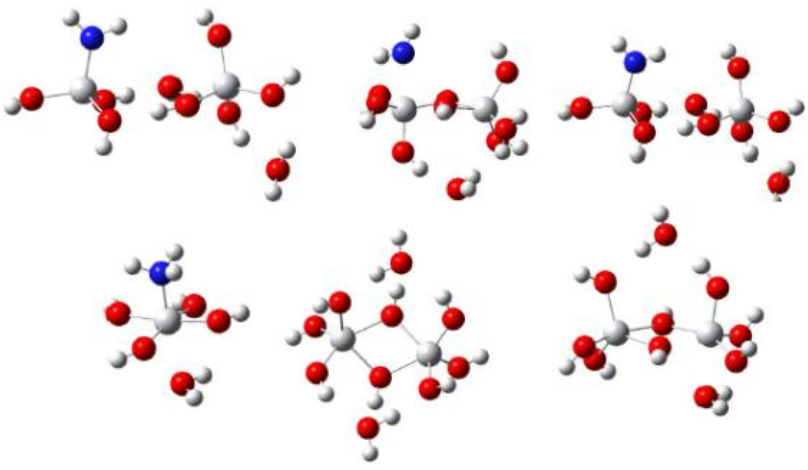
Relazione Scientifica sui Risultati dell'Attività di Ricerca svolta presso la Pennsylvania State University, Department of Mechanical and Nuclear Engineering.

Collaborazione con il Professor Adri C.T. van Duin.

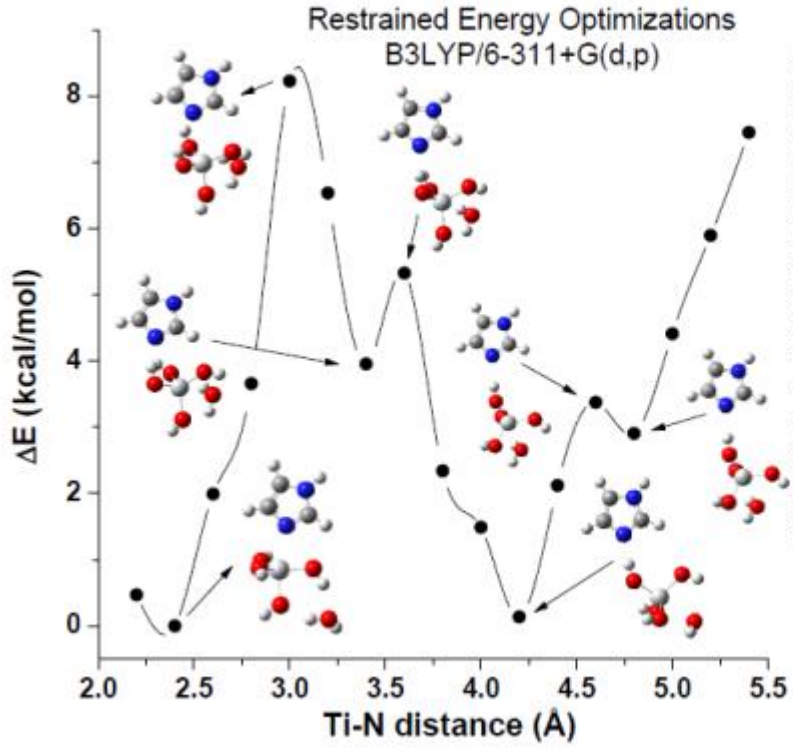
Fruitore: Susanna Monti
Istituto di afferenza del Fruitore: Istituto di Chimica dei Composti Organometallici (ICCOM) – UOS di Pisa
Titolo del programma: Development and Application of ReaxFF Description for Proteins and their Interactions with Metal Oxide Surfaces
Obiettivi Raggiunti: sviluppo dei parametri necessari per effettuare simulazioni dinamica molecolare classica, in cui sia possibile osservare la reattivita' delle specie interagenti (formazione e rottura di legami), nel caso di strutture peptidiche e proteiche in soluzione acquosa ed in contatto con substrati a base metallica. Studio di varie sequenze peptidiche in fase gassosa ed in soluzione. Studio dell'adsorbimento di glicina e diglicina su biossido di titanio.
Attività Svolta: il programma di ricerca svolto, concordato con il professor van Duin, è stato suddiviso in due diverse linee: **1)** messa a punto dei parametri per descrivere l'adsorbimento di amino acidi e peptidi su interfacce di biossido di titanio; **2)** sviluppo dei parametri per una corretta descrizione della dinamica di proteine ed enzimi in soluzione e delle reazioni coinvolgenti tali sistemi. Le linee di ricerca sono state sviluppate contemporaneamente in modo da sfruttare al massimo il periodo di permanenza presso la Pennsylvania State University.

