Electronic Supplementary Information: Efficient conformational sampling of peptides adsorbed onto inorganic surfaces: Insights from a quartz binding peptide.

Louise B. Wright[a] and Tiffany R. Walsh[b*]

a) Dept. of Chemistry and Centre for Scientific Computing, University of Warwick, Coventry, CV4 7AL, U.K.

b) Institute for Frontier Materials, Deakin University, Geelong, 3216 Vic., Australia.

Contents

*tiffany.walsh@deakin.edu.au

T-REMD replica temperatures

For each T-REMD variant applied in this work 36 replicas were simulated, at temperatures (in K) of 295.00, 297.49, 300.00, 302.52, 305.06, 307.61, 310.18, 312.77, 315.37, 317.99, 320.63, 323.28, 325.95, 328.61, 331.31, 334.03, 336.77, 339.53, 342.31, 345.10, 347.91, 350.74, 353.58, 356.45, 359.32, 362.23, 365.15, 368.11, 371.07, 374.05, 377.04, 380.05, 383.09, 386.14, 389.22 and 392.29.

Additional Methodolgy

Cluster Analysis

The Daura clustering algorithm, as implemented within the the Gromacs g_cluster code, was used to classify the adsorbed structures of S1 into clusters,[2] using a root-mean-squared deviation (RMSD) cut-off of 2 Å . The same procedure and cut-off were used to assign structures into clusters in the previous T-REMD study of S1 in solution.[1]

The RMSD distance of backbone atoms between pairs of structures belonging to the three most populated clusters in each of the REMD-based methods were carried out to quantify structural similarities. Centroid structures of each cluster (*i.e.* the structure within the cluster which had the smallest RMSD to all the other structures within the same cluster) were chosen; the clusters themsleves were identified from clustering analysis over the final 4×10^6 MD steps only. Clusters from two different REMD-based methods were classified as 'identical' if the RMSD between their centroid structures was less than the cut-off used for the original clustering analysis (2 Å). Similarities between the structures of the three most populated clusters of S1 found in this study, where the peptide was adsorbed onto the (100) α-quartz surface, and those of S1 in solution[1] were quantified in the same manner.

In addition to clustering the structures on the basis of their backbone atoms, structure cluster analysis (using the algorithm described above) was also carried out over peptide side-chain atoms, using a cut-off of 2.5 Å.

Ramachandran Analysis

Combinations of specific torsion angles are characteristic of different peptide secondary structures, such as α-helix (-60,-45), β-sheet (-135,-135), and PPII (-75, 150). To quantify whether each method sampled the possible peptide secondary structures with similar probability, ϕ-ψ phase-space was subdivided into bins centred around the characteristic range of values for each type of secondary-structure region (see ESI†, Figure E7). Over the last 4×10^6 MD steps of the trajectory, the percentage of dihedral angles found to lie within each bin was calculated. Comparisons were made between different REMD methods. Residues

contributing to the α and α_L Ramachandran signatures in each REMD method were quantified in a similar manner.

Hydrogen Bonding Analysis

Geometric criteria to define the existence of a hydrogen bond[3] have been employed here. Namely, a hydrogen bond is defined as being present if simultaneously the D–A distance is less than 3.5 Å and the HDA angle is less than 30°, where D (hydrogen bond donor) and A (hydrogen bond acceptor) are O or N atoms. Using these criteria, the average number of intra-peptide, peptide-water and peptide-quartz hydrogen bonds formed by S1 over the last 4×10^6 MD steps of each trajectory were calculated.

