

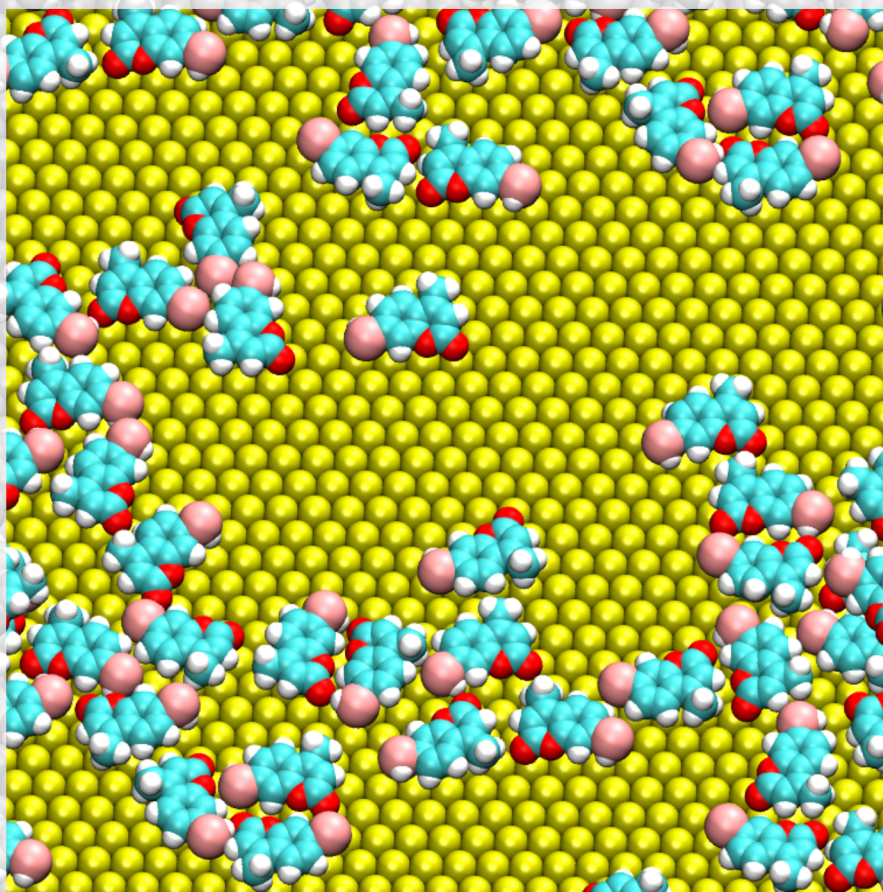


UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE

Università del Piemonte Orientale Amedeo Avogadro
Dipartimento di scienze e innovazione tecnologica (DISIT)
Ph.D. Program in Chemistry & Biology
XXXIV cycle; 2018-2021

Multiscale modelling of organic molecules interacting with solids

SSD: CHIM/02



Davide Marchi

Supervisor: Prof. M. Cossi
Ph.D. Program Coordinator: Prof. G. C. Tron



UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE

Università del Piemonte Orientale Amedeo Avogadro
Dipartimento di scienze e innovazione tecnologica (DISIT)
Ph.D. Program in Chemistry & Biology
XXXIV cycle; 2018-2021

Multiscale modelling of organic molecules interacting with solids

SSD: CHIM/02

Ph.D. Candidate:
Davide Marchi

Supervisor:
Prof. Maurizio Cossi

Ph.D. Program Coordinator: Prof. G. C. Tron



UNIVERSITÀ DEL PIEMONTE ORIENTALE
DOTTORATO DI RICERCA
IN CHEMISTRY & BIOLOGY

Via Duomo, 6
13100 – Vercelli (ITALY)

DECLARATION AND AUTHORISATION TO ANTIPLAGIARISM DETECTION


The undersigned **Marchi Davide** student of the Chemistry & Biology
Ph.D course (.xxxiv Cycle)

declares:

- to be aware that the University has adopted a web-based service to detect plagiarism through a software system called “Turnit.in”,
- his/her Ph.D. thesis was submitted to Turnit.in scan and reasonably it resulted an original document, which correctly cites the literature;

acknowledges:

- his/her Ph.D. thesis can be verified by his/her Ph.D. tutor and/or Ph.D Coordinator in order to confirm its originality.

Date: **2022-02-15** Signature: 

Contents

1	Introduction: theoretical background	8
1.1	Computer simulations in science	8
1.2	Computational chemistry	12
1.2.1	Classical and <i>ab initio</i> models	12
1.2.2	Computable quantities	13
1.3	Basics of thermodynamics and statistical mechanics . .	15
1.3.1	Classical thermodynamics	15
1.3.2	Statistical mechanics	18
1.4	Simulations of many-body systems	24
1.4.1	Statistical mechanics and molecular simulations .	24
1.4.2	Force fields	27
1.4.3	The basic molecular dynamics algorithm	29
1.4.4	Free energy calculations	31
1.5	Ab initio methods	33
1.5.1	Notes on notation	33
1.5.2	The Hartree-Fock method	34
1.5.3	Density Functional Theory	38
1.5.3.1	From the Thomas-Fermi model to the Hohenber-Kohn theorems	38
1.5.3.2	Kohn-Sham DFT	40
1.5.3.3	Exchange-correlation and non-covalent interactions	42
2	Thesis outline	48
3	Force field parametrization of Au-X interactions from ab initio calculations	51
3.1	UFF_Au	51
3.2	GROMOS_Au1	54
3.3	GROMOS_Au2	58
3.3.1	Obtaining FF parameters	58
3.3.2	Testing FF parameters on target systems	63

4	Computational study of MPA and MMC thiol monolayers on Au(111)	66
4.1	DFT free energy calculations of S-H bond dissociation . . .	69
4.2	Molecular dynamics equilibration of the monolayers	71
4.3	BAR free energy calculations from MD trajectories . . .	75
4.4	Resulting chemical potentials	77
5	Towards a traceable enhancement factor in surface-enhanced Raman spectroscopy	82
6	Structure and Stability of 7-mercapto-4-methylcoumarin SAM on Gold: an Experimental and Computational analysis	90
7	Discussion and future perspectives	127
	Bibliography	135
	Appendices	145
A	Optimizing the relaxivity at high fields: systematic variation of the rotational dynamics in polynuclear Gd-complexes based on the AAZTA ligand	146
B	Hyper-Cross-Linked Polymers for the Capture of Aromatic Volatile Compounds	161
C	Predicting the Conformation of Organic Catalysts Grafted on Silica Surfaces with Different Numbers of Tethering Chains: The Silicopodality Concept	174
D	FFfitpy script	187
	List of publications	193

In theory, practice and theory are the same.
In practice, they are not.

ANONYM

Chapter 1

Introduction: theoretical background

1.1 Computer simulations in science

The first application of computer simulations in science dates back to the mid-1940s, when Stanislaw Ulam, John von Neumann, and others, applied the Monte Carlo method on electronic computers to aid in the design of the hydrogen bomb. Their goal was to solve certain analytically unsolvable problems in neutron diffusion. [1] After the end of World War II, the dramatic improvement in computing power and availability set the stage for a parallel development in computer simulations. Today most of the natural sciences, as well as many social sciences, make extensive use of simulations.

A computer simulation can be defined as a program, running on a computer, that performs a specific algorithm in order to produce numerical data of interest. Simulations can be equation-based, meaning that they are governed by global equations (as is common in physical sciences), or agent-based, if they represent the behavior of n discrete individuals who follow local “rules of evolution” [2].

More insights can be gained by analyzing Winsberg’s definition of what simulation *studies* are: “Successful simulation studies do more than compute numbers. They make use of a variety of techniques to draw inferences from these numbers. Simulations make creative use of calculational techniques that can only be motivated extra-mathematically and extra-theoretically. As such, unlike simple computations that can be carried out on a computer, the results of simulations are not automatically reliable.” [3]

This definition contains many points which are going to be expanded upon here, but first it’s necessary to examine for what purposes simulations are used in science. A simulation can be used [2, 4]:

1. As an explanation. If a simulation is capable of reproducing a

particular aspect of the real-world system of interest, then it can be used as an explanation of the phenomenon. Given that simulations are always simplified models of reality, a simulation will be a good explanation if all the relevant features of the real-world system are included in the model. In this case, the simulation can be probed for further understanding of the mechanism giving rise to the data of interest.

2. As a prediction. It is often the case that simulations are applied when observable data is sparse, hard or impossible to obtain. Simulations can therefore complement experimental data where it lacks; they can be used to predict the future, retrodict the past, or compute the properties of systems that are not experimentally known.
3. For heuristic exploration. A simulation provides a virtual world whose structure and hidden workings can be explored in much greater detail than it would be possible in the real world. Furthermore, Frenkel and Smit [5] emphasize how simulations can be viewed as tools giving *exact* results of the application of a theoretical model. In other words, they provide an answer to the question: “What would the real world look like, if it were described perfectly by this certain theoretical model?” The heuristic advantages that simulation have over experiments, therefore, extend to changes in the description chosen for the system, as well as hypothetical interventions made upon it.

A key feature of simulations that grants them explanatory and exploratory power is *emergence*. Much has been written about emergence [6]; here the term is used loosely to signify that simulations share the unpredictability of real experiments. They must be run in order to get a result, which couldn't be predicted before the simulation was run. This is known as *epistemic opacity* [7]. A prominent example of emergent behavior in computer simulations can be found in the works of Alder and Wainwright [8] and Wood and Jacobson [9], who showed that a model composed of a periodic box of spheres having “hard” short-range repulsive forces and no attractive forces was capable of undergoing a phase transition from liquid to crystal, thus indicating the predominant role of repulsive forces in such phenomena.

The discussion up to this point should have clarified most of Winsberg's definition of what successful simulation studies are. In particular, the emergent behavior of simulations, paired with the theoretical and computational restrictions that force simulations to be *simplified* representations of the real world, explain Winsberg's statement about

simulations being motivated only “extra-mathematically and extra-theoretically”. Simulations are informed by our scientific theories about the real world, but are essentially distinct from it as well as from the theories themselves.

The crucial question still left unanswered is: how, then, can we trust simulations to give useful and reliable information about the real world? Parker [10] formalized the five strategies that can be used to increase the researcher’s rational confidence in the results of simulation. These strategies are divided in *model evaluation* and *code evaluation*, and are presented in analogy with Franklin’s [11] strategies for experimental results. These strategies are applied every day in the practitioner’s work, and are reported below in Table 1.1.

Parker’s model evaluation strategies are generally termed *validation*, and consist of evaluating the soundness of the simulation from the standpoint of the model itself, independently of the computer code performing the simulation. Code evaluation strategies, on the other hand, can also be called *verification* [2].

In the scientific literature, the results of simulations are commonly corroborated by the comparison with experimental data or justified on the basis of well-understood physical theories. This obviously depends on whether experimental data is available (as already stated, simulations are particularly useful precisely when experimental data is scarce) or whether an underlying mathematical theory exists for the system of interest. In practice, there is often no single agreed-upon method for the validation of simulations, and simulation techniques gain credentials over time in a similar way as experimental methods do, by showing applicability and robustness in different studies of the system of interest, or other systems similar to it.

It’s also reasonable to hold the computational results to different standards depending on their intended use. For example, are simulations used to explain the qualitative behavior of the system, or for a quantitative prediction of a certain property? The second instance would warrant stricter methods of validation.

The relationship between experiments and simulations is further complicated by aspects of simulations that are unique to them and do not have a correspondence in the real world. Most notably, simulations can make use of unphysical features, sometimes called *fictions*, that help in understanding a system despite being inconsistent with the accepted physical theory underlying it. Smit *et al.* [12], for example, computed the phase diagrams of *n*-alkanes at temperatures where these molecules are thermally unstable, and therefore impossible to investigate experimentally. They did so by leveraging the fact that, while real chemical bonds can become unstable and break at high temperatures, simulated

ones can be modeled so that they never dissociate.

Franklin's strategies for experiments	Model evaluation strategies	Code evaluation strategies
(1) Apparatus gives other results that match known results	Simulation output fits closely enough with various observational data	Estimated solutions fit closely enough with analytic and/or other numerical solutions
(2) Apparatus responds as expected after intervention on the experimental system	Simulation results change as expected after intervention on substantive model parameters	Solutions change as expected after intervention on algorithm parameters
(3) Capacities of apparatus are underwritten by well confirmed theory	Simulation model is constructed using well-confirmed theoretical assumptions	Solution method is underwritten by sound mathematical theorizing and analysis
(4) Experimental results are replicated in other experiments	Simulation results are reproduced in other simulations or in traditional experiments	Solutions are produced using other pieces of code
(5) Plausible sources of significant experimental error can be ruled out	Plausible sources of significant modelling error can be ruled out	Plausible sources of significant mathematical/computational error can be ruled out

Table 1.1: Parker's proposed strategies to increase the confidence in the results of simulations, in analogy with Franklin's strategies for experimental results. Taken from ref [10].

1.2 Computational chemistry

This PhD thesis concerns the application of methods of computational chemistry to systems of interest in the scientific literature. The term *computational chemistry* will be used here to signify any computer-based approach to calculate quantities related to chemical problems, including the development or improvement of such methods. Many other related terms exist and, although any clear distinction is somewhat arbitrary, a cursory definition of their relationship will be given for clarity. One might use the term *molecular modeling* in order to emphasize the computational chemist’s task of trying to obtain a *model*, or an idealized representation, of a molecular system. The term *simulation* can be preferred when the computation is meant to describe the evolution of a chemical system through time. Finally, *theoretical chemistry* is the discipline of obtaining new theories and/or models with improved performances over the existing ones [13].

1.2.1 Classical and *ab initio* models

Physics, chemistry, and biology can be thought of as three natural sciences studying the physical universe at different scales, respectively from the smaller constituents of matter to living organisms. (The need for different disciplines characterized by different practices is itself a product of *emergence*, see Section 1.1.)

At the “chemistry scale” of the universe, Nature is described in terms of atoms and molecules. The description chosen for the unit of matter in chemical terms, the atom, can be regarded as one fundamental distinction among the methods of computational chemistry. On one hand, atoms can be modeled as hard spheres with a certain mass, characterized by harsh repulsive forces when two atoms move closer than the sum of their radii. Molecular structure is defined in terms of bonds, angles, dihedral angles, and improper dihedral angles, as will be discussed in greater detail in Subsection 1.4.3. Different functional forms can be used to describe the change in energy associated with changes in molecular structure; for example, the harmonic potential is a common choice for bonds and angles. Coulombic interactions between charges and van der Waals forces are also commonly included in the computation (see Subsection 1.4.3). Models constructed this way are termed *classical*, because they employ classical mechanics to simulate chemical systems, although they still more or less implicitly acknowledge the quantum nature of atoms, e.g. by including van der Waals forces.

