DOTTORATO DI RICERCA IN FISICA E NANOSCIENZE

XXVI CICLO

Sede Amministrativa

Università degli Studi di MODENA e REGGIO EMILIA

Tesi per il Conseguimento del Titolo di
Dottore di Ricerca

Simulating the association between DNA and a gold substrate

Candidato: Marta Rosa

Relatori: Dr. Rosa Di Felice
Dr. Stefano Corni

Coordinatore del dottorato
e Direttore della scuola: Prof. Franca Manghi

Febbraio 2014

Contents

List of abbreviations

AFM	*Atomic Force Microscopy*
BD	*Brownian Dynamics*
BO	*Born-Oppenheimer*
C09	*Cooper 09*
DNA	*Deoxyribonucleic acid*
DFT	*Density Functional Theory*
DOS	*Density Of States*
FF	*Force Field*
GGA	*Generalized Gradient Approximation*
HOMO	*Highest Occupied Molecular Orbital*
LDA	*Local Density Approximation*
LJ	*Lennard Jones*
MC	*Monte Carlo*
MD	*Molecular Dynamics*
MM	*Molecular Mechanics*
MO	*Molecular Orbital*
PBE	*Pederew-Burke-Ernzerhof*
PME (PPME)	*Particle-Mesh Ewald (Particle-Particle-Mesh Ewald)*
PP	*Pseudo Potential*
PW	*Plane Waves*
QM	*Quantum Mechanics*
revPBE	*revised PBE*
XC	*exchange-correlation*

Chapter 1

Introduction

In the last 70 years the spreading of computational science definitely modified the approach to the study of physical and chemical systems.

Experimental observations and analysis of the structure and properties of condensed matter allowed the first comprehension of the mechanisms that rule the behavior of a vast number of systems. Moreover, the importance of experiments in providing the starting point and the benchmark for theoretical and computational studies has been largely assessed. Nevertheless, experimental informations are not always able to determine univocally the structure and reactions of the system under study without a further interpretation provided by a theoretical approach. At the same time, fully theoretical studies of important phenomena, based on quantum, classical and statistical mechanics, lead to equations that can not be solved analytically except for a few special cases (i.e. H atom, two body systems). The need for approximated techniques able to solve the many body problem is stringed to fill the gap between theoretical and experimental techniques. Computer simulations have indeed altered the interplay between experiment and theory.

The birth of computational science enabled the study of processes that are out of the range of experimental techniques: it is able to provide information on quantum mechanisms and on how the system evolves in time. The combination of a quantum and a classical description allows to deal with different systems, ranging from small ones dominated by quantum processes to large ones that can be completely described with classical mechanics. The main problem in the

computational approach is the balance between the approximations adopted (and consequently the accuracy of the description) and the size of the system under study, which influences the computational cost of the calculations: a fully quantum description, even with approximations, is possible only for systems of some hundred of atoms. Despite the small dimensions of the systems that can be studied, quantum analysis remains a fundamental tool to fully understand the processes that are behind experimental observations. Less accurate models, as simple empirical models based on experiments or classical and coarse grained models based on quantum calculations, can however be developed for a less computationally expensive simulation of the studied systems.

This year the awarding of the Nobel Prize to Martin Karplus, Michael Levitt and Arieh Warshel, who are the inventors of molecular simulations and the promoters of Quantum Mechanics/Molecular Mechanics (QM/MM) method[1], demonstrates how the computational approach has become of fundamental importance. The Nobel prize *"focuses on how to evaluate the variation in the energy of the real system in a accurate and efficient way for systems where relatively large geometry changes or changes in electronic configuration in a smaller part of the studied system is strongly coupled to a surrounding which is only weakly perturbed"* [2]. A fundamental problem is indeed how to describe a system accurately enough relying on the available computational power. The development of more and more powerful computational resources can not be considered as the ultimate solution, as new methods and resources in the past has always brought to new questions. Physicochemical intuition is still the key to tackle complex problems in materials and biological sciences.

In the following, I will present a combination of quantum and classical techniques, joined in a unique approach that belongs to this vast front of computational methods. I will focus on the application of these methodologies to DNA adsorbed on the Au(111) surface.

The adsorption and interaction of biomolecules with hard materials is a complex phenomenon which raises numerous difficulties due to the deep differences between the two system components, which reflects in a difference in the computational tools and approximations suitable for each of them. Proteins are actually one of the main subjects of study in biochemistry, both in solution and, more recently, interacting with surfaces. In the *Scientific Background of the Nobel Prize in Chemistry 2013* [2], Levitt's study of the folding of the protein Bovine Pancreas Trypsin Inhibitor (BPTI)[3] is cited as one of the first contributions to this research field. For what con-

cerns the interaction with surfaces, several natural processes involve the interaction between polypeptides and solid surfaces[4]. Some examples include protein—inorganic-surface interactions that regulate the growth of hard tissues[5, 6], the anti-freezing capabilities of animals living in cold environments[7, 8], the adhesion of cells to solid supports[9] and the interaction of tissues with artificial medical implants. For this reason the interaction of organic molecules with hard surfaces is becoming a hot research topic. Moreover, new computational methods were developed with this purpose.

In recent years many efforts have been addressed also to the study of DNA, both for its inherent properties, because of the main role it has in the development of life and in the understanding of many biophysical and medical processes, and for the possible nano-electronic applications, where its peculiar properties would be of great usage in the creation of new devices[10]. Its ability to create long one-dimensional polymers through a self assembling mechanism makes it a fascinating candidate for charge transport in nano-devices, and many studies have been devoted to obtain information about its capabilities as a charge conductor[10, 11, 12, 13].

The comprehension of DNA interactions with hard substrates is a fundamental step towards its exploitation in nanotechnology applications, as the construction of bio-devices is totally dependent on the mechanisms that develop at the organic/inorganic interface. Nevertheless, the understanding of the DNA behavior when interacting with surfaces is a challenging topic both from the experimental and theoretical points of view. A big effort is still needed in this direction.

The behavior of a DNA molecule adsorbed on a gold surface and the setup of suitable methods for its study are the subject of this PhD thesis.

1.1 DNA structure

DNA is a polymer made of repeating units called nucleotides. The nucleotides are made of a 5-carbon sugar called deoxyribose, a phosphate group and a molecule called *base*. There are four possible nucleotides, corresponding to four possible bases, that are called adenine, guanine, cytosine and thymine (Figure 1.1 a).

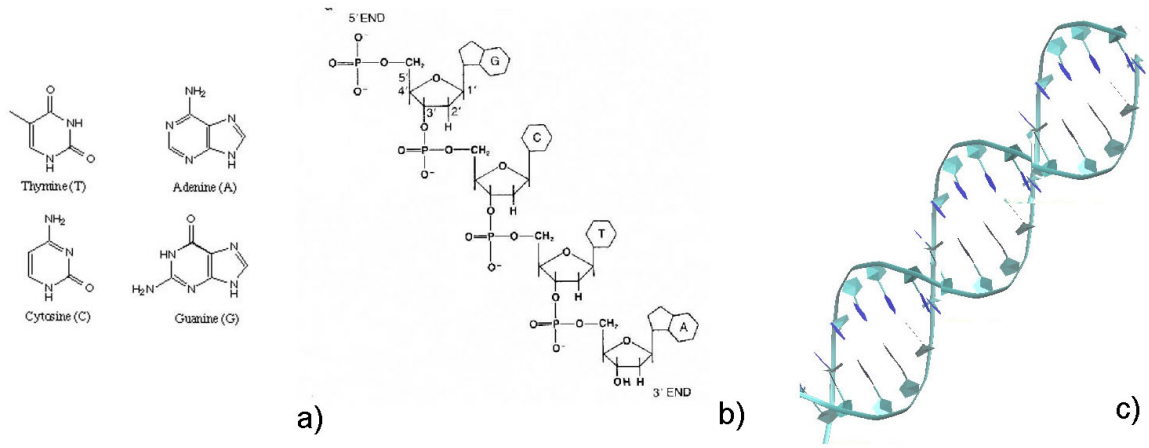


Figure 1.1: a) nucleobases b) DNA single strand structure c)DNA double helix

Nucleotides are joined to each other in a chain through the phosphate groups. A phosphate group is bonded on one side to the 3^{rd} carbon atom of one sugar and on the other side to the 5^{th} carbon atom of another sugar (Figure 1.1 b). This way of bonding is not symmetric and this allows to identify a direction of the DNA strand. The two ends are called 3' and 5' depending on what C atom is not bonded to the phosphate group.

Usually the DNA single strand does not stand alone, it interacts by complementarity with another similar strand, forming a particular twisted structure called double helix. The DNA double helix structure is stabilized by the interactions between the bases: the hydrogen bonds act between the bases belonging to different strands; base-stacking interactions act among bases belonging to the same strand, due to their heterocyclic structure. Figure 1.1c shows that the interacting bases are inside the double helix structure, and in this way they also minimize their interaction with the solvation shell when they are in solution, which is DNA natural environment. The chain formed by sugars and phosphates is called backbone and builds the external part of the structure. Because of the solubility of phosphates in solution, the backbone is naturally charged and each nucleotide carries a charge equal to -1. Apart for the twisting shape, the double helix motif is strictly characterized by two other features. First of all the fact that the directions of the two strands are opposite, which means that the strands are antiparallel; the second is that each base can not pair by H-bonds with any of the others but

with just one of them: they are classified in purines and pyrimidines, depending on the number and the shape of the aromatic rings. The purines are guanine and adenine and are made by one five and one six atoms heterocycles fused together, while cytosine and thymine are the pyrimidines, six membered rings. A purine base can only H-bond to a pyrimidine, and more explicitly cytosine pairs with guanine and thymine binds with adenine.

1.2 DNA on surfaces

The interaction of DNA with surfaces is fundamental both for the study of DNA itself than for its possible nanoelectronics applications. For what concerns the first aspect, from the medical point of view the sequencing of DNA fragments has become of great importance after the association of genetic variation with both diseases and drug response [14, 15]. Sequencing measurements imply both the study of DNA in its natural environment and its interaction with the apparatus needed for the measurement, which means both electromagnetic radiation and nano-particles and hard surfaces. To make an example, a typical array-based sensor for the sequencing of DNA consists of single-stranded oligonucleotides of different sequences attached to a surface, that are supposed to hybridize with the target single-stranded DNA, allowing to discover the unknown sequence [16].

The investigation of DNA in the context of nanoelectronic is driven by the need for novel electrical and computing devices, which bypass the limitations of Si-based technology. The electronic industry has pursued so far the miniaturization of electronic devices to reduce dimensions and increase performances. This rapid development is limited by those physical properties that become determinant at the nano scale, as heat dissipation and capacitive coupling. For this reason in recent years major efforts moved to bottom-up construction from nanoscale units that are endowed with special properties as self assembling and recognition. Biomolecules are appealing candidates as nanoscale building blocks. DNA in particular presents some fascinating features, e.g. complementary base pairing, that would allow interesting applications in electronics.

The components of electrical circuits are supported on hard inorganic surfaces. Furthermore, scanning measurement setups require deposition on surfaces[17] . Hence, the comprehension of the DNA-surface coupling become of paramount importance. The first and main question is

obviously if DNA is able to transport charge, depending on the length, sequence and environment [10, 11, 12, 13]. Focusing on the environment, one could ask which kind of structure and function modifications are induced by the interaction with an inorganic surface.
Determining the behavior of DNA at interfaces is particularly challenging, as the modifications of its structure when approaching to interfaces are potentially severe. These modifications are surely due to the strong charge that belongs to the molecule backbone (Section 1.1), but also to the dispersion interactions between the surface and the bases, if the latter are exposed to the external environment by geometry modifications. To make an example of the strong interaction between surface and DNA, we mention the results of AFM experiments on double-stranded DNA molecules deposited "horizontally" on a hard substrates: the measured height of the deposited helix is 50% of its diameter in solution[18, 19]. This measure suggests that an unfolding process is taking place.
The possibility of simulating correctly the behavior of a DNA molecule approaching to a substrate would allow a deep understanding of the interactions involved and the possibility of exploiting DNA properties for the construction of nano-devices. Such goal needs to be approached from the very origin, studying the different components of DNA molecule adsorbed on the surface to be sure of taking into account all the different aspects of the interaction, and moving then to the simulation of the whole molecule on the surface. With this purpose in mind, different computational methods have been combined in a multi-step approach able to give a correct description of the behavior of a DNA molecule adsorbed on a Au(111) surface in aqueous environment. This approach, together with the different computational methods used, is explained in details in the following chapter.

This PhD thesis is focused on the parametrization of a classical force field, able to describe the adsorption of DNA on a Au(111) surface. The study of the interaction between single DNA bases and the surface is performed through ab-initio density functional theory (DFT) calculations, and results on the geometry and electronic structure of the adsorbed molecules are presented in the following chapters. DFT results are then reproduced with molecular dynamics (MD) simulations parametrizing a suitable FF for the purpose. The FF itself is then validated comparing DFT, classical MD and experimental results on more complex structures. The parametrization and validation of the FF are the main topic of this PhD thesis, along with the results on in-vacuo and

in solution systems obtained with DFT and MD calculations. Eventually, preliminary results on docking are presented, able to determine some starting configurations for future classical MD simulations of the DNA molecule adsorbed on Au(111) surface in solution.

Chapter 2

Methodology

In this chapter the computational techniques used to investigate the adsorption of a DNA molecule on Au(111) surface will be presented.

Nowadays computer simulations have become fundamental for the study of the microscopic mechanisms that hide behind any complex process. They can both be combined with experiments to understand deeply the behavior of the system under study or used to predict new results and give new inputs for further investigations. However, to set up a simulation different ingredients are needed, which can be derived by the comparison with experiments or obtained by a theoretical description of the system.

The simulation of a complex process, such as the refolding of a molecule on a surface, requires a careful approach, with a specific attention to the different length and time scales involved. Classical molecular dynamics (MD) atomistic simulations (Section 2.3.1) are in principle the most suitable method for the description of such a process, as they have successfully been employed to describe the unfolding of biological molecules in solution. [20, 21, 22] However, the crucial point in any classical MD methodology is the parametric force field (FF), which is the potential energy necessary for the evolution of the atomic coordinates in time according to Newton's equations of motion[23, 24]. A correct description of the environment and of the objects belonging to the system is needed, as the reliability of results is deeply dependent on the correct parametrization of the FF.

Reliable FFs exist for solid, liquid and molecular materials, included DNA in solution; for what

concerns the adsorption of organic molecules on hard substrates the parametrization of suitable FFs is still in its infancy. This is even more true for DNA, as in recent years major efforts have been devoted to the interaction of proteins with surfaces.

Some FF already exist for the description of the interaction of DNA with surfaces, relying on theory or on experiments for the parametrization of polar and non polar interactions[25, 26, 27]. One of the main problems when generating a FF is to keep it general enough to be able to describe different geometries and conditions, exploiting suitable approximations for the different interactions, and to have it specific enough to catch the characteristics of each particular system. CHARMM-METAL FF [25] has been chosen by some groups for the simulation of peptides adsorbed on metals[28, 29, 30, 31, 32] and can in principle be used also for DNA on Au(111). One of the main advantages of this FF is that it is not limited to a specific molecule or surface. This means that in principle it would be possible to set up a simulation for DNA on Au(111), since both are well described separately by the FF and interactions are taken into account through approximations. On the other hand, if it is true that a FF must be as general as possible to be able to describe different conditions (e. g. in vacuo and solvated) and geometries (e. g. monolayers, single molecules, etc) it is also true that this general parametrization may induce a certain lack of accuracy. Morover, another limitation of CHARMM-METAL is that it does not take into account the polarization of the surface. This was already presented as a source of uncertainty for the adsorption of peptides on surfaces, even if its contribution was thought to be small [32]. In DNA, the backbone carries a -1e^- charge for each nucleotide: this means that image charge effects are likely non negligible for such a system.

Besides CHARMM, there are some other FFs parametrized specifically for DNA on Au, even taking into account polarization interactions. The FF developed by Zerbetto et al. [26, 27], for example, treats the distribution of charges of the system in a dynamical way, which means that at certain intervals the charges on the atoms are reassigned depending on the environment[33]. In this kind of more specific FF the parameters describing the non polar interactions of the single DNA bases with the surface are fitted exploiting experimental parameters. This contribution includes long range interaction and other interactions not included in the electrostatic model. In particular, Zerbetto et al. used monolayer experiments to fit FF parameters.

The main disadvantage of this approach is that such a detailed FF is computationally cumbersome and is probably not suitable for such a complex system as a double stranded DNA

molecule adsorbed on the surface in solution.

In our group the GolP force field was developed for the description of the interaction of proteins with the Au(111) surface [34, 35] and we have recently reviewed our computational approach in this context, as well as similar efforts by other groups[36]. Maintaining the original structure of the GolP FF, the same procedure used for proteins can be adopted for the parametrization of the interaction between DNA and gold. The parametrization of this FF relies on a multi-step approach that already gave good results for other molecules adsorbed on surfaces [37] and we chose to transfer it to the study of our present system, DNA on Au(111). The outline of the method is sketched in Figure 2.1.

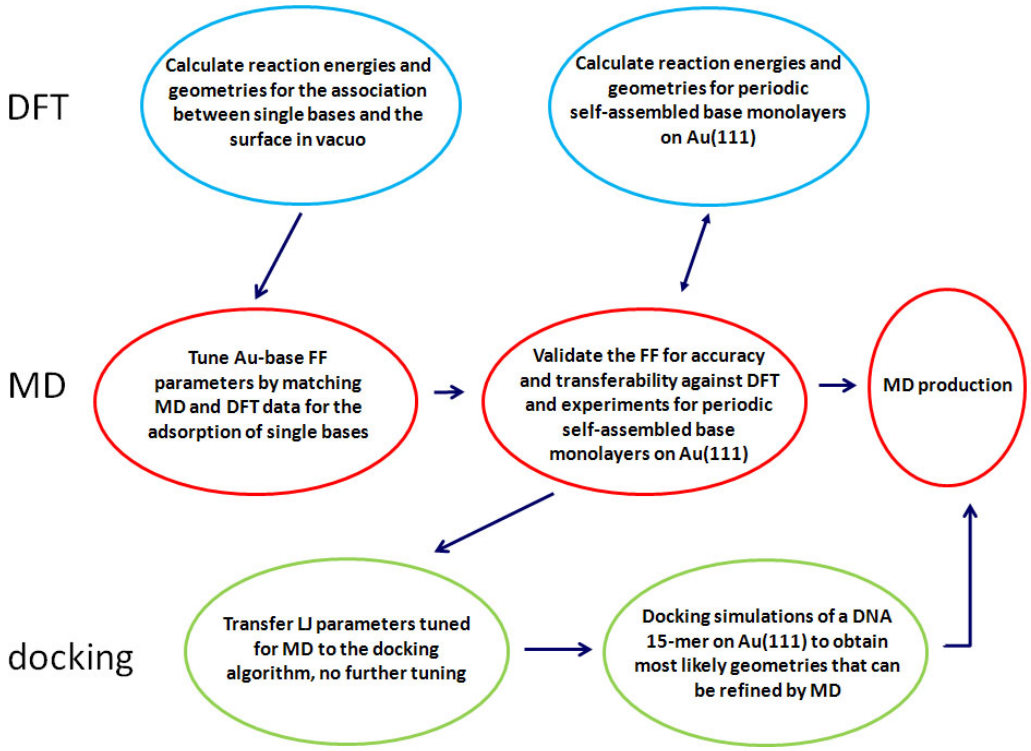


Figure 2.1: Outilne of the multistep approach used. Different colors are for different method used: DFT (blue), MD (red), Docking (green).

The FF is parametrized with respect to first principles calculations of DNA single bases adsorbed on Au(111). In particular, we started with performing quantum mechanical (QM) Density Functional Theory (DFT) calculations of single DNA bases adsorbed on gold, postponing the study

of the backbone. This choice is motivated by the fact that the backbone is charged and interacts mainly through electrostatics, while for what concerns the single bases long range dispersion interactions are fundamental and difficult to resolve in a different way.

The importance of studying the adsorption of single bases on gold is motivated first of all by the intrinsic interest in unraveling the interaction of such molecules with hard substrates, that can have applications in electronics (e.g. self-assembled structures on metals, sequencing of unknown DNA strands through single strand adsorption). Moreover, concerning our specific goal, that is the description of the adsorption of a DNA molecule, if the DNA helix really unfolds once it is adsorbed on Au, the bases that usually are hidden inside the double strand structure will get in contact with gold. Moreover, it is also possible for the bases themselves to be the promoters of the unfolding process, interacting with the surface despite their unfavored position. For all these reasons, we can not just neglect the study of single bases adsorption with the motivation that they are hidden inside the helix.

Once obtained the interaction energies and geometries through DFT calculations, it is possible to move to classical molecular dynamics simulations (MD). Different adsorption geometries analyzed with DFT calculations are reproduced through classical MD, tuning atom types LJ parameters in the FF to obtain the correct description of the system. The detailed structure of GolP FF, including the model with which polarization effects are taken into account and the choice of LJ parameters for specific atoms, is given in detail in Section 2.3.1. The validation of the FF is done through the comparison between ab initio, MD and experimental results on more complex systems.

Once the parametrization for the classical FF is completed, we are able to move to docking simulations of specific DNA molecules approaching to the Au(111) surface. In fact in our docking code gold surface is described similarly to how it is done in classical MD calculations, and LJ parameters for the interaction are inserted into the coarse-grained model, allowing to obtain a correct description of the adsorption of DNA oligomer on Au(111) (Section 2.4.2). Docking strategies are relatively inexpensive from the computational point of view, because the solvent is treated as a continuum dielectric medium and the approaching partners are rigid bodies. Thus, one normally performs many simulations, so that the system is treated statistically.

The analysis of the different adsorption geometries of DNA oligomer on Au(111) allows to understand the effect of the different components of the interaction, in particular for what concerns

the ratio between the electrostatic interaction between the backbone and the surface and the long range interaction between the hidden nucleobases and gold. Once the main interaction configurations are identified relying on docking simulations results, they can be used as starting geometries for classical MD simulations. In these final calculations atoms are let free to move in explicit solvent and the DNA structure is able to relax and possibly to unfold due to the hard surface.

In this chapter, all the different approximations and computational methods needed for this multi-step study are presented.

2.1 Many body electronic structure problem

In the study of a many body system the main problem it is how to balance an accurate description of the system under study and the computational cost of the calculations. Moreover, even without worrying of the time and resources, when dealing with a system including more than two bodies some approximations are required. What must be done, it is to choose accurately the method and approximations suitable for describing correctly the system under study.

For the study of the adsorption of nucleobases on Au(111), we decided to use a multi-step approach that relies both on quantum-mechanical methods, which treats explicitly the behavior of the electrons, and on classical molecular dynamics, that solves Newton equations of motions for nuclei, treating electrons behavior implicitly through a force field.

The starting point of both DFT and MD calculations is Born-Oppenheimer approximation[38], that allows to separate nuclei and electrons motion. The starting point for the description of a many-body system is the Schrödinger equation: lets consider a system of N nuclei described by the coordinates $R_1...R_N$=R and masses $M_1....M_N$, and N_e electrons described by the coordinates $r_1....r_n$=r, spin variables $s_1.....s_n$=s, with mass = m_e. The Hamiltonian of the system is given by:

$$\begin{aligned}\hat{H} = &\sum_{I=1}^{N} \frac{\hbar}{2M_I}\nabla_I^2 + \sum_{i=1}^{N_e} \frac{\hbar}{2m_e}\nabla_i^2 + \sum_{i>j} \frac{e^2}{4\pi\varepsilon_0}\frac{1}{|r_i - r_j|} + \\ &\sum_{I>J} \frac{e^2 Z_I Z_J}{4\pi\varepsilon_0}\frac{1}{|R_I - R_J|} - \sum_{i,I} \frac{e^2 Z_I}{|R_I - r_i|} = \\ &= T_N + T_e + V_{ee}(r) + V_{NN}(R) + V_{eN}(r,R)\end{aligned} \tag{2.1}$$

where $Z_I{}^*e$ is the charge on the I_{th} nucleus and ε_0=$8.85 * 10^{-12}$ C^2/J. To solve the complete quantum mechanical problem, time-independent Shrodinger equation must be solved:

$$[T_N + T_e + V_{ee}(r) + V_{NN}(R) + V_{eN}(r, R)]\Psi(x, R) = E\Psi(x, R) \tag{2.2}$$

where $x = (r, s)$ denotes the full collection of electron position and spin variables, and $\Psi(x, R)$ is an eingenfunction of $\hat{\mathrm{H}}$ with eingenvalue E.

The solution of the Schrödinger equation would describe exactly the system, but it is not possible to solve it without approximations when dealing with a many-body system. The fundamental idea of Born-Oppenheimer approximation is to separate the electronic motion and the nuclear motion. The physical justification for this assumption relies on the differences between the particles, that emerges in the difference in the masses or, from another point of view, in the spacial localization of the wave function. In a dynamical sense, the electrons can be regarded as particles that follow the nuclear motion adiabatically, that means that they are considered to remain in their ground state as the system evolves in space following the equations of motion.

A solution to Eq 2.2 can be proposed in the form:

$$\Psi(r, R) = \sum_n \psi_n(r, R)\chi_n(R) \tag{2.3}$$

where $\chi(R)$ is the wave function describing the evolution of the nuclei in each of the adiabatic electronic eingestates $\psi(r, R)$. $\psi(r, R)$ satisfy the time independent Schrödinger equation:

$$\hat{h}_e\psi(r, R) = [T^E + V^{NE} + V^{EE}]\psi_n^{r,R} = E^R\psi_n^{r,R} \tag{2.4}$$

Equation 2.4 has to be solved for all the different nuclear configurations. Substituting the above equation into the full Shrodinger equation we obtain:

$$\begin{aligned} &[i\hbar\frac{\delta}{\delta t} + \sum_{I=1}^{P}\frac{\hbar^2}{2M_I} - V_{nn} - E_q(R)]\chi(R) = \\ &-\sum_n\sum_I = 1^P\frac{\hbar^2}{2M_I} < \psi_q|\nabla_I^2|\psi_n > \chi_n(R, t) \\ &-2\sum_n\sum_{I=1}^{P}\frac{\hbar^2}{2M_I}\nabla_I\chi_n(R, t)\cdot < \psi_q|\nabla_I|\psi_n > \end{aligned} \tag{2.5}$$

From the equation it emerges that the separation of nuclei and electrons wave function is correct only if the off diagonal terms can be neglected. This condition depends from the electrons and

nuclei masses ratio, and its low value justifies this adiabatic approximation unless very small gaps occurs.

Starting from this results, nuclei positions can be fixed and a quantum mechanical approach can be adopted, treating explicitly the contribution of the electrons to the energy of the system and solving the electron Schrödinger equation; alternatively, the contribution of electrons to the energy can be averaged and the motion of nuclei can be treated classically, as they were moving into an external mean field.

In this thesis we first perform DFT calculations to study the interactions of the nucleobases with the surface, with the aim of obtaining information about the electronic structure and the interaction. This information allows the parametrization of a suitable FF for classical MD calculations and the study of larger systems.

In the following, the two computational methods are explained in details.

2.2 Density Functional Theory

2.2.1 Theory

Density functional theory is a quantum mechanical method able to model the electronic structure of many-body systems, widely adopted in the condensed matter physics and quantum chemistry communities as it is able to balance between accuracy and computational cost. Quantum mechanical methods deal with the electronic Shorodinger equation starting from first principles, without relying on empirical assumption. Electrons are considered explicitly; for this reason these methods have a very good accuracy but their applications are limited to systems of some hundred of atoms because of their computational cost. The central idea of DFT approach is to express the ground-state total energy as a functional of the charge density $\rho(r)$. In this way is possible to avoid solving the many-body wave function, which is a very complex object that requires long computational times to be solved, even with approximations, and to deal with a much more simple function, depending only on x, y and z. Kohn-Sham approximation is the most used to obtain an expression for the charge density. The central point of the approximation is not to deal with the true interacting system but with a non interacting one, taking care the potential acting on the electrons to be the same in the two systems. The shape of this

potential is unknown, in particular concerning exchange and correlation terms, and must be approximated in different ways depending on the systems under study.

Honemberg-Kohn theorems

The first Hohenberg-Kohn theorem (1964)[39] demonstrated that the electron density uniquely determines the external potential, and consequently the Hamiltonian operator and all the properties of the system.
The proof can easily be given assuming that there were two external potential $V_{ext}(r)$ and $V'_{ext}(r)$ differing more than a constant, each giving the same ρ(r). Conformly there are two Hamiltonians, $\hat{\mathrm{H}}$ and $\hat{\mathrm{H}}$', and wave functions ψ and ψ'. The variational principle states that

$$E_0 < \langle\psi'|\hat{H}|\psi'\rangle = \langle\psi'|\hat{H} - \hat{H}'|\psi'\rangle + \langle\psi'|\hat{H}'|\psi'\rangle = E'_0 + \rho(r)[V_{ext}(r) - V'_{ext}(r)]dr \tag{2.6}$$

where E_0 and E'_0 are the ground states energies for $\hat{\mathrm{H}}$ and $\hat{\mathrm{H}}$'. The same relation is valid starting with $E'_0 < \langle\psi|H'|\psi\rangle\rangle$. Adding the two equations the final result is: $E_0 + E'_0 < E'_0 + E_0$, that is a contradiction and consequently the initial assumption is confuted.

Now that it was demonstrated that electron density uniquely determines the potential, still it is to demonstrate that to a specific potential correspond only one density of the system. This statement is nothing but the variational principle:

$$E_0 \leq E[\rho'] = T[\rho'] + E_{Ne}[\rho'] + E_{ee}[\rho] \tag{2.7}$$

This means that to any trial density correspond an external potential and a system energy that is the ground state energy if and only if $\rho' = \rho$.
Thus the expression for the energy of the system can be written as a function of the density:

$$E[\rho] = T[\rho] + V_{ext}[\rho] + V_{ee}[\rho] = F_{HK}[\rho] + V_{ext}[\rho] \tag{2.8}$$

The interaction with the external potential is trivial:

$$V_{ext}[\rho] = \int V_{ext}\rho(r)dr \tag{2.9}$$

This theorem restricts density functional theory to studies of the ground state, as long as there is have a biounivoc relation between the potential and the density.

Eventually, relying on the two HK theorems the final statement of DFT can be written:

$$\Delta[E[\rho]] - \mu(\int \rho(r)dr - N)] = 0 \tag{2.10}$$

This equation tells us that the the ground state energy and density correspond to the minimum of a $E[\rho]$ functional, with the constraint that the density corresponds to the number of electrons in the system. μ is the Lagrange multiplier of the constraint and is known as the electronic chemical potential.
F_{HK} is an universal functional independent from the system, and knowing it would allow to solve the Shrodinger equation exactly. Since it is not known, different approximations must be done for it.

Kohn-Sham approximation

In searching an expression for F_{HK} the first thing to do is to write the kinetic energy, whose expression in terms of the electron density is not known. The main problem with kinetic energy and Thomas-Fermi approach is that the kinetic operator is non local. To solve this problem the system can be moved to a new one (S) of non interacting particles whose density is the same that the interacting ones[40].

$$T_S = -\frac{1}{2}\sum_i^N \langle\psi_i|\nabla^2|\psi_i\rangle \tag{2.11}$$

$$\rho_S = \sum_i N\psi_i = \rho \tag{2.12}$$

where ψ_i are the orbitals of the non interacting system and T_S is the kinetic energy of the non-interacting system. Of course T_S is different from the true kinetic energy of the system. The difference must be taken into account separately. The classical part of the electrostatic interaction can be explicitated, and it is written as the Hartree energy: $V_H[\rho] = \frac{1}{2}\int \frac{\rho(r_1)\rho(r_2)}{|r1-r2|}dr_1dr_2$.

In the end the following equation is obtained:

$$F[\rho] = T_S[\rho] + V_{ext}[\rho] + V_H[\rho] + E_{XC}[\rho] \tag{2.13}$$

where it has been introduced the exchange-correlation functional:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + V_{ee}[\rho] - V_H[\rho] \tag{2.14}$$

As can be seen E_{xc} includes the error made in using a non-interacting kinetic energy and the one made in treating the electron-electron interaction classically.

Eventually, the Kohn-Sham equation can be written:

$$[-\frac{1}{2}\nabla^2 + \int \frac{\rho(r')}{|r - r'|}dr' + V_{xc}]\psi(r) = \varepsilon_i\psi_i(r) \tag{2.15}$$

This set of equations describes the behavior of non-interacting electrons in an effective potential. The Kohn-Sham approach obtains the ground state energy and density of a system of non interacting electrons, with an exact correspondence with the real interacting system under study. The correspondence is exact only if the functional is known exactly, but this is not possible in such a system because of the greater computational effort that would be needed. Different approximations to the functional can be done to solve the equations, which allow to obtain excellent results for different systems if the appropriate functional is adopted.

2.2.2 Implementation

Once the method is set up, different approximations can be used to solve Kohn-Sham equation, starting from the simplest one, that teats the system as homogeneous, and moving to more complex ones, including hybrid functionals where Hartree-Fock approximation is used and exchange-correlation (xc) functionals derived from experimental data.
The main problem is always to find a balance between accuracy and resources needed, searching for a useful description of the system in a proper time. In the following, approximations used for DFT calculations in this work are presented.

LDA and GGA

Here we want to summarize brefly the first and more simple approximation which can be used to solve Kohn-Sham equation, the Local Density Approximation (LDA) [40] and the further improvement to it that we used for our calculations on DNA and surfaces, the Generalized Gradient Approximation (GGA).

LDA: as the name suggests, LDA express the energy functional in terms of local functions of the density. This approximation is derived from the calculations for the homogeneous electron gas. In this systems the electrons are subject to a constant external potential and thus the charge density is constant, $\rho = N/V$.

Using the kinetic and exchange energy densities of the non-interacting homogeneous electron gas and approximating the functional as an integral over a local function of the charge density this expression can be derived:

$$T[\rho] = 2.87 \int \rho^5/3(r)dr \tag{2.16}$$

while for xc functional a combination of Dirac form for exchange and a quantum Monte Carlo simulation for correlation are used.

Strictly, the LDA approximation is valid only for slowly varying densities, but different calculations on atoms, molecules, and solids showed that it can also be applied to these systems. In particular, much current understanding of metal surfaces comes from LDA simulations.

GGA: if LDA is thought to be 0^{th} order expansion of the density matrix in terms of the density, the more obvious improvement is the gradient expansion approximation (GEA) in which first order gradient terms in the expansion are included. The typical form of a GGA functional is:

$$E_{xc} = \int \rho(r)\varepsilon_{xc}(\rho, \nabla\rho(r))dr \tag{2.17}$$

In general GGA functional tends to improve total energies [41], atomization energies [41, 42, 43], energy barriers and structural energy differences [44] with respect to LDA functional.

A number of different functionals have been developed within the GGA family. The analysis of all these functional is beyond the purpose of this introduction. It is enough to know that each of them is more or less suitable for different systems. In particular the GGA functionals in which we are interested in our discussion are Pederew-Burke-Ernzerhof (PBE) [45], and some of its

derivations [46]. The energy functional is still dependent only from r, so the approximation is still dealing with a local expression of the density and the *correlation* part of the *xc* functional is very poorly taken into account.

Our point is that long range interactions are thought to be the main term in the interaction between DNA and surfaces. Until the recent development of non local functionals, it was not possible to take them into account in DFT calculations. For this reason in the past years dispersion interactions they have been mainly neglected, supposing the interaction of the single bases (i.e. without backbone) with surfaces to be weak enough to be ignored or evaluating them only in MD calculations, not at the QM level [47, 27]. This assumption can be perhaps justified when dealing with the DNA double helix, where electrostatic is the main source of interaction, but we can not accept it as we want to catch the possibility of an unfolding process on the surface, that would expose the bases, usually kept inside the helix, to a direct contact with gold.

For this reason, we decided to exploit a recently developed functional [48] that includes vdW interaction into the calculations, relying on first principles instead than on empirical observation. The approximations on which this functional relies are explained in the next section.

vdW-DF functional

Different functionals belongs to the vdW-DF family [48, 49, 50, 51, 52, 53], and each of them treats the calculation of exchange and correlation in a slightly different way. Nevertheless the structure of the approximation and the motivations for it are common between the methods, and we are going to expose them in this section.

The first things to worry about is the way in which the calculation for exchange energy is dealt. In fact, since correlation is calculated separately and explicitly, a valid approximation for exchange interaction alone must be adopted. A discussion was carried on that often GGA functional overestimates the binding due to the exchange interaction, and this feature compensate the missing of correlation in the *xc* term[54]. Here a specific expression for correlation is adopted, consequently an exchange functional the less overestimating as possible must be chosen. This choice is based on considerations that differs from the ones that usually influence it in a DFT calculation. Different choices can be done for different systems and we will test them

in details (Chapter 3). It is enough to say that the first choice in the construction of vdW-DF functional by its authors fell on revPBE functional [46].
Concerning the calculation of correlation energy, the first step is to divide it in two terms, a local one and a non local one.

$$E_c = E_c^0[\rho] + E_c^{nl}[\rho] \tag{2.18}$$

The non-local term is defined to include the longest-ranged or most non-local terms that give the van der Waals interaction and it approaches zero in the limit of a slowly varying system. The first term is also non-local, but approaches the LDA in the limit of a slowly varying system. Different approximations are made for the two terms. In particular, LDA approximation is used for $E_c^0[\rho]$ as the principal longest range terms are included into the non-local parto of correlation energy. Concerning $E_c^{nl}[\rho]$ the main argument for the approximation is that it is less sensitive to the details of the system's dielectric response than the short range terms, and for this reason a very simple approximation can be chosen for the dielectric function [48].
The definition of E_c^{nl}, derived from Poisson equation, is

$$E_c^{nl} = \int_0^\infty \frac{du}{2\pi} tr[ln(1 - V\widetilde{\chi} - ln\epsilon] \tag{2.19}$$

Where $\widetilde{\chi}= \widetilde{\chi}(r, r')$ is the density response to a fully self-consistent potential while V=V(r,r') is the inter-electronic Coulomb interaction, ϵ is an approximated dielectric function and u is the imaginary frequency. For the details of the derivation we refer to literature [48].
vdW-DF functional and its derivate gives very good results for a lot of different systems, including surfaces and DNA interacting with molecules and, more recently, also molecules and DNA on surfaces [55, 56, 57]. Since few specific calculations for molecules adsorbed on surfaces were done when we began to work on our system, we performed different tests on different adsorbed systems (Chapter 3) and our final conclusion is that revPBE-vdW-DF is the functional the most suitable for our purposes.

Plane waves and Pseudopotentials

Once an approximation for the *xc* functional has been chosen, Kohn-Sham equation must be solved for the selected system.

The wavefunction are usually represented as a linear combination of basis functions.

$$\psi_i = \sum C_{ij}\varphi_j \tag{2.20}$$

At this point two possible choices can be done: plane waves basis set and localized basis set. Both the choices have different pros and cons. In particular for our system we chose plane waves (PW) basis set that, between the others, have the advantage of the periodicity. In fact, since we are dealing with the interaction of a molecule with a surface which in principles should be infinite, having a periodic system allows to reduce the dimension of the simulation box, as the molecule can interact also with Au atoms belonging to nearby cells.

Other advantages of PW are the fact that they are orthogonal, the independence from atomic positions, the possibility to use Fast Fourier Transform to solve the equations. The fundamental choice that has to be done in dealing with basis set, is where to stop the summation in Equation 2.20 that is in principle infinite. Concerning PW, one parameter is enough to define the number of plane waves to be used, which is the energy cutoff. Kinetic energy associated with each plane wave is:

$$-\frac{1}{2}\nabla^2\varphi_G(r) = \frac{1}{2}||G||^2\varphi_G(r) \tag{2.21}$$

with

$$\varphi_G(r) = \frac{1}{\sqrt{\Omega}}e^{iG\cdot r} \tag{2.22}$$

The cutoff energy is $E_{cutoff} = \frac{G^2_{max}}{2}$ and increasing its value it is possible to achieve the desired convergence on the energy of the system depending on the dimension of the simulation cell and other parameters. A second parameter that must be treated carefully in the set up of the simulation is the k-point number. Once G_{max} has been chosen the volume of the sphere containing all the selected PW will be a function of G: $V_{spere} = \frac{4\pi}{3}G^3_{max}$.

Kohn-Sham equation should be solved for each value of r, but this is obviously impossible since points in space are infinite. For this reason sample points must be chosen for solving the equation.

Plane waves basis set allows to move to reciprocal space to solve the equations and, as we know, the volume of the reciprocal cell is in a relation of reciprocity with the dimension in the real space. This allows to reduce the number of points in which solving the equation while increasing the dimension of the system.

Eventually, a particularity of the plane waves approach is the description of the core electrons. In fact core electrons wave functions vary very rapidly. Consequently a proper description of the system can be obtained only increasing the energy cutoff. A possible solution is to use pseudopotential (PP) for the description of core states. Since core electrons do not participate in bonding, excitation and conductivity, they are not treated explicitly anymore, but their effect on the system is described through a potential, called pseudopotential, which is substituted to the nuclei Coulombic potential and which takes into account both the effects of the nuclei and of the core electrons.

2.3 Classical Molecular Dynamics

Classical molecular dynamics can deal with larger systems than quantum mechanical methods, involving easily more than ten thousand of atoms, due to the fact that it treats the electronic interaction through a parametrization. System studied with MD are macroscopic and comparable with samples measured in experiments, and are characterized by a huge number of degrees of freedom, both of the main object of the system itself and of the environment that surrounds it. In MD simulation the system is described through the solution of a time evolution equation discretized in consecutive time intervals. At each time steps position and velocities are calculated and the system is allowed to sample in time the phase space, giving as a result a trajectory driven by the equations of motion.

This approach is in contrast with the stochastic one, which is proper, for example, of Monte Carlo method. In these methods the configuration space is sampled through a random variation of its degrees of freedom. Eventually calculations performed with the two methods lead to the same values for average quantities, assuming that the time (MD) or the phase space sampling (MC) are long enough.

In the following, classical molecular dynamic method is presented in details.