Linea 1. Adsorbimento di Amino Acidi e Dipeptidi su TiO_2

Sono stati costruiti ed ottimizzati tramite calcoli quanto meccanici di diverso livello vari modelli molecolari a complessita' crescente che potessero rappresentare l'assorbimento di frammenti aminoacidici su biossido di titanio. Queste strutture sono state utilizzate come training set per la definizione dei parametri nel force field reattivo. Il programma puo' eseguire ottimizzazione dei parametri per riprodurre sia le energie che le geometrie delle molecole inserite nel training set. Per completare la parametrizzazione delle interazioni peptidi-titanio sono stati considerati i legami titanio-azoto, titanio-ossigeno e titanio-zolfo.
Alcuni esempi sono visibili nell'immagine riportata sotto ed alcune delle strutture inserite nel training set per definire la formazione e rottura dei legami considerati sono riportate nelle figure successive.

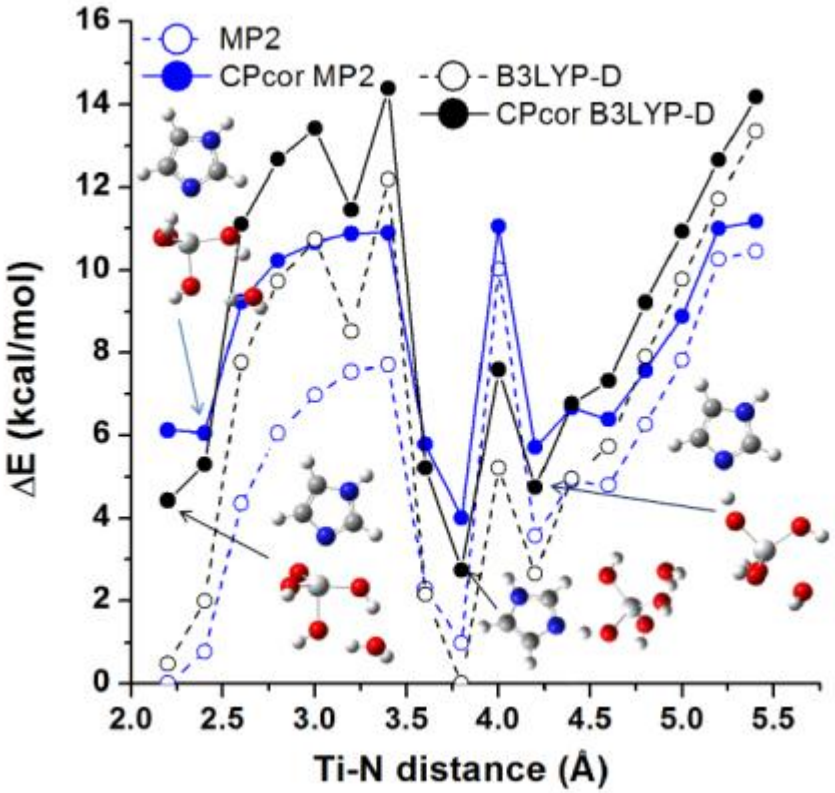


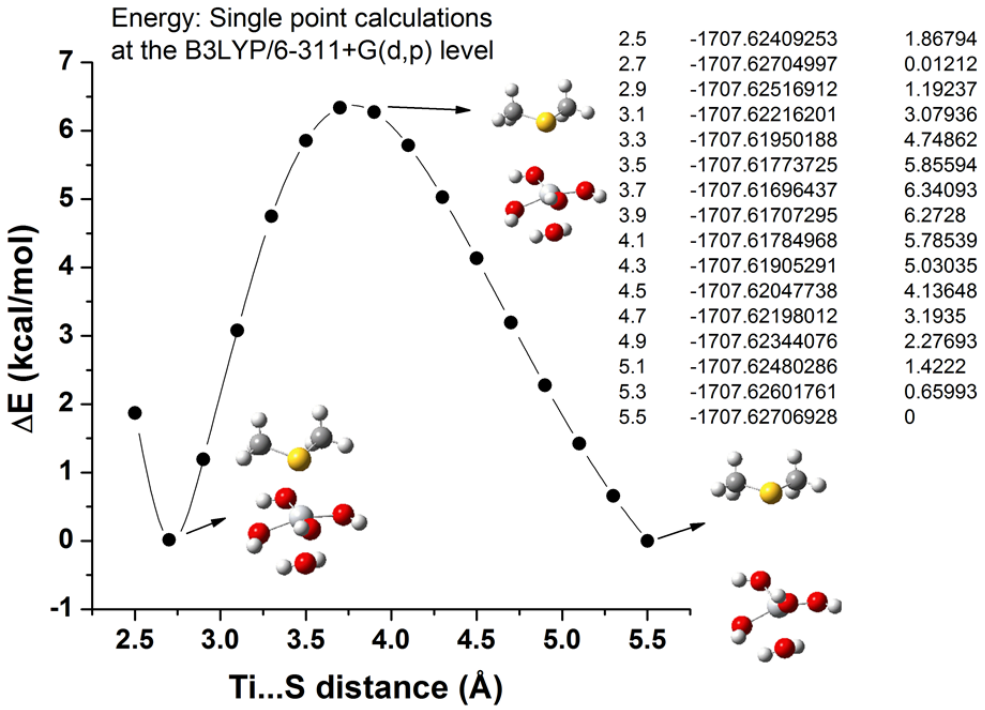
Selected Clusters Used to Parametrize the Interactions: Ti-N and Ti-O