On the other hand, models can be built *ab initio* (from first principles) by using the theory of quantum mechanics. In this case, the

internal structure of the atom is made explicit; atoms have a nucleus surrounded by electrons. All *ab initio* models make use of the Born-Oppenheimer approximation, which is the assumption that the wave functions describing atomic nuclei and electrons can be treated separately. This can be justified on the basis of the difference in mass between nuclei and electrons; nuclei being several order of magnitudes heavier, their motion is much slower and hence can be considered null from the point of view of the electrons. For a certain position of atomic nuclei, then, the full state of the system is obtained by finding the electronic wave function. This is done by solving the electronic time-independent Schrödinger equation:

$$H |\Psi_{el}\rangle = E |\Psi_{el}\rangle \quad (1.1)$$

This is an eigenvalue equation, where H is the Hamiltonian operator and E is the energy of the system.

Section 1.5 will explain the basics of the methods used to find a solution to Equation 1.1, as well as an alternative approach to electronic structure that makes use of the electron density, rather than the wave function, and is known as *density functional theory* (DFT). It is worth noting that even in *ab initio* methods several approximations must be made in order to have equations that can be solved, which is done in a numerical, iterative way.

The methods based on a quantum mechanical description, despite obviously containing less approximations, require much more computational resources compared to the classical ones, therefore limiting their application to smaller systems.

1.2.2 Computable quantities

Very rarely real-world experiments have a molecular or atomic resolution in the quantification of the physical properties they are meant to measure. Even when that's the case, experiments are almost always concerned with *large* collections of molecules. Therefore the physical observables that can be measured are either the product of collective behavior (like temperature and pressure) or averages of molecular properties over a population of molecules, each one in a different state.

In order to compute physical properties, therefore, it is apparent that a theoretical link is needed between the atomic description of the system adopted by the computational chemist, and the macroscopic or averaged properties the experimental chemist can measure. That link is provided by thermodynamics, and in particular by statistical mechanics, which emerged alongside atomic theories in the second part of the nineteenth century, ultimately explaining classical thermodynamics on

the basis of collections of atom and molecules obeying statistical laws. Due to the central importance of such theoretical framework, the basics of statistical mechanics will be illustrated in Subsection 1.3.2.

Given this, the fundamental property that must be known in order to compute useful numbers is *structure*. Molecular structure is most often formulated in terms of Potential Energy Surface (PES). Figure 1.1 gives an intuitive representation of the PES.

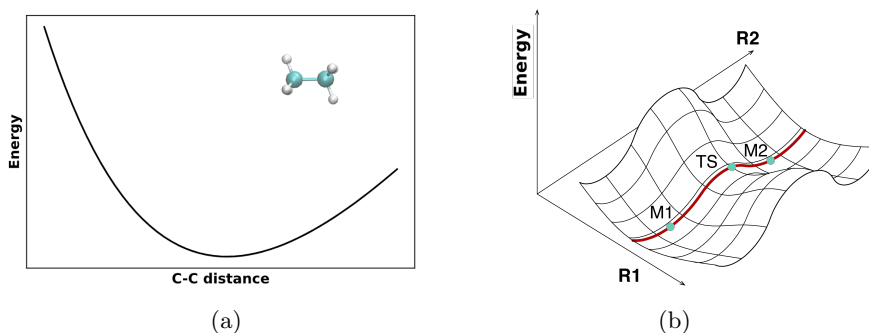


Figure 1.1: (a) A monodimensional slice of PES for an ethane molecule. The energy is represented as a function of only one coordinate: the carbon-carbon distance. (b) A bidimensional slice of PES. M1 and M2 are two minima connected by a saddle point TS.

Note how for a molecule composed of N atoms the real dimensionality of the PES is $3N - 6$, where the term 3 comes from the number of spatial dimensions and the term 6 accounts for the invariance of the energy with respect to translations and rotations of the whole structure. (For linear molecules, the energy associated with rotation along the molecular axis is negligible, and hence the degrees of freedom are $3N - 5$.)

Thus, fully determining the PES could conceivably be the first goal to pursue, and finding efficient algorithms for exploring the PES is indeed a central problem in computational chemistry [14,15]. However, full knowledge of the PES can rarely be achieved in practice, and care must be taken in the evaluation of what constitutes sufficient knowledge for the system of interest.

Knowledge of the PES is more useful in some regions than others. In particular, local minima correspond to the most stable, and hence more populated, states. The other interesting points on the PES are saddle points, see Figure 1.1, which, being the lowest energy paths between adjacent minima, correspond to the chemical concept of *transition state*.

Assuming sufficient knowledge of the PES for the system under investigation, Cramer proposes a classification of computable properties into three categories [13].

Firstly, there are properties that depend on the structure of indi-

vidual molecules, such as nuclear magnetic resonance (NMR) chemical shifts and coupling constants, electron paramagnetic resonance (EPR) hyperfine coupling constants, absorption maxima for rotational, vibrational, and electronic spectra, electron affinities, and ionization potentials.

Another class of computable quantities is formed by the physical observables arising from the collective behavior of ensembles of molecules. Notable examples of such thermodynamic quantities that are often of interest for the computational chemist are free energies of solvation, heats of formation, complexation energies, and pK_a values.

Finally, Cramer emphasizes the existence of quantities that are computable, such as partial atomic charge, and can be useful to gain insights into the system, and yet don't have an experimental counterpart.

All the quantities mentioned up to this point do not contain an explicit dependence on time. Molecular dynamics (MD) is a computational method that allows for a description of the time evolution of a chemical system. MD simulations can therefore be used to compute other properties related to time-dependent processes, such as diffusion coefficients [16]; they are also commonly used to obtain ensemble thermodynamic properties as well as to explore the PES of the system [17, 18]. A more detailed description of the MD method will be given in Subsection 1.4.3.

1.3 Basics of thermodynamics and statistical mechanics

As stated above in Subsection 1.2.2, thermodynamics is the branch of physics providing a theoretical background to compare the results of simulations with experiments. Therefore in this section the fundamental equations of thermodynamics will be introduced together with some historical context; then it will be shown how thermodynamics can be derived mathematically by a statistical approach applied to large collections of particles.

1.3.1 Classical thermodynamics

The birth of thermodynamics as a science is commonly attributed to Nicolas Léonard Sadi Carnot, a mechanical engineer and military scientist in the French army. In 1824 Carnot published his book *Reflections on the Motive Power of Fire*, a treatise whose aim was to study the physical principles underlying the workings of steam engines in order to increase their efficiency and establish the theoretical limit of such improvements. But Carnot's goal was even broader; in his own words,

“In order to consider in the most general way the principle of the production of motion by heat, it must be considered independently of any mechanism or any particular agent. It is necessary to establish principles applicable not only to steam-engines but to all imaginable heat-engines, whatever the working substance and whatever the method by which it is operated.”

Despite lacking mathematical rigor and being based on the now obsolete caloric theory, which considers heat as a self-repellent fluid, Carnot’s book laid the bases of thermodynamics by providing some key insights about the physical laws governing heat. In particular, Carnot noted how machines do work by exploiting the *flow* of heat from a hotter to a colder source. This led him to introduce the concept of thermodynamic cycles and establish a limit for the efficiency of conversion of heat into work; such upper limit can only be reached by a theoretical machine performing an ideal thermodynamic cycle, today called *Carnot cycle*, and is dependent only on the temperature of the hot and the cold reservoirs according to Equation 1.2:

$$\eta = 1 - \frac{T_{cold}}{T_{hot}} \quad (1.2)$$

(Note: this equation was never expressed mathematically by Carnot.)

The Carnot cycle represents a *reversible* thermodynamic process, meaning a process in which all heat is transformed into work and none is lost to the system’s surroundings; at the end of one cycle both the system and its surroundings will return to their original states.

In the following years, these ideas were further developed by scientist such as Benoît Paul Émile Clapeyron, William Thomson (Lord Kelvin), and especially Rudolf Clausius [19]. In 1850, Clausius introduced a function U that related the heat exchanged by a system, Q , with the work done, W :

$$dU = dQ - dW \quad (1.3)$$

Later he began calling U *energy*, and Equation 1.3 became known as the first law of thermodynamics. In 1854, by extending Carnot’s theoretical analysis of thermodynamic cycles to irreversible (that is, real) processes, Clausius showed that:

$$\int \frac{dQ}{T} \leq 0 \quad (1.4)$$

where the equality holds for reversible cycles.

In 1865 he introduced a function he called *entropy*, to which he assigned the symbol S :

$$S = \frac{dQ}{T} \text{ (reversible processes)} \quad (1.5)$$

By recognizing that entropy is a state function, that is a function that depends only on the current state of the system and not on the path taken to reach it, the study of heat could be extended beyond cyclic processes. For a system undergoing a change from state 1 to state 2:

$$S_2 - S_1 \geq \int_1^2 \frac{dQ}{T} \quad (1.6)$$

where the equality holds for reversible processes. Note that for an adiabatic process $Q = 0$, and $\Delta S \geq 0$. Equation 1.6 is thus a statement of the second law of thermodynamics.

Clausius concluded the same 1865 paper with his enunciation of the first and second laws of thermodynamics in the form:

1. The energy of the universe is constant.
2. The entropy of the universe tends to a maximum.

Equation 1.3 and Equation 1.5 can be used to derive the fundamental thermodynamic relation, a set of four differential equations that define the relationships among all thermodynamic quantities:

$$dU(S, V, N_i) = TdS - pdV + \sum_i \mu_i dN_i \quad (1.7a)$$

$$dH(S, p, N_i) = TdS + Vdp + \sum_i \mu_i dN_i \quad (1.7b)$$

$$dF(T, V, N_i) = -SdT - pdV + \sum_i \mu_i dN_i \quad (1.7c)$$

$$dG(T, p, N_i) = -SdT + Vdp + \sum_i \mu_i dN_i \quad (1.7d)$$

where μ_i is the chemical potential of the particles composing the system, N is the number of particles, V is volume, p is pressure, S is entropy, and T is temperature. The *thermodynamic potentials* U , H , F , and G are called respectively internal energy, enthalpy, Helmholtz free energy, and Gibbs free energy. It is often useful to express all other thermodynamic potentials with respect to the internal energy:

$$H = U + pV \quad (1.8a)$$

$$F = U - TS \quad (1.8b)$$

$$G = U + pV - TS \quad (1.8c)$$

The works of Clausius and others formed the bedrock of classical thermodynamics, where the term *classical* is used to signify that its

equations derive from a macroscopic treatment of thermodynamic systems.

In the last part of the 19th century and in the early 20th century, James Clerk Maxwell, Ludwig Eduard Boltzmann, and Josiah Willard Gibbs founded statistical mechanics (also known as statistical thermodynamics), a mathematical framework to describe the collective behavior of large ensembles of particles. This new branch of thermodynamics was fuelled by other concomitant advancements in the atomic description of matter. In the next Subsection it will be shown how an atomic, statistical treatment of physical systems can recover mathematically the thermodynamic equations given above.

1.3.2 Statistical mechanics

The fundamental postulate of statistical mechanics, known as the equal *a priori* probability postulate, states that for a system of known energy, every *microstate* (the specific microscopic configuration of the particles composing the system) compatible with this knowledge is equally probable. This postulate acknowledges that there's a limit to the information we can obtain about a system, and many possible microscopic arrangements correspond to the same macroscopic properties observed. To deal with this, statistical mechanics makes use of the concept of *ensemble*. An ensemble is an infinitely large set of “mental copies” of the system of interest; since every copy is under the same physical constraints, the ensemble represents all the possible microstates for the system. The assumption is that the long-time average of every physical property f measured on a single copy in the ensemble (that is, the result we expect from a measurement) will be equal to the average of f over the ensemble.

To understand how thermodynamic equations can emerge from such a statistical framework, let's consider a system we divide in N subsystems. The system has a total energy E that is shared among the subsystems, which can exchange energy among them. According to the definitions given above, the system can be considered as an ensemble of subsystems that are subject to the constraints of fixed composition, volume, and temperature (if we consider a single subsystem, all the others act as a heat bath of infinite heat capacity). This is called a *canonical* ensemble.

Every subsystem occupies a certain state i having energy E_i ; the generic n_i will give the number of subsystems occupying state i . The probability P_i for a subsystem to be in state i is by definition:

$$P_i = \frac{n_i}{N} \quad (1.9)$$

Expressed mathematically, the constraints on the system are:

$$\begin{cases} \sum_i n_i = N \\ \sum_i n_i E_i = E = NU \end{cases} \quad (1.10)$$

where U is the average energy of a subsystem.

There will be several configurations of the system that satisfy Equation 1.10. Each possible configuration will be described by a set of numbers $\{n_i\}$; on the other hand, a set $\{n_i\}$ might describe more than one configuration, obtained by “redistributing” the subsystems among the states $\{i\}$. Table 1.2 gives a simple example to clarify this point.

For a set $\{n_i\}$, the number of ways of redistributing the subsystems is given by:

$$C = \frac{N!}{\prod_i n_i!} \quad (1.11)$$

So, in order to find the most probable set $\{n_i\}$, we seek to maximize this quantity.

We can choose to maximize C by maximizing its logarithm instead. Since N is in general a very big number we can invoke Stirling’s approximation:

$$\log N! = N \log N - N \quad (1.12)$$

and, using Equation 1.9, write:

$$\begin{aligned} \log C &= N \log N - \sum_i n_i \log n_i \\ &= N \log N - \sum_i P_i N \log P_i N \\ &= N \log N - \sum_i P_i N (\log P_i + \log N) \\ &= N \log N - N \sum_i P_i \log P_i - N \log N \sum_i P_i \\ &= -N \sum_i P_i \log P_i \end{aligned} \quad (1.13)$$

We can view the right side of the result in Equation 1.13 as the constant N , the total number of subsystems, multiplying a function of the probabilities P_i . Such function is the statistical definition of entropy:

$$S = - \sum_i P_i \log P_i \quad (1.14)$$

While it still must be shown that S as defined in Equation 1.14 does in fact correspond to the classical definition of entropy, satisfying

Outcomes	$\{n_i\}$	Number of equivalent outcomes: $C = \frac{N!}{\prod_i n_i!}$
HHHH	$n_H = 4, n_T = 0$	1
HHHT \vee HHTH \vee HTHH \vee THHH	$n_H = 3, n_T = 1$	4
HHTT \vee TTHH \vee HTHT \vee THTH \vee HTTH \vee THHT	$n_H = 2, n_T = 2$	6
HTTT \vee THTT \vee TTHT \vee TTTH	$n_H = 1, n_T = 3$	4
TTTT	$n_H = 0, n_T = 4$	1

Table 1.2: A set of 4 coins can be considered as a system composed of 4 subsystems (the individual coins), where each one can exist in two different states: heads (H) and tails (T). Here, without constraints on the energy (or indeed a *definition* of energy) the two states are equally probable. Some sets $\{n_i\}$ will be more probable than others just by virtue of the fact that there exist more ways of obtaining them. In particular, here, $\{n_H = 2, n_T = 2\}$ is the most probable set, with 6 ways of redistributing the subsystems. This is just another way of stating the common-sense fact that if we toss 4 coins the most probable outcome will be 2 heads and 2 tails. Note that the same example can be formulated in terms of 4 consecutive tosses of a single coin, rather than 4 “copies” all at once.

the usual relationships with the other thermodynamic quantities, it is interesting to note that the stated problem of finding the most probable set $\{n_i\}$ has reduced to the problem of maximizing the function S . In other words, the most probable configuration adopted by the system is the one that maximizes S , the entropy. This, in a nutshell, is the microscopic explanation (mainly due to Boltzmann) of the concept of

entropy and the origin of the second law of thermodynamics.