2.3.1 Theory

General aspects

The starting point of classical MD is again Born-Openheimer approximation: the contribution of electrons is treated implicitly, and this can be done in a proper way with DFT calculations, and then the motion of nuclei is treated classically.
Classical molecular dynamics is based on Newtonian equations of motions. For each particle I in the system, and with particle here we mean nuclei, it states:

$$F_I = M_I R_I \tag{2.23}$$

where R_I and M_I are the position and mass of the I_{th} particle and F_I is the force acting on it. In a classical molecular dynamic simulation all the different forces applied on each atom of the system are not taken separately into account. For such a reason it is better to rewrite the equations of motion depending on the potential acting on the system. At this point the potential can be approximated keeping in mind two main purposes: the first is to have forces on each particle as close as possible to the real ones acting on them; the second is, as usual, to reduce the computational cost of the calculation.
The equations of motions explicitly depends on the potential and the momentum of the particle and can be written in the following way:

$$F_I = -\frac{\partial V^C(R_I)}{\partial R_I} \tag{2.24}$$

$$P_I = -\frac{\partial V^C(R_I)}{\partial R_I} \tag{2.25}$$

$$P_I = M_I R_I \tag{2.26}$$

where P_I is the linear momentum of the I_{th} particle and $V^C(R_I)$ is the classical potential function of the system.
In classical molecular dynamics simulations, the potential energy function, which is proper of the system and depends on the interactions between the particles, is described by mathematical functions that approximate the different interactions and by a set of parameters that specifies its value for the different atoms and molecules. The combination of these two component is called

the *Force Field* (FF). Different kind of systems and interaction require different approximations for the FF. The advantage of classical molecular dynamics with respect to a quantum mechanical approach is, as it was already said, the possibility of dealing with larger systems, without taking into account explicitly the effects of the electrons. On the other hand, whenever the quantum nature of the electrons and nuclei composing the system becomes the main responsible of the investigated process, classical descriptions dramatically fails. Examples of such situations are tunneling processes and high frequency vibrational modes.

Force Fields

Two steps are involved in generating a force field: the modeling of the interactions and their parametrization. The choices done for these two steps are determinant for the correct description of the system. First of all the model for each energy term must be defined; as an example, harmonic functions are usually employed to reproduce chemical bonding between atoms, while the non bonding dispersion forces are usually described through a combination of power functions. Starting from these simple approximations, more terms can be added to improve the description of the system (e.g. including three body interactions) depending on the kind of system under study. Once the different component have been described, a parametrization is needed (λ), that means that for each atom/couple/etc all the parameters of the interaction must be given (e.g. bonding length, harmonic constant etc.). Finally, a topology file is needed (T), where the atoms belonging to the same molecules are grouped. In this way they are defined as interacting trough the *bonding* potential (V_B). *Non-bonding* interactions are also linked to the topology file (V_{NB}) since for atom pairs already interacting through a bonding potential they are usually neglected or reduced.
A typical Force Field belonging to the simplest class of potentials (Class I) is expressed in the following way:

$$\begin{aligned} &V(\lambda,T) = V_B(\lambda,T) + V_{NB}(\lambda,T) = \\ &V_{bnd}(\lambda,T) + V_{ang}(\lambda,T) + V_{dih}(\lambda,T) + V_{D,R}(\lambda,T) + V_{Cou}(\lambda,T) \end{aligned} \tag{2.27}$$

The bonding potential is made up of several contributions: a two-particle interaction describing the bonds between atoms (V_{bnd}), a three-particle interaction for angular contributions (V_{ang})

defined by three atoms connected by two consecutive bonds, a four-particle interaction (V_{dih}) that accounts for proper dihedrals contribution, that are defined by four atoms connected by three consecutive bonds, and an improper dihedrals contribution that allows to keep four atoms on the same plane (Figure 2.2). Usually bond and angle potentials have an harmonic form. Class I force fields are of harmonic type:

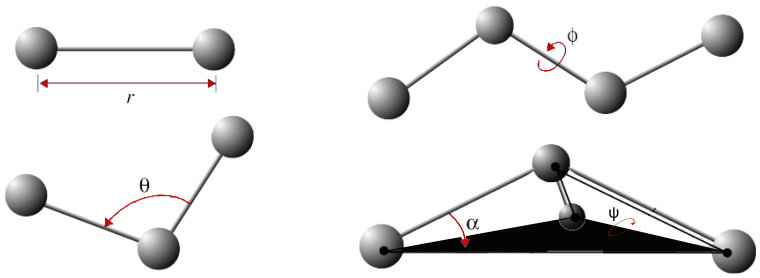


Figure 2.2: a) bonding term b) angle term c) proper dihedral term d) improper dihedral term

$$V_{bdn}(b_0, k_b) = \sum_{i}^{N_b} k_i (b_i - b_{0_i})^2 \tag{2.28}$$

$$V_{ang}(\theta_0, k_\theta) = \sum_{j}^{N_\theta} k_j (\theta_j - \theta_{j_0})^2 \tag{2.29}$$

where b_0, k_b, θ_0, k_θ are the parameters vectors for the single atoms and the couples.

For the description of solids with classical potentials, no other bonding terms are needed. When dealing with molecules, on the other hand, one must describe also the rotations of the different parts of the molecule the ones with respect to the others. This effect is described through the *torsion* potential, that can also be called *dihedral* potential. This potential can be divided in two part, *proper* and *improper dihedral* potential. Usually the proper term is made of a sum of a periodic functions while the improper term is made up of a sum of an harmonic functions. The final expression for it varies between one force field and the other.

$$V_{dih}(\lambda_{pd}, \phi_0, k_\phi = V_{pd}(\lambda_{pd}) + \sum_{l}^{N_{id}} k_l (\phi_l - \phi_{0l})^2 \tag{2.30}$$

where *id* stands for improper dihedrals and *pd* for proper dihedrals.

The *non bonding* term describes the Coulomb and the vdW interaction.

$$V_{non-bonding} = V_{Cou}(q,\epsilon) + V_{D,R}(\sigma,\epsilon) = \sum_{i,j<i}^{N_p} \frac{q_i q_j}{\epsilon |r_i - r_j|} + \sum_{i,j<i}^{N_p} [(\frac{\sigma_{ij}}{|r_i - r_j|})^{12} - (\frac{\sigma_{ij}}{|r_i - r_{j|}})^{6}] \tag{2.31}$$

where q is the vector containing all the atom charges, while ϵ is a screening constant. Concerning vdW interaction, ϵ and σ are vectors collecting all LJ parameters for each atom type. In particular, looking at the LJ potential shown in figure Figure 2.3, ϵ is the well depth while σ is the collision diameter, that is the distance at which the potential intersects the x axes.
Since we will talk more of LJ potential in the following, modifying the parametrization to fit our DFT results, it is important to point out that practically the values of σ and ϵ in the LJ potential expression are obtained combining LJ parameters defined in the topology for single atom types (e.g. O atoms bounded to C atoms, O atoms of water molecules etc.), using mixing rules, that are different for different force fields (e.g. Lorentz-Berthelot mixing rules [58, 59]).

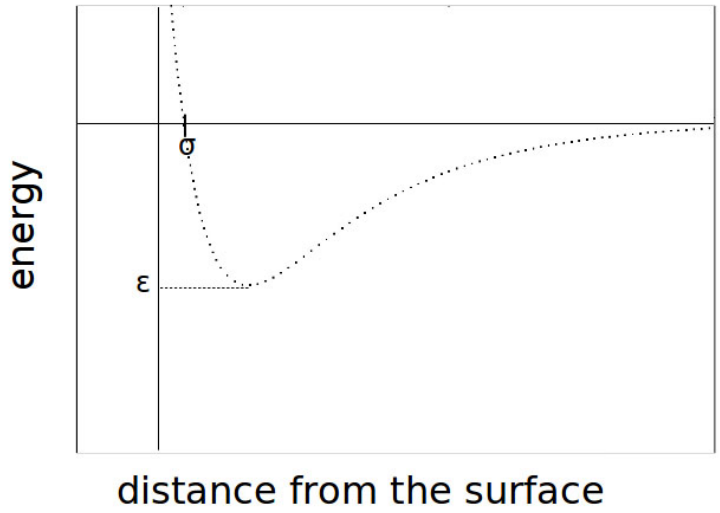


Figure 2.3: Shape of LJ potential

GolP model for Au(111) surface

Concerning the study of our specific system, we want to describe here the particular structure of the FF that was used as a starting point for the description of gold surface, GolP [34]. In fact with the aim of parameterizing a FF suitable for the description of the interaction between DNA and Au(111), we decided to use AMBER FF [60, 61] for the description of DNA molecules and GolP FF for the description of Au(111) surface, adding special parameters for the interaction between DNA and gold. AMBER FF was chosen among the others for the accuracy of the results obtained with this FF on systems involving nucleic acids and various DNA molecules. Our group has tested AMBER parm99 [62] against CHARMM for quadruplex DNA, finding a similar behavior over the length scale of 5 ns [63]. For double stranded DNA we do not expect a better performance of a different FF, and we are confident that our result are not sensitively dependent on the choice of the force field, or at least that no better force fields have been assessed so far.

GolP FF was developed in our group for studying the adsorption of proteins on the Au(111) surface and its parametrization has been modified for the description of DNA adsorption, maintaining its structure as it is suitable also for the description of this new system.

The potential energy function for the interaction between DNA bases and Au(111) surface is obtained as:

$$V^{tot}_{Au-DNA} = V^{im}_{Au-DNA} + V^{vdW}_{Au-DNA} + V^{chemisorb}_{Au-DNA} \tag{2.32}$$

V^{vdW}_{Au-DN} and $V^{chemisorb}_{Au-DNA}$ are the van der Waals and chemisorption interaction between the DNA molecule and gold, while V^{im}_{Au-DNA} is the Coulombic electrostatic interaction between the molecule and the image charges induced in gold. The surface polarization effect is described through the substitution of Au atoms with a dipole made up by two, opposite charges, connected by a rigid rod free to rotate[64]. In the absence of a molecule on the surface the average dipole moment is equal to zero, with the initial orientation of the dipoles chosen randomly. When a molecule lies above the surface with a partial atomic charge on its atoms, it forces the rods to rotate in a specific orientation, modifying the average dipole moment. vdW interactions are described through a sum of 12-6 Lennhard-Jones (LJ) interactions (2.33 between each DNA atom and properly placed interaction sites on the gold surface. In fact the other main feature

of GolP structure is the inclusion of gold virtual sites in hollow positions on the surface.

$$V_{vdW}(r_{ij}) = 4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - \frac{\sigma_{ij}}{r_{ij}})^{6}] \tag{2.33}$$

These atoms are the ones interacting with adsorbing species via the LJ terms, while for real gold atoms LJ parameters are equal to zero. Instead dipolar rods are located on each gold atom, real and virtual. This structure allows the adsorption of small molecules on top adsorption sites instead than on hollow ones. DNA bases were found to adsorb preferentially on top adsorption sites [65, 66], and for this reason we maintained the virtual site structure for our FF. If no particular interaction is defined, LJ parameters for gold atoms (σ_{Au} , ε_{Au}) are defined and combined with LJ parameters for DNA atom types, in accordance with AMBER Lorentz-Berthelot mixing rules [58, 59] ($\sigma_{ij} = (\sigma_i + \sigma_j)/2, \varepsilon_{ij} = (\varepsilon_i * \varepsilon_j)^{1/2}$). Such LJ interactions account for dispersion, repulsion and weak chemisorption effects.

2.3.2 Implementation

Interactions calculation methods and periodic boundary conditions

The main problem in simulating a macroscopic system is choosing an appropriate size for the calculation of non-bonded interaction and for a correct representation of the system, since the more the number of atoms and the larger the dimension of the cell, the more computationally expensive becomes the simulation.

As we already pointed out for DFT calculations, the first possible solution to this problem is to apply periodic boundary conditions. First of all, unless the behaviors near the walls (surface effects) are of interest, this choice allows to eliminate them and to simulate the bulk material, either dealing with a surface, with a monolayer or any kind of periodic system. Molecules in solution are allowed to freely enter and leave the cell, being substituted by their images that simultaneously enter the cell through the opposite face.

However, to avoid artifacts that may be introduced by periodic boundary conditions (e.g. interaction between images of isolated molecules, as for DFT, or errors in systems with high spatial correlation length) test calculations must be performed to choose the appropriate cell

dimension. Other expedients to use with periodic boundary conditions are the minimum image convention and the introduction of a cutoff radius, that is needed for dealing with short-range forces. The minimum image convention states that each particle interacts with only one mirror image of the other particles, that is the one at minor distance. Then it is possible to define a cutoff radius for short-range interactions, eliminating also the dependence on the box shape, and including into the calculation only the interactions with particles within the sphere defined by the cutoff radius. To be consistent with the minimum image convention the cutoff radius must be shorter than half of the shortest box vector.

Concerning the electrostatic interaction, the two approximations we just explained are not suitable for it since the r^{-1} dependence does not have a fast convergence. Moreover in evaluating it, one must deal with the fact that the Coulomb potential is both slowly decaying at high distances and strongly varying at small distances. This means that an accurate description must be provided, but on the other hand it becomes expensive since the summation can not be truncated at large distances. Several methods have been developed for the treating of this contribution at affordable computational costs. Methods based on Fourier-transform are now available on most of the MD codes and work together with the periodic boundary conditions, as they need a periodic system to be applied. The exploited idea is the one on Edwal summation, that rewrites the interaction potential as the sum of two terms:

$$\varphi(\mathbf{r}) \stackrel{\text{def}}{=} \varphi_{sr}(\mathbf{r}) + \varphi_{\ell r}(\mathbf{r}) \tag{2.34}$$

where $\varphi_{sr}(\mathbf{r})$ represents the short-range term that sums quickly in real space and $\varphi_{\ell r}(\mathbf{r})$ represents the long-range term that sums quickly in Fourier space. While the short-range term is easily evaluable defining a cutoff radius similar to the one of LJ interaction, the long-ranged part should be finite for all arguments (most notably r = 0) but may have any convenient mathematical form, most typically a Gaussian distribution. The evaluation of the long range interaction is the more difficult point of the calculation, but several methods are available for dealing with it, as Particle-Mesh Ewald (PME) [67, 68] and Particle-Particle Particle-Mesh (PPPM) [69], have been widely used in the fields of computational chemistry and biochemistry and are available on most of the MD codes.

Integrators

Once the information needed for the simulation is provided, namely the dimension and geometry of the simulation box, the force field shape and parameters, the topology and the initial coordinates of the atoms, the length of the MD simulation is chosen, defining the dimension of the time steps dt and the number them, and the MD algorithm calculates the position and velocity of all the particle at every time step, eventually producing a trajectory for the molecules. The value of the time step depends on the time scales of the system, in particularly it must be in accordance with the one of the fastest motions determined by the force fild (i.e. stretching and bending of the lighter particles). The iterative procedure performed at each time step consists in the following steps:

- *Compute forces* ($-\bigtriangledown V^C(r)$) deriving the potential energy functions and calculate them in the initial positions;

- *Integrate the equations of motions* (Eq. 2.23) from forces, previous coordinates and velocities;

The numerical integration of the equation of motion should be done for a continuous time, but as it was already said it is approximated to discrete time steps. The approximation is the responsible for the accuracy of the simulation results, and must reproduce the correct final trajectory and be able to conserve the energy despite the perturbation induced by the discretization.
The integration is based on the Taylor expansion of the position of each particle as a function of time[70]:

$$r_i(t+dt) = r_i(t)dt - \frac{-\bigtriangledown_{r_i} V^C(r(t)}{2m_i}dt^2 + \frac{\dot{r}_i}{3!}dt^3 + O(dt^4) \tag{2.35}$$

where the equation has been truncated to the third order.
Different numerical integrators have been developed based on Eq. 2.35. The simplest model belonging to the Verlet family of algorithms allows to obtain the position of the particle at time $t+dt$ from the ones at time t and $t-dt$ with an error of the order dt^4 (Eq. 2.36).

$$r(t+dt) = 2r(t) - r(t-dt) - \frac{-\bigtriangledown V^C(r(t)}{2m}dt^2 + O(dt^4) \tag{2.36}$$

Velocities are then updated separately. This leads to different accuracies for positions and momenta and may cause drift in the calculations.
Further computational schemes have been developed to improve the accuracy on the calculations. In particular the Leap-Frog algorithm was chosen. In this algorithm the velocities are first calculated at time $t + \frac{1}{2}\delta t$; these are used to calculate the positions at time $t + \delta t$. In this way the velocities leap over the positions, then the positions leap over the velocities. In such a way the velocities are calculated explicitly, even if they are not calculated at the same time.

$$v(t) = \frac{1}{2}[v(t - \frac{1}{2}\delta t) + v(t + \frac{1}{2}\delta t)] \tag{2.37}$$

and

$$r(t + \delta t)) = r(t) + v(t + \frac{1}{2}\delta t)\delta t \tag{2.38}$$

$$v(t + \frac{1}{2}\delta t) = v(t - \frac{1}{2}\delta t) + a(t)\delta t \tag{2.39}$$

It can be shown that the global error on position remains of the order of dt^2 [115], and hence Verlet like algorithms are classified as second order integrators.

Different ensambles

Equation 2.36 is suitable for working in a microcanonical ensemble, where the total energy, number of particles and volume are preserver through the simulation (NVE). It may be considered as the natural ensemble for molecular dynamics simulations, as no time-dependent external forces are considered and the Hamiltonian of the system is constant.
It is also possible to perform simulations in different thermodynamic ensembles, as NPT and NVT, but it is needed to define specific equations of motion which preserve the respective quantities .

2.4 Docking simulations

As it was already said, to describe the behavior of DNA molecules adsorbed on a hard substrate, classical MD is the best tool to be used. However all-atoms molecular modeling methods are extremely computationally expensive and this means that are limited in the sampling of length and even more important time scales (10-100 ns), while experimental studies often, as in our case, give an averaged behavior of molecules over millisecond or longer. The main cause of the computational cost of MD simulation has to be attributed to solvent molecules, as they are treated explicitly as all the other atoms. Another source of cost, caused not by the solvent but by the solute itself, is dealing with the coordinates of all the atoms, allowing the molecules to modify their shape stretching and bending the bonding between atoms. It is easy to understand that to speed up the calculations some degrees of freedom must be taken away. First of all it is possible to treat the solvent implicitly, parameterizing and modeling its effects on the molecules, and then to freeze some degrees of freedom of the molecule itself, for example constraining its internal coordinates.

This approach allows to speed up the simulation when the molecule is still approaching to the surface and looking for an overall interaction geometry. Once this is done, variations of DNA structure due to the interaction with the surface can be studied with MD, allowing all the coordinates to relax. Docking is a method that predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. In our case we are dealing with a molecule and a surface, and this system it is even simpler to describe. In fact we do not need to search for a formed complex, and to define new parameters for it, but we just want to make DNA approaching to the surface and take note of the possible adsorbed configurations, defining an adsorption distance between the two parts of the system. Since we project to let the helix completely relax with MD simulations, we don't need the internal geometry of DNA molecule to be able to relax: all the internal coordinates are then constrained and the molecule moves in implicit solvent, counting the different possible adsorption geometries on Au and the relatives energy values.

In this approach the molecule is let free to move into the solvent following Brownian Dynamics motion equations (Section 2.4.1), driven by the interaction with the surface. We decided to use ProMetCS model, that was developed in collaboration with our group for the simulation

of the interaction between proteins and gold surface, adapting it for the case of DNA[71]. Our choice is justified by the fact that, even if different models have been developed for simulations of molecules in solutions[72], these are not suitable for describing the interaction of a molecule with a surface[73, 74]. Moreover, different surfaces were found to behave in a qualitative different way (e. g. on metals a barrier is formed by the energetically unfavored displacement of the water molecules) when the interacting molecule is adsorbed[75]. This means that a specific model should be developed for Au(111) surface and it should include a microscopic characterization of the process at the protein-surface interface. Gold structure has been modeled for the adsorption of proteins similarly to GolP structure. While electrostatic interaction is calculated with general approximations suitable for any molecule adsorbed on different surfaces, LJ interaction parametrization is taken from MD calculations specifically for DNA.

2.4.1 Simulation of Diffusional Motion by Brownian Dynamics

Brownian motion is the motion of particles suspended in a fluid, that means that their mass and size are larger than those of the molecules of the solvent. These particles seem to move randomly, because of the stochastic collisions with the solvent molecules. The result of such interaction is diffusion through the liquid. The dynamics of diffusional motion are described by the Langevin equations (Eq. 2.40).

$$m\frac{d^2r}{dt^2} = -m\eta\frac{dx}{dt} + F_{random} \tag{2.40}$$

In particular, Brownian motion correspond to the limit where the inertia tends to zero, and no average acceleration takes place during the simulation run. Such motion was first observed by Brown in 1872 and formalized by Einstein[76] and Smoluchowsky[77] that showed that the displacement of a particle undergoing Brownian Motion in time Δt is given by:

$$\langle \Delta r^2 \rangle = 6D\Delta t \tag{2.41}$$

whit D the translational diffusion coefficient of the particle

$$D = \frac{k_b T}{6\pi\eta a} \tag{2.42}$$

where k_b is the Boltzman constant, T is the temperature, η is the solvent viscosity and a is the hydrodynamic radius of the particle.

In our case the Langevin equation is solved following the method of Ermak and McCammon [78] for simulation of the motion of Brownian particles. The time is discretized in intervals of Δt and the trajectory is made by the coordinates at each of them.

The translational displacement Δr of each particle is obtained as:

$$\Delta r = (k_b T)^{-1} DF\Delta t + R \tag{2.43}$$

where F is the force systematically applied by the system on the particle, while R is a random vector simulating the effect of the solvent. The random vector must satisfy $\langle R\rangle = 0$ and $\langle R^2\rangle = 6D\delta t$. The rotational displacement angle at each time step is calculated as:

$$\Delta w = (k_b T)^{-1} D_R T\Delta t + W \tag{2.44}$$

where T is the torque acting on the particle, D_R is the rotational diffusion constant of the particle and W is a random rotation angle that must satisfy the conditions $\langle W\rangle = 0$ and $\langle W^2\rangle = 6D_R\Delta t$

2.4.2 Interaction Energy Function

In ProMetCS model the interaction energy function is the sum of three separate contributions:

$$U = E_{LJ} + U_{EP} + U_{desolv} \tag{2.45}$$

that are respectively LJ energy, electrostatic energy and desolvation energy.

LJ interaction: E_{LJ} includes nonpolar and van der Waals interactions between the molecule and the surface. The LJ components is calculated with the same model used for classical MD calculations, as the force field parametrization is moved here without modifications, introducing σ and ϵ parameters for Au atoms and special interaction parameters when it is needed. This means also that the potential is calculated as a sum of Lennard-Jones 12-6 terms and that

the surface is represented as in GolP force field: the physical position of each Au atom in the upper layer of the gold cluster is replaced by two virtual atoms that occupy hollow sites. Image interaction is taken into account separately, consequently the rod model is not used here.
Calculating LJ interaction directly for each atom pair would be too expensive. Therefore LJ interaction between the molecule and the surface is saved on the nodes of a three-dimensional grid with the origin placed at the molecule centre.

Electrostatic free energy: U_{EP} is the electrostatic interaction free energy in aqueous solvent and is represented by the interaction between real charges and their image charges, placed in a symmetric position with respect to the surface. To satisfy the boundary conditions of having zero surface potential of the metal, the electrostatic potentials of the molecule and of its image must cancel each other at the surface plane. This means that the image molecule must be the exact reproduction of the real molecule with opposite charges on it, including the surroundings. In particular, as molecules are represented as charge distributions immersed in a low dielectric cavity, also this one must be reproduced for the image. Molecules are then separated by another low dielectric cavity that surrounds the surface, and the three are immersed in a high dielectric solution.
As for the LJ case computing the electrostatic interaction free energy of two macromolecules at each time step is too computationally expensive. For this reason it is initially calculated solving the numerically the Poisson-Boltzmann equation and then Effective Charge approximation for Macromolecules (ECM) is applied [79]. Just to summarize the main point of this method, electrostatic interaction is the sum of 4 terms, equal two by two since the two molecules are one the image of the other.

$$U_{EP} = 1/2U_{mol} + 1/2U_{im} + U_{mol-cavity} + U_{im-cavity} = U_{mol} + 2U_{mol-cavity} \tag{2.46}$$

The first term, U_{mol} is the interaction energy of the molecule charges with the image electrostatic potential computed in the presence of molecule, image and metal cavities. U_{im} is equal to U_{mol}, as the image and the molecule have equal charges with opposite signs. $U_{mol-cavity}$ (and U_{im_cavity} as well) describes the perturbation of the molecule electrostatic potential by the low-dielectric cavity of the image [71], and is called *electrostatic desolvation*.
As it was already said, to solve the Poisson-Boltzman equation at each step is too expensive.

For this reason the two potentials, molecule and image, are calculated and the P-B equation is solved once. Than the solutions of the equation are stored on a three-dimensional grid centered on the geometric center of the molecule. Finally effective charges q^{eff} are calculated for the molecules, imposing that the potential obtained from effective charges immersed in a uniform high dielectric medium is the same as the one of the real charge distribution inside a low dielectric cavity [79]. The electrostatic energy U_{mol} can then be calculated as the interaction of q^{eff} immersed in a uniform solvent with the electrostatic potential of the image, stored on the nodes of the electrostatic grid.

$$U_{mol} = \sum_i \Phi^{im}(r_i^{im}) q_i^{eff} \tag{2.47}$$

This energy term does not take into account the effect of the cavities around the molecule and image and around the surface. $U_{mol-cavity}$ counts for this effect when the molecule is still far from the surface. $U_{mol-cavity}$ can be accounted by the introduction of a positive energy term Φ_{ed}[80]. In particular the effect of the metal cavity is 2^4 times larger than the one of the image cavity, so only the first one is taken.

$$U_{mol-cavity} = \sum_i \Phi_{ed}^{met}(r_i) x(q_i^{eff}) = \alpha \frac{\varepsilon_s - \varepsilon_m}{\varepsilon_s(2\varepsilon_s + \varepsilon_m} \sum_j (1 + kr_{ij})^2 exp(-2kr_{ij} \frac{a_j^3}{r_{ij}^4}) x(q_i^{eff}) \tag{2.48}$$

where k is the Debye-Hückel parameter, that accounts for the ionic concentration, a choice in which will be interested in our system. ε_m is the molecule dielectric constant while ε_s is the one of the solvent; a_j is the vdW radius of the *jth* atom of the metal surface and r_{ij} is the distance between this *jth* atom and q^{eff} of the molecule. α is a scaling factor estimated to be 1.67.
Moreover, when the molecule is closer that 5.5 Å to the surface the cavity around the molecule and the one around the surface starts to merge, and no solvent is anymore in between them screening the charge-metal interaction. This means that another correction must be done, lowering the unfavoring positive term by the unscreened electrostatic interaction component between the molecule and the surface.
The final expression for the electrostatic desolvation energy is[71]:

$$U_{mol-cavity}^{corr} = U_{mol-cavity} + \sum_i \frac{(q_i^{eff})^2}{2z_i(4\pi\varepsilon_0)} (\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon(z_i)}) \tag{2.49}$$

Protein Hydrophobic Desolvation Energy: U_{desolv} finally is the term describing desol-

vation effects, that is the free energy change arising from molecule-water, surface-water and water-water interactions. It is divided in two terms, the one of the molecule and the one of the surface. Both the terms are proportional to the solvent accessible surface area of the molecule (SASA) or surface S^{desolv} through an energy coefficient [79, 71].

$$U_{desolv}^{mol} = \sum_{m} \Phi_{pd} SASA_m \tag{2.50}$$

$$U_{desolv}^{met} = \sum_{i} \Phi_{met-des} S^{desolv} \tag{2.51}$$

The energy potentials Φ_{pd} and $\Phi_{met-des}$ are computed on a three-dimensional grid and are defined as a function of the distance r from the van der Waals surface of the molecule [79] or z_i, the distance between the centre of the molecule surface atom i and the metal surface [71].

Chapter 3

Van der Waals effects in molecule at metal-surface adsorption

The first step of our multi-step approach is the quantum mechanical study of single DNA bases on Au, with the particular goal of including vdW interactions in the calculations. With the aim of determining the functional that is best suitable for the description of DNA bases adsorbed on Au, we performed test calculations on different systems, evaluating the performances of the various functionals.

As already discussed (Section 2.2.2), recent theoretical developments allowed to incorporate vdW interactions in DFT calculations: new functionals based on theory or on experiments [48, 49, 53, 81, 82, 83] give more and more accurate results on some dispersion-interactions-dominated systems, e. g. aromatic molecules adsorbed on metals[84, 85].

Different functionals were found to give better results in the treatment of different systems, e.g. gas-phase molecules, metallic surfaces, graphene layers. Concerning the adsorption on surfaces, major efforts have been focused on the improvement of the description of physisorption on metals, as the vdW interaction is fundamental in such cases and aromatic molecules on metals are one of the reference system for evaluating the performances of these new functionals [55, 56, 57].

In particular, we were interested in comparing the performances of different functionals belonging to the vdW-DF family[48, 49, 50, 51, 52, 53] and to chose the most suitable for the study of DNA bases on Au. After the release of the first functional of the family, which separately treats

exchange and correlation contributions in a fully theoretical way [48], different modifications have been proposed to improve its performances on different kind of systems, mainly modifying the choice for the exchange functional and the parameters for the long range interactions. At present, four functionals have been developed and implemented in the quantum-espresso package [86], namely vdW-DF[48], vdW-DF2[52], C09-vdW-DF and C09-vdW-DF2[53]. In the first two functionals, exchange is calculated according to the revPBE [46] functional, while in the others the C09 exchange functional is used [53]. The difference between vdW-DF and vdW-DF2 models for correlation is just a matter of parametrization.

The first step of our study was to evaluate the description of the Au(111) surface, and the improvement relative to GGA treatments. The correct reproduction of the lattice parameter can be of extreme importance if the geometry of adsorbed molecules is influenced by the underlying geometry of the metal. If it is the case, a wrong lattice parameter would force to select between two inaccurate choices, either to force the experimental lattice parameter putting the system in artificial pressure conditions, or to adopt the DFT relaxed parameter, creating errors in the adsorption geometry. Thus we assessed the functionals on the evaluation of the equilibrium lattice parameter of fcc gold. The same calculation were done with LDA, PBE [45], and vdW-DF family functionals[48, 49, 50, 51, 52, 53].

As expected, an excellent result is obtained with LDA, while the PBE and vdW-DF(2) functionals overestimate the lattice parameter. On the contrary, the C09-vdW-DF(2) functionals are able to reproduce the correct lattice parameter: this result is not surprising, because these functionals are built to reduce short-range repulsion by construction [53].

We then evaluated the performances of vdW-DF functionals for the adsorption of small molecules on Au(111). Two extreme regimes are defined for the adsorption of molecules on surfaces: the case of weak overlap of electron orbitals between the adsorbate and the substrate is called physisorption; in this case long range interactions are often the only forces that bind the molecule to the surface. When covalent or ionic bonding between the molecule and the surface dominates the process of adsorption, the interaction is defined as a chemisorption, and the overall influence of vdW forces is assumed to be weak. Starting from these two extreme situations, different intermediate regimes can realize when both vdW interactions and short-range ones contribute significantly to the bonding.

The correct description of all the interactions involved in an adsorption process is practically

impossible to achieve and different functionals succeed in correctly describing different terms of the forces involved. For a reliable description of DNA bases on Au(111), we searched for a functional that encompasses the chemisorption and physisorption regimes. To this aim, test calculations on the adsorption of C_6H_6 (physisorption) and NH_3 (chemisorption) on Au(111) were performed.

Good results were obtained for C_6H_6 adsorbed on Au(111): in particular, the computed value of adsorption energy is within 1 kcal/mol from the experimental value for all the vdW-DF functionals, except vdW-DF-C09. While the revPBE exchange functional embedded in vdW-DF(2) gives better results on the energy, it is also true that the vdW-DF functional overestimates binding distances of molecules adsorbed on surfaces. This overestimation is improved by all the other functionals of the vdW-DF family.

The C09 exchange functional substantially overestimates the binding energy of NH_3 molecule, while the vdW-DF(2) functionals give good agreement with experiments. The explanation lies in the construction of the C09 exchange functional, that reduces short-range repulsion. The consequence is that for chemisorbed molecules the attractive interaction is too strong and the binding energy is overestimated.

Finally calculations for cytosine adsorbed on Au(111) were performed, both with vdW-DF and vdW-DF2-C09 functionals. These two functionals were chosen because the first one gives good results for the adsorption energy both in physisorption and chemisorption regimes and has a smaller lattice parameter than vdW-DF(2), while the second one gives the correct lattice parameter with a greater accuracy in the energy value than vdW-DF-C09.

The values of adsorption energy of cytosine molecule adsorbed on Au(111) computed with these two functionals differ by 6kcal/mol. Comparing this result with those obtained on C_6H_6 on Au(111) and NH_3 on Au(111), it is possible to state that vdW-DF2-C09 calculations probably overestimate the absolute value of the energy: plots of the density of states and molecular orbitals show an interaction between the base and the surface that is not only long-range. Results on NH_3 on Au(111) show that the short-range part of the interaction is badly evaluated with the vdW-DF(2)-C09 functional. Results obtained on these test systems motivated us to chose the vdW-DF functional as the most suitable one for the study of DNA bases adsorbed on Au(111).

The results of this work are presented in a manuscript in preparation, that is attached as an

integral part of this PhD dissertation.

Van der Waals effects in molecule/metal adsorption

Marta Rosa, Stefano Corni, and Rosa Di Felice*

Center S3, CNR Instituto of Nanoscience,

Via Campi 213/A, 41125 Modena, Italy

(Dated: November 6, 2013)

Abstract

We present the results of plane-wave pseudopotential periodic density functional theory calculations on the geometries, energetics and electronic structure of small molecules on Au(111). The chosen molecules – benzene, cytosine and ammonia – are representative of different adsorption regimes and interaction strengths. The chosen substrate is a prototype noble metal surface that is widely employed as a support for organic materials. We assess the relevance of van der Waals effects in the adsorption process and the accuracy of different density functionals that have been recently developed to embody such effects. We find that there is no unique functional that is optimal for any system. In fact the functionals that were found to describe more accurately gold geometry, fail in describing correctly chemisorption of small molecules on surfaces. In particular vdW-DF functional was found to describe satisfactorily both chemisorption and physisorption processes, while for a correct description of the surface C09 exchange functional gives better results.

Usage: Secondary publications and information retrieval purposes.

PACS numbers: May be entered using the `\pacs{#1}` command.

Structure: You may use the `description` environment to structure your abstract; use the optional argument of the `\item` command to give the category of each item.

I. INTRODUCTION

The adsorption of small molecules on metallic surfaces is gaining increasing interest in various fields of science and technology. On one hand, it is the configuration of relevance in molecular electronic devices. On the other hand, molecules must be adsorbed on flat (metal) substrates for the investigation of their intrinsic properties with imaging and spectroscopy tools.

For instance: When the individual electron orbitals of pentacene are imaged with a scanning tunneling microscope, the organic molecules are deposited on an inorganic substrate;[1] The adsorption of DNA molecules on surfaces is fundamental, *e.g.* for DNA sequencing.[2–4] In organic- and bio-molecular electronics the possible applications are virtually endless: molecules are used to transport current between electrodes,[5–9] as bonding between organic and inorganic materials (e.g., thiol or amino bridges[9–11]), to create self assembled monolayers on metals[12–15] and in other configurations. The realization and performance of such devices is crucially affected by the organic-inorganic interaction, which strongly motivates interest in this topic.

When a molecule approaches a surface the adsortpion mechanism is described by two extreme regimes and the range of intermediate situations. The case of weak adsorbate-substrate coupling with no overlap of electron orbitals is called physisorption; in this case, which is typical of aromatic molecules, long-range dispersion interactions are practically the only terms that "bind" the molecules to the surface. When covalent or ionic bonding between the molecule and the surface dominates the process of adsorption, the interaction is defined as chemisorption; in this case, which is typical of molecules with a reactive functional group (e.g., thiols), van der Waals (vdW) interactions terms are not dominant and can be neglected to a first approximation. Between these two extreme conditions, there exist intermediate regimes in which both long-range and short-range interactions contribute significantly to the bonding.

Small molecules chemisorbed on metallic surfaces, where long-range interactions are usually negligible, can be satisfactorily described by density functional theory (DFT) calculations and have been the subject of a variety of studies in the past years.[16–18] However, the lack of vdW interactions in traditional DFT functionals prevented the computational investigation of physisorbed systems.[19] Recent theoretical developements enabled the incorporation of vdW interactions in DFT calculations, through the formulation of improved exchange-

correlation functionals. Successive generations of new functionals (all developed in the past 5 years) based on theory or on experiments[20–25] yield increasingly accurate reasults on some critical systems dominated by dispersion interactions, e. g. aromatic molecules adsorbed on metals.[26,27] Unfortunately, despite the significant progress in the description of such critical systems, there is still confusion about which is the best functional to compute the structural and electronic properties of a certain molecule/inorganic-surface interface. In fact, different functionals seem more suitable to describe different types of systems, such as gas-phase molecules, metallic surfaces, graphene layers. Major efforts have been devoted to improve the description of physisorption of organic molecules on metals, as the vdW interaction is fundamental in such cases; aromatic molecules on metals thus constitute benchmark systems for evaluating the preformances of these new functionals.[28–30] DFT calculations with vdW functionals substantially improve the results on the adsorption energy of C_6H_6 on Au relative to GGA functionals, giving values that are within 0.5 eV from experimental data.

To shed light on the variety of available vdW functionals and their different performance on different systems, in this work we present a systematic evaluation of such functionals[20–22,31–33] on different adsorption systems that represent the physisorption regime (C_6H_6/Au(111)), the chemisorption regime (NH_3/Au(111)) and an intermediate regime (cytosine/Au(111)). After the first functional of the family (vdW-DF),[20] different modifications have been done to it to improve the DFT performance on different kinds of systems. For instance, some of them are able to reproduce experimental lattice parameters for different metals, while usually GGA functionals strongly overestimate them.[34,35] These functionals may be best suitable for the adsorption of molecules on surfaces, thanks to the correct description of the host substrate. However, the results obtained so far indicate that the DFT performance with vdW functionals varies substantially depending on the adsorption mechanism. More specifically, problems arise when the adsorption mechanism is in the regime between chemisorption and physisorption, or when it is not known. In these cases it is not yet clear wich one of the possible functionals gives the better accuracy. With the aim of selecting the best functional for the study of DNA bases on Au(111), we performed calculations on different systems, comparing our resuls with experimental and theoretical data. DNA bases are known to interact with gold mainly through long range interactions,[19,36] but the geometry of the bases allows also direct binding of some reactive atoms, e.g. oxygens and non-saturated hydrogens. For this reason, the importance of short-range interactions should not be excluded *a priori*.

Indeed, using the vdW-DF functional,[20] we found that both long-range and short-range interactions play an important role in the adsorption of cytosine, guanine, thymine and adenine on Au(111).[37,38] In the following, we present the results of DFT calculations of benzene and ammonia on Au(111) with the different available functionals of the vdW-family and formulate our rationale for the choice of vdW-DF to handle DNA on gold.

METHOD

Local-density and gradient-corrected DFT calculations of bulk Au, the Au(111) clean surface, gas-phase molecules and molecule@Au(111) interfaces are performed with the quantum-espresso package,[39] using the LDA[40] and PBE[41] exchange-correlation functionals, plus four different functionals of the vdW-DF family.[20,22,33] A plane wave basis set with a cutoff of 25 Ry on the wave functions is chosen and the electron-ion interaction is described with ultra-soft pseudopotentials.[42] The Brillouin zone (BZ) sums are calculated including 2 Monkhorst-Pack special k points in the irreducible wedge. All the atomic coordinates are relaxed until the forces vanish within 0.05 eV/Å. After atomic optimization, a series of single-point self-consistent calculations at frozen internal coordinates of the molecule and surface and varying molecule-surface distance is performed, to avoid that the system is trapped in local minima.

We use periodically repeated supercells with large orthorhombic unit cells that ensure negligible interactions between neighboring replicas, as inferred by tests calculations with larger unit cells. Specifically, the size of the unit cells for C_6H_6@Au(111) and NH_3@Au(111) interfaces fit Au(111) slab with $4 \times 2\sqrt{3}$ periodicity and four layers, with 16 Au atoms in each layer. Due to the larger size of cytosine, cytosine@Au(111) interfaces are modeled with $6 \times 3\sqrt{3}$ periodic unit cells and four layers, with 36 Au atoms per layer. The vacuum thickness is in all cases at least 11 Å. The in-plane distance between periodic molecules is at least 7 Å in the case of C_6H_6 and NH_3 adsorbates, at least 10 Å in the case of cytosine adsorbates. For a given molecular species, the exact size of the unit cell is different for any of the functionals, fixed according to the computed equilibrium lattice constant of *fcc* bulk gold with that same functional.

Four different functionals of the vdW-DF family have been so far implemented in the quantum-espresso package, namely vdW-DF,[20] vdW-DF2[33], C09-vdW-DF[22] and C09-vdW-

DF2.[34] In the functionals vdW-Df and vdW-DF2 the exchange part is revPBE,[43] while in the functionals C09-vdW-DF and c)9-vdW-DF2 the exchange part is C09.[22] The difference between the vdW-DF and vdW-DF2 functionals (as well as between C09-vdW-DF and c09-vdW-DF2) is in the correlation part.[33]

A crucial choice for doing calculations with functionals of the vdW-DF family concerns the numerical pesudopotential. A numerical pseudopotential is specified not only by the analytical form (norm-conserving, non-norm-conserving, Kleinman-Bylander factorization, etc.) but also by the functional that has been used to calculate the values on the grid.[44,45] Rigorously, a pseudopotential that is used in a vdW-DF (vdW-DF2, C09-vdW-DF, vdW-C09-DF2) calculation of a target interface system should be generated with the same functional in an atomic DFT calculation. Pseudopotentials generated with any of the functionals of the vdW-DF family are not available in the most used databases. Database pseudopotentials (e.g., `http://www.quantum-espresso.org`) that were widely tested for use in a variety of materials and systems were generated with the PBE functional. In principle, one could generate pseudopotentials consistent with the functionals of the vdW-DF family (revPBE or C09 exchange): we have done so and performed test calculations to benchmark PBE psudopotentials. The generation and test of a pseudopotential is a delicate matter, and the risk of using a not well tested functional would be justified only by a sensibly negative such benchmark. Our tests on Au bulk, C_6H_6@Au(111) and cytosine@Au(111) indicate a good performance of the psudopotentials generated in atomic calculations with the PBE functionals and then used in materials calculations employing the vdW-DF functional, against pseudopotentials generated in atomic calculations with the revPBE functional.[37] Similar results were obtained by other groups on graphite and C_6H_6 adsorbed on a graphite (0001) surface.[46] Thus, when it is not differently stated, calculations are performed with PBE pseudopotentials.