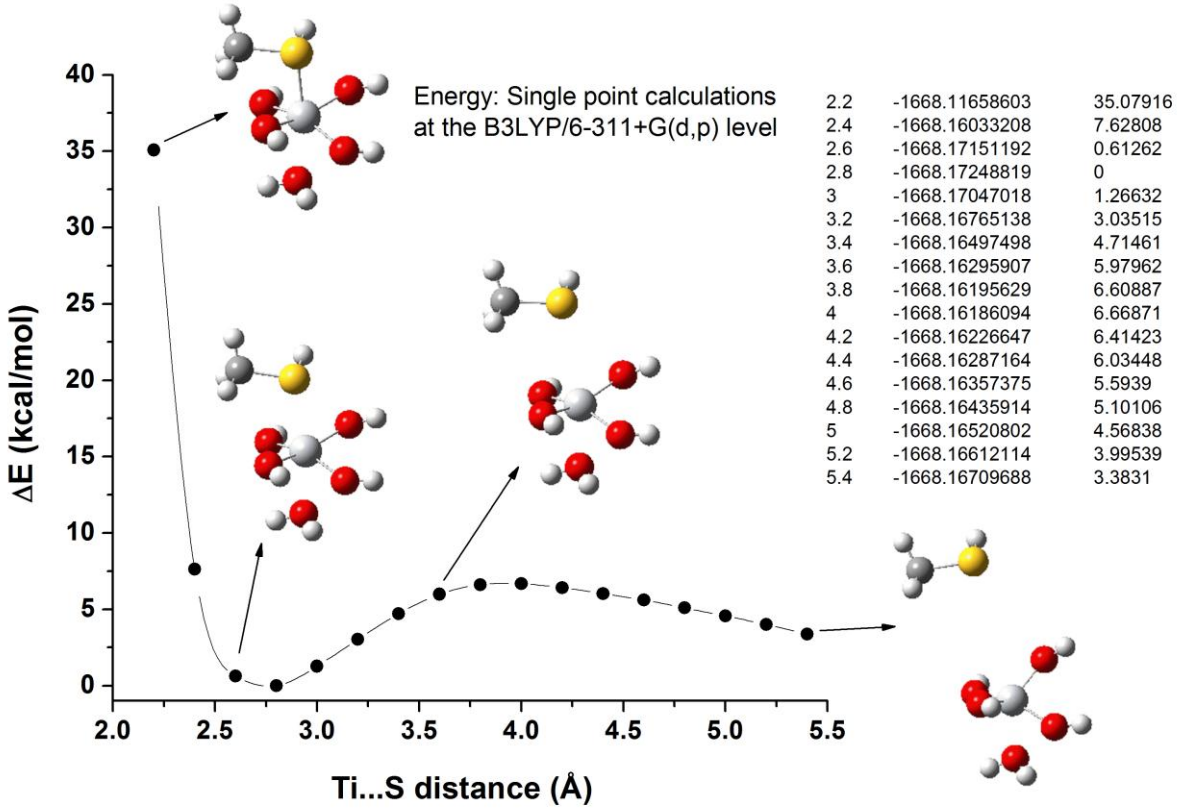
2.2	-1455.88278256	0.46758
2.4	-1455.88352769	0
2.6	-1455.88035416	1.99142
2.8	-1455.87769660	3.65907
3	-1455.87041577	8.22786
3.2	-1455.87311099	6.53658
3.4	-1455.87722341	3.956
3.6	-1455.87503860	5.32699
3.8	-1455.87979791	2.34047
4	-1455.88115486	1.48897
4.2	-1455.88331216	0.13525
4.4	-1455.88014845	2.12051
4.6	-1455.87814791	3.37587
4.8	-1455.87889887	2.90463
5	-1455.87649774	4.41136
5.2	-1455.87413077	5.89666
5.4	-1455.87165358	7.45112

The geometry was realxed keeping fixed the Ti-N distance





2.5	-1707.62409253	1.86794
2.7	-1707.62704997	0.01212
2.9	-1707.62516912	1.19237
3.1	-1707.62216201	3.07936
3.3	-1707.61950188	4.74862
3.5	-1707.61773725	5.85594
3.7	-1707.61696437	6.34093
3.9	-1707.61707295	6.2728
4.1	-1707.61784968	5.78539
4.3	-1707.61905291	5.03035
4.5	-1707.62047738	4.13648
4.7	-1707.62198012	3.1935
4.9	-1707.62344076	2.27693
5.1	-1707.62480286	1.4222
5.3	-1707.62601761	0.65993
5.5	-1707.62706928	0



2.2	-1668.11658603	35.07916
2.4	-1668.16033208	7.62808
2.6	-1668.17151192	0.61262
2.8	-1668.17248819	0
3	-1668.17047018	1.26632
3.2	-1668.16765138	3.03515
3.4	-1668.16497498	4.71461
3.6	-1668.16295907	5.97962
3.8	-1668.16195629	6.60887
4	-1668.16186094	6.66871
4.2	-1668.16226647	6.41423
4.4	-1668.16287164	6.03448
4.6	-1668.16357375	5.5939
4.8	-1668.16435914	5.10106
5	-1668.16520802	4.56838
5.2	-1668.16612114	3.99539
5.4	-1668.16709688	3.3831

Oltre a distanze di legame sono stati considerati angoli di legame, ed una simile procedura di parametrizzazione e' stata ripetuta anche per questi.