We can now reformulate the problem in the language of Lagrange multipliers. We want to maximize S , a function of the probabilities P_i as shown in Equation 1.14, under the two constraints already written in Equation 1.10:

1. We can equivalently rewrite the first constraint in terms of probabilities, rather than the number of subsystems:

$$G_1 = \sum_i P_i - 1 = 0 \quad (1.15)$$

2. While the second can be written as:

$$G_2 = \sum_i P_i E_i = U \quad (1.16)$$

Let's now apply the method of Lagrange multipliers:

$$F = S + \alpha \left[\sum_i P_i - 1 \right] + \beta \left[\sum_i E_i P_i - U \right] \quad (1.17)$$

We solve for P_i by imposing:

$$\frac{\partial F}{\partial P_i} = 0 \quad (1.18)$$

and find:

$$\begin{aligned} \log P_i + 1 + \alpha + \beta E_i &= 0 \\ \log P_i &= -(1 + \alpha) - \beta E_i \\ P_i &= \frac{1}{Z} e^{-\beta E_i} \end{aligned} \quad (1.19)$$

where in the last step we have defined $Z = e^{1+\alpha}$. Finally, we can plug this result into the constraints equations (Equation 1.15 and Equation 1.16), and get:

$$Z = \sum_i e^{-\beta E_i} \quad (1.20)$$

for the first constraint, while for the second:

$$\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = U \quad (1.21)$$

We call Z the *partition function*. Equation 1.21 suggests that by knowing the partition function we would be able to compute U , which, it

bears repeating, is the average value of the energy we would get from a measurement. It can be shown that, in fact, knowledge of the partition function implies knowledge of *every* thermodynamic quantity of the system.

We can obtain an equation for entropy by using the relations found until now:

$$\begin{aligned}
 S &= - \sum_i P_i \log P_i \\
 &= \sum_i \frac{1}{Z} e^{-\beta E_i} [\beta E_i + \log Z] \\
 &= \beta U + \frac{1}{Z} \log Z \sum_i e^{-\beta E_i} \\
 &= \beta U + \log Z
 \end{aligned} \tag{1.22}$$

and by differentiating the result we find:

$$\begin{aligned}
 dS &= \beta dU + U d\beta + \frac{\partial \log Z}{\partial \beta} d\beta \\
 &= \beta dU
 \end{aligned} \tag{1.23}$$

From the thermodynamic definition of temperature, $T \equiv \frac{dU}{dS}$, we get:

$$\beta = \frac{1}{T} \tag{1.24}$$

so we can rewrite Equation 1.22 as:

$$U - TS = -T \log Z \tag{1.25}$$

We have already defined the Helmholtz free energy as $U - TS$, see Equation 1.8, and thus we have found an equation for the Helmholtz free energy as a function of Z :

$$F = -T \log Z \tag{1.26}$$

Equation 1.26 has been obtained for a system with constant volume, temperature, and composition. All the other thermodynamic potentials listed in Equation 1.8 can be obtained by the same principles for ensembles with different constraints.

Some clarifications and details were omitted for clarity of exposition, and will now be given here.

Firstly, a note on units. In the mathematical derivation above, temperature was expressed as having units of energy. This is appropriate,

since temperature is ultimately the average translational kinetic energy of the particles composing the system. To convert this microscopic quantity into the macroscopic, measurable temperature, a universal constant is needed:

$$\beta = 1/T = 1/k_B T_K \quad (1.27)$$

k_B is called *Boltzmann constant* and is equal to $1.380649 \times 10^{-23} \text{ J/K}$. It was first computed by Max Planck [20] and is currently used in the definition of the kelvin unit (symbol K) for the absolute temperature.

Secondly, the partition function was written in terms of the possible microstates of the whole system, but a partition function could be written also in terms of the energy levels occupied by each individual particle. If the particles are interacting weakly enough that the energy of the system is given by the sum of the energies of the particles, the total partition function will be:

$$\begin{aligned} Z &= \sum_i e^{-\beta E_i} \\ &= \sum_i e^{-\beta \sum_{n=1}^N \epsilon_{ni}} \\ &= \sum_i e^{-\beta \epsilon_{1i}} \sum_i e^{-\beta \epsilon_{2i}} \dots \sum_i e^{-\beta \epsilon_{Ni}} \\ &= z^N \end{aligned} \quad (1.28)$$

where ϵ is the energy of a single particle, z is the molecular partition function, and N is the total number of particles. In Equation 1.28 it is assumed that all particles are the same (since their partition function is taken to be equal), but for quantum particles that are *truly* the same, meaning that they are indistinguishable *in principle*, the total partition function should be multiplied by a factor $\frac{1}{N!}$ to avoid over-counting the number of microstates.

The molecular partition function can be further factorized, since the total molecular energy ϵ can be expressed as the sum of electronic, translational, rotational, and vibrational terms:

$$z = z_{el} z_{trans} z_{rot} z_{vib} \quad (1.29)$$

Analytic expressions for the individual terms can be derived, but will not be given here.

For real systems, Equation 1.28 is in general not true, and many-body terms appear in the Hamiltonian:

$$H = \sum_{n=1}^N \frac{p_n^2}{2m} + \sum_{n < k} u_{nk} \quad (1.30)$$

We can write an expression for the partition function of the interacting system by considering that every microstate is ultimately expressed as a function of the positions and momenta of each particle, which can vary continuously:

$$Z = \frac{1}{N!h^{3N}} \int \int e^{-\beta \sum_n \frac{p_n^2}{2m}} e^{-\beta \sum_{n < k} u_{nk}} d\mathbf{p}_1 \dots d\mathbf{p}_N d\mathbf{r}_1 \dots d\mathbf{r}_N \quad (1.31)$$

where $(\mathbf{p}_n, \mathbf{r}_n)$ indicate respectively the three spatial degrees of freedom of momenta and position for each particle with index n , and h is the Planck constant (used here only to make Z dimensionless). We can rewrite this in terms of the ideal partition function Z_0 and a correction term Q called configuration integral:

$$\begin{aligned} Z &= Z_0 Q \\ Z_0 &= \frac{V^N}{N!h^{3N}} \int e^{-\beta \sum_n \frac{p_n^2}{2m}} d\mathbf{p}^N \\ Q &= \frac{1}{V^N} \int e^{-\beta \sum_{n < k} u_{nk}} d\mathbf{r}^N \end{aligned} \quad (1.32)$$

The configuration integral can be approximated by a series expansion:

$$Q = 1 + \frac{N}{V} \alpha_1 + \frac{N(N-1)}{2V^2} \alpha_2 + \dots \quad (1.33)$$

from which it's possible to derive the Van der Waals equation of state by limiting the expansion to the first corrective term.

1.4 Simulations of many-body systems

1.4.1 Statistical mechanics and molecular simulations

In Subsection 1.3.2 it was shown that knowing the partition function of a system would allow to compute all its thermodynamic properties. In practice, knowing the partition function of a system would mean knowing the energy of all its possible configurations, or *microstates*. All systems of real interest are complex enough that the enumeration of all possible microstates is impossible; hence, simulations are used as a tool to explore the *phase space* of possible microstates. A simulation is assumed to correctly sample from the set of accessible microstates, and if the number of sampled microstates is large enough, adequate approximations of the *true* physical quantities of the system can then be computed.

There are two ways of generating microstates for a simulation:

- Molecular dynamics (MD) runs simulate the evolution of a system in time by applying the classical laws of motion that govern each atom. The microstates sampled are thus connected in time, and can be considered part of a microcanonical (constant NVE) ensemble, although, as it will be discussed in Subsection 1.4.3, MD runs can also be performed in the canonical (NVT) and NPT ensemble. The system is assumed to be *ergodic*, meaning that a long-time average for a generic property A will equal the ensemble average. Then, the exact analytic expression for the average of A ,

$$\langle A \rangle = \int \int d\mathbf{p}^N d\mathbf{r}^N A(\mathbf{p}^N, \mathbf{r}^N) \rho(\mathbf{p}^N, \mathbf{r}^N) \quad (1.34)$$

where $\rho(\mathbf{p}^N, \mathbf{r}^N)$ is the probability of a microstate with momenta \mathbf{p}^N and positions \mathbf{r}^N , and for a NVT ensemble it's the so called *Maxwell-Boltzmann* distribution derived in Subsection 1.3.2,

$$\rho(\mathbf{p}^N, \mathbf{r}^N) = \frac{1}{Z} e^{-\beta E(\mathbf{p}^N, \mathbf{r}^N)} \quad (1.35)$$

is approximated by the MD simulation as:

$$\langle A \rangle = \frac{1}{M} \sum_{i=1}^M A(\mathbf{p}^N, \mathbf{r}^N) \quad (1.36)$$

where M is the number of time steps being simulated. Note that Equation 1.34 is not valid for *all* thermodynamic properties, and in particular it's applicable only to properties for which the value $A(\mathbf{p}^N, \mathbf{r}^N)$ of a *particular microstate* is defined. Notable exceptions are the entropy and the free energy; this point will not be expanded upon further here, but more will be said in Subsection 1.4.4.

- Monte Carlo (MC) simulations, on the other hand, generate microstates that do *not* represent the time evolution of a system. A MC simulation is composed of M iterations; at every iteration, a “trial move” (which can consist, for example, in the translation or rotation of a molecule) is applied to the system. Since the move is applied randomly, the simulation would explore only a tiny portion of the vast $3N$ -dimensional phase space of possible positions for each atom, without much utility. So, the potential energy of the configuration obtained by applying the trial move is computed, and if it is lower than the energy of the previous configuration, the move is accepted and applied; if instead the

energy is higher, the *Boltzmann factor* of the energy difference is calculated:

$$B = e^{-\beta(E_{new}(\mathbf{r}^N) - E_{old}(\mathbf{r}^N))} \quad (1.37)$$

and compared to a random number b generated between 0 and 1. If $B > b$, the move is accepted, otherwise the move is rejected and the previous configuration is used for the next iteration. This way, the MC simulation will generate preferentially low-energy configurations (with the possibility of generating, at each step, a higher energy configuration according to the magnitude of the “uphill” jump), correctly sampling from a NVT ensemble. So, for a generic property A , the simulation computes the average as:

$$\langle A \rangle = \frac{1}{M} \sum_{i=1}^M A(\mathbf{r}^N) \quad (1.38)$$

for M Monte Carlo iterations. Notation in Equation 1.38 makes explicit how, in a MC simulation, the generic property A depends only on the position of the particles, rather than their position *and* their momenta. The reason why no information is lost by only considering the positions of the particles is the separability of the partition function already shown in Equation 1.32. The integral over the momenta contained in the total partition function can be solved analytically:

$$\int e^{-\beta \sum_n \frac{p_n^2}{2m}} d\mathbf{p}^N = (2\pi m k_B T)^{3N/2} \quad (1.39)$$

Thermodynamic properties can thus be written as a sum of an ideal contribution due to the momenta (computed analytically) and a contribution due to the interaction between the particles, which depends only on their position and is estimated numerically by the simulation.

A common goal of molecular simulations, regardless of the method adopted, is to simulate *macroscopic* systems and their properties. Thus, *periodic boundary conditions* are often applied: the simulation box is replicated (in one or, more often, multiple dimensions) so that, provided that the simulation box is big enough to begin with (see Subsection 1.4.2 below), each particle will experience the same potential it would in the macroscopic bulk. Figure 1.2 shows a periodic simulation box of liquid water equilibrated through a MD run in the NPT ensemble.

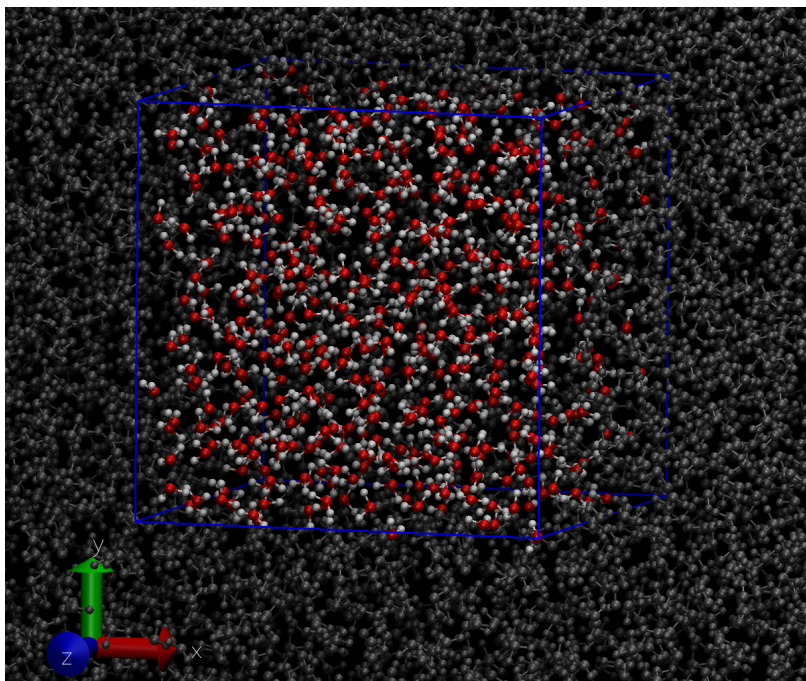


Figure 1.2: A periodic simulation box of liquid water equilibrated through a MD run in the NPT ensemble (300 K, 1 atm); the periodic boundaries are shown in blue, while the water molecules of the periodic images are shown in gray.

1.4.2 Force fields

The equations of statistical mechanics derived in Subsection 1.3.2 are valid independently of the laws of physics that determine the energy of the system in a particular configuration (microstate). Hence, one could in principle use quantum mechanics (see Section 1.5) to determine the energies for many-body simulations, and indeed this is one possible approach [21]. More commonly, however, due to the high computational cost of *ab initio* methods, many-body simulations make use of a simplified description of the system, in which electrons are not treated explicitly, and a set of functional forms and parameters, known as a *force field*, is employed to describe all the inter and intra-molecular interactions.