Binding Residue and Motif Analysis

Structures belonging to each of the three most populated clusters for each trajectory were analysed to determine which residues mediate surface binding. A residue was categorised as being bound to the surface if the average residue-surface separation (z) of any of its atoms was less than 3 Å. As in previous works, the top of the quartz surface ($z = 0$) was defined as the average z coordinate of silicon atoms in the top layer of the quartz slab.[4–6]

In addition, for direct comparison between the different variants of REMD-based simulation conducted here and a previous bioinformatics study,[5,7] dyad and triad sequence motifs contributing to peptide binding have been identified. This characterisation has been conducted over all the peptide structures in last 2×10^6 MD steps of each trajectory, not only those which belong to the most stable clusters. At each time frame, the criteria for a binding dyad/triad was met if the residue-surface separations of two/three sequential residues were all less than the cut-off value of 3 Å.

Energetic Analysis

To investigate whether the most populated clusters identified by the REMD-based simulations represented low-energy peptide-quartz configurations, the relative potential energies of a

number of systems were probed by carrying out several additional standard MD simulations of 5 ns duration. This was particularly important to help identify whether helical structures were in fact low-energy adsorbed states of S1, or just an artefact of the initial replica population chosen. Hence, six different snapshots of S1 adsorbed onto the (100) α-quartz surface, representative of both extended and helical structures, and solution clusters 1 to 4,[1] were taken from different REMD simulations. Each system was otherwise identical (*e.g.* in terms of periodic cell dimensions and chemical composition). Simulations were conducted in the *NVT* ensemble, at a temperature of 300 K. The average potential energy of the entire system was calculated over the last 2 ns of the trajectory with a block averaging method used to estimate the errors[8] in each case.

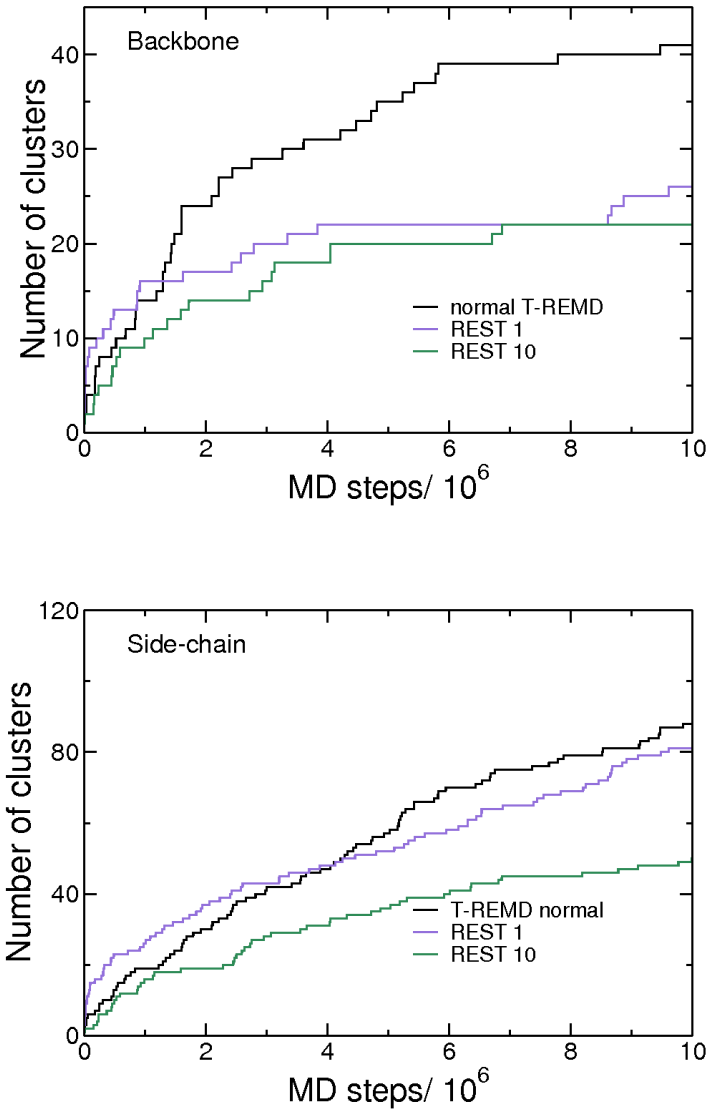

Figure E1: Number of S1 backbone (top) and side-chain (bottom) clusters found in REST10 (REST with an exchange attempted every 10000 MD steps), REST 1 (REST with an exchange attempted every 1000 MD steps) and T-REMD normal simulations as a function of MD steps.

