COMPARATIVE DENSITY FUNCTIONAL THEORY, DENSITY FUNCTIONAL TIGHT BINDING AND FORCE FIELD STUDIES OF AMINOACIDS, PEPTIDES AND PEPTIDE-TITANIA INTERFACES: BRINGING AB INITIO ACCURACY TO BIOMOLECULAR SIMULATIONS

LI WENXUAN

(*B.Eng, Huazhong University of Science and Technology*)

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF ENGINEERING
DEPARTMENT OF MECHANICAL ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2016

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Li Wenxuan

20 June 2016

Acknowledgements

The results contained in this thesis were presented in the following papers and conference presentations:

Wenxuan Li, Konstantinos Kotsis, Sergei Manzhos: Comparative force-field molecular dynamics and density functional tight binding study of cell-penetrating peptides - Lycosin-I, *Golden Jubilee Chemistry Conference, Singapore*

H. Tan, W. Luo, L. Wei, B. Chen, W. Li, L. Xiao, S. Manzhos, Z. Liu, S. Liang: Quantifying the Distribution of the Stoichiometric Composition of Anti-cancer Peptide Lycosin-I on Lipid Membrane with Single Molecule Spectroscopy. *The Journal of Physical Chemistry B,* **2016**, *120*(12), pp 3081-3088

W. Li, K. Kotsis, S. Manzhos: Comparative Density Functional Theory and Density Functional Tight Binding Study of Arginine and Arginine-Rich Cell Penetrating Peptide TAT Adsorption on Anatase TiO_2. Submitted *Physical Chemistry Chemical Physics*, **2016**

I would like to thank my supervisor, Dr. Sergei Manzhos, for the patient guidance, encouragement and advice in the last two years. I am so lucky to have such a kind supervisor who not only taught me how to be a good researcher, but also how to be a good man. And I sincerely appreciate all his help. I would also like to thank Dr. Konstantinos Kotsis and our group members for their continued help and support on me.

Contents

Summary

Force field based Molecular Dynamics (MD) techniques are widely applied in the field of biomolecular simulations, while quantum mechanical (QM) techniques are only routinely applied to small model systems such as individual amino acids, due to their high CPU cost. Even though force-field MD simulations are most widely used, QM calculations (practically, at the Density Functional Theory, DFT, level) are still desired, because they provide a better accuracy, especially when reactions are modeled. They are also necessary for many applications, as they can provide electron density, charge distribution, and electronic and optical properties which are in principle impossible to obtain with force fields. Especially for mixed organic-inorganic systems, the advantages of QM methods over force fields are significant. In this thesis, we study the performance of the ab initio based density functional tight binding (DFTB) method, which is feasible even for large biomolecule-containing systems, but whose performance strongly depends on proper parameterization and therefore needs to be benchmarked. We compare the performance of DFTB with a common force field for the simulation of large biomolecules, specifically, the cell penetrating peptide Lycosine-I which has shown aggregation-dependent anticancer activity, the understanding of which required simulations. We also present a comparative DFT-DFTB study of bioinorganic interfaces: arginine, arginine dipeptide, and the arginine-rich TAT cell-penetrating peptide on TiO_2. While there is good agreement in the structures and relative energies of arginine (Arg) and of peptide conformers between DFT and DFTB, adsorption geometries

and energies are noticeably different between the two methods for Arg adsorbed on TiO_2. We relate this difference to the difference in electronic structures resulting from the two methods (DFT and DFTB) and specifically to the band alignment between the molecule and the oxide. We show that the band alignment of TAT and of TiO_2 modeled with DFTB is qualitatively correct but that with DFT using the PBE functional is not. This is specific to the modeling of large molecules where the highest occupied molecular orbital (HOMO) is close to the conduction band of the oxide. We therefore report a case where the approximate DFT-based method DFTB (with which the correct band structure can be effectively obtained) performs better than the DFT itself with a functional approximation feasible for the modeling of large bio-inorganic interfaces, i.e. GGA (as opposed to hybrid functionals which are impractical at such a scale). Our results highlight the utility of the DFTB method for the modeling of bioinorganic interfaces not only from the CPU cost perspective but also from the accuracy point of view.

List of Tables

Table 1. Mean Absolute Difference of bond lengths (Å) and angles (°) for each amino acid (MAD_{bond} and MAD_{angle}) between the Gaussian setup and the DFTB with 3ob-2-1 parameter set, as well as the energy difference between the conformer and the original structure ($\Delta E_{conformer}$, in eV) in both computational schemes.

	MAD_{bond}	MAD_{angle}	$\Delta E_{conformer}$	
			DFT	DFTB
Alanine	0.006	0.57	/	/
Arginine	0.006	1.42	0.03	0.02
Asparagine	0.005	1.86	0.02	0.19
Aspartic Acid	0.006	1.12	0.47	0.21
Cysteine	0.005	1.87	0.08	0.01
Glutamic Acid	0.004	1.33	0.02	0.04
Glutamine	0.006	1.20	0.03	0.04
Glycine	0.005	1.45	/	/
Histidine	0.008	0.56	0.11	0.03
Isoleucine	0.007	0.76	0.01	0.00
Leucine	0.009	1.68	0.50	0.14
Lysine	0.006	0.55	0.01	0.03
Methionine	0.007	0.59	0.01	0.16
Phenylalanine	0.005	0.65	0.07	0.04
Proline	0.011	0.92	/	/
Serine	0.009	3.02	0.06	0.08
Threonine	0.009	1.36	0.07	0.06
Tryptophan	0.006	0.67	0.35	0.17
Tyrosine	0.005	0.69	0.13	0.04
Valine	0.005	2.03	0.24	0.01
Average over all	**0.007**	**1.22**	**0.13**	**0.07**

Table 2. Mean Absolute Difference of bond lengths (Å) and angles (°) for Arginine (MAD_{bond} and MAD_{angle}) and the length of the peptide bond ($\Delta B_{pep\text{-}bond}$) between the Gaussian setup and the other three computational schemes, as well as the energy difference between the conformer shown in Figure 1 (a-c) and the cyclic structure shown in Fig. 1(d) ($\Delta E_{conformer}$, in eV).

	MAD_{bond}	MAD_{angle}	$\Delta B_{pep\text{-}bond}$	$\Delta E_{conformer}$
DFT/G09/6-31g+(d,p)	/	/	/	0.07
DFT/G09/6-31g++(2d,2p)	0.002	0.08	-0.004	0.07
DFT/SIESTA[a]	0.011	0.44	0.008	0.34
DFT/SIESTA[b]	0.011	0.43	0.011	0.35
DFTB/Mio-1-1	0.013	1.30	0.006	0.20
DFTB/Matsci-0-3	0.033	1.77	-0.028	0.15

[a] Broader basis functions (PAO.EnergyShuft = 0.001 Ry)

[b] Narrower basis functions (PAO.EnergyShuft = 0.002 Ry)

Table 3. Binding energies (eV) of four putative dimer configurations in vacuum.

Dimer	ΔE_{MD}	ΔE_{DFTB}
a	-2.1	-2.69
b	-0.67	-1.05
c	-1.36	-1.6
d	-1.98	-2.08

Table 4. Adsorption energies E_{ads} (in eV) of Arginine in different configurations on anatase (101) surface of TiO_2 for adsorption via the carboxylic group. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For bidentate configurations, the two bond lengths are O_{mol}-Ti; for monodentate, they are O_{mol}-Ti and H-O_{surf}. The data are for DFTB calculations with two parameterizations (matsci-0-3 and tiorg-0-1*) and for DFT calculations with different choices of DZP basis parameters (using PAO.EnergyShift of 0.001 and 0.002 Ry)

System	E_{ads}, eV	O_{mol}-Ti, Å	O_{mol}-Ti/H-O_{surf}, Å
DFTB/Matsci-0-3			
BB	-1.78	2.23	2.23
M1	-0.97	2.25	1.63
M2	-1.08	2.26	1.62
DFTB/Tiorg-0-1*			
BB	-0.97	2.10	2.11
M1	-0.99	1.92	1.00
M2	-1.07	1.98	0.99
SIESTA (0.001Ry)			
BB	-1.02	2.04	2.06
M1	-1.14	2.14	1.51
M2	-1.21	2.12	1.53
SIESTA (0.002Ry)			
BB	-1.00	2.05	2.07
M1	-1.09	2.14	1.50
M2	-1.17	2.13	1.55

Table 5. Adsorption energies E_{ads} (in eV) of Arginie in different configurations on anatase (101) surface of TiO_2 via the amine groups. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For H-O configurations, the two bond lengths are O_{surf}-H; for NTi configurations, they are N_{mol}-Ti. (The bond length cannot be defined in the same way for the HO configuration obtained with matsci-0-3 due to a very different resulting geometry.)

System	E_{ads}, eV	O_{surf}-H/N_{mol}-Ti, Å	O_{surf}-H/N_{mol}-Ti, Å
DFTB/Tiorg-0-1*			
H-O	0.04	1.97	2.08
N-Ti	-0.22	3.49	3.78
DFTB/Matsci-0-3			
H-O	-2.16	/	/
N-Ti	-1.27	2.27	5.11
SIESTA/0.001Ry			
H-O	-0.26	2.14	2.11
N-Ti	-1.55	2.12	3.89
SIESTA/0.002Ry			
H-O	-0.21	2.14	2.09
N-Ti	-1.65	2.08	4.08

Table 6. Adsorption energies E_{ads} (in eV) of Arginine dipeptide in different configurations on anatase (101) surface of TiO_2. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For bidentate configurations, the two bond lengths are O_{mol}-Ti; for monodentate, they are O_{mol}-Ti and H-$O_{surf.}$

System	E_{ads}, eV	O_{mol}-Ti, Å	O_{mol}-Ti/H-O_{surf},
DFTB/Tiorg-0-1*			
BB	-1.01	2.10	2.11
M1	-0.92	1.99	1.00
M2	-1.13	1.99	0.99
DFT/PBE			
BB	-0.88	2.07	2.05
M1	-1.13	2.13	1.54
M2	-1.07	2.18	1.47

Table 7. Adsorption energies E_{ads} (in eV) of TAT in different configurations on anatase (101) surface of TiO_2. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For bidentate configurations, the two bond lengths are O_{mol}-Ti; for monodentate, they are O_{mol}-Ti and H-O_{surf}

System	E_{ads}, eV	O_{mol}-Ti, Å	O_{mol}-Ti/H-O_{surf}, Å
DFTB/Tiorg-0-1*			
BB	-1.28	2.11	2.11
M1	-0.99	1.98	1.00
M2	-1.37	1.99	0.99

List of Figures

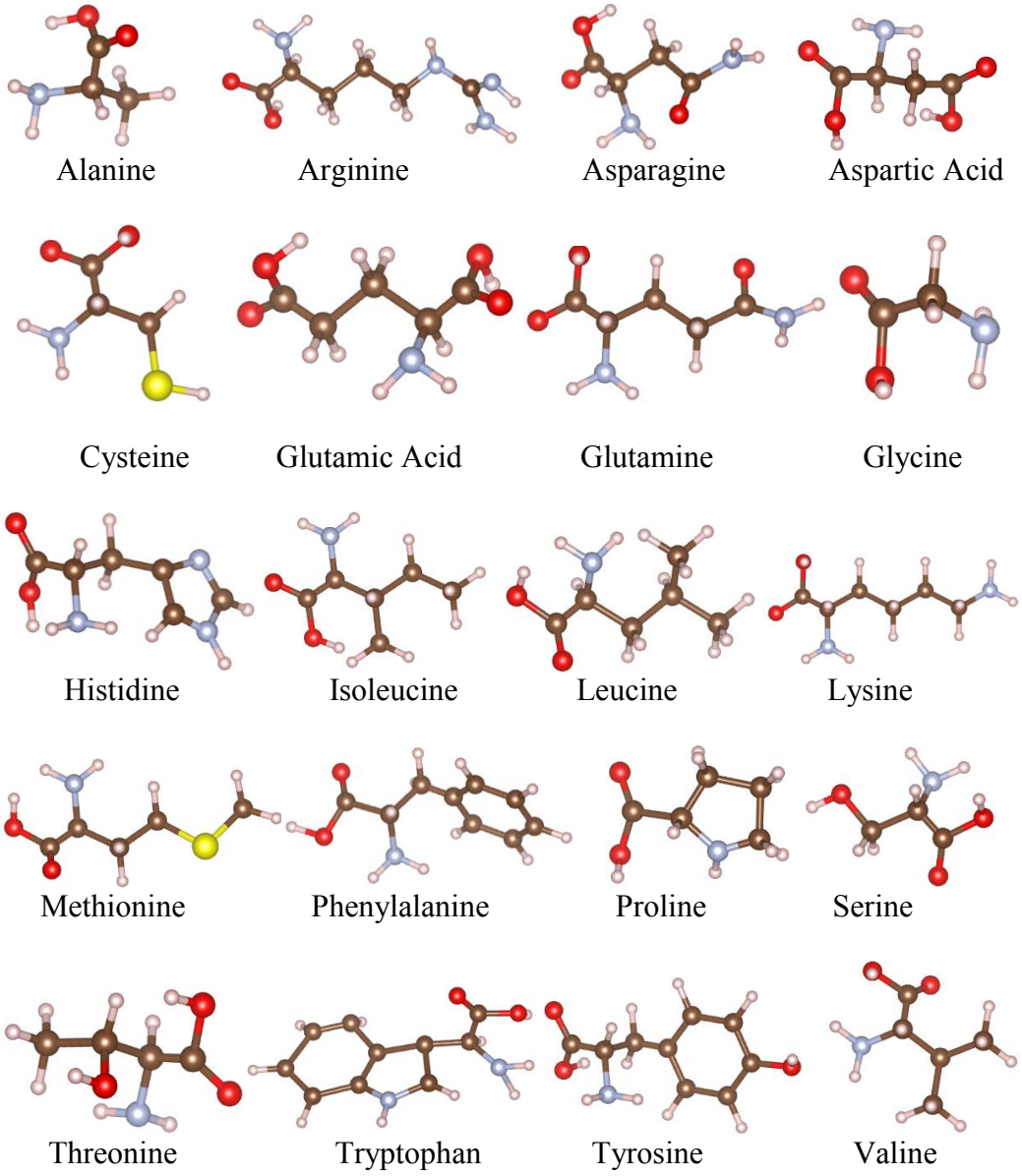


Figure 1. Optimized structures of 20 amino acids obtained with Gaussian09. The atom color code here: C, brown; O, red; H, light grey; N, gray; S, yellow. Visualization here by VESTA.

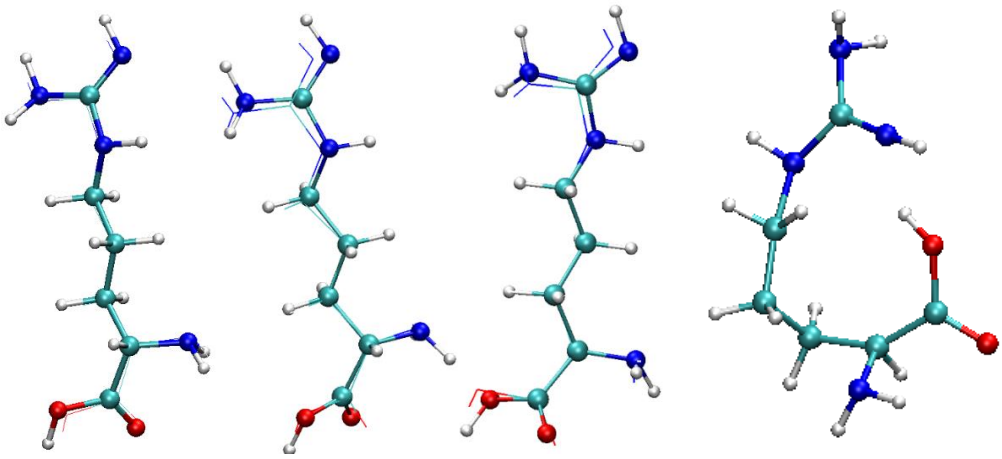


Figure 2. Wireframe optimized structures of Arginine with (a) DFT/SIESTA, (b) DFTB/mio-1-1, and (c) DFTB/matsci-0-3, overlaid with ball and stick optimized structures from DFT/Gaussian. Panel (d) shows the lowest-energy structure from Ref. 119. The atom color code here and elswehere with VMD: C, green; O, red; H, light grey; N, blue.

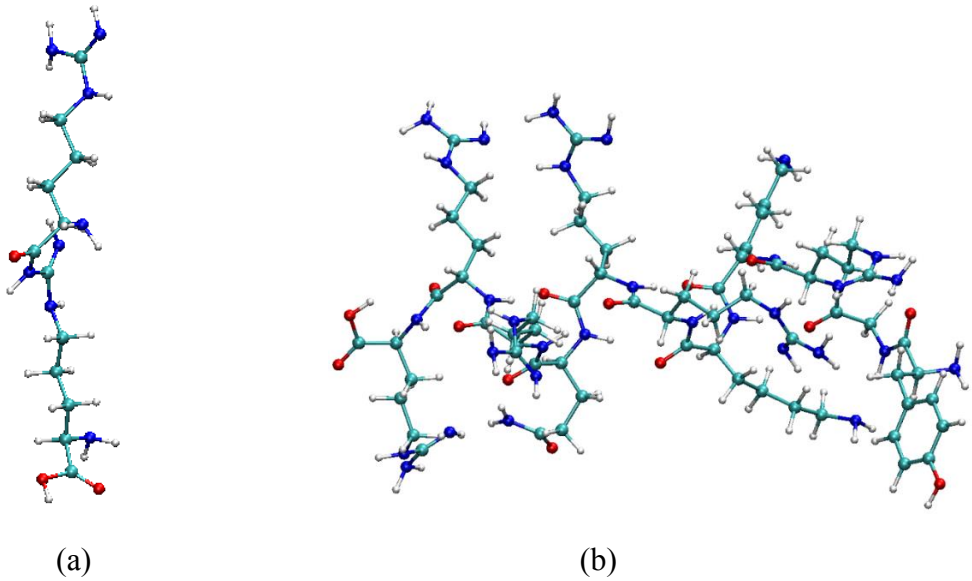


Figure 3. Structures of Arginine dipeptide (a) and the TAT peptide (b). The structures computed with Gaussian 09 are shown and are visually similar for all computational setups used here. Visualization here by VMD.

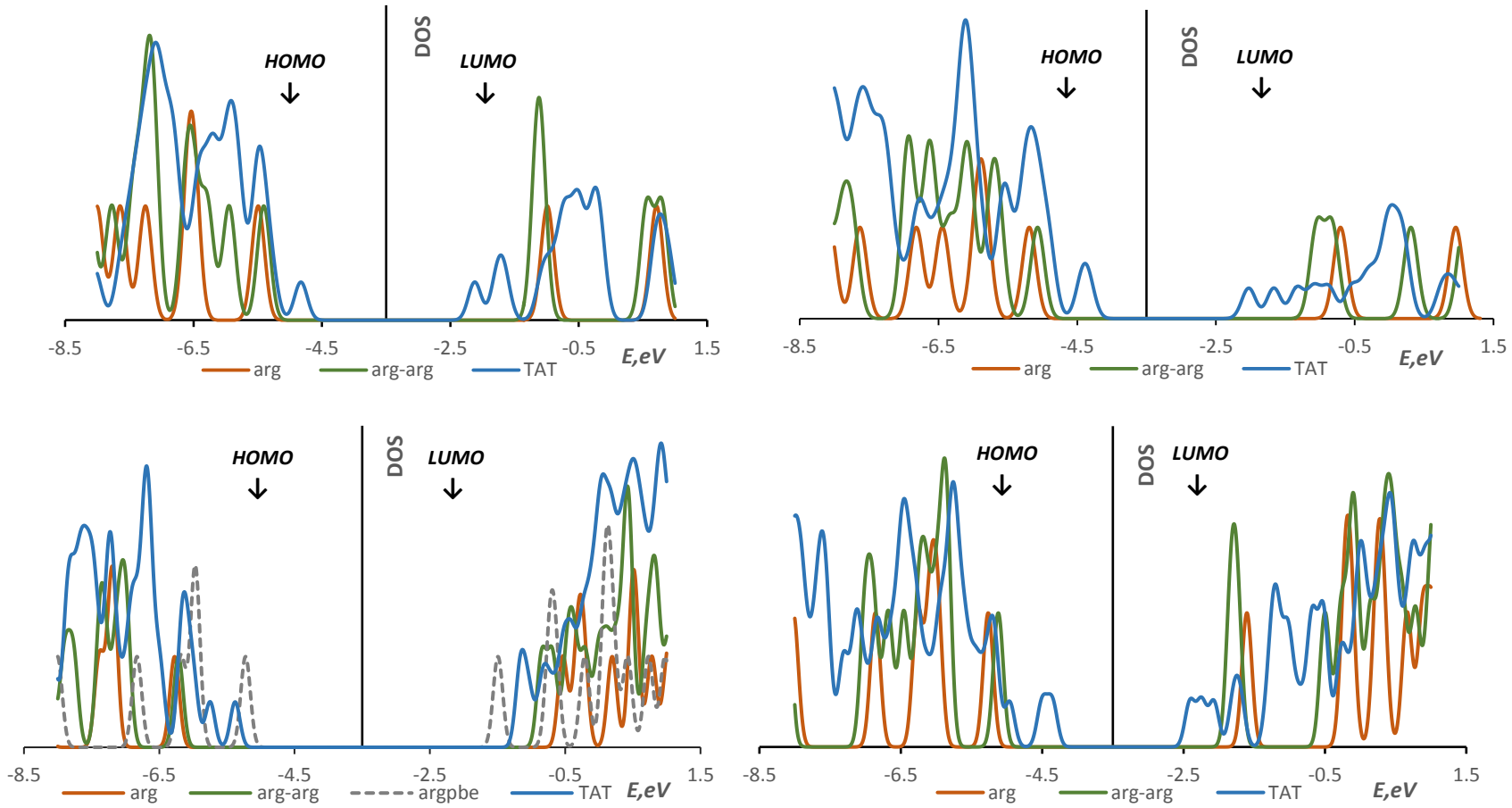
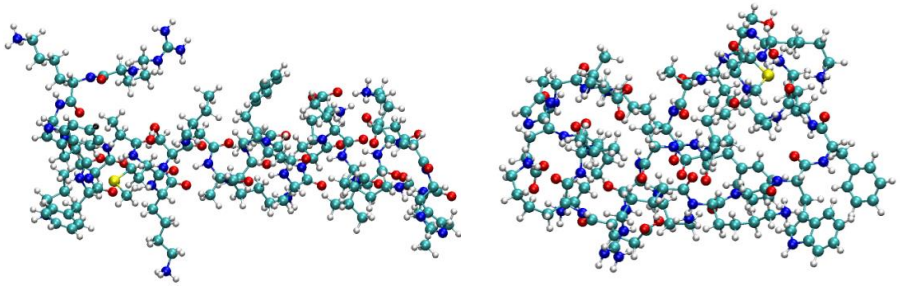
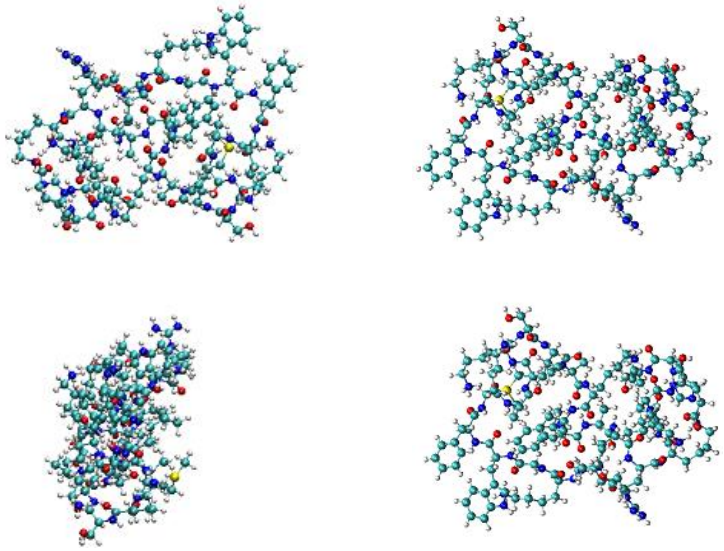


Figure 4. Densities of states of Arg, Arg dipeptide (arg-arg) and TAT computed with DFTB (top left: matsci-0-3 and top right: mio-1-1 parameter sets) and DFT (bottom left: G09/B3LYP, bottom right: SIESTA/PBE). A Gaussian broadening of 0.1 eV is applied. Approximate positions of HOMO and LUMO energies are indicated with arrows (i.e. HOMO is the highest-energy peak before the gap and LUMO is the lowest-energy peak after the gap; the y axis has been positioned inside the gap).

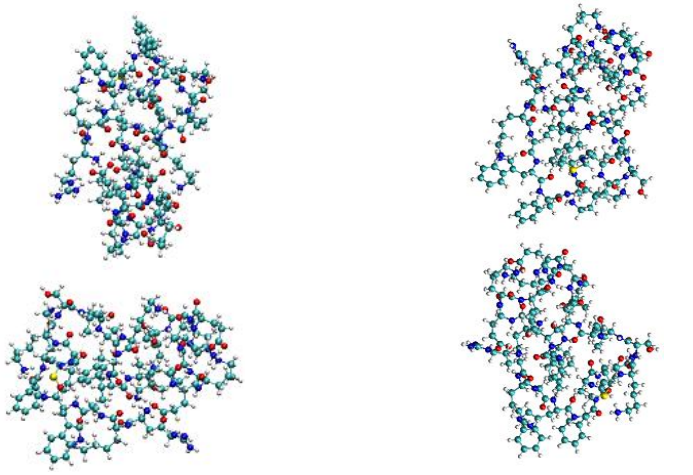


(a) (b)

Figure 5. Initial structure (a) and folded structure (b) of Lycosin-I. Visualization here by VMD.