Dopo aver verificato l'affidabilita' dei parametri tramite simulazioni di sistemi non contenuti nel training set e confronto con calcoli quanto meccanici e' stato eseguito lo studio dell'adsorbimento ed interazione dell'amino acido glicina su una superficie di biossido di titanio (rutilo(110)) gia' studiata in precedenza con tecniche classiche non reattive. Riportiamo qui di seguito parte del lavoro in via di sottomissione.

The Drop Evaporation Approach

The strategy chosen to avoid system bias and define a configuration of the deposited glycine layers where the molecules were randomly distributed on top of the substrate, consisted in the creation of a spherical solute droplet (**Figure S1**), which was placed in the empty space between the surface and its periodic z-image and then evaporated, by means of high temperature dynamics in such a way that the molecules could be deposited over the surface. The starting model of the droplet had an initial radius of about 15 Å and was composed of two different shells: an inner shell (where the radius was around 10 Å) made of zwitterionic molecules (47 in all) and an external layer surrounding the central core, which was around 5 Å thick and contained 136 non-ionic species. All the molecules were randomly orientated but were accurately arranged one with respect to the other by using the xfit [1] procedure (available in the SYBYL [2]), which find the best fit between two molecules and can be employed to pack molecules in a cluster.

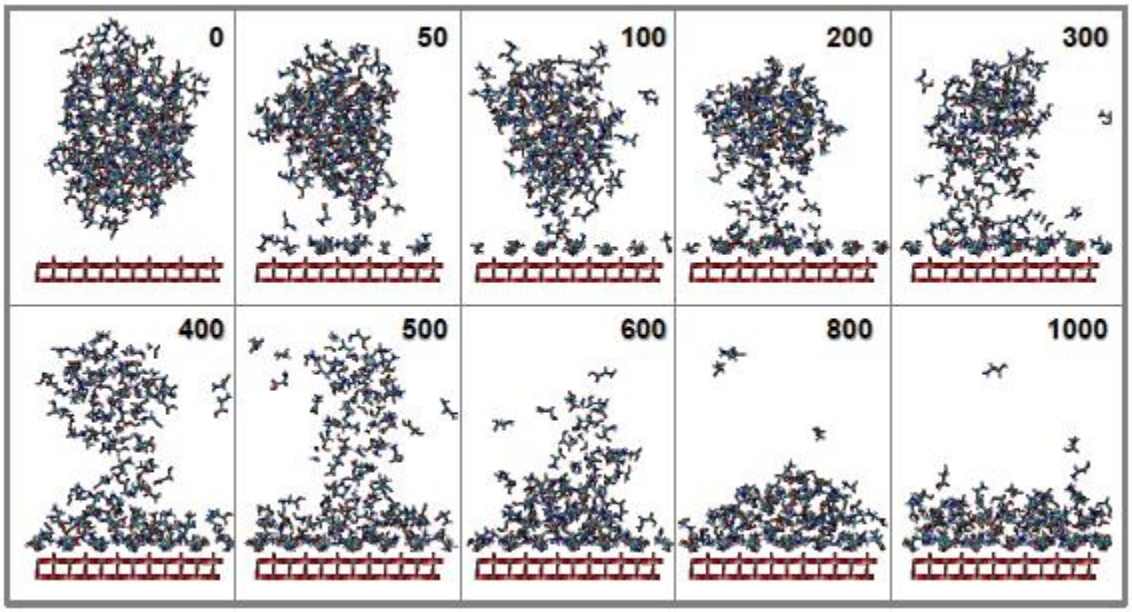


Figure S1: Droplet Evaporation Dynamics. Snapshots extracted from the high temperature dynamics (800 K) depicting the disruption of the droplet and settlement of the various glycine species on the interface. The starting structure (t=0 ps) has been obtained after energy minimization and equilibration at T=300 K of the initial spherical model.