Chemical bonds are included in the description through the definition of attractive forces for which different functional forms can be used, but the harmonic potential is a common choice. Molecular structure is further defined by energy terms for *angles* formed by three atoms and *dihedral angles* between two planes; for a set of four atoms bonded in the sequence A-B-C-D, a dihedral angle is identified between a plane pass-

ing through atoms A-B-C and another plane passing through atoms B-C-D. If, in contrast, there are three atoms all bonded to a central fourth atom, the two planes are said to form an *improper dihedral angle*. Finally, non-bonded interactions are treated with functions that depend on the distance between pairs of atoms, and are commonly divided in electrostatic and van der Waals interactions. A possible explicit expression for the potential energy of a system modelled through a force field is thus:

$$\begin{aligned}
 E = & \sum_i^{\text{bonds}} k_i (r_i - r_{i,0})^2 + \sum_i^{\text{angles}} k'_i (\theta_i - \theta_{i,0})^2 + \sum_i^{\text{dihedrals}} k''_i [1 + d_i \cos(n_i \phi_i)] \\
 & + \sum_{\nu > \mu}^{\text{particles}} \left(4\epsilon_{\nu\mu} \left[\left(\frac{\sigma_{\nu\mu}}{r_{\nu\mu}} \right)^{12} - \left(\frac{\sigma_{\nu\mu}}{r_{\nu\mu}} \right)^6 \right] + \frac{q_\nu q_\mu}{4\pi\epsilon_0 r_{\nu\mu}} \right)
 \end{aligned}
 \tag{1.40}$$

where each term represents, in order, bonds, angles, dihedral angles, and non-bonded terms (van der Waals and electrostatic); common functional forms were chosen for each.

Non-bonded terms are evaluated, in principle, between each pair of particles, even those that are far enough that their contribution is negligible. Thus, in order to limit the computational cost, a *cutoff* is often applied, and all pairwise interactions beyond the chosen cutoff radius r_c are considered to be null. This is especially necessary when periodic boundary conditions are applied, since care must be taken so that each particle will not interact with its own periodic image, an artifact that would give unphysical features in the simulation. For a periodic cubic box, this implies that the cutoff must always be less than half the box side. Several correction methods exist to treat interactions, such as the coulomb charge-charge interaction, that have significant contributions even beyond the cutoff [22].

In force field simulations, every atom or particle has a *type* associated to it. For an atom, the type is certainly determined by the element of the atom, but it can include more information about its chemical nature and that of its surroundings (for example, to differentiate sp² and sp³ carbon atoms). Different parameters, fitted to experimental data or *ab initio* calculations (see Chapter 3), are assigned to each atom type or pair. The force field approach greatly speeds up calculations, but it inevitably raises issues of performance (accuracy compared to experimental data and/or higher-level computations) and transferability (applicability of the force field to a large number of systems, at least with respect to the data set used for the fitting procedure).

The basic force field form shown above can be improved with the inclusion of more sophisticated terms, especially for the non-bonded

interactions, which are critical for many systems of scientific interest [23]; examples include polarizable electrostatic interactions [24] and nonadditive many-body interactions [25].

Conversely, force fields can be parametrized to purposefully giving up detail to simulate matter at a different scale, e.g. to model the mechanical properties of a solid or study very big molecules such as polymers or biomolecules. This is referred to as *coarse-grained* modelling, meaning that matter is described at a macroscopic (non-atomistic) level [26]. Coarse-grained models permit to apply computational techniques to a wider range of problems that would be inaccessible for atomistic simulations, although it should be noted that in recent years atomistic MD has been applied to systems composed of up to a billion atoms, that is an entire gene of DNA [27].

1.4.3 The basic molecular dynamics algorithm

A MD run is typically structured in the following way:

1. A starting geometry is generated in some way by the user and supplied to the MD code, together with the force field and all the information necessary to specify the parameters of the simulation.
2. Initial velocities are assigned to each particle randomly; the velocities can follow a uniform distribution, or already be selected from the Maxwell-Boltzmann distribution, where the probability for the velocity of particle i along the x direction is given by:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} e^{-\frac{m_i v_{ix}^2}{2k_B T}}.$$
3. Time is considered in discrete timesteps δt , usually in the order of 1 fs, but chosen by the user. The forces acting on each particle are computed at every timestep according to the potentials of the force field employed (or, indeed, *ab initio* calculations), and the classical laws of motion ($F = ma$) are used to compute the corresponding accelerations for each particle. Knowing the positions, velocities, and accelerations for each particle at time t , new positions and velocities can be computed for every particle at time $t + \delta t$. The cycle is repeated until the system has reached *equilibrium*, that is until a set of monitored properties becomes stable in time; the properties chosen to monitor equilibration can depend on the system being simulated, but generally include thermodynamic quantities such as energy, temperature, and pressure.
4. Now that the system has been equilibrated, the so called *production* run begins. The laws of motion are integrated as before until the number of timesteps set by the user is reached.

5. The properties of interest are computed from the production MD trajectory; many thermodynamic properties, such as the total potential and kinetic energies, temperature, and pressure, are commonly already computed by the MD code and printed to a output file that contains both the values at each timestep and the total averages.

As stated in Subsection 1.4.1, the “traditional” MD algorithm described above samples from the microcanonical ensemble (constant NVE). For MD runs to more closely resemble the conditions of an actual experiment, it’s often desirable to perform MD simulations in the canonical (NVT) ensemble or the NPT ensemble.

Several methods exist to apply thermostats and barostats to MD simulations [28]. A basic approach would be to scale the velocities of the particles or the volume of the box at every step in order to enforce the desired temperature or pressure, respectively. More sophisticated methods, such as the one proposed by Berendsen *et al.* [29], weakly couple the system to an external bath at constant temperature or pressure.

The change in temperature between two timesteps given by the Berendsen thermostat is:

$$\Delta T = \frac{\delta t}{\tau} \left(T_{bath} - T(t)_{system} \right) \quad (1.41)$$

where τ is a parameter that determines the strength of the coupling to the heat bath. The velocities will be scaled by a factor λ such that:

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left(\frac{T_{bath}}{T(t)_{system}} - 1 \right) \quad (1.42)$$

note that for $\tau = \delta t$ this reduces to the simple velocity rescaling method.

For the barostat, we have:

$$\Delta P = \frac{\delta t}{\tau_p} \left(P_{bath} - P(t)_{system} \right) \quad (1.43)$$

and the volume is scaled by a factor:

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_p} \left(P(t)_{system} - P_{bath} \right) \quad (1.44)$$

with κ being a constant.

1.4.4 Free energy calculations

In chemistry, the importance of knowing the change in free energy associated with a process of interest cannot be over-emphasized, as this change determines whether the process will occur spontaneously or not according to the relations:

$$\Delta F = \Delta U - T\Delta S \leq 0 \quad (\text{constant } T, V) \quad (1.45a)$$

$$\Delta G = \Delta U - T\Delta S + P\Delta V \leq 0 \quad (\text{constant } T, P) \quad (1.45b)$$

Thus, being able to accurately compute the free energy of a system is a goal of obvious importance. A difficulty arises from the fact that the free energy is an *entropic*, rather than a *mechanical* thermodynamic property [22]. From the perspective of computation, the difference between mechanical properties, such as the internal energy U , and entropic properties, is that the former depend on the derivative of the partition function, while the latter depend on the partition function itself. The relevant equations for the internal energy U and the Helmholtz free energy F are:

$$U = \frac{k_B T^2}{Z} \frac{\partial Z}{\partial T} \quad (1.46a)$$

$$F = -k_B T \log Z \quad (1.46b)$$

From Subsection 1.3.2 and Subsection 1.4.1, ignoring normalization constants:

$$\begin{aligned} U &= \frac{k_B T^2}{Z} \frac{\partial Z}{\partial T} \\ &= \int \int d\mathbf{p}^N d\mathbf{r}^N E(\mathbf{p}^N, \mathbf{r}^N) \rho(\mathbf{p}^N, \mathbf{r}^N) \end{aligned} \quad (1.47)$$

$$\begin{aligned} F &= -k_B T \log Z \\ &= k_B T \log \left(\int \int d\mathbf{p}^N d\mathbf{r}^N e^{-\frac{E(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}} \rho(\mathbf{p}^N, \mathbf{r}^N) \right) \end{aligned} \quad (1.48)$$

Equation 1.47 and Equation 1.48 contain a crucial difference: due to the exponential $e^{-\frac{E(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}}$ in Equation 1.48, microstates with high energies will have non-negligible contributions to F despite having low probability ρ , while this doesn't happen for U . Now, since MD and MC simulations sample preferentially from lower-energy states according to the Boltzmann distribution, as described in Subsection 1.4.1, they are generally inadequate to compute accurate values of free energy. This

prevents from computing the difference of free energy between states α and β as simply $\Delta F = F_\beta - F_\alpha$. Instead, specific methods must be used.

One widely used method is that proposed by Charles Bennett [30], known as *Bennett acceptance ratio* (BAR) method. If the difference between states α and β can be expressed as a difference between two Hamiltonians H_α and H_β acting on the same phase space, then we can exploit the fact we are interested in the *difference* between the free energies of the two states, rather than the free energy of each. The difference in free energy will be:

$$\begin{aligned}\Delta F &= F_\beta - F_\alpha = -k_B T \log \frac{Z_\beta}{Z_\alpha} \\ &= -k_B T \log \left(\frac{\int \int d\mathbf{p}^N d\mathbf{r}^N e^{-\frac{H_\beta(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}}}{\int \int d\mathbf{p}^N d\mathbf{r}^N e^{-\frac{H_\alpha(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}}} \right)\end{aligned}\quad (1.49)$$

and multiplying times 1 in the form $\left(1 = e^{\frac{H_\alpha(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}} e^{-\frac{H_\alpha(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}} \right)$ we get:

$$\Delta F = -k_B T \log \left(\frac{\int \int d\mathbf{p}^N d\mathbf{r}^N e^{-\frac{H_\beta(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}} e^{\frac{H_\alpha(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}} e^{-\frac{H_\alpha(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}}}{\int \int d\mathbf{p}^N d\mathbf{r}^N e^{-\frac{H_\alpha(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}}} \right)\quad (1.50)$$

which can be written as an ensemble average:

$$\Delta F = -k_B T \log \left\langle \frac{e^{[H_\alpha(\mathbf{p}^N, \mathbf{r}^N) - H_\beta(\mathbf{p}^N, \mathbf{r}^N)]}}{k_B T} \right\rangle_\alpha\quad (1.51)$$

The condition that α and β span the same phase space might seem more restrictive than it is. Indeed, Equation 1.51 can be used to compute experimentally interesting quantities such as solvation and protein-ligand binding free energies. This is usually done by performing a series of simulations where the Hamiltonian is altered through a parameter λ , which assumes a specific value for each simulation and varies between 0 and 1. The two states α ($\lambda = 0$) and β ($\lambda = 1$) are thus connected by a series of intermediate steps, ensuring a good overlap between the phase spaces of steps at λ and $\lambda + \delta\lambda$ and making Equation 1.51 valid. (This method is sometimes called MBAR, *multistate* Bennett acceptance ratio.)

For example, λ might modulate the intermolecular interactions between a solute and the solvent molecules in a MD run: at $\lambda = 0$ the

solute interacts normally with its surroundings, while at $\lambda = 1$ the solute is fully decoupled and doesn't interact with the other molecules. For each value of λ , the potential energy of every microstate visited by the system during the simulation is evaluated for λ and $\lambda + \delta\lambda$, and Equation 1.51 is used to compute the change in free energy associated with the step ($\lambda \rightarrow \lambda + \delta\lambda$); ΔF_{solv} is then obtained by summing all contributions from simulations at different values of λ (with a change of sign, since we are simulating the process of a molecule *disappearing* from the solvent bulk, rather than a solvation process). In mathematical notation:

$$\Delta F_{(\lambda \rightarrow \lambda + \delta\lambda)} = -k_B T \log \left\langle \frac{e^{[H_{(\lambda)} - H_{(\lambda + \delta\lambda)}]}}{k_B T} \right\rangle_{\lambda} \quad (1.52a)$$

$$\Delta F_{solv} = - \sum \Delta F_{(\lambda \rightarrow \lambda + \delta\lambda)} \quad (1.52b)$$

1.5 Ab initio methods

In 1929, Nobel laureate Paul Dirac noted: “The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.” [31]

In the same years, electronic structure methods seeking approximate solutions began to emerge, laying the foundations of modern *ab initio* computational chemistry. One of such methods was that proposed by Hartree [32], with later important corrections from John Slater [33] and Vladimir Fock [34]. The Hartree-Fock (HF) method, as it is now known, aims at solving the Schrödinger equation by an iterative, *self-consistent*, procedure. A different approach based on the electron density emerged alongside HF and quickly became popular in the scientific community due to its accuracy and computational efficiency. This method, called density functional theory (DFT) will be presented in detail in Subsection 1.5.3.

Although the focus of this section will be on DFT, a brief introduction of the HF algorithm will be given in Subsection 1.5.2 due to the close relationship, both historical and mathematical, that these two approaches share.

1.5.1 Notes on notation

In quantum mechanics, physical states are represented by vectors in a complex Hilbert space, that is a vector space with infinite dimensions

for which the inner product is defined. Throughout this Section the so called *Dirac* or *bra-ket* notation is used interchangeably with the “wave function” notation, when useful for clarity. In Dirac notation, a state is defined by a *ket* $|\Psi\rangle$; physical observables are associated with Hermitian operators that satisfy eigenvalue equations of the type:

$$H|\Psi\rangle = E|\Psi\rangle \quad (1.53)$$

where the generic eigenvalue, E in this case, is the value of the observable (energy) in state $|\Psi\rangle$.