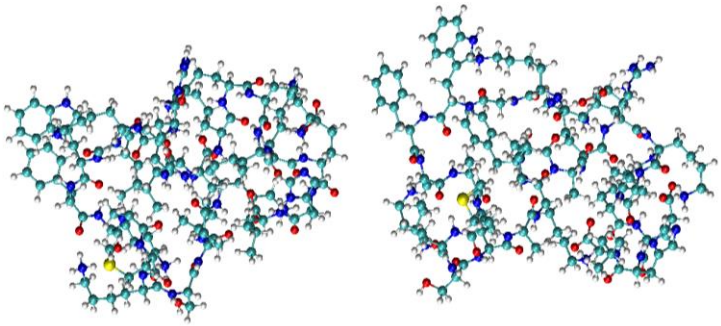


(a) (b)

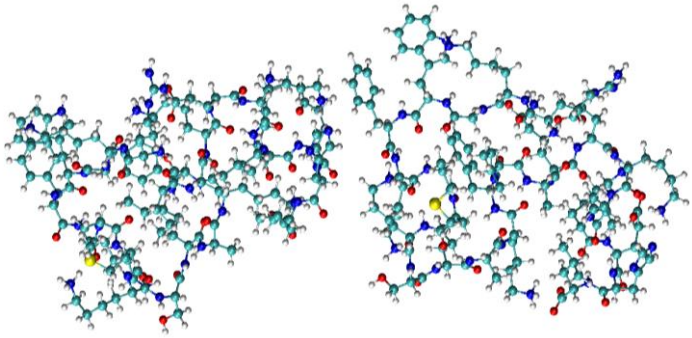


(c) (d)

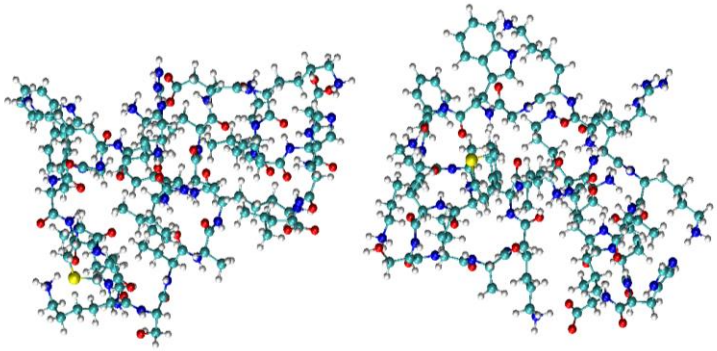
Figure 6. Structures of four putative dimer configurations. Visualization by VMD.



(a)

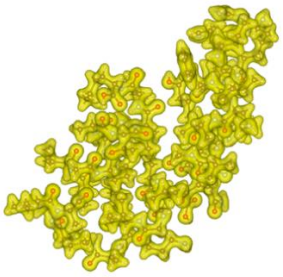


(b)

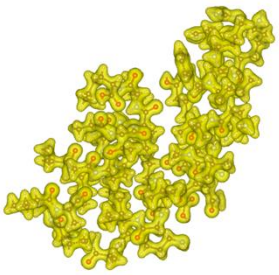


(c)

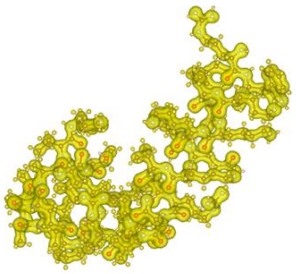
Figure 7. Snapshot of a point along the MD trajectory of a peptide dimer simulation: initial position (a), bonded (b), and dissociated (c) state. Water molecules are omitted for clarity. Visualization by VMD.



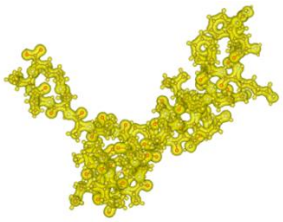
CecropinA-1



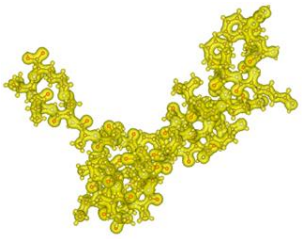
CecropinA-2



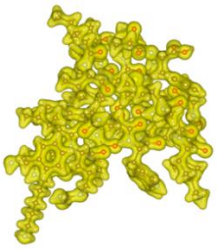
CecropinA-3



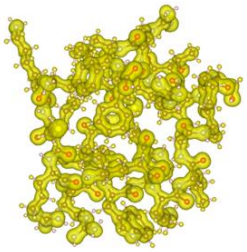
CecropinA-4



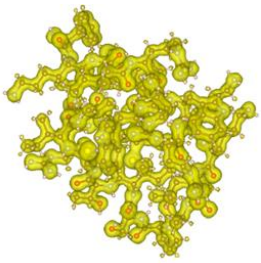
CecropinA-5



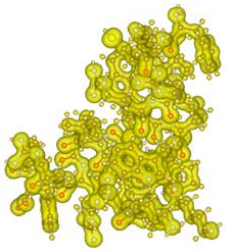
LfcinB-1



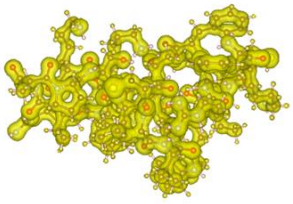
LfcinB-2



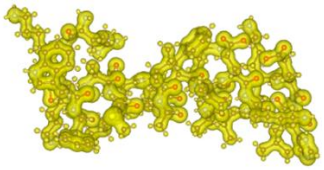
LfcinB-3



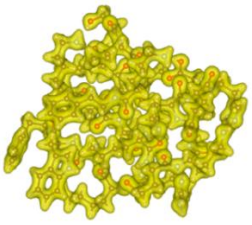
LfcinB-4



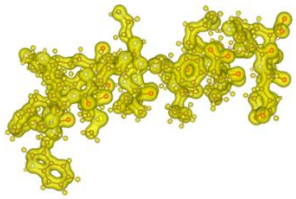
LfcinB-5



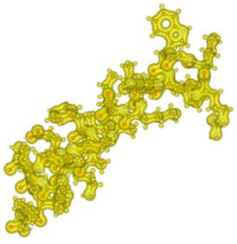
Lyscoin-I-1



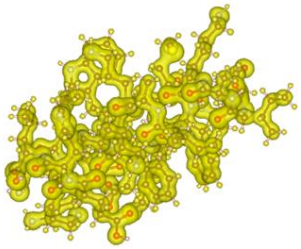
Lycosin-I-2



Lycosin-I-3



Lycosin-I-4



Lycosin-I-5

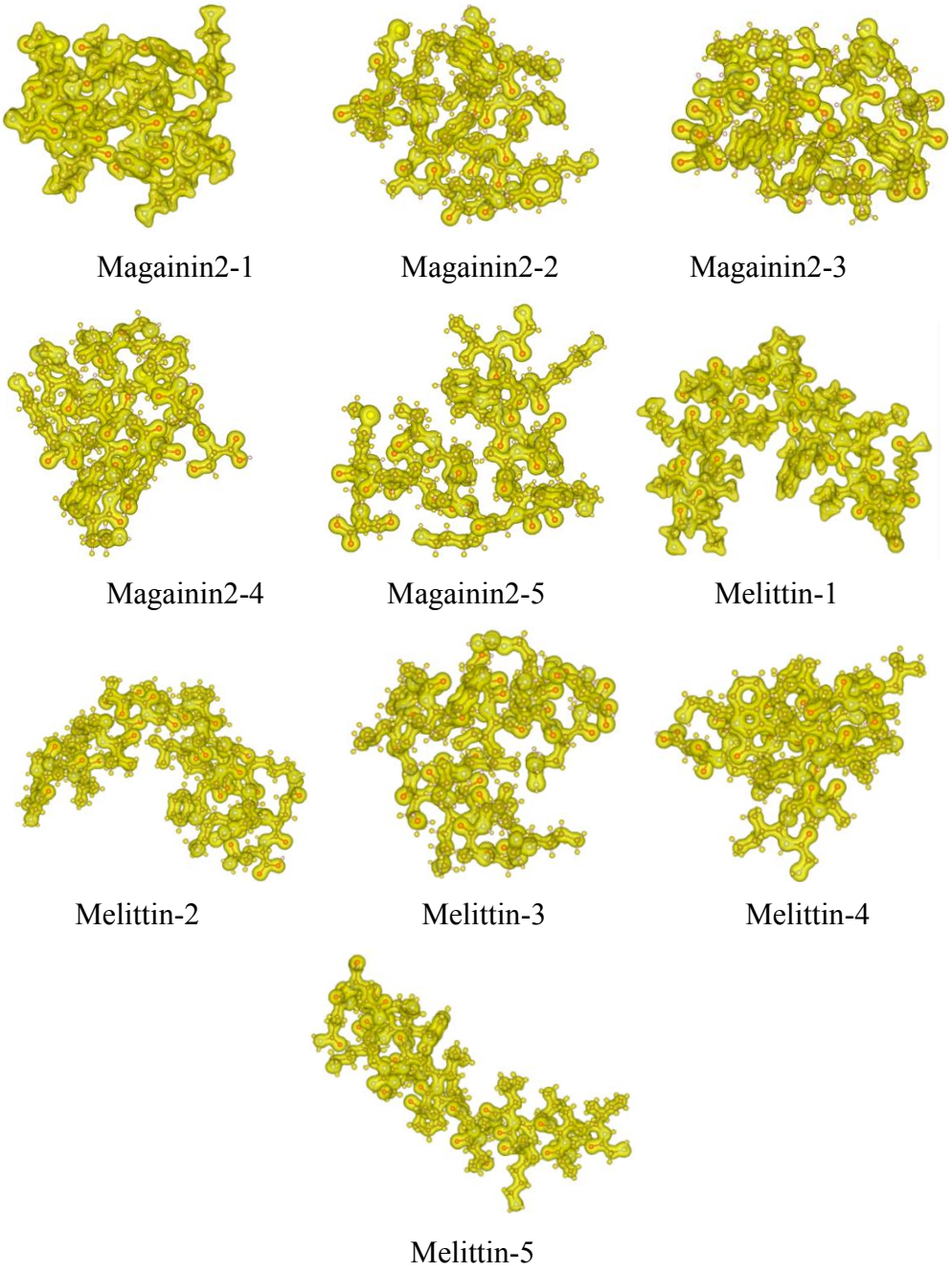


Figure 8. Total electron density distribution of cell-penetrating peptides. Visualization here by SIESTA.

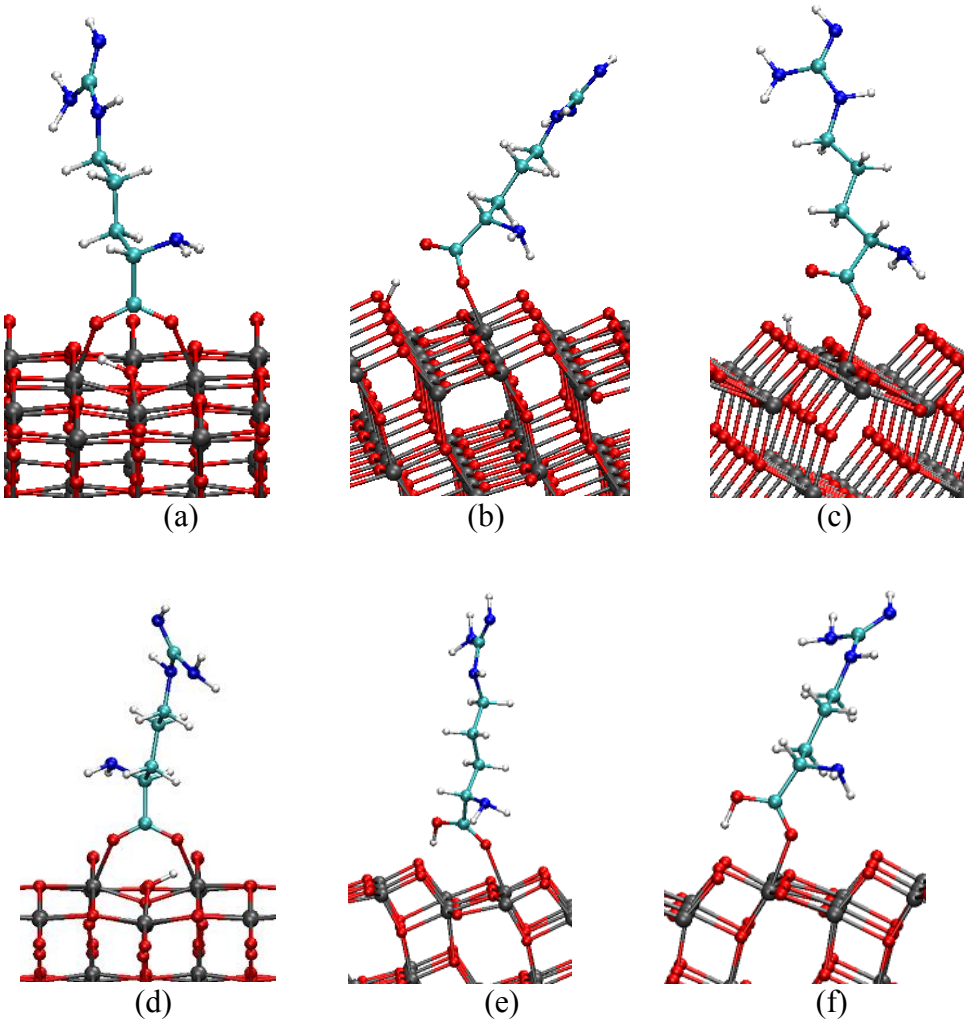


Figure 9. Optimized structures of BB (a), M1 (b) and M2 (c) configurations of Arg on TiO_2 from DFTB tiorg-0-1* and BB (d), M1 (e) and M2 (f) from SIESTA. The atom color code here and elswehere: C, green; O, red; H, light grey; N, blue.; Ti, dark grey. Visualization here by VMD.

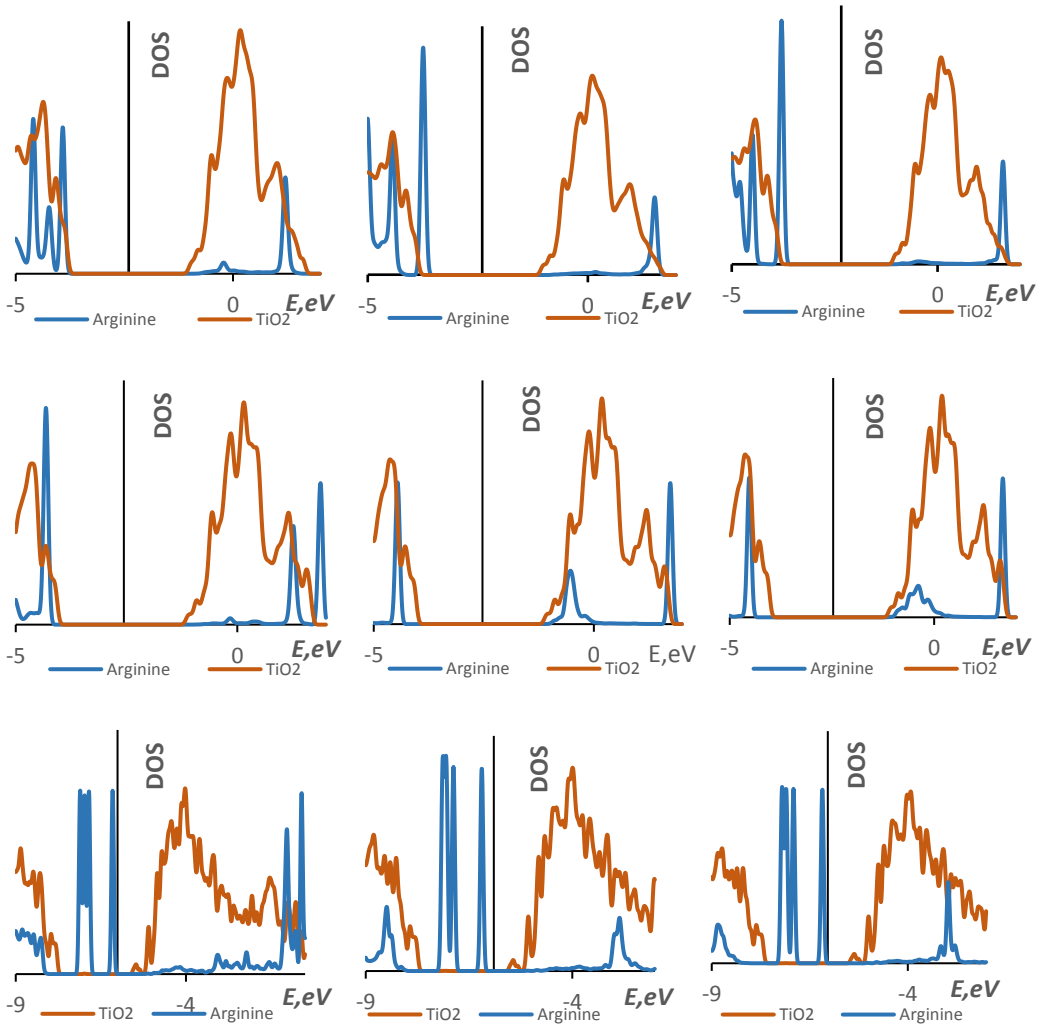


Figure 10. Molecule- and substrate- projected density of states of (left to rigth) BB, M1 and M2 obtained in DFTB with the tiorg-0-1* parameters (top row), matsci-0-3 (middle row), and with DFT (bottom row). The *y* axis crosses the energy axis at the Fermi energy.

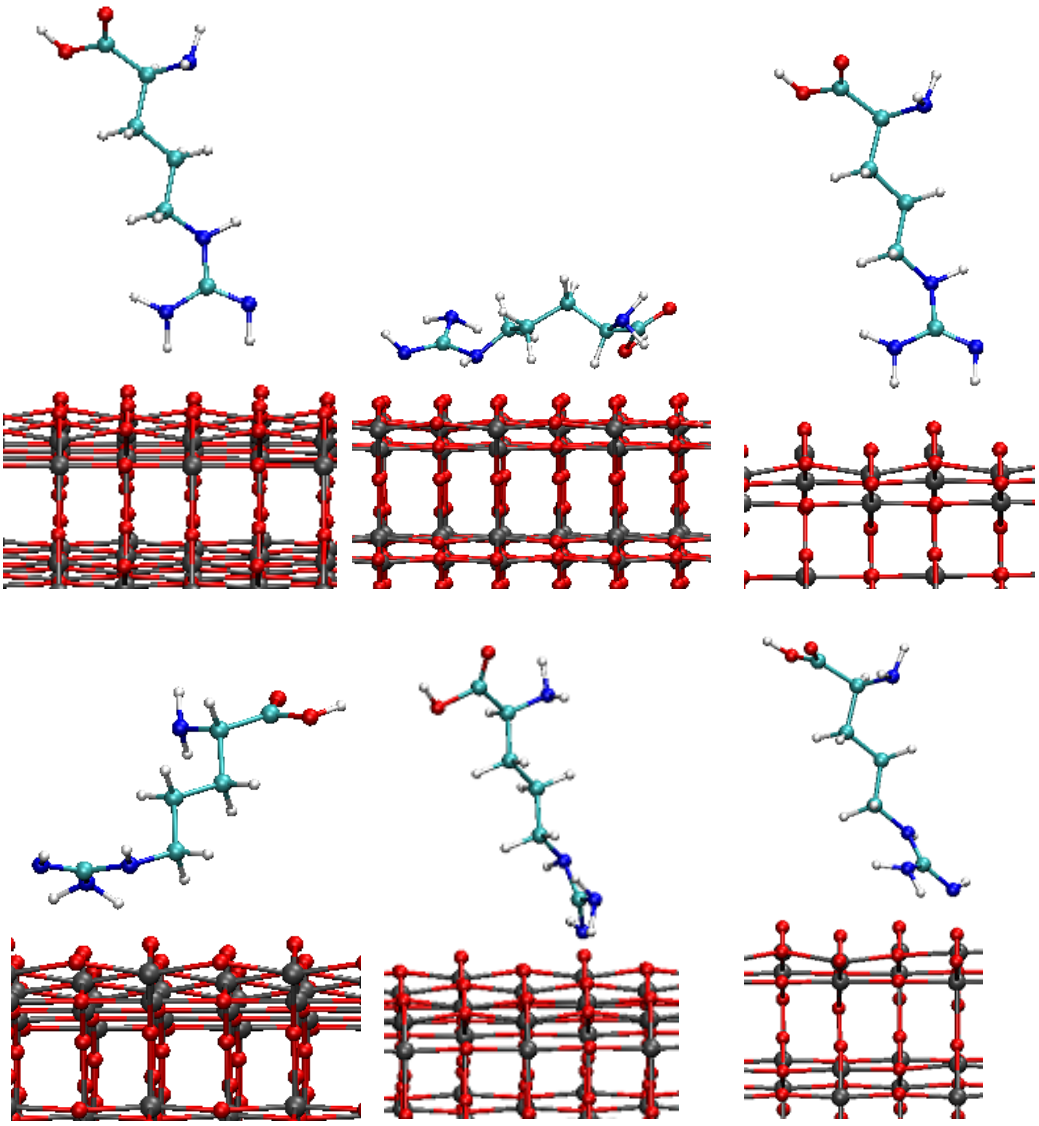


Figure 11. Optimized structure of Arg adsorbed on TiO_2 through the amine groups: configurations defined in the text as HO (top row) and NTi (bottom row) obtaibned with tiorg-0-1* (left), matsci-0-3 (middle) and with DFT (right). Visualization here by VMD.

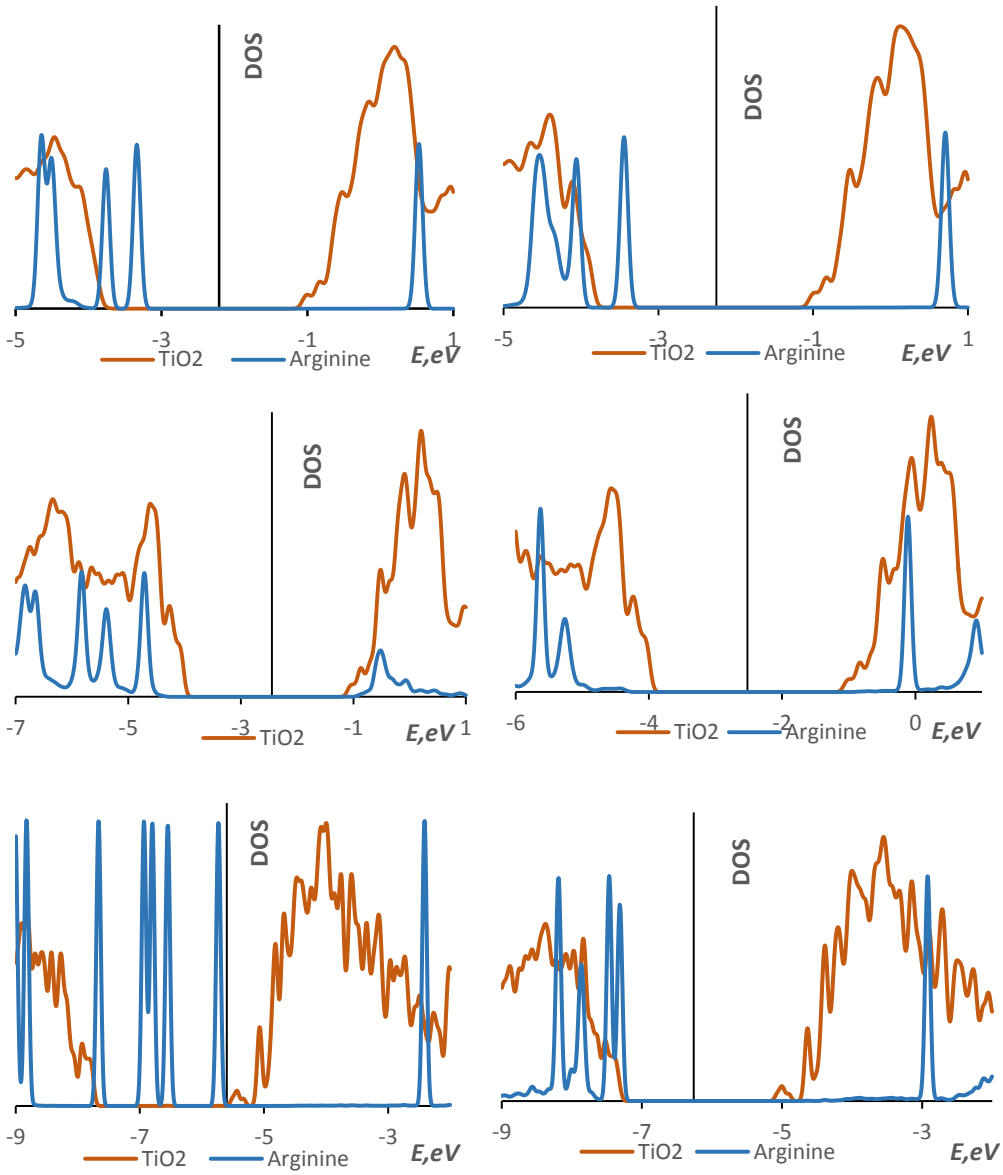


Figure 12. Molecule- and substrate- projected density of states of HO (left) and NTi (right) configurations of Arg on TiO_2 obtained in DFTB with the tiorg-0-1* parameters (top row) and matsci-0-3 DFT (middle row) and with DFT (bottom row). The *y* axis crosses the energy axis at the Fermi energy. Visualization here by VMD.