Such a configuration was chosen because, even though glycine zwitterions are rarely observed in the gas phase, a few molecules might be present inside a droplet and be adsorbed among the molecules of the first or second glycine deposited layer, depending on the competition with the other species. This can be simulated using standard classical MD but, when a reactive representation of the whole system is employed a more sophisticated picture of the mechanisms which might take place along the path towards the surface during the adsorption process can be obtained. Surrounding the central portion of the droplet (self-assembled zwitterions - GLY_ZW) with non-ionic glycine (GLY_0), which are stable in the gas phase, would probably delay or prevent the intra-molecular exchange of protons, which, instead, would have been observed if a single amino acid had been placed inside an empty cell. Before exploring the system dynamics, a preliminary phase consisted in the relief of bad solute-solute interactions, which was performed by energy minimization. Then, NVT molecular dynamics simulations at T=300 K were carried out in order to equilibrate the whole model and randomize size and shape of the droplet, as well as mutual

arrangement of the constituting units. The coordinates of all the atoms in the slab were frozen during all the simulation runs. The initial almost spherical shape of the solute changed to an ellipsoidal elongated structure where the molecules maintained a compact arrangement being connected by a dense network of intermolecular hydrogen bonds. At T= 300 K the droplet was quite stable and it was not possible to observe migration of the various units towards the surface. Some of the peripheral members showed the tendency to leave the supra-molecular structure but the process was very slow and could not take place in a computationally reasonable amount of time. Thus, in order to speed up droplet disassembly and adsorption of glycine onto the surface, it was necessary to increase the simulation temperature to 800 K. The system was gradually heated to 800 K, during 500 ps, and then simulated at that temperature for about 5 ns which were sufficient to observe complete evaporation of the droplet and adsorption of all glycine molecules onto the substrate. Then, the last configuration sampled during high temperature dynamics was chosen as starting structure for the subsequent low temperature simulation (T=300 K). The system was firstly equilibrated for 2 ns and then the production run was carried out for 20 ns to explore preferential arrangements of the SAM on the surface and inter-monomer organization.

Classical Non-Reactive Simulations

Inspection of the evolution of the radius of gyration (**Figure S2**) during the simulations at T=800 K (**Figure S2** left-hand side) and 300 K (**Figure S2** right-hand side) revealed that the droplet disruption process at T=800 K was quite fast and took place in less than 1 ns (**Figure S1**). Both glycine species (GLY_0 and GLY_ZW) were randomly distributed and highly mobile during the high temperature dynamics but moved towards the surface and adopted a stable organization after cooling (at 300 K). The average radius of gyration of the outer shell changed from 20.2±0.7 to 18.9±0.05 Å, whereas <Rgyr> of GLY_zw remained almost constant (at about 14.7 Å) but with a shrunk range of explored values. It could be speculated that the diffusion of GLY_ZW was more contained because of its location relative to the droplet and, as a consequence, a delayed adsorption. The arrangement on the interface was maintained for the whole simulation time and no migration of the adsorbed units to farther distances from the surface was observed. This is confirmed by the trend of the distance of the center of mass of each species from the plane containing the five coordinated titanium sites. As shown in **Figure S3**, glycine-surface separation reached an equilibrium value after about 800 ps and then oscillated around it until the end of the sampling process. It should be noticed that during this initial phase the molecules of the droplet core (GLY_zw) approached the surface and laid on top of the first adsorbed layer (which is mainly composed of GLY_0), but tended to insert among the units closest to the substrate and engage in direct interactions with the inorganic interface.

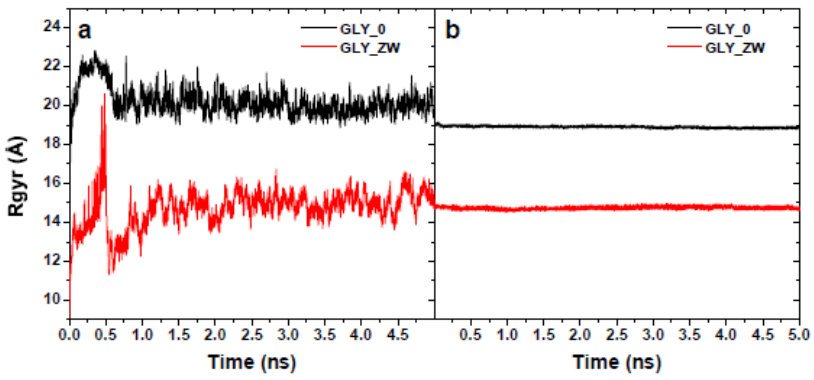


Figure S2. Radius of gyration as a function of the simulation time during droplet disruption and after adsorption onto the titanium layer.

