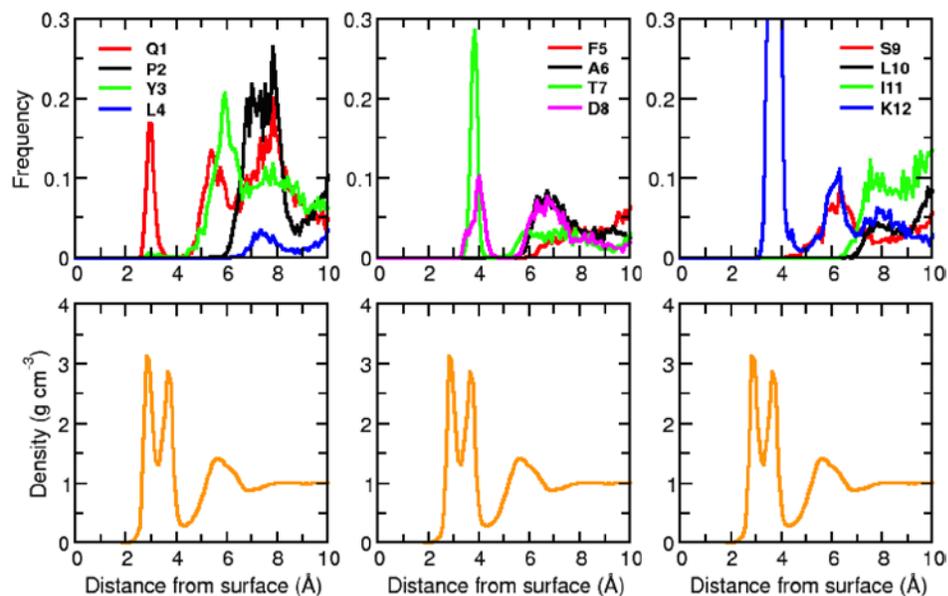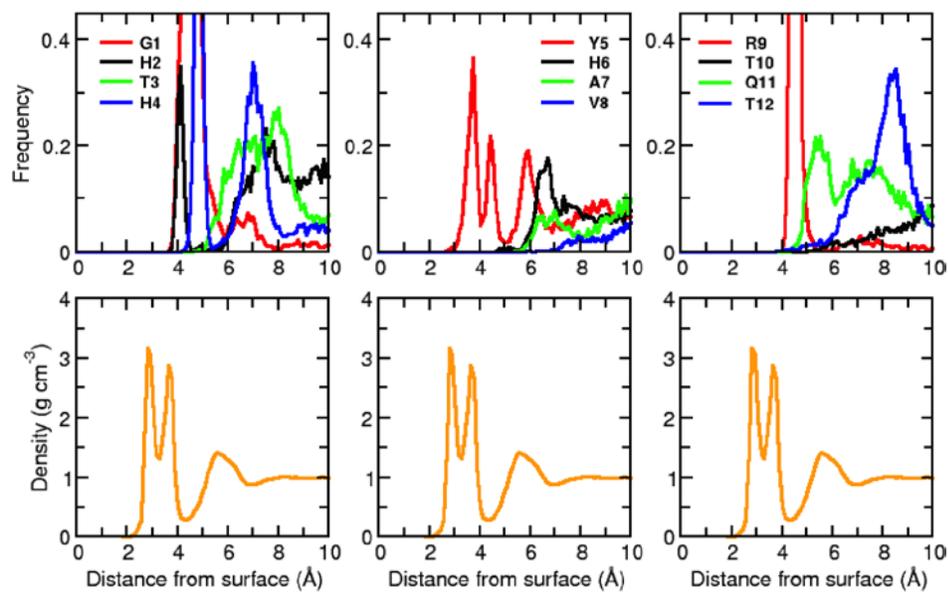
Figure S8: Histograms of residue-surface contact for peptides **a)** Ti-1 and **b)** Ti-2, generated from REST-only simulation data. The vertical interfacial water density profile (orange) is shown in each case for comparison.
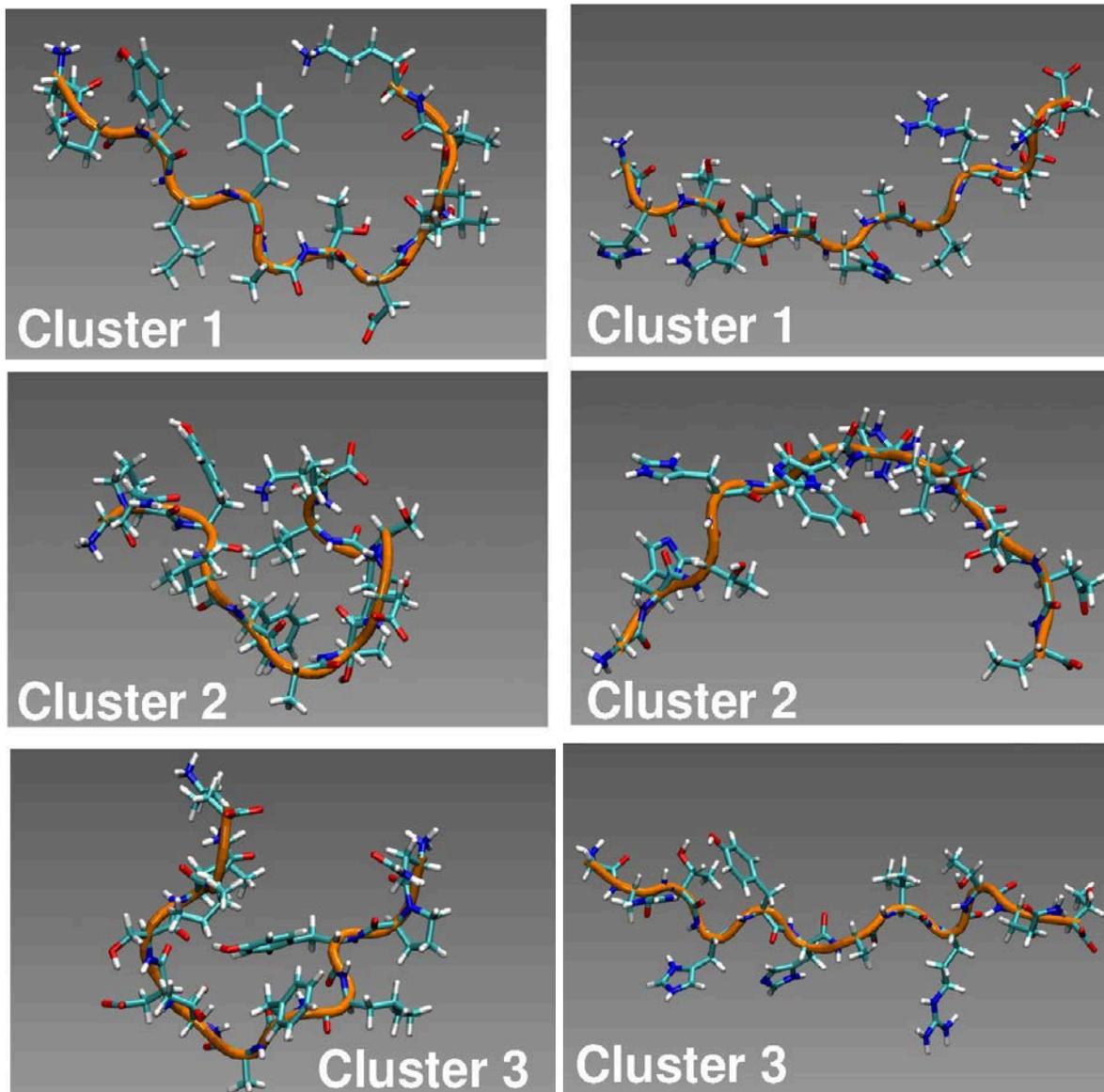