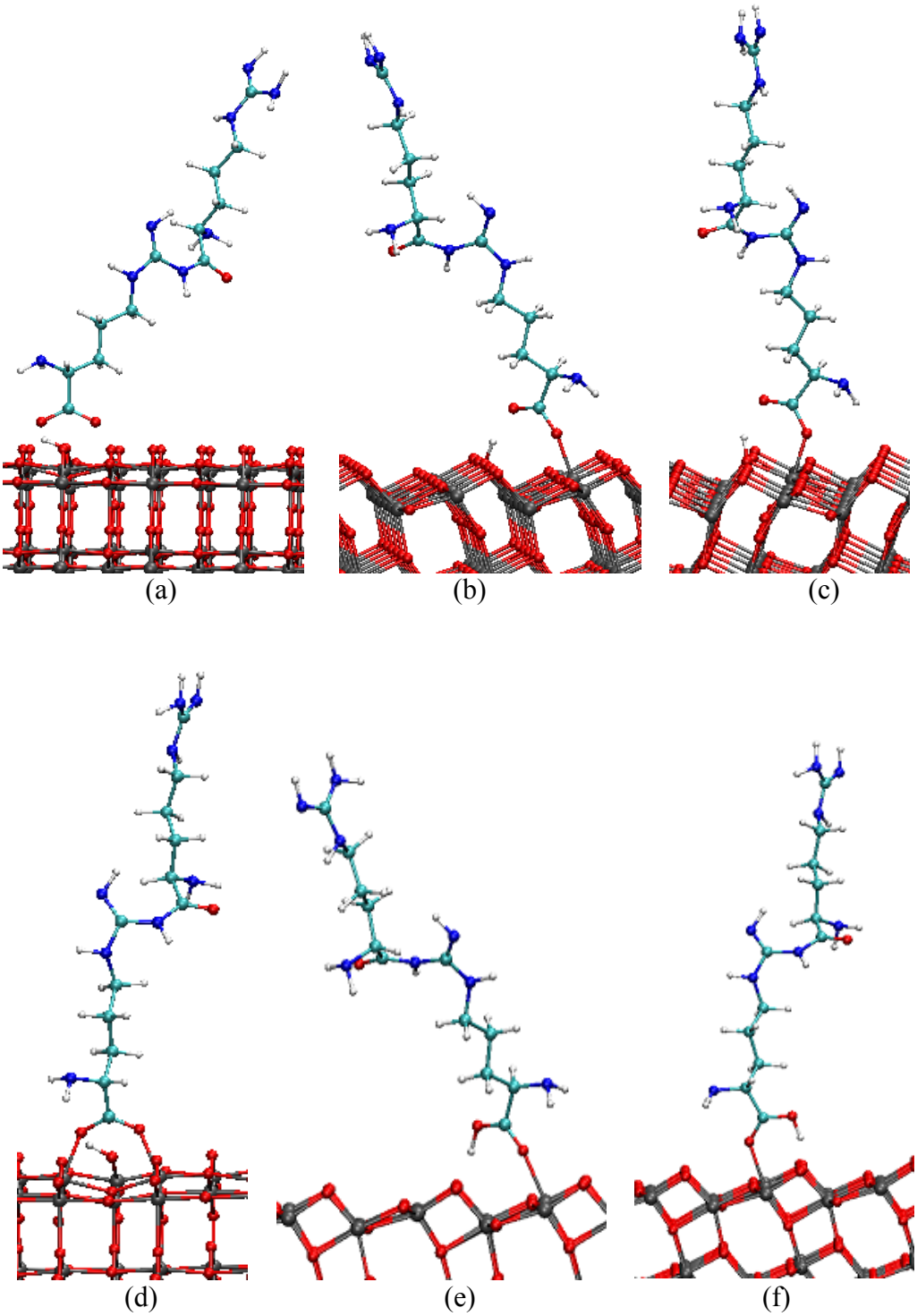


Figure 13. Optimized structures of BB (a), M1 (b) and M2 (c) for arginine dipeptide adsorbed on anatase TiO_2 obtained with DFTB tiorg-0-1* and BB (d), M1 (e) and M2 (f) from SIESTA. Visualization here by VMD.

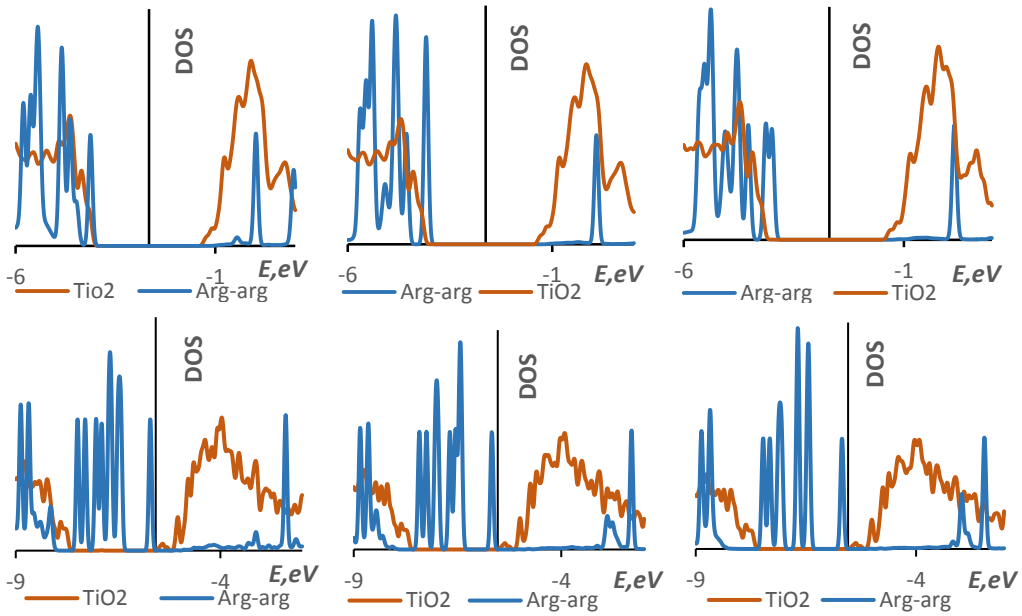
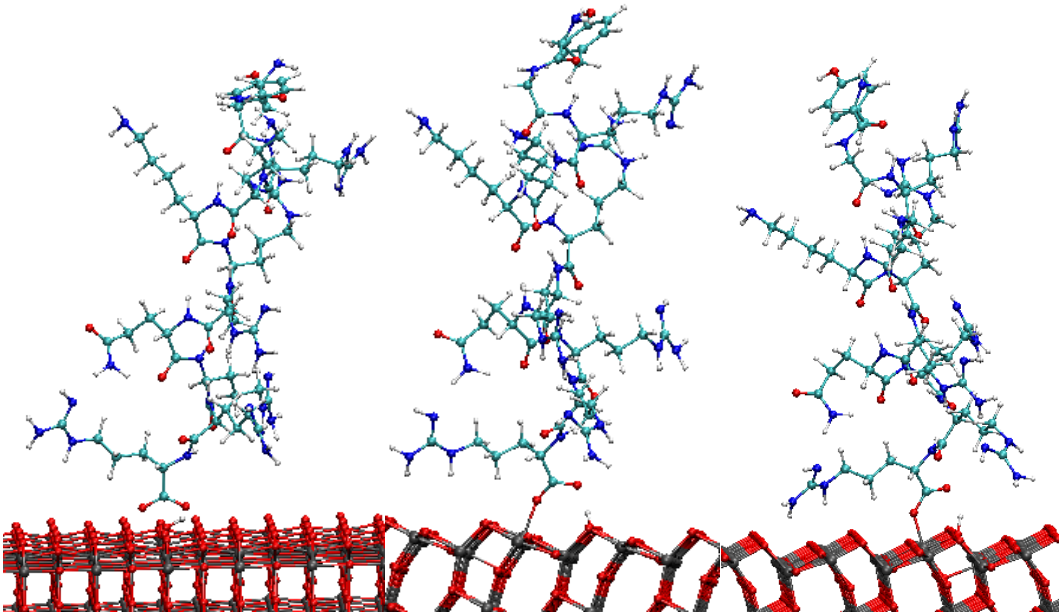


Figure 14. Molecule- and substrate- projected density of states of BB, M1 and M2 configurations of Arg_2 on TiO_2 obtained in DFTB with the tiorg-0-1* setup (top row) and with DFT (bottom row). The *y* axis crosses the energy axis at the Fermi energy.



(a) (b) (c)

Figure 15. Optimized structures of BB (a), M1 (b) and M2 (c) configurations of TAT adsorbed on anatase TiO_2 obtained with DFTB using tiorg-0-1* parameter set. Visualization here by VMD.

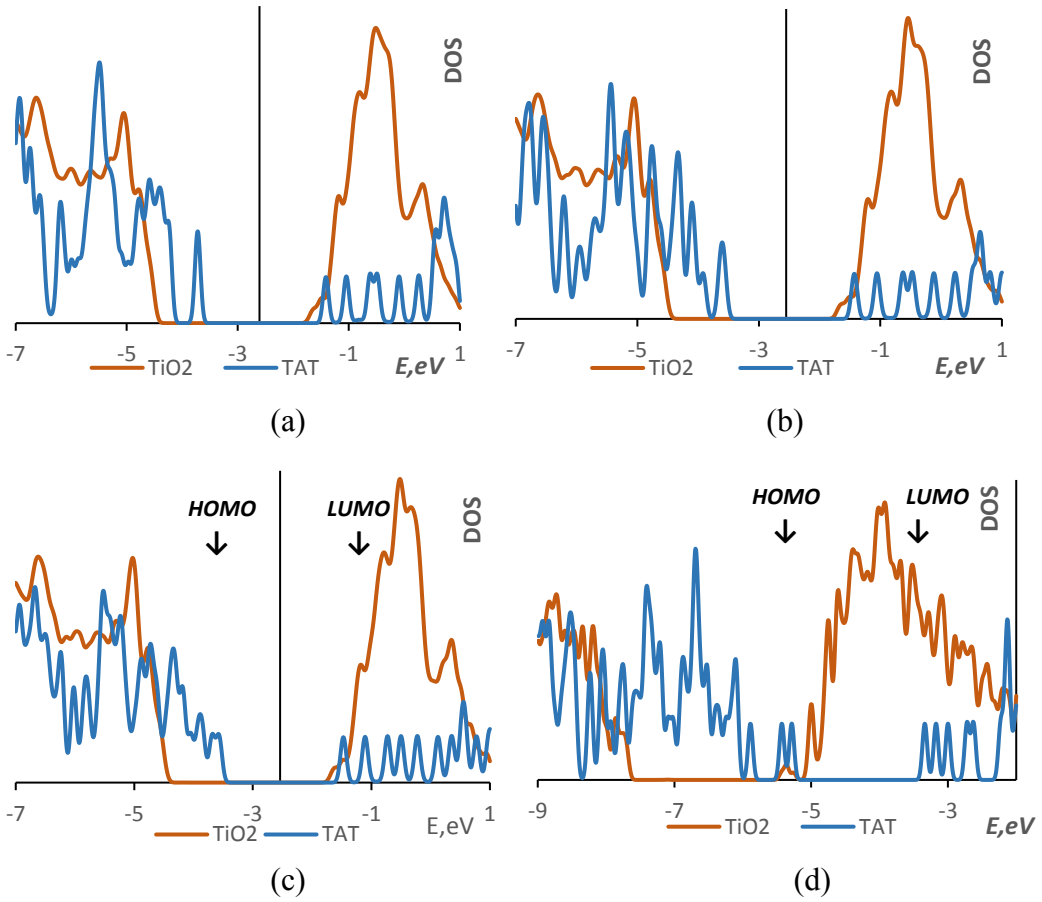


Figure 16. (a-c) Molecule- and substrate- projected density of states of BB, M1 and M2 configurations of TAT on TiO_2 obtained in DFTB with the tiorg-0-1* parameters. The *y* axis crosses the energy axis at the Fermi energy. (d) The simulated band alignment between the TAT and the (101) anatase surface. Approximate positions of molecular HOMO and LUMO energies are indicated with arrows.

List of Abbreviations

CBM	Conduction Band Minimum
CPP	Cell-Penetrating Peptide
DFT	Density Functional Theory
DFTB	Density Functional Tight Binding
DOS	Density of States
EAM	Embedded-Atom Method
FF	Force Field
GGA	Generalized Gradient Approximation
HOMO	Highest Occupied Molecular Orbital
LJ	Lennard-Jones
LUMO	Lowest Unoccupied Molecular Orbital
MD	Molecular Dynamics
NPT	Isothermal-Isobaric
PBC	Periodic Boundary Conditions
PBE	Perdew-Burke-Ernzerhof
PCM	Polarizable Continuum Model
PDOS	Projected Density of States
PP	Pseudopotentials
QM	Quantum Mechanics
TAT	Transactivator of Transcription
TB	Tight Binding
VBM	Valence Band Maximum

1 Introduction

Biomolecules are organic molecules which can form the basic structural constituent of a living cell. The organic compounds such as amino acids, nucleotides and monosaccharides can serve as building blocks of complex biomolecules[1], such as proteins, enzymes and nucleic acids. Those biomolecules are critical to life. Among them are peptides which have received prominence in molecular biology, because peptides not only allow the creation of peptide antibodies in animals without the need of purifying the protein of interest,[2] but also have become instrumental in mass spectrometry, allowing the identification of proteins of interest based on peptide masses and sequence. A special type of peptides are cell-penetrating peptides which are short peptides that facilitate cellular uptake of various molecules. For example, Lycosin-I (a kind of antimicrobial peptide derived from the venom of the spider Lycosa singorensis) is made up of by 24 amino acids with the sequence RKGWFKAMKSIAKFIAKEKL KEHL, and this kind of peptide is reported to be able to potently inhibit tumor cell growth in vitro.[3] Even though many experiments have been conducted to study their structures and properties, it is difficult with experimental techniques to provide a direct view of the molecular geometries at the atomic scale. Experiments also do not directly provide the mechanism of interactions among molecules or between biomolecules and other matter. Thus, computational simulation methods are usually applied to provide insights that are not fully accessible through experimental techniques, as well as to assess the accuracy of experimental data.

Biomolecules interacting with inorganic materials are important, as such interactions happen in living organisms; such interactions also can give rise to materials which are promising in the areas of nanotechnology and biomedicine[4-5], hence, bioinorganic materials and their interfaces have attracted increasingly significant attention. Applications of materials containing interfaces of metals or oxides with biomolecules range from dentistry[6] to drug delivery[7-8] to detection or sensing of molecules[9-14] and their conformers[15-16]. Studies of interactions of biomolecules with oxides (and specifically titania considered in this work) are important for understanding biocompatibility of implants,[17] toxicity of environmental particles,[18] and for the development of essays and sensors.[19-20] Basic information about interaction of large biomolecules with metallic or oxide surfaces can sometimes be obtained by studying interactions with individual aminoacids.[21-32] While both biomolecules and bioinorganic interfaces have been extensively studied experimentally,[15,17,23-25,27-32] experimental techniques do not provide a direct view of the molecular motions during adsorption, do not directly measure adsorption strength or geometry and do not directly provide the mechanism of interaction. Simulations are therefore necessary to provide a detailed understanding of the interactions between the biomolecules and inorganic matter, which is necessary for rational rather than ad hoc design of new materials with desired functionality.

For molecular modeling, there exist a multitude of methods of different levels of accuracy and different CPU cost, starting from classical force field based molecular dynamics methods (FFMD)[33-37] to first-principles quantum chemical

methods, such as density functional theory (DFT)[38-39] and various wavefunction based methods such as MP2[40-41] or the highly accurate coupled-cluster (CC)[42-44] methods. The wavefunction methods are of limited applicability to biomolecules as computational cost scales rapidly with the number of electrons N_e ($\sim N_e^4$ and $\sim N_e^6$ for MP2 and CCSD, respectively). DFT scales as $\sim N_e^3$ and is feasible for rather large molecules with 10^{2-3} atoms such as peptides. Order-*N* scaling DFT methods[45-47] have been developed but suffer from large prefactors. Due to their enormous CPU and memory cost advantage, force fields have been dominating the field of biomolecular simulation. The most commonly used types of force field such as CHARMM[48-53] or AMBER[54] represent key bonding and non-bonding interactions with simple parameterized functions (e.g. quadratic for covalent bonds or simple trigonometric functions for dihedral angles). Therefore, limited orders of coupling among nuclear degrees of freedom are included, and the simplicity of the functional form significantly limits the accuracy; patent failures of such FFs have been documented.[55]

Because bioinorganic interfaces consist of an organic and an inorganic subsystems, their modeling poses challenges, as highly accurate methods have been developed for the modeling of individual components but not necessarily of their interfaces. For the modeling of inorganic solids and their surfaces, DFT as well as force fields are also used. However, both DFT and FF schemes which are optimal for solids are not the same as those for molecules. Specifically in DFT, as periodic boundary conditions (PBC) are typically applied to model bulk and surfaces, plane wave expansions of the orbitals and density are often used, with

which it is necessary to use pseudopotentials (PP). This is often not necessary in molecular simulations. For the same reason, GGA (generalized gradient approximation) functionals such as the Perdew-Burke-Ernzerhof (PBE)[56] functional are typically used, while generally more accurate hybrid functionals are easier to apply to molecules. The net result is that it is difficult to use highly accurate DFT setups for bioinorganic interfaces. Even with a less costly setup such as a GGA functional and large core pseudopotentials, the computational cost of modeling interfaces with DFT is large; specifically, slab surface models of lateral extent of about 30 Å are not routinely computable. However, such large slab sizes are required for an efficient modeling of bioinorganic interfaces in view of the large sizes of biomolecules.

Force fields[57] used for metals and semiconductors also differ significantly from those typically used for molecules and, specifically, biomolecules. This has to do in particular with the fact that while for molecular systems, interaction energies rapidly converge in an expansion over orders of coupling[58-61] of intramolecular coordinates, this is generally not the case in metals and semiconductors. Two-body terms are usually described by empirical pairwise interatomic parametrized potentials consisting of a long-range Coulomb potentials, short-range Morse or Buckingham potentials, and Lennard-Jones potentials for the repulsive and Van Der Waals interactions.[62-66] Coupling terms are often described by density embedding schemes, such as the embedded-atom method (EAM) by Daw and Baskes.[67] As a result, it is difficult to model mixed systems with good accuracy also using force fields.

Overall, the accuracy of force field modeling can be much worse than that of DFT with errors in interaction energies and kinetic barriers on the order of tenth of an eV or even 1 eV.[68-69] Besides, FF simulations in principle cannot provide electronic structure and electron density information. It is an important limitation, as e.g. information about propensity to fold can be obtained by analyzing electron density.[70-71] Charges on atoms can be obtained from first principles if density is available. Further, interactions of biomolecules with inorganic matter can result in bond reformation which is difficult to account for with simple force fields. Reactive force fields[69,72] need to be used which are much more complex, often require tuning to a specific system, and may not approach ab initio accuracy.[69,73] Therefore, it is desirable to be able to perform electronic structure calculations on biomolecules and biomolecule containing interfaces, but such calculations are usually too costly at the DFT level.

DFTB (Density Functional Tight Binding)[74-82] is very attractive for the modeling of such interfaces because of a CPU cost which is about 3 orders of magnitude lower than that of DFT and amenability to near-linear scaling for large systems. The 3 orders of magnitude cost advantage is sufficient to routinely model biomolecules as well as sufficiently large slabs (surfaces) which can accommodate large biomolecules. DFTB therefore allows bringing ab initio accuracy and the ability to compute electronic properties to the modeling of systems of size which typically called for FF modeling. DFTB is an approximate DFT method which can provide high accuracy for system types for which it was parameterized, and it can be parametrized specifically for bioinorganic interfaces.

There already exist DFTB parameterizations for organic molecule-surface interactions, for instance DFTB was successfully employed to model acetic acid on TiO_2.[83] PTCDA (3,4,9,10-perylene tetracarboxylic dianhydride) on S-passivated GaAs,[84] and glycine on silica.[85] While as a method DFTB is certainly suitable for modeling of bioinorganic interfaces with DFT-like accuracy, a specific parameterization might not be.[86]

1.1 Purpose of Thesis and Problem Statement

The purpose of this thesis is to benchmark DFTB for biomolecules and Bioinorganic interfaces. We start with the modeling of small biomolecules like amino acids. Many experiments have been conducted to study their structures and properties,[87-90] and they are small enough to be computable ab initio with accurate methods like DFT with hybrid functionals. Therefore, we can perform a reliable benchmarking of DFTB performance for individual amino acids. To be specific, in this section we apply both DFT and DFTB methods to study the optimization of all 20 amino acids.

Next we will focus on a large biomolecule and specifically the anti-cancer toxin peptide Lycosin-I, which has also been studied through experiments by our collaborators in Hunan Normal University. Lycosin-I is able to potently inhibit tumor cell growth in vitro, and suppresses various tumor growth in vivo when tested in human cancer xenograft models.[3] The preliminary experimental results demonstrated that the Lycosin-I peptide can indeed interact with cell membrane and then be internalized into cytoplasm of cancer cells to initiate the programmable cell death.[3] It was also deduced by experimental measurements

that the majority of Lycosin-I displayed a monomer state in diverse solutions regardless of the concentration of peptide and the incubation time.[91] In our work, we compare force-field molecular dynamics (FFMD), which is the most widely used in the field of biomolecular simulations, with density functional tight binding (DFTB) method when studying the conformers of Lycosin-I and to prove the weak binding of Lycosin-I molecules between themselves through computational simulations. Furthermore, we apply DFTB to study the electron structures of 5 different kinds of cell-penetrating peptides, which in principle is impossible for MD simulations.

In the final section, we aim to assess the performance of DFTB with two parameter sets: matsci-0-3[92] and a combination of mio-1-1[74] and tiorg-0-1[93] when modeling arginine-rich cell-penetrating peptides and their interactions with an inorganic surface. Arginine-rich cell-penetrating peptides are important for their capability to act as delivery vehicles,[94] which means they can deliver certain small molecular drugs and other therapeutic agents into the cytoplasm. Examples include the short arginine-rich transduction domain of the HIV-1 transactivator of transcription (TAT) protein.[95] It is worth noting that TAT has attracted the most attention as a prototypical example that has many of the essential characteristics of the arginine-rich cell-penetrating peptides.[96] Materials which contain titanium (Ti) are often found in biological environment, specifically, Ti is a widely used implant material, and their surfaces are usually oxidized under ambient conditions.[97] Thus understanding mechanistic details of the interactions between TiO_2 and Arginine is useful to understand the way peptides interact with oxides,

which may help improve the design of safe nanomaterials that could find use in nanomedicine.[98] Because of the significant computational cost of DFT modeling of a peptide-titania interface, the strategy we adopt is to study in detail the adsorption of arginine and arginine dipeptide on TiO_2 with both DFT and DFTB and to relate adsorption properties to the band structure and specifically band alignment of the molecule with titania, we then consider TAT adsorption on titania by modeling it directly with DFTB, and simulating indirectly the band alignment expected with DFT.

2 Methodology

2.1 Molecular Dynamics

Molecular dynamics (MD) is a computer simulation method for studying the physical movements of atoms and molecules, and is thus a type of N-body simulation. The atoms and molecules are allowed to interact giving rise to the dynamical evolution of the system. In classical MD simulations, the atoms move according to the Newtonian equations of motion. [34]

$$m_\alpha \ddot{\vec{r}}_\alpha = -\frac{\partial}{\partial \vec{r}_\alpha} U_{total}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N), \quad \alpha=1,2\ldots N, \qquad (1)$$

where m_α is the mass of atom α, $\vec{r}_\alpha$ is its position, and U_{total} is the total potential energy that depends on all atomic positions of N atoms. The potential energy, represented through the MD "force field," is the most crucial part of the simulation because it must faithfully represent the interaction between atoms, yet be cast in the form of a simple mathematical function that can be calculated quickly. The set of equations (1) together with initial conditions in principle

allows computing positions of all atoms at any point in time within numeric errors.

In our work, the CHARMM22[99] force field was used, which had been specially optimized for biomolecules including DNA, RNA and lipids. The form of CHARMM force filed is:

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{Coulomb} \quad (2)$$

The first three terms describe the stretching, bending, and torsional bonded interactions,

$$U_{bond} = \sum_{bonds} k_b(b - b_0)^2 \quad (3)$$

$$U_{angle} = \sum_{angles} k_\theta (\theta - \theta_0)^2 \quad (4)$$

$$U_{dihedral} = \sum_{dihedrals}[\sum_j k_{\varphi,j}(1 + \cos(n_j\varphi - \delta_j))] \quad (5)$$

The bond and angle terms are harmonic, with force constants k_b and k_θ and equilibrium values b_0 and θ_0. The dihedral potential is a sum of sinusoids with force constant $k_{\varphi,j}$, multiplicity n, and offset δ, and j can range from 1 to 6.[100]

The final two terms in eq. (2) describe interactions between nonbonded atom pairs:

$$U_{vdW} = \sum_{nonbonded\ pairs\ i,j} \varepsilon_{ij}[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2(\frac{R_{min,ij}}{r_{ij}})^6] \quad (6)$$

$$U_{Coulomb} = \sum_{nonbonded\ pairs\ i,j} \frac{q_i q_j}{\varepsilon_D r_{ij}} \quad (7)$$

A Lennard-Jones (LJ) potential models the van der Waals interactions, where ε_{ij}

is the potential energy minimum between two particals separated by r_{ij}, and $R_{min,ij}$ is the position of this minimum. Q_i and q_j are the partial atomic charges for the Coulombic term; the dielectric constant ε_D equals 1 in explicit solvent simulations.[100] As is clear from eqs. (3-6), the interactions are here modeled with very simple functions. Even two-body interactions are modeled very approximately (anharmonicity is not included in eq. (3)). Multi-body effects are also modeled only selectively (through 3-body and 4-body angles only) with simple functions. Effects such as bond softening (or hardening) due to chemical environment or other multibody effects are excluded.

2.2 Quantum Mechanics and Density Functional Theory

In contrast to MM, quantum mechanics (QM) approaches compute the function *U* as the electronic energy, using some approximation to the electronic Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m}\sum_{i=1}^{N}\nabla_i^2+\sum_{i=1}^{N}V(\boldsymbol{r}_i)+\sum_{i=1}^{N}\sum_{j<i}U(\boldsymbol{r}_i,\boldsymbol{r}_j)\right]\Psi=E\Psi \qquad (8)$$

Where for an N-electron system, $\hbar$ is the Plank constant, m is the mass, E is the total energy, V is the potential energy from the external field (e.g. due to positively charged nuclei), Ψ is the wave function and U is the electron-electron interaction energy.