Below is given the relationship between the ket and the wave function of the system (which is the projection of $|\Psi\rangle$ onto the position basis $\{|\mathbf{r}\rangle\}$), as well as all the other equivalent notations used in this Section:

$$\Psi(\mathbf{r}) = \langle \mathbf{r} | \Psi \rangle \quad (1.54a)$$

$$|\Psi\rangle = \int d\mathbf{r} \Psi(\mathbf{r}) |\mathbf{r}\rangle \quad (1.54b)$$

$$H\Psi(\mathbf{r}) = \langle \mathbf{r} | H | \Psi \rangle \quad (1.54c)$$

$$\langle \Psi | \Psi \rangle = \int \Psi^*(\mathbf{r}) \Psi(\mathbf{r}) d\mathbf{r} \quad (1.54d)$$

$$\langle \Psi | H | \Psi \rangle = \int \Psi^*(\mathbf{r}) H\Psi(\mathbf{r}) d\mathbf{r} \quad (1.54e)$$

No special symbols are used to signal that the Hamiltonian H is an operator, or to distinguish operators acting on kets rather than scalar fields (wave functions); the complex conjugate symbol (*) will also be dropped, leading to the notations:

$$\langle \Psi | H | \Psi \rangle = \int \Psi(\mathbf{r}) H\Psi(\mathbf{r}) d\mathbf{r} \quad (1.55a)$$

$$\langle \Psi | \Psi \rangle = \int \Psi^2(\mathbf{r}) d\mathbf{r} \quad (1.55b)$$

1.5.2 The Hartree-Fock method

By invoking the Born-Oppenheimer approximation (see Subsection 1.2.1), the electronic, time-independent Schrödinger equation can be written as:

$$H|\Psi_{el}\rangle = E|\Psi_{el}\rangle \quad (1.56)$$

This equation will have multiple solutions, corresponding to the different energy levels possible for the system. We use the term *ground state* to indicate the state with the lowest energy, E_0 .

$$H|\Psi_0\rangle = E_0|\Psi_0\rangle \quad (1.57)$$

The set of $\{\Psi_i\}$ eigenfunctions, each corresponding to an energy E_i , are orthonormal. This is expressed notationally by the use of the Kronecker delta:

$$\int \Psi_i \Psi_j d\mathbf{r} = \delta_{ij} = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases} \quad (1.58)$$

The orthogonality comes from the fact that the Hamiltonian H is a Hermitian operator, and the additional condition is satisfied if the eigenfunctions are properly normalized. Given this, we can use the set $\{\Psi_i\}$ as a basis set to write a function Φ as a linear combination:

$$\Phi = \sum_i c_i \Psi_i \quad (1.59)$$

We take Φ to represent an approximation for Ψ_0 , the object we are most interesting in obtaining (together with the corresponding E_0). Note that the coefficients $\{c_i\}$, as well as the eigenfunctions $\{\Psi_i\}$, are all unknown, so no real progress has been made yet. However, from this it can be easily proven that:

$$\frac{\int \Phi H \Phi d\mathbf{r}}{\int \Phi^2 d\mathbf{r}} \geq E_0 \quad (1.60)$$

where the denominator equals 1 if Φ is normalized. This is known as the *variational principle*, and it's the theoretical bedrock upon which *ab initio* computational chemistry methods are built. The variational principle states that the expectation value for the Hamiltonian applied on a trial wave function Φ will be always greater than (or equal to) the true value for the energy of the ground state. Given that we may express Φ as a linear combination of a set of *arbitrary* basis functions (through a change of basis), the variational principle hints at the possibility of iteratively refining an initial guess for the ground-state wave function, and provides a mean to assess the performance of such approximate solutions.

The way the Hartree-Fock method *actually* produces incremental guesses for the ground-state wave function is by first approximating the total wave function as a *Slater determinant* formed by *molecular spinorbitals*, which are one-electron wave functions with three spatial coordinates and one spin coordinate:

$$\Phi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_N(2) \\ \dots & \dots & \dots & \dots \\ \chi_1(N) & \chi_2(N) & \dots & \chi_N(N) \end{vmatrix} \quad (1.61)$$

Each spinorbital is the product of a *molecular orbital*, dependent only on space coordinates, and a spin function:

$$\chi_i(a) = \phi_i(a) \eta_i(a) \quad (1.62)$$

where the index a refers to the (spatial or spin) coordinates of electron a and subscript i labels the molecular orbitals and spinorbitals. Due to the Pauli exclusion principle, each molecular orbital can be occupied at most by two electrons with opposite spin. This is enforced mathematically by ensuring the total wave function is *antisymmetric*, that is, it changes sign upon interchanging the space and spin coordinates of two electrons. The use of a determinant as in Equation 1.61, first proposed by John Slater [35], satisfies this property.

By considering systems with paired electrons in every molecular orbital (called *closed-shell* systems), it can easily be shown that an antisymmetrized product such as Equation 1.61 for representing the total wave function can be constructed from molecular orbitals (rather than spinorbitals), which themselves form an orthonormal set.

The molecular orbitals are then constructed as a linear combination of basis functions:

$$\phi_i = \sum_{\mu=1}^n c_{\mu i} \psi_{\mu} \quad (1.63)$$

where the coefficients $c_{\mu i}$ are to be determined.

In his seminal 1951 paper [36], Roothaan described a matrixial equation whose solution would provide the optimal values for the coefficients. The equation was obtained by applying the method of Lagrange multipliers to the problem of the minimization of electronic energy as given by:

$$\int \Phi H \Phi d\mathbf{r} = E \quad (1.64)$$

with constraints that result from the orthonormality of the molecular orbitals.

The Roothaan matrixial equation is given below:

$$\mathbf{FC} = \mathbf{SCE} \quad (1.65)$$

where \mathbf{F} is composed of the matrix elements of the *Fock* (Hamiltonian) operator, \mathbf{C} is a matrix in which each column contains the molecular orbital coefficients for one orbital, \mathbf{S} is the overlap matrix with elements $S_{\mu\nu} = \langle \mu | \nu \rangle = \int \psi_{\mu} \psi_{\nu} d\mathbf{r}$, and \mathbf{E} is the diagonal matrix of orbital energies.

There will be N values ϵ_i for Equation 1.65 such that:

$$\mathbf{Fc}_i = \epsilon_i \mathbf{Sc}_i \quad (1.66)$$

where \mathbf{c}_i are now eigenvectors of \mathbf{F} corresponding to the orbital coefficients and ϵ_i are the associated orbital energies. Solutions are found by writing:

$$\begin{aligned}\mathbf{F}\mathbf{c} &= \epsilon\mathbf{S}\mathbf{c} \\ (\mathbf{F} - \epsilon\mathbf{S})\mathbf{c} &= 0\end{aligned}\tag{1.67}$$

and imposing:

$$\det(\mathbf{F} - \epsilon\mathbf{S}) = \begin{vmatrix} F_{11} - \epsilon S_{11} & F_{12} - \epsilon S_{12} & \dots & F_{1n} - \epsilon S_{1n} \\ F_{21} - \epsilon S_{21} & F_{22} - \epsilon S_{22} & \dots & F_{2n} - \epsilon S_{2n} \\ \dots & \dots & \dots & \dots \\ F_{n1} - \epsilon S_{n1} & F_{n2} - \epsilon S_{n2} & \dots & F_{nn} - \epsilon S_{nn} \end{vmatrix} = 0\tag{1.68}$$

It must be now crucially noted how \mathbf{F} is dependent upon the coefficients in \mathbf{c} . The elements $F_{\mu\nu}$ are given by, in Dirac notation:

$$F_{\mu\nu} = \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| \nu \right\rangle - \sum_m^{nuclei} Z_m \left\langle \mu \left| \frac{1}{r_m} \right| \nu \right\rangle + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \mu\nu | \lambda\sigma \rangle - \frac{1}{2} \langle \mu\lambda | \nu\sigma \rangle \right]\tag{1.69}$$

As before, the lower-case Greek letters span the basis set functions. The first term represents the kinetic energy of the electrons; the second term is the coulomb attraction between the nuclei and the electrons; the last term represents the coulomb repulsion between electrons and it takes into account the *exchange energy* term that arises from the Pauli exclusion principle. The double notation stands for:

$$\langle \mu\nu | \lambda\sigma \rangle = \int \int \psi_\mu(1) \psi_\nu(1) \frac{1}{r_{12}} \psi_\lambda(2) \psi_\sigma(2) d\mathbf{r}(1) d\mathbf{r}(2)\tag{1.70}$$

The electron-electron repulsion clearly depends on the occupation of the orbitals, and this dependency is included in the terms $P_{\lambda\sigma}$ of the *density matrix* \mathbf{P} , which have the form:

$$P_{\lambda\sigma} = 2 \sum_i^{occupied} c_{\lambda i} c_{\sigma i}\tag{1.71}$$

The factor of two appears because we are considering singlet states, in which each orbital is doubly occupied.

Since we are solving Equation 1.68 to find the eigenvalues ϵ_i of \mathbf{F} corresponding to the molecular orbital energies and ultimately find the coefficients \mathbf{c}_i forming the corresponding eigenvectors, but the same coefficients appear in the definition of \mathbf{F} , the procedure must be carried out iteratively:

- First, a guess for the density matrix is made.
- Then, Equation 1.68 is solved to obtain new vectors \mathbf{c}_i .
- The new vectors \mathbf{c}_i are used to construct an updated density matrix; Equation 1.68 is solved again.
- The loop is repeated until convergence (which could be defined according to different criteria) is reached.

The method as outlined above is the basis for more precise calculations (so called *post-Hartree-Fock* methods) that include *electronic correlation* through different computational techniques. This extra energy term is not captured by Hartree-Fock theory because electron repulsion is computed as the repulsion of static orbitals, and no instantaneous correlation in the motion of the electrons is included.

Although these improved methods are needed to compute adequately accurate energies for systems bigger than the hydrogen atom, they are not the focus of this thesis and so they will not be discussed here.

1.5.3 Density Functional Theory

1.5.3.1 From the Thomas-Fermi model to the Hohenberg-Kohn theorems

Density functional theory (DFT) is an *ab initio* electronic structure method that represents an alternative to molecular orbital theories such as HF and post-HF. Instead of solving the Schrödinger equation to find the ground-state wave function of the system, DFT focuses on the electron density. The electron density represents the probability, accounting for the total number of electrons, of finding an electron in an infinitesimal volume ($dx dy dz$). However, the electron density can also be thought of as a *physical observable*, as it is observed in electron diffraction and X-ray crystallography experiments [37].

The roots of DFT can be traced back to the model of Thomas and Fermi [38, 39]. This model expresses the total energy of an atom as:

$$E(\rho(\mathbf{r})) = T(\rho(\mathbf{r})) + U_{Ne}(\rho(\mathbf{r})) + U_{ee}(\rho(\mathbf{r})) \quad (1.72)$$

where T is the kinetic energy of the electrons, U_{Ne} is the coulomb attraction between nuclei and electrons, and U_{ee} is the coulomb electron-electron repulsion.

It can be intuitively understood how, among the terms in Equation 1.72, expressing the kinetic energy of the electrons as a function of the electron density alone would entail the biggest approximations. Furthermore, as shown in Subsection 1.5.2, the electron-electron repulsion

term should include a correction due to the Pauli exclusion principle, a fact not taken into consideration by this model. Indeed, although the Thomas-Fermi model has found applications due to its simplicity and versatility, the approximations made were too rough for accurate chemical calculations.

The birth of modern DFT happened almost forty years later, thanks to the seminal paper of Pierre Hohenberg and Walter Kohn [40]. Hohenberg and Kohn gave DFT a rigorous foundation by first showing that for electrons moving in an external potential $V(\mathbf{r})$ (given by the presence of atomic nuclei at fixed positions), there will be a *unique* electron density $\rho(\mathbf{r})$ compatible with $V(\mathbf{r})$. Hence, the ground-state electron density uniquely determines the potential and consequently the energy and every other property of the system. The proof of this statement, known as the first HK theorem, is simple, and proceeds by *reductio ad absurdum*.

We assume that two different potentials, V_a and V_b , are both compatible with the same ground-electron density ρ_0 . V_a and V_b will determine two different Hamiltonians we call H_a and H_b , each one associated with a ground-state wave function, $\Psi_{0,a}$ and $\Psi_{0,b}$ respectively. The variational principle (Equation 1.60) then imposes:

$$E_{0,a} < \langle \Psi_{0,b} | H_a | \Psi_{0,b} \rangle \quad (1.73)$$

from which:

$$\begin{aligned} E_{0,a} &< \langle \Psi_{0,b} | H_a - H_b + H_b | \Psi_{0,b} \rangle \\ &< \langle \Psi_{0,b} | H_a - H_b | \Psi_{0,b} \rangle + \langle \Psi_{0,b} | H_b | \Psi_{0,b} \rangle \\ &< \langle \Psi_{0,b} | V_a - V_b | \Psi_{0,b} \rangle + E_{0,b} \end{aligned} \quad (1.74)$$

Hence, we obtain:

$$E_{0,a} < \int [V_a(\mathbf{r}) - V_b(\mathbf{r})] \rho_0(\mathbf{r}) d\mathbf{r} + E_{0,b} \quad (1.75)$$

but since the labels are completely arbitrary, it must also be:

$$E_{0,b} < \int [V_b(\mathbf{r}) - V_a(\mathbf{r})] \rho_0(\mathbf{r}) d\mathbf{r} + E_{0,a} \quad (1.76)$$

Adding Equation 1.75 and Equation 1.76 we get:

$$E_{0,a} + E_{0,b} < E_{0,b} + E_{0,a} \quad (1.77)$$

This inconsistency proves that $V(\mathbf{r})$ must indeed be unique for a given electron density of the ground state.

The second piece of the theoretical bedrock laid by Hohenberg and Kohn, known as the second HK theorem, is the application of the variational principle to the ground-state electron density, in analogy with HF theory, in which the principle is applied to the ground-state wave function. In fact, since a trial density ρ' defines an Hamiltonian H' and a corresponding wave function Ψ' , we can use Ψ' as a trial wave function for the expectation value of the *true* Hamiltonian H (the one generated from the true external potential $V(\mathbf{r})$). By virtue of the variational principle, we have:

$$\langle \Psi' | H | \Psi' \rangle = E(\rho') \geq E_0(\rho_0) = \langle \Psi_0 | H | \Psi_0 \rangle \quad (1.78)$$

Like for the HF wave function described in Subsection 1.5.2, then, Equation 1.78 tells us that, in principle, a method can be devised to obtain the correct ground-state energy from a trial electron density by iteratively refining an initial guess.

1.5.3.2 Kohn-Sham DFT

Although the HK theorems give DFT a sounder theoretical basis, they still don't address the problems highlighted for the Thomas-Fermi model, that is the need for an accurate explicit description of the total energy as a functional of the electron density.