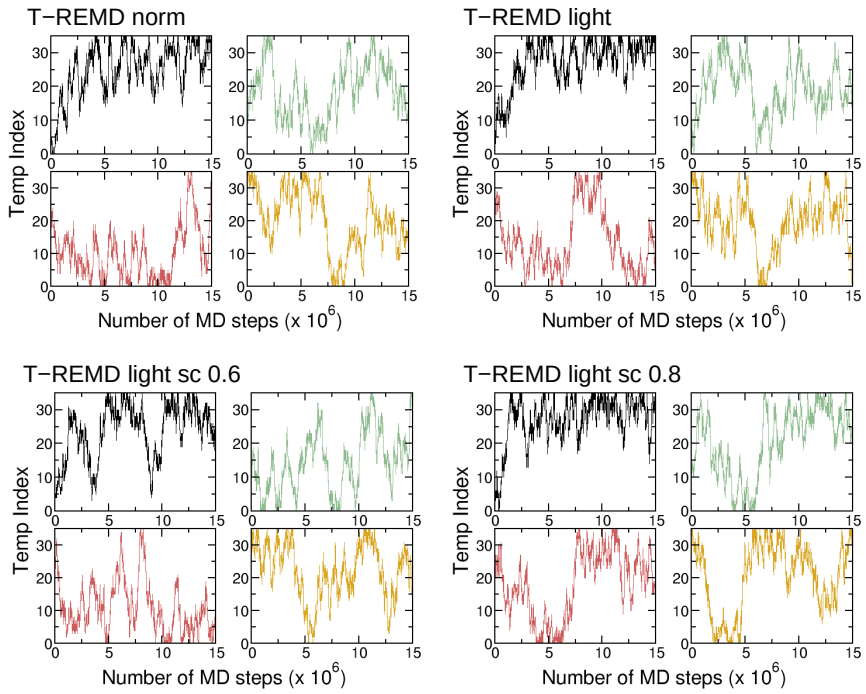

Figure E2: Temperature mobilities for 4 out of the 36 replicas in each T-REMD simulation. Replicas shown are (clockwise from top left) 1 (black), 12 (green), 24 (red) and 36 (gold).