Computing U of Eq. (1) from Eq. (8) in principle obviates the sources of error in force fields mentioned above. The solution of Eq. (8) is orders of magnitude costlier than the recall of an analytic function *U*. In fact, for molecular sizes typical of biomolecules, a solution involving Ψ is prohibitively expensive, and an

approximate method described in the next section is the only practical approximation. This is in particular due to the very large dimensionality of Ψ which depends on the position of all electrons in the system:$\Psi = \Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N)$.

For decades, ab initio materials and molecular (except for very small molecules) simulations have been dominated by Kohn-Sham (KS-) DFT (Density Functional Theory)[38,39] as the only practical option. DFT[38-39] expresses the ground state energy of any system with N_e electrons (E of eq. (8)) as a functional of its electron density$\rho(\boldsymbol{r})$: $E = E[\rho(\boldsymbol{r})]$. In KS-DFT, ρ is computed as $\rho = \sum_{i=1}^{N_e} \rho_i(\boldsymbol{r})$, where $\rho_i = |\phi_i(\boldsymbol{r})|^2$. The ϕ_i (called orbitals) can be found from the KS equation

$$-\frac{1}{2}\nabla^2\phi_i + V_{eff}(\boldsymbol{r}|\rho)\phi_i = \epsilon_i\phi_i \qquad (9)$$

With

$$V_{eff}([\rho];\vec{r}) = v(\vec{r}) + \int \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} d\vec{r}' + \frac{\delta E_{XC}[\rho]}{\delta\rho} \qquad (10)$$

$$\rho(\vec{r}) = \sum_{i=1}^{N} |\phi_i|^2 \qquad (11)$$

Where the effective potential V_{eff} includes an approximation to exchange and correlation (XC), with which the one-particle eq. (9) can model the real system of interacting electrons.

In KS-DFT, it is necessary to compute at least as many orbitals as there are electrons. Further, the KS equation must be solved iteratively (as V_{eff} depends on ρ). The solution involves large basis set expansions of the orbitals. As a result, the scaling of CPU cost is $O(N_e^3)$. Simulations are therefore limited to very small

model systems, typically of dozens to hundreds of atoms and become unwieldy for the modeling of large organic molecules like cell-penetrating peptides (CPP). DFT simulations for individual amino acids are, however, not computationally costly and are performed here.

As a QM method, DFT automatically considers all couplings between nuclear degrees of freedom, and its accuracy is not compromised by reliance on simple functions like those of eqs. (3-6). DFT can naturally treat mixed (e.g. organic-inorganic) systems and interfaces for which different types of force fields are typically developed. The exchange-correlation term *is* approximated, but errors due to that approximation are typically much smaller than errors of force fields. As a result, the accuracy of DFT calculations is much higher than that of force-fields mentioned above (e.g. accuracy of interaction energies of better than 0.1 eV are typically achievable). In fact, many force fields are fitted to DFT data. It would be desirable to bring DFT accuracy to the simulations of large biomolecules and interfaces.

2.3 Density Functional Tight Binding

The tight binding (TB) approaches work on the principle of treating electronic wavefunction of a system as a superposition of atom-like wavefunction (known as LCAO, or Linear Combination of Atomic Orbitals, approach). Beyond the confines of a minimal LCAO basis, valence electrons are tightly bound to the cores and are not allowed to delocalize. The TB energy for N electrons, M atoms system is:

$$E_{TB} = \sum_i^N \epsilon_i + \frac{1}{2}\sum_{j\neq k}^M v^{j,k}\left(\left|R_j - R_k\right|\right) \qquad (12)$$

This separation of one-electron energies (essentially on-site energies) and interatomic distance-dependent potential $v^{j,k}$ constitutes the TB method. ϵ_i are eigenvalues of the KS equation and solved variationally using atom-like (minimum, single-zeta) AO basis set, leading to a secular equation:

$$|H - \epsilon S| = 0 \qquad (13)$$

Where H and S are Hamiltonian and overlap matrices in the basis of the AO functions.[78] H and S are constructed using nearest-neighbor relationships; typically only nearest-neighbor interactions are considered. When calculating the one-electron integrals, the Slater-Koster (SK) approximation is used and the integrals are converted into two-center one electron integrals.[75]

Density Functional Tight Binding (DFTB) method is an approximation to DFT.[38-39] The method makes use of an optimized atomic orbital minimal basis set and approximated the effective Kohn−Sham (KS) potential in a rigorous two-center approximation. Thus, terms of three or more centers are not present in the KS matrix, allowing the a priori calculation and tabulation of all matrix elements, employing the so-called Slater−Koster technique.[75] The tabulated matrix elements as well as other data needed for DFTB calculations (such as polynomial repulsive parameters) form the DFTB parameter sets which are stored in the so-called Slater-Koster files; different versions of these parameters (named 3ob-2-1, matsci-0-3, mio-1-1, and tiorg-0-1) are used below. Because of the algorithmic simplicity of the method, it has been straightforward to implement it early on massively

parallel high-performance computers;[101-102] all computationally intensive operations of DFTB are available in standard linear algebra packages and thus highly optimized for virtually all computer architectures.

2.4 Software and Computational Parameters

For the modeling of amino acids, DFT calculations of 20 kinds of amino acids were performed with the Gaussian 09[103] (G09) program using the B3LYP[104-107] functional and the 6-31g+(d,p) basis set for the purpose of comparing the results of the self-consistent charge Density Functional Tight Binding method[74-82] with a quantitatively accurate DFT setup. SCC-DFTB calculations were performed with the DFTB+[74] software using the 3ob-2-1[108-109] parameter set (Slater-Koster file which is designed for organic and biological applications).

For the modeling of the cell-penetrating peptides, Molecular Dynamics (MD) simulations were performed using the CHARMM force field[48-53] and the NAMD[110] software. Isothermal-isobaric (NPT) simulations were performed at 300 K and ambient pressure in water. A rectangular simulation cell with periodic boundary conditions was used. The cell size was 50x50x50 Å. The time step size was 1 fs. The initial peptide structure was generated using PEP-FOLD[111-112] and optimized in NAMD. The resulting structure is shown in Figure 5(b). Simulations were performed in water and vacuum.

A reference ab initio calculation of the peptide using the self-consistent charge Density Functional Tight Binding method was also performed, and the DFTB+ software was used with the 3ob-2-1 parameter set. That this setup provides

accuracy similar to DFT is shown by comparative simulations of individual aminoacids. The dispersion interactions were modeled with UFF[113]. Due to the higher computational cost and unavailability of a continuum solvent model in DFTB+, DFTB simulations were only performed in vacuum. Besides, the waveplot program from the DFTB+ package was used to provide the charge distributions of 5 kinds of cell-penetrating peptides.

For the modeling of bioinorganic interfaces, DFT calculations for isolated arginine and TAT molecules were performed with the Gaussian 09 (G09) program using the B3LYP functional and the 6-31g+(d,p) basis set for the purpose of comparing the DFTB results with a quantitatively accurate DFT setup. For arginine, we have confirmed that the 6-31g+(d,p) basis set provides converged values vs. 6-31g++(2d,2p) (see Table 2).

DFT calculations of the adsorption of Arg on TiO_2 were performed using the periodic slab surface model of titania. The SIESTA code[114] was used with the PBE[55] functional and a DZP (double-ξ polarized) basis. Troullier-Martins pseudopotentials[115] were used to account for core electrons. Basis functions of different width were used (as defined by specifying the PAO.EnergyShift parameter, which has the meaning of the extent of destabilization of atomic energy levels with a given basis vs. exact levels, and thereby it effectively controls the basis width) to quantify possible artifacts due to the use of an atom-centered basis set. The tolerance for the relative change in the density matrix was set to 1×10^{-5}. The anatase TiO2 band gap with this scheme is 2.4 eV[116] and is largely preserved for the (101) surface and the adsorbate systems (see DOS plots

below). The (101) surface of TiO_2 was modelling with a 4-Ti layer slab (144 atoms) of lateral size 11.3×10.3 Å and 26.0 Å of vacuum for arginine and 37.0 Å of vacuum for arginine dipeptide. The Brillouin zone was sampled with a 2×2×1 Monkhorst-Pack[117] grid point and a 200 Ry cutoff was used for the Fourier expansion of the electron density. The positions of atoms in the lower half of the slab were fixed at bulk positions, and all other atoms and lattice vectors were allowed to relax until forces were below 0.02 eV/Å and stresses below 0.1 GPa, respectively. The dipole correction[118-119] in the direction normal to the surfaces was checked but found to be insignificant.

SCC-DFTB calculations on the interfaces were performed with the DFTB+ software. We used two different parametrizations, matsci-0-3[92] and a combination of mio-1-1[74] and tiorg-0-1[93] (this combination will be referred to below as tiorg-0-1* for simplicity). The anatase band gap with both matsci-0-3 and tiorg-0-1 is 3.2 eV,[93,120] in good agreement with experimental estimates of 3.2-3.3 eV.[121] This gap is largely preserved for the (101) surface and the adsorbate systems (see DOS plots below). A 4-Ti layer slab of lateral size 20.59×20.95 Å and 13.4 Å of vacuum was used for arginine on TiO_2; the Brillouin zone was sampled with a 2×2×1 Monkhorst-Pack[117] grid of points. For TAT on TiO_2, a larger slab of 41.5×43.8 Å and vacuum layer of 60 Å were needed for the proper relaxation of the TAT molecule on the anatase surface, and only the tiorg-0-1* parametrization was used in the calculations because with these parameters we obtained more reliable results for arginine adsorption on anatase (101) surface. Here, the Brillouin zone was sampled at the Γ point due to the large size of the cell.

The adsorption energy is computed as:

$$E_{ads} = E_{titania/arginine} - E_{titania} - E_{arginine} \qquad (14)$$

where $E_{titania/arginine}$ is the total energy of the bioinorganic interface system, $E_{titania}$ the total energy of the clean anatase (101) surface and $E_{arginine}$ the total energy of the isolated arginine molecule. A negative E_{ads} value therefore corresponds to favored adsorption.

Previous ab initio studies imposed different charge states on the amino acids (e.g. on arginine molecule), including neutral.[21-22,122-130] Different codes used here allow for different treatments of solvent effects (Polarizable Continuum Model (PCM)[131-135] in Gaussian, only explicit solvation in SIESTA and DFTB+, which is not practical for interfaces used here) which tend to produce different results beyond the differences induced by a specific DFT/DFTB approximation to the electronic structure. Therefore, to enable proper comparison between the two DFT setups and the DFTB setup, all systems (molecules and interfaces) were modelled in vacuum and in neutral state. For the same reasons, we do not include dispersion corrections; without such corrections, we obtain relatively strong chemisorption except for adsorption via amino groups in DFTB; in this latter case we confirmed that inclusion of dispersion corrections does not change the results (changing the adsorption energy on the order of 0.1 eV). We also focus our study on single molecule adsorption, i.e. the arginine molecule and the dipeptide arginine as well as the larger arginine-rich peptide molecule in form of TAT. Only the linear conformer of Arg was considered, similarly to all previous works studying

arginine adsorption[22,125-128,130] on TiO_2, and because it is the conformation it assumes in TAT (see Figure 3(b)).

Trial TAT structures were generated using PEP-FOLD and optimized with DFTB, whereby the lowest energy conformer was chosen and used in all calculation.

3 Results

3.1 Amino Acids

3.1.1 DFT and DFTB studies of 20 Amino Acids

Both DFT and DFTB calculations were performed for a set of 20 kinds of amino acids. In Figure 1, the structures of these molecules optimized with DFT (using Gaussian setup with the hybrid B3LYP functional) are shown. Due to a large numbers of molecules, DFTB optimized geometries are not shown here, but their structures are similar between these two methods. The differences in bond lengths of C-C, C-N, and C-O bonds and angles involving these atoms for each kind of amino acid between these two computational schemes are measured through VESTA[136] here and elsewhere and they are given in Table1. The good agreement in bond lengths can be observed and only angular coordinates have noticeable differences between these two methods. It can be concluded that these two computational schemes provide largely similar structural information. The differences in energy between two conformers of 17 amino acids in each computational scheme are also shown in Table 1. The observed differences in

Table 1 between DFT and DFTB can be an indication of what kind of accuracy can be expected from DFTB when modeling larger biomolecules.

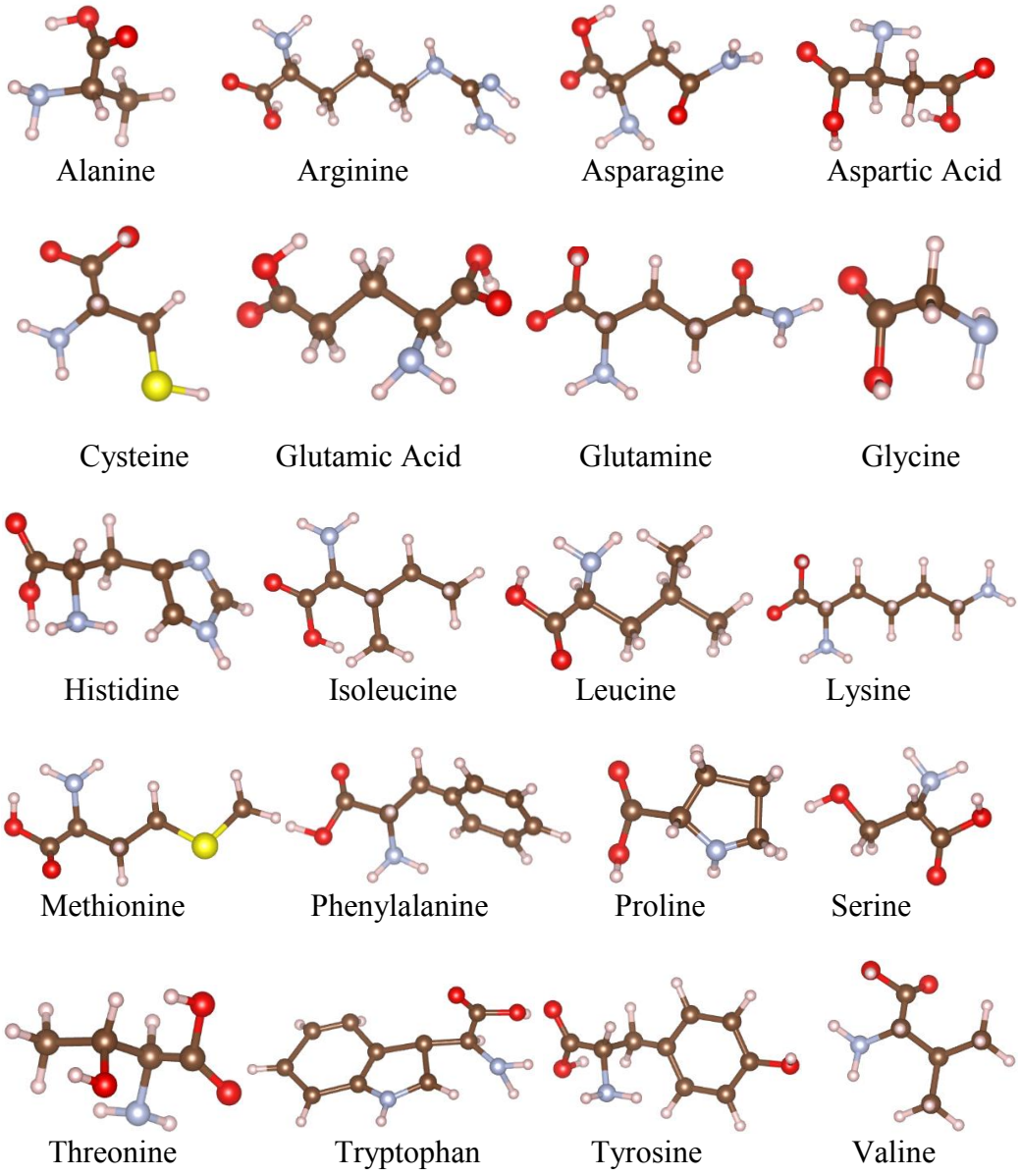


Figure 1. Optimized structures of 20 amino acids obtained with Gaussian09. The atom color code here: C, brown; O, red; H, light grey; N, gray; S, yellow. Visualization here by VESTA.

Table 1. Mean Absolute Difference of bond lengths (Å) and angles (°) for each amino acid (MAD_{bond} and MAD_{angle}) between the Gaussian setup and the DFTB with 3ob-2-1 parameter set, as well as the energy difference between the conformer and the original structure ($\Delta E_{conformer}$, in eV) in both computational schemes.

	MAD_{bond}	MAD_{angle}	$\Delta E_{conformer}$	
			DFT	DFTB
Alanine	0.006	0.57	/	/
Arginine	0.006	1.42	0.03	0.02
Asparagine	0.005	1.86	0.02	0.19
Aspartic Acid	0.006	1.12	0.47	0.21
Cysteine	0.005	1.87	0.08	0.01
Glutamic Acid	0.004	1.33	0.02	0.04
Glutamine	0.006	1.20	0.03	0.04
Glycine	0.005	1.45	/	/
Histidine	0.008	0.56	0.11	0.03
Isoleucine	0.007	0.76	0.01	0.00
Leucine	0.009	1.68	0.50	0.14
Lysine	0.006	0.55	0.01	0.03
Methionine	0.007	0.59	0.01	0.16
Phenylalanine	0.005	0.65	0.07	0.04
Proline	0.011	0.92	/	/
Serine	0.009	3.02	0.06	0.08
Threonine	0.009	1.36	0.07	0.06
Tryptophan	0.006	0.67	0.35	0.17
Tyrosine	0.005	0.69	0.13	0.04
Valine	0.005	2.03	0.24	0.01
Average over all	**0.007**	**1.22**	**0.13**	**0.07**

3.1.2 Molecular Calculations of Arginine, Arginine Dipeptide and TAT

As the adsorption of arginine based molecules on an oxide surface will be studied in a later chapter, it is important to know their molecular geometry and electronic structure. In this section, the molecular and electronic structures of arginine, arginine dipeptide and TAT are compared, as obtained with DFT and DFTB.

In Figure 2, the structures of these molecules optimized with DFTB (using the mio-1-1 parameter set which is matched with tiorg-0-1 used for TiO_2 calculations, as well as the matsci-0-3 parameter set) and with DFT using two setups: the higher-accuracy Gaussian setup with the hybrid B3LYP functional suitable for molecules, as well as the SIESTA with the PBE functional suitable for studying adsorption on TiO_2 are shown. The wireframe structures in Figure 1 (a-c) are computed with, respectively, DFT/SIESTA, DFTB/mio-1-1, and DFTB/mastic-0-3 and are overlaid with structures computed with the more accurate Gaussian setup. The agreement is good with noticeable differences observed only in angular coordinates. The differences in bond lengths and angles for Arginine between these computational schemes are given in Table 2 together with differences in structural parameters related to the bond between two amino acids in the dipeptide (Figure 3). These comparisons are expected to hold as well for TAT (also shown in Figure 3) and other peptides. From Table 2 and Figure 2 one can conclude that all four computational schemes provide largely similar structural information, in particular, that DFTB provides similar accuracy to DFT and especially mio-1-1

parameter set can provide more similar results to DFT compared with matsci-0-3. The largest differences are observed in dihedral angles and can reach 29 degrees with DFTB (matsci-0-3). Table 1 also shows differences in energy between two conformers of Arg (shown in Figure 2(a-c) and 2(d), respectively). The differences observed in Table 2 between the most accurate computational setup we used here (i.e. Gaussian) and which is typical of those used for the modeling of small molecules, and the DFT setup (SIESTA) typical of those used for the modeling of bulk materials and interfaces is an indication of what kind of accuracy can be expected from DFT when modeling bio-inorganic interfaces; this should be kept in mind when comparing DFT and DFTB results for adsorbate systems in the next section. Table 2 also shows that our G09 and SIESTA results are converged with respect to the choice of the basis.

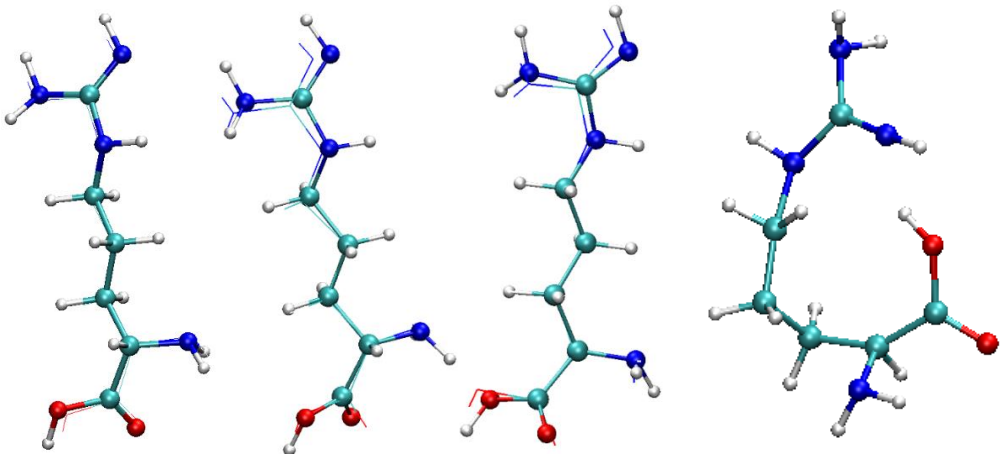
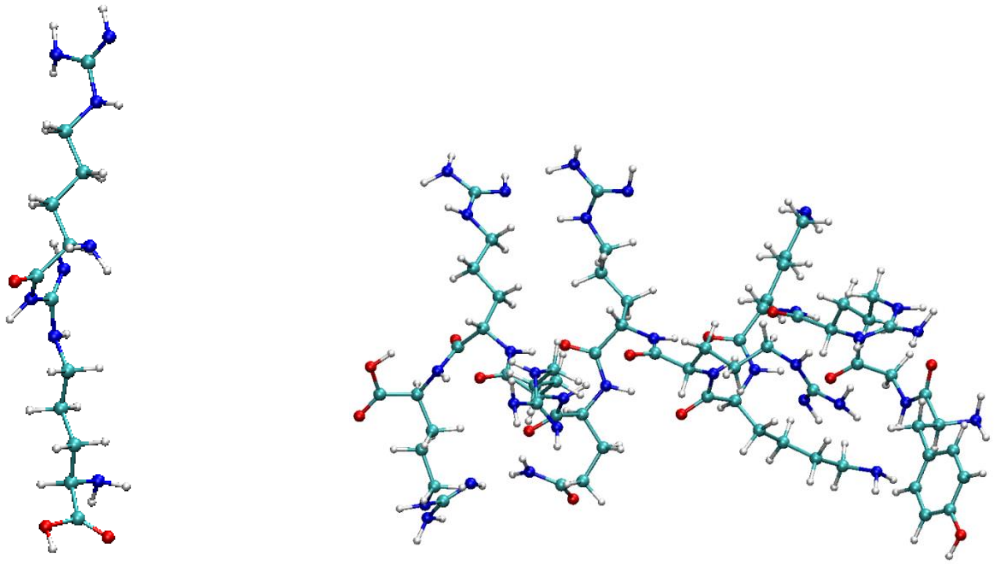


Figure 2. Wireframe optimized structures of Arginine with (a) DFT/SIESTA, (b) DFTB/mio-1-1, and (c) DFTB/matsci-0-3, overlaid with ball and stick optimized structures from DFT/Gaussian. Panel (d) shows the lowest-energy structure from Ref. [122]. The atom color code here and elswehere with VMD[137]: C, green; O, red; H, light grey; N, blue.



(a) (b)

Figure 3. Structures of Arginine dipeptide (a) and the TAT peptide (b). The structures computed with Gaussian 09 are shown and are visually similar for all computational setups used here. Visualization here by VMD.

Table 2. Mean Absolute Difference of bond lengths (Å) and angles (°) for Arginine (MAD_{bond} and MAD_{angle}) and the length of the peptide bond ($\Delta B_{pep\text{-}bond}$) between the Gaussian setup and the other three computational schemes, as well as the energy difference between the conformer shown in Figure 1 (a-c) and the cyclic structure shown in Fig. 1(d) ($\Delta E_{conformer}$, in eV).

	MAD_{bond}	MAD_{angle}	$\Delta B_{pep\text{-}bond}$	$\Delta E_{conformer}$
DFT/G09/6-31g+(d,p)	/	/	/	0.07
DFT/G09/6-31g++(2d,2p)	0.002	0.08	-0.004	0.07
DFT/SIESTA[a]	0.011	0.44	0.008	0.34
DFT/SIESTA[b]	0.011	0.43	0.011	0.35
DFTB/Mio-1-1	0.013	1.30	0.006	0.20
DFTB/Matsci-0-3	0.033	1.77	-0.028	0.15

[a] Broader basis functions (PAO.EnergyShuft = 0.001 Ry)

[b] Narrower basis functions (PAO.EnergyShuft = 0.002 Ry)

In Figure 4, the densities of states (DOS) of the Arg, Arg dipeptide, and TAT following from the four methods are compared. We can see that the HOMO-LUMO gap (the energy difference between the highest occupied molecular orbital and the lowest unoccupied molecular orbital) narrows with the increase of the size of the molecule due to both an increase of the HOMO energy and decrease of the LUMO energy. Specifically, the HOMO of TAT is higher than that of Arg by 0.66, 0.85, 0.82, and 0.90 eV with DFTB/matsci-0-3, DFTB/mio-1-1, DFT/G09, and DFT/SIESTA, respectively, while the LUMO is lowered by 1.13, 1.32, 0.54, and 0.81 eV, respectively. What we will show is consequential to the modeling of peptide adsorption on oxides is the fact that the gap of a large biomolecule, TAT in this case, is much smaller than that of individual aminoacids, and its HOMO is much higher. We will show in what follows that this has a significant effect on band alignment with titania and also affects the performance and applicability of DFTB and DFT setups. The band gap obtained with SIESTA is much smaller than with Gaussian or DFTB due to the use of the PBE functional, as expected[138]. The dotted line on the "Gaussian" panel of Figure 4 illustrates this by showing the DOS obtained for Arg in Gaussian with the PBE functional.

By now, we have benchmarked DFTB for small biomolecules, but it is also important to know what kind of accuracy DFTB can achieve for large biomolecules. Therefore, in next section, we will focus on the simulation of large biomolecules, such as cell-penetrating peptides, using both DFTB and FFMD methods.