Figure S9: Representative structures of the three most populated clusters for Ti-1 (left) and Ti-2 (right) in solution predicted from REST-only simulations. Water not shown for clarity.
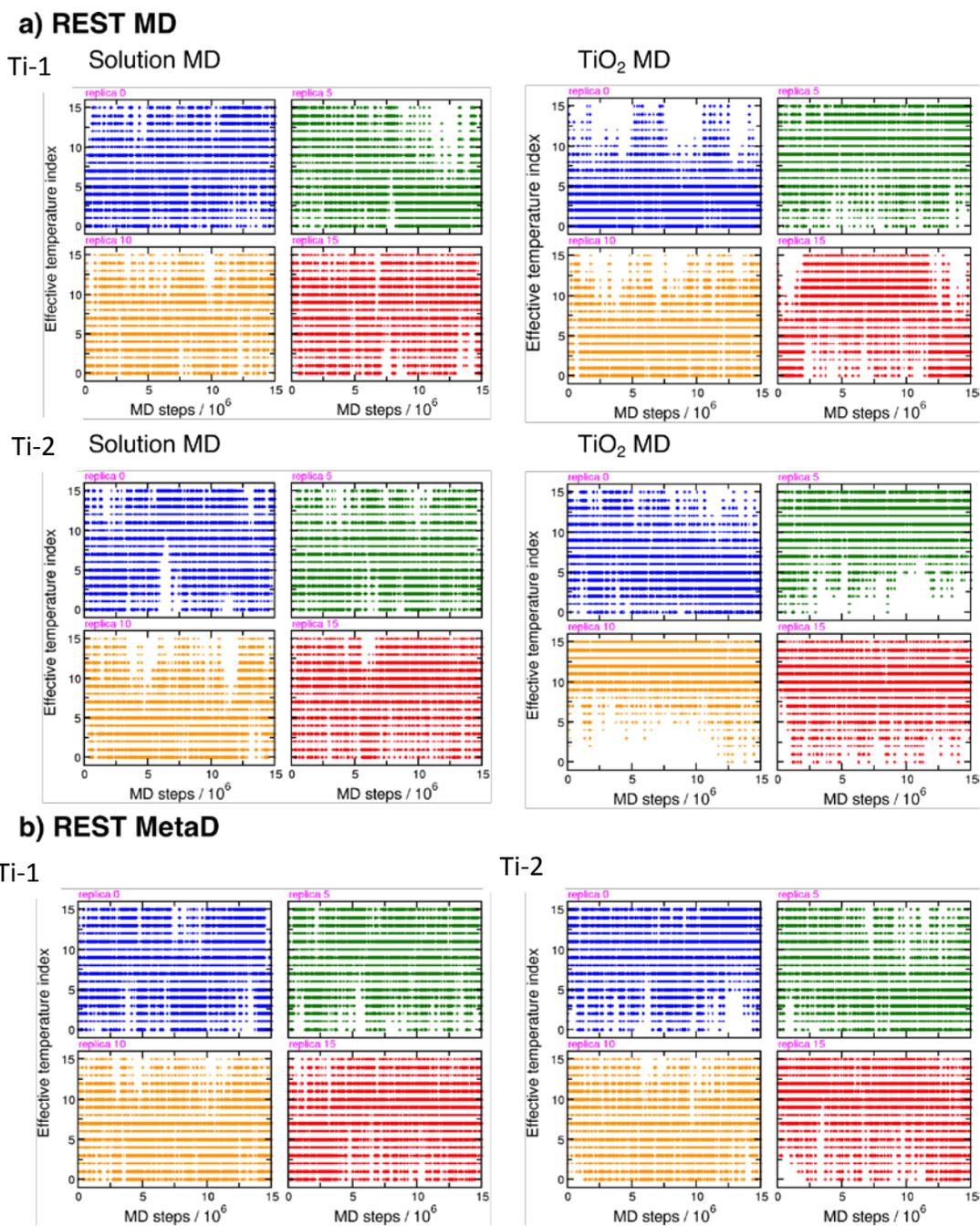
Figure S10: Mobility of replicas through "effective temperature" space for the **a)** REST-only and **b)** REST-MetaD simulations. Blue, green, orange and red