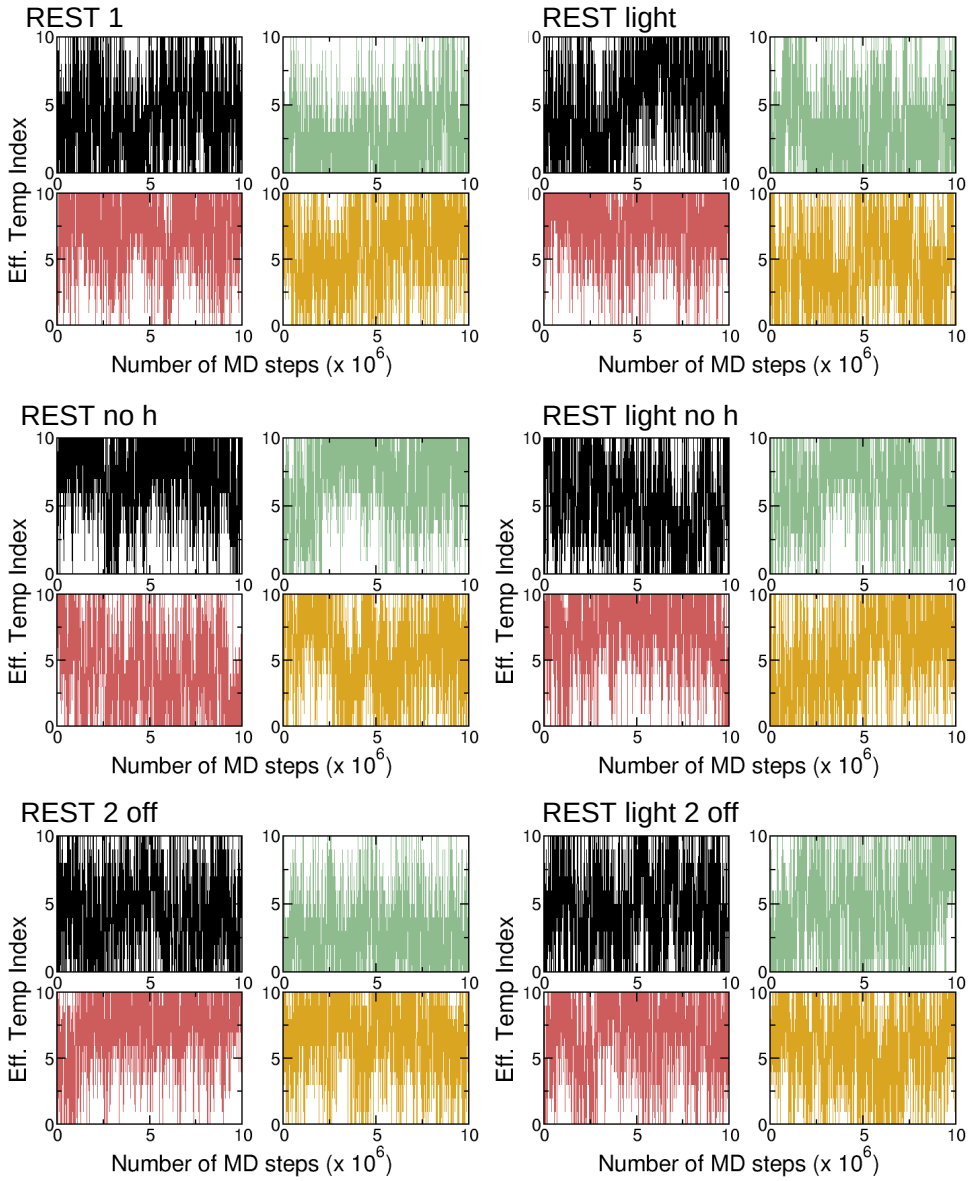

Figure E3: Effective temperature mobilities for 4 out of the 11 replicas in each REST simulation. Replicas shown are (clockwise from top left) 1 (black), 4 (green), 8 (red) and 11 (gold).