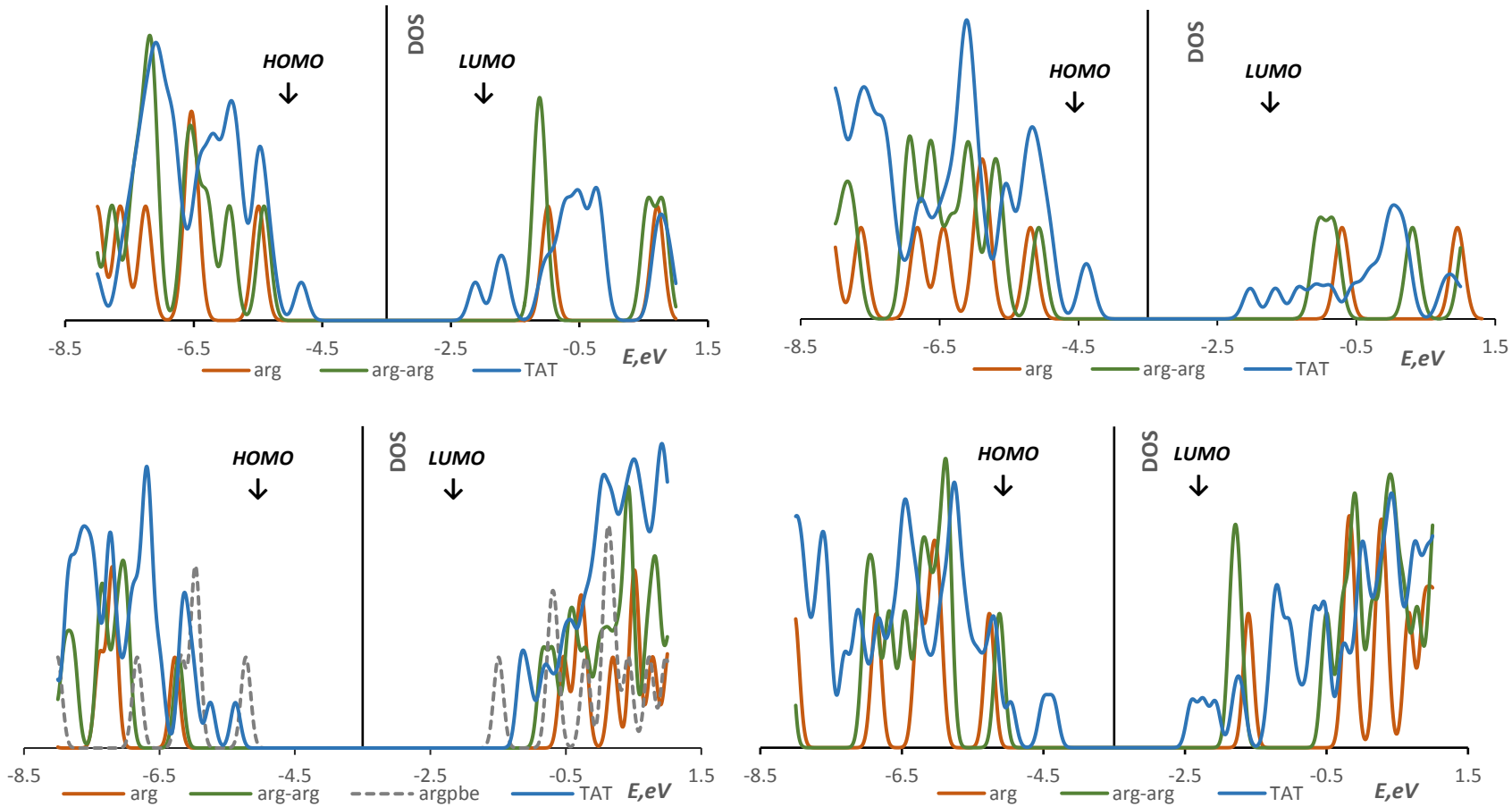
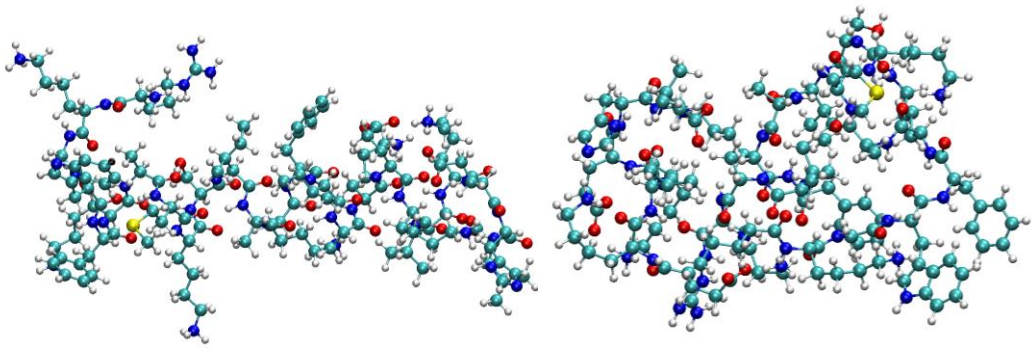


Figure 4. Densities of states of Arg, Arg dipeptide (arg-arg) and TAT computed with DFTB (top left: matsci-0-3 and top right: mio-1-1 parameter sets) and DFT (bottom left: G09/B3LYP, bottom right: SIESTA/PBE). A Gaussian broadening of 0.1 eV is applied. Approximate positions of HOMO and LUMO energies are indicated with arrows (i.e. HOMO is the highest-energy peak before the gap and LUMO is the lowest-energy peak after the gap; the *y* axis has been positioned inside the gap).

3.2 Cell-Penetrating Peptide Lycosin-I

3.2.1 FFMD and DFTB studies of Monomer

The initial structure of Lycosin-I was obtained from PEP-FOLD based on its amino acid sequence (RKGWFKAMKSIAKFIAKEKLKEHL), as shown in Figure 5 (a). Molecular dynamics simulations on the monomer were performed in vacuum and water with NAMD. The vacuum simulation resulted in folding after 13 ps, as shown in Figure 5 (b). The difference of total energy after optimization between initial structure and folded structure is small (0.12 eV), but folded structure has lower total energy, which means its structure is more preferred. No folding was observed in water within the timeframe of the simulation (500ps) but simulations of the folded structure in water in NAMD didn't show any unfolding. The folding was also confirmed by the DFTB calculation. Due to the higher computational cost, DFTB simulations were only performed in vacuum. The folding was observed after 25 ps in NVT simulations and no unfolding was observed thereafter (up to 500 ps). The folded structures obtained in NAMD and DFTB simulations were similar.



(b) (b)

Figure 5. Initial structure (a) and folded structure (b) of Lycosin-I. Visualization here by VMD.

3.2.2 FFMD and DFTB studies of Peptide Dimers

As simulations of the folded structure in water in NAMD did not show any unfolding, four putative dimer configurations with different mutual orientations of the folded monomer were generated to study their self-interaction in vacuum, as shown in Figure 6. In these dimer configurations, the molecules were initially located a few Angstrom from each other. Both FFMD and DFTB calculations were performed on them in vacuum, and their binding energies in each computational scheme are shown in Table 3. FFMD shows in general similar binding energies compared with DFTB for all four configurations, specifically, the order of binding strength among dimer configurations is the same. Besides, both of these two method indicate that the (a) dimer configuration in Figure 6 owns the strongest interaction.

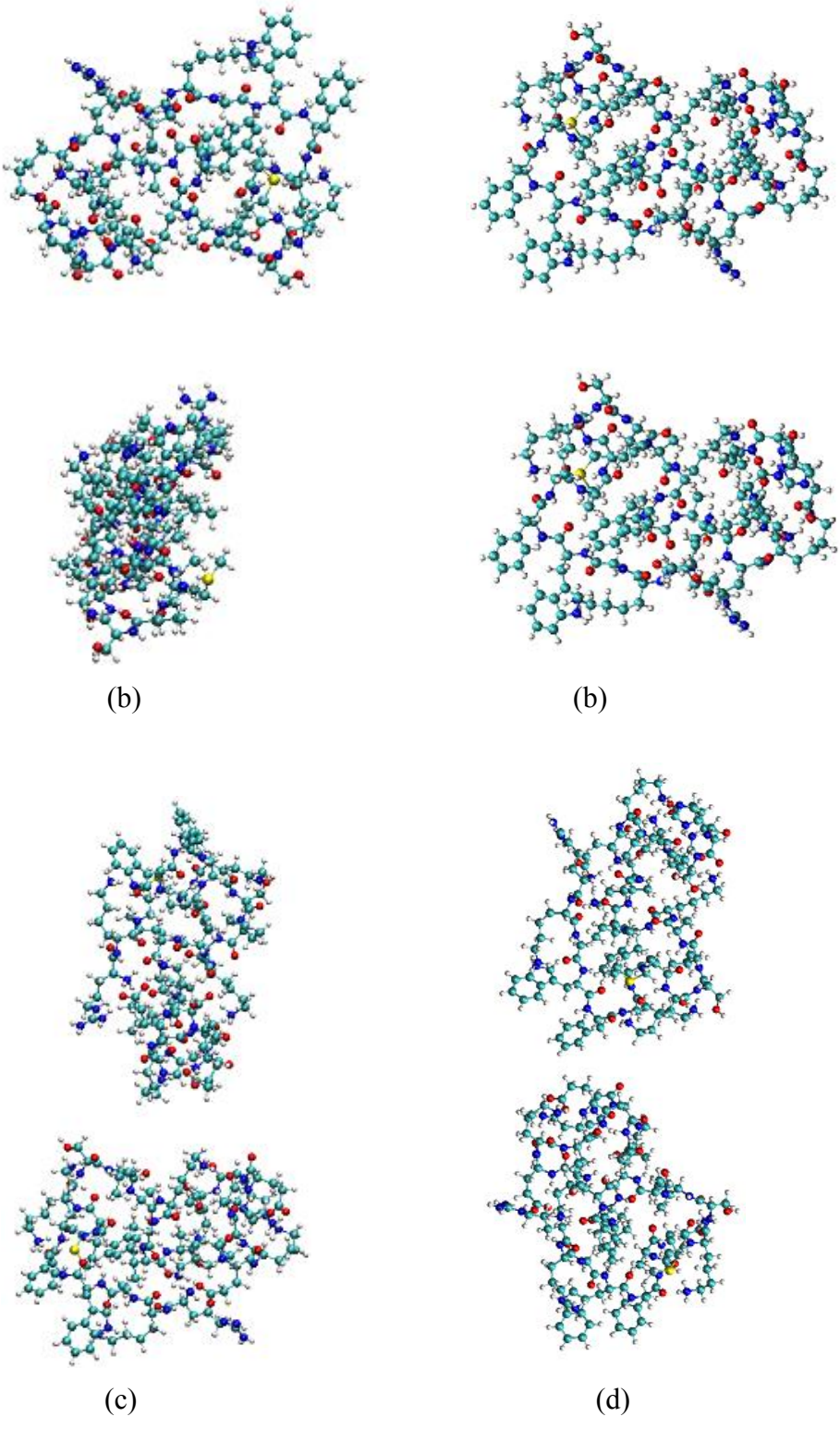
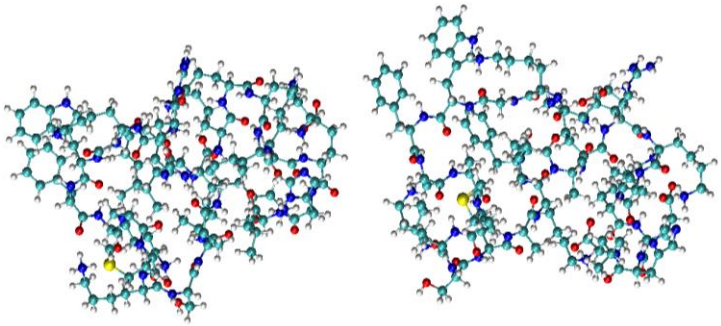


Figure 6. Structures of four putative dimer configurations. Visualization by VMD.

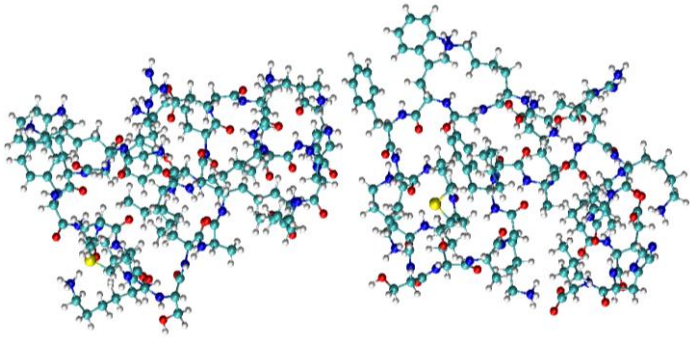
Table 3. Binding energies (eV) of four putative dimer configurations in vacuum.

Dimer	ΔE_{MD}	ΔE_{DFTB}
a	-2.1	-2.69
b	-0.67	-1.05
c	-1.36	-1.6
d	-1.98	-2.08

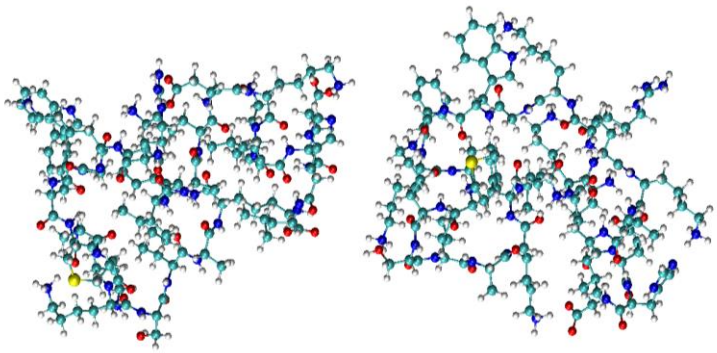
It was found that the majority of Lycosin-I displayed a monomer state in diverse solutions regardless of the concentration of the peptide and the incubation time through experiments.[139] Therefore, FFMD simulations were also performed on dimers in water. However, we noted that it was impossible for us to compute an exact ΔE like we did in vacuum, because that would require a prohibitively expensive sampling of all possible orientations of the solvent molecules. Therefore we only study the trajectory of the peptides. The initial monomer separation of a few Å (Figure 7a) was "noncommittal" in the sense that it was small enough for the molecules to come close if the binding was strong; alternatively, they would be able to drift apart if it was weak. A representative snapshot along the MD trajectory of a dimer simulation is shown in Figure 7. What we observed is that monomers are able to repeatedly approach each other (Figure 7b) and to drift apart (Figure 7c), which implies that any dimers formed can also be dissociated at room temperature and the observed weak clustering behavior of the CPP in solution is corroborated by MD simulations.



(a)



(b)



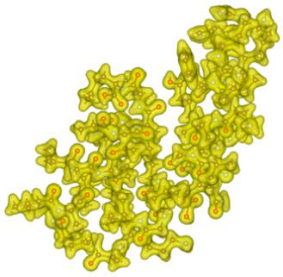
(c)

Figure 7. Snapshot of a point along the MD trajectory of a peptide dimer simulation: initial position (a), bonded (b), and dissociated (c) state. Water molecules are omitted for clarity. Visualization by VMD.

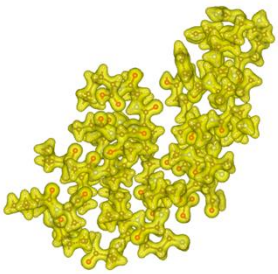
3.2.3 Charge distributions of cell-penetrating peptides

Compared with the results of FFMD simulation for cell-penetrating peptides, DFTB can provide similar structural information and interaction energies, which means DFTB can be used to validate FFMD simulation. However, FFMD is in principle unable to provide electronic structures and electron density information while DFTB provides electronic properties in the same way as DFT. Specifically, electron density distributions are very important because their topology can be used to analyze peptides' propensity to fold.[70-71] Therefore, in order to obtain the electron density information of cell-penetrating peptides, we performed DFTB calculations for a set of 5 cell-penetrating peptides and compute their electronic structures respectively.

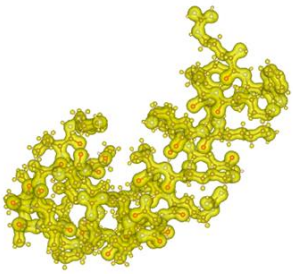
The 5 cell penetrating peptides studied in this section were Lycosin-I, Magainin2, Melittin, CecropinA and LfcinB respectively. FFMD simulations were performed for each kind of peptide at first and 5 different conformers for each peptide were chosen from MD simulations. Then DFTB calculations were performed for each conformers and waveplot program was finally used to generate charge density cube files[140] containing data for molecular orbitals and total electron density distribution based on the results obtained with DFTB+. The cube files for each peptide are listed in Figure 8. The total electron density distribution of each cell-penetrating peptide is shown and they will be used by our collaborators to analyze the propensity to fold in future work.



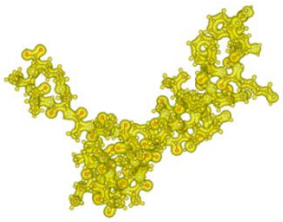
CecropinA-1



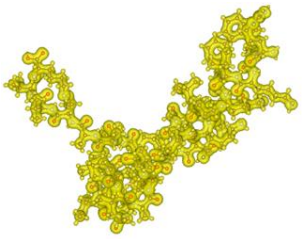
CecropinA-2



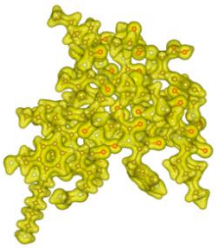
CecropinA-3



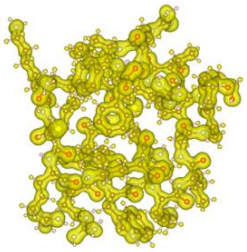
CecropinA-4



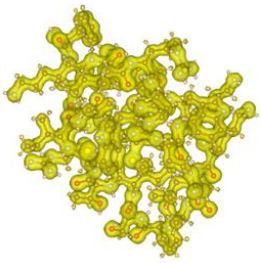
CecropinA-5



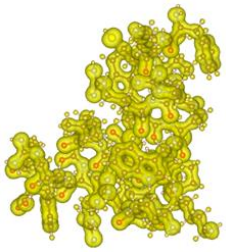
LfcinB-1



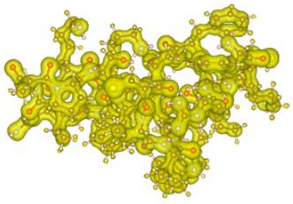
LfcinB-2



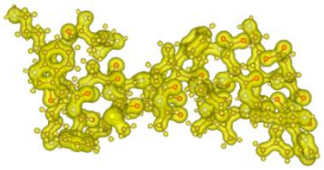
LfcinB-3



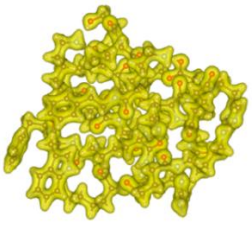
LfcinB-4



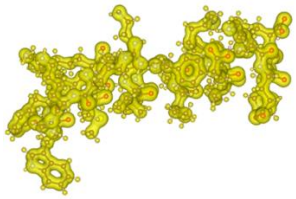
LfcinB-5



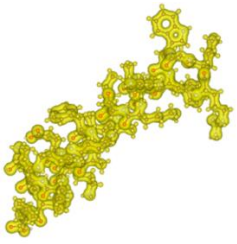
Lyscoin-I-1



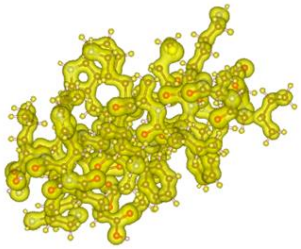
Lycosin-I-2



Lycosin-I-3



Lycosin-I-4



Lycosin-I-5

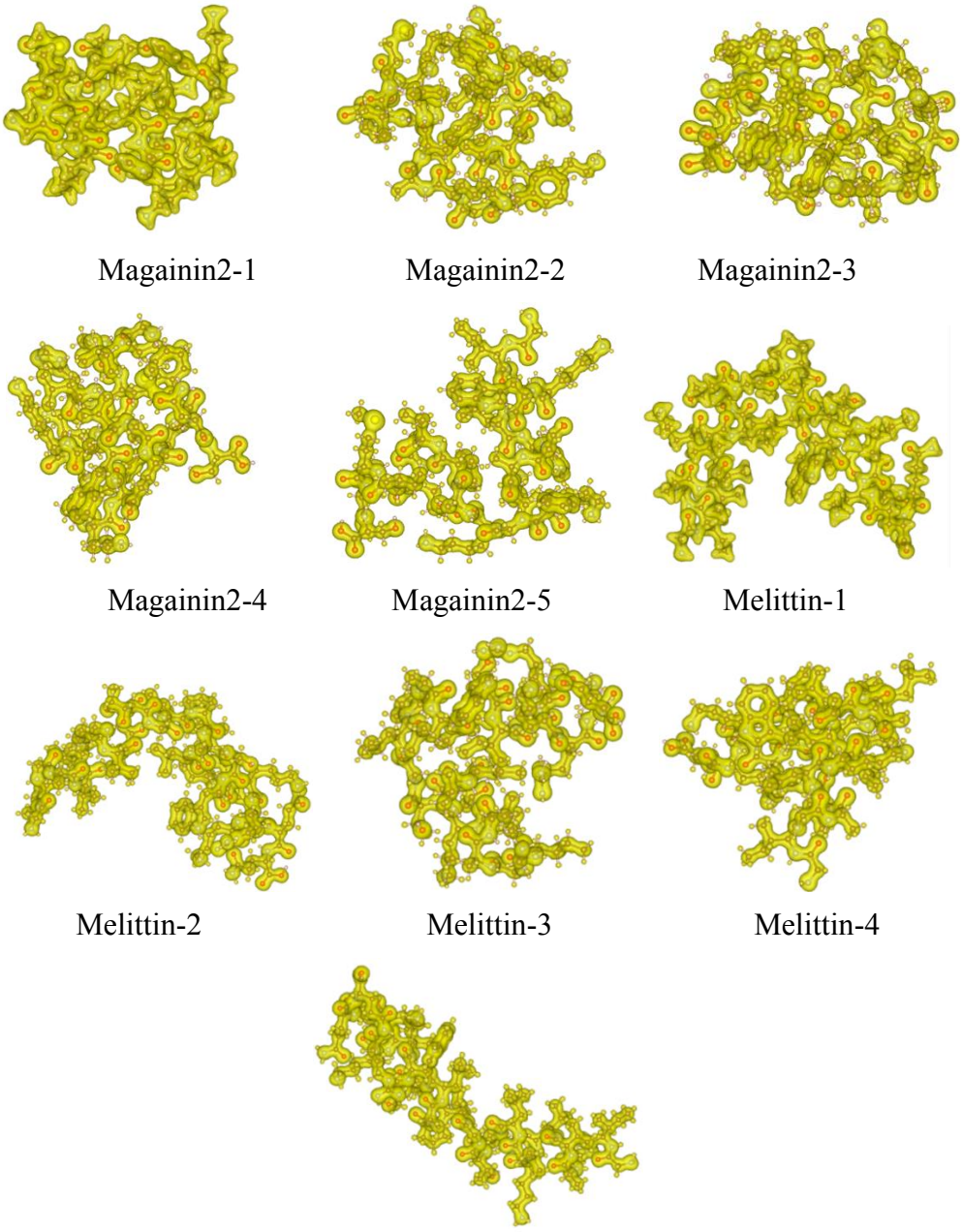


Figure 8. Total electron density distribution of cell-penetrating peptides. Visualization here by SIESTA.

3.3 Interfaces of Arginine and Arginine-Rich Cell-Penetrating Peptide TAT on Anatase TiO_2

3.3.1 Arginine-Titania Interfaces

In former sections, it is shown what kind of accuracy DFTB can achieve for biomolecular simulation, so in this section, we are going to present what kind of accuracy DFTB can achieve for bioinorganic interfaces. Two adsorption regimes were considered: adsorption via the carboxylic group and via the amine groups. The adsorption configurations via the –COOH groups obtained with DFTB and DFT (SIESTA) are shown in Figure 9. The corresponding adsorption energies (E_{ads}) and key bond lengths between the molecule and TiO_2 are given in Table 4. Similar to the molecular calculations, we have checked that adsorption geometries and energies obtained in SIESTA are sufficiently stable with respect to the choice of basis parameters, see Table 4 where results with two bases are given.

Three types of adsorption configurations are considered: a bidentate (BB) and two monodentate (M1 and M2) configurations, shown in Figure 9 (a, d) and Figure 9 (b, c, e, f), respectively. The DFTB structures in Figure 9(a-c) are obtained with the tiorg-0-1* parameterization and are visually similar to those with matsci-0-3 except that with matsci-0-3, in monodentate configurations, the hydrogen atom of the carboxylic group remains on the molecule (as in Figure 4(e-f)) while it dissociates with tiorg-0-1*.

In Table 4, significant differences in adsorption geometries and energies are observed among the three computational schemes. Specifically, the difference in the carboxylic group's hydrogen coordination between tiorg-0-1* and matsci-0-3 (DFTB) is reflected in different distances labeled as $H\text{-}O_{surf}$ in Table 4. Here, the

H-O_{surf} distances obtained in DFTB with matsci-0-3 are closer to the DFT results than those with tiorg-0-1* because with matsci-0-3 the carboxyl's H atoms remained on the molecule, similar to DFT. However, in previous ab initio works on molecules adsorbed on anatase (101) via the COOH group, H was or was not transferred from the carboxyl group to the nearest surface oxygen depending on the scheme employed.[141-142] The H transfer between the carboxylic group's and the surface's O atoms is therefore not a good criterion to accept or reject the parameterization; the O_{mol}-Ti distance is more important in this case.