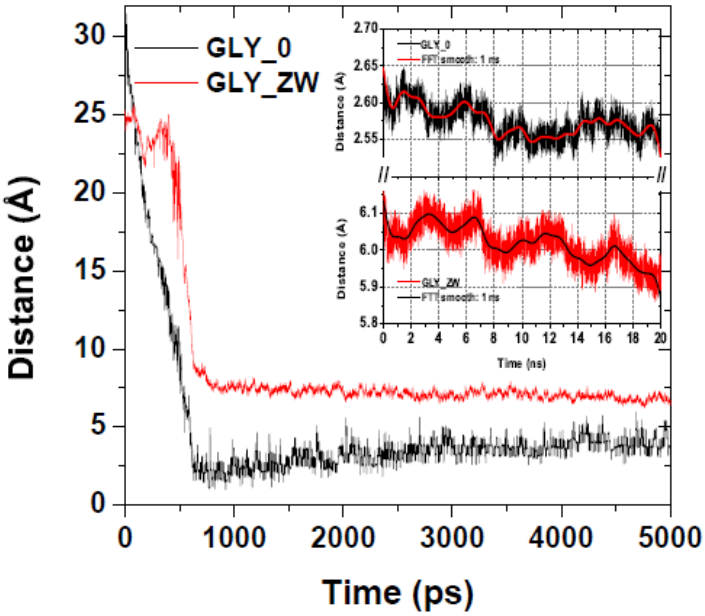


Figure S3. Glycine-surface separation as a function of the simulation time during the high temperature dynamics and after adsorption on the layer of all the molecules.

The values of Rgyr reflect the different location and number of the two species and together with the distribution of translational kinetic energies (**Figure S4**) show that the outer shell had the tendency to move more slowly in comparison with the faster restructuring and relocation of the post adsorbed species (GLY_ZW). This could be due to the quick adsorption of GLY_0 and its stable settlement on the substrate. However, visual inspection of the trajectories at low temperature revealed that only local and limited structural reorganization of the molecules belonging to the second layer took place due to a dense and dynamic network of intermolecular hydrogen bonds.

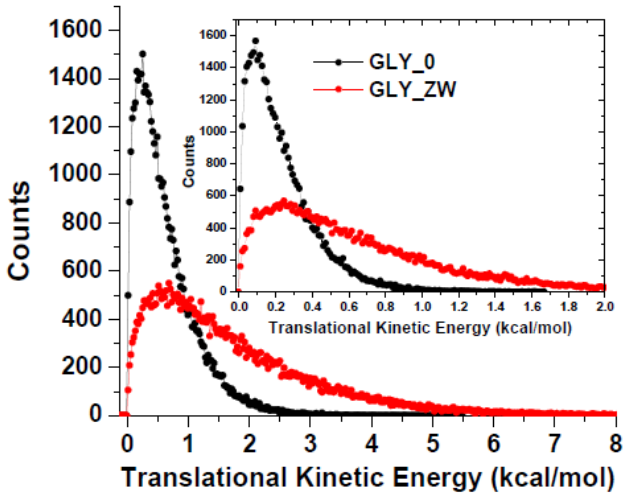


Figure S4. Distribution of the Translational Kinetic Energy of the center of mass of the molecules during the adsorption process and after settlement on the surface (in the inset).

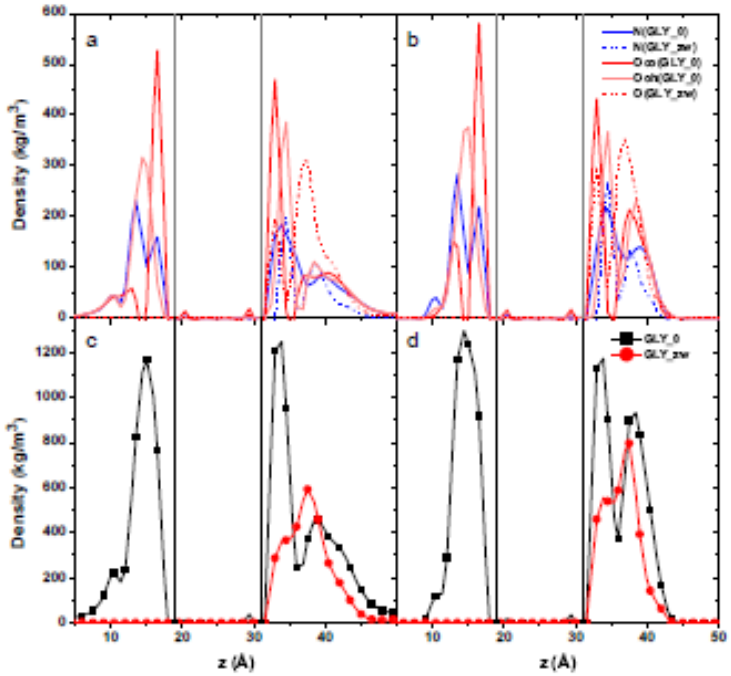


Figure S5. Glycine density as a function of the z coordinate during the high temperature dynamics (a,c) and after adsorption (MD simulations at 300 K). The distribution of the glycine center of mass is shown in c and d, whereas atom profiles are displayed in a and b.

In **Figure S5** glycine densities as a function of the z coordinate for both simulations at 800 and 300 K are showed together with atom-based local densities. The position of the two interfaces are represented by vertical grey lines and are located at z=18.97 and 31.04 Å, respectively.

[1] Blanco, M. J. Comp. Chem. 1991, 12, 237–247.

[2] SYBYL Molecular Modelling Software, Version 7.2; TRIPOS Associates, St. Louis, MO, 2005.