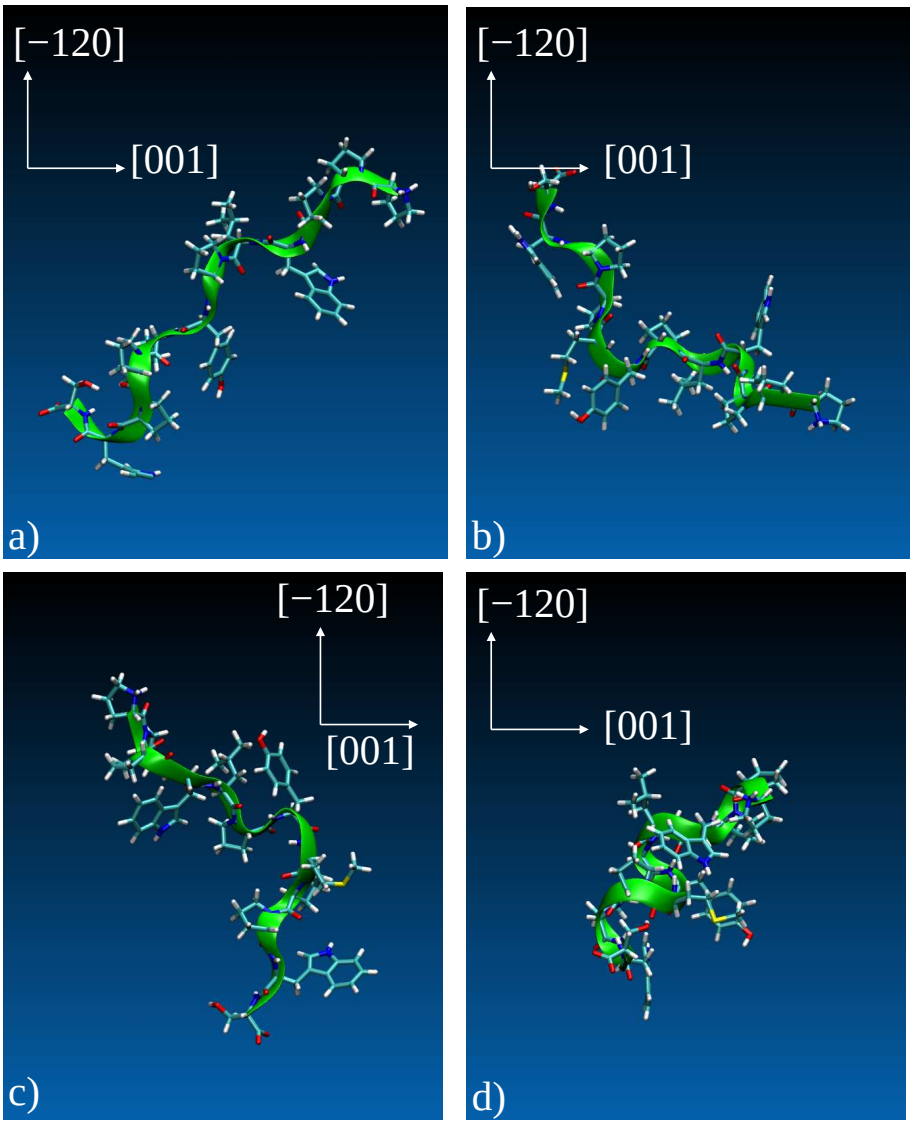

Figure E4: Commonly observed structures of S1 in the most populated clusters after REMD simulations of the peptide adsorbed onto the (100) α-quartz surface: **a)** extended, **b)** extended with kinks (akin to solution clusters 2 and 4 in Reference [2]), **c)** 'C' shaped (akin to solution cluster 1) and **d)** helix (found as a heavily populated cluster using the 'REST 2 off', 'REST light' and 'REST light 2 off' methods). Water and surface atoms have been omitted for clarity.