The formation energy is calculated as the difference between the total energy of the interface ($E_{tot}^{interface}$) and the sum of total energies of the clean surface ($E_{tot}^{substrate}$) and gas-phase molecule ($E_{tot}^{adsorbate}$): $E_{form} = E_{tot}^{interface} - \left[E_{tot}^{substrate} + E_{tot}^{adsorbate}\right]$. The calculations of the isolated sub-systems are done using the same periodic unit cell as in the interface calculation. A negative value of the formation energy stands for exothermic adsorption. The adsorption energy is the opposite of the formation energy and is positive for exothermic reactions.

RESULTS AND DISCUSSION

A DFT calculation of a surface or interface should be done using a two-dimensional unit cell that is a multiple of the equilibrium bulk lattice constant computed consistently with the methodology of choice, so that the system does not experience a fake pressure in the simulation. Therefore, it is important to check if the functionals of the vdW-DF family give correct results for the equilibrium lattice parameter of bulk gold. If this is not the case, since asorption geometries are influenced by the structure of the substrate, one would obtain unreliable results for interfaces by using a unit cell for the interface calculation that maps the size of the theoretical equilibrium lattice constant. Alternatively, one could choose the experimental lattice constant, but in this case the interface system would be subjected to an artificial condition of pressure.

Our results for the equilibrium lattice constant of bulk Au, computed with the functionals of the vdW-DF family, the LDA and PBE functionals, are collected in Table I. The values reported in Table I represent the minimum of the energy curve constructed, for each selected functional, by doing several single-point self-consistent electronic structure calculations of fcc bulk gold at different values of the unit cell around the experimental lattice constant (4.08 Å [47]).

As expected on the basis of previous work,[48,49] an excellent result is obtained with LDA. PBE, vdW-DF and vdW-DF2 calculations yield results that overestimate the experimental value by 2÷8%, while PBE and vdW-DF(2) functionals overestimate the lattice parameter; the pseudopotential used in vdW-DF calculations has a negligible effect on this overestimation. The C09-vdW-DF and C09-vdW-DF2 functionals, with the use of an optimized exchange part that by construction reduces short-range repulsion,[22] are able to correct for this overestimation and give an excellent agreement relative to the experimental value. The main driving force for implementing modifications to the first functional of the vdW-DF family, the vdW-DF functional, was to improve defective results for the description of surfaces, as well as for molecule-molecule and molecule-surface distances. While the different vdW-DF2 parametrization of correlation should be effective in tuning the adsorption distance of molecules on surfaces, the C09 exchange functional succeds in calculating correct lattice parameters of bulk solids. For this reason, it is important to determine its accuracy for simulations of small molecules on metal surfaces, at interfaces characterized by differ-

Method	pseudopotential	a_{eq} (Å)
exp	N/A	4.08
DFT-LDA	LDA	4.06
DFT-PBE	PBE	4.17
DFT-vdW-DF	PBE	4.24
DFT-vdW-DF	revPBE	4.26
DFT-vdW-DF2	PBE	4.40
DFT-C09-vdW-DF	PBE	4.08
DFT-C09-vdW-DF2	PBE	4.07

Table I. Equilibrium lattice parameter for gold bulk in Å (a_{eq}) calculated with different functionals. The property reported in the second column is the functional adopted in the atomic calculation of the pseudopotentials that is now used for the materials calculation, according to the first row. The pseudopotentials are taken from the quantum-espresso database and the atomic calculations for pseudopotential generation are not done by us.

ent kinds of interactions, dominated by dispersion or chemical bonding. We thus chose the C_6H_6/Au(111) as a prototype example of the adsorption of an aromatic molecule, with the resulting interface intuitively dominated by dispersion interactions. The NH_3/Au(111) system was chosen as example of the adsorption of a molecule with a reactive center that can in principle give origin to chemical bonding. The cytosine/Au(111) systems represent an intermediate regime. It is important to establish the performance of the chosen functionals in such diverse interaction regimes.

Figure I reports the formation energy of the C_6H_6 interface as a function of the molecule-surface distance obtained from single-point DFT-vdW-DF calculations at fixed internal coordinates of the substrate and adsorbate. The left inset shows a top view of the optimized geometry. The plot in Figure I illustrates that the use of PBE or revPBE pseudopotentials is not effective in such calculations, thus reinforcing our choice of using PBE pseudopotentials for all the other systems and functionals.

The adsorption energy calculated with functionals of the vdW-DF family for the different target interfaces is reported in Table II. The adsorption energy is obtained in all cases with the same procedure as in Figure I, namely full geometry optimization followed by single-point

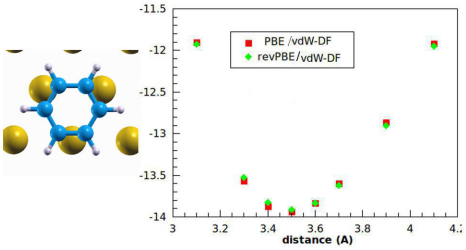


Figure 1. Formation energy as a function of molecule-surface distance at fixed coordinates of the molecule and of the surface, obtained from single-point self-consistent DFT-vdW-DF calculations with different pseudopotentials, as indicated in the legend. PBE/vdW-DF (revPBE/vdW-DF) refers to DFT-vdW-DF calculations of the interface done with pseudopotentials generated from atomic calculations with the PBE (revPBE) exchange-correlation functional. The inset illustrates a top view of the optimized adsorption configuration for benzene on Au(111), in which the molecule lays with its aromatic plane parallel to the substrate plane (horizontal adsorption). This relaxed configuration was obtained by DFT-vdW-DF calculations with PBE-generated pseudopotentials.

self-consistent DFT calculations at frozen internal coordinates and variable molecule-surface distance. Two different adsorption sites were tested with the vdW-DF functional for both C_6H_6@Au(111) and NH_3@Au(111). C_6H_6 finds equilibrium conditions when the center of the aromatic cycle is above a *fcc* site or a *top* site, with adsorption energy of 13.7 kcal/mol and 12.9 kcal/mol, respectively. NH_3 finds equilibrium conditions with the N atom above the same locations of the Au(111) triangular lattice, but it is more stable at the *top* site (10.5 kcal/mol) that at the *fcc* site (5.7 kcal/mol). In the most favorable configuration of the C_6H_6@Au(111) interface the molecule lies horizontal on the surface with its center above a *fcc* site. In the most favorable configuration of the NH_3@Au(111) interface the molecule points the N atom to the surface above a *top* site and the H atoms opposite to the surface. These results are in agreement with published theoretical[50,51] and experimental data.[52] For the cytosine@Au(111) interface we performed calculations with the O atom above *fcc*, *top* and *bridge* sites: we find both horizontal and vertical configurations, with the O atom adsorbed on top of a substrate gold atom (Figure 2).[37,38] The results presented in Table II pertain to the most favorable conditions for each interface.

The data in Table II reveals that the different functionals of the vdW-DF family yield different accuracies for interfaces with different chemical nature, as compared to available

	C_6H_6/Au(111)		NH_3/Au(111)		cytosine/Au(111)	
	E_{ads}	d_{eq}^{int}	E_{ads}	d_{eq}^{int}	E_{ads}	d_{eq}^{int}
experiment	13.8÷14.6[30,52]		7.7÷10.0[51]			
vdW-DF (PBE)	13.7	3.4	10.5	2.6	19.7^h,19.2^v	2.7^h,2.6^v
vdW-DF (revPBE)	13.8	3.4				
vdW-DF2 (PBE)	13.0	3.2	11.7	2.6		
C09-vdW-DF (PBE)	23.1	3.0	18.4	2.4		
C09-vdW-DF2 (PBE)	15.6	3.1	15.0	2.4	25.0^h,25.3^v	2.2^h,2.2^v

Table II. Adsorption energy and molecule-surface distance from DFT calculations with different functionals, for C_6H_6/Au(111), NH_3/Au(111) and cytosine/Au(111). Experimental values and related references are also reported. E_{ads} is the eadsorption energy in kcal/mol. d_{eq}^{int} is the molecule-surface equilibrium distance calculated between the closest atoms in the substrate and in the adsorbate, in Å. In the first column, we indicate in brackets the functional used in the atomic DFT calculation for the generation of the pseudopotential adopted by us for the interface calculation. The superscripts ^{h} and ^{v} for cytosine indicate horizontal and vertical adsorption geometries, respectively (see Figure 2).

experimental values for benzene and ammonia. The experimental values are desorption energies from temperature desorption (TPD) experiments.[52,53]. Before presenting our analysis, we wish to comment on the limits of comparing computed adsorption energies at a precise coverage to measured desorption energies in experiments where the coverage is variable.
TPD estimates of the desorption energy are derived from the measured desorption rate $N(t)$, according to Redhead formula:[54]

$$N(t) = \nu N_S \sigma^n exp(-H_d/k_B T_d), \tag{1}$$

where ν, σ, H_d, k_B and T_d are the pre-exponential factor, the surface coverage, the desorption enthalpy, the Boltzmann constant and the desorption temperature, respectively. The evaluation of the desorption enthalpy from the Redhead formula requires approximations that limit its accuracy. It is assumed that the desorption enthalpy is independent of surface coverage and that desorption follows a first-order kinetics ($n = 1$ in equation 1),[55]. The value of the pre-exponential factor is normally assigned as 10^{-13} s^{-1}, irrespectively of adsorbate and substrate species and coverage. The values for teh desorption enthalpy obtained

in this way are commonly used as the experimental reference in benchmark of computational approaches.[56–58] A more detailed evaluation has been done for some systems[57], with the conclusion that the Redhead formula underestimates the strength of the interaction. A correction has been recently proposed for the C_6H_6@Au(111) system.[55] Nevertheless, a general agreement on the pre-exponential factor has not been reached[57,58] and the refined models have had limited application so far, only to the adsorption of straight-chain alkanes to surfaces different from Au. Because C_6H_6 is "topologically" different from alkanes (a cycle rather than a linear chain), probably a different refinement of the Redhead formula should be found and this was not done so far. Thus, we compare our adsorption energies to the only available sets of data. Despite the energy value extrapolated by TPD experiments probably underestimates the real one, we think that this difference is not large enough to our qualitative conclusion on the performance of the functionals.

For the system C_6H_6/Au(111) the adsorption energy calculated with all functionals is in good agreement with the experimental range, within 1 kcal/mol, except for the value obtained with the C09-vdW-DF functional, which is severely overestimated by 8.4 kcal/mol. The performance is sensibly weaker for NH_3, for which all results are above the experimental range; vdW-DF and vdW-DF2 overestimate the highest boundary of the experimental range by 0.5 and 1.7 kcal/mol, respectively, while C09-vdW-DF and C09-vdW-DF2 are appreciably off, by 8.4 and 5 kcal/mol, respectively. We can conclude that overall the functionals of the vdW-DF family that include the C09 exchange part perform worse than those with the revPBE exchange part for the computation of adsorption energies, much worse for our prototype bonding interface – NH_3@Au(111) – than for our prototype dispersive interface – C_6H_6@Au(111). For cytosine on Au(111) experimental data on single adsorbed molecules are not available; yet, considering the larger adsorption energies obtained from C09-vdW-DF calculations, we can reasonably say that functionals with the C09 exchange part systematically overestimate binding energies for any range of interface interactions. This is quite reasonable, because the C09 exchange functional has been devised to reduce the short-range repulsion.[22] This evidence was already found for the adsorption of graphene on Au(111)[34] and our findings now prove that it is quite more general and applies to a variety of molecular adsorbates with diverse chemical nature and consequent diverse coupling to the host substrate. DNA bases on Au(111) were extensively investigated by us with the vdW-DF functional.[37,38] Although experimental data on single adsorbed cytosine on Au(111) is not

available for direct comparison in Table II, our vdW-DF results for the adsorption energies are in good agreement with experimental data on monolayer formation.[59]

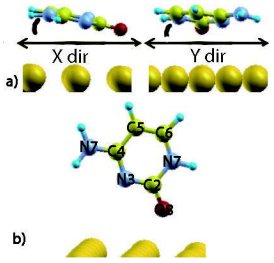


Figure 2. Optimized adsorption configurations for cytosine/Au(111)

The adsorption strength and mechanism can be deeper investigated by analyzing the electronic structure obtained by DFT at the equilibrium configuration. It is interesting to find out whether different functionals that give different adsorption energies for a given system also alter the electronic structure. We tackle this issue focusing on cytosine@Au(111), for which we recently published a detailed work performed with the vdW-DF functional.[37,38] For this system we identified an adsorption mechanism intermediate between chemisorption and physisorption, with the formation of bonding orbitals between the O8 and N3 atoms of cytosine and Au atoms of the substrate.[37] The same result is obtained also with C09-vdW-DF2. The insets in Figure I show for the two functionals the contribution to the charge density for approximatively the same energies (-2.33 eV vs -2.37 eV) plotted with the same isosurface value. In the case of C09-vdW-DF2 there is the formation of a bonding orbital between both O8 and N3 atoms and the surface, while in the case of vdW-DF the bonding orbital is only between N3 atom and gold. Plots obtained with different isosurface values or for different energies show bonding orbitals between O8 and gold also in the case of vdW-DF functional.[38] Consequently the difference between the two functionals is only quantitative and it is probably due to the closer distance of O8 atom to the surface in the case of C09-vdW-DF2 geometry, which is a consequence of a stronger interaction. Comparing the DOS plot it can be seen that the difference between the gas-phase and the adsorbed DOS plots is stronger for C09-vdW-DF2 functional, particularly in the range between 0 and -3 eV. The gas-phase peaks are more spread in the adsorbed molecule DOS and small shift happens for both the functionals.

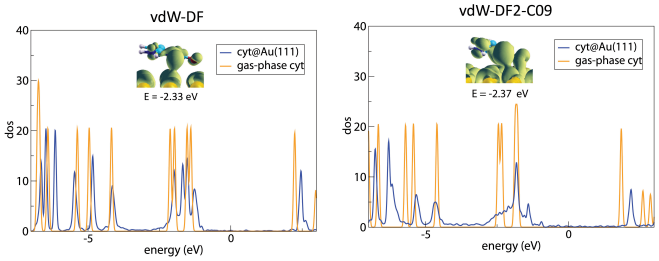


Figure 3. Density of states of cytosine adsorbed horizontally on the Au(111) surface, computed with the vdW-DF functional. The Fermi level of the system is set at the origin of the energy scale and the deepest energy level, which is associated to the same orbital in the gas-phase molecule and in the adsorbed molecule, is used for alignment of the various curves. The orange (blue) line is the DOS of the gas-phase molecule (total DOS of the interface projected onto the molecule). In the inset, a bonding orbital is shown

Comparing this results with the ones obtained on C6H6 and NH3 adsorption, it is possible to say that vdW-DF2-C09 calculation probably overestimates the absolute value of the adsorption energy and the strength on the interaction in general. This hypothesis is sustained first of all by the value of the distance between the surface and the molecule: 2 Åseems to be a too small distance for such a weak interaction. Moreover experiments on the adsorption of monolayers on Au(111)[60] and molecular dynamics calculations[59] show how horizontal adsorption configuration is preferred respect to vertical one, while C09-vdW-DF2 inverts this result. Finally calculations performed with vdW-DF functional on DNA bases proved a fair agrement with experiments, demonstrating that vdW-DF gives a good description of the system.[59] This result suggests a bad description of the short range interaction in the system by vdW-DF2-C09 functional.

On the other hand we must not forget that this good result on the adsorption of DNA bases depends strongly from the fact that DNA bases adsorption does not depends from the geometry of the underlying lattice and from the fact that in this comparison there was no interest in reproducing the correct adsorption distance. On the contrary vdW-DF2-C09 would be probably the better choice for describing the adsorption of a fully physisorbed molecule, even more if the adsorption geometry depends from the lattice geometry.

CONCLUSION

We performed calculations on molecules adsorbed on Au(111) surface through different adsorption mechanism, verifying the performances of different functionals able to take into account vdW interaction. Different functionals show different pros and cons and the final choice depends on the system under study. We demonstrated that in the case of a system where both short and long range interaction are, or are supposed to be, involved, functionals that were thought to be optimized for the adsorption of molecules on surfaces (i. e. C09-vdW-DF(2)) have some difficulties to provide the correct energy and distance, as they are not able to deal properly with the short range part of the interaction.
On the contrary, vdW-DF functional seems to be a good candidate to correctly describe such kind of system due to the good results obtained on DNA bases on Au(111), but only if lattice geometry does not influence the adsorption geometry.

---

* rosa.difelice@unimore.it; Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA

[1] J. Repp, G. Meyer, S. M. Stojković, A. Gourdon, and C. Joachim, Physical Review Letters (PRL) **94**, 026803 (2005).

[2] T. M. Herne and M. J. Tarlov, Journal of the American Chemical Society **119**, 8916 (1997).

[3] S. M. Schreiner, A. L. Hatch, D. F. Shudy, D. R. Howard, C. Howell, J. Zhao, P. Koelsch, M. Zharnikov, D. Y. Petrovykh, and A. Opdahl, Analytical chemistry **83**, 4288 (2011).

[4] M. Zwolak and M. Di Ventra, Reviews of Modern Physics **80**, 141 (2008).

[5] S. Aradhya, M. Frei, M. Hybertsen, and L. Venkataraman, Nature Mater. **11** (2012).

[6] L. Venkataraman, J. E. Klare, C. Nuckolls, M. S. Hybertsen, and M. L. Steigerwald, Nature **442** (2006).

[7] A. Nitzan and M. A. Ratner, Science **300** (2003).

[8] C. Z. M. A. Reed, C. J. Muller, T. P. Burgin, and J. M. Tour, Science **287** (1997).

[9] H. Cohen, C. Nogues, R. Naaman, and D. Porath, Proc. Natl. Acad. Sci. USA **102** (2005).

[10] D. Porath, A. Bezryadin, S. de Vries, and C. Dekker, Nature **403** (2000).

[11] S. Y. Quek, L. Venkataraman, H. J. Choi, S. G. Louie, M. S. Hybertsen, and J. Neaton, Nano letters **7**, 3477 (2007).

[12] A. Ulman, Chem. Rev. **96**, 1533 (1996).

[13] C. Joachin, J. K. Gimzewski, and A. Aviram, Nature **408** (2000).

[14] A. Kühnle, T. R. Linderoth, B. Hammer, and F. Besenbacher, Nature **415** (2002).

[15] R. Otero, M. Schöck, L. M. Molina, E. Lægsgaard, I. Stensgaard, B. Hammer, and F. Besenbacher, Angewandte Chemie International Edition **44**, 2270 (2005).

[16] R. D. Felice, A. Selloni, and E. Molinari, J. Phys. Chem. B **111**, 1151 (2003).

[17] C. Jing and Y. Fang, Chemical physics **332**, 27 (2007).

[18] H. Grönbeck, A. Curioni, and W. Andreoni, Journal of the American Chemical Society **122**, 3839 (2000).

[19] S. Piana and A. Billic, J. Phys. Chem. B **110**, 23467 (2006).

[20] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Physical review letters **92**, 246401 (2004).

[21] T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard, and D. C. Langreth, Physical Review B **76**, 125112 (2007).

[22] V. R. Cooper, Physical Review B **81**, 161104 (2010).

[23] J. Antony and S. Grimme, Physical Chemistry Chemical Physics **8**, 5287 (2006).

[24] A. D. Becke and E. R. Johnson, The Journal of chemical physics **127**, 154108 (2007).

[25] A. Tkatchenko and M. Scheffler, Physical review letters **102**, 073005 (2009).

[26] S. D. Chakarova-Käck, E. Schröder, B. I. Lundqvist, and D. C. Langreth, Physical review letters **96**, 146107 (2006).

[27] T. S. Chwee and M. B. Sullivan, J. Chem. Phys. **137** (2012).

[28] J. Klimes, D. R. Bowder, and A. Michaelides, J. Phys.:Condens. Matter **22** (2010).

[29] W. Liu, J. Carrasco, B. Santra, A. Michaelides, M. Scheffler, and A. Tkatchenko, Phys. rev. B **86** (2012).

[30] J. Wellendorff, A. Kelkkanen, J. J. Mortensen, B. I. Lundqvist, and T. Bligaard, Top. Catal. **53** (2010).

[31] D. C. Langreth, M. Dion, H. Rydberg, E. Schröder, P. Hyldgaard, and B. I. Lundqvist, International journal of quantum chemistry **101**, 599 (2005).

[32] A. Puzder, M. Dion, and D. C. Langreth, The Journal of chemical physics **124**, 164105 (2006).

[33] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Physical Review B **82**, 081101 (2010).

[34] I. Hamada and M. Otani, Physical Review B **82**, 153412 (2010).

[35] G. Li, I. Tamblyn, V. R. Cooper, H.-J. Gao, and J. B. Neaton, Physical Review B **85**, 121409 (2012).

[36] N. Sandig and F. Zerbetto, Chem. Commun. **46**, 667 (2010).

[37] M. Rosa, S. Corni, and R. Di Felice, The Journal of Physical Chemistry C **116**, 21366 (2012).

[38] M. Rosa, S. Corni, and R. Di Felice, Journal of Chemical Theory and Computation **9**, 4552 (2013).

[39] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, *et al.*, Journal of Physics: Condensed Matter **21**, 395502 (2009).

[40] W. Kohn and L. J. Sham, Physical Review **140**, A1133 (1965).

[41] J. P. Perdew, K. Burke, and M. Ernzerhof, Physical review letters **77**, 3865 (1996).

[42] D. Vanderbilt, Physical Review B **41**, 7892 (1990).

[43] Y. Zhang and W. Yang, The Journal of chemical physics **109**, 2604 (1998).

[44] G. Ortiz and P. Ballone, Physical Review B **43**, 6376 (1991).

[45] E. L. Shirley, R. M. Martin, G. B. Bachelet, and D. M. Ceperley, Physical Review B **42**, 5057 (1990).

[46] I. Hamada and S. Yanagisawa, Physical Review B **84**, 153104 (2011).

[47] R. W. G. Wyckoff, *Crystal Structures*, 2nd ed., Vol. 1 (IntersciencePublishers, New York, 1963).

[48] K. Terakura, T. Oguchi, T. Mohri, and K. Watanabe, Phys. Rev. B **51** (1995).

[49] A. Khein, D. J. Singh, and C. J. Umrigar, Physical Review B **51**, 4105 (1995).

[50] K. Tonigold and A. Groß, The Journal of chemical physics **132**, 224701 (2010).

[51] B. D. Kay, K. R. Lykke, J. R. Creighton, and S. J. Ward, The Journal of chemical physics **91**, 5120 (1989).

[52] D. Syomin, J. Kim, B. E. Koel, and G. B. Ellison, The Journal of Physical Chemistry B **105**, 8387 (2001).

[53] M. E. Coltrin and B. D. Kay, J. Chem. Phys. **89** (2988).

[54] P. Redhead, Vacuum **12**, 203 (1962).

[55] W. Liu, V. G. Ruiz, G.-X. Zhang, B. Santra, X. Ren, M. Scheffler, and A. Tkatchenko, New Journal of Physics **15**, 053046 (2013).

[56] R. Z. Lei, A. J. Gellman, and B. Koel, Surface science **554**, 125 (2004).

[57] C. T. Campbell and J. R. Sellers., J. Am. Chem. Soc. **134** (2012).

[58] K. A. Fichthorn and R. A. Miron, Physical review letters **89**, 196103 (2002).

[59] M. Rosa, S. Corni, and R. Di Felice, in preparation.

[60] R. E. A. Kelly, M. Lukas, L. N. Kantorovich, R. Otero, W. Xu, M. Mura, E. Laegsgaard, I. Stensgaard, and F. Besenbacher, J. Chem. Phys **193** (2008).

Chapter 4

A Density Functional Theory Study of Cytosine on Au (111)

With the aim of parameterizing a FF for molecular dynamics simulations of DNA on surfaces, the first thing to do is to study how single bases adsorb on hard substrates with ab-initio calculations. The interaction between nucleobases and the Au(111) surface is expected to be dominated by dispersion interactions. Yet, studies have been performed so far neglecting this contribution to energy, both claiming it to be negligible or because it was not practically possible to include it into DFT calculations[47]. The performances of the vdW-DF functional [48] in describing the adsorption of small molecules on surfaces has already been successfully tested. The first step of our study on DNA adsorption on Au(111) was to perform calculations with and without the inclusion of long range interactions (namely with the vdW-DF functional and the PBE functional) for one of the DNA bases, comparing the results obtained also with experiments, to point out the main differences between the performances of the two functionals, the agreement with experiments and the most suitable procedure for extending the same study to the other three DNA bases.

Cytosine was chosen between the four bases as it is the one that better represents the particular features of DNA bases: its structure includes one O atom and one not saturated N atom, the ones that are supposed to interact with the surface. Adenine has no O atoms, while thymine has two of them, in addition to the CH_3 group that induces steric hindrance. Guanine could

have been a good choice as well, but we decided to start with cytosine as it is smaller and has a smaller phase space on Au(111).

Comparison between PBE and vdW-DF computational results reveals differences both in adsorption energy and geometry. The starting configuration in the calculations performed with both the functionals is with the molecule parallel to the surface, at a distance of 3.4 Å from gold atoms. On the Au(111) surface three symmetry sites are present and are interesting for the adsorption: top, that is positioned above a gold atom; bridge that is in the middle of the line connecting two neighboring gold atoms; center/fcc, that is in the centre of the triangle made by three neighboring gold atoms. To evaluate the different adsorption affinities and the mobility of the base on the surface we performed calculations for the base adsorbed on the three sites, using the O atom of cytosine as reference for the position.

Comparing PBE and vdW-DF calculations the most striking difference is the value of the adsorption energy: including vdW interaction it is approximatively five times larger than without vdW. This result attests that long range interactions are not negligible and are fundamental for a correct description of the system. Moreover, looking at the geometry, it is possible to see that not only a quantitative, but also a qualitative difference arises: namely, there is a major shift in the adsorption process. In fact using the PBE functional cytosine is adsorbed parallel with respect to the surface, at a distance of 3.5 Å. On the contrary, vdW-DF cytosine is adsorbed tilted with respect to the surface, with the centre of mass at a distance of 3.3 Å and with the O atom at a distance of 2.8 Å from the top gold plane, possibly short-range interactions with gold. test calculations proved that this difference is not an artifact of the starting condition, but a genuine result. Comparing vdW-DF and PBE energies for the vertical adsorption geometries, we found another qualitative difference that must be pointed out: the vertical configuration is preferred in calculations performed without dispersion interactions, while including long range forces the horizontal configuration is favored, as C atoms belonging to the aromatic ring have an important role in the adsorption. This result is in agreement with AFM experiment on monolayers, that show nucleobases adsorbed horizontally with respect to the surface[87, 88, 89, 90, 91, 92]. A deeper analysis of this evidence will be performed in the following chapter for the four bases. The electronic density of states reveals minor perturbations of the electrostatic structure upon adsorption. This fact would suggest a weak interaction between the molecule and the surface. In fact, in the case of chemisorption the Newns-Anderson model[93] predicts that the interaction

between the localized molecular HOMO (Higher Occupied Molecular Orbital) and the narrow d band of the metal produces hybrid orbitals of both bonding and antibonding type, while in our case HOMO and HOMO-1 peaks are superposed to the d bands of gold, with siglht changes relative to the gas-phase molecule. Surprisingly, not in agreement with this picture of negligible coupling, plotting the charge density distributions we discover that MOs (Molecular Orbitals) in the energy range of gold d bands are able to create bonding orbitals with the metal. Bonding orbitals are formed by various cytosine molecular orbitals with the Au d bands, but the energy level of a hybrid orbital undergos only a slight shift relative to the gas phase. Our findings point out a kind of interaction stronger than pure physisorption, in particular between O (and not saturated N) and gold, but comparable to effective chemisorption.

By comparing charge density results between PBE and vdW-DF calculations, we notice that PBE calculations are not able to point out this effect. This means that the vdW interaction allows the molecule to go close to the surface as a consequence of the long range interaction of the heterocycle with Au, so that it can eventually exploit O-Au and N-Au short-range interactions.

The results obtained on cytosine confirm the importance of vdW effects in dealing with the adsorption of nucleobases on Au(111), both quantitatively and qualitatively, while the accordance with experiment confirm the accuracy of the vdw-DF functional for the description of the cytosine/Au(111) system.

The results of this work have been published in the Journal of Physical Chemestry C and the reprint is attached as an integral part of this PhD dissertation.

THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

pubs.acs.org/JPCC

A Density Functional Theory Study of Cytosine on Au(111)

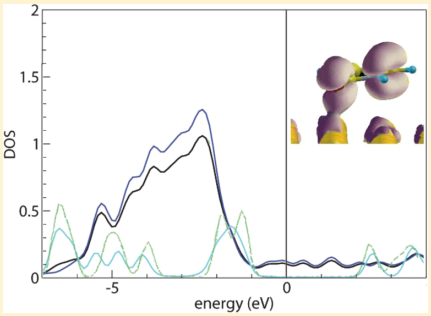
Marta Rosa,[†,‡] Stefano Corni,[†] and Rosa Di Felice*[,†]

[†]Center S3, CNR Institute of Nanoscience, Via Campi 213/A, 41125 Modena, Italy

[‡]Department of Physics, University of Modena and Reggio Emilia, Modena, Italy

Supporting Information

ABSTRACT: The adsorption of cytosine on Au(111) is investigated using density functional theory with the nonlocal van der Waals density functional. Test calculations performed on the benzene stacked dimer and on a benzene molecule adsorbed on Au(111) allow us to assess the methodology and reveal the accuracy and predictivity of the van der Waals density funcional relative to experimental outcome. Our results for cytosine on Au(111) indicate that the inclusion of dispersion interactions is crucial for the treatment of this system. In fact, such terms enhance the value of the adsorption energy and also affect the cytosine bonding geometry: in particular, we find that a tilted geometry is always favorable relative to a parallel geometry, which was not found in standard density functional theory investigations. The combined new data for energetics and geometry lead to conclusions that contrast the common opinion that the surface−molecule interaction is negligible in the process of monolayer formation.



INTRODUCTION

The investigation of small molecules on surfaces is a viable way to gain insights into the general principles of self-assembly. In particular, the self-assembly of DNA bases on template surfaces is a hot topic in relation to the emergence of life under prebiotic conditions.[1] For its connection to this extremely fascinating topic, the adsorption of DNA bases on inorganic surfaces has been recently investigated with modern experimental and theoretical techniques. In particular, atomistic computational methods have been proven to be an effective tool to unravel crucial features linked to the geometry and to the electronic structure.[2−6] Among the DNA bases, cytosine deserves special attention because it is not able to create a self-assembled monolayer on, e.g., Au(111): in fact, at low coverage, cytosine prefers filament structures with a high mobility.[7,8]

Understanding the adsorption of DNA bases on inorganic surfaces is also a crucial step to grasp the mechanisms of interaction of entire DNA molecules with hard substrates, which is a key issue in different contexts, such as the development of DNA microarray techniques,[9] portable cost-effective DNA sequencing,[10] targeted drug delivery,[11] formation of self-assembled superstructures,[3,8] and others. Although huge research efforts have been invested in recent years in this field, the awareness on the interaction mechanisms is still elusive and deserves further attention. To gain fundamental insights into the problem of DNA/surface coupling, one may again resort to computational approaches. Computational studies at an ab initio level are unfeasible because of the huge size of the target systems. In order to parametrize simplified affordable effective models, e.g., for molecular dynamics (MD) and docking algorithms, it is necessary to elucidate how the individual nucleobases adsorb on hard substrates at the fundamental level. Therefore, an ab initio study of a nucleobase on an inorganic surface brings insight on the specific system and simultaneously produces the key ingredients to enable the multistep simulation of more sophisticated complex systems based on nucleic acids and host surfaces. Such an approach (ab initio + MD + docking) was recently proposed for proteins on surfaces.[12] Moved by the above considerations, we have undertaken a first-principle density functional theory (DFT) study of cytosine on Au(111). The ultimate goal is the generation of a classical force field for atomistic MD simulations of a DNA molecule on Au(111): we remark that such a force field, derived from electronic structure calculations, is not currently available. The MD simulation should, in turn, be the link to large-scale docking simulations in a multistep approach to DNA/surface interaction.[12] This intermediate work is devoted to revealing the microscopic nature of the interaction between cytosine and the Au(111) surface, and particularly to quantifying the importance of dispersion interactions in order to develop a reliable multistep method.

The coupling between nucleobases and the Au(111) surface is expected to be dominated by dispersion interactions. The shortcomings of DFT in this respect have limited the amount of work done and the accuracy of theoretical descriptions. A recent attempt[13] to include a van der Waals (vdW) term in the total energy in calculations on supramolecular structures on gold indicates the importance of this contribution, even if it can be encompassed in just the simplest form. This evidence

Received: June 14, 2012
Revised: September 11, 2012
Published: September 17, 2012

prompted us to investigate cytosine on Au(111) including the vdW interaction, which was not done in previous studies of the same system.[3,14] We could do so by benefiting of a recent implementation of a vdW-corrected functional for DFT calculations, named vdW-DF.[15−20] Indeed, we reveal significant differences in the adsorption energies computed with and without the vdW term, which is particularly relevant in view of developing a force field. In addition, we even find qualitative difference in the adsorption geometries with respect to previous calculations done without vdW interaction, with our results in better agreement with experimental data. In particular, we find that a tilted geometry is always favorable relative to a parallel geometry: this has implications for monolayer formation and contradicts a common opinion that the surface−molecule interaction gives only a minor contribution to the monolayer formation energy.

METHOD

Computational Approach. We performed gradient-corrected DFT calculations of cytosine adsorbed at the Au(111) surface and in the gas phase with the quantum-espresso package[21] (version 4.3), using the Perdew−Burke−Ernzerhof (PBE) exchange correlation functional[22] and the vdW-DF functional.[16,17,20] We chose a plane wave basis set with a cutoff of 25 Ry, and we described the electron−ion interaction with ultrasoft pseudopotentials.[23] The surface was modeled with a slab of four Au layers with a periodically repeated $6 \times 3\sqrt{3}$ surface supercell, with 36 atoms per layer. The lateral distance between two neighboring cytosine replicas was at least 11 Å, and the vacuum thickness in the direction perpendicular to the surface was at least 14 Å, enough to ensure no interaction between the molecule and the periodic images. The Brillouin zone (BZ) sums were calculated including 2 Monkhorst−Pack special k points in the irreducible wedge.[24,25] All the atomic coordinates of the molecule and of the four Au layers were relaxed until the forces vanished within 0.05 eV/Å. We also performed test calculations for guanine in the gas-phase with the same methodology adopted for adsorbed cytosine. Atomic charges and the amount of charge transfer were evaluated through Löwdin population analysis.[26]

For the tests on benzene/Au(111) and on the benzene dimer, we used the same computational details and accuracy criterion, with smaller supercells. For benzene/Au(111), we adopted the same supercell as in previous calculations of this system, namely a slab with 4 substrate layers with a $4 \times 2\sqrt{3}$ surface periodicity (16 atoms/layer),[27] in which the lateral replicas are separated by about 7 Å. For the benzene dimer we used a cubic supercell in which neighboring replicas are separated by about 10 Å in any spatial direction. These tests were aimed at assessing the performance of vdW-DF versus other DFT functionals and, within the vdW-DF scheme, of PBE[22] versus revPBE[28] pseudopotentials.

Throughout the work, for comparing different levels of theoretical description, we first relaxed the system at each level until the forces vanished within 0.05 eV/Å and then performed single-point electronic structure calculations at the same level keeping the system coordinates frozen and varying only the distance between the surface and the molecule (or between the two benzene molecules in the dimer).

We compared the adsorption energies of the investigated iterfaces to each other to determine the most favorable structures. Specifically, we evaluated the adsorption energy for each structure as: $E_{ads} = E_{int} - E_{mol} - E_{clean}$, where E_{int} is the total energy of a given interface, E_{mol} is the total energy of the relaxed gas-phase isolated molecule, and E_{clean} is the total energy of the relaxed clean Au(111) surface. The reaction of interface formation is favorable if $E_{ads} < 0$, in which case we speak of an adsorption energy gain. The higher the adsorption energy gain, the more likely the interface.

The vdW-DF Approach. The lack of dispersion interactions is a well-known shortcoming of standard exchange-correlation functionals in DFT. While this is a minor problem in crystalline inorganic materials where the interatomic interactions are essentially bonding-like, it is undoubtedly a major issue in organic materials, molecules, and molecules on inorganic surfaces. Despite the importance of vdW interactions, many studies have been done so far for molecules on surfaces by standard DFT. Here we have adopted a new approach that self-consistently includes vdW terms in electronic structure calculations. We show that the approach not only gives significant adsorption energy differences relative to standard DFT, but also qualitatively changes the adsorption landscape for cytosine on Au(111).

The vdw-DF approach implies by construction the choice of the revPBE exchange correlation functional, which minimizes the error in the exchange part due to fictitious contributions in the vdW regime.[16] We have adopted this choice in all our calculations for atomic optimization and self-consistent electronic structure determination of molecules and interfaces. However, we could not generate vdW-DF pseudopotentials for the atoms that appear in our structures. Therefore, we had to use PBE or revPBE pseudopotentials, both of which do not exactly match the self-consistent calculations for the target systems. Well-tested PBE pseudopotentials are available in databases for all the elements in our systems. We produced revPBE pseudopotentials with a utility in the quantum-espresso package. We then tested the performance of vdW-DF calculations in conjunction with either PBE or revPBE pseudopotentials. We performed such tests on two benchmark systems: a benzene molecule adsorbed on Au(111), and a benzene dimer in vacuum. In the following, we label the calculations with different functionals and pseudopotentials as PBE/PBE, PBE/vdW-DF, and revPBE/vdW-DF; in each label, the first functional denotes the exchange and correlation terms adopted in the calculations for the generation of the pseudopotentials, and the second functional denotes the exchange and correlation terms adopted in our DFT calculations for the molecules and interfaces. Note that this is different from testing the use of different exchange terms in the vdW-DF functional.[18,29,30] Furthermore, this problem of matching the vdW-DF functional with the functional used in the pseudopotential generation was not encountered in previous work for the self-consistent use of the vdW-DF functional,[17,29] because of the use of the projector augmented wave (PAW) approach.

First of all we tested the vdw-DF approach with PBE and revPBE pseudopotentials for Au bulk, because the correct description of the lattice constant of solids is a tricky point of any DFT implementation beyond the local density approximation (LDA). The lattice constant of bulk gold obtained by us is 4.24 Å, 4.26 Å, and 4.06 Å with PBE/vdW-DF, revPBE/vdW-DF, and LDA/LDA, respectively. Our PBE/vdW-DF value, which was obtained with a PBE pseudopotential but with the revPBE exchange in the vdW-DF functional for the bulk calculation, is comparable to that obtained by other investigators with the PBE exchange in the vdW-DF func-

tional.[31] The PBE/vdW-DF and revPBE/vdW-DF values are overestimated relative to the experimental (4.08 Å)[32] and LDA (4.06 Å) data. The overestimation is even higher than our PBE/PBE result (4.17 Å).

Then we performed test calculations for the eclipsed benzene dimer, which was suggested[30] to be very sensible to the choice of the exchange part, PBE or revPBE, in the vdW-DF functional. After relaxing the structure of the parallel dimer as explained in the previous subsection, we performed a series of single-point electronic structure calculations at fixed intra-benzene coordinates, varying only the vertical distance between the two benzene molecules in the dimer to obtain the equilibrium distance and the energetics. We did so with both PBE/vdW-DF and revPBE/vdW-DF. The results, illustrated in Figure 1, show that the PBE and revPBE pseudopotentials, used

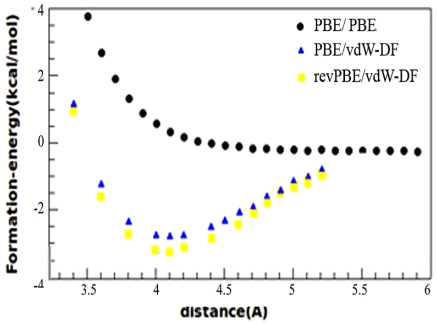


Figure 1. Formation energy as a function of the interbenzene vertical distance for a C_6H_6 dimer computed by DFT with various functionals, with and without vdW terms. As expected, PBE/PBE does not even give an equilibrium distance, while the vdW terms correctly give an energy minimum.

within the self-consistent vdW-DF scheme, perform very similar to each other, giving the same equilibrium distance and dimer formation energy, which are also in agreement with other calculations.[30] The dimer formation energy (−2.85 kcal/mol in PBE/vdW-DF results) is overestimated relative to CCSD(T) (−1.43 kcal/mol[16,33,34]) and QMC (−0.5 kcal/mol[35]) data. Yet, we obtain an improvement with respect to the MP2 description, which yields a formation energy of −4.66 kcal/mol.[16,33,34] Note that the PBE/PBE description yields a very shallow minimum at a large interbenzene distance.

The last part of the test to choose the details of our vdW-DF approach was devoted to a benzene molecule on Au(111), for which DFT results with and without vdW interactions are available[27] for benchmark. The results for the formation energy from single-point electronic structure calculations at different surface−molecule distances are reported in Figure 2. We note that both PBE/vdW-DF and revPBE/vdW-DF approaches yield the same accuracy for the equilibrium distance and formation energy; a deeper analysis, not shown, also indicates the same structural data at equilibrium. Our value for the formation energy at the equilibrium distance, 13.9 kcal/mol, is in good agreement with experimental results (13.8−14.7 kcal/mol), better than in previous vdW calculations.[27,31,36,37] As the two tested approaches give the same results for the adsorption of an aromatic molecule on Au(111), which is quite similar to our target system cytosine/Au(111), we decided to eventually use the PBE/vdW-DF method. The reason for this choice is multiple: (i) the database PBE pseudopotentials are well tested

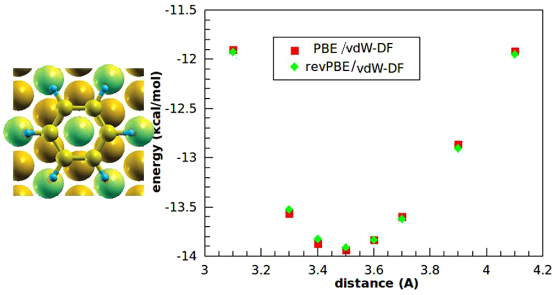


Figure 2. Adsorption energy for the C_6H_6/Au(111) system versus the molecule−surface vertical distance, calculated with the PBE/vdW-DF and revPBE/vdW-DF methods. The left panel illustrates a top view of the optimized geometry of the system: the aromatic ring of the benzene molecule is horizontal with respect to the surface, and the center of the ring lies above a top-layer Au atom. The large green and golden spheres represent Au atoms on the first layer and on the other three layers, respectively; the small yellow ad cyan spheres represent C and H atoms in the molecule, respectively.

for several systems; (ii) PBE pseudopotentials are the most reliable for solids; and (iii) PBE/PBE results have been published for similar systems and are useful as benchmarks.[24,31,38]

RESULTS AND DISCUSSION

Gas-Phase Cytosine and Guanine. The total energy of isolated cytosine in the same supercell and with the same computational accuracy as the cytosine/Au(111) systems is needed in order to compute the formation energy. The electronic structure obtained at the same theoretical level is a starting point to understand the modifications induced by the substrate.