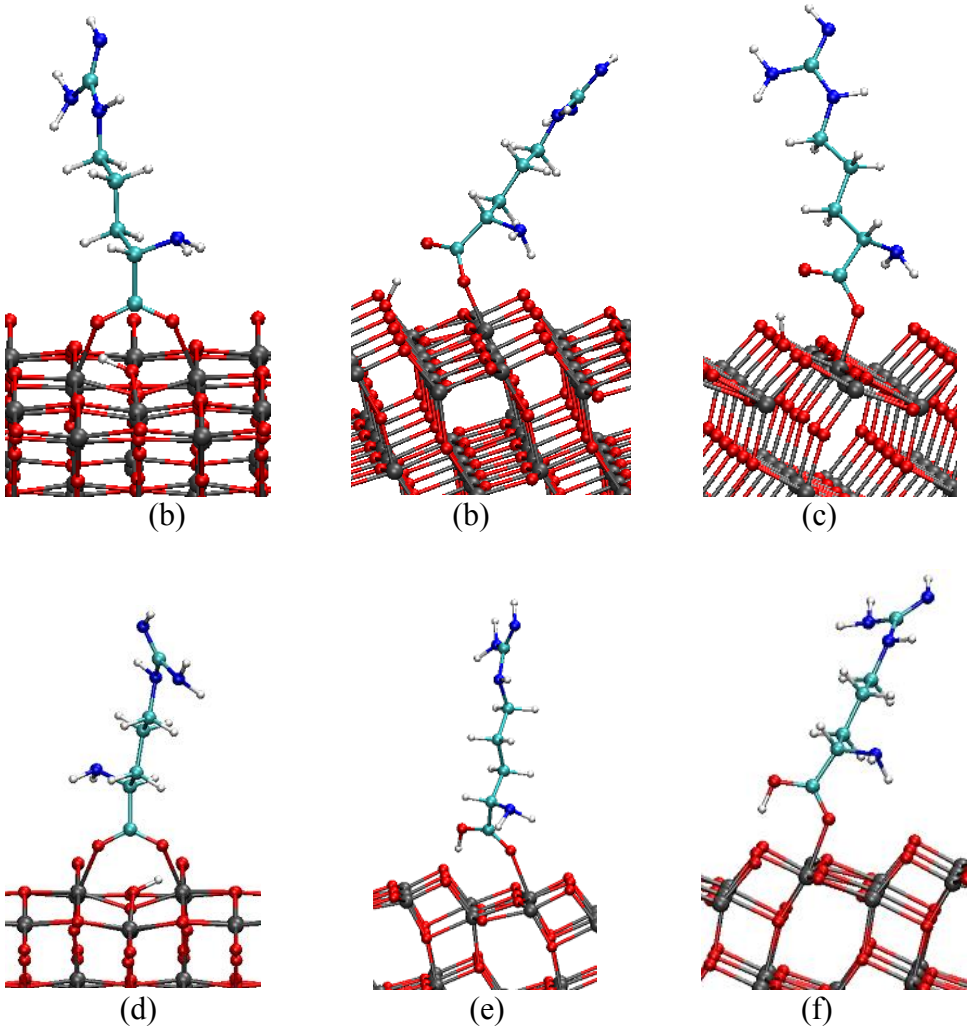


Figure 9. Optimized structures of BB (a), M1 (b) and M2 (c) configurations of Arg on TiO_2 from DFTB tiorg-0-1* and BB (d), M1 (e) and M2 (f) from SIESTA. The atom color code here and elswehere: C, green; O, red; H, light grey; N, blue.; Ti, dark grey. Visualization here by VMD.

The O_{mol}-Ti bond lengths, which determine the electronic coupling between the molecule and the surface, with tiorg-0-1* are closer to the DFT results than those with matsci-0-3. The O_{mol}-Ti bond lengths with tiorg-0-1* straddle those obtained with DFT; specifically, in the BB configuration, they are slightly higher, and in the monodentate configurations, lower.

The adsorption energies obtained with tiorg-0-1* are closer to the DFT results than those obtained with matsci-0-3: E_{ads} is about -1.0, -1.0, and -1.1 eV in BB, M1, and M2 configurations, respectively, vs. the DFT values of -1.0, -1.1, and -1.2 eV, respectively. Specifically, the order of adsorption strength among different configurations is reproduced reasonably well. In contrast, matsci-0-3 results in the strongest adsorption in BB (E_{ads}=-1.8 eV), followed by M1 (-1.0 eV) and M2 (-1.1 eV). The differences are therefore qualitative. Overall, for a single aminoacid, tiorg-0-1* provides more similar adsorption energies and geometries to DFT and will therefore be used below in calculations involving the dipeptide and the TAT peptide.

Figure 10 shows the molecule- and titania- projected density of states (PDOS) obtained for different adsorption configurations with the three computational schemes. Both DFTB with tiorg-0-1* and DFT have molecular HOMO in the bandgap and LUMO in the conduction band, except that in the BB configuration with tiorg-0-1*, the molecular HOMO remains just below the valence band maximum (VBM). In contrast, the molecular HOMO is inside the VB with matsci-0-3 for all adsorption configurations. All three approaches put the molecular LUMO in the conduction band of titania.

Table 4. Adsorption energies E_{ads} (in eV) of Arginine in different configurations on anatase (101) surface of TiO_2 for adsorption via the carboxylic group. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For bidentate configurations, the two bond lengths are O_{mol}-Ti; for monodentate, they are O_{mol}-Ti and H-O_{surf}. The data are for DFTB calculations with two parameterizations (matsci-0-3 and tiorg-0-1*) and for DFT calculations with different choices of DZP basis parameters (using PAO.EnergyShift of 0.001 and 0.002 Ry[114])

System	E_{ads}, eV	O_{mol}-Ti, Å	O_{mol}-Ti/H-O_{surf}, Å
DFTB/Matsci-0-3			
BB	-1.78	2.23	2.23
M1	-0.97	2.25	1.63
M2	-1.08	2.26	1.62
DFTB/Tiorg-0-1*			
BB	-0.97	2.10	2.11
M1	-0.99	1.92	1.00
M2	-1.07	1.98	0.99
SIESTA (0.001Ry)			
BB	-1.02	2.04	2.06
M1	-1.14	2.14	1.51
M2	-1.21	2.12	1.53
SIESTA (0.002Ry)			
BB	-1.00	2.05	2.07
M1	-1.09	2.14	1.50
M2	-1.17	2.13	1.55

From Table 4, we also see that the monodentate configurations obtained with tiorg-0-1* (i.e. those for which the molecular HOMO is in the gap, similar to the DFT results) have smaller O_{mol}-Ti bond lengths than DFT-optimized systems, while the BB configuration (i.e. that for which the molecular HOMO is in the VB) – larger, similar to the overestimation of the O_{mol}-Ti bond lengths with matsci-0-3 (with which the molecular HOMO is in the VB in all configurations). We therefore see that the match between DFT and DFTB results depends on the qualitative match of their respective band structures, specifically, the band alignment between the molecule and the substrate. A similar observation has been made in other comparative DFT-DFTB studies of molecule-titania interfaces[83,86]. We note that the DFT model puts the molecular HOMO only 0.62, 0.69, and 0.75 eV below the CBM for the BB, M1, and M2 configurations, respectively.

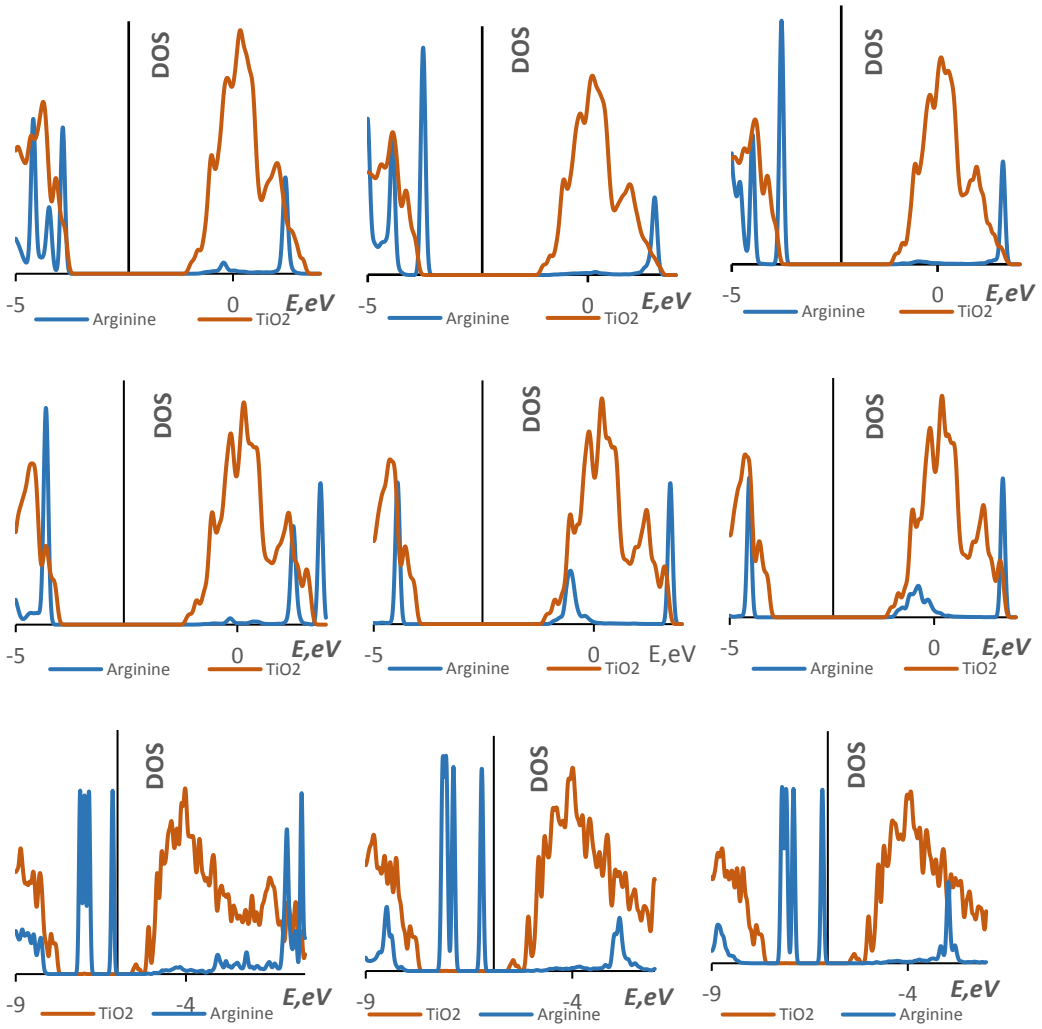


Figure 10. Molecule- and substrate- projected density of states of (left to rigth) BB, M1 and M2 obtained in DFTB with the tiorg-0-1* parameters (top row), matsci-0-3 (middle row), and with DFT (bottom row). The *y* axis crosses the energy axis at the Fermi energy.

Table 5 lists adsorption geometries and energies obtained with the three computational schemes for adsorption via the amine group, for which the geometries are shown in Figure 11. The adsorption can result in molecule's H atoms coordinating to surface oxygen atoms (these configurations are labeled as HO) or in molecule's N atom binding to Ti atom (these configurations are labeled as NTi). In Figure 12 shown are titania- and molecule- projected densities of states obtained with the three approaches. Similar to the adsorption via the carboxylic group, DFTB calculations with the tiorg-0-1* parameter set put the molecular HOMO into the bandgap, while those with matsci-0-3 put the HOMO inside the valence band. The band alignment obtained with tiorg-0-1* matches therefore qualitatively that obtained with DFT. The order of adsorption energies between HO and NTi configurations obtained in DFTB with tiorg-0-1* also matches that obtained with DFT although the magnitude of E_{ads} is much weaker. The order of E_{ads} with matsci-0-3 does not match that of DFT. The conformation assumed by the molecule in the OH configuration with matsci-0-3 is also very different to that obtained with DFT and with tiorg-0-1* (even though both tiorg-0-1* and matsci-0-3 calculations used the same DFT-based initial configuration). Similarly to the adsorption via COOH, there is therefore correspondence between ability to qualitatively reproduce the band alignment and adsorption properties. Similarly to the adsorption via COOH, DFT puts the HOMO very close to the CBM of titania, only 0.23 eV below it, in the HO configuration, while the DOS observed with the NTi configuration shows the HOMO near VBM. The PDOS of the NTi configuration is therefore different (quantitatively) to those of other

adsorption configurations. This is because of the geometry the molecule assumes in this configuration. This was confirmed by the comparison of the DOS of a single arginine molecule in vacuum in its equilibrium configuration and in configurations it assumes during adsorption.

Table 5. Adsorption energies E_{ads} (in eV) of Arginie in different configurations on anatase (101) surface of TiO_2 via the amine groups. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For H-O configurations, the two bond lengths are O_{surf}-H; for NTi configurations, they are N_{mol}-Ti. (The bond length cannot be defined in the same way for the HO configuration obtained with matsci-0-3 due to a very different resulting geometry.)

System	E_{ads}, eV	O_{surf}-H/N_{mol}-Ti, Å	O_{surf}-H/N_{mol}-Ti, Å
DFTB/Tiorg-0-1*			
H-O	0.04	1.97	2.08
N-Ti	-0.22	3.49	3.78
DFTB/Matsci-0-3			
H-O	-2.16	/	/
N-Ti	-1.27	2.27	5.11
SIESTA/0.001Ry			
H-O	-0.26	2.14	2.11
N-Ti	-1.55	2.12	3.89
SIESTA/0.002Ry			
H-O	-0.21	2.14	2.09
N-Ti	-1.65	2.08	4.08

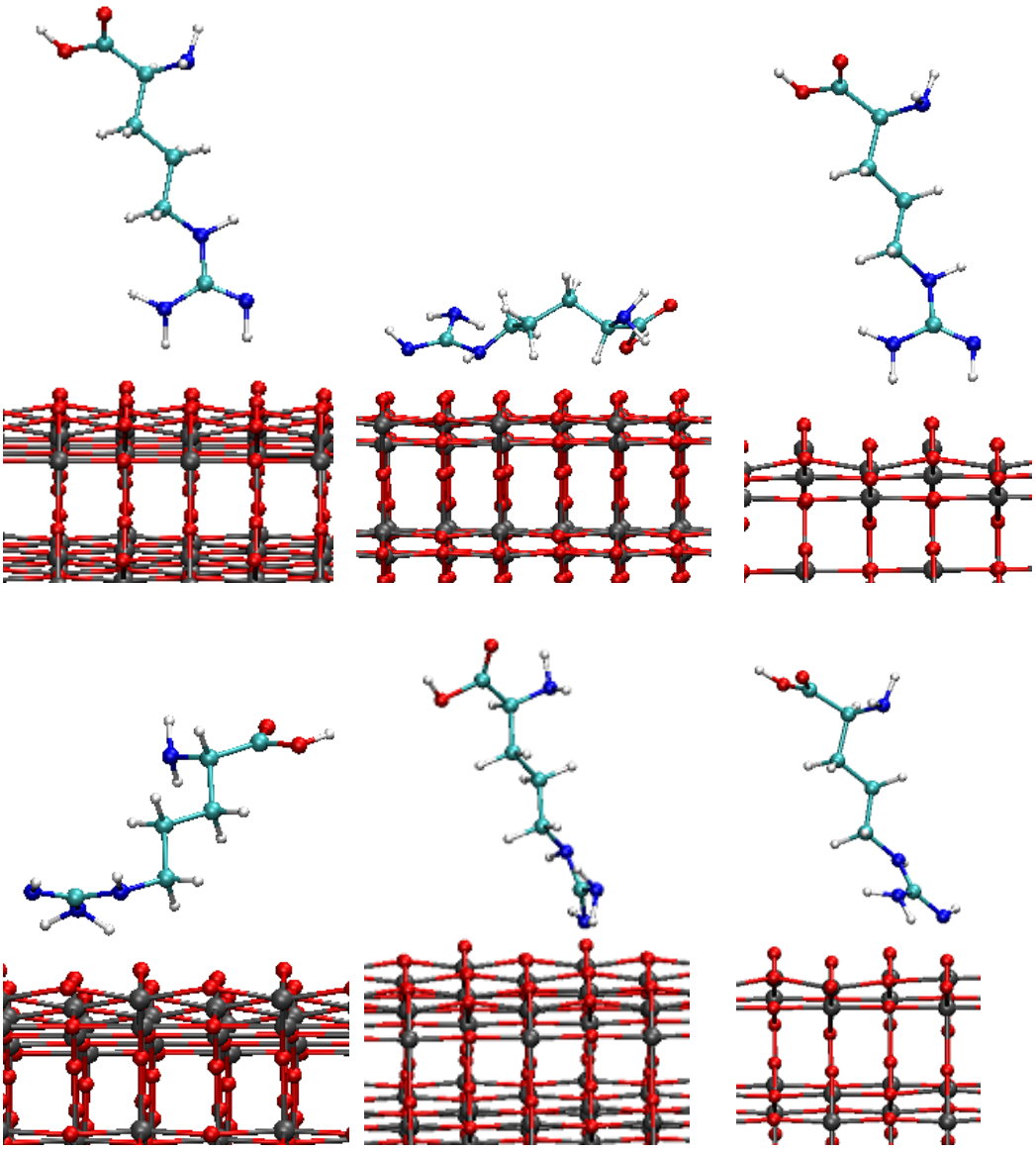


Figure 11. Optimized structure of Arg adsorbed on TiO_2 through the amine groups: configurations defined in the text as HO (top row) and NTi (bottom row) obtained with tiorg-0-1* (left), matsci-0-3 (middle) and with DFT (right). Visualization here by VMD.

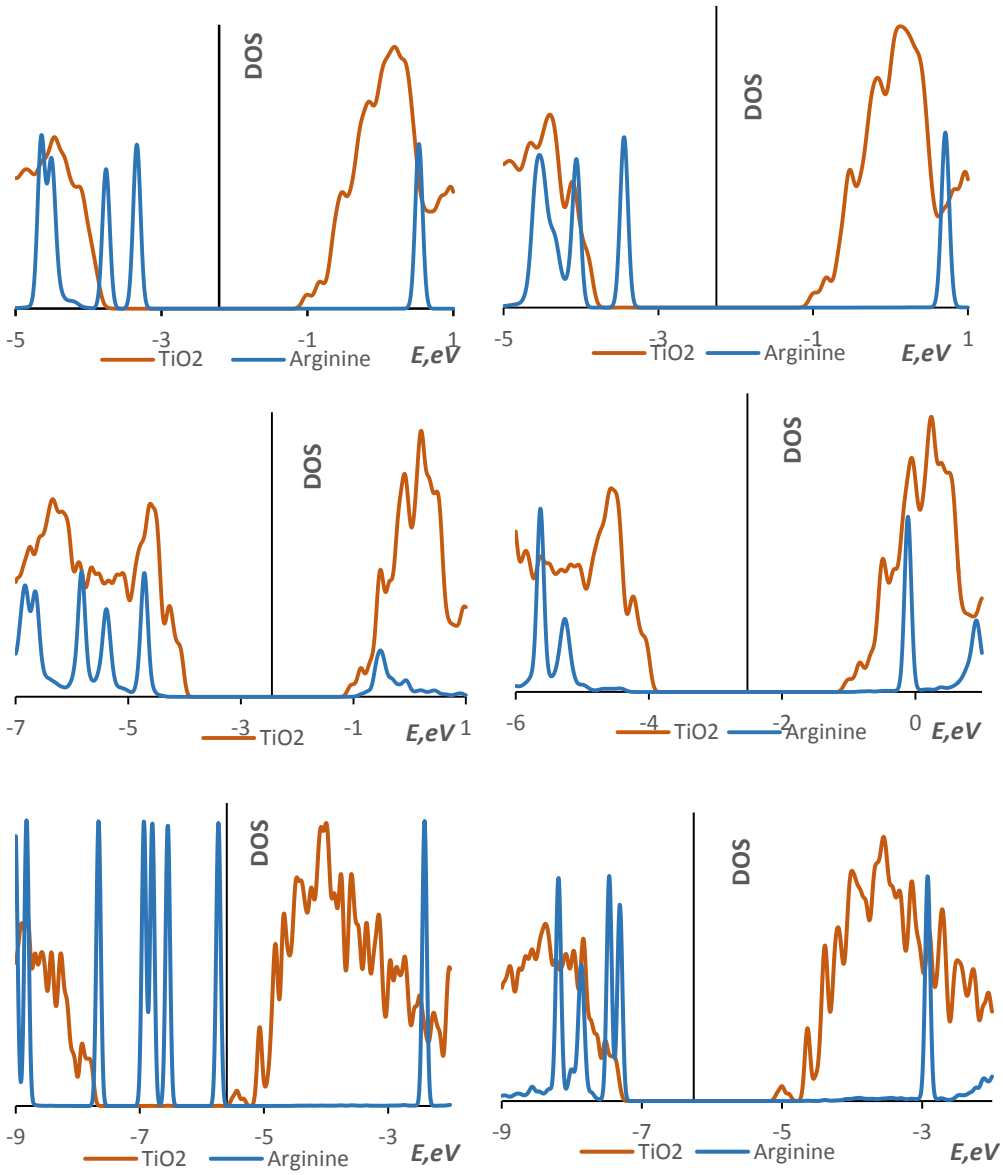


Figure 12. Molecule- and substrate- projected density of states of HO (left) and NTi (right) configurations of Arg on TiO_2 obtained in DFTB with the tiorg-0-1* parameters (top row) and matsci-0-3 DFT (middle row) and with DFT (bottom row). The *y* axis crosses the energy axis at the Fermi energy. Visualization here by VMD.

3.3.2 Arginine Dipeptide-Titania Interfaces

We have computed the adsorption of Arginine dipeptide on titania with DFT and DFTB. Only adsorption via the COOH group was computed and for DFTB, only the tiorg-0-1* parameterization was used, as only at this juncture qualitative agreement was observed for Arginine adsorption considered above. This is sufficient for our purposes. The adsorption geometries obtained with both methods are shown in Figure 13; adsorption energies and key bond lengths are listed in Table 6. The level of agreement between DFT and DFTB is similar to that obtained for monomer adsorption considered above, in terms of both adsorption energies and bond lengths at the interface. The order of adsorption energies is somewhat violated in DFTB in that in predicts the weakest adsorption in M1 configuration, but the differences on the order of 0.1 eV are similar to what was seen with the monomer.

Figure 14 shown the partial densities of states for all adsorption configurations obtained with the two methods. Similarly to the monomer case, in DFTB there is a qualitative agreement in the band alignment in that the molecular HOMO enters the band gap (about 0.1 and 0.2 eV above the VBM for M1 and M2, respectively) in the monodentate configurations, while in the BB regime it remains slightly lower than the VBM (by 0.03 eV). We note that with DFT, the HOMO becomes very close to the conduction band minimum, lower than CBM by only 0.23, 0.25, and 0.24 eV in BB, M1, and M2 configurations, respectively.

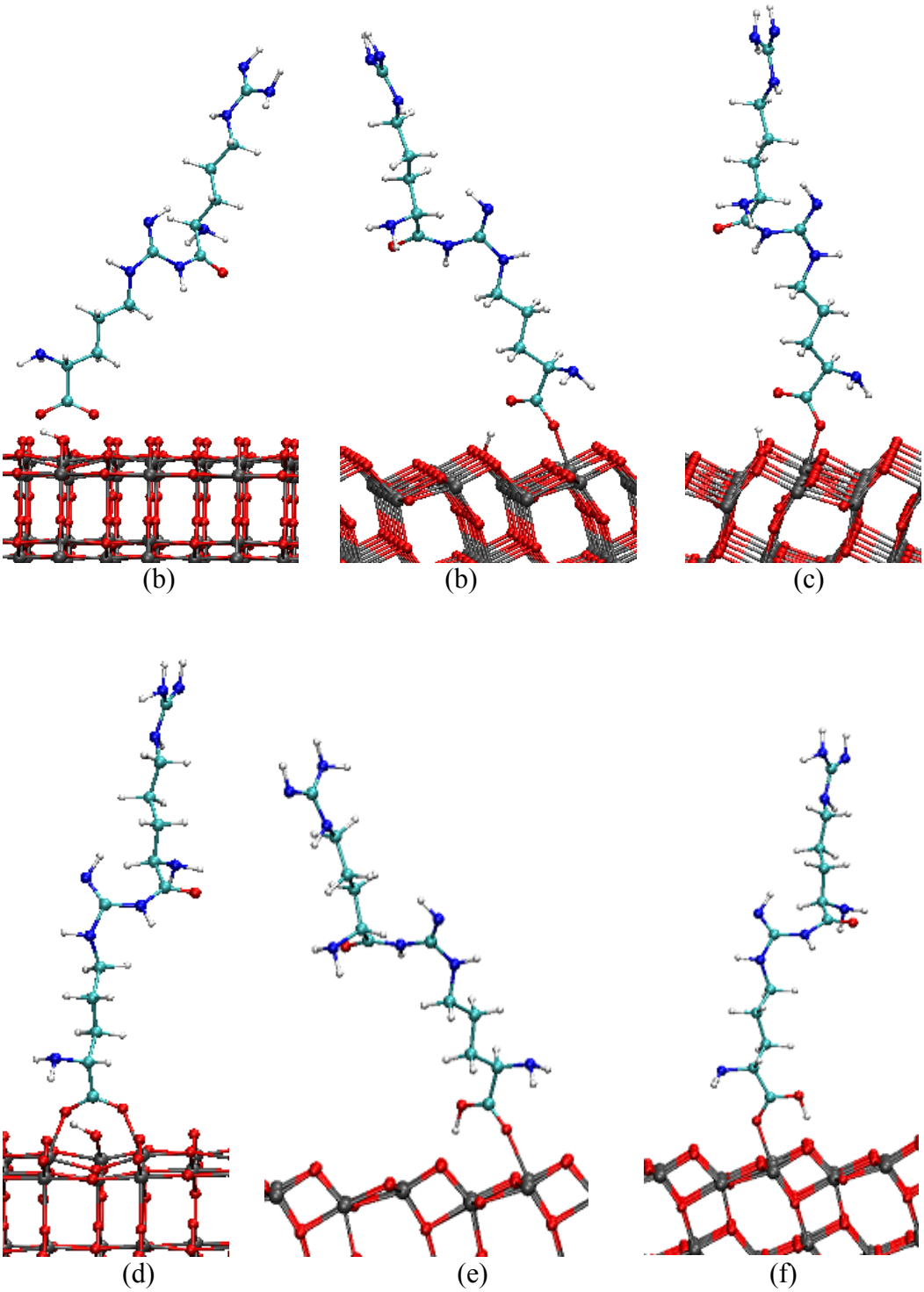


Figure 13. Optimized structures of BB (a), M1 (b) and M2 (c) for arginine dipeptide adsorbed on anatase TiO_2 obtained with DFTB tiorg-0-1* and BB (d), M1 (e) and M2 (f) from SIESTA. Visualization here by VMD.