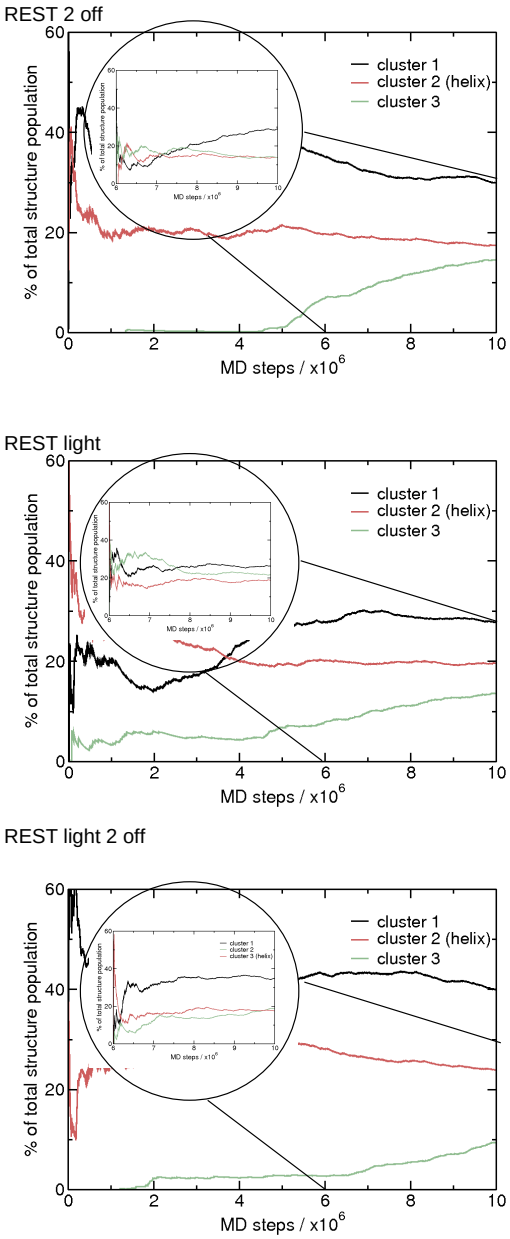

Figure E5: Percentage of the total structure population belonging to each of the three most populated clusters as a function of MD steps shown for 'REST 2 off', 'REST light' and 'REST light 2 off' simulations. Data shown for cluster analysis performed over the entire trajectory and over the last $4x10^6$ MD steps only (inset). Clusters comprised of helical structures are shown in red.

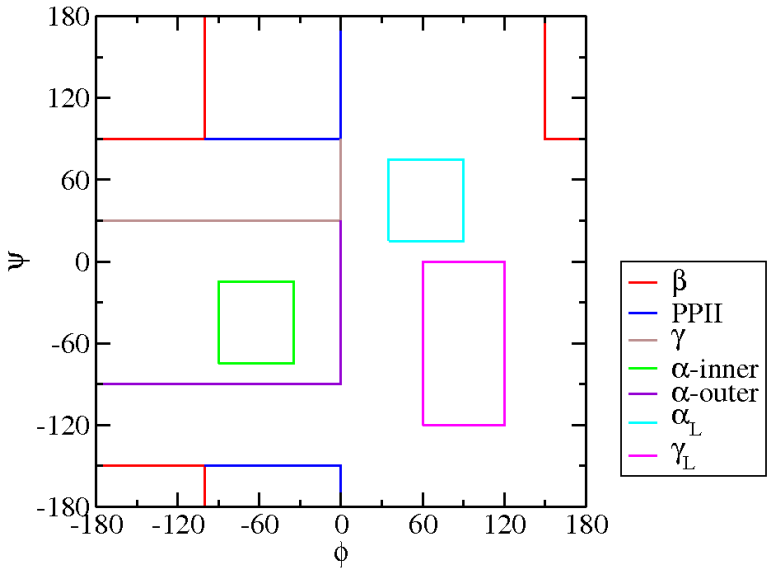

Figure E6: Boundaries in ϕ/ψ space demarcing the principle regions in a Ramachandran plot.

method	% helical-type structures	intra peptide	pept-water	pept-quartz
normal T-REMD	0.0	0.49	27.37	4.89
light T-REMD	0.0	0.23	27.84	4.95
light sc 0.6 T-REMD	0.0	0.42	28.09	4.90
light sc 0.8 T-REMD	0.0	0.27	27.53	5.49
REST1	10.4	0.68	28.16	4.77
REST no h	0.0	0.15	27.97	4.61
REST 2 off	13.9	0.76	28.39	3.80
REST light	18.8	1.01	27.24	4.56
REST light no h	0.0	0.23	27.92	5.41
REST light 2 off	17.5	1.14	27.36	4.64

Table E1: Average numbers of hydrogen bonds per MD frame.