Dipeptide Adsorption in Water Solution

In order to analyze the stability and reactive dynamics of configurations where water molecules were present, twelve glycine dipeptides, in their zwitterionic form, were attached to different regions (randomly chosen) of the interface, far from each other, through a bi-dentate coordination, which was expected to be stronger than the mono-dentate adsorption mode. The complex was solvated adding about 1400 water molecules on top of the slab and the system was equilibrated at T=300 K using non-reactive dynamics. Water was described through the modified TIP3P force field and its density was adjusted to the bulk value in a series of NVT simulations and energy minimizations, as already done in previous works.32,68 The final thickness was about 15 Å. Classical non-reactive simulations were perform just to define an equilibrated configuration which could be used as starting structure in the reactive dynamics. The results of this simulation will not be exposed in detail because they are very similar to data found in previous studies, but representative findings will be reported and compared with the reactive description. An initial observation which deserves to be made before discussing the outcome of this investigation is that, in line with earlier works all the molecules remained stuck to the titanium layer until the end of the simulation which was about 5 ns long. The portion of the chain far from the surface was more flexible than the carboxyl+peptide bond segment (closer to the surface) and could oscillate more freely in solution.

Reactive Force Field Simulations

Differently form the classical conventional approaches, ReaxFF updates bond orders and the related descriptors every time step, which, for this reason, should be substantially small, that is in the range 0.1-0.5 fs. All the simulations were performed at T=300 K in the NVT ensemble using the Berendsen thermostat78 with a relaxation constant of 0.1 ps and the equations of motion were solved with the Verlet leapfrog algorithm using a time step of 0.25 fs. To remove bad contacts and to adapt the systems to the new description, energy minimizations were carried out in the NVE ensemble for 25 ps. Then, the whole systems were equilibrated at T=300 for 35 ps and then simulated at this temperature for 100 ps without any restraints on their different components. The goal of these calculations was to investigate glycine adsorption properties, inter-glycine interactions and proton exchange with the surface, both in the gas phase and in solution, and compare the results with the findings of experimental and theoretical studies. Reactive MD simulations were performed both when the droplet was far from the surface and when the peptide layers were deposited on the substrate (starting from the last configuration sampled during standard classical dynamics). As already noticed, in the case of conventional simulations, the adsorption process at T=300 K was very slow and it was impossible to describe the entire peptide surface adhesion mechanism. However, it was very interesting to analyze the first few picoseconds of the droplet evaporation and the reaction of the peripheral units between themselves and with the titanium oxide interface. As far as simulations in solution are concerned, the last configuration of the diglycine solvated system obtained through standard dynamics was used as starting structure for the reactive investigation and subjected to ReaxFF MD simulations in order to explore the influence of water on pre-adsorbed peptides and water-surface interactions. Considering that reaction times are very short the system was simulated for 60 ps.

Linea 2: Peptidi e Proteine in Fase Gassosa ed in Soluzione

Anche in questo caso riportiamo qui di seguito parte del lavoro in via di scrittura.

Force Field Parametrization and Validation

The force field parametrization consisted in fitting both geometry and energy differences from quantum mechanics calculations on a large number of molecular species which were relevant to the simulation of amino acid, peptide and protein structures. Starting from the parameters developed in an earlier work [1], a three-stage approach, where molecular models of increasing complexity were progressively added to the training set, was followed to improve the systems representation.

The first stage consisted in including the first ten minimum energy conformations of each amino acid (that is two hundreds conformers in all), designed and kindly provided by Zijing Lin and co-workers, into the previous training set. It is not our intention to describe here the huge computational work performed by the authors, which is well documented in a number of articles but a brief account could be useful to outline the main features and, above all, the refinement technique which is at the base of the complex parametrization procedure.

Briefly, the added amino acid structures were obtained through systematic quantum mechanical studies carried out at the DFT-B3LYP/6-311++G** level of theory. The authors conducted exhaustive conformational searches building all combinations of internal single-bond rotamers and performed energy optimization of the generated conformations without any constraint. Subsequent checks, comprising harmonic frequencies and single point calculations at a higher level of theory (MP2/6-311G(2df,p)//B3LYP/6-311++G**), allowed for the identification of the lowest energy conformers. The

addition of all these structures was aimed at improving the prediction of both backbone and side chain conformations. The force field was trained against the QM-derived set of energies, valence bond, torsional angles and charge parameters, which were all optimized to reproduce the QM data. ReaxFF final geometries matched quite satisfactorily all the data in the training set.

The second stage comprised examination and improvement of the force field performance to reproduce amino acids terminated with N-Acetyl and N-methylamide functional groups (dipeptides), tripeptides and tetrapeptides, not included in the training set. Fit-for-purpose QM optimized sets of molecules were found in the literature as supplementary material or in the form of well organized databases where also other useful information could be found.