We performed a full atomic relaxation of a cytosine molecule in vacuum with the PBE/PBE and PBE/vdW-DF methods. Our structural results are reported in Table 1 and compared to MP2 calculations.[39] Deviations of our results from reference data are similar with and without vdW corrections, as could be expected for single gas-phase molecules where vdW terms should not be effective. We find deviations from MP2 results below 2% for the bond lengths and below 0.75% for the planar angles. The differences between the PBE/PBE and PBE/vdW-DF methods are below 0.75%.

The electronic density of states (DOS) of cytosine obtained at the PBE/PBE and PBE/vdW-DF levels is plotted in Figure 3a, where the insets illustrate PBE/vdW-DF frontier orbitals. The DOS curves in Figure 3 are aligned at the deepest occupied orbital and the PBE/vdW-DF highest occupied orbital (HOMO) is taken as the origin of the energy scale. We note just minor energy shifts between the two computational levels, which do not affect the shape of the molecular orbitals.

Although our attention in this work is focused on cytosine/Au(111), we also validated the methodology on another DNA base for a better assessment of this approach in DNA-related systems. The guanine DOS computed with both PBE/PBE and PBE/vdW-DF is presented in Figure 3b. We note negligible differences in the energy range corresponding to occupied orbitals, with small shifts but no major rearrangements of the peaks (levels). The lowest-energy unoccupied orbitals have, instead, a larger spread in the PBE/vdW-DF results, and there are inversions of levels between the two descriptions. In

Table 1. Bond Lengths and Angles[a] for Cytosine in Vacuum and Adsorbed on Au(111) from Our PBE/PBE and PBE/vdW-DF Calculations and Reference MP2 Data[39]

	gas-phase cytosine			Cyt/Au(111)	
	MP2	PBE/PBE	PBE/vdW-DF	PBE/PBE	PBE/vdW-DF
Bond Length (Å)					
N1−C1	1.413	1.436	1.445	1.432 (−0.3%)[b]	1.431 (−1.0%)
C2−O8	1.219	1.233	1.232	1.238 (+0.4%)	1.247 (+1.2%)
C2−N3	1.376	1.376	1.381	1.372 (−0.3%)	1.373 (−0.6%)
N3−C4	1.312	1.332	1.333	1.336 (+0.3%)	1.343 (+0.8%)
C4−N7	1.367	1.366	1.370	1.362 (−0.3%)	1.367 (−0.2%)
C4−C5	1.429	1.439	1.442	1.439 (0.0)	1.439 (−0.2%)
C5−C6	1.352	1.366	1.365	1.369 (+0.2%)	1.365 (0.0)
C6−N1	1.351	1.359	1.363	1.357 (−0.2%)	1.366 (+0.2%)
Angle (Degrees)					
O8−C2−N1	118.9	118.1	118.3	118.3	118.2
O8−C2−N3		125.7	125.6	125.6	125.1
C2−N3−C4	119.9	120.1	120.0	120.0	119.8
N3−C4−N7	117.0	117.0	116.7	116.7	116.7
N7−C4−C5	118.5	119.3	119.1	119.1	119.5
C4−C5−C6	116.2	116.3	116.5	116.5	116.7
N1−C2−N3	116.0	116.2	116.1	116.1	116.8
C5−C6−N1	119.6	119.9	119.8	119.8	119.9
N3−C4−C5	124.4	124.1	124.2	124.2	123.8
C6−N1−C2	123.8	123.4	121.3	121.3	122.9

[a]We adopt in this table and in the whole text the standard numbering of atoms in pyrimidine heterocycles: O8 is the oxygen atom attached to the C2 carbon atom; N7 is the nitrogen atom attached to the C4 carbon atom. We also show the atom labels later in the text. [b]The values in parentheses indicate the percentage variation of the bond lengths in adsorbed cytosine relative to gas-phase cytosine.

particular, the energy of the lowest unoccupied molecular orbital (LUMO) is smaller in PBE/vdW-DF than in PBE/PBE, resulting in a smaller PBE/vdW-DF HOMO−LUMO gap.

To better understand the origin of the inversion between unoccupied orbitals, we tried to change the calculation parameters, as the convergence request of energy and forces. The values of these parameters do not affect the shape of the MOs. Then we also performed additional single-point electronic structure calculations of gas-phase guanine. We took the two relaxed atomic configurations, resulting from PBE/PBE and PBE/vdW-DF structural optimizations, and for each of them we performed single-point electronic structure calculations with PBE/PBE and PBE/vdW-DF. The aim was to see whether the orbital inversion is an effect of the different structure or of the different theoretical level and possibly an artifact of vdW-DF. We also compared these results with others obtained with the use of the cc-pVTZ basis set of localized orbitals and different xc functionals (B3LYP, BHandH, M06-2x) in the software NWChem.[40] The NWChem results show that the LUMO, LUMO+1, and LUMO+2 energy values and orbital shapes are influenced by the choice of xc functional. The results of our calculations show that both changing the molecule structure and including the vdW interaction into the calculation may cause orbital inversion. Thus, virtual orbital inversion is not an artifact of vdW-DF but is a typical effect of changing the theory level also in other approaches. It is known that in systems with close energy levels[41] the inversion of virtual orbitals can create problems/errors in describing the interaction between the molecule and a substrate, if there is a charge transfer between surface and molecule. We do not expect conspicuous charge transfer between the Au(111) surface and the DNA bases. This point should, however, be accurately checked to ensure the reliability of the models for DNA/surface docking, which is beyond the scope of this work.

Cytosine/Au(111). By means of PBE/vdW-DF calculations, we optimized the atomic coordinates of cytosine on Au(111) with the O8 atom on bridge, center, and top adsorption sites. Selected optimized geometries are illustrated in Figure 4.

The initial configuration of the system was prepared by calculating the molecule−surface optimal vertical distance. This procedure was done for the bridge adsorption site with both PBE/PBE and PBE/vdW-DF, which also allows us to estimate the effect of vdW interactions in the specific system of interest. As explained in the Method section, we first optimized the atomic coordinates of the cytosine/Au(111) system starting from an arbitrary configuration with the molecule flat at 3.4 Å

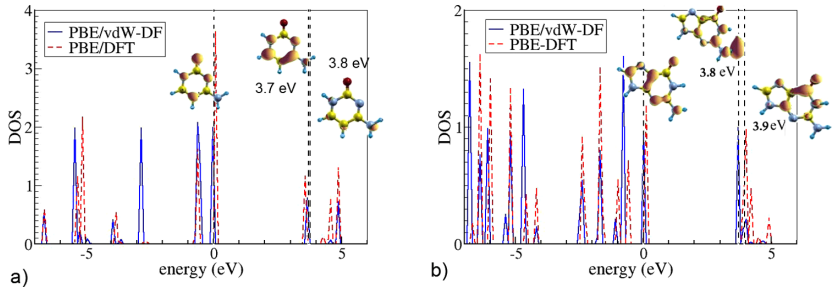


Figure 3. Electronic DOS of gas-phase cytosine (a) and guanine (b) molecules, calculated with (blue line) and without (red dots) the vdW interaction. For both guanine and cytosine separately, the DOS curves obtained in PBE/PBE and PBE/vdW-DF calculations are aligned by aligning the PBE/PBE and PBE/vdW-DF energies of the deepest occupied orbital, which is immune to vdW corrections and should therefore not suffer from the choice of the vdW-DF functional. In each panel, the HOMO resulting from the PBE/vdW-DF calculation is set as the origin of the energy scale. The energies reported for the unoccupied molecular orbitals are assigned according to this alignment. The insets show HOMO, LUMO, and LUMO +1 (marked by the vertical dashed lines) obtained from the PBE/vdW-DF calculation.

dx.doi.org/10.1021/jp305833c | J. Phys. Chem. C 2012, 116, 21366−21373

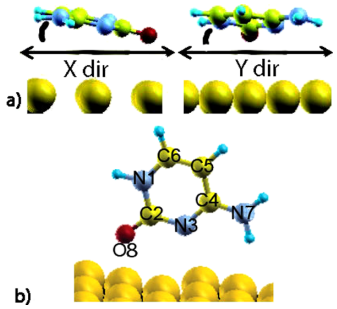


Figure 4. (a) Side view of the PBE/vdW-DF optimized configuration of horizontal cytosine on Au(111). The images show the two tilt directions. (b) Side view of the PBE/vdW-DF optimized configuration of vertical cytosine on Au(111) with atomic labels. The O8 atom of cytosine is on top of a Au substrate atom in both geometries.

above the (111) top substrate plane, and then we performed a series of single-point calculations at frozen internal coordinates by varying the molecule−surface vertical distance. The results for the formation energy are shown in Figure 5. The minima of

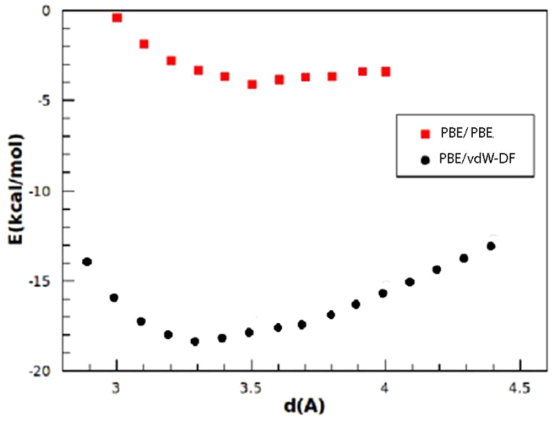


Figure 5. Adsorption energy versus cytosine−Au vertical distance for horizontal cytosine on Au(111) with the O8 atom in bridge position, starting from a parallel configuration of the molecule. The vertical distance is measured between the center of mass of the molecule and the average surface height. Black (red) circles (squares) denote PBE/vdW-DF (PBE/PBE) results. The PBE/vdW-DF results are in much better agreement with experimental data[3,41] that report adsorption energies between 25 and 36 kcal/mol in the high-coverage regime. Our results thus show that half of the adsorption energy is due to long-range dispersive interactions between the molecules and the surface.

the curves in Figure 5 give the equilibrium distance and formation energy. Note that PBE/PBE yields a very shallow minimum at 3.5 Å, while PBE/vdW-DF produces a deeper minimum at the distance of 3.3 Å, which is more compliant with a variety of results on aromatic systems and heterocycles.[42,43] As a pristine outcome of this work, we thus find a remarkable effect of vdW interactions, which were not so far evaluated for cytosine/Au(111).[1,2] We elaborate more on this issue in the following paragraphs.

The equilibrium distance of 3.3 Å was fixed as the starting condition for PBE/vdW-DF geometry optimizations of cytosine/Au(111) at different adsorption sites. At each adsorption site, we started with the molecular plane parallel to the surface plane, which we label horizontal adsorption geometry (Figure 4a), as opposed to vertical when the molecular plane is initially perpendicular to the surface plane (Figure 4b). In this work we focus on horizontal adsorption to unravel the effects of dispersion interactions. The computed equilibrium formation energy gains at the PBE/vdW-DF level, summarized in Table 2 for all the investigated configurations, show that the different adsorption sites are practically equivalent, as the formation energy difference is less than 1 kcal/mol. This result confirms the high mobility of cytosine molecules measured for low coverages, already reported from DFT calculations.[2] From the data in Table 2, we also draw another important conclusion, at odds with previous DFT reports:[3] the most favorable adsorption configuration is with the molecule tilted relative to the surface plane when the O8 atom is on a top (111) lattice site, while the DFT result indicates a geometry with the molecular plane parallel to the surface. This is a major effect of the vdW correction, which not only changes the gain, but substantially changes the adsorption geometry, which has always been a controversy with respect to experiments.[3]

To verify our results, we performed cross-check calculations as follows: (1) starting from the PBE/PBE equilibrium distance of 3.5 Å, we optimized the geometry at the PBE/PBE level; (2) starting from the PBE/vdW-DF distance of 3.3 Å, we optimized the geometry at the PBE/PBE level; (3) starting from the PBE/PBE distance of 3.5 Å, we optimized the geometry at the PBE/vdW-DF level. In all cases (1−3), the cytosine ring was initially parallel to the surface plane.

In case 1, the optimized cytosine remains parallel to the substrate. In case 3, the optimized cytosine end up in the same optimized configuration as starting from the smaller distance of 3.3 Å. In case 2, the optimized cytosine ends up in a vertical configuration. These cross-checks thus indicate that the PBE/vdW-DF tilted geometry is not an artifact of a starting distance too close to the surface, but is a stable result of this work, a real local minimum, in agreement with theoretical and experimental data in the monolayer regime.[3,38] Instead, the PBE/PBE parallel geometry that is obtained starting from the distance of 3.5 Å is simply due to the fact that at that distance there is no bound state, and the molecule practically does not feel the substrate and consequently remains in the initial state.

The major stability of the horizontal configuration with respect to the vertical one was already demonstrated for adsorption of nucleobases on Au(100),[14] even without the inclusion of vdW interaction in the calculations. However, for the Au(111) surface, the DFT result is not only quantitatively but also qualitatively wrong. The relative stability between adsorption configurations with different molecular orientations will be discussed thoroughly in a forthcoming work.

Interestingly, a test calculation of a different configuration reveals another fundamental qualitative change of the vdW-DF picture relative to DFT. We carried out PBE/PBE and PBE/vdW-DF geometry optimizations starting with the molecular plane perpendicular to the substrate and the O8 atom at a top lattice site (Figure 4b, vertical configuration). The PBE/vdW-DF geometry and energy gain are reported in Table 2. The data in Table 2 show that at the PBE/vdW-DF level, the vertical configuration is almost degenerate with the horizontal configuration, with an energy difference of only 0.7 kcal/mol. This trend is different when evaluating the energetics at the PBE/PBE level: the vertical adsorption geometry is favorable by

Table 2. Adsorption Energy and Molecule-Surface Relative Configuration (Distance and Inclination) for Cytosine Adsorbed on Au(111)

		final adsorption site[a]	final vertical distance (Å)	final tilt angle (x dir) (degrees)	final tilt angle (y dir) (degrees)	adsorption energy gain (kcal/mol)
horizontal cyt/Au(111)	PBE/vdW-DF	bridge	3.2	4.0	1.0	18.5
		center	3.1	9.0	3.5	18.5
	PBE/PBE	top	2.7	14.0	4.5	19.3
		bridge	3.3	0	0	4.5
vertical cyt/Au(111)	PBE/vdW-DF	top	2.6			18.6
	PBE/PBE	top	2.3			15.5

[a]The adsorption site is the site of the Au(111) triangular lattice on top of which the O8 atom of cytosine resides.

11 kcal/mol. The vdW contibution in the vertical configuration is about 3 kcal/mol, while it is about 14 kcal/mol in the horizontal configuration. The conclusion is that even if the adsorption energy gain is practically the same in the horizontal and *vertical* configurations, the interactions that are behind the two behaviors are different: the horizontal case is dominated by dispersion terms, which are completely lost in PBE/PBE, while in the vertical case the O8 atom interacts with the Au surface more strongly, which is also captured in PBE/PBE.

These results show that at the PBE/vdW-DF level, the vertical and horizontal configurations are practically degenerate, modifing the trend of PBE/PBE calculations. This is another major effect of the vdW terms, which we originally report in this work and will be important for theoretical studies of similar systems. On the contrary, DFT studies reported a major stability of the horizontal configuration with respect to the vertical one for adsorption of nucleobases on Au(100).[14] Including the vdW interaction in the calculations allows us to understand the different contributions to the interaction energy, and in the case of adsorption on Au(111), PBE/PBE completely misrepresents the relative stability of the horizontal adsorption geometry.

The PBE/vdW-DF optimized configuration with the lowest formation energy is illustrated in side view in Figure 4a. The changes of bond lengths and bond angles between gas-phase and Au(111)-adsorbed cytosine are tiny, mostly below 0.6% (Table 1), indicating a weak adsorption regime. There are, however, slightly larger deviations for some bonds. In particular, the length of the C2−O8 bond increases by 1.2%: this variation upon adsorption is a consequence of the weak interaction of the O8 atom with the surface. Similar trends were recently reported from the results of PBE/PBE calculations on the similar system cytosine/Au(100).[14] In the latter system Bogdan and co-workers also noted a shortening of the C4−N7 bond. These bond length changes in their work were accompanied by electron loss at N7 and electron gain at O8, which was interpreted as charge transfer between the two atoms. In our system cytosine/Au(111), we find at the PBE/vdW-DF level that both N7 and O8 gain electrons, 0.046 e^- and 0.055 e^-, respectively, with a total electron gain in the whole cytosine of 0.2 e^- (Löwdin population analysis). Thus, our results indicate that charge transfer occurs between the surface and the adsorbate and is not a purely intramolecular charge redistribution: in fact, there is a charge loss at the outermost substrate layer of 0.29 electrons and a total charge gain at the adsorbate of 0.20 electrons (the missing charge is due to spillout effects and transfer to the bulk). Although the Löwdin population analysis does not yield quantitatively accurate results for the atomic charges, it is reliable to detect relative trends and has been used in the past in this manor (e.g., refs 24 and 38). We have also checked that Bader's method[44] for computing atomic charges produces similar results.

By comparing the DOS of isolated and adsorbed cytosine for the most favorable structure of Figure 4a, we detect only minor perturbations in the electronic structure upon adsorption. Nevertheless, the differences are not negligible. In particular, around the upper edge of the Au *d* bands, namely between 2 and 1 eV below the Fermi level, the DOS shows a redistribution of the cytosine peaks. This is reported in Figure 6; see also the Supporting Information for a DOS with a finer Gaussian broadening. Interestingly, we note DOS molecular features in energy ranges where the Au DOS is also non-negligible. We inspected such energy ranges to search for hybrid molecule−Au orbitals that would reveal some electronic coupling. In fact, the small distance of 2.7 Å (Table 2) between the O8 atom of cytosine and the nearest Au surface atom suggests the possibility of an interaction mechanism beyond pure physisorption.

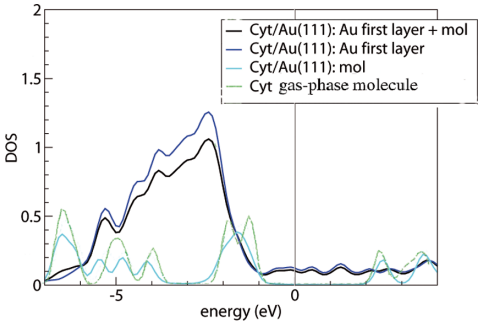


Figure 6. DOS for cytosine adsorbed on the Au(111) surface, computed by PBE/vdW-DF. The Fermi level is set at the origin of the energy scale. The deepest energy level, which is the same for the gas-phase molecule and for the adsorbed system, is used for alignment of the various curves. The green line is the DOS of the gas-phase molecule. The cyan, blue, and black lines represent the projections of the total cytosine/Au(111) DOS on cytosine, the top Au layer and the sum of cytosine with the top Au layer, respectively, as indicated in the legend. The projected DOS is computed by projecting the total DOS onto atomic orbitals and then summing over all the projections that constitute the subsystems of interest.

Figure 7 shows isosurface plots of relevant orbitals of the cytosine/Au(111) system. The systematic analysis of all the single-particle electron wave functions reveals the formation of

dx.doi.org/10.1021/jp305833c | J. Phys. Chem. C 2012, 116, 21366−21373

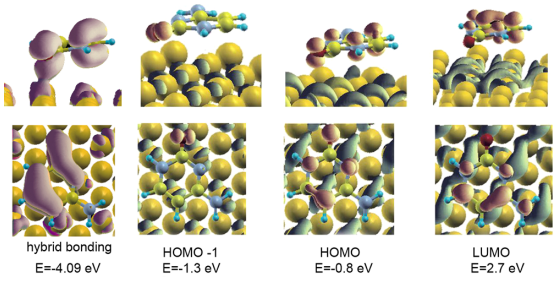


Figure 7. Isosurface plots of representative hybridized orbitals of cytosine adsorbed on Au(111) at the bridge site of the triangular substrate lattice. The leftmost panel shows an example of a bonding orbital between cytosine and the host surface, with a pink isosurface. The other panels show examples of hybrid, but not bonding, orbitals formed with the Au surface by the HOMO-1, HOMO, and LUMO: the isosurface portion localized on the surface is blue; the isosurface portion localized on the molecule is red.

bonding orbitals, of which the leftmost panel in Figure 7 encloses just an illustrative example.

The Newns–Anderson model for atomic and molecular chemisorption on metal surfaces predicts that the interaction between the localized molecular orbitals and the narrow *d* band of the metal produces hybrid orbitals of both bonding and antibonding type. The metal–HOMO bonding orbitals give a peak in the DOS at the lower-energy edge of the metal *d* bands and the metal–HOMO antibonding orbitals give a peak in the DOS at the upper-energy edge of the metal *d* bands.[45] This mechanism is found, e.g., in thiols chemisorbed on Au(111)[46] and was accurately discussed in the case of cysteine/Au(111).[24] In the case of cytosine adsorbed on Au(111) that we treat in this article, the situation is different. We reveal the formation of bonding orbitals, but the overall electronic hybridization mechanism does not comply with the Newns–Anderson picture. We do not detect bonding–antibonding splitting of the HOMO, which in the case of cysteine/Au(111) is on the order of 4 eV. Bonding orbitals are formed by various cytosine molecular orbitals with the Au *d* bands, but the energy level of the hybrid orbitals undergo only a slight shift relative to the gas-phase case. For instance, the bonding orbital at energy −4.09 eV shown in Figure 7 is found at the energy of −4.06 in gas-phase cysteine, and no corresponding antibonding orbital in cytosine/Au(111) is found between the upper edge of the Au *d* and the Fermi level.

Our findings point out a kind of interaction that is stronger than pure physisorption and that causes the formation of hybrid orbitals between the molecule and the surface. Yet, the mechanism is not that typical of molecular chemisorption. At some energies at which electron states appear in both the isolated cytosine (particularly, with a component on O8) and the clean Au(111) surface, the interaction is strong enough to allow for charge sharing and the formation of bonding orbitals. The analysis of the electron wave functions gives a remarkable evidence of adsorbate–substrate interaction beyond the pure dispersion regime, despite the fact that dispersion interactions are crucial for a correct description of the system. This evidence, along with results accumulating from recent works,[38,47−49] contradicts the common belief that homo- and heterocycles adsorb on metal surfaces by solely dispersion interactions. In addition, PBE/PBE calculations were not able to point out this effect, as the lack of vdW contribution does not bring the molecule close enough to the surface to allow the interaction between the O8 atom and Au. This is the same effect that causes the molecule to be tiled relative to the surface and the elongation of the C2–O8 bond length, as oxygen lowers towards the surface feeling the interaction with Au.

SUMMARY

In this paper we have examined the adsorption of cytosine on the Au(111) surface by means of DFT periodic supercell calculations of the structure and electronic properties, with particular attention given to revealing the effects of the vdW interaction.

Our results confirm the accuracy of the vdw-DF functional for the description of the cytosine/Au(111) system and show the importance of the vdW interaction in describing the adsorption of DNA bases on gold surfaces. We find that the PBE/vdW-DF method for the adsorption of cytosine on Au(111) in the horizontal configuration yields adsorption energy gains that are 4 times larger than those obtained with the PBE/PBE method. In addition and even more importantly, the PBE/vdW-DF equilibrium configuration is slightly tilted, in line with experiments on monolayer formation and at odds with previous DFT results, showing the predominance of dispersion terms in the interaction energy between the molecule and the surface.

Charge density and DOS analysis show evidence of hybridization between the molecule and the surface and the formation of weak bonding between the O8 atom and Au atoms, but excludes the formation of bonding–antibonding splitting indicative of chemisorption.

Our results are important for energetic quantities that are relevant in the determination of a suitable DNA–surface force field to describe the interaction between DNA and surfaces, while the vdW interaction was never taken into account in previous DFT studies that yielded energetic data.

ASSOCIATED CONTENT

Supporting Information

Plot of the DOS as in Figure 6 but with a finer Gaussian broadening. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rosa.difelice@unimore.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by the European Commission through project "DNA-Nanodevices" (Contract # FP6-029192), by the ESF through the COST Action MP0802, by the Italian Institute of Technology through project MOPROSURF and the Computational Platform, and by Fondazione Cassa di Risparmio di Modena through Progetto Internazionalizzazione 2011. The ISCRA staff at CINECA (Bologna, Italy) is acknowledged for computational facilities and technical support.

REFERENCES

(1) Sowerby, S. J.; Stockwell, P. A.; Heckl, W. M.; Petersen, G. B. *Orig. Life Evol. Biosph.* **2000**, *30*, 81.

(2) Ortmann, F.; Schmidt, W. G.; Bechstedt, F. *Phys. Rev. Lett.* **2005**, *95*, 186101.
(3) Piana, S.; Billic, A. *J. Phys. Chem. B* **2006**, *110*, 23467−23471.
(4) Seino, K.; Schmidt, W. G.; Preuss, M.; Bechstedt, F. *J. Phys. Chem. B* **2003**, *107*, 5031.
(5) Kilina, S.; Tretiak, S.; Yarotski, D. A.; Zhu, J.-X.; Modine, N.; Taylor, A.; Balatsky, A. V. *J. Phys. Chem. C* **2007**, *111*, 14541.
(6) Mignon, P.; Ugliengo, P.; Sodupe, M. *J. Phys. Chem. C* **2009**, *113*, 13741.
(7) Kelly, R. E. A.; Lukas, M.; Kantorovich, L. N.; Otero., R.; Xu, W.; Mura, M.; Laegsgaard, E.; Stensgaard, I.; Besenbacher, F. *J. Chem. Phys.* **2008**, *129*, 187707.
(8) Otero, R.; Lukas, M.; Kelly, R. E. A.; Xu, W.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Science* **2008**, *319*, 312−315.
(9) Wong, K.-Y.; Pettitt, B. M. *Theor. Chim. Acta* **2001**, *106*, 233.
(10) Zwolak, M.; Di Ventra, M. *Rev. Mod. Phys.* **2008**, *80*, 141.
(11) Hughes, G. A. *Nanomedicine* **2005**, *1*, 22.
(12) Di Felice, R.; Corni, S. *J. Phys. Chem. Lett.* **2011**, *2*, 1510.
(13) Nguyen, M.-N.; Pignedoli, C. A.; Treier, M.; Faselb, R.; Passerone, D. *Phys. Chem. Chem. Phys.* **2010**, *12*, 992.
(14) Bogdan, D.; Morari, C. *J. Phys. Chem. C* **2012**, *116*, 7351.
(15) Langreth, D. C.; Dion, M.; Rydberg, H.; Schroder, E.; Hyldgaard, P.; Lundqvist, B. I. *Int. J. Quantum Chem.* **2005**, *101*, 599−610.
(16) Dion, M.; Rydberg, H; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.
(17) Thonhauser, T.; Cooper, V. R.; Li, S.; Puzder, A.; Hyldgaard, P.; Langreth, D. C. *Phys. Rev. B* **2007**, *76*, 125112.
(18) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.
(19) Tkatchenko, A.; Romaner, L.; Hofmann, O. T.; Zojer, E.; Ambrosch-Draxl, C.; Scheffler, M. *MRS Bull.* **2010**, *35*, 435.
(20) Roman-Perez, G.; Soler, J. M. *Phys. Rev. Lett.* **2009**, *103*, 096102.
(21) Giannozzi, P.; et al. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.
(22) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(23) Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, R7892.
(24) Di Felice, R.; Selloni, A.; Molinari, E. *J. Phys. Chem. B* **2003**, *107*, 1151.
(25) In a previous work on the adsorption of an amino acid on the Au(111) surface,[24] four k points were employed to sample the two-dimensional Brillouin Zone, after convergence tests. In that work, the supercell in the direct space contained 12 Au atoms/layer. In this work, we have 36 Au atoms/layer, which means that the two-dimensional Brillouin Zone is smaller and can be sampled with a smaller number of k points. The ratio should be roughly 1/3, which we approximate by excess choosing 1/2. We believe that this should be sufficient, because of the large size of the supercell.
(26) (a) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365. (b) Löwdin, P.-O. *Adv. Quantum Chem.* **1970**, *5*, 185.
(27) Wellendorff, J.; Kelkkanen, A.; Mortensen, J. J.; Lundqvist, B. I.; Bligaard, T. *Top. Catal.* **2010**, *53*, 378.
(28) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 890.
(29) Klimeš, J.; Bowler, D. R.; Michaelides, A. *Phys. Rev. B* **2011**, *83*, 195131.
(30) Cooper, V. R.; *Phys. Rev. B* **2010**, *81*, 161104(R).
(31) Johnston, K.; Harmandaris, V. *J. Chem. Phys. C* **2011**, *115*, 1407.
(32) Landolt-Bornstein, *Structure Data of Elements and Intermetallic Phases*; New Series; Springer: Berlin, 1971; Vol. *IIIb*.
(33) Tsuzuki, S.; Honda, K; Mikami, M; Tanabe, K *J. Am. Chem. Soc.* **2002**, *124*, 104.
(34) Puzder, A; Dion, M; Langreth, D. C. *J. Chem. Phys.* **2006**, *124*, 164105.
(35) Sorella, S.; Casula, M.; Rocca, D. *J. Chem. Phys.* **2007**, *127*, 014105.
(36) Syomin, D.; Kim, J.; Koel, B. E.; Ellison, G. B. *J. Phys. Chem. B* **2001**, *105*, 8387−8394.
(37) Wetterer, S.; Lavrich, D.; Cummings, T.; Bernasek, S.; Scoles, G. *J. Phys. Chem. B* **1998**, *102*, 9266−9275.
(38) Iori, F.; Corni, S.; Di Felice, R. *J. Phys. Chem. C* **2008**, *112*, 13540.
(39) Kobayashi, R. *J. Phys. Chem. A* **1998**, *102*, 10813−10817.
(40) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. *Comput. Phys. Commun.* **2010**, *18*, 1477.
(41) Ostblom, M.; Liedberg, B.; Demers, L. M.; Mirkin, C. A. *J. Phys. Chem. B* **2005**, *109*, 15150−15160.
(42) Tautz, F. S. *Prog. Surf. Sci.* **2007**, *82*, 479.
(43) Henze, S. K. M.; Bauer, O.; Lee, T.-L.; Sokolowski, M.; Tautz, F. S. *Surf. Sci.* **2007**, *601*, 1566.
(44) Tang, W.; Sanville, E.; Henkelman, G. *J. Phys.: Condens. Matter* **2009**, *21*, 084204.
(45) Hammer, B.; Nørskov, J. K. Theory of Adorption and Surface Reactions. In *Chemisorption and Reactivity of Supported Clusters and Thin Films*; Lambert, R. M., Pacchioni, G., Eds.; Kluwer Academic Publishers: The Netherlands, 1997.
(46) Vargas, M. C.; Giannozzi, P.; Selloni, A.; Scoles, G. *J. Phys. Chem. B* **2001**, *105*, 9509.
(47) Ferretti, A.; Calzolari, A.; Di Felice, R.; Ruini, A.; Molinari, E.; Baldacchini, C.; Betti, M. G. *Phys. Rev. Lett.* **2007**, *99*, 046802.
(48) Lorente, N.; Hedouin, M. F. G.; Palmer, R. E.; Persson, M. *Phys. Rev. B* **2003**, *68*, 155401.
(49) Toyoda, K.; Hamada, I.; Lee, K.; Yanagisawa, S.; Morikawa, Y. *J. Chem. Phys.* **2010**, *132*, 134703.

Chapter 5

The interaction of DNA bases with the Au(111) surface

After assessing DFT details for cytosine adsorbed on Au(111), we extended our work to all the bases. For each DNA base we performed vdW-DF calculations for different adsorption sites and geometries, searching for a correct description of the possible short-range interaction between O and N atoms and gold. Again, as was done for cytosine, the most favorable configurations were re-optimized with the PBE functional, in order to clarify the vdW effects. The coherence of the results obtained and the agreement that emerges between theory and experiments demonstrate the importance of vdW interaction when dealing with such systems. First of all, all four bases were found to adsorb preferentially horizontally on the surface, in agreement with experimental results on monolayers [87, 88, 89, 90, 91, 92] and contrary to what is obtained with PBE calculations. The adsorption order of the bases is in agreement with the results obtained from temperature desorption experiments [94]. This result is a consequence of the vdW interaction inclusion, as PBE calculations fail in reproducing the correct adsorption order. Guanine is the base that interacts most strongly with the surface, followed by adenine, cytosine and thymine. Different components of the energy are determinant for the final adsorption energies of different bases: in particular the difference between adenine and cytosine stands in the dimension of the molecule and in the presence of O atom; the consequences are a larger dispersion term for adenine v.s. the stronger short-range interaction of cytosine O atom. The final proximity of

adsorption energy values, in perfect agreement with desorption experiments, is a consequence of the correct evaluation of the ratio between these two interaction. Finally, formation energy at different adsorption sites were found to differ by less than 1.5 kcal/mol, confirming literature results on mobility on Au surface[27]. The large mobility of DNA bases on gold could be a consequence of weak interaction between the two: the adsorption energy values obtained with vdW-Df calculations, systematically larger with respect to PBE ones, rejects this interpretation. In conclusion, the picture drawn by our calculations seems finally to allow a correct description of the interaction of DNA with surfaces, a result that was never obtained before. We then performed a deeper study of the systems reproduced, analyzing of the electronic structure. It already emerged from the study on cytosine how this further deepening in the analysis is important to gain a real knowledge on the computed system, often disregarded in the major concern of reproducing experimental results. The electronic DOS and the visualization of wave functions show that O and not saturated N atoms are able to create bonding orbitals with the gold surface in all the different configurations, as we already presented for horizontal cytosine in Chapter 4. The most pronounces DOS changes upon adsorption regard guanine and cytosine. These two bases are the ones adsorbed tilted with respect to the surface, with their internal structure more sensibly modified by the interaction with gold. Nevertheless, for all the bases the MO in the energy range of gold d bands create bonding orbitals with the surface, without the splitting in a bonding-antibondig couple predicted for chemisorption.

In conclusion, the adsorption mechanism on Au(111) disclosed in Chapter 4 for cytosine has been generalized to all the nucleobases.

The work has been published in 2013 in the Journal of Chemical Theory and Computation and the reprint is attached as an integral part of this PhD dissertation.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Interaction of Nucleic Acid Bases with the Au(111) Surface

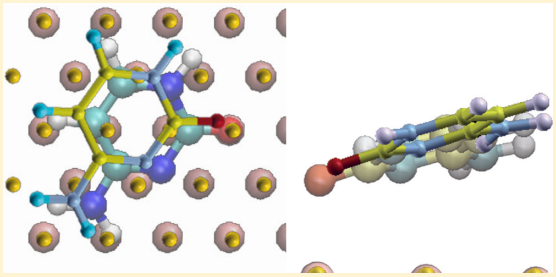
Marta Rosa,[†,‡] Stefano Corni,[†] and Rosa Di Felice*[,†]

[†]Center S3, CNR Institute of Nanoscience, Via Campi 213/A, 41125 Modena, Italy

[‡]Department of Physics, University of Modena and Reggio Emilia, 41125 Modena, Italy

Supporting Information

ABSTRACT: The fate of an individual DNA molecule when it is deposited on a hard inorganic surface in a "dry" environment is unknown, while it is a crucial determinant for nanotechnology applications of nucleic acids. In the absence of experimental approaches that are able to unravel the three-dimensional atomic structure of the target system, here we tackle the first step toward a computational solution of the problem. By using first-principles quantum mechanical calculations of the four nucleobases on the Au(111) surface, we present results for the geometries, energetics, and electronic structure, in view of developing a force field that will enable classical simulations of DNA on Au(111) to investigate the structural modifications of the duplex in these non-native conditions. We fully characterize each system at the individual level. We find that van der Waals interactions are crucial for a correct description of the geometry and energetics. However, the mechanism of adsorption is well beyond pure dispersion interactions. Indeed, we find charge sharing between the substrate and the adsorbate, the formation of hybrid orbitals, and even bonding orbitals. Yet, this molecule−surface association is qualitatively distinct from the thiol adsorption mechanism: we discuss such differences and also the relation to the adsorption mechanism of pure aromatic molecules.



1. INTRODUCTION

DNA, the molecule of life, has been a fascinating research subject since its discovery.[1,2] The determination of its three-dimensional structure in 1953[3] boosted investigations to understand the miriad of biological and medical phenomena in which it is involved, such as DNA−protein interactions for genetic transcription and replication, the evolution of the species, cancer development and treatment, genetic diseases due to mutations and many others. Research on nucleic acids has always been inherently interdisciplinary, encompassing biology, chemistry, and medicine.

In recent years, new applications of DNA are envisaged, which require the usage of physics and nanotechnology tools. In fact, given its intrinsic task of storing and translating information, scientists have started to ask the question whether its physico−chemical nature makes it able to store and transfer charges as well. Concepts and achievements in this direction are summarized in recent reviews.[4−7] Several new techniques (e.g., DNA chips) and applications (e.g., molecular electronics) require the interaction of DNA molecules with nonliving entities, such as inorganic substrates and fluorophores. The interaction of DNA molecules with inorganic surfaces is particularly puzzling, because the "hard" environment, so different from the natural solution environment of the cells, may induce denaturation and unfolding. To mention only a couple of general examples: DNA molecules are deposited "horizontally" on metal or insulating substrates for scanning tunneling microscopy/spectroscopy (STM/STS)[8−10] and atomic force microscopy (AFM)[4,11,12] investigations of the morphology and electronic structure; thiol-functionalized DNA molecules are attached "perpendicularly" to metal electrodes to measure the charge transport capabilities.[13−16] The height of double-stranded DNA molecules deposited "horizontally" on a hard substrate, measured by AFM, is about 50% of the diameter in solution,[11,12] which suggests possible unfolding of the nucleic acid in such experimental conditions. However, no confirmation of this hypothesis exists. Knowledge of the structure of DNA molecules on a hard substrate is extremely important to determine their ability to conduct charges in setups relevant for nanotechnologies, because the electronic structure and transfer rates are extremely sensitive to the conformational details.[17−20] This knowledge is not accessible through conventional methods for the determination of the three-dimensional atomic structure of biological molecules, such as X-ray or NMR, which are not compatible with the substrate environment. Therefore, we have undertaken a long-term plan to investigate this problem by computational means.

Classical molecular dynamics (MD) atomistic simulations are in principle a method of choice. They have successfully been employed to describe the unfolding of biological molecules in solution.[21−23] However, the crucial point in any classical MD methodology is the parametric force field (FF), which is the potential energy necessary for the evolution of the atomic

Received: March 26, 2013
Published: August 27, 2013

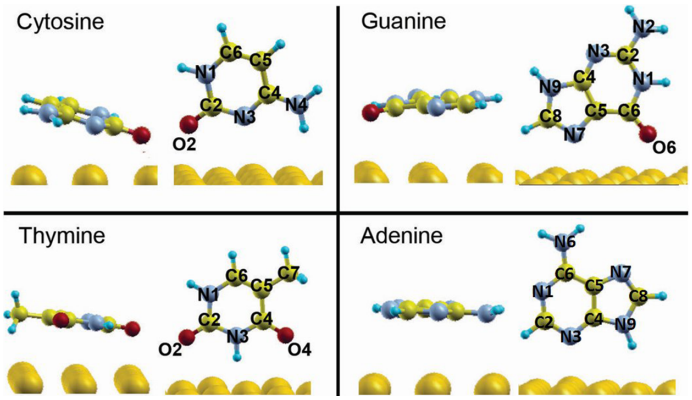


Figure 1. Optimized vdW−DF three-dimensional structures for the most favorable horizontal (left part in each panel) and vertical (right in each panel) adsorption configurations among those considered by us for cytosine (C), guanine (G), thymine (T), and adenine(A) on Au(111). Only the top atomic plane of the substrate is shown (large golden spheres). Small yellow, cyan, red, and turquoise spheres represent C, N, O, and H atoms of the molecule.

coordinates in time according to Newton's equations.[24,25] Reliable force fields exist for liquid, solid, and molecular materials, but the description of the interaction between molecules and hard inorganic surfaces is still in its infancy. Our group has developed force fields for the interaction of proteins with the Au(111) surface,[26,27] and we have recently reviewed our computational approach in this context, as well as similar efforts by other groups.[28] We now plan to extend this approach to the interaction of nucleic acids with the Au(111) surface, which requires specific parametrization due to different functional groups and heterocycles of the bases. As described elsewhere,[28] the first step of a viable multistep methodology is the derivation of a classical force field by means of density functional theory (DFT) calculations of small molecules on the target surface. The outcome of this approach sheds light not only on the energetics that is the basis of the force field but also on the adsorption mechanisms and charge redistribution.[29] This article pursues this objective for the adsorption of DNA bases on the Au(111) surface. Our research connects to a variety of lively investigations for the development of DNA microarray techniques,[30] portable cost-effective DNA sequencing,[31] targeted drug delivery,[32] formation of self-assembled superstructures,[33−35] and the development of life under prebiotic conditions.[36] It is also related to recent studies to assess atomistic computational methods for unraveling crucial features linked to the geometry and to the electronic structure.[33,35,37−39]

We have already described the "horizontal" adsorption of cytosine on Au(111) by DFT as a model system to prove the crucial role of van der Waals (vdW) interactions in the description of DNA bases on surfaces.[40] We have shown that several configurations are quasi-isoenergetic and need to be considered for the derivation of a new classical FF. The four bases cytosine, guanine, thymine, and adenine need to be studied one by one on the same footing, because experimental data show that they behave in a very different way on Au(111).[41−47] Guanine and adenine self-assemble in monolayer structures.[43] Cytosine forms filaments rather than self-assembled monolayers.[45−47] Thymine behaves in an intermediate way, choosing filament or monolayer configurations depending on the molecular density on the surface.[44]

We find that the inclusion of vdW interactions affects differently the adsorption of the four bases, modifying more the internal geometries and adsorption configurations of cytosine and guanine than those of thymine and adenine. We reveal significant differences in the adsorption energies computed with and without vdW terms, which is particularly relevant in view of developing a force field. Our results on the adsorption energies of cytosine (C), guanine (G), thymine (T), and adenine (A) on Au(111) give adsorption strengths in the order $T < C \sim A < G$, which is in line with the outcome of desorption experiments, $T < C \leq A < G$:[48] this is a remarkable success of the vdW dscription, which could not be achieved with gradient-corrected (GGA) or hybrid DFT functionals. Furthermore, our results show that "horizontal" adsorption geometries are favorable with respect to "vertical" orientations for each of the four bases: this has implications for the abilities of monolayer formation. The outcome of our investigation points to an adsorption mechanism that entails adsorbate−substrate electronic mixing and thus cannot be ascribed to solely dispersion interactions.