In 1965 Kohn, working with Lu Jeu Sham, proposed a solution that is still by far the most used today [41]. To see how the energy terms are treated by Kohn-Sham DFT (KS DFT), we return to the expression of the total energy as a functional of the density:

$$E(\rho(\mathbf{r})) = T(\rho(\mathbf{r})) + U_{Ne}(\rho(\mathbf{r})) + U_{ee}(\rho(\mathbf{r})) \quad (1.79)$$

The nuclear-electron interaction can be computed exactly as:

$$U_{Ne}(\rho(\mathbf{r})) = \sum_m^{nuclei} \int \frac{Z_m \rho(\mathbf{r})}{|\mathbf{R}_m - \mathbf{r}|} d\mathbf{r} \quad (1.80)$$

An expression for *classical* electron-electron repulsion (that is, by ignoring the effect of exchange and instantaneous electron correlation, in addition to self-repulsion) can be given as:

$$J(\rho(\mathbf{r})) = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad (1.81)$$

Finally, KS DFT computes the kinetic energy in the non-interacting limit, as if the kinetic energy of each electron does not depend on the presence of the others. Mathematically, this is equivalent to saying

that KS DFT assumes that the electron density corresponds to a wave function composed by a single Slater determinant. In other words, this is the Hartree-Fock kinetic energy:

$$T_{Slater}(\rho(\mathbf{r})) = \sum_{i=1}^N \left\langle \phi_i \left| -\frac{1}{2} \nabla^2 \right| \phi_i \right\rangle \quad (1.82a)$$

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2 \quad (1.82b)$$

where ϕ_i are called the *Kohn-Sham orbitals*. Then, all corrections due to the approximations described above are included by defining the *exchange-correlation energy* as:

$$E_{xc}(\rho(\mathbf{r})) = [U_{ee}(\rho(\mathbf{r})) - J(\rho(\mathbf{r}))] + [T(\rho(\mathbf{r})) - T_{Slater}(\rho(\mathbf{r}))] \quad (1.83)$$

Hence, the total KS energy is:

$$E_{KS}(\rho(\mathbf{r})) = T_{Slater}(\rho(\mathbf{r})) + U_{Ne}(\rho(\mathbf{r})) + J(\rho(\mathbf{r})) + E_{xc}(\rho(\mathbf{r})) \quad (1.84)$$

In perfect analogy with the Hartree-Fock method (see Subsection 1.5.2), the KS orbitals are expressed as a linear combination of basis functions:

$$\phi_i = \sum_{\mu=1}^n c_{\mu i} \psi_{\mu} \quad (1.85)$$

Each orbital satisfies the eigenvalue equation for a one-electron Hamiltonian operator:

$$h_i \phi_i = E_i \phi_i \quad (1.86a)$$

$$h_i \sum_{\mu=1}^n c_{\mu i} \psi_{\mu} = E_i \sum_{\mu=1}^n c_{\mu i} \psi_{\mu} \quad (1.86b)$$

where:

$$h_i = -\frac{1}{2} \nabla_i^2 - \sum_m^{nuclei} \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_m|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' + V_{xc} \quad (1.87)$$

and:

$$V_{xc} = \frac{\partial E_{xc}}{\partial \rho} \quad (1.88)$$

By minimizing the expectation value of the energy with respect to the coefficients, a secular equation like Equation 1.68 can be obtained;

the matrix elements of the KS operator (corresponding to the Fock operator in HF) are given by:

$$K_{\mu\nu} = \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| \nu \right\rangle - \sum_m^{nuclei} Z_m \left\langle \mu \left| \frac{1}{r_m} \right| \nu \right\rangle + \left\langle \mu \left| \int \frac{\rho(\mathbf{r}')}{r'} d\mathbf{r}' \right| \nu \right\rangle + \langle \mu | V_{xc} | \nu \rangle \quad (1.89)$$

Note how the orbitals ϕ_i that we are seeking by solving Equation 1.86 for each orbital are needed for the definition of ρ , which in turn appears in Equation 1.89. This implies that, like for HF, an iterative self-consistent procedure must be carried out.

1.5.3.3 Exchange-correlation and non-covalent interactions

The method shown above is *exact*; the problem is that, of course, E_{xc} and V_{xc} are not known, and thus approximate expressions must be employed. In fact, the search for exchange-correlation potentials of increasing accuracy is a current research topic in DFT.

A very important thing to note is that, by employing approximations for E_{xc} , DFT can no longer be considered variational. Hence, there is not a clear path toward more accurate DFT calculations, and the choice of a particular functional for a system must be corroborated by previous computational studies on similar systems and/or experimental data.

The exchange (E_x) and correlation (E_c) energy terms are in general treated separately, allowing for different computational approaches for each. Note that exchange-correlation functionals are commonly referred to by prepending the acronym identifying the *exchange* functional to the acronym of the *correlation* functional, so, for example, BP86 uses Becke's B functional [42] for exchange and Perdew's P86 correlation functional. It's common to divide exchange-correlation functionals into four categories.

- The first category, also the first one to appear chronologically, makes use of the local (spin) density approximation (L(S)DA). LDA functionals compute the exchange-correlation energy at every point exclusively from the value of $\rho(\mathbf{r})$ in that point. It's common to employ the uniform electron gas approximation, a system for which the exchange term can be computed exactly. In 1980, Ceperley and Alder [43] used quantum Monte Carlo techniques to compute the total energy of fully interacting uniform electron gases at different densities. In the same year, Vosko, Wilk, and Nusair [44] used that data by subtracting the analytical exchange energy from the total energy to get a value for the

correlation energy. They proposed different fitting schemes, and the corresponding functionals are generally referred to in the literature by their initials (VWN). Although LDA functionals give very good results for molecular geometries, they suffer from large systematic errors in the computation of energy values.

- An obvious path to improve LDA functionals stems from the consideration that the electron density in a molecule is far from being a uniform electron gas; hence, exchange-correlation functionals should depend on the *gradient* of the density as well, a term that is generally added as a correction to the LDA energy. Functionals employing this correction are called GGA (generalized gradient approximation) functionals.
- Significant improvements over GGA functionals can be obtained with meta-GGA (MGGA) functionals, which continue the Taylor expansion of the exchange-correlation functional to include the second derivative of the density.
- The last category of functionals is that of *hybrid* functionals. Hybrid functionals include Hartree-Fock exchange by making use of the *adiabatic connection* method. If we imagine switching on the inter-electronic repulsion through a parameter λ , where $\lambda = 0$ means no interaction and $\lambda = 1$ means full interaction, the exchange-correlation energy is given by:

$$E_{xc} = \int_0^1 \langle \Psi(\lambda) | V_{xc} | \Psi(\lambda) \rangle d\lambda$$

One does not need an analytic expression for V_{xc} and Ψ as a function of λ in order to evaluate this integral. Instead, we can use the HF exchange energy to compute the left endpoint of the curve ($\lambda = 0$), since E_x^{HF} is the *exact* exchange for a non-interacting system. Then, an approximation for the right endpoint E'_{xc} can be made with the choice of some DFT functional, and we can write:

$$E_{xc} = E_x^{HF} + z(E'_{xc} - E_x^{HF}),$$

with z being an empirical parameter to be optimized. Becke was the first to generalize this approach by introducing two more empirical constants with his B3PW91 functional [45]:

$$E_{xc}^{B3PW91} = (1 - a)E_x^{LDA} + aE_x^{HF} + bE_x^B + E_c^{LDA} + cE_c^{PW91}$$

Note that, compared to the previous equation, here $a = 1 - z$. Stevens *et al.* [46] later modified this functional while keeping

the same values of a, b, c but using LYP instead of PW1 as a correlation functional:

$$E_{xc}^{B3LYP} = (1 - a)E_x^{LDA} + aE_x^{HF} + bE_x^B + (1 - c)E_c^{LDA} + cE_c^{LYP}$$

Due to its accuracy and broad applicability, B3LYP has been the most used functional to date, but the use of empirical constants always implies that performances might be substantially worse for systems or properties too different from those used for the fitting procedure.

In general, the performance of functionals follows John Perdew’s “Jacob’s Ladder” [47]: hybrid > meta-GGA > GGA > LDA. A comprehensive benchmark review by Mardirossian et al. on 200 functionals validates this statement [48]; this study reports the following average ranking for the best performing functional of each category: hybrid (12/200), meta-GGA (46/200), GGA (98/200), LDA (173/200).

There are many instances in which the effects of electron correlation and exchange must be considered in a broader sense in order to compute chemically useful quantities; namely, for system in which non-covalent interactions play an important role, such as for the intermolecular interactions of many molecules or the intramolecular interactions of large molecules. Non-covalent interactions are of three types [49]: the *electrostatic* interaction is the interaction (either repulsive or attractive) between two static charge distributions; *induction* effects arise from the deformation of electron distributions due to the electric field of the surroundings, and are always attractive; lastly, *dispersion* interactions are always attractive, and are due to the instantaneous correlation of the motion of electrons in different regions of space.

These terms can be understood in the context of perturbation theory; if a pair of interacting molecules A and B is far enough apart that the overlap of their wave functions can be ignored, the total wave function is simply the product:

$$\Psi_{ij}^{AB} = \Psi_i^A \Psi_j^B \quad (1.94)$$

with each Ψ_i satisfying $H\Psi_i = E_i\Psi_i$. The unperturbed Hamiltonian is:

$$H_0 = H^A + H^B \quad (1.95)$$

The perturbation is given by the electrostatic interactions between A and B:

$$H' = \sum_{a \in A} \sum_{b \in B} \frac{e_a e_b}{4\pi\epsilon_0 r_{ab}} \quad (1.96)$$

The first-order correction, $\langle \Psi_{00}^{AB} | H' | \Psi_{00}^{AB} \rangle$, will be just the electrostatic interaction between A and B. The second-order terms are obtained by considering mixed states in which one or both the molecules are in excited states:

$$E^{(2)} = - \sum'_{ij} \frac{\langle 00 | H' | ij \rangle \langle ij | H' | 00 \rangle}{E_{ij} - E_{00}} \quad (1.97)$$

where the prime on the sum indicates that it includes all states except $ij = 00$. Equation 1.97 can be split into:

$$E_{ind}^A = - \sum_{i \neq 0} \frac{\langle 00 | H' | i0 \rangle \langle i0 | H' | 00 \rangle}{E_{i0} - E_{00}} \quad (1.98a)$$

$$E_{ind}^B = - \sum_{j \neq 0} \frac{\langle 00 | H' | 0j \rangle \langle 0j | H' | 00 \rangle}{E_{0j} - E_{00}} \quad (1.98b)$$

$$E_{disp} = - \sum_{i \neq 0, j \neq 0} \frac{\langle 00 | H' | ij \rangle \langle ij | H' | 00 \rangle}{E_{ij} - E_{00}} \quad (1.98c)$$

At short distances, when the wave functions of the two molecules have non-negligible overlap, exchange corrections must be taken into account both for the first-order energy of the perturbation expansion (*exchange*) and the second-order terms (*exchange-induction* and *exchange-dispersion*). Note that analogous considerations can be made for *intra*-molecular interactions in large molecules.

Due to their local or semi-local nature, traditional density functionals are unable to describe these non-covalent interactions, which arise from highly non-local electron correlation effects [50–53]. As a consequence, numerous techniques have been developed over the years to include induction/dispersion effects in DFT calculations. These methods can be grouped into four categories: [54]

1. MGGA/hybrid functionals that contain many empirical parameters fitted to reproduce non-covalent as well as covalent effects, such as the family of Minnesota functionals [55].
2. Atom-centered potentials [56] that use the same computational framework traditionally used for effective core potentials (ECPs) meant to treat the less chemically important core electrons with an *effective potential* that doesn't include them explicitly in the HF/KS equations. These *dispersion-correcting potentials* (DCPs) are simple attractive potentials that are meant to compensate for the repulsive nature of traditional functionals, such as B3LYP, at short distances where highly-correlated post-HF methods predict an attractive interaction.

3. Van der Waals (vdW) density functionals that build from the adiabatic connection formula to include non-local correlation effects. [57]
4. Additive corrections to the KS calculations, such as Grimme's D3 [58] (and D4 [59]) correction and the XDM model [60]. These methods contain empirical parameters fitted to large data sets but are also founded on theoretical considerations that make them widely applicable. Parameters are assigned to each atom and combined to obtain 2-bodies and possibly 3-bodies interaction energies.

Each of the methods enumerated above has been shown to perform well in some situations and fail in others. This adds to what already stated for the introduction of E_{xc} approximations: the lack of a variational principle leads to methods that must be validated somehow for the studied system before the results can be considered reliable. The price paid in this regard for using DFT is compensated by the ability of computing energies for chemically-interesting systems at the precision of highly-correlated post-HF methods, but at a fraction of the computational cost.

All models are wrong,
but some are useful.

George Box

Chapter 2

Thesis outline

The work of this thesis consisted in applying computational modeling methods to systems of experimental interest. Both *ab initio* density functional theory (DFT) and molecular dynamics (MD) techniques have been used, often together in a *multiscale* approach, providing models that could be probed to assist in the interpretation of experimental data.

The bulk of the research was conducted on self-assembled monolayers (SAMs) of thiol organic molecules on gold surfaces. SAMs are supramolecular structures that form spontaneously, guided by a strong interaction between the *headgroup* of the molecules and a surface. Intermolecular van der Waals forces further contribute to the self-organization and stabilization of the system.

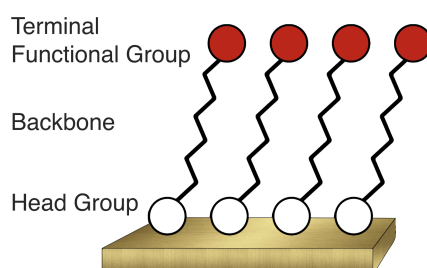


Figure 2.1: Schematic representation of a SAM of thiols on a gold substrate.

The tailorability of the organic molecules used, both at the level of their backbones and terminal functional groups, combined with the ease with which the self-assembly process is obtained, makes SAMs a subject of profound interest in surface science. Since their introduction by Nuzzo and Allara in the 1980's [61], thiol (and, especially, alkanethiol) SAMs on gold have been the subject of thousands of papers and many reviews [62–65]. SAMs are studied for potential applications,

among others, in nanolithography [66, 67], biosensing [68, 69], and electronics [64]. Furthermore, in the context of this thesis, thiols forming SAMs on gold surfaces are studied as model systems for determining the enhancement factor (EF) in surface-enhanced Raman spectroscopy (SERS).

SERS is a technique that makes use of nanostructured metal surfaces, such as gold, to greatly enhance the Raman signal of adsorbed molecules. If we denote with I_{NR} and I_{SERS} the intensities of the normal and surface-enhanced Raman signals respectively, and use the symbols N_{NR} and N_{SERS} for the number of molecules contributing to the two signals, the enhancement factor is defined as:

$$EF = \frac{I_{SERS}/N_{SERS}}{I_{NR}/N_{NR}} \quad (2.1)$$

According to ref. [70] N_{SERS} is given by:

$$N_{SERS} = \mu_M \mu_{mol} A_M A_{eff} \quad (2.2)$$

where μ_M is the surface density of the SERS-active nanostructures on the substrate (which is usually a silicon or glass surface), μ_{mol} is the surface density of the molecules adsorbed, A_M is the metallic surface area of a nanostructure, and A_{eff} is the effective area of the confocal scattering volume of the probing laser. For a quantitative determination of the enhancement factor of a SERS substrate it is therefore critical to have an accurate value for each of these terms.

The initial aim of this work was to determine the optimal density (μ_{mol} in Equation 2.2) for a SAM of 7-mercapto-4-methylcoumarin (MMC) on gold through computational means. The method was then applied to 3-mercaptopropionic acid (MPA) and the study extended to describe in detail the thermodynamics of SAM formation resulting from the models. The thesis is structured as follows:

- Chapter 3 describes the development of force field parameters fitted to DFT B3LYP-D3 calculations to model the gold/thiol interactions; tests of the parameters on target molecules MMC and MPA are reported as well.
- Chapter 4 contains a thermodynamic study of SAM formation for systems of MMC and MPA. Models of SAMs were obtained both with the molecule as a thiol (undissociated S-H bond) and as a thiyl radical (S-H is dissociated, forming a covalent S-Au bond); this was done to elucidate the differences in the SAMs resulting from the two different species, since there exists some controversy in the scientific literature about the nature of the S-Au bond, with

some works [71–74] challenging the most commonly held view of thiols dissociating at the interface to form covalent S-Au bonds. The chemical potential of each species as a function of the SAM density was computed through DFT and Bennett acceptance ratio (BAR) calculations.