A brief account illustrating the collected samples deserves to be made in order show the consistency and variety of the whole dataset. As far as dipeptides is concerned, Kaminsky and Jensen [3] sampled and optimized at the DFT and MP2 levels with an augmented double-zeta basis set the conformational degrees of freedom of glycine, alanine, serine and cysteine. They obtained the best estimate of the relative conformational energies by means of an extrapolation procedure of the MP2 energy results (extrapolation to the basis set limit and CCSD(T) correction - see literature for details) and used these data as reference to check the performance of eight different force fields. The comparison revealed that only half of the conformations were identified by force fields based on fixed partial charges, whereas those including multipoles and polarizability, could reproduce most of the structures. These structures were not included in ReaxFF training set but were used as test case to assess the validity of the reactive force field parametrization. Hobza and co-workers performed theoretical studies on the performance of different levels of theory in comparison with the CCSD(T)/CBS benchmarks for Phe-Gly-Phe (FGF), Trp-Gly (WG), Trp-Gly-Gly (WGG), Phe-Gly-Gly (FGG), Gly-Gly-Phe (GGF) and Gly-Phe-Ala (GFA) peptides [4].

On the basis of experimental observations they restricted the benchmark study to fifteen different conformers for each species, which were selected following a procedure defined in earlier investigations. Geometry optimizations were performed at the RI-MP2/cc-pVTZ level of theory and also by means of the AMBER ff99 force field with the HF/6-31G* RESP charges. In their conclusion the authors did not recommend the use of this force field for investigations in the gas phase but suggested that a reliable strategy could be to use three different levels of theory, namely tight-binding DFT-D, RI-MP2 or M06-2X, and MP2/CBS for the study of the potential energy surface of the peptide, the reoptimization of selected structures and single-point energy calculations, respectively.

Some of these structures, were included in ReaxFF training set, whereas the others were used as test case to assess the validity of the developed parameters.

A series of polypeptide conformations, made of alanine residues, were employed by Martin Head-Gordon and co-workers [5] to appraise the overall performance of the RI-TRIM MP2 method against other techniques in the prediction of the relative energies of the structures. The main sample consisted of twenty seven different conformations of the alanine tetrapeptide, but octapeptide, and hexadecapeptide structures were also present among the deposited data. All the geometries were optimized at the HF/6-31G** level and single-point energy calculations at higher levels with different basis sets were performed as well. The authors concluded that the accurate prediction of the relative energies of the series of tetrapeptides implies calculations with the cc-pVTZ and cc-pVQZ basis sets and then extrapolation to the ccpV(TQ)Z limit. Also in this case the conformations did not become members of the training set, but were used to check the force field performance.

[1] O. Rahaman, A. C. T. van Duin, W. A. Goddard, III, D. J. Doren J. Phys. Chem. B, 2011, 115, 249

[2] Huang, Z.; Lin, Z. J. Phys. Chem. A 2005, 109, 2656; Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Harañczyk, M.; Gutowski, M. J. Phys. Chem. A 2006, 110, 12282; Huang, Z.; Yu, W.; Lin, Z. J. Mol. Struct.: THEOCHEM 2006, 801, 7; Huang, Z.; Lin, Z.; Song, C. J. Phys. Chem. A 2007, 111, 4340; Rai, A. K.; Song, C.; Lin, Z. Spectrochimica Acta Part A 2009, 73, 865; Leng, Y.; Zhang, M.; Song, C.; Chen, M.; Lin, Z. J. Mol. Struct.: THEOCHEM 2008, 858, 52; Meng, L.; Lin, Z. Comput. Theore. Chem. 2011, 976, 42; Xu, X.; Lin, Z. J. Mol. Struct.: THEOCHEM 2010, 962, 23; Zhang, M.; Lin, Z. J. Mol. Struct.: THEOCHEM 2005, 760, 159; Chen, M.; Huang, Z.; Lin, Z. J. Mol. Struct.: THEOCHEM 2005, 719, 153; Huang, Z.; Yu, W.; Lin, Z. J. Mol. Struct.: THEOCHEM 2006, 758, 195; Chen, M.; Lin, Z. J. Chem. Phys. 2007, 127, 154314; Yu, W.; Lin, Z.; Huang, Z. ChemPhysChem 2006, 7, 828; Zhang, M.; Huang, Z.; Lin, Z. J. Chem. Phys. 2005, 122, 134313; Yu, W.; Liang, L.; Lin, Z.; Ling, S.; Haranczyk, M.; Gutowski, M. J. Comput. Chem. 2009, 30, 589.
[3] J. Kaminský, F. Jensen J. Chem. Theory Comput., 2007, 3 , 1774
[4] Valdés, H.; Pluhackova, K.; Hobza, P. J. Chem. Theory Comput. 2009, 5, 2248; Valdés, H.; Pluhackova, K.; Pitonak, M.; Rezac, J.; Hobza, P. Phys. Chem. Chem. Phys. 2008,
10, 2747; Reha, D.; Valdés, H.; Vondrášek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. Chem. Eur. J. 2005, 11, 6803.
[5] Head-Gordon, M. et al. J. Chem. Theory Comput. 2005, 1, 862

Il force field e' stato ottimizzato ed ora e' in corso di raffinamento. Allo scopo sono state scelte piccole proteine, spesso utilizzate come case study per i metodi classici non reattivi , da simulare in soluzione.