2. METHOD

We performed gradient corrected DFT calculations of the four DNA bases adsorbed at the Au(111) surface (Figure 1) with the quantum−espresso package version 5.0,[49] using the PBE exchange correlation functional[50] and the vdW−DF functional.[51] The surface was modeled with a slab of four Au layers with a periodically repeated $6 \times 3\sqrt{3}$ surface supercell, with 36 atoms per layer. The lateral size of the supercell was 17.58 × 15.22 Å^2 and 18.00 × 15.59 Å^2 in PBE and vdW−DF calculations of the base/Au(111) interfaces, respectively, obtained from the corresponding bulk fcc Au equilibrium lattice parameter (4.14 Å and 4.24 Å, respectively).

The adsorbate−adsorbate lateral distance between two neighboring replicas was at least 10 Å and the vacuum thickness in the direction perpendicular to the surface was at least 11 Å. Tests in a larger supercell were performed, to check that the chosen size assures no interaction between the molecule and its periodic images. We chose a plane wave basis set with a cutoff of 25 Ry,[29] and we described the electron−ion interaction with ultrasoft pseudopotentials.[52] The valence shells were 2*s* and 2*p* for C, N, and O; 5*d* and 6*s* for Au with a scalar relativistic treatment that was assessed on surface chemistry and

dx.doi.org/10.1021/ct4002416 | J. Chem. Theory Comput. 2013, 9, 4552−4561

Table 1. Calculated Adsorption Energies, Molecule−Surface Distances and Orientations: The Quantities without (with) PBE Superscript Result from vdW−DF (PBE) Calculations[a]

structure[b]	α^c (deg)	d^d (Å)	$d^{PBE,c}$ (Å)	E_{ads} (kcal/mol)	E_{ads}^{PBE} (kcal/mol)
		Horizontal Configurations			
$_h$G@Au(111)^t	83.0	3.0		**23.3**	
$_h$G@Au(111)^f	87.0	3.2		23.1	
$_h$G@Au(111)^b	86.0	3.2	3.3	23.0	5.9
$_h$A@Au(111)^b	87.0	3.2		**19.7**	
$_h$A@Au(111)^t	87.0	3.2	3.3	19.4	3.9
$_h$C@Au(111)^t	76.0	2.7		**19.7**	
$_h$C@Au(111)^b	86.0	3.1	3.4	18.2	4.5
$_h$C@Au(111)^f	81.0	3.1		18.0	
$_h$T@Au(111)$^{t+f}$	84.5	3.2÷3.6	3.3÷4.0	**17.2**	2.7
$_h$T@Au(111)^f	85.6	3.2÷		17.1	
$_h$T@Au(111)$^{f+t}$	90.0	3.4÷3.4		17.1	
$_h$T@Au(111)$^{t+t}$	84.1	3.2÷3.4		17.0	
$_h$T@Au(111)$^{b+f}$	86.6	3.4÷3.6		16.9	
		Vertical Configurations			
$_v$G@Au(111)$^{t+N7t}$	0.0	2.6	2.9	18.6	10.9
$_v$G@Au(111)$^{b+N7t}$	0.0	2.6		18.1	
$_v$G@Au(111)$^{b+N2t}$	41.0	3.0	3.1	16.6	7.9
$_v$G@Au(111)$^{f+N2t}$	41.0	3.0		16.5	
$_v$G@Au(111)^{N3t}	30.0	3.3		15.2	
$_v$A@Au(111)^{N3t}	0.0	2.7	2.5	17.4	10.9
$_v$A@Au(111)^{N1t}	0.0	2.8		14.4	
$_v$C@Au(111)$^{t+N3t}$	32.0	2.6	2.2	19.2	15.5
$_v$C@Au(111)^t	22.0	2.5		12.2	
$_v$C@Au(111)^f	0.0	2.4		10.9	
$_v$C@Au(111)^b	0.0	2.3		10.0	
$_v$T@Au(111)$^{t+b}$	0.0	3.1÷3.3	3.4÷3.5	11.0	2.6
$_v$T@Au(111)$^{f+b}$	0.0	3.2÷3.2		11.0	
$_v$T@Au(111)^{O4t}	0.0	2.8		10.1	
$_v$T@Au(111)^{O2t}	0.0	2.8		7.5	

[a]Bold font marks the lowest-energy vdW−DF structures for C, G, T, and A on Au(111). [b]Labeling conventions are explained in the text. [c]α is the inclination angle of the molecule relative to the vertical axis, from vdW−DF calculations. [d]d is the molecule−surface distance, evaluated for horizontal (vertical) configurations as the distance between the most reactive (nearest) molecular center and the Au atom below it in the outermost substrate plane.

nanostructure formation.[53,54] The Brillouin zone (BZ) sums were calculated including 2 Monkhorst−Pack special **k** points in the irreducible wedge. All the atomic coordinates were relaxed until each force vanished within 0.05 eV/Å. The technical details (basis set, Brillouin zone sums, pseudopotentials) were extensively validated elsewhere for cytosine/Au(111),[40] and based on the similarity of all the computed interfaces, we are confident that they are extendable with the same accuracy to all the systems in this work. The thickness of the Au slab is appropriate for describing interface effects in such extended structures.[55,56]

For comparing different levels of theoretical description, we first relaxed the system at each level until the forces vanished within 0.05 eV/Å and then performed single-point electronic structure calculations at the same level keeping the system coordinates frozen and varying only the distance between the surface and the molecule.

In the following, we denote with the term "formation energy" the difference between the total energy of the interface and the total energies of the isolated constituents. We use the term "adsorption energy" for the opposite of the formation energy: thus, the adsorption energy is negative when the formation of the interface is endothermic, it is positive when the formation of the interface is exothermic.

The lack of dispersion interactions is a well-known shortcoming of standard exchange−correlation functionals in DFT, and this is undoubtedly a major issue in organic materials, molecules, and molecules on inorganic surfaces. Despite the importance of vdW interactions, many studies were done so far for molecules on surfaces by DFT with standard GGA functionals. In this work, we have adopted an up-to-date approach that self-consistently includes vdW terms in electronic structure calculations.[40] Specifically, among recently proposed functionals to treat van der Waals effects within DFT calculations,[51,57−59] we have used the vdW−DF functional.[51] The use of this functional is becoming increasingly benchmarked and accepted for the accurate DFT description of interfaces between organic layers and metal substrates.[60−65] The method was assessed by us on the model system cytosine@Au(111), and here, we exploit the outcome of those tests.[40] In this work, we apply the same approach to the whole set of DNA bases adsorbed on Au(111) in both "horizontal" and "vertical" orientations relative to the substrate.

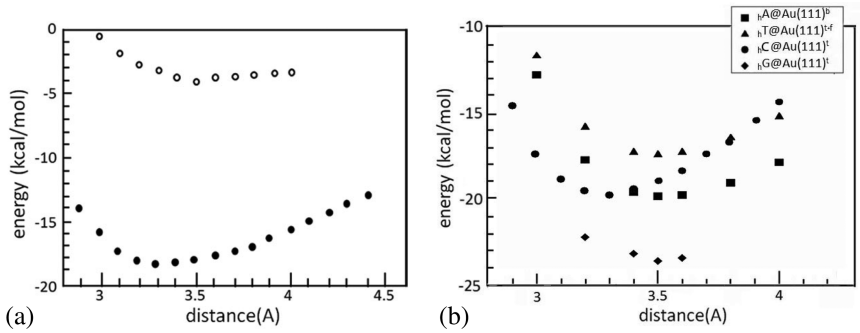


Figure 2. Formation energy versus molecule−Au vertical distance. The vertical distance is measured between the center of mass of the molecule and the average surface height. (a) PBE (circles) and vdW−DF (dots) results for ${}_h$C@Au(111)^b. The vdW−DF results are in much better agreement than the PBE ones with experimental data.[33,48] Half of the adsorption energy is due to long−range dispersive interactions between the molecules and the surface. (b) vdW−DF results for the four bases in lowest−energy horizontal configurations, as in the legend.

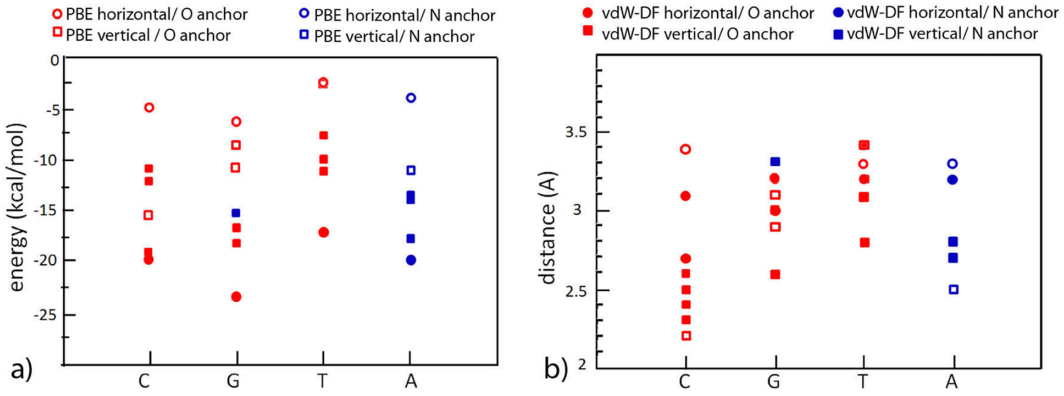


Figure 3. Formation energies in kcal/mol (a) and molecule−surface distances in Å (b) from vdW−DF calculations, for all the structures that we have considered, according to the values in Table 1. Multiple symbols (dots, circles, full squares, open squares) for a certain base report the multiple data from Table 1: accordingly, there are four full squares and three dots for C and similarly for the other bases. Red (blue) symbols represent geometries where an O (N) atom of the molecule is the closest to the surface.

3. RESULTS AND DISCUSSION

3.1. Energetics and Relaxed Geometries. Different adsorption sites (top, bridge, fcc) and adsorption geometries (horizontal and vertical) were sampled for each of the four DNA bases (see Figure S1, Supporting Information). The most favorable geometries are illustrated in Figure 1, which also defines the atom labeling that is used throughout the article.

We considered 13 horizontal configurations, obtained by placing the most reactive atomic center (an atom or a group of atoms) of each base on the three possible sites of the Au(111) triangular lattice. The most reactive atomic center is O2 and O6 for cytosine and guanine, respectively: three horizontal configurations were sampled for each of these two bases. We will label them in the following ${}_h$C@Au(111)^t, ${}_h$C@Au(111)^b, ${}_h$C@Au(111)^f, ${}_h$G@Au(111)^t, ${}_h$G@Au(111)^b, ${}_h$G@Au(111)^f, where the right superscript indicates the adsorption site for the O atom (t for top, b for bridge, f for fcc) and the left subscript h indicates a horizontal configuration (later we use the left subscript v for vertical configurations). Thymine has two most reactive atomic centers, O2 and O4: we considered five horizontal geometries ${}_h$T@Au(111) to allow for simultaneous reaction of these two centers. The labels used for the horizontal ${}_h$T@Au(111) structures are ${}_h$T@Au(111)$^{t+f}$, ${}_h$T@Au(111)$^{f+t}$, ${}_h$T@Au(111)$^{t+t}$, ${}_h$T@Au(111)^f, ${}_h$T@Au(111)$^{b+f}$, where the right superscript t+t ($t + f$, $f + t$, $b + f$) indicates that O2 and O4 are above equivalent top (inequivalent top and fcc, fcc and top, bridge and fcc) sites, while the right superscript f indicates that O2 only is above an fcc site of the Au(111) substrate. The most reactive center of adenine is N3; considering hindrance constraints, we identified only two horizontal configurations, ${}_h$A@Au(111)^b and ${}_h$A@Au(111)^t, namely with the N3 atom above a bridge and top position. For the vertical configurations, we considered more possibilities, because different inclinations are possible so that more than one atomic centers of the adsorbate reacts with the Au(111) substrate (see Figure S1, Supporting Information, and Table 1). The labels adopted for vertical configurations follow similar criteria as those of horizontal configurations, namely, (1) the superscript t, b, or f indicates that the most reactive atom center (O2 in C, O6 in G, and N3 in A) is above a top, bridge or fcc position of the Au(111) lattice; (2) thymine has two equally reactive atoms, O2 and O4, so we distinguish two vertical configurations in which one O atom is binding to gold (superscripts O2t and O4t); (3) to denote structures in which multiple molecular centers are close to the substrate we use the symbol +, and we explicitly indicate which atom is at which position.

dx.doi.org/10.1021/ct4002416 | J. Chem. Theory Comput. 2013, 9, 4552−4561

For each system, we fixed an initial geometry with the molecule at a certain distance from the surface and relaxed all the atomic coordinates. The molecule was exactly parallel (90-degrees tilt angle) and perpendicular (0-degrees tilt angle) to the surface in the initial conditions for the horizontal and vertical configurations, respectively. The initial molecule–surface distance, as measured from the most reactive (closest) atom of the adsorbate to the Au(111) lattice site below it for the horizontal (vertical) molecular orientations, was 3.3 Å (2.8 Å). These values were chosen on the basis of our recent results for C@Au(111).[40] The internal geometry of the adsorbate in all the initial conditions was flat: this implies that for horizontal configurations the starting adsorbate–substrate distance coincides with the distance between the plane of the molecule and the surface.

To verify that the system is not trapped in an irrelevant local minimum by the geometry optimization algorithm, for some configurations, we did a series of single-point calculations at frozen internal coordinates, starting from the optimized structure and varying only the molecule–surface vertical distance over a broad range of distances. The results of this procedure are illustrated for exemplifying cases in Figure 2. In Figure 2a, for the system ${}_{h}\mathrm{C@Au(111)}^{b}$, we show the relative performance of the vdW–DF and PBE functionals in the self-consistent calculations. The minima of the curves in Figure 2a give the equilibrium distance and formation energy. Note that PBE calculations yield a very shallow minimum at 3.5 Å, while vdW–DF calculations produce a deeper minimum at a distance between the center of the molecule and the surface of 3.3 Å, which is more compliant with a variety of results on aromatic systems and heterocycles.[66,67] Similar trends as those in Figure 2a are found for all the bases, which means that long-range interactions are fundamental for the correct description of the adsorption of DNA bases on surfaces. In Figure 2b, we visualize the relative behavior of the different bases: this plot identifies the order of adsorption strengths between the various structures. The same concepts are visualized in a different manner in Figure 3.

The analysis of Figures 2 and 3, along with Table 1 and the consideration of all the equilibrium structures, allow us to draw some general conclusions about the relative behavior of the four nucleobases on Au(111) and about the performance of density functional theory for these systems.

- DFT calculations that include the long-term vdW interaction reveal that a horizontal adsorption geometry is always preferred over a vertical mode, practically degenerate in the case of cytosine.[40] In fact, the full circles in Figure 3 lie always at lower/same energies than the full squares. This is a genuine result of our approach, while the relative energetics is erroneously described if the long-range vdW interaction is not included.[33] We remark that experimental results on monolayers indicate that DNA bases are horizontal relative to the substrate:[41,44,45,47] thus, our results are in line with observations.
- The adsorption energy from vdW–DF calculations is systematically larger than that from PBE calculations, by three to six times. This was already observed for horizontal C@Au(111) and is confirmed here over a much larger sample. There is one single exception, the case of vertical cytosine, which is separately discussed later.
- The molecule–surface distance (Table 1) is similar in horizontal and vertical adsorption configurations for the same base. It is also similar for the different bases. It is usually larger from PBE than from vdW–DF calculations, except for structures ${}_{v}\mathrm{C@Au(111)}^{t+N3t}$ and ${}_{v}\mathrm{A@Au(111)}^{N3t}$. vdW–DF results on adsorption distances are in fair agreement with a variety of results on aromatic systems and heterocycles,[66,67] while PBE results fail.
- The values of the interaction energy and the analysis of the electronic structure suggest a moderately strong molecule/substrate coupling, beyond the pure van der Waals regime, that would hint to a smaller distance between the molecule and the surface. This is, in fact, in line with the already known overestimation of equilibrium distances characteristic of the vdW–DF functional.[51,57,68,69]
- The vdW interaction has impact not only on the adsorption energies and molecule–surface distances but also affects the relative orientation of the adsorbate to the substrate. This is particularly true for cytosine[40] and guanine in horizontal configurations, in which the molecular plane is tilted with respect to the surface plane, while this is not the case in PBE calculations. The importance of van der Waals terms in DFT calculations of DNA bases on Cu(111) was recently analyzed, with qualitatively similar conclusions.[70]
- O atoms are the most reactive centers in the DNA bases, followed by N atoms. Cytosine, guanine, and thymine preferentially interact with the Au(111) surface through O atoms (O2, O6, and O2+O4, respectively), and also the N atoms that are not saturated by H atoms take part in the binding mode.[40] Adenine can interact with the Au(111) surface through N1, N3, and N6. O and N atoms overall prefer top adsorption sites, where they are able to form bonding orbitals with the gold *d* bands (see subsection 3.2). As a result of their favored position outside the aromatic ring, O atoms are the ones that stay closer to the surface.
- The adsorption energy of the horizontal configurations is fairly independent of the adsorption site: 17, 20, 20, and 23 kcal/mol for T, C, A, and G, respectively (Table 1). In fact, although O and N prefer top adsorption sites, the spread of adsorption energy among the different sampled geometries for a particular base and a particular orientation is rather small (1.5 kcal/mol), indicating a rather flat energy profile for lateral mobility, in line with mobility data from molecular dynamics simulations.[71] In vertical configurations there is more spreading, especially when N atoms are involved in the adsorption mode.
- Guanine is the DNA base that interacts most strongly with the surface, while thymine interacts most weakly: the order of adsorption strength is $T < C \sim A < G$. This result is in agreement with the trend from experimental results on monolayers,[48] while in PBE calculations the desorption order is $A < T < C < G$.

A common behavior among the different nucleobases can also be traced by comparing the adsorption geometries of each base on the surface and in the gas phase (Tables S1–S5 in Supporting Information). First of all, the distortion of the adsorbed molecule relative to the gas-phase molecule are more marked in vertical configurations than in horizontal configurations. In all the adsorption systems treated in this work the

C−O bond length is elongated when the molecule is close to the surface. Guanine is the molecule that undergoes the largest changes in bond lengths, while the opposite extreme is thymine. In all the bases, bond lengths and angles change less than 1.5% upon adsorption. These structural changes are comparable in PBE and vdW−DF calculations. In the remainder of this subsection, we discuss specifically each base.

Cytosine. For C@Au(111) in horizontal configurations, we refer the reader elsewhere.[40] We only summarize here the salient factors: (1) the preferred adsorption mode is at the top site, though the energy difference for the other two possible choices is tiny; (2) the molecule is tilted by 76° with respect to the vertical axis, namely 14° relative to the surface plane.

Among the viable vertical configurations, cytosine likes to approach the substrate with both the O2 and N3 atoms. In fact, as already noted, nitrogen atoms that are not saturated by H atoms can interact with the Au(111) surface if they are close enough to it. This behavior was recently described for other systems, such as imidazole on Au(111).[29]

The ${}_{v}C@Au(111)^{t+N3t}$ configuration is almost isoenergetic with the horizontal ones, within 0.4 kcal/mol; instead, the other bases adopt horizontal configurations more favorably (by 4.7, 6.2, and 2.3 kcal/mol for G, T, and A, respectively). In this configuration, both the O2 and N3 atoms are located above top sites of the Au(111) triangular lattice. Comparing the PBE and vdW−DF results for the adsorption energy of ${}_{v}C@Au(111)^{t+N3t}$, we argue that the interaction is mainly short-range. In fact, the small energy difference of 3.7 kcal/mol indicates that PBE performs rather well, which is not true in situations where vdW effects are strong (as in the other configurations, for which the difference in formation energy between vdW−DF and PBE calculations is larger). This good performance of the PBE functional is peculiar of configurations with a concomitant binding of O and N atoms, as we also describe for *G@Au*(111). Despite such a similarity between the PBE and vdW−DF adsorption energy values, the computed molecule−surface distances at the two different levels of theory are quite dissimilar. Normally vdW−DF distances are smaller than PBE distances, although they are still overestimated relative to experimental data;[51,57,68,69] yet, the ${}_{v}C@Au(111)^{t+N3t}$ structure is an exception to this trend, with a PBE distance of 2.2 Å, much smaller than the vdW−DF value of 2.6 Å. We are confident that the correct description of the geometry is closer to the vdW−DF result than to the PBE result.

Both the energetics and the electronic structure of the peculiar ${}_{v}C@Au(111)^{t+N3t}$ configuration are very similar to those of the other cytosine adsorption geometries. This suggests for all the sampled cytosine interfaces the occurrence of the same kind of short-range interaction between the adsorbate and the surface. The inclusion of the short-range vdW coupling in the DFT functional is crucial to attain a correct description: in fact, the vdW−DF coupling allows the molecule to get close enough to the surface in order to feel the correct surface field and establish the bonding between Au and O/N atoms. Therefore, both short-range and long-range effects are important for the binding mode of cytosine on Au(111).

For a deeper inspection of the relative energetics between C@Au(111) structures, we performed some test calculations starting from intermediate configurations between horizontal and vertical, with both O and N atoms close to the surface and the molecule tilted at different angles. These tests indicate that the molecule favors a horizontal orientation. Thus, although horizontal and vertical geometries have similar adsorption energies and can coexist, the energy profile is such that the horizontal adsorption mode is selected more frequently, or on a larger portion of the substrate.

Guanine. For what concerns horizontal configurations, guanine on Au(111) behaves very similarly to cytosine. The three adsorption sites available to the O6 atoms are practically isoenergetic, within 0.3 kcal/mol. The molecule is not exactly parallel to the surface plane. The O−Au vertical distance is 3.0 Å at the top site.

We calculated five vertical configurations, with the O atom, a N atom, or atoms of both species oriented toward the surface. The ${}_{v}G@Au(111)^{t+N7t}$ structure has the lowest formation energy among the vertical guanine adsorption geometries: it is almost degenerate with ${}_{v}G@Au(111)^{b+N7t}$ and has a formation energy difference of 2÷3 kcal/mol relative to the other ${}_{v}G@Au(111)$ structures (Table 1). The lowest energy vertical structure for guanine has both the O2 and N7 atoms above top sites of the substrate lattice and interacting with Au; this situation is comparable to that of the ${}_{v}C@Au(111)^{t+N3t}$ configuration discussed above.

As for cytosine, also for guanine the preferred (by 5÷8 kcal/mol) adsorption orientation is horizontal, with a tilt angle between 3 and 7 degrees relative to the surface plane.

Thymine. Thymine is the base that has the weaker interaction with the Au(111) surface, despite the presence of two O atoms that are potential strong binders to Au: the horizontal configurations have adsorption energies between 16.9 and 17.2 kcal/mol and the vertical configurations between 7.5 and 11.0 kcal/mol. The CH_3 group in our optimized structures is oriented with one H atom on the plane of the molecule, one above and one below (Figure 1). Calculations were done to check other possible orientations of the CH_3 group that would imply a minor steric hindrance and the possibility for thymine to get closer to the surface. Independently of the starting orientation of the CH_3 group, this rotates back to a situation with high steric hindrance. This is the origin of the comparatively weak interaction of thymine with the Au(111) surface.

The deepest energy minumum is found for the ${}_{h}T@Au(111)^{t+f}$ horizontal configuration with the O2 atom at a top site and the O4 atom at a fcc site. All the other computed horizontal thymine structure, including ${}_{h}T@Au(111)^{f+t}$ in which O2 and O4 are simply exchanged at fcc and top sites, are almost isoenergetic, within 0.3 kcal/mol.

In all the horizontal ${}_{h}T@Au(111)$ configurations, thymine is slightly inclined relative to the surface plane, by 2÷6 degrees. The O2 and O4 atoms of thymine remain quite far from the surface with respect to the O2 atom of cytosine, consistently with the smaller energy gain upon adsorption.

The vertical T@Au(111) configurations are higher in energy by at least 6.2 kcal/mol than the preferred horizontal structure. The lowest-energy vertical configuration is with both the O2 and O4 atoms close to the surface (at either top+bridge or fcc +bridge sites), followed by that with the O2−C bond perpendicular to the (111) plane.

Adenine. Adenine is the only DNA base without an O atom, so the choice of strucutures was based on locating the N3 and N7 atoms above lattice sites.

Similarly to the other bases, the difference in adsorpton energies between different sites is quite small, practically vanishing for the horizontal configurations and within 2.5 kcal/mol for the vertical configurations.

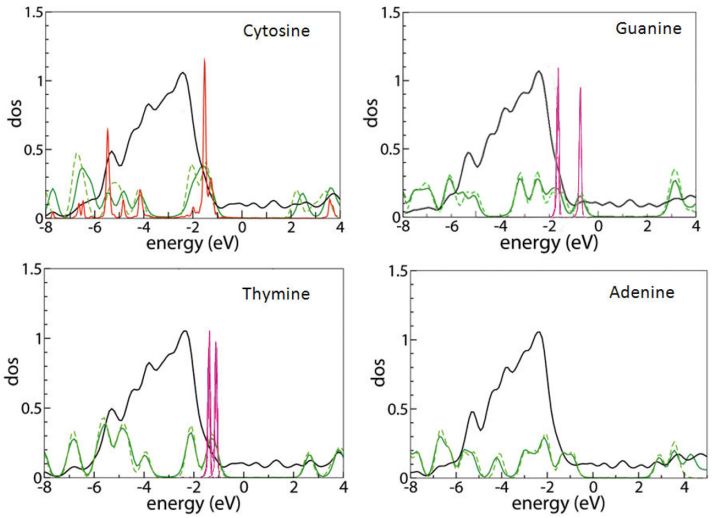


Figure 4. Density of states for the four bases adsorbed horizontally on the Au(111) surface, computed with the vdW−DF functional. In each plot, the Fermi level of the system is set at the origin of the energy scale and the deepest energy level, which is associated to the same orbital in the gas-phase molecule and in the adsorbed molecule, is used for alignment of the various curves. The green dashed (solid) line is the DOS of the gas-phase molecule (total DOS of the interface projected onto the molecule). The black line represents the sum of the projections of the total base/Au DOS on the adsorbate and on the outermost substrate layer. The red line in the top left panel is the projection of the cytosine/Au(111) DOS on the O2 atom. The black, red, and solid green projected DOS curves are computed by projecting the total DOS onto atomic orbitals and then summing over all the projections that constitute the subsystems of interest. The magenta line in the top right and bottom left panels is the sum of the projections of the total DOS onto the HOMO and HOMO − 1 molecular states of the gas-phase guanine and thymine.

Even in the absence of O atoms, adenine adheres significantly to the Au(111) surface. The adsorption energy of the horizontal structures is comparable to that of the horizontal $_h$C@Au(111) structures, even slightly higher. Among the sampled vertical configurations, adenine prefers an orientation with the NH_2 group opposite to the surface, so that the N3 atom, which is particularly reactive, can get close to a top site.

We end this section with an interesting analogy to another technologically relevant system, namely DNA bases on a graphene layer.[61,62] Stacked configurations of DNA bases on graphene can occur during translocation of a single-stranded DNA molecule through a nanopore and may affect the conductance of the graphene sheet in a way that can be exploited for fast DNA sequencing.[72] A graphene sheet is a metal, as the Au(111) surface is, which is the source of the analogy; note, however, that the metallic nature has a different physico−chemical nature in the two substrates. The structural analysis presented above basically concludes that DNA bases adsorb on Au(111) in a horizontal orientation, with binding strength in the order G (23 kcal/mol) > C ∼ A (20 kcal/mol) > T (17 kcal/mol), with the values of the interaction energy dominated by van der Waals effects. These features are almost exactly matched in the interfaces between DNA bases and graphene, where the horizontal adsorption orientation is also preferred and the binding energies are also around 20 kcal/mol. The only slight difference between base/Au(111) and base/graphene interfaces is the order of adsorption strength. In fact, in base/graphene interfaces, it was found that the order is G (22 kcal/mol) > A (20 kcal/mol) > T (19 kcal/mol) > C (17 kcal/mol).[61] The honeycomb lattice of graphene systematically prefers adsorption configurations in which the reactive atom in the base is above the center of a hexagon. Instead, we find that the (111) lattice of the gold surface accommodates the O atom of G and C at a top site: while for G the binding energy at different Au(111) sites is the same (Table 1), there is an energy gain of 2 kcal/mol for C from the fcc to the top site. This explains the larger stability of C at the Au(111) surface than at the graphene sheets and accounts for the different stability order. Note, however, that the energy differences are within 1−2 kcal/mol, which is at the limit of resolution of this type of calculations. The analogy traced here suggests a possible general behavior for the binding energy of DNA bases on metal surfaces and opens the way to a variety of possible applications of DNA/crystal interfaces.

3.2. Electronic Structure. We show in Figure 4 the plots of the electronic density of states (DOS) for the most favorable structures $_h$C@Au(111)^t, $_h$G@Au(111)^t, $_h$T@Au(111)$^{t+f}$, and $_h$A@Au(111)^b. The color code is explained in the figure caption. We note that the DOS of thymine and adenine is not significantly affected by the interaction with gold (similarity of the green solid and dashed lines), while in cytosine and guanine, we observe a redistribution of the DOS peaks for the adsorbed molecules in the energy range of the Au *d* bands. This finding is in line with the energetic and structural results. In fact, thymine and adenine exhibit negligible relaxation of the bond lengths and angles upon adsorption on Au(111) relative to the gas phase, and thymine is the DNA base that has the smallest adsorbate−substrate interaction energy. Cytosine and guanine, instead, undergo larger atomic distortions upon adsorption and gain more energy. These effects are not biased by the use of the vdW−DF functional (see Supporting Information, Figure S5).

The projections of the total DOS onto molecular states (magenta lines in Figure 4) indicate that the HOMO lies above the Au *d* bands for all the computed interfaces, even if shown

only for guanine and thymine. The same is true for the HOMO − 1, with the exception of the G@Au(111) interface that we address separately below. Hybridization occurs between the HOMO and HOMO − 1 of cytosine, thymine and adenine and *d* orbitals of gold; however, the shape of such molecular orbitals is not affected by the interaction and no bonding orbitals are formed between the base and gold.

We note that, interestingly, DOS molecular features are present in energy ranges where the Au DOS is also non-negligible. We inspected such energy ranges to search for hybrid molecule−Au orbitals that would reveal some electronic coupling, encouraged by the findings in similar or related systems.[29,40,55] The systematic analysis of all the single-particle electron wave functions reveals the formation of bonding orbitals between the adsorbate and the substrate in the energy range of Au *d* bands. This is true for the four bases, for both horizontal and vertical configurations, despite the fact that the DOS of T@Au(111) and A@Au(111) interfaces does not show appreciable differences relative to the DOS of the respective gas-phase molecule. We find that the formation of bonding orbitals requires the participation of a molecular orbital with a charge component on the most reactive atoms O and N. Bonding orbitals (Figure 5) are accomplished in the energy

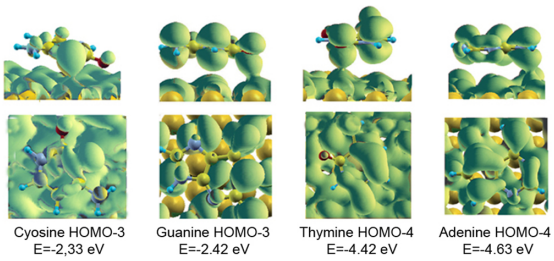


Figure 5. Isosurface plots of representative bonding orbitals formed by cytosine, guanine, thymine, and adenine with gold. These representative examples pertain to the lowest-energy interfaces (gray cells in Table 1).

range between −6.0 and −1.5 eV, namely in the domain of the Au *d* bands (Figure 4). For all the computed interfaces, the HOMO − 3 and HOMO − 4 are mostly responsible for the formation of bonding orbitals with Au. In the case of G@Au(111), also, the HOMO−1 of guanine forms bonding orbitals with *d* orbitals of Au, because of the relative energy of this orbital in the gas phase and the Au bands (Figure 4, top right panel).

The mechanism for electronic hybridization that we find in these systems is different from that characteristic of chemisorption systems.[55] The Newns−Anderson model for atomic and molecular chemisorption on transition metal surfaces predicts that the interaction between a localized atomic or molecular orbital and the narrow *d* band of the metal produces hybrid orbitals of both bonding and antibonding type. In the case of molecular chemisorption, the HOMO of the molecule is responsible for this mechanism.[55] The metal−HOMO bonding orbitals give a peak in the DOS of the molecule/metal interface at the lower−energy edge of the metal *d* bands and the metal−HOMO antibonding orbitals give a peak in the DOS at the upper−energy edge of the metal *d* bands.[73] This mechanism is found, for example, in thiols chemisorbed on Au(111)[56] but also in the case of molecular adsorbates with O and N anchoring groups instead of S.[74] It was accurately discussed in the case of cysteine/Au(111).[55] In the case of DNA bases adsorbed on Au(111) that we treat in this article, we have an alternative situation. We reveal the formation of bonding orbitals, but the overall electronic hybridization mechanism does not comply with the Newns−Anderson picture. We do not detect any bonding−antibonding splitting of the HOMO. Nay, the HOMO is not responsible at all for bonding-like hybridization. The molecular orbitals that form hybrid molecule−metal bonding orbitals are actually those that in the gas-phase molecule have energies coincident with those of Au *d* orbitals; such molecular orbitals do not show bonding−antibonding splitting but only a slight shift with respect to the isolated components.

On the other hand, this adsorption mechanism is also different from that typical of purely aromatic molecules. For pentacene on Cu(100),[75] for instance, it was shown that the slight electronic hybridization between the molecule and the surface occurs via partial filling of the LUMO orbital. This is not accompanied by the formation of any bonding states between occupied molecular orbitals and metal *d* orbitals.

4. SUMMARY

In this paper, we have investigated by means of first-principles electronic structure calculations several interfaces formed by cytosine, guanine, thymine, and adenine with the Au(111) surface, with the objectives of interpreting the adsorbate−substrate interaction mechanism and developing a classical force field for related systems constituted of DNA oligomers on Au(111). We showed that vdW effects are main determinants to attain a correct description of the computed interfaces, both qualitatively and quantitatively. We gave insights into the adsorption mechanism of these heterocycles, which is different from that typical of homocycles[75] on one hand and of non-planar molecules (e.g., thiols)[55] on the other hand. These results are the basis for enabling classical molecular dynamics simulations of DNA on gold, through the generation of a tailored force field that will include adsorbate/substrate interaction and image charge effects.[26,27] In fact, the vdW−DF results on the structure and energetics of G, C, A, and T on Au(111) are being exploited to develop a first-generation AMBER-like force field to describe DNA/gold interfaces. The procedure is conceptually similar to what we recently did for protein/Au(111) interfaces,[26,29] but using data suitable to the adsorption of nucleobases rather than amino acids.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S4 illustrate the most relevant optimized interfaces that appear in Table 1: Figure S1 for C@Au(111), Figure S2 for G@Au(111), Figure S3 for T@Au(111), Figure S4 for A@Au(111). Figure S5 illustrates the total DOS of $_h$G@Au(111)^t computed with the PBE and vdW−DF functionals. Tables S1−S5 report the changes in molecular bond lengths and bond angles when the molecule goes from the gas phase to the adsorbed phase, in selected most favorable configurations: Tables S1 and S2 for C@Au(111), Table S3 for G@Au(111), Table S4 for T@Au(111), Table S5 for A@Au(111). This material is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rosa.difelice@unimore.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by the European Commission through project "DNA-Nanodevices" (Contract No. FP6-029192), by the ESF through the COST Action MP0802, by the Italian Institute of Technology through project MOPROSURF and the Computational Platform, by Fondazione Cassa di Risparmio di Modena through Progetto Internazionalizzazione 2011. The ISCRA staff at CINECA (Bologna, Italy) is acknowledged for computational facilities and technical support. Fengzhu Sun and Remo Rohs are gratefully acknowledged for their crucial support of extended visits in Rohs laboratory at the University of Southern California in Los Angeles during the final stages of this work.

REFERENCES

(1) Miescher, F. *Medicinischchemische Untersuchungen* **1871**, *4*, 441−460.
(2) Dahm, R. *Am. Sci.* **2008**, *96*, 320−327.
(3) Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 737−738.
(4) Porath, D.; Cuniberti, G.; Di Felice, R. *Top. Curr. Chem.* **2004**, *237*, 183−227.
(5) Endres, R. G.; Cox, D. L.; Singh, R. R. P. *Rev. Mod. Phys.* **2004**, *76*, 195−214.
(6) Mallajosyula, S. S.; Pati, S. K. *J. Phys. Chem. Lett.* **2010**, *1*, 1881−1894.
(7) Genereux, J. G.; Barton, J. K. *Chem. Rev.* **2010**, *110*, 1642−1662.
(8) Shapir, E.; Calzolari, A.; Cavazzoni, C.; Ryndyk, D.; Cuniberti, G.; Kotlyar, A. B.; Di Felice, R.; Porath, D. *Nat. Mater.* **2008**, *7*, 68−74.
(9) Shapir, E.; Sagiv, L.; Molotsky, T.; Kotlyar, A. B.; Di Felice, R.; Porath, D. *J. Phys. Chem. C* **2010**, *114*, 22079−22084.
(10) Lindsay, S. M.; Thundat, T.; Nagahara, L. A. *J. Microsc.* **1988**, *152*, 213−220.
(11) de Pablo, P. J.; Moreno-Herrero, F.; Colchero, J.; Gómez Herrero, J.; Herrero, P.; Baró, A. M.; Ordejón, P.; Soler, J. M.; Artacho, E. *Phys. Rev. Lett.* **2000**, *85*, 4992−4995.
(12) Kotlyar, A. B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath, D. *Adv. Mater.* **2005**, *17*, 1901−1905.
(13) Porath, D.; Bezryadin, A.; de Vries, S.; Dekker, C. *Nature* **2000**, *403*, 635−638.
(14) Cohen, H.; Nogues, C.; Naaman, R.; Porath, D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11589−11593.
(15) Boon, E. M.; Barton, J. K. *Curr. Opin. Struct. Biol.* **2002**, *12*, 320−329.
(16) Slinker, J. D.; Muren, N. B.; Gorodetsky, A. A.; Barton, J. K. *J. Am. Chem. Soc.* **2010**, *132*, 2769−2774.
(17) Varsano, D.; Garbesi, A.; Di Felice, R. *J. Phys. Chem. B* **2007**, *111*, 14012−14021.
(18) Migliore, A.; Corni, S.; Varsano, D.; Klein, M. L.; Di Felice, R. *J. Phys. Chem. B* **2009**, *113*, 9402−9415.
(19) Woiczikowski, P.; Kubar, T.; Gutierrez, R.; Caetano, R.; Cuniberti, G.; Elstner, M. *J. Chem. Phys.* **2009**, *130*, 215104/1−14.
(20) Woiczikowski, P.; Kubar, T.; Gutierrez, R.; Cuniberti, G.; Elstner, M. *J. Chem. Phys.* **2010**, *133*, 035103/1−12.
(21) Daggett, V.; Levitt, M. *J. Mol. Biol.* **1993**, *232*, 600−618.
(22) Tirado−Rives, J.; Orozco, M.; Jorgensen, W. L. *Biochem.* **1997**, *36*, 7313−7329.
(23) Toofanny, R. D.; Daggett, V. *WIREs Comput. Mol. Sci.* **2012**, *2*, 405−423.
(24) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987; pp 1−385.
(25) van Gusteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed.* **1990**, *29*, 992−1023.
(26) Iori, F.; Di Felice, R.; Molinari, E.; Corni, S. *J. Comput. Chem.* **2009**, *30*, 1465−1476.
(27) Wright, L. B.; Rodger, P. M.; Corni, S.; Walsh, T. *J. Chem. Theor. Comput.* **2013**, *9*, 1616−1630.
(28) Di Felice, R.; Corni, S. *J. Phys. Chem. Lett.* **2011**, *2*, 1510−1519.
(29) Iori, F.; Corni, S.; Di Felice, R. *J. Phys. Chem. C* **2008**, *112*, 13540−13545.
(30) Wong, K.-Y.; Pettitt, B. M. *Teor. Chim. Acta* **2001**, *106*, 233−235.
(31) Zwolak, M.; Di Ventra, M. *Rev. Mod. Phys.* **2008**, *80*, 141−165.
(32) Hughes, G. A. *Nanomedicine* **2005**, *1*, 22−30.
(33) Piana, S.; Billic, A. *J. Phys. Chem. B* **2006**, *110*, 23467−23471.
(34) Otero, R.; Lukas, M.; Kelly, R. E. A.; Xu, W.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Science* **2008**, *319*, 312−315.
(35) Ortmann, F.; Schmidt, W. G.; Bechstedt, F. *Phys. Rev. Lett.* **2005**, *95*, 186101/1−4.
(36) Sowerby, S. J.; Stockwell, P. A.; Heckl, W. M.; Petersen, G. B. *Origins Life Evol. Biospheres* **2000**, *30*, 81−99.
(37) Seino, K.; Schmidt, W. G.; Preuss, M.; Bechstedt, F. *J. Phys. Chem. B* **2003**, *107*, 5031−5035.
(38) Kilina, S.; Tretiak, S.; Yarotski, D. A.; Zhu, J.-X.; Modine, N.; Taylor, A.; Balatsky, A. V. *J. Phys. Chem. C* **2007**, *111*, 14541−14551.
(39) Mignon, P.; Ugliengo, P.; Sodupe, M. *J. Phys. Chem. C* **2009**, *113*, 13741−13749.
(40) Rosa, M.; Corni, S.; Di Felice, R. *J. Phys. Chem. C* **2012**, *116*, 21366−21373.
(41) Kelly, R. E. A.; Xu, W.; Lukas, M.; Otero, R.; Mura, M.; Lee, Y-J.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Small* **2008**, *4*, 1494−1500.
(42) Lukas, M.; Kelly, R. E. A.; Kantorovich, L. N; Otero, R.; Xu, W.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *J. Chem. Phys.* **2009**, *130*, 024705/1−9.
(43) Xu, W.; Kelly, R. E. A.; Henkjan, G.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Small* **2009**, *5*, 1952−1956.
(44) Xu, W.; Kelly, R. E. A.; Otero, R.; Schöck, M.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Small* **2007**, *3*, 2011−2014.
(45) Kelly, R. E. A.; Lukas, M.; Kantorovich, L. N.; Otero, R.; Xu, W.; Mura, M.; Laegsgaard, E.; Stensgaard, I.; Besenbacher, F. *J. Chem. Phys.* **2008**, *129*, 187707.
(46) Otero, R.; Lukas, W. X. M.; Kelly, R. E. A.; Xu, W.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Science* **2008**, *319*, 312−315.
(47) Kelly, R. E. A.; Lukas, M.; Kantorovich, L. N.; Otero, R.; Xu, W.; Mura, M.; Laegsgaard, E.; Stensgaard, I.; Besenbacher, F. *J. Chem. Phys.* **2008**, *129*, 184707/1−13.
(48) Ostblom, M.; Liedberg, B.; Demers, L. M.; Mirkin, C. A. *J. Phys. Chem. B* **2005**, *109*, 15150−15160.
(49) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. *J. Phys.: Condens. Matter* **2009**, *21*, 395502/1−19.
(50) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(51) Dion, M.; Rydberg, H.; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401/1−4.
(52) Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, R7892−R7895.
(53) Akinaga, Y.; Nakajima, T.; Hirao, K. *J. Chem. Phys.* **2001**, *114*, 8555−8564.
(54) Fernandez, E. M.; Soler, J. M.; Balbas, L. C. *Phys. Rev. B* **2008**, *73*, 235433/1−8.