correspond to mobilities for the 1st, 6th, 11th and 16th replicas respectively.
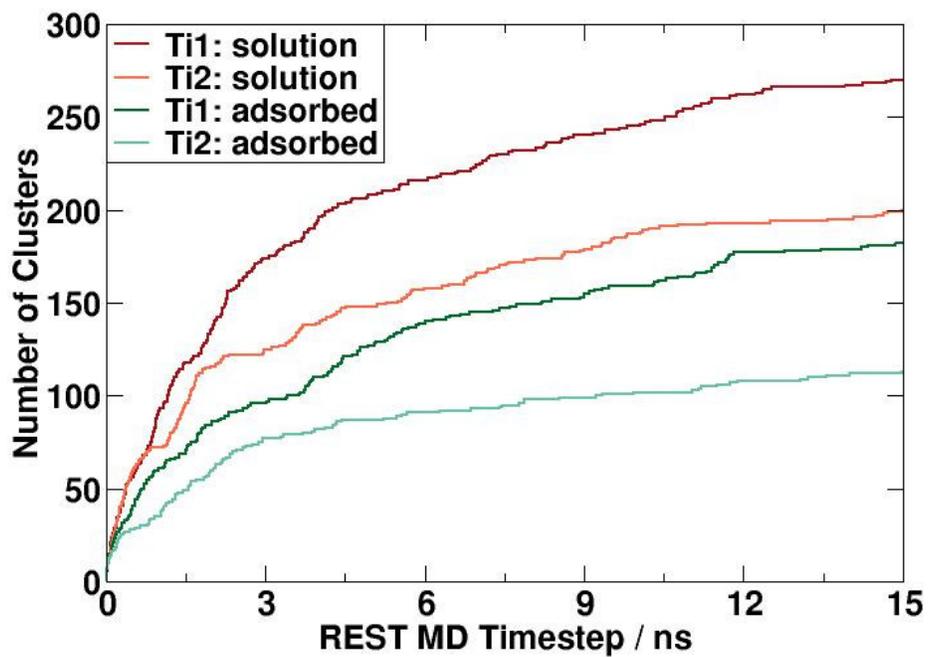
Figure S11: Number of distinct clusters as a function of REST-only MD simulation timestep, for Ti-1 and Ti-2, both in-solution and adsorbed at the aqueous titania interface.
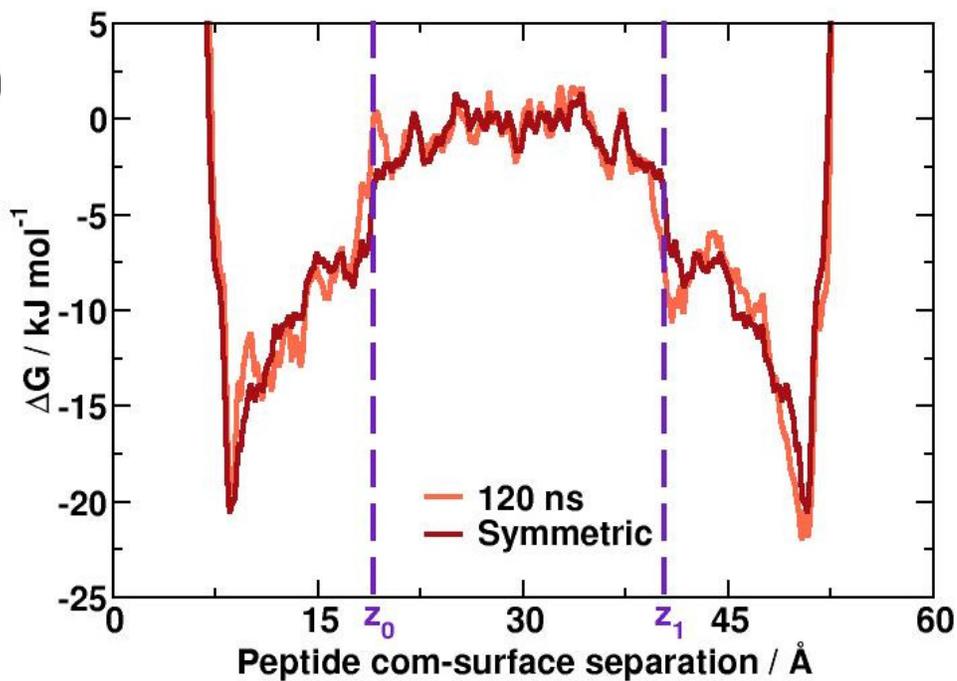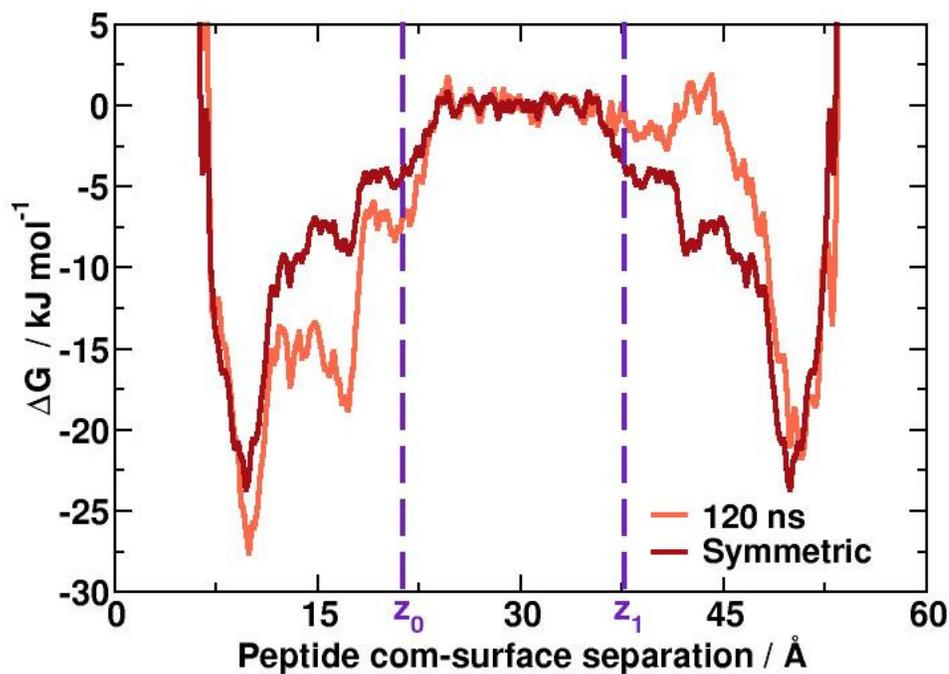