- Chapter 5 consists in the results published (DOI: 10.1039/d0tc04364h) for a combined experimental/MD determination of monolayer density for a SAM of 7-mercapto-4-methylcoumarin on a gold substrate. From this, experimental and computational EFs were calculated and compared.
- Chapter 6 is composed by a manuscript currently in preparation for submission in which some of the results in Chapter 4 are presented, comparing expected monolayer densities from free energy calculations to experimental values.
- Chapter 7 contains a discussion of the overall results of these studies, and proposes several points of scientific interest that future simulations might be able to address.

Modeling techniques have also been applied to other systems, in collaboration with several experimental research groups. Internal rotational barriers for Gd-AAZTA multimers have been computed at the DFT level to investigate their dependence upon the number of monomers, a crucial factor to consider in their design as MRI contrast agents (DOI: 10.1039/d1qi00904d). DFT calculations have also been applied to models of toluene adsorbed in the porous framework of hypercrosslinked polymers (HCPs); computed NMR shifts aided in the elucidation of guest-host characteristic interactions for aromatic molecules in HCPs (DOI: 10.1021/acsapm.9b01000). Finally, MD simulations were carried out to predict the conformation of organic catalysts grafted on silica substrates, and it was shown how the number of tethering chains might affect the conformation adopted by the catalyst and hence potentially its catalytic activity (DOI: 10.1021/acs.jpcc.1c06150). The published papers containing all these additional results are attached at the end of this thesis, in Appendix A, Appendix B, and Appendix C.

Chapter 3

Force field parametrization of Au-X interactions from ab initio calculations

During the course of my PhD I developed three sets of force field (FF) parameters, each one building on the previous, to describe the bonded and non-bonded interactions between a small set of organic molecules and a gold (111) surface. In all three cases, a common and widely applicable FF (either UFF [75] or GROMOS 54A7 [76]) was extended to include Au-X pairwise interactions by fitting to DFT B3LYP-D3 calculations.

The methodologies followed to obtain the FF parameters for each set are presented below in the order they were developed. Here they have been named UFF_Au, GROMOS_Au1, and GROMOS_Au2.

3.1 UFF_Au

The UFF_Au parameters were developed to accurately reproduce the S-Au covalent bond energy, as well as Au-X non-bonded energies, for a monolayer of 7-mercapto-4-methylcoumarin (MMC) on a gold (111) surface.

First, a model for the gold surface was obtained by extracting a cluster composed of two atomic layers (26 and 20 gold atoms) from a periodic three-layers thick gold surface, previously optimized in CRYSTAL17 at the DFT/PW91 level with the basis set and pseudopotentials for core electrons of Hay and Wadt [77].

Then, one MMC molecule was optimized at the DFT/B3LYP [46] level close the upper face of the gold cluster using the Gaussian16 suite [78]. All gold atoms were kept frozen during the geometry optimization. Dunning's correlation consistent cc-pVDZ basis set [79] was used for all the light atoms, and Hay-Wadt pseudopotentials and basis sets were used for S and Au. Dispersion energies were included through Grimme's

D3 method and parameters [58].

The distance between the sulfur atom and the gold cluster was then varied as shown in Figure 3.1 while keeping the molecular geometry fixed, and single-point energies were computed at the same level described above; the basis set superposition error (BSSE) was corrected through the counterpoise method, as implemented in Gaussian16. Note that since the model is meant to describe a covalent bond, both the MMC molecule and the gold slab were considered in a doublet spin state for the computation of the interaction energy.

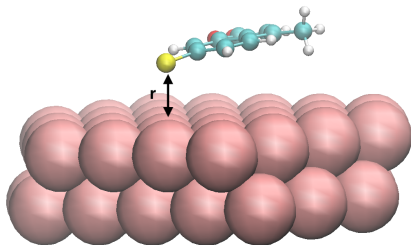


Figure 3.1: Model system used to compute MMC-gold interaction energies; the distance r was varied by keeping everything else rigid.

Au-X pairwise interactions were included in UFF_Au by means of a Lennard-Jones (LJ) curve for every element in the MMC molecule, while no electrostatic charges are explicitly present in the model. For the sulfur atom, two kinds of S-Au interactions were defined; one refers to the covalent bond and involves only one S-Au couple, while the other refers to non-bonded interactions between sulfur and all the remaining gold atoms. We chose the LJ potential over the more commonly used harmonic potential so that molecules would be allowed to leave the gold surface during a MD run, if this lowers the system energy.

For every distance sampled by the geometry scan, an ad hoc python script was used to obtain the force field interaction energies from single-point molecular mechanics calculations performed with the LAMMPS code [80]; the LJ parameters were then updated in an iterative way and fitted to the DFT energies. The trust region reflective algorithm was used for the fitting procedure, as implemented in the *optimize* module of the SciPy library [81]. The python script, here named FFfitpy, is attached in Appendix D.

Figure 3.2 shows the comparison between the DFT energy curve for

the scan and UFF_Au energies, while Table 3.1 contains the optimized LJ parameters for each pair.

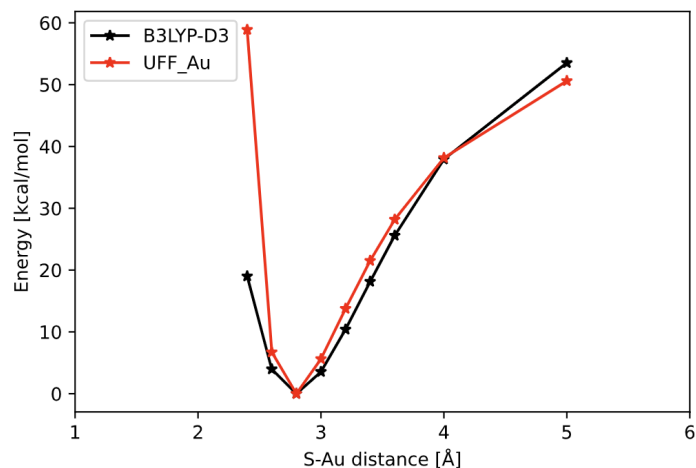


Figure 3.2: Energy curve resulting from the fitting procedure compared to DFT B3LYP-D3 calculations.

Pair	ϵ [kcal/mol]	σ [Å]
Au-C	0.050	3.172
Au-O	0.050	3.024
Au-S (bonded)	24.50	2.495
Au-S (non-bonded)	4.203	3.247

Table 3.1: Lennard-Jones parameters fitted for the UFF_Au set.

Note that an important feature of the curves shown in Figure 3.2 is that the use of a LJ functional form to approximate DFT B3LYP-D3 energies entails an accuracy trade-off. Although both energy curves are governed by the same asymptotic relation in the long-range regime ($E \propto -\frac{1}{R^6}$), the same set of LJ parameters cannot adequately describe the overall shape of the DFT curve. A judicious choice of the distances

sampled by the energy scan is therefore necessary in order to have a good overlap of the DFT and fitted energies in the regions of interest, mainly around the energy minimum. Also, a good agreement at medium to long gold-molecule distances was deemed more important than agreement in the repulsive branch of the curve, since this represents an “energy wall” in any case, the degree of its steepness having limited impact during a MD run.

3.2 GROMOS_Au1

The parameters of UFF_Au were later refined in the development of GROMOS_Au1. The first change that the name chosen for the FF implies is the use of GROMOS (version 54A7) instead of UFF for the description of intramolecular and intermolecular interactions for the organic part. This was done for convenience, since UFF is implemented in the LAMMPS code but not in GROMACS [82]. For the purpose of our simulations, these energy terms are secondary with respect to molecule/gold interactions. In addition, both force fields are able to account for them fairly [76, 83], so this change is expected to have a minor effect on our results. More importantly, different atom types were introduced, while in UFF_Au only one atom type was used for each element. Furthermore, FF parameters were developed also for thiols interacting with gold in an undissociated state (S-H bond intact). As a consequence, the procedure in Section 3.1 was modified as detailed below.

A model for the gold surface was obtained by cleaving a cluster of 60 gold atoms (5x4x3) from the same periodic structure described in Section 3.1.

The molecule set used for the fitting procedure is reported in Table 3.2, together with the atom types and symbols used in the force field. In order to obtain a set of LJ parameters with minimal correlation, the fit was carried out incrementally, determining the optimal FF parameters for one or two atom types at each step.

A geometry optimization was performed at the DFT B3LYP-D3 level, using the Gaussian16 suite, for each molecule in Table 3.2 in close proximity with the surface of the gold cluster. Then, the distance between the molecule and the gold slab was varied incrementally, while keeping the molecular geometry fixed. At each step, the energy of interaction between the molecule and the slab was computed. All *ab initio* calculations were performed at the same level of Section 3.1, but the cc-pVDZ basis set was used for sulfur instead of LANL2DZ.

For benzene and benzenethiol, additional analogous scans were performed while keeping the plane of the aromatic ring perpendicular to

Molecule	Atom types fitted	Atom types symbol
benzene	aromatic C, aromatic H	CA, HA
benzenethiol	S thiol, H in thiol	SH, HB
methylbenzenethiol	sp^3 C, H bound to aliphatic C	C3, H
7-mercapto-4-methylcoumarin	oxygen	O

Table 3.2: Molecule set, atom types and symbols used to fit the GROMOS_Au1 force field parameters.

the gold surface. This geometry is meant to mimic the preferred orientation in densely packed self-assembled monolayers (SAMs), a target system that the force field should describe correctly (see Figure 3.3).

Finally, a rigid energy scan was performed also for MMC as a thiyl radical species, forming a covalent S-Au bond with the gold surface. Contrary to UFF_Au, no special parameters were defined for bonded S-Au pairs; instead, a single set of LJ parameters was used to describe the interaction of the S atom with every Au atom in the slab. The local nature of the S-Au bond is thus partially lost in the model; on the other hand, without the arbitrary definition of S-Au bonded pairs, during MD equilibrations the adsorbed molecules can “glide” on the gold surface, moving laterally to minimize intermolecular repulsions and mimicking a self-assembly process.

The energies resulting from the fit are shown in Figure 3.4, while the optimized force field parameters are reported in Table 3.3.

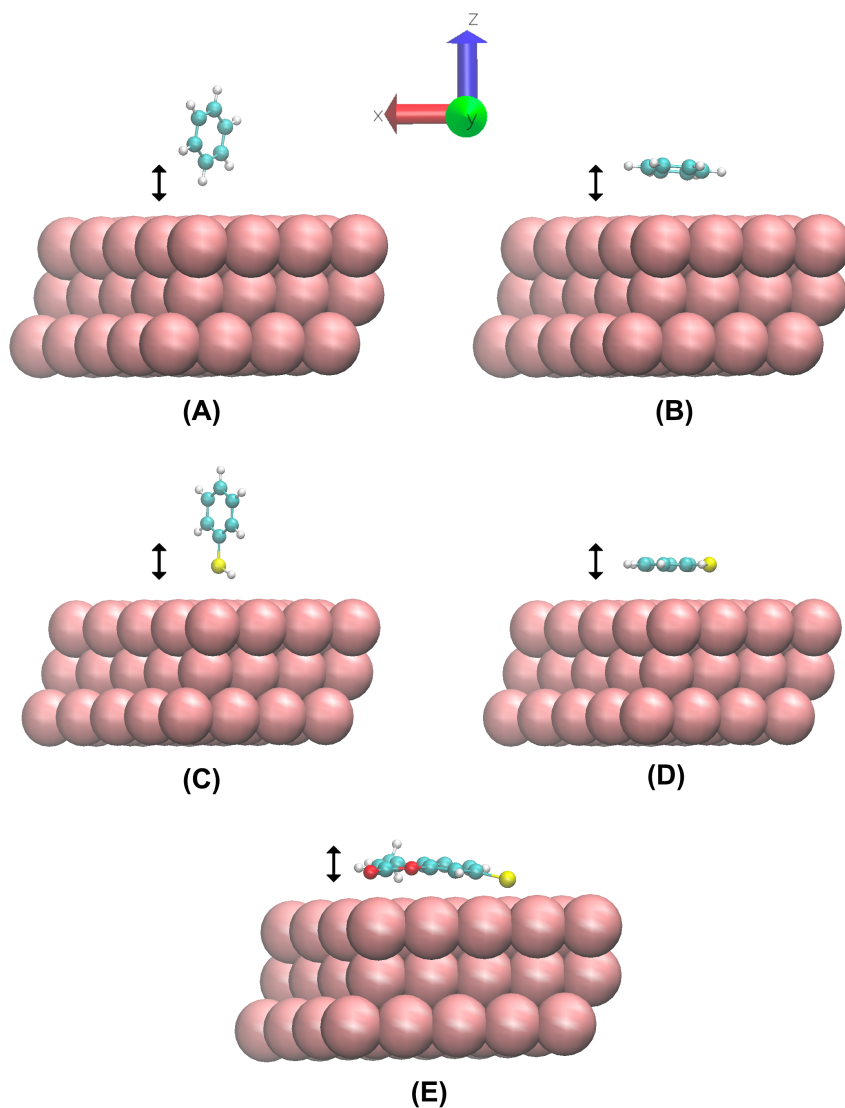


Figure 3.3: Some of the model systems used in the energy scans for the fitting of GROMOS_Au1 force field parameters. The distance of the molecule from the slab was varied incrementally along the z axis, perpendicular to the surface. Benzene in vertical (A) and horizontal (B) configuration; benzenethiol in vertical (C) and horizontal (D) configuration; (E) MMC thiyl.

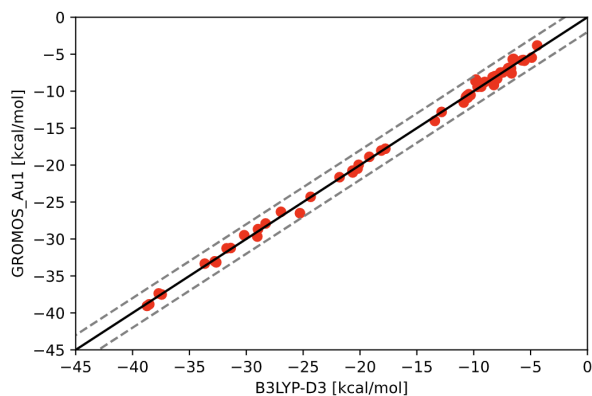


Figure 3.4: FF energies for each atom type fitted to DFT B3LYP-D3 calculations. The black solid line denotes 1:1 correlation, while the grey dotted lines represent the ± 2 kcal/mol mark.

Pair	ϵ [kcal/mol]	σ [\AA]
Au - CA	0.571	3.228
Au - HA	0.020	2.975
Au - SH	1.169	3.243
Au - HB	0.500	2.555
Au - C3	0.775	2.996
Au - H	0.100	3.114
Au - O	0.146	3.435
Au - S (thiyl)	13.406	2.230

Table 3.3: Lennard-Jones parameters fitted for the GROMOS_Au1 set.