(55) Di Felice, R.; Selloni, A.; Molinari, E. *J. Phys. Chem. B* **2003**, *107*, 1151−1156.
(56) Vargas, M. C.; Giannozzi, P.; Selloni, A.; Scoles, G. *J. Phys. Chem. B* **2001**, *105*, 9509−9513.
(57) Thonhauser, T.; Cooper, V. R.; Li, S.; Puzder, A.; Hyldgaard, P.; Langreth, D. C. *Phys. Rev. B* **2007**, *76*, 125112/1−11.
(58) Langreth, D. C.; Dion, M.; Rydberg, H.; Schroder, E.; Hyldgaard, P.; Lundqvist, B. I. *Int. J. Quantum Chem.* **2005**, *101*, 599−610.
(59) Tkatchenko, A.; Romaner, L.; Hofmann, O. T.; Zojer, E.; Ambrosch-Draxl, C.; Scheffler, M. *MRS Bull.* **2010**, *35*, 435−442.
(60) Vanin, M.; Mortensen, J. J.; Kelkkanen, A. K.; Garcia-Lastra, J. M.; Thygesen, K. S.; Jacobsen, K. W. *Phys. Rev. B* **2010**, *81*, 081408/1−4.
(61) Cho, Y.; Min, S. K.; Yu, J.; Kim, W. Y.; Tkachenko, A.; Kim, K. S. *J. Chem. Theor. Comput.* **2013**, *9*, 2090−2096.
(62) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2722−2729.
(63) Mura, M.; Gulans, A.; Thonhauser, T.; Kantorovich, L. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4759−4767.
(64) Atodiresei, N.; Caciuc, V.; Lazić, P.; Blügel, S. *Phys. Rev. Lett.* **2009**, *102*, 136809/1−4.
(65) Johnston, K.; Harmandaris, V. *J. Chem. Phys. C* **2011**, *115*, 14707−14717.
(66) Tautz, F. S. *Prog. Surf. Sci.* **2007**, *82*, 479−520.
(67) Henze, S. K. M.; Bauer, O.; Lee, T-L.; Sokolowski, M.; Tautz, F. S. *Surf. Sci.* **2007**, *601*, 1566−1573.
(68) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101(R)/1−4.
(69) Puzder, A.; Dion, M.; Langreth, D. C. *J. Chem. Phys.* **2006**, *124*, 164105/1−8.
(70) Bogdan, D.; Morari, C. *J. Phys. Chem. A* **2013**, *117*, 4669−4678.
(71) Rapino, S.; Zerbetto, F. *Langmuir* **2005**, *21*, 2512−2518.
(72) Min, S. K.; Kim, W. Y.; Cho, Y.; Kim, K. S. *Nature Nanotechnol.* **2011**, *6*, 162−165.
(73) Hammer, B.; Nørskov, J. K. *In Theory of Adorption and Surface Reactions, in Chemisorption and Reactivity of Supported Clusters and Thin Films*; Lambert, R. M., Pacchioni, G., Eds.; Kluwer Academic Publishers: The Netherlands, 1997; pp 285−351.
(74) Gori, P.; Contini, G.; Prosperi, T.; Catone, D.; Turchini, S.; Zema, N.; Palma, A. *J. Phys. Chem. B* **2008**, *112*, 3963−3970.
(75) Ferretti, A.; Calzolari, A.; Di Felice, R.; Ruini, A.; Molinari, E.; Baldacchini, C.; Betti, M. G. *Phys. Rev. Lett.* **2007**, *99*, 046802/1−4.

Chapter 6

Release of the new GolDNA-AMBER force field

The DFT study on the interaction of nucleobases with gold allowed to start the parametrization of a suitable FF for the adsorption of DNA on Au(111). Our main concern in this parametrization is to reproduce correctly DFT calculations maintaining the FF as general as possible, as it should be transferable to the study of more complex systems. For validation, we followed a sequence of three steps.

- we tuned the classical LJ parameters for the DNA at gold interface on the training set of structures, namely the systems consisting of isolated bases adsorbed on Au(111); the LJ parameters were adjusted until FF structural relaxation gave the same geometries and energetics as DFT on the training set;
- we proved the portability of our new FF, by showing that FF structural relaxation gives results in agreement with DFT and experimental data on systems that do not belong to the training set; specifically, for this task we considered periodic self-assembled monolayers of guanine and adenine on Au(111)[87, 89];
- we applied our new FF to the simulation of base adsorption in water solution, revealing an interesting enthalphy-entropy balance that resembles the adsorption mechanism of peptide groups on gold[35], thus giving an additional proof of portability non only to

different materials systems but also different environments.

The good results obtained on in vacuo and in solution systems validate the approach adopted for the FF parametrization. The FF itself is able to reproduce experimental results on complex systems and it proved to be suitable for the study of an entire DNA oligomer on Au(111) in vacuo and in solution, to investigate the possible deleterious effect of a hard inorganic substrate on the helical folding.

This work has not yet been published, but a manuscript is almost final for submission for publication in an international journal and is attached as an integral part of this PhD dissertation. It is articulated as follows: (1) the Introduction puts our activities in context with open questions and current knowledge on the effect of hard inorganic surfaces on nucleic acids and its relevance; (2) the Method section describes the classical model that stems from the GolP FF[34] and its relation to other FFs for molecule-surface interactions, and contains the computational details of quantum and classical simulations; (3) the Results pertain to self-assembled base monolayers on gold in vacuo and to isolated bases on gold in water; (4) the Supporting Information contains the explanation of the procedure for tuning the LJ parameters at the DNA at Au interface.

The FF parametrization was performed reproducing with MD simulations the adsorption energies and geometries of DFT calculations, with particular attention to the energetic order of the different configurations. Atoms that were found to interact strongly with the surface in DFT calculations have been specifically parametrized for the interaction with gold atoms.
Results on adenine and guanine monolayer adsorption on Au(111) are in fair agreement between ab-initio, MD calculations and experiments, and allow both to confirm the roboustness of the FF chosen and to evaluate the effects of varying coverage in the evaluation of energy in desorption experiments. In addition, we investigated also the electronic structure of guanine monolayer, comparing the results obtained with the ones of the single bases adsorption: this is an important added value of quantum calculations that if too often neglected in surface studies. This is one of the rare cases in which the ab initio electronic structure of a fully supported monolayer is presented: in this particular case, we find the same adsorption mechanism that pertains to isolated DNA bases, namely an interaction regime that is intermediate between purely long-range and purely short-range. While early prejudices considered the adsorption of

molecular homo- and hetero-cycles to be pure physisorption, recent evidences accumulate that there is an electronic partnership between adsorbate and substrate[95, 96]. Our findings add on to such evidences and reveal a modulation of the electronic coupling depending on the chemical nature of the adsorbate.

Finally, we carried on a study on the adsorption of single bases on gold in aqueous conditions. In this case a comparison between experimental and computational results could not be done because of the lacking of experimental data, but the similarities with the results obtained with the same method on peptides adsorbed on Au(111)[35] validate the approach and the results obtained. The results on this system are particularly interesting, as the adsorption potential energy value obtained for the four bases is positive, meaning that the adsorption of single bases on gold is enthalpically unfavored. Nevertheless, longer simulations (100 ns, 300K) demonstrate that desorption from the surface is not favored. Free energy calculation for cytosine adsorbed on Au(111) allowed to confirm that the entropic contribution is determinant in understanding the behavior of bases in solution: it was found that the molecule has to overcome a free energy barrier to adsorb on the surface, due both to the strong interaction of water molecules with gold and nucleobase surfaces and to the screening of cytosine-gold interaction performed by the solvent. Free energy calculations are able to describe the particular adsorption process of DNA bases in water, revealing the presence of a free energy barrier close to the surface due to the competition between water and DNA molecules for the interaction with gold. GolDNA offers a computationally accessible way to study the adsorption of DNA on Au(111).

Enthalpy-Entropy Tuning in the Adsorption of Nucleobases at the Au(111) Surface

Marta Rosa,[†,‡] Stefano Corni,[†] and Rosa Di Felice[*,†,¶]

Center S3, CNR Institute of Nanoscience, Via Campi 213/A, 41125 Modena, Italy

E-mail: rosa.difelice@unimore.it

Abstract

The interaction of DNA molecules with hard substrates is of paramount importance both for the study of DNA itself and for the variety of possible technological applications. Interaction with inorganic surfaces strongly modifies the helical shape of DNA. Hence, an accurate understanding of DNA structure and function at interfaces is a fundamental question with enormous impact in science and society. This work sets the fundamentals for the simulation of entire DNA oligomers on gold surfaces in dry and wet conditions.

Thanks to the new GolDNA-AMBER force field, which was derived from first principles and includes dispersion interactions and polarization effects, we simulated self-assembled guanine and adenine monolayers on Au(111) in *vacuo* and the adsorption of all nucleobases on the same substrate in aqueous conditions. The periodic monolayers obtained from classical simulations match very well those from first principle calculations and experiments, assessing the robustness of the force field and motivating the application to more complex systems for which quantum calculations are not affordable and experiments are elusive. The energetics of nucleobases on Au(111) in solution reveal fundamental physico-chemical effects: we find that

[*]To whom correspondence should be addressed
[†]Center S3, CNR Institute of Nanoscience, Via Campi 213/A, 41125 Modena, Italy
[‡]Department of Physics, University of Modena and Reggio Emilia, Modena, Italy
[¶]Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA

the adsorption paradigm shifts from purely enthalpic to dominantly entropic by changing the environment and aggregation phase.

Introduction

Research on DNA has nowadays trespassed the boundaries of biological and medical laboratories and is becoming increasingly interdisciplinary, covering chemical sciences, physical sciences and engineering/nanotechnologies. In this process, the interaction of nucleic acids with hard inorganic surfaces has assumed a primary importance.

Even within the context of life sciences, one can think of applications that require the formation of interfaces between DNA/RNA and inorganic materials: for instance, array-based sensors for DNA sequencing consist of single-stranded oligonucleotides of different sequences attached to a surface through anchoring functional groups.[1] The principle of operation is the complementarity of such single-stranded monolayers with the target sequences. In this kind of devices, the double helix does not contact the host substrate directly but through a linker chain that is responsible for the attachment. A similar configuration, with DNA molecules "standing" on the surface or between electrodes and attached through thiol linkers, has been successfully adopted to measure charge currents through DNA.[2–6]

An alternative configuration relevant in nanotechnologies embodies DNA molecules "lying" on the surface so that the backbone or bases directly contact the substrate. This is relevant for a novel vision of DNA sequencing based on physical measurements, namely translocation through nanopores.[7,8] It is mainly related to measurements of the morphology and of the electronic structure through scanning microscopes, as well as to electrical transport measurements of molecules deposited onto substrates.[9–11] The electrical characterization of DNA molecules is instrumental to the development of DNA-based molecular electronics and is still controversial.[12–15]

Modifications of the helix motif when a DNA molecule comes into contact with a hard substrate

can be particularly severe. As an example of such a dramatic deviation, we refer to the results of atomic force microscope (AFM) experiments on double-stranded DNA molecules "lying" on a hard substrate: in this condition, the measured height of a DNA molecule is about 50% of the diameter of the double helix in solution.[11,16] This finding suggests the possible occurrence of partial unfolding, which would have strong consequences on the electrical performance. In fact, it is becoming steadily clear that the ground-state and excited-state electronic structure and the charge transfer rates in nucleobase stacks depend in a very sensitive way on the details of the atomic spatial arrangement.[17–22]

This brief overview introduces the concept that the knowledge and control of DNA structure-function properties and their dependence on the environment (e.g., substrate *versus* solution) are fundamental for understanding some intrinsic features of nucleic acids that impact applications in different research fields. Experimental structural resolution of DNA molecules at interfaces is unfeasible, because conventional techniques for structure determination are specific for biological molecules or inorganic crystals. Molecular simulations thus become a powerful tool for bypassing the lack of atomic details. Classical molecular dynamics (MD) atomistic simulations are in principle the method of choice. They have successfully been employed to describe the unfolding of biological molecules in solution.[23–26] However, the development of force fields (FFs) suitable for the description of the adsorption of (bio)molecules at surfaces is still immature and studies of DNA/inorganic interfaces are consequently hindered. Lee and Schatz were able to perform simulations of DNA oligomers attached to metal nanoparticles,[27] tackling a "standing" configuration with a linker that mediates the DNA-surface connection: the presence of a linker substantially neutralizes the problem of defining a FF that encompasses the DNA-gold interaction. In order to enable simulations of "lying" configurations of DNA oligomers on a substrate, FF parameters that account for this interaction cannot be ignored. In the following, we present insights from simulations of nucleobases on a gold surface in different aggregation states and different environments, with an effective FF that has been originally developed for studying peptide adsorption on Au(111) and now tailored to study DNA adsorption on Au(111). The positive benchmark of classical sim-

ulations against quantum calculations and experimental studies is a convincing assessment of this methodology and launches the ability to predict folding motifs of double-stranded DNA oligomers on the gold surface.

The target systems of our investigation are adenine and guanine full periodic monolayers on Au(111) in *vacuo* and single bases on Au(111) in water solution. The results on the monolayers comply with experimental knowledge, proving that the classical model is robust beyond the specifically parametrized systems for which it was developed and can therefore be confidently used to simulate a variety of DNA structures on the same surface. Furthermore, they allow us to understand the limits of the Redhead approximation,[28] which is widely adopted for extracting values of desorption energy from the desorption rates measured in temperature programmed desorption (TPD) experiments. Encouraged by this success for the portability of the force field, we sampled a more complex environment condition that includes bulk water. We find that the environment has dramatic effects on the adsorption thermodynamics: indeed, while the monolayer formation on the substrate in *vacuo* is driven by enthalpy, the adsorption of individual bases (e.g., a dilute coverage regime) in solution is driven by entropy. These findings stimulate a wider exploration of aggregation states of DNA on Au(111) in diverse ambient conditions. The entropic principle may have dramatic effects on the deposition of DNA oligomers on gold from solution drops and its understanding will eventually open the way to control sample preparation conditions in order to minimize the disruption of the helical folding motif.

Method

The classical model

Survey of bio-inorganic force field development

To the best of our knowledge, four different approaches are amenable for evolvement in order to enable simulations of DNA oligomers on gold. CHARMM-METAL[29] was parametrized to reproduce bulk metal density and surface tension; it is compatible with CVFF[30] and a number of other biomolecular FFs. GolP[31–33] was specifically developed for describing the interaction of proteins with the Au(111) surface with inclusion of image charge effects; it is compatible with OPLS[34] and CHARMM.[35] Zerbetto and coworkers[36,37] developed a FF for the adsorption on the Au(111) surface of a variety of molecules, including DNA bases. Piana and Bilic[38] developed special Lennard-Jones (LJ) parameters for the adsorption of DNA on Au(111), based on results of post-Hartree-Fock Møller-Plesset perturbation theory (MP2) calculations.

The CHARMM-METAL FF has been used to study the adsorption of peptides on Au(111) surfaces[39–43] and can in principle be used also for DNA on Au(111). Yet, we believe that it is not ideal for the description of DNA adsorption on a metal surface because of two negative built-in features. First, it neglects the dynamic polarization of the metal atoms in the Au surface. This shortcoming was presented as a source of uncertainty for the adsorption of peptides on surfaces, although its contribution was supposed to be small.[43] In the case of DNA, the backbone carries a $1e^-$ charge for each nucleotide and therefore the polarizability of the substrate is not a negligible effect. Second, CHARMM-METAL was not specifically parametrized to capture individual atom-gold interactions and consequently it does not include terms to account for the weak chemisorption of some specific atomic species. This problem was identified in the study of peptide adsorption[33] and becomes dominant in the case of DNA adsorption. In fact, density functional theory (DFT) calculations with the inclusion of van der Waals (vdW) interactions pointed out that guanine, cy-

tosine, adenine and thymine are weakly chemisorbed on Au(111) via O and N atoms.[44,45]

The FF developed by Zerbetto and coworkers[36] is customized for the adsorption of DNA bases on Au(111). Metal polarization effects are taken into account through a charge-equilibration model that modifies the charge on the atoms during the simulation.[46] Moreover, long-range and weak chemisorption effects are taken into account through the fitting of interaction parameters with experimental results on nucleobase monolayers in *vacuo*. The main disadvantage of this approach is the computational burden: as a matter of fact, such a detailed FF is affordable only for very short simulation times for a system as complex as a DNA oligomer on the surface and it is prohibitive to investigate a water solution environment.

Piana and Bilic were able to develop a FF that includes the vdW interaction between the bases and gold in an indirect way.[38] Namely, they obtained FF parameters for individual atoms from MP2 calculations of small molecules adsorbed on a planar four-atom gold cluster and then transfered them to nucleobase atoms. This approach, which was restricted by the computational cost of MP2 calculations, has the limitation of not including the specificity of the bases for what concerns adsorption on gold surfaces. Moreover, the model for the surface does not take into account polarization effects. Polarization effects can be negligible for systems like self-assembled base monolayers, which do not include the charged backbone, but are certainly relevant for the adsorption of realistic oligomers, which is our general target.

Finally, GolP accounts for polarization of the metal atoms and can be parametrized for all the specific interactions found in DNA bases, as it was previously done for peptides. GolP exists in two versions: the original GolP is compatible with OPLS[32] and the later GolP-CHARMM is compatible with CHARMM.[33] For this work we decided to use a GolP-like FF that is designed for nucleic acids rather than proteins and is compatible with AMBER, which is acknowledged as the most reliable classical description of DNA in solution:[47–52] we will call the new FF GolDNA-AMBER. Some authors have addressed the problem if FF parameters derived for biomolecules in solution are directly transferable to biointerfacial simulations.[55,56] This question is still unanswered. Yet,

at present, a FF for gold-DNA interaction that is consistent with an existing bio-organic FF, such as AMBER, is the only viable way to predict adsorption equilibria from computational simulations.

GolDNA-AMBER

In GolDNA-AMBER the potential energy function for the interaction between DNA bases and the Au(111) surface is written as[32]

$$V_{Au-DNA}^{tot} = V_{Au-DNA}^{im} + V_{Au-DNA}^{vdW} + V_{Au-DNA}^{bonding}. \quad (1)$$

V_{Au-DNA}^{vdW} and $V_{Au-DNA}^{bonding}$ are the van der Waals and bonding interaction terms between atoms in the DNA base and gold atoms. V_{Au-DNA}^{im} is the Coulombic electrostatic interaction between the molecule and the image charges induced in gold.

The surface polarization effect is described through a dipole located in the position of each gold atom. One such dipole is constituted of two opposite charges, connected by a rigid rod free to rotate around one end.[31] In the absence of a molecule on the surface the average dipole moment is equal to zero, with the initial orientation of the dipoles chosen randomly. When a molecule lies above the surface, the partial atomic charge on its atoms forces the rods to preferentially rotate in a specific orientation, modifying the average dipole moment. This model is computationally efficient[32,57,58] and able to reproduce continuous electrostatics results.[31] It was demonstrated for GolP[31,32] and GolP-CHARMM[33] that including polarization effects facilitates the transferability of parameters derived in the gas phase to the condensed phase (e.g., this was demonstrated for the AMOEBA FF[59] that is specific for biomolecules). This means that parameters fitted to in-*vacuo* reference data are suitable for the study of molecules adsorbed on gold in solution. GolDNA-AMBER gold atoms are held frozen during the simulation, therefore it is not necessary to define an interaction between them. This means that this FF is not useful for describing the gold surface itself, but is focused on the adsorption of molecules on it. The Au(111) $22 \times \sqrt{3}$ surface reconstruction,

in particular, has not been taken into account, as only marginally small differences exist between the reconstructed and ideal surfaces,[60] which are negligible for the purposes of this work. Moreover, an optimization of interaction parameters of water molecules allows GolDNA-AMBER to describe DNA adsorption not only in *vacuo* but also in aqueous conditions.[33] The parametrization of GolDNA-AMBER relies on a multi-step approach based on DFT results. Details are reported in the Supporting Information.

Bonding and vdW interactions are together described through a sum of 12-6 Lennard-Jones interactions (Equation 2) between each DNA atom and interaction sites on the gold surface:

$$V^{vdW}(r_{ij}) = 4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}]. \quad (2)$$

In GolDNA-AMBER, as in GolP and GolP-CHARMM, the topmost layer of the surface slab includes real gold atoms and virtual interaction sites. Virtual sites are placed in hollow positions of the real lattice and are the ones interacting with adsorbing species via the LJ terms, while the LJ parameters of real gold atoms are set to zero. Dipolar rods are located only on gold atoms. This formulation allowed us to achieve the correct description of the adsorption of small molecules at top sites rather than at hollow sites.[32,33,61] This is important for this study, because also DNA bases prefer top adsorption sites,[45] although they are characterized by a pronounced lateral mobility.[36] Our results reported later satisfy both these conditions, which we have posed as a requirement for the positive evaluation of our classical model.

If no particular interaction is defined, LJ parameters for gold atoms (σ_{Au}, ε_{Au}) are combined with LJ parameters for DNA atoms, in accordance with AMBER Lorentz-Berthelot mixing rules – σ_{ij} = $(\sigma_i + \sigma_j)/2$, $\varepsilon_{ij} = (\varepsilon_i * \varepsilon_j)^{1/2}$. Ad-hoc LJ parameters can be defined for the interaction between some atoms and the Au surface. Such adjusted LJ interactions account for dispersion, repulsion and weak chemisorption effects. Details of our choice of GolDNA-AMBER parameters, along with the tuning procedure, are given in the Supporting Information.

Computational approach

DFT calculations

We performed plane-wave pseudopotential DFT calculations with the quantum-espresso package version 5.0,[62] using the vdW-DF functional.[63] The two adenine and guanine self-assembled monolayers, which we label A_{1ML}@Au(111) and G_{1ML}@Au(111), respectively, were simulated in the repeated supercell approach: the size of the supercells was chosen according to periodicities revealed in experimental data but was optimized by DFT. Each supercell contained a single nucleobase layer at full coverage (four guanines, twelve adenines) and a gold slab with four layers. We adopted a kinetic energy cutoff of 25 Ry for the plane-wave expansion of the wave functions (200 Ry for the charge density)[61] and we described the electron-ion interaction with ultrasoft pseudopotentials.[64] Brillouin zone sums were calculated including 2 Monkhorst-Pack special **k** points in the irreducible wedge. In calculations for structural optimization all the atomic coordinates were relaxed until each force component vanished within 0.05 eV/Å.

The lateral size of the supercells (parameters $d1$, $d2$ and γ, Figure 1) for guanine and adenine self-assembled monolayers on Au(111) was fixed to respect the periodicity of the overlayers observed in experimental studies (Figure 1), while being a multiple of the triangular lattice of the gold substrate. The triangular lattice constant of Au(111) was set to 3.0 Å,[44] which is the equilibrium value computed by DFT with the vdW-DF functional, consistently with the monolayer calculations. This value expresses an overestimation by about 4% of the experimental value. The size of the base superstructures was optimized starting from the observed periodicities[71,73], after checking that the relaxed gas-phase adenine dimer matches the measured adenine monolayer.[71] Structural optimization of the adenine dimer was carried out in a $6\times4\sqrt{3}$ lateral supercell (in units of the substrate triangular lattice vectors), resulting in a lateral distance between periodic replicas at least of 10 Å. The perpendicular distance between neighboring replicas was always larger than this quantity. The relaxed cells of the free-standing adenine and guanine gas-phase monolayers were computed as $d1 \times d2 = 15.8 \times 12.6$ Å^2 and 15.8×15.4 Å^2, respectively, both with $\gamma = 90°$.

These are in agreement with the experimental determinations of $(15.6 \pm 1.0) \times (12.5 \pm 1.0)$ Å^2 with γ of $88 \pm 5^\circ$ for A_{1ML}@Au(111) and $(15.0 \pm 1.0) \times (15.0 \pm 1.0)$ Å^2 with γ of $90 \pm 5^\circ$ for G_{1ML}@Au(111). These optimized cell sizes did not allow us to find two supercells that are suitable for DFT calculations, namely that match multiples of the vdW-DF Au lattice parameter (3.0 Å). For this reason we slightly adjusted the monolayer cells maintaining $d1$, $d2$ and γ in the experimental error range. The lateral supercells for the DFT calculations of the adsorbed monolayer were then fixed as follows: γ was 90° for both the cells; the A_{1ML}@Au(111) cell was chosen as $d1 \times d2 = 15.6 \times 13.0$ Å^2, that repeated three times matches a $3\sqrt{3} \times 13$ ($15.6\times$ 39.0 Å^2) supercell of the Au(111) lattice; the G_{1ML}@Au(111) cell was chosen as 15.0×15.6 Å^2, that matches a $5 \times 3\sqrt{3}$ supercell of the Au(111) lattice. The differences in energies for gas-phase monolayers in optimized and adjusted cells are below 1 kcal/mol. More details on the choice of the supercells are added in the Supporting Information (Table S6).

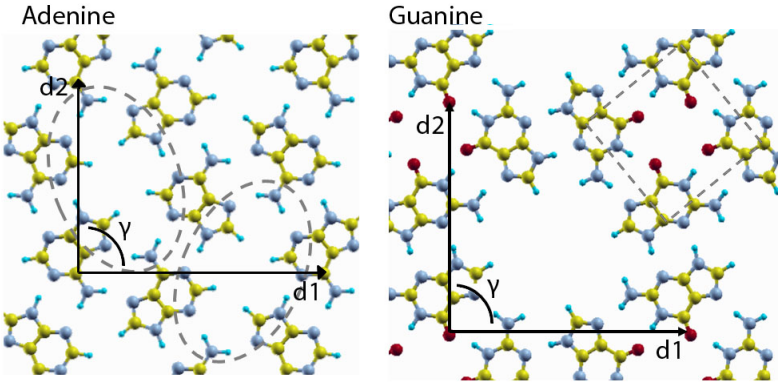


Figure 1: The periodicity of adenine[71] and guanine[73] monolayers as reported from scanning tunneling microscope measurements. $d1$, $d2$ and γ are the unitary cell parameters. The gray dashed line in each panel marks the repeating unit of the monolayer.

The quantity that we benchmark against experimental values of the desorption energy is the computed monolayer formation energy of the interface relative to the isolated components, normalized to one adsorbate molecule, which is

$$E_{form}[X_{1ML}@Au(111)] = \frac{E_{tot}[X_{1ML}@Au(111)] - E_{tot}[Au(111)] - N \times E_{tot}[X]}{N}, \qquad (3)$$

where: $X = A, G$; N is the number of bases in the unit supercell for the monolayer calculation; $E_{tot}[X_{1ML}@Au(111)$ is the calculated total energy of the self-assembled monolayer on the substrate, $E_{tot}[Au(111)]$ and $E_{tot}[X]$ is the total energy of the clean substrate and gas-phase nucleobase X computed with exactly the same method as for the adsorbed monolayer (including the same supercell size). A negative value of E_{form} is representative of an exothermic reaction. Desorption from a stable phase requires an energy cost and therefore the measured desorption energy (E_{des}) is a positive value.

Classical MD calculations

All the FF-based simulations were performed with the Gromacs 4 MD simulation package[65] in the NVT ensemble using a Nosé-Hoover thermostat. The Particle Mesh Ewald (PME)[66] electrostatic summation was used with a real-space cutoff at 11 Å. A force-switched cutoff starting at 9 Å and ending at 10 Å was used for Lennard-Jones nonbonded interactions. The amber99sb[52] version of the AMBER FF was applied to the nucleic acids and the GolDNA-AMBER FF was applied to the substrate coupled to the bases. Each MD simulation was preceded by a standard minimization-equilibration protocol (details in the Supporting Information).

Monolayers in *vacuo*. The lateral size of the supercells adopted in classical simulations was $9\sqrt{3} \times 13$ for A_{1ML}@Au(111) and $10 \times 9\sqrt{3}$ for G_{1ML}@Au(111), multiple of the DFT supercells. The substrate was simulated by a gold slab with five layers at the vdW-DF equilibrium lattice constant. If N is the number of bases in the DFT calculation of the X_{1ML}@Au(111), the number of bases in the FF calculation is M=3N for X=A and M=6N for X=G.

The initial configurations for A_{1ML}@Au(111) and G_{1ML}@Au(111) were chosen with the molecular adlayers close to the surface, in the equilibrium structures calculated by DFT. The two geometries were subjected to a 20-ps annealing procedure to determine the FF-optimized geometries. A L-BFGS[53,53] minimization was performed for a system with just one guanine or adenine far from the surface (to simulate the clean surface and isolated molecule). The bonding and LJ interaction

energies were then calculated as:

$$E_{form}^{LJ+bonding} = \frac{E^{LJ+bonding}[X_{1ML}@Au(111)]}{M} - E^{LJ+bonding}[X + Au(111)] \quad (4)$$

where $E^{LJ+bonding}[X_{1ML}@Au(111)]$ is the LJ and bonding energy of the adsorbed monolayer and $E^{LJ+bonding}[X + Au(111)]$ is the LJ and bonding energy of the system composed of the surface plus one desorbed guanine or adenine. In the latter term, bonding interactions occur within the adsorbate layer.

To evaluate the Coulomb contribution to the formation energy of the adsorbed monolayer, the procedure is more subtle because of the rod model:[32] it also requires short finite-temperature MD runs, not only structural minimization. For each of our systems A_{1ML}@Au(111) and G_{1ML}@Au(111) we carried out three 50-ps MD simulations at constant temperature of 300 K, in which the only mobile species was that of the polarization rods: (1) one run with the correct partial charges on all atoms of both the adsorbate and the substrate, with Coulomb energy $E^{Coul}[X_{1ML}@Au(111)]$; (2) one run with the correct partial charges on the adsorbate atoms and the substrate charges set to zero, with Coulomb energy $E^{Coul}[X_{1ML}@Au(111), q_{sub} = 0]$; (3) one run with the correct partial charges on the substrate atoms and the adsorbate charges set to zero, with Coulomb energy $E^{Coul}[X_{1ML}@Au(111), q_{ads} = 0]$. The electrostatic contribution of the free-standing monolayer to the entire formation energy was determined by subtracting the Coulomb energy of the single base (evaluated from the calculation with the molecule far from the surface, mentioned above) from the Coulomb energy of the monolayer calculation with surface charges set to zero. The Coulomb contribution to the formation energy of the adsorbed monolayer was then computed as

$$E_{form}^{Coul} = \frac{1}{M}\left(\frac{E^{Coul}[X_{1ML}@Au(111)] - E^{Coul}[X_{1ML}@Au(111), q_{sub} = 0] - E^{Coul}[X_{1ML}@Au(111), q_{ads} = 0]}{2} + E^{Coul}[X_{1ML}@Au(111), q_{sub} = 0] - M \times E^{Coul}[X + Au(111)]\right) \quad (5)$$

In Equation 5 the factor 1/2 comes from the derivation of the electrostatic free energy of formation as half of the Coulomb energy,[31] which gives a good approximation of the electrostatic polarization energy. Eventually, the FF formation energy that should be compared to the DFT value and to the experimental desorption energy (Table 1) results from the sum of $E_{form}^{LJ+bonding}$ (Equation 4) and E_{form}^{Coul} (Equation 5).

For the adsorbed guanine monolayer a 10-ns MD simulation was performed at a temperature of 400 K to check the stability of the minimized configuration.

We also calculated the formation energy of an incomplete monolayer, $G_{\frac{1}{6}ML}$@Au(111), in vacuo. This was obtained with four guanines in the same supercell that was used for the G_{1ML}@Au(111), with the guanines linked by the same H-bond pattern that characterizes the unit motif of the complete monolayer (Figure 1).

Single bases in water. The substrate was modeled with a $8\times 5\sqrt{3}$ slab containing five Au layers (80 atoms/layer). The lateral size was 24.0×26.0 Å^2. The perpendicular size was about 40 Å: it was adjusted for each base when filling the simulation box with water molecules, in order to maintain the correct density of bulk water in the centre of the box[67].

We performed two 10-ns simulations for each DNA base at the temperature of 300 K: one MD simulation started with the nucleobase adsorbed on the surface in a parallel configuration and the other started with the nucleobase located in the centre of the box immersed in bulk water and constrained at that distance to prevent adsorption. The cumulative bonding and LJ contibution to the formation energy were calculated by subtracting the bonding and LJ component of the solvated molecule from that of the adsorbed molecule (Equation 6):

$$E_{form}^{LJ+bonding} = E_{int}^{LJ+bonding}[X@Au(111)] - E_{sol}^{LJ+bonding}[X + Au(111)], \quad (6)$$

where $E_{int}^{LJ+bonding}$ and $E_{sol}^{LJ+bonding}$ are the energies of the X@Au(111)$_{H_2O}$ interface and of the system with the solvated molecule, respectively. The Coulomb formation energy was calculated

as for monolayers, according to Equation 7. The electrostatic component for the adsorbed and solvated geometries was evaluated running the simulation again on the obtained trajectories, first with the charges of the nucleobases and water molecules equal to zero, then with the charges of gold atoms equal to zero. Then, similarly to Equation 6, the energy of the constrained system was subtracted from that of the adsorbed system.

$$E_{form}^{Coul} = \frac{1}{2}\Bigg(E_{surf}^{Coul}[X@Au(111)] - E_{surf}^{Coul}[X@Au(111), q_{sub}=0] - E_{surf}^{Coul}[X@Au(111), q_{base+wat}=0] - E_{sol}^{Coul}[X@Au(111)] - E_{sol}^{Coul}[X@Au(111), q_{sub}=0] - E_{sol}^{Coul}[X@Au(111), q_{X+wat}=0] \Bigg) \tag{7}$$

To describe the adsorption mechanism of nucleobases on gold in a water solution, we computed the free energy of interface formation, adopting a strategy that gave good results in applications to aminoacids.[68] Note that the vast majority of computational studies of biomolecular systems rely on total energy analysis, while the determination of the free energy requires specific skills and methods and is not routinely done in molecular simulations. The scheme rests on calculating the potential of mean force from restrained simulations and allowed us to compute free energies of interacting particles in an accurate and computationally affordable way.[69] The different states of interest were identified through a reaction coordinate, which in our case was defined as the distance between the center of mass (COM) of the nucleobase and the gold first layer. The molecule was restrained in the direction perpendicular to the surface at different COM-surface distances and each value of the reaction coordinate was sampled by MD for 7.5 ns at 300 K. The force needed to keep the molecule at the fixed distance from gold was averaged and integration of this force versus distance gave us the free energy of association. The integration of the force plot was performed using the *Grace* software[70] version 5.1.22. The space accessible to the base on the surface was sampled in the range 2.8÷13.8 Å of COM-surface distance. The initial span of 4 Å was divided

into steps of 0.25 Å; the following span of 2 Å was divided into steps of 0.5 Å; the final span of 5 Å was divided in steps of 1 Å. To monitor the accuracy, we also sampled smaller distances and steps, covering the range 2.70÷3.05 with tiny intervals of 0.05 Å.

Results and discussion

Self-assembled guanine and adenine monolayers on Au(111)

DNA bases self-assemble in periodic overlayers on the Au(111) surface, with a two-dimensional periodicity for guanine and adenine. The structure of such bio-inorganic interfaces has been intensively studied by scanning tunneling microscope (STM) measurements in combination with DFT calculations.[71–77] These investigations convey a description of the structure of guanine and adenine monolayers on Au(111), which was a precious pre-condition for our investigation, avoiding the need for sampling many different supercell sizes. Data on desorption energies for base monolayers on Au(111) also exist, from TPD infrared reflection absorption spectroscopy (TPD-IRAS) and TPD mass spectroscopy (TPD-MS) experiments.[78] By calculating formation energies for the computed monolayer structures that map the STM images, we are in place to benchmark QM and MM results against experimental data of desorption energies, for validation of the GolDNA-AMBER FF and its use for prediction of the behavior of nucleobases on the surface immersed in a water solution. We find, indeed, that the force field is robust and transferable to systems different from those used for the parametrization. We discuss the results of DFT calculations, the performance of FF simulations and finally the approximations that are implicit in comparing our formation energies to experimental desorption energies, which are not measured directly but are indirectly estimated from measured quantities.

Besenbacher and coworkers[71,73] predicted the atomic structures of guanine and adenine self-assembled monolayers on Au(111) by combining STM images and DFT results not accounting for vdW terms in the functional. They generated several regular monolayer structures, charac-

terized by different possible patterns of hydrogen bonds. They carried out DFT calculations of a number of unsupported monolayers that best fit the STM images, optimizing the lateral size (Figure 1). They eventually ranked the computed structures in energy and found a fair agreement of the size of the lowest-energy structures relative to the experimental measurements. All the DFT calculations were done on gas-phase structures, not adsorbed on the surface, under the assumption that the base-gold interaction is negligible.

We actually showed that the base-gold coupling is not negligible and goes beyond pure dispersion interactions; in fact, we even found electronic hybridization.[44,45] It is therefore important to correctly describe the presence of the substrate in order to attain a quantitative description of the adsorption energies and structures. We discussed in our recent work the importance of using a functional that is able to estimate van der Waals effects in DFT calculations of single nucleobases on Au(111) in *vacuo* and we start our analysis by monitoring the importance of these effects in the self-assembled monolayers, by assessing vdW-DF results against PBE results.

We evaluated the length of covalent bonds and of H-bonds for adenine gas-phase dimers (A_2) with the vdW-DF and PBE functionals and for the same dimers adsorbed on Au with the vdW-DF functional. Results show that the bonding lengths are equivalent in the three cases. This evidence is an indication that van der Waals effects and the presence of the surface are not strong factors in determining the inherent geometry of the monolayers. The negligible effect of the surface is in line with the large lateral mobility of single DNA bases adsorbed horizontally on Au(111) emerging from our DFT calculations (quasi-degeneracy of various adsorption sites within 1.5 kcal/mol).

We then checked the possible role of the vdW treatment in the energetic ranking of different unsupported adenine monolayers that were found compatible with STM data.[71] Specifically, we evaluated with the vdW-DF functional the relative energy for the three most favorable geometries identified in agreement with morphological measurements, following the procedure previously suggested.[71] In these vdW-DF calculations, the structure was not relaxed and the supercell size was fixed according to the published work, though this is not fully consistent with our choice of den-

sity functional. We found the same global minimum as that identified without the vdW treatment (Table S5 in Supporting Information). We further checked the accuracy of the lateral size of the periodic monolayer unit. To this aim, we performed a series of self-consistent calculations for the unsupported monolayers at variable supercell size, by changing the parameters $d1$, $d2$ and γ. The computed equilibrium supercell parameters are still in agreement with STM measurements and the energetic order is still in agreement with the published DFT results (see Supporting Information). These tests, performed by us at a higher level of theory, validate the structure and energetics obtained with the simpler approach: consequently, we concluded that the inherent geometries of the guanine and adenine monolayers proposed by Besenbacher and coworkers are faithful for use in the following calculations for the surface-supported monolayers. The same procedure was followed for calculating the equilibrium values of $d1$, $d2$ and γ for the guanine unsupported monolayer (Table S6 in Supporting Information). The adjustment to comply with the substrate lattice registry and size, as well as the supercells for FF calculations, are described in the section Method.

The formation energies of the adsorbed monolayers obtained from our vdW-DF and FF calculations are reported in 1. This is a pristine result of this work: The DFT values could not be evaluated earlier by considering only the unsupported monolayers; the classical calculations were missing a proper FF for the monolayer-substrate interactions. Table 1 also compiles the formation energies of interfaces between single bases and Au(111) in water, which are elaborated in the next section.

Before commenting on the accuracy of FF simulations, we include the analysis of the electronic structure from DFT, which is another original ingredient of this article. Electronic hybridization and bonding orbitals were detected at the interfaces between individual bases and Au(111) in *vacuo*: it is a fundamental question to unravel whether this property extends to the larger scale of periodic monolayers.

In Figure 2 we illustrate the electronic density of states (DOS) of the G_{1ML}@Au(111) system, compared to that of the ${}_hG^t$@Au(111) interface (namely, a single guanine adsorbed on Au(111)

Table 1: Calculated formation energies by DFT and MM+MD and measured desorption energies, in kcal/mol.