Table 6. Adsorption energies E_{ads} (in eV) of Arginine dipeptide in different configurations on anatase (101) surface of TiO_2. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For bidentate configurations, the two bond lengths are O_{mol}-Ti; for monodentate, they are O_{mol}-Ti and H-$O_{surf.}$

System	E_{ads}, eV	O_{mol}-Ti, Å	O_{mol}-Ti/H-O_{surf},
DFTB/Tiorg-0-1*			
BB	-1.01	2.10	2.11
M1	-0.92	1.99	1.00
M2	-1.13	1.99	0.99
DFT/PBE			
BB	-0.88	2.07	2.05
M1	-1.13	2.13	1.54
M2	-1.07	2.18	1.47

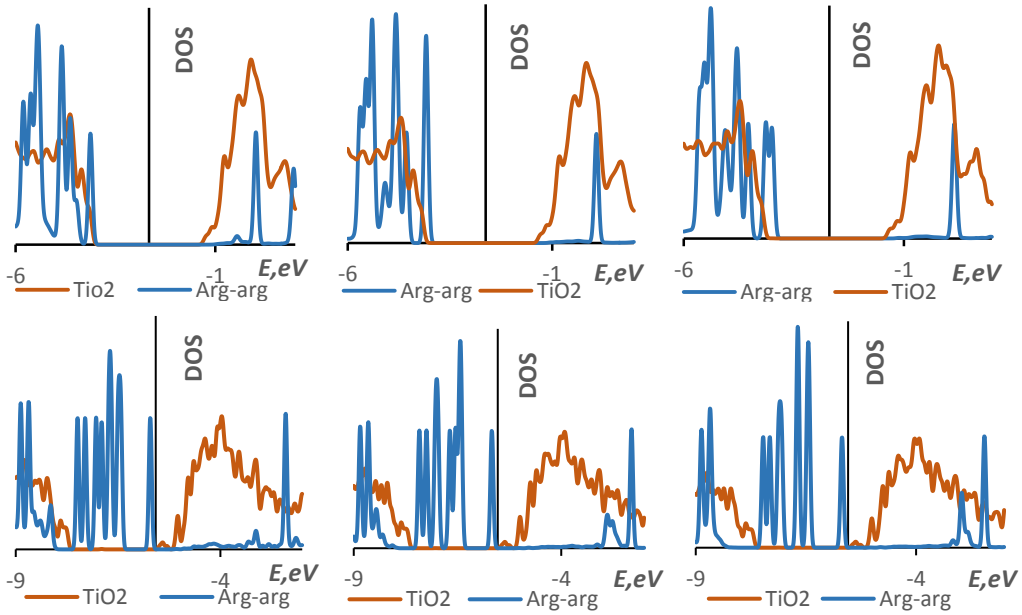


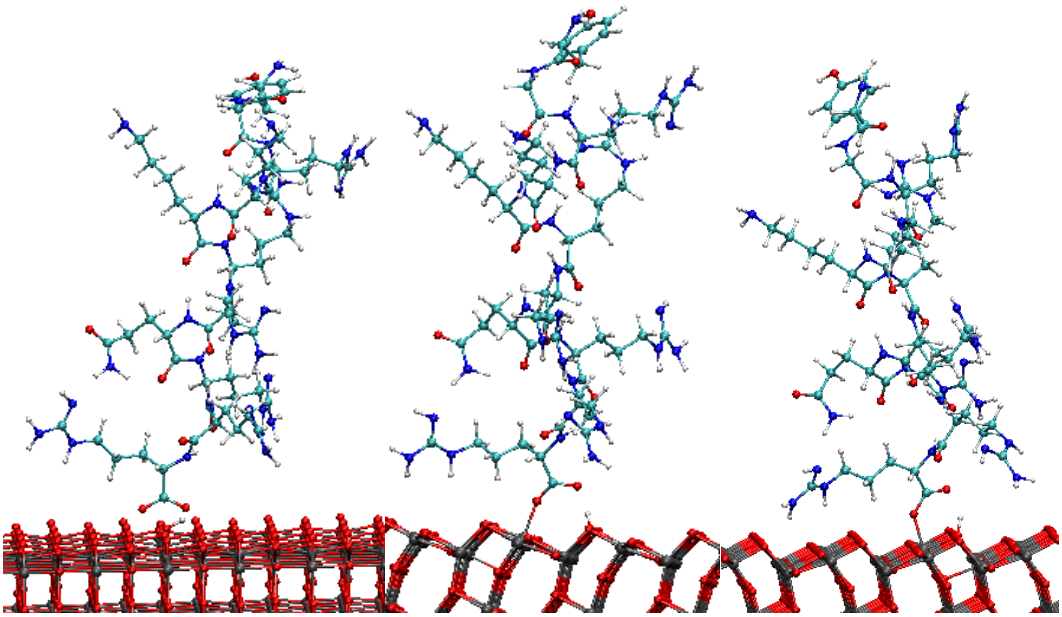
Figure 14. Molecule- and substrate- projected density of states of BB, M1 and M2 configurations of Arg_2 on TiO_2 obtained in DFTB with the tiorg-0-1* setup (top row) and with DFT (bottom row). The *y* axis crosses the energy axis at the Fermi energy.

3.3.3 TAT-Titania Interface

We have computed the adsorption of TAT on TiO_2 in DFTB (with the tiorg-0-1* parameters) via a –COOH group. The adsorption geometries are shown in Figure 15 and adsorption energies are listed in Table 7. Due to a large number of angular degrees of freedom in the peptide with shallow potential, the force convergence of 0.02 eV/Å was not achieved here; we have confirmed that the energy and the DOS were converged. Due to the large size of the molecule, other amino acids contribute to binding to the surface, as can be seen in the figure, which contributes to a strengthening of the adsorption compared to Arg and Arg_2. The key O_{mol}-Ti bond lengths are similar to those obtained with the small molecules. What interests us here is the band alignment between the molecule and the surface. In Figure 15(a-c), we show the PDOS for these adsorbate systems. Similar to the case of individual peptide, the molecular HOMO is in the bandgap and LUMO in the conduction band. To compare this band alignment to that obtained with DFT/PBE, due to the prohibitive computational cost of the TAT/TiO_2 system (models shown in Figure 15 have more than 2500 atoms), we compare the DOS of the molecule and the anatase (101) surface computed individually and aligned based on the alignment obtained with Arg and Arg_2 and the known destabilization of TAT's HOMO vs the HOMO of Arg and Arg_2. This band alignment is shown in the last panel of Figure 16. As was expected from the analysis of Figures 4 and 10 together, the molecular HOMO enters the conduction band. Indeed, the HOMO of free TAT is 0.91 eV above that of Arg (with PBE, see Figure 4), and the HOMO of Arg on the anatase (101) surface is only about

0.6-0.7 eV lower than the CBM of the oxide (see Figure 10). The HOMO of free TAT is about 0.76 eV above that of Arg_2 (with PBE, see Figure 4), and the HOMO of Arg_2 on the anatase (101) surface is only about 0.24 eV lower than the CBM of the oxide (see Figure 14). It is therefore expected that with DFT/PBE, TAT's HOMO would enter the CB and this is exactly what is seen in Figure 16(d). This band alignment would have significant practically important consequences: the molecule would be oxidized.

The electronic structure resulting from DFT/PBE is therefore qualitatively different from that resulting from DFTB. However, in this case it is not the DFTB but the DFT calculation which must be wrong. The high quality DFT calculations of TAT in section 1.2 (in Gaussian 09 and using the B3LYP functional) predict a HOMO level of about -5.0 eV, as can be seen in Figure 4.



(b) (b) (c)

Figure 15. Optimized structures of BB (a), M1 (b) and M2 (c) configurations of TAT adsorbed on anatase TiO_2 obtained with DFTB using tiorg-0-1* parameter set. Visualization here by VMD.

While computational studies using hybrid functionals widely vary in their estimates of the CBM of anatase and its clean (101) surface[143-145], the latest high-quality experimental[145-146] and hybrid functional based DFT calculations using the HSE06[147-148] functional put the CBM of the anatase (101) surface at -5.10 eV.[145] [149] We computed the HOMO of TAT with HSE06 (in Gaussian 09) and obtained a value of -5.20 eV. That is to say, the HOMO of TAT should remain below the CBM of anatase (101). The DFTB achieves that, in part because the DFTB parameterization was done in a way that effectively reproduced the bandgap of TiO_2.[93] The band alignment is expected to be correctly reproduced with an appropriate hybrid functional, but, as explained in the Introduction, such calculations are not practical for modeling of interfaces involving large

biomolecules, due to the high CPU cost of computing exact exchange. The practical DFT approach for large interfaces remains that using the GGA approximation. In Ref. [86], we showed that DFTB with matsci-0-3 and tiorg-0-1* parameterizations does not produce a qualitatively correct band alignment of a *small* dye adsorbed on the anatase (101) surface of titania, while DFT/PBE does. It was not surprising to find a system where DFTB fails; after all, it is an approximation to DFT which is not expected to perform well in all cases.

Table 7. Adsorption energies E_{ads} (in eV) of TAT in different configurations on anatase (101) surface of TiO_2. The bond length for bonding between the molecule's and surface atoms are also given (in Å). For bidentate configurations, the two bond lengths are O_{mol}-Ti; for monodentate, they are O_{mol}-Ti and H-O_{surf}

System	E_{ads}, eV	O_{mol}-Ti, Å	O_{mol}-Ti/H-O_{surf}, Å
DFTB/Tiorg-0-1*			
BB	-1.28	2.11	2.11
M1	-0.99	1.98	1.00
M2	-1.37	1.99	0.99

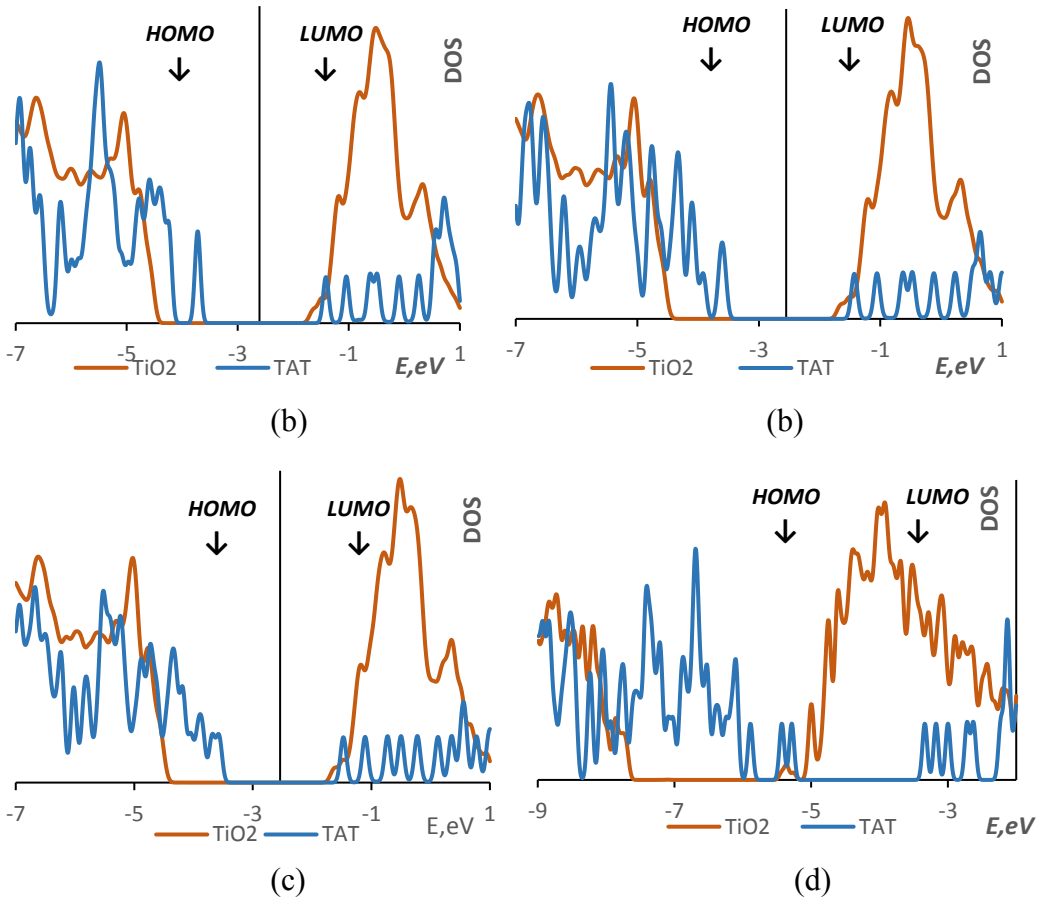


Figure 16. (a-c) Molecule- and substrate- projected density of states of BB, M1 and M2 configurations of TAT on TiO_2 obtained in DFTB with the tiorg-0-1* parameters. The *y* axis crosses the energy axis at the Fermi energy. (d) The simulated band alignment between the TAT and the (101) anatase surface. Approximate positions of molecular HOMO and LUMO energies are indicated with arrows.

Here, we have a somewhat less expected situation where DFT/PBE fails, but DFTB does not. This is specific to the adsorption of large molecules (i.e. all biomolecules) where HOMO is closer to the CBM. This makes the system more sensitive to the description of the band gap of the semiconductor substrate, which is known to be erroneous with GGA[144,150-153]. A practically important conclusion of the above is that DFTB is a good approach to model biomolecule-

semiconductor interfaces, and that is not only due to the CPU cost advantage of three orders of magnitude but also because it can effectively, via parameterization, achieve qualitatively correct band alignment which may not be possible with GGA.

4 Conclusions

We have conducted comparative DFTB and DFT studies on biomolecules and bioinorganic interfaces as well as a comparative DFTB and Force Field MD study on cell-penetrating peptides. The accuracy of DFTB is established by the modeling of biomolecules and it is found that DFTB can bring ab initio accuracy to biomolecular simulation which usually calls for Force Field MD. Meanwhile, through the simulation of bioinorganic interfaces, we highlight the utility of the DFTB method for the modeling of bioinorganic interfaces not only from the CPU cost perspective but also from the accuracy point of view.

Specifically, we started from the modeling of small biomolecules with both DFT and DFTB methods. Compared with the DFT results, DFTB shows in general good agreements in structural information and conformation energies which means DFTB can reach relatively similar ab initio accuracy in small biomolecules simulation. Therefore, in the simulation of cell-penetrating peptides, we bring in DFTB calculations in order to testify its applicability for large biomolecules simulation. FFMD is compared to DFTB simulation and it is found that their results match qualitatively well, which means DFTB can be feasible and trustworthy to bring ab initio accuracy to large biomolecules simulation. We then proceeded to compute FFMD simulation of Lycosin-I in solution, which helped

corroborate the weak clustering behavior observed through experiments. We also computed total charge density distribution of 5 cell-penetrating peptides, which is useful for the analysis of the propensity to fold.

Finally, we have conducted a comparative computational study of interactions of biomolecules with an oxide surface. Such mixed bioinorganic systems present difficulties for ab initio modelling due to their sheer size. Specifically, the large size of biological molecules necessitates the use of large slabs to model the substrates, which makes more accurate ab initio schemes impractical. For example, the TAT/TiO_2 model used here has more than 2500 atoms. For most laboratories, the only practical and relatively accurate ab initio approach to model such system remains DFT with the GGA approximation. Even with a GGA functional (i.e. without the additional cost of exact exchange), the CPU cost is substantial. The DFTB method – an approximate density functional based method - therefore looks very promising for the modeling of such system, offering a three orders of magnitude speedup. Previous works reported both successes and failures of DFTB with specific parameterizations for the modeling of organic-inorganic interfaces.

In this work, we have compared the performance of DFT to DFTB for the modeling of adsorption on the widely use substrate TiO_2 of *both a small* biomolecule (an aminoacid, Arg) and a *large* biomolecule (the TAT cell-penetrating peptide). We have found that the quality of DFTB calculations depends on their ability to reproduce qualitatively the correct band alignment between the molecule and the surface, which in this case signifies the molecular

HOMO in the titania bandgap and LUMO in the conduction band. It is the qualitative band alignment (rather than e.g. where exactly in the gap the HOMO is located) which determines possible electron donation between the molecule and the substrate as well as hybridizations (e.g. a molecular HOMO in the VB or LUMO in the CB will hybridize much more significantly with the titania states than an orbital in the gap). These are expected to directly influence the energies and geometries via band energies and geometries which are optimal for a given bandstructure. Specifically, for systems and with a parameterization where a correct band alignment is obtained, we obtained a decent agreement in adsorption energies and geometries between DFT and DFTB.

Importantly, by comparing absorption of biomolecules of different size – an aminoacid, a dipeptide, and a real-sized peptide – we have discovered a seemingly counterintuitive phenomenon whereby it is the GGA DFT that fails while the DFTB is at least qualitatively correct. Specifically, due to a large destabilization of the HOMO of a large peptide vs a single aminoacid combined with an underestimation of the titania bandgap typical of GGA functionals, DFT is predicted to result in HOMO entering the conduction band, which would effectively ionize the molecule. DFTB, on the other hand, is able to reproduce a correct bandstructure precisely because it is an approximate method with which one can *effectively* reproduce the correct band gap. This effect is specific to large molecules such as biomolecules and may not have been appreciated due to a tendency to perform ab initio modeling on simplified, abridged systems. To get a

correct band alignment for such a bioinorganic interface with DFT would require the use of (range-separated) hybrid functionals, which is prohibitive for most labs.

5 References

(1) Hanessian, S., Reflections on the Total Synthesis of Natural Products: Art, Craft, Logic, and the Chiron Approach. In *Pure Appl. Chem.*, 1993; Vol. 65, p 1189.

(2) Bulinski, J. C., Peptide Antibodies: New Tools for Cell Biology. In *Int. Rev. Cytol.*, G.H. Bourne, K. W. J.; Friedlander, M., Eds. Academic Press: 1986; Vol. Volume 103, pp 281-302.

(3) Liu, Z.; Deng, M.; Xiang, J.; Ma, H.; Hu, W.; Zhao, Y.; Li, D. W. C.; Liang, S., A Novel Spider Peptide Toxin Suppresses Tumor Growth Through Dual Signaling Pathways. *Current Molecular Medicine* **2012,** *12* (10), 1350-1360.

(4) Busseron, E.; Ruff, Y.; Moulin, E.; Giuseppone, N., Supramolecular Self-Assemblies as Functional Nanomaterials. *Nanoscale* **2013,** *5* (16), 7098-7140.

(5) Shiba, K., Exploitation of Peptide Motif Sequences and Their Use in Nanobiotechnology. *Curr. Opin. Biotechnol.* **2010,** *21* (4), 412-425.

(6) Jones, F. H., Teeth and Bones: Applications of Surface Science to Dental Materials and Related Biomaterials. *Surf. Sci. Rep.* **2001,** *42* (3–5), 75-205.

(7) Cheng, Y.; C. Samia, A.; Meyers, J. D.; Panagopoulos, I.; Fei, B.; Burda, C., Highly Efficient Drug Delivery with Gold Nanoparticle Vectors for in Vivo Photodynamic Therapy of Cancer. *J. Am. Chem. Soc.* **2008,** *130* (32), 10643-10647.

(8) Cheng, Y.; Meyers, J. D.; Broome, A.-M.; Kenney, M. E.; Basilion, J. P.; Burda, C., Deep Penetration of a PDT Drug into Tumors by Noncovalent Drug-Gold Nanoparticle Conjugates. *J. Am. Chem. Soc.* **2011,** *133* (8), 2583-2591.

(9) Shiang, Y.-C.; Huang, C.-C.; Chang, H.-T., Gold Nanodot-based Luminescent Sensor for the Detection of Hydrogen Peroxide and Glucose. *Chem. Commun.* **2009,** (23), 3437-3439.

(10) Wang, Y.; Wang, Y.; Zhou, F.; Kim, P.; Xia, Y., Protein-Protected Au Clusters as a New Class of Nanoscale Biosensor for Label-Free Fluorescence Detection of Proteases. *Small* **2012,** *8* (24), 3769-3773.

(11) Huang, C.-C.; Chiang, C.-K.; Lin, Z.-H.; Lee, K.-H.; Chang, H.-T., Bioconjugated Gold Nanodots and Nanoparticles for Protein Assays Based on Photoluminescence Quenching. *Anal. Chem.* **2008,** *80* (5), 1497-1504.

(12) Chen, C.-T.; Chen, W.-J.; Liu, C.-Z.; Chang, L.-Y.; Chen, Y.-C., Glutathione-bound Gold Nanoclusters for Selective-binding and Detection of Glutathione S-transferase-fusion Proteins from Cell Lysates. *Chem. Commun.* **2009,** (48), 7515-7517.

(13) Triulzi, R. C.; Micic, M.; Giordani, S.; Serry, M.; Chiou, W.-A.; Leblanc, R. M., Immunoasssay Based on the Antibody-conjugated PAMAM-dendrimer-gold Quantum Dot Complex. *Chem. Commun.* **2006,** (48), 5068-5070.

(14) Shiang, Y.-C.; Lin, C.-A.; Huang, C.-C.; Chang, H.-T., Protein A-conjugated Luminescent Gold Nanodots as a Label-free Assay for Immunoglobulin G in Plasma. *Analyst* **2011,** *136* (6), 1177-1182.

(15) Li, H.; Rothberg, L., Colorimetric Detection of DNA Sequences Based on Electrostatic Interactions with Unmodified Gold Nanoparticles. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (39), 14036-14039.

(16) Li; Rothberg, L. J., Label-Free Colorimetric Detection of Specific Sequences in Genomic DNA Amplified by the Polymerase Chain Reaction. *J. Am. Chem. Soc.* **2004,** *126* (35), 10958-10961.
(17) Fleming, G. J.; Adib, K.; Rodriguez, J. A.; Barteau, M. A.; Idriss, H., Proline Adsorption on TiO_2(1 1 0) Single Crystal Surface: A Study by High Resolution Photoelectron Spectroscopy. *Surf. Sci.* **2007,** *601* (24), 5726-5731.
(18) Nel, A.; Xia, T.; Mädler, L.; Li, N., Toxic Potential of Materials at the Nanolevel. *Science* **2006,** *311* (5761), 622-627.
(19) Xie, Q.; Zhao, Y.; Chen, X.; Liu, H.; Evans, D. G.; Yang, W., Nanosheet-based Titania Microspheres with Hollow Core-shell Structure Encapsulating Horseradish Peroxidase for a Mediator-free Biosensor. *Biomaterials* **2011,** *32* (27), 6588-6594.
(20) Yang, Z.; Zong, X.; Ye, Z.; Zhao, B.; Wang, Q.; Wang, P., The Application of Complex Multiple Forklike ZnO Nanostructures to Rapid and Ultrahigh Sensitive Hydrogen Peroxide Biosensors. *Biomaterials* **2010,** *31* (29), 7534-7541.
(21) Buonocore, F.; Arcangeli, C.; Gala, F.; Zollo, G.; Celino, M., Adsorption of Modified Arg, Lys, Asp, and Gln to Dry and Hydrated ZnO Surface: A Density Functional Theory Study. *J. Phys. Chem. B* **2015,** *119* (35), 11791-11797.
(22) Agosta, L.; Zollo, G.; Arcangeli, C.; Buonocore, F.; Gala, F.; Celino, M., Water Driven Adsorption of Amino Acids on the (101) Anatase TiO_2 Surface: An Ab Initio Study. *Phys. Chem. Chem. Phys.* **2015,** *17* (3), 1556-1561.
(23) Thomas, A. G.; Jackman, M. J.; Wagstaffe, M.; Radtke, H.; Syres, K.; Adell, J.; Lévy, A.; Martsinovich, N., Adsorption Studies of p-Aminobenzoic Acid on the Anatase TiO_2(101) Surface. *Langmuir* **2014,** *30* (41), 12306-12314.
(24) Zhao, X.; Zhao, R. G.; Yang, W. S., Scanning Tunneling Microscopy Investigation of l-Lysine Adsorbed on Cu(001). *Langmuir* **2000,** *16* (25), 9812-9818.
(25) Barlow, S. M.; Kitching, K. J.; Haq, S.; Richardson, N. V., A Study of Glycine Adsorption on a Cu{110} Surface Using Reflection Absorption Infrared Spectroscopy. *Surf. Sci.* **1998,** *401* (3), 322-335.
(26) Chen, Load-Bearing Capacity of Masonry Arch Bridges Strengthened with Fibre Reinforced Polymer Composites. *Adv. Struct. Eng*, **2002,** *5* (1), 37-44.
(27) Barlow, S. M.; Raval, R., Complex Organic Molecules at Metal Surfaces: Bonding, Organisation and Chirality. *Surf. Sci. Rep.* **2003,** *50* (6–8), 201-341.
(28) Ihs, A.; Liedberg, B.; Uvdal, K.; Törnkvist, C.; Bodö, P.; Lundström, I., Infrared and Photoelectron Spectroscopy of Amino Acids on Copper: Glycine, l-alanine and β-alanine. *J. Colloid Interface Sci.* **1990,** *140* (1), 192-206.
(29) Williams, J.; Haq, S.; Raval, R., The Bonding and Orientation of the Amino Acid l-alanine on Cu{110} Determined by RAIRS. *Surf. Sci.* **1996,** *368* (1–3), 303-309.
(30) Löfgren, P.; Krozer, A.; Lausmaa, J.; Kasemo, B., Glycine on Pt(111): a TDS and XPS study. *Surf. Sci.* **1997,** *370* (2–3), 277-292.
(31) Cavalleri, O.; Gonella, G.; Terreni, S.; Vignolo, M.; Floreano, L.; Morgante, A.; Canepa, M.; Rolandi, R., High Resolution X-ray Photoelectron