Figure S12: Symmetrized final free energy profiles of **a)** Ti-1 and **b)** Ti-2 adsorbed at the aqueous aqueous titania interface. z_0 and z_1 mark the extent of the 'adsorbed' zone for the peptide defined in Equations (1-5) in **Section S1: Additional Methodology**.
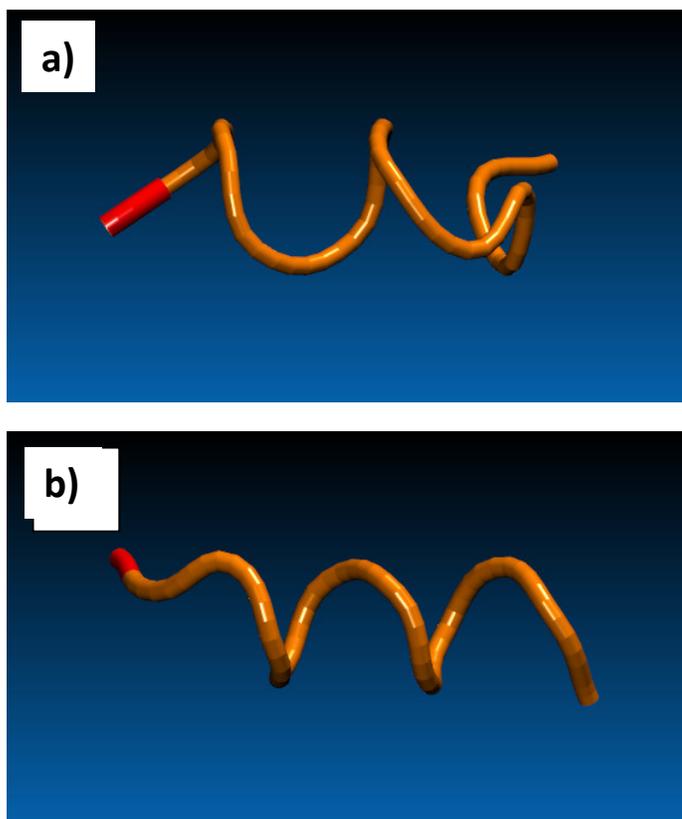