System		[†]Computed E_{form}		[%]Measured E_{des}
		[*]DFT	[@]FF	
Adsorbed full monolayers in *vacuo*	A_{1ML}@Au(111)	-34.6	-34.8	29.6±0.7
	G_{1ML}@Au(111)	-41.2	-42.8	30.5±0.5
Adsorbed partial monolayers in *vacuo*	$G_{\frac{1}{6}ML}$@Au(111)		-31.3	
Adsorbed bases in water solution	A@Au(111)$_{H_2O}$		1.4±0.4	
	G@Au(111)$_{H_2O}$		0.6±0.4	
	T@Au(111)$_{H_2O}$		3.0±0.8	
	C@Au(111)$_{H_2O}$		1.2±0.4	

[†] The negative values of the formation energy indicate favorable adsorption reactions.
[*] The DFT formation energy is evaluated by Equation 3.
[@] The FF formation energy is evaluated by the sum of Equations 4 and 5 for the monolayers in *vacuo* and the sum of Equations 6 and 7 for the single bases in water.
[%] The measured desorption energy is taken from published work.[78] The positive values of the desorption energy indicate that desorption is an unfavorable process and must be activated.

in *vacuo* in a horizontal configuration with the O atom on top of a Au atom) from our previous study.[45] The similarity between the solid and dashed green lines testifies that the molecular electron states are similarly distributed at the two interfaces. In the HOMO energy range (-1 eV) the two curves are practically identical. Shape differences emerge in the energy range -1.5÷-4.0 eV, where the monolayer has a more disperse guanine-projected DOS.

We note in Figure 2 energy intervals where the projected DOS on both guanine and Au is finite, which is a possible index of electronic coupling. The analysis of the electronic structure for ${}_h\mathrm{G}^t$@Au(111), as well as for the similar systems with the other bases, revealed indeed a special kind of hybridization between adsorbate and substrate. By inspecting the electronic functions in relevant energy ranges, between 0.0 eV and -4.0 eV, we find the same electronic mixing for the periodic guanine monolayer G_{1ML}@Au(111) as for the non-periodic G@Au(111) interface. The guanine HOMO forms hybrid orbitals with the substrate around -1.0 eV, with nodes in the interface region. The guanine HOMO-1 forms bonding orbitals with the substrate in the energy range -2.0÷-4.0 eV: one such example is shown in 3. The mechanism for this molecule-substrate pairing was extensively discussed elsewhere: it is intermediate between the Newns-Anderson

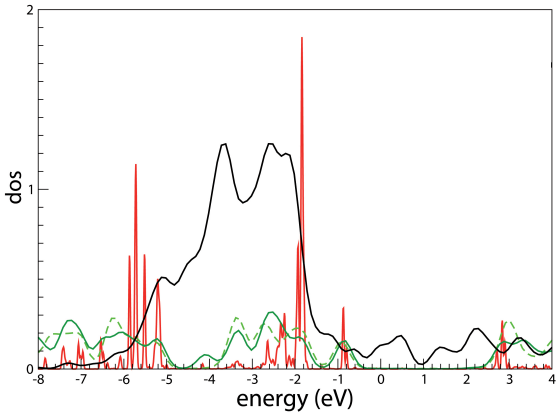


Figure 2: DOS of the optimized G_{1ML}@Au(111) and ${}_hG^t$@Au(111) structures. The origin of the energy scale is set at the Fermi level of the monolayer. The deepest occupied energy level, which is associated to the same orbital in the guanine monolayer and in the isolated guanine molecule, is used for alignment of the different curves. The green solid line is the projection of the total G_{1ML}@Au(111) DOS on the guanine molecules of the monolayer, normalized to one guanine. The black line is the projections of the total G_{1ML}@Au(111) DOS on the outermost substrate layer. The red line is the projection of the total G_{1ML}@Au(111) DOS on the O atoms of the monolayer, normalized to one O atom. The green dashed line is the projection of the total ${}_hG^t$@Au(111) DOS on the guanine molecule. The projected DOS curves are computed by projecting the total DOS onto atomic orbitals and then summing over all the projections that constitute the subsystems of interest.

chemisorption model and pure pysisorption.[45] Similar findings apply to the adenine monolayer A_{1ML}@Au(111).

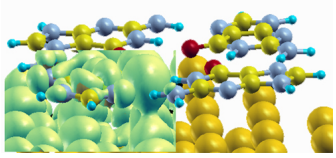


Figure 3: Isosurface plot of an orbital orbital of G_{1ML}@Au(111) at -3.0 eV, which shows a bonding character between the adsorbate and the substrate. The molecular component of this orbital stems from the first occupied orbital below the HOMO of isolated guanine.

The values in Table 1 show that the classical results are in good agreement with DFT results, within 1.6 kcal/mol that is the FF accuracy. We conclude that the GolDNA-AMBER force field,

which parameters were obtained by fitting classical to quantum energies for interfaces of single bases with gold in *vacuo*, is robust beyond the specific parametrized systems. In particular, it is transferable to systems where interaction occurs not only between the adsorbate and the substrate but also within the adsorbate. Confident in this positive feedback, we use it in the next section to investigate the role of an aqueous environment in nucleobase-gold reactions, with a plethora of implications for biological problems.

Yet, let us also discuss the relation between computational results and experimental data. Both DFT and FF calculations overestimate the measured formation energies by 20÷40%. This discrepancy calls for a deeper inspection: it is important to explain how the experimental desorption energy is indirectly extracted from the measurements.

TPD estimates of the desorption energy are derived from the measured desorption rate $N(t)$ according to the Redhead formula:[28]

$$N(t) = \nu N_S \sigma^n exp(-H_d/k_B T_d), \tag{8}$$

where: ν is a pre-exponential factor that depends on atomic masses and bond strength, N_S is the number of available surface sites, σ is the surface coverage, n is the order of the desorption reaction, H_d is the formation enthalpy, k_B is the Boltzmann constant and T_d is the desorption temperature. The evaluation of the desorption enthalpy is usually done assuming that it is independent of surface coverage and that desorption follows a first-order kinetics ($n = 1$ in Equation 8),[82] using a standard value of $\nu = 10^{13}\ s^{-1}$, assuming that the partition function for the free and adsorbed molecule are the same, apart for the degree of freedom along the dimension normal to the surface.[80] The latter assumption is controversial:[80–83] an agreement on a suitable evaluation method for a more accurate pre-exponential factor is elusive, but it is widely accepted that it should depend on the size of the adsorbate molecule. A more accurate evaluation of both ν and H_d can be achieved by performing desorption experiments with different coverages and temperatures and fitting the results with the Polanyi-Wigner equation,[81] which is equivalent to the Redhead equation

with a coverage-dependent H_d. For the evaluation of desorption energies of nucleobase monolayers on Au(111), the simple Redhead formula was adopted with a fixed $\nu = 10^{13}\ s^{-1}$ and a coverage-independent H_d:[78] we argue that this approximation introduces errors in the experimental estimates, which are partly the source of the apparent discrepancy with our calculations. During the desorption process a progressive detachment of the molecules takes places, but in the Redhead estimation of the desorption energy from the measured rate the energy itself is considered to be independent of the coverage. This assumption yields a systematic underestimation of the effective desorption energy when adsorbate-adsorbate interactions are favorable, as in the cases studied here.

To understand whether this assumption is substantially responsible for the deviation of the computed formation energies (Table 1) from the experimental data of desorption energies, we calculated the formation energy of an interface at lower guanine coverage. We considered the single guanine unit of Figure 1, which contains the four guanine molecules marked by the dashed rectangle: we refer to this system as $G_{\frac{1}{6}ML}$@Au(111), because the coverage is $\frac{1}{6}$ relative to the full guanine monolayer (see section Method). We made this choice because experimental data indicate that the molecules first adsorb in units of four G molecules and then cover all the surface.[78] The formation energy per guanine, relative to gas-phase guanine molecules, is much smaller than the formation energy of the monolayer G_{1ML}@Au(111) and is a slight overestimation of the reported experimental value, by only 3%, which is a great improvement with respect to the 40% overestimation for the monolayer. The formation energy of -31.3 kcal/mol for $G_{\frac{1}{6}ML}$@Au(111) is a lower limit for the desorption energy in the case of a partial coverage of the surface. We can safely state that the Redhead approximations in the evaluation of the desorption energy from TPD experiments cause an underestimation of the real value, because of the too low value that is conventionally adopted for the pre-exponential factor and of the missing dependence on the surface coverage in the right-hand side of Equation 8. We do not claim a guess for the correct value of the experimental desorption energy for the monolayers, but we can estimate a similar agreement for guanine as for adenine, namely a precision within 5kcal/mol. For what concerns the geometry,

we correctly describe the structure proposed on the basis of STM imaging by both DFT and FF calculations.

Single bases in water

The FF computed values of formation energy are reported in Table 1. The errors of each component of the energy were estimated by the block averaging technique as implemented in the *g_analyze* Gromacs utility.[65,84] The formation energy is positive for all four adsorbed systems, i.e. the adsorbed configurations are enthalpically unfavored with respect to the solvated bases. This result differs from what we know for the adsorption of nucleobases on the same substrate in *vacuo*[45] (Supporting Information). Nevertheless, we find that once the molecule is adsorbed on the surface, it does not desorb during the whole MD simulation, indicating a preference for the adsorbed configuration.

We then performed two longer MD simulations, 100 ns at 300 K, with guanine and cytosine starting in the middle of the cell immersed in bulk water. In both cases the molecule adsorbed on the surface after approximately 30 ns and it did not desorb during the rest of the simulation: this evidence suggests that the free energy should be representative of the adsorption energetics rather than only the enthalpic term. We speculated that the free energy of formation should be negative if the interface is formed from the solvated molecule. This implies that the entropic term, which accounts for the entropy change in adsorbing the DNA base and desorbing the corresponding water molecules from the gold and molecular surfaces, contributes sizably to the free energy of the adsorption process and favors the adsorbed configuration over the solvated one. Indeed, the adsorption energies in Table 1 are tiny, so that the entropy easily becomes determinant. To verify this picture, we computed the free energy of interaction for cytosine as an illustrative example, by means of MD potential of mean force calculations.

The results are illustrated in Figure 4, which reports the free energy of formation of the interface between cytosine and Au(111) in a water solution, as a function of the cytosine(COM)-surface

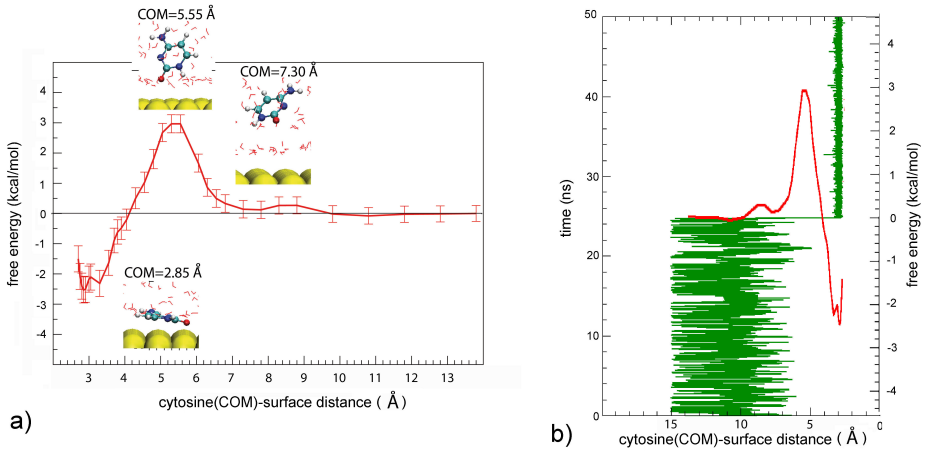


Figure 4: (a) Plot of the free energy landscape for C@Au(111)$_{H_2O}$; insets show snapshots from restrained MD runs for the free-energy calculation. (b) Superposition of the free energy profile (red) to the cytosine(COM)-Au distance in time (green): the latter is extracted from the 100-ns free-MD trajectory and fluctuates always in the same manner after 25 ns.

distance. The lowest-free-energy minimum occurs when the COM of the molecule is at a distance of 2.85 Å from the surface, confirming the preference of the molecule for the adsorbed configuration. The free energy difference between the adsorbed and the solvated configuration is equal to -2.6 ± 0.3 kcal/mol. This result demonstrates that the adsorption of cytosine on Au(111) in water is determined by the entropic contribution, in contrast with the results obtained for the system in *vacuo* where the enthalpic contribution alone explains the C@Au(111) interface. The adsorption process can be further unraveled by analyzing the cytosine orientation with respect to the surface at different COM-surface distances. The insets in 4a visualize various configurations that occur in the restrained MD runs, with the COM of the adsorbate at different distances from the substrate. At the distance of $2.85 \div 3.05$ Å two minima are present: the first one corresponds to cytosine adsorbed horizontally on the surface, slightly tilted in the same geometry found in *vacuo* with DFT calculations. In this case the molecule is adsorbed just above the surface, with its reactive O atom below any water molecules (see inset at 2.85 Å) and the geometry is consequently similar to the one obtained in *vacuo*. The second minimum is practically degenerate with the first one. It is probably due to another adsorption configuration that exists also in *vacuo*,[45] very close in energy to

the most favorable configuration but with the molecular plane almost perpendicular to the surface. At a distance of 5.55 Å there is a maximum of the free energy. The inset at 5.55 Å shows that the O2 atom of cytosine and some water molecules are at the same distance from the substrate. This geometry is strongly unfavored, as water is displaced from the gold and cytosine surfaces but the gold-cytosine interaction is not strong enough to balance the unfavorable de-hydration. From DFT calculations we know that the N3-Au interaction complements the O2-Au interaction, thus fortifying the adsorption mechanism. However, at such a large distance from the surface N3 is well screened by water molecules and does not react with gold. At the distance of 7.30 Å we find a shallow local minimum, characterized by a thick water layer between substrate and adsorbate. This analysis allowed us to better understand the balance between the enthalpic and entropic contributions in the adsorption process. The interaction of water molecules with both the nucleobase and the gold surface is strong enough with respect to the interaction between the nucleobase and the surface to favor the desorbed configuration with respect to the adsorbed one. The fundamental contribution to the adsorption process comes from entropy: the adsorption of the nucleobase on gold allows water molecules to diffuse in the solution, supplying a positive entropic contribution to the free energy. This contribution determines the final free energy profile and the system preference for the adsorbed configuration. Once the molecule is adsorbed on the surface, desorption is unfavored not only by the difference in the free energy value between the two configurations, but more notably by the presence of a free energy barrier. The barrier is due to the competition between water and cytosine for gaining the surface: the relative cytosine-Au and water-Au distances are responsible for the shape of the free energy profile.

The plot in Figure 4b superposes the free energy profile to the cytosine(COM)-gold distance as a function of time. The latter is from the free-MD trajectory of 100 ns starting with solvated cytosine, during which the maximum distance between the cytosine and the periodic image of the surface was 1.5 nm. From this plot we learn that the molecule spent some time in water before eventually overcoming the barrier and getting adsorbed on the surface, where it remained afterward for the entire duration of the MD run.

This interplay between entropy and enthalpy is further analyzed in Figure 5, which displays the densities of water molecules and of cytosine O2 atoms along the reaction coordinate d, which is the water(O)-Au and cytosine(COM)-Au distance, respectively. The blue area represents the distribution of water molecules on Au(111) from an independent free MD simulation (7.5 ns at 300 K) in the absence of cytosine, in the same $8 \times 5\sqrt{3}$ supercell: the first and most intense peak of water density occurs at a water(O)-Au distance that coincides with the cytosine(COM)-Au distance at the lowest-free-energy minimum (Figure 4). This is an additional evidence that the cytosine and water molecules are competing for the surface at that distance and that the adsorbed molecule must replace a layer of interacting water molecules. The cyan area represents the O2 density during the restrained MD simulation with the cytosine(COM)-surface distance of 5.55 Å, a condition that corresponds to the free-energy barrier: therefore, as we already inferred above, the energy barrier corresponds to a situation in which the O2 atoms approach locations that are the preferred ones for water and must displace the water molecules before getting closer to the substrate. In other words, the barrier is due to the unfavored displacement of water from the interface. The gray area represents the O2 distribution during the restrained MD simulation with the cytosine(COM)-surface distance of 7.50 Å, which corresponds to the shallow local minimum of the free energy (Figure 4). In this condition the O2 peak coincides with the second water density peak and the O2 atoms maximize the interaction with water molecules.

Comparison between experimental and computational results could not be done for DNA bases adsorbed on Au(111) in aqueous conditions because of the lack of experimental data. Yet, the similarities of our results to those obtained using a similar procedure for peptides adsorbed on Au(111)[33] validate the approach and the outcome.

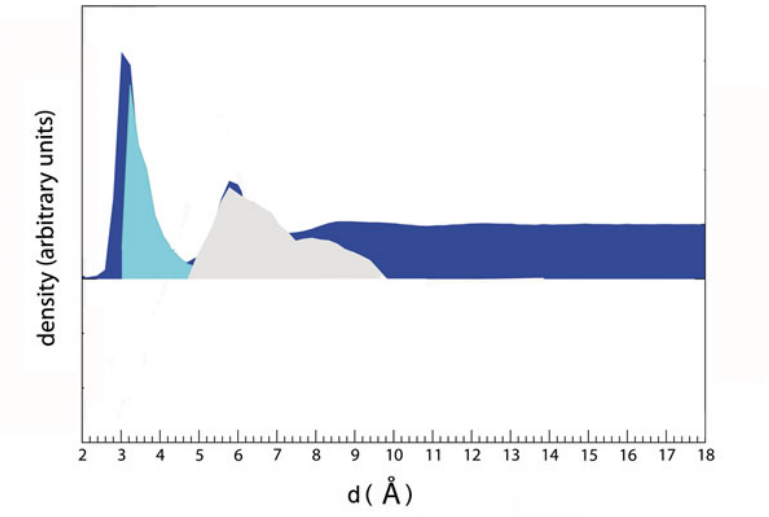


Figure 5: Relation between free energy profile and adsorbate densities. Blue: density of water molecules along the direction perpendicular to the surface, in the absence of any other adsorbate species. Cyan: density of cytosine O2 atoms when the cytosine COM is restrained at 5.55 Å from the surface (this cytosine location corresponds to the maximum of the free energy in Figure 4a). Gray: density of cytosine O2 atoms when the cytosine COM is restrained at 7.30 Å from the surface (this cytosine location corresponds to the shallow minimum of the free energy immediatel after the barrier in Figure 4a). The reaction coordinate on the horizontal axis is the water(O)-Au distance for the blue curve, the cytosine(COM)-Au distance for the cyan and gray curves.

Summary

In conclusion, thanks to the derivation of the tailored GolDNA-AMBER FF, we were able to study the adsorption of DNA bases on the Au(111) surface both in *vacuo* and in aqueous conditions.

Our quantum and classical results for the monolayer formation energies and geometries are unprecedented. They were obtained with a special functional for vdW effects and a tailored force field, respectively. They are in agreement with experimental data. They allow us to confirm the robustness of the GolDNA-AMBER force field and to rationalize the effects of varying coverage on the evaluation of desorption energy from TPD experiments.

The free energy calculations give us insights into the particular adsorption mechanism of DNA bases on Au(111) in water. We find that the interaction between nucleobases and the surface, which is the essential factor regulating the molecule adsorption in *vacuo*, is balanced by the interaction of water molecules with both the adsorbate and the substrate. Consequently, the adsorption process is driven by the entropic contribution to the free energy, which favors the adsorption of the bases and the desorption of water molecules. Our results reveal the presence of a free energy barrier at relatively small molecule-surface distance, due to the competition between water and DNA molecules for the interaction with gold.

GolDNA-AMBER offers a computationally accessible way to study the adsorption of DNA bases on the Au(111) surface, in different geometries, both in *vacuo* and in solution, while accounting for the dynamic polarization of the gold atoms. Its extension to the simulation of DNA oligomers laying flat on gold is a stepping stone towards revealing the effect of an inorganic surface on the state of folding of the double helix.

Acknowledgement

This work was funded by the Italian Institute of Technology through project MOPROSURF and the Computational Platform, by Fondazione Cassa di Risparmio di Modena through Progetto Internazionalizzazione 2011. The ISCRA staff at CINECA (Bologna, Italy) is acknowledged for computational facilities and technical support.

Supporting Information Available

Text on the description of the procedure that was followed to set the parameters of the GolDNA-AMBER FF. Table S1: LJ parameters for Au interacting with DNA bases. Table S2: DFT and FF formation energies of the training set of base@Au(111) structures in *vacuo*. Table S3: LJ parameters for Au interaction with water, for different water models. Table S4: energetics of

water adsorption on Au(111) by DFT and FF calculations. Tables S5 and S6: additional data on the energetics of periodic adsorbed adenine and guanine monolayers. Figure S1: training set of structures for the derivation of the GolDNA-AMBER FF. Figure S2: correlation between vdW-DF and FF computed formation energies for the training set. Figure S3: adsorption structures for water on Au(111). This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Dovichi, N. J.; Zhang, J. *Angew. Chem. Int. Ed.* **2000**, *39*, 4463.

(2) Porath, D.; Bezryadin, A.; de Vries, S.; Dekker, C. *Nature **2000**, 403, 635.*

*(3) Cohen, H.; Nogues, C.; Naaman, R.; Porath. Proc. Natl. Acad. Sci. USA **2005**, 102, 11589.*

*(4) Xu, B.; Zhang, P.; Li, X.; Tao, N. Nano Lett. **2004**, 4, 1107.*

*(5) Kang, N.; Erbe, A.; Scheer, E. New J. Phys. **2008**, 10, 023030.*

*(6) Boon, E. M.; Ceres, D. M.; Drummond, T. G.; Hill, M. G.; Barton, J. K. Nature Biotech. **2000**, 18, 1096.*

(7) Zwolak, M.; Di Ventra, M. Rev. Mod. Phys. ***2008***, 80*, 141.*

(8) He, H.; Scheicher, R. H.; Pandey, R.; Rocha, A. R.; Sanvito, S.; Grigoriev, An.; Ahuja, R. J. Phys. Chem. C ***2008***, 112*, 3456.*

*(9) Shapir, E.; Cohen, H.; Calzolari, A.; Cavazzoni, C.; Ryndyk, D.; Cuniberti, G.; Kotlyar, A. B.; Di Felice, R.; Porath, D. Nature Mater. **2008**, 7, 68.*

*(10) Shapir, E.; Sagiv, L.; Molotsky, T.; Kotlyar, A. B.; Di Felice, R.; Porath, D. J. Phys. Chem. C **2010**, 114, 22079.*

(11) de Pablo, P.; Moreno-Herrero, F.; Colchero, J.; Gomez-Herrero, J.; Herrero, P.; Baró, A. M.; Ordejón, P.; Soler, J. M.; Artacho, E. Phys. Rev. Lett. ***2000****, 85, 4992.*

(12) Porath, D.; Cuniberti, G.; Di Felice, R. Top. Curr. Chem. ***2004****,* 237*, 183.*

(13) Endres, R.G.; Cox, D.L.; Singh, R.R.P. Rev. Mod. Phys. ***2004****,* 76*, 195.*

(14) Mallajosyula, S. S.; Pati, S. K. J. Phys. Chem. Lett. ***2010****,* 1*, 1881.*

(15) Genereux, J. G.; Barton, J. K. Chem. Rev. ***2010****,* 110*, 1642.*

(16) Kotlyar, A.B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath D. Adv. Mater. ***2005****,* 17*, 1901.*

(17) Varsano, D.; Garbesi, A.; Di Felice, R. J. Phys. Chem. B ***2007****,* 111*, 14012.*

(18) Migliore, A.; Corni, S.; Varsano, D.; Klein, M. L.; Di Felice, R. J. Phys. Chem. B ***2009****,* 113*, 9402.*

(19) Woiczikowski, P.; Kubar, T.; Gutierrez, R.; Caetano, R.; Cuniberti, G.; Elstner, M. J. Chem. Phys. ***2009****,* 130*, 215104.*

(20) Woiczikowski, P.; Kubar, T.; Gutierrez, R.; Cuniberti, G.; Elstner, M. J. Chem. Phys. ***2010****,* 133*, 035103.*

(21) Grozema, F. C.; Tonzani, S.; Berlin, Y. A.; Schatz, G. C.; Siebbeles, L. D. A.; Ratner, M. A. J. Am. Chem. Soc. ***2008****, 130, 5157.*

(22) Patwardhan, S.; Tonzani, S.; Lewis, F. D.; Siebbeles, L. D. A.; Schatz, G. C.; Grozema, F. C. J. Phys. Chem. B ***2012****, 116, 11447.*

(23) Daggett, V.; Levitt, M. J. Mol. Biol. ***1993****,* 232*, 600.*

(24) Tirado-Rives, J.; Orozco, M.; Jorgensen, W.L. Biochem. ***1997****,* 36*, 7313.*

(25) Toofanny R. D.; Daggett V. WIREs Comput. Mol. Sci. ***2012****,* 2*, 405-423.*

(26) *Peréz, A.; Orozco, M.* Angew. Chem. Int. Ed. ***2010***, *49, 4805 âĂŞ4808*

(27) *Lee, O-S.; Schatz, G. C.* J. Phys. Chem. C ***2009***, 113, *2316âĂŞ2321*

(28) *Redhead, P. A.* Vacuum, ***1962*** 12.

(29) *Heinz, H; Vaia, R. A.; Farmer, B. L.; Naik, R. R.* J. Ohys. Chem. C ***2008***, 112, *17281.*

(30) *Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Proteins: Structure, Function and Genetics* ***1988***, *4, 31.*

(31) *Iori, F; Corni, S.* J. Comp. Chem. ***2008***,*29, 1656.*

(32) *Iori, F.; Di Felice, R.; Molinari, E.; Corni, S.* J. Comput. Chem. ***2009***, 30, *1465.*

(33) *Wright, L. B.; Rodger, P. M.; Corni, S.; Walsh, T.* J. Chem. Theo. Comput. ***2013***, 9, *1616-1630.*

(34) *Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J.* J. Am. Chem. Soc., ***1996***, 118, *11225.*

(35) *Brooks, B. R.; Brooks, C. L. III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch,S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. J. Comp. Chem.* ***2009***, *30, 1545.*

(36) *Rapino, S.; Zerbetto, F.* Langmuir ***2005***, 21, *2512.*

(37) *Sändig, N.; Zerbetto, F.* Chem. Commun. ***2010***, 46, *667-676.*

(38) *Piana, S.; Billic, A.* J. Phys. Chem. B ***2006***, 110, *23467-23471.*

(39) *Heinz, H.; Farmer, B. L.; Pandey, R. B.; Slocik, J. M.; Patnaik, S. S.; Pachter, R; Naik, R. R.* J. Am. Che. Soc. ***2009***, 131, *9704.*

(40) Heinz, H.; Jha, K, C.; Luettmer-Strathmann, J.; Farmer, B. L.; Naik, R. R. J. R. Soc.,Interface ***2010***, 8, *220.*

(41) Yu, J.; Becker, M. L.; Carri, G. A. Small ***2010***, 6, *2242.*

(42) Yu, J.; Becker, M. L.; Carri, G. A. Langmuir ***2012***, 28, *1408.*

(43) Feng, J.; Slocik, J. M.; Sarikaya, M.; Naik, R. R.; Farmer, B. L.; Heinz, H. Small ***2012***, 8, *1049.*

(44) Rosa, M.; Corni, S.; Di Felice, R. J. Phys. Chem. C ***2012***, 116, *21366-21373.*

(45) Rosa, M.; Corni, S.; Di Felice, R. J. Chem. Theo. Comput. ***2013***, 9, *4552.*

(46) Ercolessi, F.; Parinello, M.; Tosatti, E. Philos. Mag. A ***1988***, *58, 213.*

(47) Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I. R.; Merz, K. M. Jr.; Ferguson, D. M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J. Am. Chem. Soc. ***1995***, *117, 5179.*

(48) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Biophys. J. ***2007***, *92, 3817.*

(49) Perèz, A.; Luque, F. J.; Orozco, M. Acc. Chem. Res. ***2012***, *45, 196.*

(50) Sorin, E.J; Pande, V.S. Biophysical J. ***2005***, 88, *2472-2493.*

(51) DePaul, A. J.; Thompson, E. J.; Patel, S.S.; Haldeman, K.; Sorin E. J. Nucleic Acids Research ***2010***, 38, *4856-4867.*

(52) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Proteins: Struct., Funct., Bioinf. ***2006***, *65, 712.*

(53) Byrd, R. H.; Lu, P.; Nocedal J. J. Scientif. Statistic. Comput. ***1995***, 16, *1190âĂŞ1208.*

(54) Zhu, C.; Byrd, R. H.; Nocedal, J. Trans. Math. Softw. ***1997***, *23, 550âĂŞ560.*

(55) Collier, G.; Vellore, N. A.; Yancey, J.A.; Stuart, S. J.; Latour, R. A. Biointerphases ***2012***, 7, *24.*

(56) Vellore, N. A.; Yancey, J. A.; Collier, G.; Latour, R. A. Langmuir ***2010***, 26, *7396.*

(57) Barone, V.; Casarin, M.; Forrer, D.; Monti, S.; Prampolini, G J. Phys. Chem. C ***2011***, 115, *18434.*

(58) Mendonca, A. C. F.; Malfreyt, P.; Padua, A. A. H.; J. Chem. Theory Comput ***2012***, 8, *3348.*

(59) Ren, P.; Ponder, J. W. J. Phys. Chem. B ***2004***, 108, *13427.*

(60) Harten, U.; Lahee, A. M.; Toennies, J. P.; Woll, C. Phys. Rev. Lett. ***1985***, 54, *2619*

(61) Iori, F.; Corni, S.; Di Felice, R. J. Phys. Chem. C ***2008***, 112, *13540-13545.*

(62) Giannozzi, P. et al. J. Phys.: Condens. Matter ***2009***, 21, *395502.*

(63) Dion, M.; Rydberg, H; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. Phys. Rev. Lett. ***2004***, 92, *246401.*

(64) Vanderbilt, D. Phys. ReV. B ***1990***, 41, *R7892.*

(65) a) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. J. Chem. Theo. Comput. ***2008***, 4, *435-447. b) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, A. E. M.; Berendsen, H. J. C.* J. Comp. Chem. ***2005*** 26

(66) a) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L.G. J. Chem. Phys ***1995***, 103, *8577-8592*

(67) Jorgensen, W. L.; Jenson, C. J. Comp. Chem. ***1997***, 19, *1179-1186*

(68) Hoefling, M.; Iori, F.; Corni, S.; Gottschalk, K. E. Langmuir ***2010***, 26, *8347-8351.*

(69) Trzesniak, D.; Kunz, A. P. E.; van Gunsteren, W. F. ChemPhysChem ***2007***, 8, *162-169.*

(70) http://plasma-gate.weizmann.ac.il/Grace/

(71) Kelly, R. E. A.; Xu, W.; Lukas, M.; Otero, R.; Mura, M.; Lee, Y-J.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. Small ***2008***, 4, *1494-1500.*

(72) Lukas, M.; Kelly, R.E.A.; Kantorovich, L. N; Otero, R.; Xu, W.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. J. Chem. Phys. ***2009***, 130, *024705.*

(73) Xu, W.; Kelly, R.E.A.; Henkjan, G.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. Small ***2009***, 5, *1952-1956.*

(74) Xu, W.; Kelly, R.E.A.;Otero, R.; Schö ck , M.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. Small ***2007***, 3, *2011-2014.*

(75) Kelly, R. E. A.; Lukas, M.; Kantorovich, L. N.; Otero, R.; Xu, W.; Mura, M.; Laegsgaard, E.;Stensgaard, I.; Besenbacher, F. J. Chem. Phys ***2008***, 129, *187707.*

(76) Otero, R.; Lukas, W. X. M.; Kelly, R. E. A.; Xu, W.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. Science ***2008***, 319, *312-315.*

(77) Kelly, R. E. A.; Lukas, M.; Kantorovich, L. N.; Otero. R.; Xu, W.; Mura, M.; Laegsgaard, E.; Stensgaard, I.; Besenbacher, F. J. Chem. Phys. ***2008***, 129, *187707.*

(78) Ostblom, M.; Liedberg, B.; Demers, L. M.; Mirkin, C.A. J. Phys. Chem. B ***2005***, 109, *15150-15160.*

(79) Perdew, J. P.; Burke, K.; Ernzerhof, M. Phys. Rev. Lett. ***1996***, 77, *3865.*

(80) Tait, S. L.; DohnÃąlek, Z.; Campbell, C. T.; Kay, B. D. J. Chem. Phys. ***2005*** 122

(81) Campbell, C. T.; Sellers, J. R. J. Am. Chem. Soc., ***2012*** 134.

(82) Liu, W.; Ruiz, V.G.; Zhang, G. X.; Santra1, B.; Ren, X.;Scheffler, M.; Tkatchenko, A. New Journal of Physics, ***2013***, 15.

(83) Fichthorn, K. A.; Miro, R. A. Phys. Rev. Lett., ***2002***, 89*, 19.*

(84) Hess, B. J. Chem. Phys. ***2002***, 116*, 209âĂŞ217.*

Supporting Information

Marta Rosa,[†,‡] Stefano Corni,[†] and Rosa Di Felice[*,†,¶]

Center S3, CNR Institute of Nanoscience, Via Campi 213/A, 41125 Modena, Italy

E-mail: rosa.difelice@unimore.it

FF parametrization

The procedure that we followed for the parametrization of the GolDNA-AMBER force field (FF) is similar to that developed by us for protein/Au interfaces.[1–3] It is based on the benchmark of classical molecular mechanics (MM) and molecular dynamics (MD) simulations against density functional theory (DFT) calculations of individual bases on Au(111) in vacuo. The formation energies and geometries obtained by MM+MD were compared to those obtained by DFT and the Lennard-Jones (LJ) parameters in the new force field were tuned until the outcomes of the classical and quantum methods match. Because some O and N atoms in the nucleobases are particularly reactive and induce weak chemisorption at base@Au(111) interfaces,[4,5] specific LJ parameters were defined for such atoms, bypassing Lorentz-Berthelot mixing rules. In the following, we explain the details of the computations that were performed to implement the whole procedure and FF tuning.

[*]To whom correspondence should be addressed
[†]Center S3, CNR Institute of Nanoscience, Via Campi 213/A, 41125 Modena, Italy
[‡]Department of Physics, University of Modena and Reggio Emilia, Modena, Italy
[¶]Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA

Computational approach.

The results of DFT calculations that were used to tune the force field are published elsewhere.[4,5] We concisely report the main computational details here and refer the reader to the original papers. We performed pseudopotential plane-wave periodic DFT calculations of the four DNA bases adsorbed at the Au(111) surface with the quantum-espresso package version 5.0,[6] using the PBE exchange correlation functional[7] and the vdW-DF functional.[8] The adsorbate-adsorbate lateral distance between two neighboring replicas was at least 10 Å and the vacuum thickness in the direction perpendicular to the surface was at least 11 Å. Tests in a larger supercell were performed, to check that the chosen size assures no interaction between the molecule and its periodic images. We chose a plane wave basis set with a cutoff of 25 Ry[1] and we described the electron-ion interaction with ultrasoft pseudopotentials.[9] The Brillouin zone (BZ) sums were calculated including 2 Monkhorst-Pack special **k** points in the irreducible wedge. All the atomic coordinates were relaxed until each force vanished within 0.05 eV/Å.

All FF-based calculations were performed using the Gromacs 4 package[10]. The surface was modeled with five Au layers at the vdW-DF equilibrium lattice constant of 3.0 Å, with a periodically repeated $7\times 4\sqrt{3}$ cell (56 atoms per layer): the lateral extension of the supercell was 21.0 x 20.8 Å, while the azimuthal size was 70Å, which corresponds to a vacuum thickness between the adsorbed molecule and the nearest periodic image of approximately 60 Å. The calculations were carried out in the NVT ensemble using a Nose-Hoover thermostat. Particle Mesh Ewald (PME)[11] electrostatic summation was used with a real-space cutoff at 11 Å, whereas a force-switched cutoff starting at 9 Å and ending at 10 Å was used for Lennard-Jones nonbonded interactions.

To identify the minimum energy structure(s) of each DNA base when adsorbed, 40 ps simulated annealing MD runs were carried out. The temperature of the gold dipole rods was maintained at 300 K throughout, while that of the adsorbate was linearly decreased from 300 to 1 K in the 40 ps. Different initial starting positions were tested with the molecule 30 Å far from the surface to check for possible final different equilibrium configurations. When local metastable config-

urations with bases close to the surface found by DFT calculations were not obtained with this classical protocol, we explicitly assumed them as starting conditions for 20-ps runs of classical simulated annealing (namely). In some of these cases, we have also constrained molecule atoms in the directions parallel to the surface. The formation energy was evaluated from LJ, bonding and Coulomb contributions as explained in the article, with the total number of adsorbed molecules $N = 1$ (Equations 4 and 5). For the evaluation of the electrostatic contribution, the MD simulations we 50 ps long at 300 K. During such simulations, all atoms were kept fixed in their initial relaxed positions with only the gold dipoles able to freely rotate.

Fine tuning of the GolDNA-AMBER parameters.

Ten adsorption geometries between those obtained from DFT[5] were selected as the training set for the FF parametrization: for each base the preferred horizontal and vertical configurations were included. In addition, two other energetically competitive vertical configurations were taken into account in the case of guanine. The training set is visualized in Figure S1, where: the bases are indicated with their initial (C, G, A, T); the configurations are labeled as, e.g., ${}_hC@Au(111)^t$); in the configuration label, the left subscript ${}_v$ or ${}_h$ indicates *vertical* or *horizontal* adsorption and the right superscript indicates the adsorption site – t for top, f for fcc and b for bridge; when not differently stated, the indicated adsorption site pertains to the most reactive atom(s) – O2 in C, O6 in G, N3 in A and O2 and O4 in T (e.g. ${}_hC@Au(111)^t$ indicates cytosine adsorbed horizontally with the O2 atom on top of a substrate Au atom). To denote structures in which a different/second atom is adsorbed on a specific site, the name of the atom and its adsorption site (e.g. ${}_vC@Au(111)^{t+N7t}$) is indicated in the superscript.

The classical FF parameters were gradually tuned via a trial-and-error procedure so that MM+MD structural optimizations would reproduce the DFT geometries and formation energies. We adopted the following criteria to evaluate the performance of the classical *versus* quantum calculations: (i) the adsorption energies of a given base@Au(111) system must be as close as possible between FF

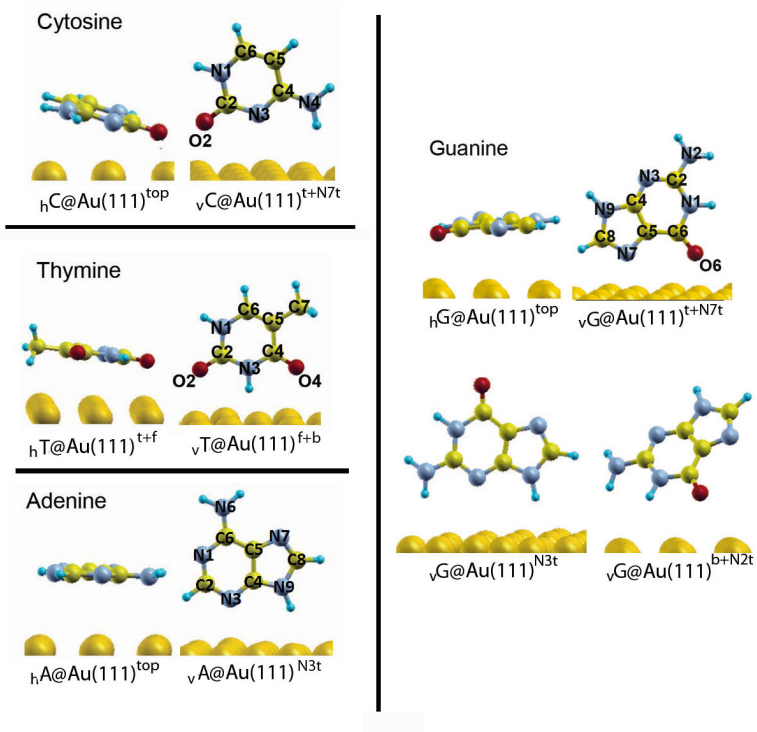


Figure S1: DNA bases adsorbed on Au(111) in vacuo:[5] training set for FF parameter tuning.

and DFT, within a chosen accuracy, for representative configurations reported in Table 1; (ii) the relative order of adsorption strength for the four bases (A < T < C < G) that we obtained from DFT must be reproduced by classical FF simulations; (iii) the relative order of the different adsorption configurations for each base must be the same from FF and DFT calculations, in particular the order between horizontal (more favorable) and vertical (less favorable) configurations must be maintained. We pursued a force field as general as possible, to be suitable to describe situations in which DNA adsorbs on gold not only in dry but also in wet conditions. For this reason we acted only on Lennard-Jones (LJ) parameters, not modifying the other interactions in ffamber99sb[12] and GolP.[2] We started by defining σ_{Au} and ε_{Au} for Au atoms and calculating all the LJ base-gold parameters via AMBER Lorentz-Berthelot mixing rules. The initial values for the two gold parameters were equal to those of GolP-OPLS and then they were modified to reproduce DFT results on adsorbed bases.

Tuning Au LJ parameters without any specific N-Au and O-Au interaction terms (data not reported), the horizontal configuration energies are well reproduced within 2 kcal/mol. Adsorption geometries are in qualitative agreement with DFT calculations, with cytosine and guanine tilted relative to the surface exploiting the interaction between O and Au. For vertical configurations, on the contrary, FF formation energies are sensibly larger than the DFT formation energies. Thus, we started to define specific LJ parameters for the N-Au and O-Au interactions, bypassing the Lorentz-Berthelot mixing rules.[13] In doing so, we were able to modify the classical formation energies of the configurations without affecting those of horizontal configurations that were already adjusted. We maintained the use of mixing rules for the intra- and inter-molecular LJ interactions, but we defined one specific parameter for the interaction of O atoms (O2 in cytosine, O2 and O4 in thymine, O6 in guanine) with Au atoms, one specific parameter for the interaction of the N3 atom in cytosine, guanine and adenine with Au, one specific parameter for the interaction for the N7 atom in guanine and adenine with Au. Atom labels are as in Figure S1. The specific LJ terms O2-Au, O4-Au, O6-Au, N3-Au and N7-Au were the same in all the bases. Particular attention was required by guanine, because the O6-Au interaction plays a strong role in the rela-

tive stability of vertical adsorption configurations. We were able to find only one set of these LJ parameters that gave good FF performance for guanine, adenine and cytosine. No unique choice could also account for the correct energetics of thymine. This peculiarity of thymine is probably due to its methyl group, which confers a high steric hindrance forcing thymine to remain too far from the substrate. The DFT results sustain this interpretation. In fact, in DFT relaxed geometries the O2-surface distance is 2.7 Å in C@Au(111) and the O2,O4-Au distances are 3.2÷3.6 Å in T@Au(111): consequently, the O-Au interaction is of a different kind in thymine and we eventually tuned different LJ parameters for it. We solved the problem of thymine parametrization by maintaining the original AMBER parameters for the O atoms of thymine and expoiting Lorentz-Berthelot mixing rules to calculate the interaction between the molecule and the Au surface. The optimized parameters are shown in Table S3 and the resulting energies are presented in Table S2. The FF calculated adsorption energies show a fair agreement with quantum results (see also Figure S2). The difference in formation energy between FF and DFT results for a given interface is within 2 kcal/mol: the worst case is that of vertical adenine ${}_{v}\mathrm{A}@\mathrm{Au}(111)^{N3t}$, which, however, does not compromise a fair accuracy. Furthermore, the criteria about the energetic order are respected. Thus, we conclude that we obtained a classical force field that is able to describe the interaction of DNA bases with the Au(111) surface and is the starting point towards classical MD simulations of entire DNA oligomers on Au(111) in dry and wet environments. Structural data (not shown) support the same conclusion. It is particularly nice that the FF reproduces the DFT result that horizontal adsorption configurations are more energetically favorable than vertical configurations, which in quantum calculations is a specific outcome of the inclusion of vdW interaction. The non-specific LJ parameters for GolDNA-AMBER are very similar to those for GolP.[2] Small deviations can be attributed to the differences between AMBER and OPLS. This result is an important corroboration of the portability of the GolP construction.