method	cluster	sol 1	sol 2	sol 3	sol 4
T-REMD normal	1	0.262	0.247	0.591	0.213
T-REMD normal	2	0.382	0.351	0.764	0.277
T-REMD normal	3	0.504	0.563	0.165	0.674
T-REMD light	1	0.385	0.365	0.788	0.267
T-REMD light	2	0.317	0.316	0.670	0.279
T-REMD light	3	0.280	0.246	0.437	0.442
T-REMD light 0.6 sc	1	0.388	0.366	0.786	0.274
T-REMD light 0.6 sc	2	0.296	0.320	0.652	0.273
T-REMD light 0.6 sc	3	0.352	0.344	0.716	0.292
T-REMD light 0.8 sc	1	0.352	0.329	0.764	0.213
T-REMD light 0.8 sc	2	0.195	0.291	0.582	0.328
T-REMD light 0.8 sc	3	0.330	0.318	0.658	0.272
REST1	1	0.394	0.350	0.787	0.235
REST1	2	0.341	0.322	0.668	0.281
REST1	3	0.293	0.313	0.678	0.304
REST no h	1	0.147	0.279	0.579	0.314
REST no h	2	0.323	0.163	0.644	0.198
REST no h	3	0.348	0.408	0.301	0.563
REST 2 off	1	0.348	0.328	0.702	0.262
REST 2 off	2	0.373	0.366	0.498	0.418
REST 2 off	3	0.260	0.129	0.587	0.278
REST light	1	0.337	0.312	0.696	0.279
REST light	2	0.365	0.339	0.783	0.209
REST light	3	0.364	0.350	0.511	0.395
REST light no h	1	0.372	0.316	0.760	0.199
REST light no h	2	0.175	0.336	0.506	0.391
REST light no h	3	0.462	0.528	0.225	0.621
REST light 2 off	1	0.370	0.341	0.770	0.259
REST light 2 off	2	0.377	0.318	0.729	0.276
REST light 2 off	3	0.353	0.341	0.455	0.449

Table E2: Average RMSD of backbone atom positions of all structures in the three most populated clusters identified by each REMD-based method from the centriod structure of the four most populated solution clusters of S1.[1] Clusters for which the RMSD was less than the cut-off used in the backbone cluster analysis (2 Å) are highlighted in red, whilst those for which it lay in the range 2–3 Å are highlighted in green.

n	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
n	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
n	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_w	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_w	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_w	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_0.6	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_0.6	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_0.6	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_0.8	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_0.8	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
l_0.8	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_1	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_1	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_1	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re nh	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re nh	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re nh	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re 2 off	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re 2 off	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re 2 off	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l nh	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l nh	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l nh	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l 2 off	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l 2 off	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C
re_l 2 off	N	P1	P2	P3	W4	L5	P6	Y7	M8	P9	P10	W11	S12	C

Table E3: Residues underlined are those which are on average bound to the quartz surface. Results are shown for the three most populated backbone clusters identified by clustering analysis carried out over the last $4x10^6$ MD steps of each simulation. (Index: ‘n’= normal T-REMD, ‘l_w’= light T-REMD, ‘l_0.6’= light sc 0.6 T-REMD, ‘l_0.8’= light sc 0.8 T-REMD, ‘re_1’= REST1, ‘re nh’= REST no h, ‘re 2 off’= REST 2 off, ‘re_l’= REST light, ‘re_l nh’= REST light no h,‘re_l 2 off’= REST light 2 off.)

	1	2	3	4
normal T-REMD	4,5 (38.4)	7,8 (27.2)	6,7 (22.6)	10,11 (18.9)
light T-REMD	3,4 (34.5)	5,6 (29.9)	6,7 (29.3)	4,5 (22.1)
light 0.6 sc T-REMD	5,6 (33.3)	6,7 (25.9)	4,5 (24.3)	3,4 (22.9)
light 0.8 sc T-REMD	10,11 (40.5)	4,5 (37.6)	6,7 (25.1)	7,8 (18.2)
REST1	4,5 (32.3)	7,8 (25.2)	10,11 (18.5)	6,7 (16.6)
REST no h	10,11 (36.4)	4,5 (32.8)	5,6 (26.3)	7,8 (23.6)
REST 2 off	6,7 (30.1)	10,11 (12.8)	4,5 (11.1)	5,6 (8.0)
REST light	4,5 (38.8)	6,7 (38.0)	10,11 (8.7)	9,10 (7.4)
REST light no h	10,11 (57.3)	6,7 (51.6)	3,4 (41.4)	7,8 (39.1)
REST light 2 off	4,5 (30.7)	7,8 (22.1)	10,11 (16.6)	6,7 (14.5)