Figure S13: Most populated clusters determined from implicit-solvent REMD simulations of **a)** Ti-1 and **b)** Ti-2. Only the backbone conformation is shown for clarity – red indicates the position of the N-terminus.
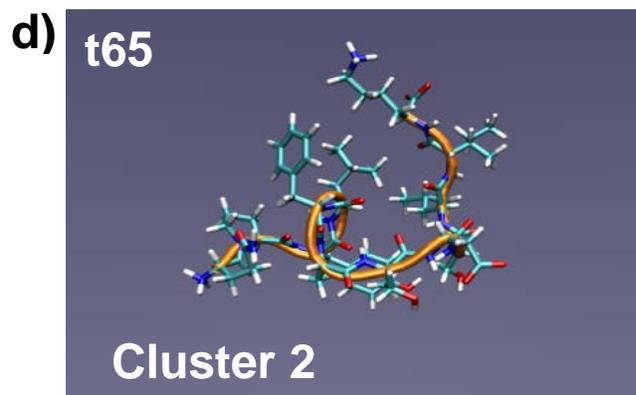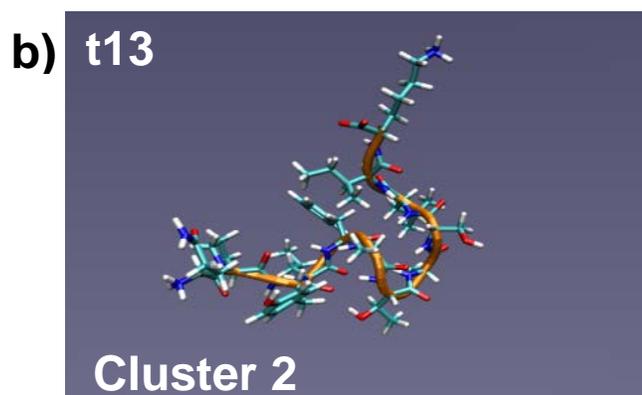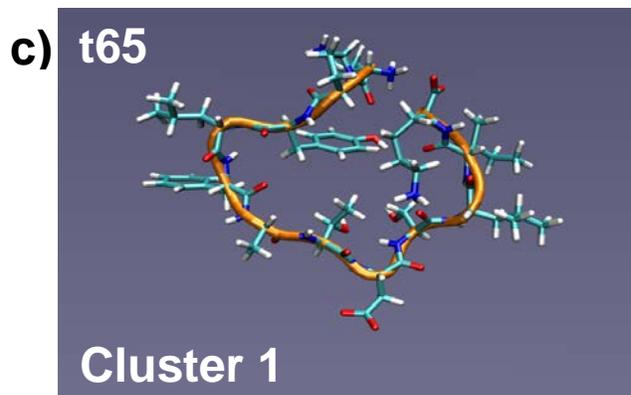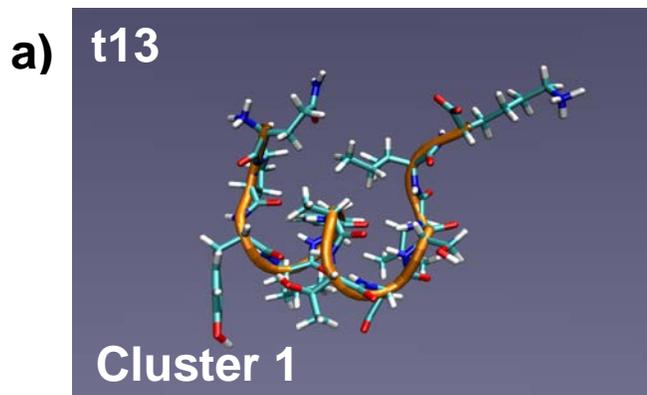