Spectroscopy of l-cysteine Self-assembled Films. *Phys. Chem. Chem. Phys.* **2004,** *6* (15), 4042-4046.
(32) Nicklin, R. E. J.; Cornish, A.; Shavorskiy, A.; Baldanza, S.; Schulte, K.; Liu, Z.; Bennett, R. A.; Held, G., Surface Chemistry of Alanine on Ni{111}. *J. Phys. Chem. C* **2015,** *119* (47), 26566-26574.
(33) Monticelli, L.; Tieleman, D. P., Force Fields for Classical Molecular Dynamics. *Methods Mol. Biol.* **2013,** *924*, 197-213.
(34) Frenkel, D.; Smit, B., *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press: San Diego, 1996.
(35) Leach, A. R., *Molecular Modelling: Principles and Applications*. Pearson Education: Singapore;Harlow, England;, 2001; Vol. 2nd.
(36) Ramachandran, K. I.; Deepa, G.; Namboori, K., *Computational Chemistry and Molecular Modeling: Principles and Applications*. Springer: Berlin, 2008; p 1-397.
(37) Alder, B. J.; Wainwright, T. E., Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959,** *31* (2), 459-466.
(38) Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Phys. Rev.* **1964,** *136* (3B), B864-B871.
(39) Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965,** *140* (4A), A1133-A1138.
(40) Møller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934,** *46* (7), 618-622.
(41) Head-Gordon, M.; Pople, J. A.; Frisch, M. J., MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988,** *153* (6), 503-506.
(42) Kümmel, H. G., A Biography of the Coupled Cluster Method. *Int. J. Mod. Phys.* **2003,** *17* (28), 5311-5325.
(43) Shavitt, I.; Bartlett, R. J., *Many-body Methods in Chemistry and Physics: MBPT and Coupled-cluster Theory*. Cambridge University Press: Cambridge, 2009.
(44) Čížek, J., On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966,** *45* (11), 4256-4266.
(45) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C., Introducing ONETEP: Linear-scaling Density Functional Simulations on Parallel Computers. *J. Chem. Phys.* **2005,** *122* (8), 084119.
(46) Bowler, D. R.; Miyazaki, T.; Gillan, M. J., Recent Progress in Linear Scaling Ab Initio Electronic Structure Techniques. *J. Phys.: Condens. Matter* **2002,** *14* (11), 2781.
(47) Bowler, D. R.; Choudhury, R.; Gillan, M. J.; Miyazaki, T., Recent Progress with Large-Scale Ab Initio Calculations: the CONQUEST Code. *Phys. Status Solidi B* **2006,** *243* (5), 989-1000.
(48) Patel, S.; Brooks, C. L., CHARMM Fluctuating Charge Force Field for Proteins: I Parameterization and Application to Bulk Organic Liquid Simulations. *J. Comput. Chem.* **2004,** *25* (1), 1-16.
(49) Patel, S.; Mackerell, A. D.; Brooks, C. L., CHARMM Fluctuating Charge Force Field for Proteins: II Protein/solvent Properties from Molecular Dynamics

Simulations Using a Nonadditive Electrostatic Model. *J. Comput. Chem.* **2004,** *25* (12), 1504-1514.
(50) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2005,** *1* (1), 153-168.
(51) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B., Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010,** *6* (3), 774-786.
(52) Ponder, J. W.; Case, D. A., Force Fields for Protein Simulations. In *Adv. Protein Chem.*, Academic Press: 2003; Vol. 66, pp 27-85.
(53) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W., Modeling Electrostatic Effects in Proteins. *Biochim. Biophys. Acta, Proteins Proteomics* **2006,** *1764* (11), 1647-1676.
(54) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.
(55) Ditzler, M. A.; Otyepka, M.; Šponer, J.; Walter, N. G., Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In. *Acc. Chem. Res.* **2010,** *43* (1), 40-47.
(56) Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996,** *77* (18), 3865-3868.
(57) Vitalini, F.; Mey, A. S. J. S.; Noé, F.; Keller, B. G., Dynamic Properties of Force Fields. *J. Chem. Phys.* **2015,** *142* (8), 084101.
(58) Li, G.; Wang, S.-W.; Rosenthal, C.; Rabitz, H., High Dimensional Model Representations Generated from Low Dimensional Data Samples. I. Mp-Cut-HDMR. *J. Math. Chem.* **2001,** *30* (1), 1-30.
(59) Manzhos, S.; Carrington, T., A Random-Sampling High Dimensional Model Representation Neural Network for Building Potential Energy Surfaces. *J. Chem. Phys.* **2006,** *125* (8), 084109.
(60) Manzhos, S.; Nakai, K.; Yamashita, K., Three-Body Interactions in Clusters CO–(pH_2)n. *Chem. Phys. Lett.* **2010,** *493* (4–6), 229-233.
(61) Rauhut, G., Efficient Calculation of Potential Energy Surfaces for the Generation of Vibrational Wave Functions. *J. Chem. Phys.* **2004,** *121* (19), 9313-9322.
(62) Pedone, A.; Malavasi, G.; Menziani, M. C.; Cormack, A. N.; Segre, U., A New Self-Consistent Empirical Interatomic Potential Model for Oxides, Silicates, and Silica-Based Glasses. *J. Phys. Chem. B* **2006,** *110* (24), 11780-11795.
(63) Kutteh, R.; Avdeev, M., Initial Assessment of an Empirical Potential as a Portable Tool for Rapid Investigation of Li+ Diffusion in Li+-Battery Cathode Materials. *J. Phys. Chem. C* **2014,** *118* (21), 11203-11214.
(64) Kerisit, S.; Chaka, A. M.; Droubay, T. C.; Ilton, E. S., Shell Model for Atomistic Simulation of Lithium Diffusion in Mixed Mn/Ti Oxides. *J. Phys. Chem. C* **2014,** *118* (42), 24231-24239.

(65) Yildirim, H.; Greeley, J.; Sankaranarayanan, S. K. R. S., Effect of Concentration on the Energetics and Dynamics of Li Ion Transport in Anatase and Amorphous TiO_2. *J. Phys. Chem. C* **2011,** *115* (31), 15661-15673.
(66) Xiong, H.; Yildirim, H.; Shevchenko, E. V.; Prakapenka, V. B.; Koo, B.; Slater, M. D.; Balasubramanian, M.; Sankaranarayanan, S. K. R. S.; Greeley, J. P.; Tepavcevic, S.; Dimitrijevic, N. M.; Podsiadlo, P.; Johnson, C. S.; Rajh, T., Self-Improving Anode for Lithium-Ion Batteries Based on Amorphous to Cubic Phase Transition in TiO_2 Nanotubes. *J. Phys. Chem. C* **2012,** *116* (4), 3181-3187.
(67) Daw, M. S.; Baskes, M. I., Embedded-atom Method: Derivation and Application to Impurities, Surfaces, and Other Defects in Metals. *Phys. Rev. B* **1984,** *29* (12), 6443-6453.
(68) Janesko, B. G.; Barone, V.; Brothers, E. N., Accurate Surface Chemistry beyond the Generalized Gradient Approximation: Illustrations for Graphene Adatoms. *J. Chem. Theory Comput.* **2013,** *9* (11), 4853-4859.
(69) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001,** *105* (41), 9396-9409.
(70) Jenkins, S., Quantum Topology Phase Diagrams for Molecules, Clusters, and Solids. *Int. J. Quantum Chem.* **2013,** *113* (11), 1603-1608.
(71) Figueredo, F. A.; Maza, J. R.; Kirk, S. R.; Jenkins, S., Quantum Topology Phase Diagrams for the Cis- and Trans-isomers of the Cyclic Contryphan-Sm Peptide. *Int. J. Quantum Chem.* **2014,** *114* (24), 1697-1706.
(72) Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W.-Q.; Goddard, W. A., Development of the ReaxFF Reactive Force Field for Describing Transition Metal Catalyzed Reactions, with Application to the Initial Stages of the Catalytic Formation of Carbon Nanotubes. *J. Phys. Chem. A* **2005,** *109* (3), 493-499.
(73) Bae, G.-T.; Aikens, C. M., Improved ReaxFF Force Field Parameters for Au–S–C–H Systems. *J. Phys. Chem. A* **2013,** *117* (40), 10438-10446.
(74) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998,** *58* (11), 7260-7268.
(75) Slater, J. C.; Koster, G. F., Simplified LCAO Method for the Periodic Potential Problem. *Phys. Rev.* **1954,** *94* (6), 1498-1524.
(76) Papaconstantopoulos, D. A., *Handbook of the Band Structure of Elemental Solids*. Plenum Press: New York, 1986.
(77) Mishin, Y.; Farkas, D.; Mehl, M. J.; Papaconstantopoulos, D. A., Interatomic Potentials for Monoatomic Metals from Experimental Data and Ab Initio Calculations. *Phys. Rev. B* **1999,** *59* (5), 3393-3407.
(78) Koskinen, P.; Mäkinen, V., Density-Functional Tight-Binding for Beginners. *Comput. Mater. Sci.* **2009,** *47* (1), 237-253.
(79) Papaconstantopoulos, D. A., Tight-Binding Hamiltonians. In *Alloy Phase Stability*, Stocks, G. M.; Gonis, A., Eds. Springer Netherlands: 1989; Vol. 163, pp 351-356.

(80) Yang, W., Direct Calculation of Electron Density in Density-functional Theory. *Phys. Rev. Lett.* **1991,** *66* (11), 1438-1441.
(81) Yang, W.; Lee, T. S., A Density-matrix Divide-and-conquer Approach for Electronic Structure Calculations of Large Molecules. *J. Chem. Phys.* **1995,** *103* (13), 5674-5678.
(82) Thomas, F.; Gotthard, S.; Marcus, E.; Thomas, N.; Christof, K.; Marc, A.; Michael, S.; Zoltán, H.; Aldo, D. C.; Sándor, S., Atomistic Simulations of Complex Materials: Ground-State and Excited-State Properties. *J. Phys.: Condens. Matter* **2002,** *14* (11), 3015.
(83) Heckel, W.; Elsner, B. A. M.; Schulz, C.; Müller, S., The Role of Hydrogen on the Adsorption Behavior of Carboxylic Acid on TiO_2 Surfaces. *J. Phys. Chem. C* **2014,** *118* (20), 10771-10779.
(84) Szűcs, B.; Hajnal, Z.; Scholz, R.; Sanna, S.; Frauenheim, T., Theoretical Study of the Adsorption of a PTCDA Monolayer on S-passivated GaAs(l 0 0). *Appl. Surf. Sci.* **2004,** *234* (1–4), 173-177.
(85) Zhao, Y. L.; Köppen, S.; Frauenheim, T., An SCC-DFTB/MD Study of the Adsorption of Zwitterionic Glycine on a Geminal Hydroxylated Silica Surface in an Explicit Water Environment. *J. Phys. Chem. C* **2011,** *115* (19), 9615-9621.
(86) Manzhos, S., Comparative Density Functional Theory and Density Functional Tight Binding Study of 2-Anthroic Acid on TiO_2. *Chem. Phys. Lett.* **2016,** *643*, 16-20.
(87) Rungsawang, R.; Ueno, Y.; Tomita, I.; Ajito, K., Angle-Dependent Terahertz Time-Domain Spectroscopy of Amino Acid Single Crystals. *J. Phys. Chem. B* **2006,** *110* (42), 21259-21263.
(88) Bertini, I.; Duma, L.; Felli, I. C.; Fey, M.; Luchinat, C.; Pierattelli, R.; Vasos, P. R., A Heteronuclear Direct-Detection NMR Spectroscopy Experiment for Protein-Backbone Assignment. *Angewandte Chemie* **2004,** *116* (17), 2307-2309.
(89) Cabella, C.; Karlsson, M.; Canapè, C.; Catanzaro, G.; Colombo Serra, S.; Miragoli, L.; Poggi, L.; Uggeri, F.; Venturi, L.; Jensen, P. R.; Lerche, M. H.; Tedoldi, F., In Vivo and in Vitro Liver Cancer Metabolism Observed with Hyperpolarized [5-(13)C]Glutamine. *J. Magn. Reson.* **2013,** *232*, 45-52.
(90) Soon, P. C.; Xu, X.; Zhang, B.; Gruppi, F.; Canary, J. W.; Jerschow, A., Hyperpolarization of Amino Acid Precursors to Neurotransmitters with Parahydrogen Induced Polarization. *Chem. Commun.* **2013,** *49* (46), 5304-5306.
(91) Tan, H.; Luo, W.; Wei, L.; Chen, B.; Li, W.; Xiao, L.; Manzhos, S.; Liu, Z.; Liang, S., Quantifying the Distribution of the Stoichiometric Composition of Anti-cancer Peptide Lycosin-I on Lipid Membrane with Single Molecule Spectroscopy. *J. Phys. Chem. B* **2016**.
(92) Luschtinetz, R.; Frenzel, J.; Milek, T.; Seifert, G., Adsorption of Phosphonic Acid at the TiO_2 Anatase (101) and Rutile (110) Surfaces. *J. Phys. Chem. C* **2009,** *113* (14), 5730-5740.
(93) Dolgonos, G.; Aradi, B.; Moreira, N. H.; Frauenheim, T., An Improved Self-Consistent-Charge Density-Functional Tight-Binding (SCC-DFTB) Set of Parameters for Simulation of Bulk and Molecular Systems Involving Titanium. *J. Chem. Theory Comput.* **2010,** *6* (1), 266-278.

(94) Vivès, E.; Brodin, P.; Lebleu, B., A Truncated HIV-1 Tat Protein Basic Domain Rapidly Translocates through the Plasma Membrane and Accumulates in the Cell Nucleus. *J. Biol. Chem.* **1997,** *272* (25), 16010-16017.
(95) Frankel, A. D.; Pabo, C. O., Cellular Uptake of the TAT Protein from Human Immunodeficiency Virus. *Cell* **1988,** *55* (6), 1189-1193.
(96) Schmidt, N.; Mishra, A.; Lai, G. H.; Wong, G. C. L., Arginine-Rich Cell-Penetrating Peptides. *FEBS Letters* **2010,** *584* (9), 1806-1813.
(97) Hanawa, T., Reconstruction and Regeneration of Surface Oxide Film on Metallic Materials in Biological Environments. *Corros. Rev.* **2003,** *21* (2-3), 161-181.
(98) Teow, Y.; Asharani, P. V.; Hande, M. P.; Valiyaveettil, S., Health Impact and Safety of Engineered Nanomaterials. *Chem. Commun.* **2011,** *47* (25), 7025-7038.
(99) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586-3616.
(100) Pastor, R. W.; MacKerell, A. D., Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2011,** *2* (13), 1526-1532.
(101) Haugk, M.; Elsner, J.; Heine, T.; Frauenheim, T.; Seifert, G., A parallel code for a self-consistent charge density functional based tight binding method: Total energy calculations for extended systems. *Comput. Mater. Sci.* **1999,** *13* (4), 239-251.
(102) Wahiduzzaman, M.; Oliveira, A. F.; Philipsen, P.; Zhechkov, L.; van Lenthe, E.; Witek, H. A.; Heine, T., DFTB Parameters for the Periodic Table: Part 1, Electronic Structure. *J. Chem. Theory Comput.* **2013,** *9* (9), 4006-4017.
(103) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
(104) Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98* (7), 5648-5652.

(105) Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988,** *37* (2), 785-789.
(106) Vosko, S. H.; Wilk, L.; Nusair, M., Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980,** *58* (8), 1200-1211.
(107) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994,** *98* (45), 11623-11627.
(108) Gaus, M.; Goez, A.; Elstner, M., Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013,** *9* (1), 338-354.
(109) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q., Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **2014,** *10* (4), 1518-1537.
(110) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.
(111) Shen, Y.; Maupetit, J.; Derreumaux, P.; Tufféry, P., Improved PEP-FOLD Approach for Peptide and Miniprotein Structure Prediction. *J. Chem. Theory Comput.* **2014,** *10* (10), 4745-4758.
(112) Thévenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P., PEP-FOLD: An Updated De Novo Structure Prediction Server for Both Linear and Disulfide Bonded Cyclic Peptides. *Nucleic Acids Res.* **2012,** *40* (W1), W288-W293.
(113) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992,** *114* (25), 10024-10035.
(114) José, M. S.; Emilio, A.; Julian, D. G.; Alberto, G.; Javier, J.; Pablo, O.; Daniel, S.-P., The SIESTA Method for Ab Initio Order- N Materials Simulation. *J. Phys.: Condens. Matter* **2002,** *14* (11), 2745.
(115) Troullier, N.; Martins, J. L., Efficient Pseudopotentials for Plane-wave Calculations. *Phys. Rev. B* **1991,** *43* (3), 1993-2006.
(116) Legrain, F.; Malyi, O.; Manzhos, S., Insertion energetics of lithium, sodium, and magnesium in crystalline and amorphous titanium dioxide: A comparative first-principles study. *J. Power Sources* **2015,** *278*, 197-202.
(117) Monkhorst, H. J.; Pack, J. D., Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* **1976,** *13* (12), 5188-5192.
(118) Triggiani, L.; Muñoz-García, A.; Agostiano, A.; Pavone, M., First-Principles Study of Trimethylamine Adsorption on Anatase TiO_2 Nanorod Surfaces. *Theor Chem Acc* **2015,** *134* (10), 1-11.
(119) Tillotson, M. J.; Brett, P. M.; Bennett, R. A.; Grau-Crespo, R., Adsorption of Organic Molecules at the TiO_2(110) Surface: The Effect of Van Der Waals Interactions. *Surf. Sci.* **2015,** *632*, 142-153.

(120) Manzhos, S.; Giorgi, G.; Yamashita, K., A Density Functional Tight Binding Study of Acetic Acid Adsorption on Crystalline and Amorphous Surfaces of Titania. *Molecules* **2015,** *20* (2), 3371.
(121) Tang, H.; Berger, H.; Schmid, P. E.; Lévy, F.; Burri, G., Photoluminescence in TiO2 anatase single crystals. *Solid State Commun.* **1993,** *87* (9), 847-850.
(122) Rak, J.; Skurski, P.; Simons, J.; Gutowski, M., Low-Energy Tautomers and Conformers of Neutral and Protonated Arginine. *J. Am. Chem. Soc.* **2001,** *123* (47), 11695-11707.
(123) Schlund, S.; Müller, R.; Graβmann, C.; Engels, B., Conformational Analysis of Arginine in Gas Phase—A Strategy for Scanning the Potential Energy Surface Effectively. *J. Comput. Chem.* **2008,** *29* (3), 407-415.
(124) Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Harañczyk, M.; Gutowski, M., Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *J. Phys. Chem. A* **2006,** *110* (44), 12282-12291.
(125) Brandt, E. G.; Lyubartsev, A. P., Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogues and a Titanium Binding Peptide on the TiO_2 (100) Surface. *J. Phys. Chem. C* **2015,** *119* (32), 18126-18139.
(126) Sowmiya, M.; Senthilkumar, K., Adsorption of RGD Tripeptide on Anatase (0 0 1) Surface – A First Principle Study. *Comput. Mater. Sci.* **2015,** *104*, 124-129.
(127) Costa, D.; Pradier, C.-M.; Tielens, F.; Savio, L., Adsorption and Self-Assembly of Bio-Organic Molecules at Model Surfaces: A Route towards Increased Complexity. *Surf. Sci. Rep.* **2015,** *70* (4), 449-553.
(128) Koch, R.; Lipton, A.; Filipek, S.; Renugopalakrishnan, V., Arginine Interactions with Anatase TiO_2 (100) Surface and the Perturbation of 49Ti NMR Chemical Shifts – a DFT Investigation: Relevance to Renu-Seeram Bio Solar Cell. *J Mol Model* **2011,** *17* (6), 1467-1472.
(129) Singla, P.; Riyaz, M.; Singhal, S.; Goel, N., Theoretical Study of Adsorption of Amino Acids on Graphene and BN Sheet in Gas and Aqueous Phase with Empirical DFT Dispersion Correction. *Phys. Chem. Chem. Phys.* **2016,** *18* (7), 5597-5604.
(130) Zhang, H.-p.; Lu, X.; Luo, X.-g.; Lin, X.-y.; Zhou, Y.-f., Effects of O-deficiency on the Interaction between Rutile and Arg: A Density Functional Theory Study. *Physica E Low Dimens Syst. Nanostruct.* **2014,** *61*, 83-89.
(131) Miertuš, S.; Scrocco, E.; Tomasi, J., Electrostatic Interaction of a Solute with a Continuum. A Direct Utilizaion of Ab Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981,** *55* (1), 117-129.
(132) Miertuš, S.; Tomasi, J., Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes. *Chem. Phys.* **1982,** *65* (2), 239-245.
(133) Barone, V.; Cossi, M.; Tomasi, J., Geometry Optimization of Molecular Structures in Solution by the Polarizable Continuum Model. *J. Comput. Chem.* **1998,** *19* (4), 404-417.

(134) Tomasi, J.; Mennucci, B.; Cancès, E., The IEF Version of the PCM Solvation Method: An Overview of a New Method Addressed to Study Molecular Solutes at the QM Ab Initio Level. *J. Mol. Struct. Theochem* **1999,** *464* (1–3), 211-226.
(135) Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105* (8), 2999-3094.
(136) Momma, K.; Izumi, F., VESTA: a three-dimensional visualization system for electronic and structural analysis. *J. Appl. Crystallogr.* **2008,** *41* (3), 653-658.
(137) Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-38.
(138) Landmann, M.; Rauls, E.; Schmidt, W. G., The Electronic Structure and Optical Response of Rutile, Anatase and Brookite TiO_2. *J. Phys.: Condens. Matter* **2012,** *24* (19), 195503.
(139) Tan, H.; Luo, W.; Wei, L.; Chen, B., Quantifying the Distribution of the Stoichiometric Composition of Anticancer Peptide Lycosin-I on the Lipid Membrane with Single Molecule Spectroscopy. *J. Phys. Chem. B* **2016,** *120* (12), 3081-3088.
(140) Bourke, P., Gaussian Cube Files. **2003**.
(141) Anselmi, C.; Mosconi, E.; Pastore, M.; Ronca, E.; De Angelis, F., Adsorption of Organic Dyes on TiO_2 Surfaces in Dye-Sensitized Solar Cells: Interplay of Theory and Experiment. *Phys. Chem. Chem. Phys.* **2012,** *14* (46), 15963-15974.
(142) Chan, M.; Carrington, T.; Manzhos, S., Anharmonic Vibrations of the Carboxyl Group in Acetic Acid on TiO_2: Implications for Adsorption Mode Assignment in Dye-Sensitized Solar Cells. *Phys. Chem. Chem. Phys.* **2013,** *15* (25), 10028-10034.
(143) Buckeridge, J.; Butler, K. T.; Catlow, C. R. A.; Logsdail, A. J.; Scanlon, D. O.; Shevlin, S. A.; Woodley, S. M.; Sokol, A. A.; Walsh, A., Polymorph Engineering of TiO_2: Demonstrating How Absolute Reference Potentials Are Determined by Local Coordination. *Chem. Mater.* **2015,** *27* (11), 3844-3851.
(144) Martsinovich, N.; Jones, D. R.; Troisi, A., Electronic Structure of TiO_2 Surfaces and Effect of Molecular Adsorbates Using Different DFT Implementations. *J. Phys. Chem. C* **2010,** *114* (51), 22659-22670.
(145) Scanlon, D. O.; Dunnill, C. W.; Buckeridge, J.; Shevlin, S. A.; Logsdail, A. J.; Woodley, S. M.; Catlow, C. R. A.; Powell, M. J.; Palgrave, R. G.; Parkin, I. P.; Watson, G. W.; Keal, T. W.; Sherwood, P.; Walsh, A.; Sokol, A. A., Band Alignment of Rutile and Anatase TiO_2. *Nat Mater* **2013,** *12* (9), 798-801.
(146) Xiong, G.; Shao, R.; Droubay, T. C.; Joly, A. G.; Beck, K. M.; Chambers, S. A.; Hess, W. P., Photoemission Electron Microscopy of TiO_2 Anatase Films Embedded with Rutile Nanocrystals. *Adv. Funct. Mater.* **2007,** *17* (13), 2133-2138.
(147) Heyd, J.; Scuseria, G. E.; Ernzerhof, M., Hybrid Functionals Based on a Screened Coulomb Potential. *J. Chem. Phys.* **2003,** *118* (18), 8207-8215.
(148) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E., Influence of the Exchange Screening Parameter on the Performance of Screened Hybrid Functionals. *J. Chem. Phys.* **2006,** *125* (22), 224106.

(149) Peter, D.; Jolla, K.; Bálint, A.; Thomas, F.; Ladislaw, K., Water Splitting and the Band Edge Positions of TiO_2 . *Electrochim. Acta* **2016**, Submitted.
(150) Albuquerque, A. R.; Bruix, A.; Sambrano, J. R.; Illas, F., Theoretical Study of the Stoichiometric and Reduced Ce-Doped TiO_2 Anatase (001) Surfaces. *J. Phys. Chem. C* **2015,** *119* (9), 4805-4816.
(151) Tosoni, S.; Lamiel-Garcia, O.; Fernandez Hevia, D.; Doña, J. M.; Illas, F., Electronic Structure of F-Doped Bulk Rutile, Anatase, and Brookite Polymorphs of TiO_2. *J. Phys. Chem. C* **2012,** *116* (23), 12738-12746.
(152) Caravaca, M. A.; Casali, R. A.; Miño, J. C., Prediction of Electronic, Structural and Elastic Properties of the Hardest Oxide: TiO_2. *Phys. Status Solidi B* **2009,** *246* (3), 599-603.
(153) Liu, G.; Yin, L.-C.; Wang, J.; Niu, P.; Zhen, C.; Xie, Y.; Cheng, H.-M., A Red Anatase TiO_2 Photocatalyst for Solar Energy Conversion. *Energy Environ. Sci.* **2012,** *5* (11), 9603-9610.