Table E4: Top 4 binding dyad motifs in the REMD-based simulations, calculated over the last $2x10^6$MD steps of each trajectory. Percentage of trajectory for which the dyad makes binding contact is given in parenthesises. Dyads present in S1 which were found to be enriched dyads in quartz binding peptides[5] are highlighted as follows: PW, WL, PP, LP, (given in order of reducing Motif Enrichment Factor).

	1	2	3	4
normal T-REMD	4,5,6 (15.6)	10,11,12 (14.0)	5,6,7 (12.2)	3,4,5 (8.2)
light T-REMD	6,7,8 (13.9)	5,6,7 (13.0)	10,11,12 (11.8)	7,8,9 (8.5)
light 0.6 sc T-REMD	2,3,4 (17.3)	4,5,6 (12.9)	1,2,3 (11.8)	10,11,12 (9.9)
light 0.8 sc T-REMD	10,11,12 (14.6)	6,7,8 (10.8)	9,10,11 (6.5)	4,5,6 (5.4)
REST1	4,5,6 (10.7)	2,3,4 (6.5)	3,4,5 (5.4)	1,2,3 (5.2)
REST no h	10,11,12 (26.4)	4,5,6 (18.9)	2,3,4 (11,4)	6,7,8 (10.7)
REST 2 off	4,5,6 (5.9)	9,10,11 (4.5)	10,11,12 (4.3)	6,7,8 (2.5)
REST light	3,4,5 (14.9)	2,3,4 (8.1)	10,11,12 (5.6)	6,7,8 (2.9)
REST light no h	6,7,8 (40.3)	10,11,12 (33.5)	2,3,4 (9.5)	4,5,6 (4.7)
REST light 2 off	10,11,12 (8.2)	4,5,6 (6.1)	6,7,8 (5.0)	3,4,5 (4.0)

Table E5: Top 4 binding triad motifs in the REMD-based simulations, calculated over the last $2x10^6$MD steps of each trajectory. Percentage of trajectory for which the dyad makes binding contact is given in parenthesises. Dyads present in S1 which were found to be enriched dyads in quartz binding peptides[5] are highlighted as follows: PPW, PWL and WLP (given in order of reducing Motif Enrichment Factor).

References

[1] R. Notman, E. E. Oren, C. Tamerler, M. Sarikaya, R. Samudrala, and T. R. Walsh, *Biomacromolecules*, 2010, **11**, 3266.

[2] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, *Angew. Chem. Int. Ed.*, 1999, **38**, 236.

[3] P. Jedlovszky, J. P. Brodholt, F. Bruni, M. A. Ricci, A. U. Soper, and R. Vallauri, *J. Chem. Phys.*, 1998, **108**, 8528.

[4] R. Notman and T. R. Walsh, *Langmuir* , 2009, **25**, 1638.

[5] E. E. Oren, R. Notman, I. W. Kim, J. S. Evans, T. R. Walsh, R. Samudrala, C. Tamerler, and M. Sarikaya, *Langmuir* , 2010, **26**, 11003.

[6] L. B. Wright and T. R. Walsh, *J. Phys. Chem. C*, 2012, **116**, 1933.

[7] E. E. Oren, C. Tamerler, D. Sahin, M. Hnilova, U. O. S. Seker, M. Sarikaya, and R. Samudrala, *Bioinformatics* , 2007, **23**, 2816.

[8] B. Hess, *J. Chem. Phys.* , 2002, **116**, 209.