Figure S14: Top-two most populated clusters generated from the post-implicit-solvent REMD simulations, using standard MD simulation in explicit solvent. **a)** and **b)** Ti-1_t13; **c)** and **d)** Ti-1_t65.
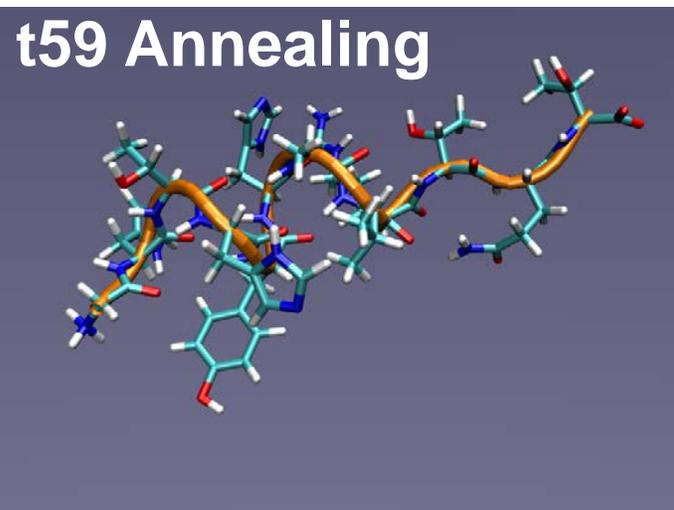
Figure S15: Most populated cluster for Ti-2 generated from the post-implicit-solvent REMD simulations, using annealing MD simulation in explicit solvent.