With the aim of simulating a solution environment for the adsorption of DNA bases on Au(111), we also parameterized the water@Au(111) interface, following an assessed procedure.[14] Car-Parrinello Molecular Dynamics (CPMD) simulations have shown that interfacial water molecules

Table S1: LJ parameters for Au with Lorentz-Berthelot mixing rules (top) and with specific O-Au and N-Au terms (bottom).

Lorentz-Berthelot		σ_{Au} (nm)	ε_{Au} (kcal/mol)
		0.31	0.68
specific		σ_{X-Au} (nm)	ε_{X-Au} (kcal/mol)
	X=O2,O4	0.28	1.68
	X=N3	0.27	1.62
	X=N7	0.29	1.58

Table S2: Calculated adsorption energies for DNA bases adsorbed on Au(111) in vacuo in different configurations.

	ΔE_{MD} $kcal/mol$	ΔE_{DFT} $kcal/mol$
${}_{h}\mathrm{C@Au(111)}^{t}$	-19.7	-19.7
${}_{v}\mathrm{C@Au(111)}^{t+N3t}$	-18.8	-19.2
${}_{h}\mathrm{G@Au(111)}^{t}$	-24.6	-23.3
${}_{v}\mathrm{G@Au(111)}^{N3t}$	-15.6	-15.2
${}_{v}\mathrm{G@Au(111)}^{t+N7t}$	-18.3	-18.6
${}_{v}\mathrm{G@Au(111)}^{f+N2t}$	-16.1	-16.7
${}_{h}\mathrm{T@Au(111)}^{t+f}$	-17.0	-17.2
${}_{v}\mathrm{T@Au(111)}^{f+b}$	-11.5	-11.0
${}_{h}\mathrm{A@Au(111)}^{t}$	-19.6	-19.9
${}_{v}\mathrm{A@Au(111)}^{N3t}$	-15.4	-17.4

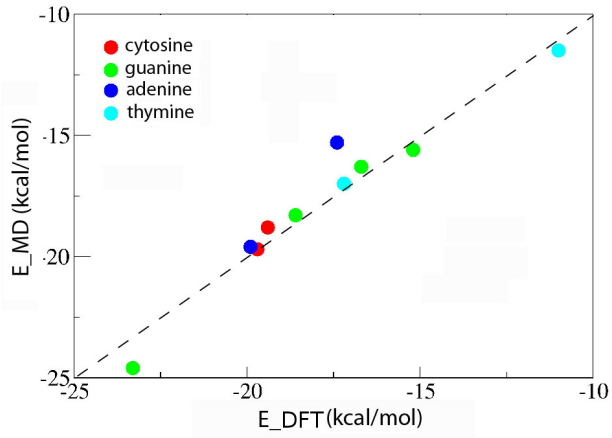


Figure S2: Correlation between FF and vdW-DF(DFT) results for the formation energy of DNA bases adsorbed on Au(111) in vacuo in different configurations.

act as hydrogen bond donors to other water molecules, but this behavior could not be classically simulated without defining special interaction parameters for water molecules with Au.[15] Effectively, without a special parametrization for water, MD calculations may give an incorrect geometry for water molecules, with one of the two H atoms pointing towards the surface,[14] while the correct DFT geometry is with O atom adsorbed on top adsorption site (Figure S3).

For what concerns the adsorption energy, if special parameters for water O atoms are not defined the strength of the interaction is underestimated by FF calculations with respect to DFT results (-3.43 kcal/mol with FF vs -4.79 kcal/mol with DFT). A slight correction to GolP was initially introduced by defining a specific H_{wat}-Au LJ component.[15] A refined parametrization in GolP-CHARMM includes H_{wat}-Au and O_{wat}-Au LJ components.[14] Following the same approach, we assessed the GolDNA-AMBER FF against vdW-DF DFT calculations of single water molecules on Au(111), using the SPC(/E) and TIP3P water models.[16,17] We started the iterative adjustment adopting the specific LJ parameters for H_{wat}-Au proposed by Cicero and coworkers,[15] and tuning the LJ parameters for O_{wat}-Au different models. Instead of defining independent $\sigma_{O_{wat}-Au}$ and $\varepsilon_{O_{wat}-Au}$ parameters for each water model, we prefer to obtain them via Lorentz-Berthelot mixing rules as:

$$\begin{aligned} \sigma_{O_{wat}-Au} &= \frac{\sigma_{O_{wat}} + \sigma_{Au^{wat}}}{2} \\ \varepsilon_{O_{wat}-Au} &= \sqrt{\varepsilon_{O_{wat}} \times \varepsilon_{Au^{wat}}} \end{aligned} \tag{1}$$

where σ_{Au}^{wat} and ε_{Au}^{wat} are the adjustable parameters, common to all water models (the superscript *wat* is used to distinguish these parameters from the generic Au ones). $\sigma_{O_{wat}}$ and $\varepsilon_{O_{wat}}$ are the standard (model-specific) O_{wat} parameters. Clearly, a better FF vs MD agreement could be obtained by using independent $\sigma_{O_{wat}-Au}, \varepsilon_{O_{wat}-Au}$ pairs, one per each water model. However, there is no physical reason to justify such multiplication of parameters, and we therefore chose to compromise between accuracy (still satisfactory) and complexity of the force field.

Comparing final DFT and MD results we can see that in the optimized DFT H_2O@Au(111) ge-

ometry the plane of the water molecule is slightly tilted (by approximately 35°) with respect to the surface normal, while in the optimized FF H_2O@Au(111) geometry the plane of the water molecule is perpendicular with respect to the surface (i.e., tilt angle 0°). We should point out that the dependence of the DFT formation energy on the tilt angle is very small; thus, we decided to accept the FF vertical configuration as viable. A comparison between DFT and MD result is presented in Table S4; the agreement between the two methods is within 0.15 kcal/mol. In this work, we decided to use the TIP3P water model because it is well tested in AMBER simulations of nucleic acids in solution. Yet, we remark that this GolDNA-AMBER parametrization is also compatible with the SCP water model.

Table S3: LJ parameters of the GolDNA FF for different water models.

TIP3P		
	σ_{X-Au} (nm)	ε_{X-Au} (kcal/mol)
X=O_{wat}	0.342	1.236
X=H_{wat}	0.340	0.010
SCP		
	σ_{X-Au} (nm)	ε_{X-Au} (kcal/mol)
X=O_{wat}	0.343	1.250
X=H_{wat}	0.340	0.010

Table S4: Adsorption energy upon the adsorption of a single water molecule on Au(111), as obtained from DFT and with our new FF compatible with TIP3P, SPC and SPC/E. z_O is the distance from the O atom of the water molecule from the surface in the optimized configuration.

	E_{form} (kcal/mol)	z_O (Å)
vdW-DF	-4.79	2.9
TIP3P	-4.65	3.0
SPC	-4.65	3.0
SPC/E	-4.81	3.0

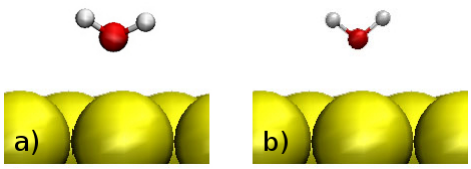


Figure S3: The most favorable orientation of a single water molecule adsorbed onto Au(111) by vdW-DF calculations (a) and the new FF (b) with water LJ parameters.

Periodic guanine and adenine monolayers on Au(111)

Table S5: Comparison between formation energies of different adenine monolayers that have been proposed on the basis of experimental data. The monolayers are identified with the labels used in the experimental works.[18,19] $A(2)_1$ and $A(2)_2$ refer to the same AFM geometry and different DFT reconstructions, while $A(4)_1$ corresponds to an AFM geometry obtained at a different temperature (that is the one reproduced in this work).

monolayer[18,19]	Energy (kcal/mol)		
	PBE functional PBE Au lattice	vdW-DF functional PBE Au lattice	vdW-DF functional vdW-DF Au lattice
$A(2)_1$	-20.3	-18.0	-18.8
$A(2)_2$	-17.9	-15.6	-16.2
$A(4)_1$	-20.9	-18.9	-19.7

Table S6: Lattice parameters and monolayer formation energies per molecule from our work and publishe works.[18,19] Our vdW-DF values are given both for optimized gas-phase monolayers and for the final geometry selected for the adsorbed calculations.

monolayer[18,19]	geometry and energy	DFT no vdW[18,19]	DFT with vdW optimized	DFT with vdW adjusted
adenine	d1 (Å)	12.2	12.6	13.0
	d2 (Å)	15.4	15.8	15.6
	γ (degrees)	90.3	90	90
	energy (kcal/mol)	-20.9	-19.7	-19.4
guanine	d1 (Å)	–	15.8	14.99
	d2 (Å)	–	15.4	15.58
	γ (degrees)	–	90	90
	energy (kcal/mol)	-26.2	-24.4	- 23.4

References

(1) Iori, F.; Corni, S.; Di Felice, R. *J. Phys. Chem. C* **2008**, *112*, 13540-13545.

(2) Iori, F.; Di Felice, R.; Molinari, E.; Corni, S. *J. Comput. Chem.* **2009**, *30*, 1465.

(3) Iori, F; Corni, S. *J. Comp. Chem.* **2008**,29, 1656.

(4) Rosa, M.; Corni, S.; Di Felice, R. *J. Phys. Chem. C* **2012**, *116*, 21366-21373.

(5) Rosa, M.; Corni, S.; Di Felice, R. *J. Chem. Theo. Comput.* **2013**, *9*, 4552.

(6) Giannozzi, P. et al. *J. Phys.: Condens. Matter* **2009**, *21*, 395502

(7) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. ReV. Lett.* **1996**, *77*, 3865.

(8) Dion, M.; Rydberg, H; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(9) Vanderbilt, D. *Phys. ReV. B* **1990**, *41*, R7892.

(10) a) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theo. Comput.* **2008**, *4*, 435-447. b) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, A. E. M. Berendsen, H. J. C. *J. Comp. Chem.* **2005** *26*

(11) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L.G. *J. Chem. Phys* **1995**, *103*, 8577-8592

(12) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.

(13) For matching FF and DFT date we must take care of two different LJ components. One is the long-range VdW interaction, which is responsible for the horizontal adsorption preference and is acting in the same way on the different atoms of the molecules. This interaction is well reproduced with the combination of LJ parameters of molecules and gold atoms, without the

need for a special attention. On the other hand, O atoms and unsaturated N atoms interact with the surface also by stronger short-range interactions, evidenced in DFT calculations by the formation of bonding orbitals and by the large absolute values of the adsorption energies. This weak bonding of the molecules needs tailored LJ parameters in GolDNA-AMBER.[2]

(14) Wright, L. B.; Rodger, P. M.; Corni, S.; Walsh, T. *J. Chem. Theo. Comput.* **2013**, *9*, 1616-1630.

(15) Cicero, G.; Calzolari, A.; Corni, S.; Catellani, A. *J. Phys. Chem. Lett.* **2011**, 2, 2582.

(16) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *B. Pullman, Ed., Reidel, Dordrecht* **1981** 331

(17) Jorgensen, W. L.; Chandrasekhar, J; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983** *79*

(18) Kelly, R. E. A.; Xu, W.; Lukas, M.; Otero, R.; Mura, M.; Lee, Y-J.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Small* **2008**, *4*, 1494-1500.

(19) Lukas, M.; Kelly, R.E.A.; Kantorovich, L. N; Otero, R.; Xu, W.; Laegsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *J. Chem. Phys.* **2009**, *130*, 024705.

Chapter 7

Docking of DNA molecule on Au(111)

Docking simulations were performed with the aim of understanding the forces that drive the adsorption of the DNA molecule on the surface. The analysis of the final adsorption configurations allows to evaluate the effect of the different components of the interaction (i.e. electrostatic, LJ). The obtained adsorption configurations can than be used as possible starting configurations for molecular dynamics simulations. The advantage of docking is that it treats implicitly the effect of the solvent, and constrains the internal degrees of freedom of the molecule, thus reducing the computational cost of the simulation and allowing the study of large systems. This method is not suitable for the description of the DNA unfolding process, but allows to find the adsorption geometries for DNA on Au(111), and to distinguish the effects of the different energies components.

We employ the SDA software, which implements Brownian dynamics according to the theory in Section 2.4. The SDA software was developed and is maintained by the group of Prof. Rebecca Wade (www.h-its.org/english/homes/wade) in Hidelberg, where I spent 3 month in 2012. During this visit, I worked with group members to learn the software and perform docking simulations jointly including the developments necessary to treat DNA-Au association.

The first step of the study was to include LJ parameters obtained with DFT and MD calculations in the docking simulation and to verify the agreement between the two methods. A

simple test, done excluding all the energy contributions apart from LJ and starting a simulation with the single bases already adsorbed on the surface, allowed to record the LJ interaction. As expected, the agreement between MD and docking is excellent (7.1).

base	docking	MD
cyt	-14.4	-14.1
gua	-19.3	-19.7
thy	-14.7	-14.4
ade	-17.8	-18.2

Table 7.1: LJ interaction obtained from docking and MD calculations. Energies are given in kcal/mol

Then, assuming we have a reliable description of the Au(111) substrate as a host of DNA adsorption, we proceeded to simulate DNA oligomers docking onto Au(111). We chose a standard length of 15 base pairs, which behaves well in MD and docking simulations, and uniform sequences poly(dG-poly(dC) and poly(dA)-poly(dT), for which experimental characterization on gold exists[16]. This choice of sequences allows us to distinguish between the behavior of the four bases comparing adsorption energies and geometries of the two DNA molecules.
The choiche of the length must comply with some requirements. In docking simulations the internal degrees of freedom of the molecule are constrained; as a consequence the molecule itself must be short enough to be suitable for the rigid body approximation. On the other hand we need a molecule long enough to appreciate the effects of the different components of the energy (i.e. desolvation, LJ, electrostatic).
As we already mentioned in Chapter 2, desolvation energy is a positive term that accounts for the unfavorable displacement of water molecules from DNA and metal surfaces when the two components get in contact. This interaction is balanced by electrostatic and LJ interactions, that favor the DNA adsorbed configuration with respect to the desorbed one. The three contributions obviously depend on the geometry and length of the molecule. Before generating docking structures, we tried to understand if the balance between the three energy components depends on the molecular length, which would be a bias to our simulations.
To this aim preliminary docking simulations were performed in general conditions, with poly(dA)-

poly(dT) molecules of different length. The clustering of the obtained trajectories was performed in a simple way, collecting all the geometries with a RMSD value difference lower than 5 Å. The obtained adsorption configurations can be divided roughly in two groups: geometries with the molecule adsorbed parallel with respect to the surface and geometries with the molecule adsorbed tilted with respect to the surface. These two groups can be further divided depending on the rotation of the DNA molecule with respect to the surface; details are given in the following paragraph. Here the orders of magnitude of desolvation (HD) and LJ energies are compared (7.2. Electrostatic interaction is not reported as it depends on the external conditions and it is grows approximatively linearly with the number of nucleotides that touch the surface.

n base pairs	parallel			tilted		
	LJ	HD	ratio	LJ	HD	ratio
8	-22.5	26.7	-0.84	-75	45	-1.67
15	-44.5	-46.8	-0.95	-79	45	-1.76
20	-60.5	-67.6	-0.9	-85	52	-1.64

Table 7.2: LJ and desorption energy values from DNA adsorbed on Au(111) with DNA molecule of different length and uniform sequence poly(dA)-poly(dT). The two groups with parallel and tilted molecules (relative to the surface). Each group is composed of approximately 10^6 structures and the energy values are the one of the structure with lower total energy. The energies are in kcal/mol

As can be seen in 7.2, the ratio between the two components of the energy is similar through the different lengths, even if it is not constant because of the twisted shape of DNA: for the tilted configurations LJ is dominant with respect to desolvation, while for the parallel configurations the two energies are very close to each other, with HD always larger than LJ. The two components of the energy in the parallel adsorption geometry grow approximatively linearly with the length of the helix, despite its twisted shape. These tests indicate that the length of the molecule does not substantially modify the ratio between the energy components. We chose 15 bp as the optimal length: keepeng the molecule as short as possible, with the aim of avoiding non-physical adsorption geometries due to the rigid body approximation, this lenght allows the

molecule to touch the surface in three or four points in the parallel adsorption configuration, giving the opportunity to distinguish the behavior of purines and pyrimidines (i.e. the molecule can be rotated in a configuration with purines-pyrimidines-purines adsorbed on the surface or in the opposite way. The molecule can also adsorb on the surface in four points, equally distributed between purines and pyridines).

Once the length of the helix has been chosen, electrostatic parameters need to be defined. The adsorption of a DNA molecule horizontally on a metallic surface in experiments is usually performed in two possible ways: the surface is positively charged to attract DNA[97], or DNA is adsorbed in two consecutive steps. In two-step protocols, DNA single strands are adsorbed on the surface in a solution with high concentration of ions, then the complementary strand are added in the solution. The high ionic concentration neutralizes the electrostatic repulsion between the two strands, so that hybridization occurs between the two complementary strands and duplex DNA molecules are formed[16].

With the idea of comparing our simulations with experiments, we decided to perform calculations both in high (500mM) and low (10mM) ionic concentration, and to deal with the charged surface only in case of need. Moreover, for the simulation performed in high ionic strength (ios), a further consideration applies: DNA charges are screened by the implicit ions in solution only when the solvent is in between the molecule and the surface. When the molecule is 5 Å or less close to gold, the solvent is supposed to be already displaced, but the possibility of ions trapped between DNA and metal is not taken into account. For this reason we decided to perform some calculations with half of the real charge on the molecule atoms, to simulate this screened interaction.

Two clustering methods can be used. One approach consists in a bottom-up aggregation: a threshold RMSD value is defined and all the geometries within that value from the representative one (i.e. the one with the most negative energy) are grouped together. The other approach consists of a top-down splitting (hierarchical based on a reference structure): a number of final clusters is defined and the software groups all the configurations in such a way to fulfill the requirement. The second method has the obvious disadvantage that, if the requested number of clusters is two small, very different geometries can be grouped together; on the other hand the first method fails in grouping together very similar geometries that differ only for small rotations. This latter defect is particularly problematic in the case of the DNA molecule, which

is symmetric with respect to the central axis. For these reasons, we tested both clustering methods for our systems. Results reported in 7.3 are obtained with different clustering methods, depending on the simulation conditions. For simulations with ios=10mM the hyerarchical clustering is used, as the RMSD-based clustering method fails in grouping together parallel adsorption configurations that differ only for small rotations around the central axis that do not modify the geometry of the molecule. For simulations with ios=500mM the RMSD method is used, as the hierarchical clustering method fails in distinguishing between the vertical configurations and the tilted one that touches the surface with the purine strand. It is important to stress that, despite these difficulties in the clustering process, dues to the fact that the software is optimized for proteins and not for DNA, the overall description of the system behavior (i.e. possible and preferred adsorption geometries) is the same from the two methods.

geometry	**ClSize**	**E-av**	**ClSize**	**E-av**
poly(dA)-poly(dT)				
	ios=10mM		ios=500mM	
tilted A strand	62%	-33.9	*27 %*	*-16.5*
tilted T strand	35%	-33.2	*18%*	*-16.5*
parallel A-T-A	3%	-32.5	-	-
vertical	-	-	*55%*	*-13.1*
poly(dC)-poly(dG)				
	ios=10mM		ios=500mM	
tilted G strand	94%	-44.7	*22%*	*-26.7*
tilted C strand	4%	-36.2	*10%*	*-27.2*
parallel C-G-C	2%	-36.2	-	-
vertical	-	-	*68%*	*14.9*

Table 7.3: Energies of the different adsorption geometries obtained at different ionic strength values for the two DNA molecules. ClSize is the relative population the cluster with respect to the total, E-av is the average total energy of the cluster in kcal/mol

Figure 7.1 shows three distinguishable adsorption configurations that realize in the different

simulation conditions: the DNA molecule is adsorbed parallel to the surface, tilted with respect to the surface or approximately vertically. The parallel and tilted geometries can be further grouped depending on the rotation of the molecule: the tilted DNA molecule can touch the surface with the purine or the pyrimidine strand; in a parallel adsorption geometry, the molecule touches the surface in three locations and therefore the final configuration can be either purines-pyrimidines-purines or the opposite (namely A-T-A or T-A-T and C-G-C or G-C-G).

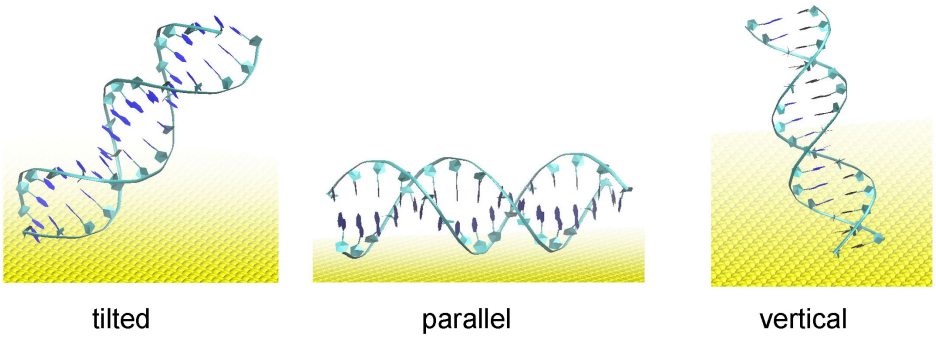


Figure 7.1: possible docking configurations of a 15bp poly(dA)-poly(dT) molecule on Au(111)

The results reported in 7.3 show that adsorption geometries are similar for the two sequences. At low ionic strength the most populated cluster is the one with the DNA molecule tilted and the purine strand touching the surface, followed by the symmetric geometry, that is the tilted one with the pyrimidine strand adsorbed on gold. The molecule is adsorbed horizontally on the surface only at low ionic strength, while at high ionic concentration this geometry is substituted by the vertical adsorption, that is also the most populated due to the ions screening of electrostatic interaction unfavorings the backbone adsorption. When the simulation is performed at high ionic concentration with half of the charge on DNA atoms, with the aim of mimic the trapping of ions between the molecule and the surface, results obtained are very similar to the ones obtained at ios=500 mM: the molecules are adsorbed only in tilted and vertical geometries. From these results we deduce that the electrostatic interaction between the molecule and the surface is determinant and hinders the parallel adsorption of the molecule. In the preliminary calculations done for choosing the length of the molecule, we already noticed that in the parallel adsorption configuration the desolvation and LJ interactions provide similar contributions to the total energy; as a consequence the contribution of electrostatic is fundamental. Calculations

done with a charged surface would surely change the balance between the populations of the different adsorption orientations, as the parallel configuration would gain importance due to the increase of the electrostatic interaction. As we already mentioned, charging the metallic surface is one of the experimental techniques used for inducing DNA adsorption on metals. For the moment we are not interested in performing this further simulation, as the calculations with the neutral surface already allowed us to obtain a parallel adsorption geometry suitable as input for MD simulations.

It is interesting to note that DNA molecules are adsorbed horizontally on the surface only in one of the two possible rotations: poly(dA)-poly(dT) geometry is "A-T-A" , poly(dC)-poly(dG) geometry is "C-G-C". At a first sight the poly(dC)-poly(dG) geometry is in contrast with DFT results, as we know that purines interact with the surface more strongly than pyrimidines, with the global adsorption order for the four nucleobases that is G > A $\sim$ C > T. Consequently one would expect the purine strand to touch the surface twice to maximize the interaction. Nevertheless when the DNA molecule is adsorbed horizontally on the surface, the individual nucleobases are oriented perpendicularly with respect to gold, rotated differently depending on their position on the strand. The nucleobases energetic order for vertical adsorption is C > G > A > T. This means that for both the adsorbed DNA molecules the parallel rotation is determined by the interaction of the nucleobases with the surface, with the nucleobase that interacts the most with gold that touches the surface two times. From these results it emerges that, even if it is the electrostatic interaction between the surface and the backbone that allows the parallel adsorption, the interaction between the internal bases and the surface is determinant for the final geometry.

This is an encouraging result towards exploring the state of folding of DNA adsorbed on Au(111): these calculations suggest that the bases iside the helix interact with the surface despite the strong electrostatic influence of the backbone. This means that the folding can be affected by nucleobases-surface interaction.

For what concerns the tilted adsorption geometries, the energy difference between pyrimidine and purine strand adsorption configurations is very small, even if the cluster dimension favors the latter geometry. It must not be forgotten that the rigid position of the molecule is an artifact of the rigid body approximation. Since this tilted configuration is the most frequent in the docking simulations, it is interesting to adopt it as a starting geometry for MD simulations.

In the MD simulation the molecule will be allowed to relax, and it will be possible to disclose if, even with the molecule starting tilted with respect to gold, it will then evolve to a parallel orientation. This starting configuration could even favor an unfolding process.

Finally, the vertical adsorption configuration is a consequence of the lower interaction between the molecule and the surface at ios=500mM. This geometry is less interesting for what concerns the study of the behavior of DNA on gold. Experimentally DNA molecules are usually adsorbed vertically within self-assembled monolayers using functional groups (thiols) for strengthening the molecule-surface bond. This docking vertical geometry is probably unstable for isolated non-functionalized molecules and it is not reproduced experimentally. For this reason we are not interested in adopting it as a starting configuration for MD simulations.

Comparing the average energy values of the different adsorption geometries for the two DNA sequences, in particular for what concerns the parallel adsorption, we note that poly(dC)-poly(dG) interacts more strongly with the surface, with this difference that has to be attributed to the interaction of the bases with the surface, as the electrostatic interaction depends on the backbone and is the same for the two sequences. This is a further evidence of the fact that nucleobase interaction with the surface influences the molecule adsorption geometry, despite the strong effect of the charged backbone.

The results obtained from docking calculations are still too general to be compared with experiments, while similar calculations performed with a higher level of theory have not been published. Nevertheless the fact that the same configurations are obtained through all the different calculations for these short sequences, together with the fact that the obtained geometries are the more intuitive when dealing with the simple shape of the DNA molecule, is enough to validate the results obtained. Eventually, they provide a clear evidence of the fact that LJ interaction between the bases and the surface is fundamental for the adsortpion process, eves when the helix is still folded.

Chapter 8

Conclusions

Within this thesis work an extensive study of DNA bases adsorbed on Au(111) in different conditions has been performed exploiting different computational techniques. Density functional theory calculations have been employed for the study of single DNA bases adsorbed on gold in vacuo. The incorporation of the vdW interaction into DFT calculations through specific exchange-correlation functionals has been evaluated on test systems, and the effects of its inclusion into the calculations for DNA bases on Au have been assessed. From our calculations it clearly emerges that a correct description of this system can be obtained only with the inclusion of dispersion interactions into the calculations, as the difference between results with and without vdW is both quantitative and qualitative. The electronic structure of the adsorbed systems was also investigated and this analysis revealed an interaction both of long and short range kind between DNA bases and gold, which is a pristine result of this study.

On the basis of DFT calculations a FF suitable for the description of DNA adsorption on gold was parametrized. DFT results for single bases were reproduced with the FF fitting the adsorption energies and geometries of the different adsorption configurations. The developed FF was then used to study the adsorption of DNA monolayers on gold, comparing the results obtained with DFT calculations and experiments on the same systems. We performed a detailed study of monolayer structures, including the ab-initio electronic structure of a fully supported monolayer: this study allowed us to recognize the same adsorption mechanism of isolated DNA bases on Au, supporting the recent findings that in the case of molecular homo- and hetero-cycles ad-

sorbed on surfaces there is an electronic partnership between adsorbate and substrate, contrary to what is commonly thought. MD calculations on adsorbed monolayers and comparison with experiments allowed us to validate the computational approach adopted and the parametrized FF and to prove the roboustness and transferability of the FF.

The adsorption of single DNA bases on Au in aqueous conditions was studied through classical MD simulations with new FF. Free energy calculations of the adsorption process for the cytosine molecule revealed an interesting balance between enthalpy and entropy contributions. The nucleobase adsorption process is enthalpically unfavored due to the displacement of water molecules from the metal and molecule interfaces. Nevertheless, we find that the entropic contribution to the free energy is fundamental for determining the final preference for the adsorbed configuration and that, due to the presence of water, DNA molecules have to overcome a free energy barrier to be adsorbed on the surface.

Finally, docking calculations were performed, including LJ parameters of the MD FF into the description of the system. DNA 15-mers with uniform sequences were adsorbed on the surface, and two possible adsorption configurations were determined. From the analysis of the adsorbed geometries it emerged clearly that LJ interaction between the nucleobases inside the helix and the surface influences the adsorption mechanism and the final geometry, despite the strong electrostatic interaction of the backbone with the surface. Moreover, these configurations are valid models as starting geometries for classical MD simulations of a DNA molecule adsorbed on Au. This work intentionally stops before the practical MD simulation of a DNA oligomer on gold: indeed, the aim of my research activities was the development of the enabling methodology, which aim was successfully attained. In fact, the validated FF parameters, plus the knowledge of docking structures that were obtained using such parameters, are ready for use in massive simulations.

The study performed in this thesis lays the foundations for the correct understanding of the behavior of DNA interacting with surfaces. The study of these kind of systems is of great importance in the process of developing new nanotechnology devices based on the interaction between organic and inorganic elements. The development of a FF suitable for the description of DNA adsorption is the starting point for the study of innumerable structures, while DFT calculations allowed us to understand the fundamental interactions between nucleobases and surfaces.

Bibliography

[1] A. Warshel and M. Levitt. *Journal of molecular biology*, 103, 1976.

[2] *http* : *//www.nobelprize.org/nobel_prizes/chemistry/laureates/*2013/*advancedchemistryprize*2013.*pdf*

[3] M. Levitt and A. Warshel. *Nature*, 253, 1975.

[4] M. Sarikaya, C. Tamerler, A. K.-Y. Jen, K. K. Schulten, and F. Baneyx. *Nature Mat.*, 3, 2003.

[5] L. Addadi and S. Weiner. *Angew. Chemie Int Ed*, 31(2), 1992.

[6] A. M. Belcher, X. C. Wu, R. J. Christensen, P. K Hansma, G. D. Stucky, and D. E. Morse. *Nature*, 381, 1996.

[7] K. V. Ewart, Q. Lin, and C. L. Hew. *Cell. Mol. Life Sci.*, 55, 1999.

[8] Y. C. Liou, A. Tocilj, P. L. Davies, and Z. Jia. *Nature*, 206, 2000.

[9] D. J. Juliano, S. S. Saavedra, and G. A. Truskey. *J. Biomed. Mater. Res.*, 27, 1993.

[10] D. Porath G. Cuniberti and R. Di Felice. *Top. Curr. Chem.*, 237, 2004.

[11] R. G. Endres, D. L. Cox, and R. R. P. Singh. *Rev. Mod. Phys*, 76, 2004.

[12] S. S. Mallajosyula and S. K. Pati. *J. Phys. Chem. Lett.*, 1, 2010.

[13] J. G. Genereux and J. K. Barton. *Chem. Rev.*, 110, 2010.

[14] E. R. Mardis. *Nature*, 470, 2011.

[15] M. Choi, U. I. Scholl, W. Jia, T. Liu, I. R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloglu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane, and R. P. Lifton. *PNAS*, 106, 2009.

[16] S. A. Schreiner, A. L. Hatch, D. F. Shudy, D. R. Howard, C. Howell, J. Zhao, P. Koelsch, M. Zharnikov, D. Y. Petrovykh, and A. Opdahl. *Anal. Chem.*, 83, 2011.

[17] E. Shapir, A. Calzolari, C. Cavzzoni, G. Cuniberti, A. B. Kotlyar, R. Di Felice, and D. Porath. *Nature Mat.*, 7, 2008.

[18] P. J. de Pablo, F. Moreno-Herrero, J. Colchero, J. Gó mez Herrero, P. Herrero, A. M. Baró, P.Ordejó, J. M. Soler, and E. Artacho. *Phys. Rev. Lett.*, 85, 2000.

[19] A. B. Kotlyar, N. Borovok, T. Molotsky, H. Cohen, E. Shapir, and D. Porath. *Adv. Mater.*, 17, 2005.

[20] V. Dagget and M. Levitt. *J. Mol. Biol.*, 232, 1993.

[21] J. Tirado-Rivers, M. Orozco, and W. L. Jorgensen. *Biochem*, 36, 1997.

[22] R. D: Toofanny and V. Daggett. *WIREs Comput. Mol. Sci.*, 2, 2012.

[23] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. 1987.

[24] W. F. van Gusteren and H. J. C. Berendsen. *Angew. Chemie Int Ed*, 29, 1990.

[25] H. Heinz, R. A. Vaia, B. L. Farmer, and R. R. Naik. *J. Phys. Chem. C*, 112, 2008.

[26] N. Sandig and F. Zerbetto. *Chem. Commun.*, 46, 2010.

[27] S. rapino and F. Zerbetto. *Langmuir*, 21, 2005.

[28] H. Heinz, B. L. Farmer, R. B. Pandey, J. M. Slocik, S. S. Patnaik, R. Pachter, and R. R. Naik. *J. Am. Chem. Soc.*, 131, 2009.

[29] H. Heinz, K. C Jha, J. Luettmer-Strathmann, B. L. Farmer, and R. R. Naik. *J. R. Soc.,Interface*, 8, 2010.

[30] J. Yu, M. L. Becker, and G. A. Carri. *Langmuir*, 28, 2012.

[31] J. Yu, M. L. Becker, and G. A. Carri. *Small*, 6, 2010.

[32] J. Feng, J. M. Slocik, R. R. Naik, B. L. Farmer, and H. Heinz. *Small*, 8, 2012.

[33] F. Ercolessi, M. Parrinello, and E. Tosatti. *Philos. Mag. A.*, 58, 1988.

[34] F. Iori, R. Di Felice, E. Molinari, and S. Corni. *J. Comp. Chem.*, 30, 2009.

[35] L. B. Wright, P. M. Rodger, S. Corni, and T. Walsh. *J. Chem. Theo. Comput*, 9, 2013.

[36] R. Di Felice and S. Corni. *J. Phys. Chem. Lett.*, 2, 2011.

[37] F. Iori, R. Di Felice, and S. Corni. *J. Phys. Chem. Lett*, 2, 2011.

[38] M. Born and R. Oppenheimer. *Ann. Phys. (Leipzig)*, 84, 1927.

[39] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136, 1964.

[40] W. Kohn and L. Sham. *Phys. Rev.*, 140, 1965.

[41] J. P. Perdew, J. A. Chevary adn S. H. Vosko adn K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. *Phys. Rev. B*, 46, 1992.

[42] A. D. Becke. *J. Chem. Phys.*, 96, 1992.

[43] E. I. Proynov, E. Ruiz, A. Vela, , and D. R. Salahub. *Int. J. Quantum Chem.*, 61, 1995.

[44] B. Hammer, K. W. Jacobsen, and J. K. Nørskov. *Phys. Rev. Lett.*, 70, 1993.

[45] J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77, 1996.

[46] Y. Zhang and W. Yang. *Phys. Rev. Lett.*, 80, 1998.

[47] S. Piana and A. Billic. *J. Phys. Chem. B*, 110, 2006.

[48] M. Dion, H. Rydberg, E. Schroder, D. C. Langreth, and B. I. Lundqvist. *Phys. Rev. Lett.*, 92, 2004.

[49] T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard, and D. C. Langreth. *Phys. Rev. B*, 76, 2007.

[50] D. C. Langreth, M. Dion, H. Rydberg, E. Schroder P.Hyldgaard, and B. I. Lundqvist. *Int. J. Quantum Chem.*, 101, 2005.

[51] A. Puzder, M. Dion, and D. C. Langreth. *J. Chem. Phys.*, 124, 2006.

[52] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth. *Phys. Rev. B*, 82, 2010.

[53] V. R. Cooper. *Phys. Rev. B*, 81, 2010.

[54] X. Wu, M. C. Vargas, S. Nayak, V. Lotrich, , and G. Scoles. *J. Chem. Phys*, 115, 2001.

[55] J. Klimes, D. R. Bowder, and A. Michaelides. *J. Phys.:Condens. Matter*, 22, 2010.

[56] W. Liu, J. Carrasco, B. Santra, A. Michaelides, M. Scheffler, and A. Tkatchenko. *Phys. Rev. B*, 86, 2012.

[57] J. Wellendorff, A. Kelkkanen, J. J. Mortensen, B. I. Lundqvist, and T. Bligaard. *Top. Catal.*, 53, 2010.

[58] H. A. Lorentz. *Annalen der Physik*, 12, 1881.

[59] D. Berthelot. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 126, 1898.

[60] E. J. Sorin and V. S. Pande. *Biophysical Journal*, 88, 2005.

[61] A. J. DePaul, E. J. Thompson, S. S. Patel, K. Haldeman, and E. J. Sorin. *Nucleic Acids Research*, 38, 2010.

[62] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. *J. Am. Chem. Soc.*, 117, 1995.

[63] M. Cavallari, A. Calzolari, A. Garbesi, and R. Di Felice. *J. Phys. Chem. B*, 110, 2006.

[64] F. Iori and S. Corni. *J. Comp. Chem.*, 29, 2008.

[65] M. Rosa, S. Corni, and R. Di Felice. *J. Phys. Chem. C*, 116, 2012.

[66] M. Rosa, S. Corni, and R. Di Felice. *J. Chem. Theo. and Comp.*, 9(10), 2013.

[67] T. Darden, D. York, and L. Pedersen. *J Chem Phys*, 98, 1993.

[68] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen. *J Chem Phys*, 103, 1995.

[69] R.W. Hockney and J.W. Eastwood. *Computer simulation using par- ticles.* Institute of Physics, 2992.

[70] F. Smith and D. Frenkel. *Understanding molecular simulations.* Wiley, 2000.

[71] D. B. Kokh, S. Corni, P. J. Winn, M. Hoefling, K. E. Gottschalk, and R. C. Wade. *J. Chem. Theory Comp.*, 6, 2010.

[72] M. Feig and C. L. Brooks. *Curr. Opin. Struct. Biol.*, 14, 2004.

[73] R. A. Latour. *Biointerphases*, 3, 2008.

[74] Y. Sun and R. Latour. *J. Comput. Chem.*, 27, 2006.

[75] A. V. Verde, J. M. Acres, and J. K. Maranas. *Biomacromolecules*, 10, 2009.

[76] A. Einstein. *Ann. Phys.*, 17, 1905.

[77] M. V. Smoluchowsky. *Ann. Phys.*, 21, 1906.

[78] D. L. Ermak and J. A. McCammon. *J. Chem. Phys.*, 69, 1978.

[79] R. R. Gabdoulline and R. C. Wade. *J. Phys. Chem.*, 100, 1996.

[80] A. H. Elcock, R. R. Gabdoulline, R. C. Wade, and J. A. McCammon. *J. Molecular Biology*, 291, 1999.

[81] J. Antony and S. Grimme. *Phys. Chem. Chem. Phys.*, 8, 2006.

[82] A. D. Becke and E. R. Johnson. *J. Chem. Phys.*, 127, 2007.

[83] A. Tkatchenko and M. Scheffler. *Phys. Rev. Lett.*, 102, 2009.

[84] S. D. Chakarova-Kack, E. Schroder, B. I. Lundqvist, and D. C. Langreth. *Phys. Rev. Lett.*, 96, 2006.

[85] T. S. Chwee and M. B. Sullivan. *J. Chem. Phys.*, 137, 2012.

[86] P. Giannozzi et al. *Condens. Matter*, 21, 2009.

[87] R. E. A. Kelly, W. Xu, M. Lukas, R. Otero, M. Mura, Y-J Lee, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, and F. Besenbacher. *Small*, 4, 2008.

[88] M. Lukas, R. E. A. Kelly, L. N. Kantorovich, R. Otero, W. Xu, E. Laegsgaard, I. Stensgaard, and F. Besenbacher. *J. Chem. Phys.*, 130, 2009.

[89] W. Xu, R. E. A. Kelly, G. Henkjan, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, and F. Besenbacher. *Small*, 5, 2009.

[90] W. Xu, R. E. A. Kelly, R. Otero, M. Schockand, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, and F. Besenbacher. *Small*, 3, 2009.

[91] R. E. A. Kelly, M. Lukas, L. N. Kantorovich, R. Otero, W. Xu, M. Mura, E. Laegsgaard, I. Stensgaard, and F. Besenbacher. *J. Chem. Phys*, 193, 2008.

[92] R. Otero, M. Lukas, R. E. A. Kelly, W. Xu, E. Laegsgaard, I. Stensgaard, N. Kantorovich, and F. Besenbacher. *Science*, 319, 2008.

[93] B. Hammer and J. K. Norkov. *Theory of Adorption and Surface Reactions*. Kluwer Academic Publishers, The Netherlands, pacchioni, g., eds. edition, 1997.

[94] M. Ostblom, B. Liedberg, L. M. Demers, and C. A. Mirkin. *J. Phys. Chem. B*, 109, 2005.

[95] A. Ferretti, C. Baldacchini, A. Calzolari, R. Di Felice, A. Ruini, E. Molinari, and M. G. Betti. *Phys. Rev. Lett.*, 99, 2007.

[96] N. Atodiresei, V. Caciuc, P. Lazic, and S. Blugel. *Phys. Rev. Lett.*, 102, 2009.

[97] E. Palecek. *Talanta*, 56, 2002.