References

(1) Fischer, H., Polikarpov, I., Craievich, A. F., Average Protein Density Is a Molecular-Weight-Dependent Function. *Protein Science* **2004,** *13*, 2825-2828.

(2) Predota, M.; Bandura, A. V.; Cummings, P. T.; Kubicki, J. D.; Wesolowski, D. J.; Chialvo, A. A.; Machesky, M. L., Electric Double Layer at the Rutile (110) Surface. 1. Structure of Surfaces and Interfacial Water from Molecular Dynamics by Use of Ab Initio Potentials. *J Phys Chem B* **2004,** *108*, 12049-12060.

(3) Nose, S., A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol Phys* **1984,** *52*, 255-268.

(4) Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Physical Review A* **1985,** *31*, 1695 - 1697.

(5) Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993,** *98*, 10089-10092.

(6) T. Terakawa, T. K., S. Takada, On Easy Implementation of a Variant of the Replica Exchange with Solute Tempering in Gromacs. *J. Comput. Chem.* **2011,** *32*, 1228-1234.

(7) L. B. Wright, T. R. W., Efficient Conformational Sampling of Peptides Adsorbed onto Inorganic Surfaces: Insights from a Quartz Binding Peptide. *Phys. Chem. Chem. Phys.* **2013,** *15*, 4715-4726.

(8) Tang, Z. H.; Palafox-Hernandez, J. P.; Law, W. C.; Hughes, Z. E.; Swihart, M. T.; Prasad, P. N.; Knecht, M. R.; Walsh, T. R., Biomolecular Recognition Principles for Bionanocombinatorics: An Integrated Approach to Elucidate Enthalpic and Entropic Factors. *Acs Nano* **2013,** *7*, 9632-9646.

(9) Palafox-Hernandez, J. P.; Tang, Z. H.; Hughes, Z. E.; Li, Y.; Swihart, M. T.; Prasad, P. N.; Walsh, T. R.; Knecht, M. R., Comparative Study of Materials-Binding Peptide Interactions with Gold and Silver Surfaces and Nanostructures: A Thermodynamic Basis for Biological Selectivity of Inorganic Materials. *Chem Mater* **2014,** *26*, 4960-4969.

(10) 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.

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

(12) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., Peptide Folding: When Simulation Meets Experiment. *Angew Chem Int Edit* **1999,** *38*, 236-240.

(13) Khandogin, J., Brooks, C. L. , Toward the Accurate First-Principles Prediction of Ionization Equilibria in Proteins. *Biochemistry* **2006,** *45*, 9363-9373.

(14) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M., Progress in the Prediction of Pk(a) Values in Proteins. *Proteins-Structure Function and Bioinformatics* **2011,** *79*, 3260-3275.

(15) Schneider, J.; Colmbi Ciacchi, L. , Specific Material Recognition by Small Peptides Mediated by the Interfacial Solvent Structure. *J Am Chem Soc* **2012,** *134*, 2407-2413.

(16) Wright, L. B.; Palafox-Hernandez, J. P.; Rodger, P. M.; Corni, S.; Walsh, T. R., Facet Selectivity in Gold Binding Peptides: Exploiting Interfacial Water Structure. *Chemical Science* **2015,** *6*, 5204-5214.