3.3 GROMOS_Au2

With the development of GROMOS_Au2, GROMOS_Au1 was further extended to include a greater number of atom types; in addition, a new method for generating model geometries for the fitting procedure was used to better sample the different orientations that thiol molecules adsorbed on gold can adopt during a molecular dynamics simulation.

3.3.1 Obtaining FF parameters

The method used to obtain GROMOS_Au2 is fully described below, while a visual summary of the process is given in Figure 3.5.

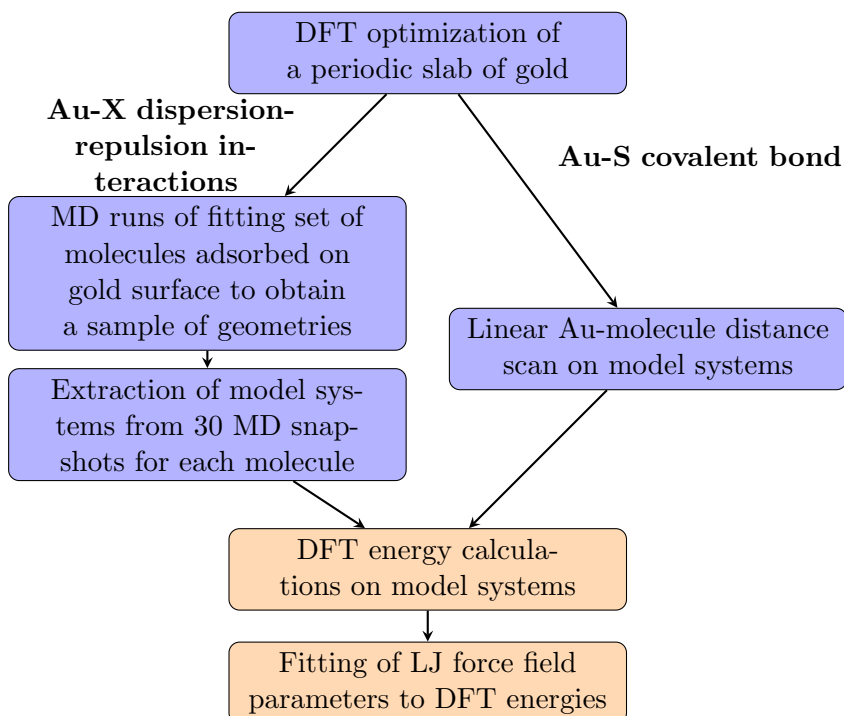


Figure 3.5: General workflow of the force field fitting procedure; the steps of the method for obtaining the model systems are shown in blue, while the steps regarding the fitting itself are shown in orange.

The set of molecules used for the force field fit is shown below in Table 3.4. As for GROMOS_Au1, the procedure was carried out incrementally, and at each step the optimal parameters for one atom type, or at most two, were obtained by following the order in Table 3.4.

For each molecule in Table 3.4 a molecular dynamics (MD) run, composed of 1 ns equilibration and 2 ns production, was performed using GROMACS. The simulations were carried out with a periodic slab

Molecule	Atom types fitted	Atom types symbol
methane, propane	sp^3 C, H bound to aliphatic C	C3, H
benzene	aromatic C, aromatic H	CA, HA
dimethyl ether, oxane	ether O	OE
acetone	sp^2 C, carbonyl O	C2, OC
propanethiol, benzenethiol	S thiol, H in thiol or hydroxyl	SH, HB
propionic acid	O in hydroxyl	OH

Table 3.4: Molecule set, atom types and symbols used to fit the GROMOS_Au2 force field parameters.

of 1728 gold atoms (24x24x3), whose geometry was always kept frozen, and a number of molecules sufficient to form a complete multilayer over the gold surface. All MD runs were performed *in vacuo* in the NVT ensemble at 300 K with a timestep of 1 fs. The bonded parameters and the non-bonded interactions for the organic component were taken from the GROMOS 54A7 force field, while the provisional non-bonded LJ parameters for the molecule–gold slab interactions were adapted from GROMOS_Au1. Bonds to hydrogen were constrained using the LINCS algorithm [84].

From each MD run, thirty equally-spaced snapshots were extracted. These were used to obtain smaller model systems, composed of a cluster of sixty gold atoms (5x4x3) and one molecule chosen randomly among those in close proximity to the surface. A representative set of the model systems thus obtained is shown below in Figure 3.6.

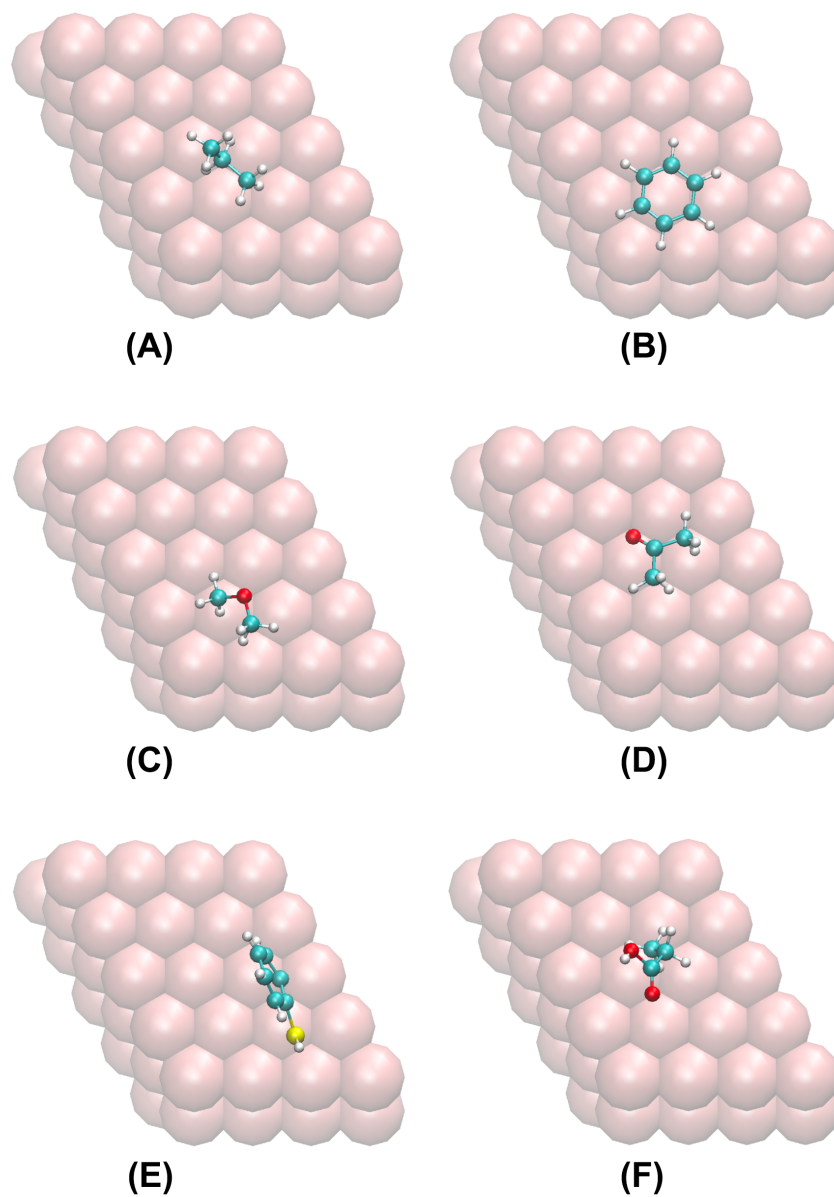


Figure 3.6: Representative model systems extracted from MD runs. Not all molecules from the fitting set are shown. For each molecule, thirty such snapshots were obtained. (A) propane, (B) benzene, (C) dimethyl ether, (D) acetone, (E) benzenethiol, (F) propionic acid.

As in the previous force fields, besides non-bonded Van der Waals interactions, a LJ function was used also to describe the S-Au covalent bond. For this case, model systems were obtained in a way analogous to the one adopted for UFF_Au1 and GROMOS_Au1. One propanethiyl molecule and one benzenethiyl molecule in doublet spin state were optimized at the DFT/B3LYP level in proximity of the upper surface of a cluster of sixty gold atoms (5x4x3); the cc-pVDZ basis set was used for all elements except for Au, for which LANL2DZ was used instead, and dispersion energies were included through the D3 correction. The geometries are shown below in Figure 3.7. Then, a rigid scan was performed by varying the S-Au distance. By following this procedure, we ensured to sample geometries close to the S-Au optimal bond distance, thus improving the force field performance in these configurations. The sulfur atoms covalently bound to gold were treated as different atom types, with different LJ parameters, depending on whether they were bound to aliphatic rather than aromatic carbon atoms. The atom type symbols adopted are, respectively, SL and SR.

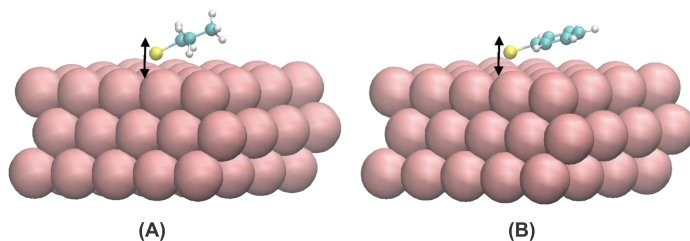
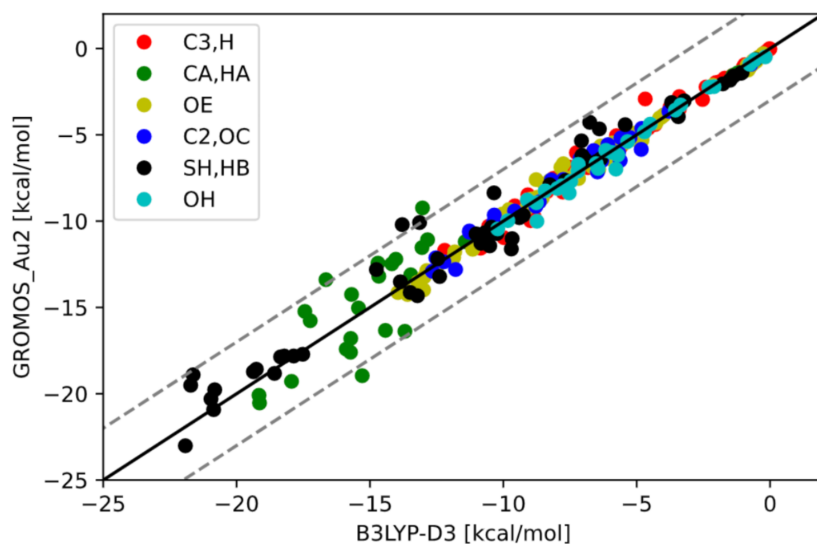


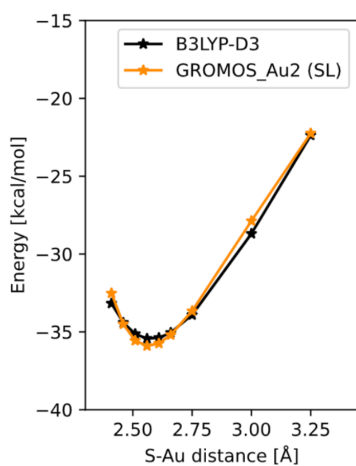
Figure 3.7: Model systems used for the parametrization of the S-Au covalent bond. The S-Au distance was varied while keeping the molecular structure fixed at each step. (A) propanethiyl, (B) benzenethiyl.

DFT molecule-gold interaction energies were then computed at the same level described above and the FFfitpy script (see Appendix D) was used to obtain optimal FF parameters.

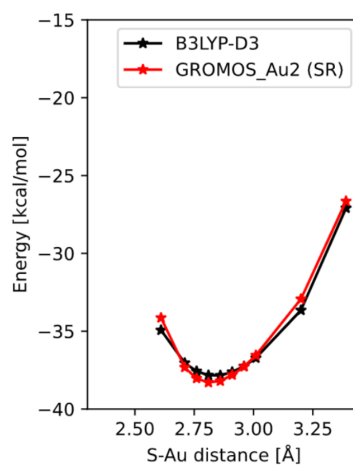
The energies resulting from the fitting procedure are reported below in Figure 3.8. The mean unsigned error (MUE) compared to DFT calculations in Figure 3.8(A) is 0.58 kcal/mol, while the root-mean-square error (RMSE) is 0.95 kcal/mol. Of all the configurations sampled, 95.9% are within ± 2 kcal/mol of a 1:1 correlation with DFT energies, and 98.5% are within ± 3 kcal/mol. All data points beyond this range correspond to model systems containing aromatic compounds; the larger discrepancy is observed even when considering the mean *relative* errors instead of the absolute values. This effect could be due to the greater directionality of the Au-CA interaction, which cannot be reproduced as well by a simple LJ functional form.



(A)



(B)



(C)

Figure 3.8: Energies resulting from the fit of each atom type in the force field, compared to DFT B3LYP-D3 energies. (A) shows the result for all non-bonded terms; the black solid line denotes 1:1 correlation, while the grey dotted lines represent the ± 3 kcal/mol mark. (B) and (C) show the linear scan for the determination of SL-Au and SR-Au bond parameters.

3.3.2 Testing FF parameters on target systems

Since the main molecules of interest for this study are 7-mercapto-4-methylcoumarin (MMC) and 3-mercaptopropionic acid (MPA), the force field parameters derived above were tested on these two molecules. The thiol configuration was tested by comparing FF energies against B3LYP-D3 DFT calculations on clusters extracted from MD runs, exactly as described before. The result is reported here in Figure 3.9.

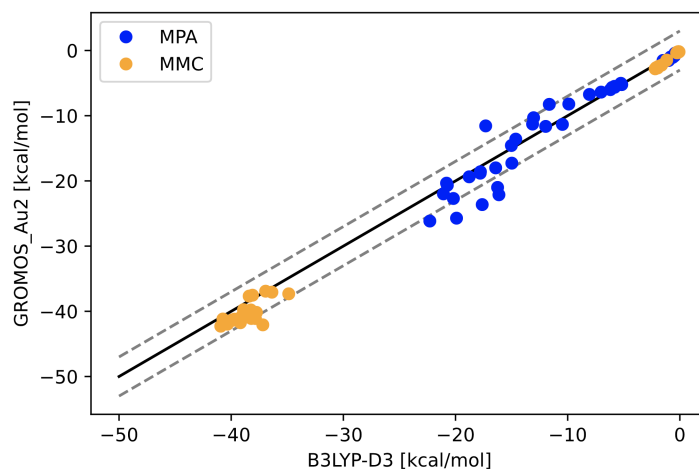


Figure 3.9: The black solid line denotes 1:1 correlation; the grey dotted lines represent the ± 3 kcal/mol mark.

From Figure 3.9 it can be seen how MMC energies seem to suffer from a small systematic error, the FF energies being overestimated by about 1 kcal/mol on average; this effect might be due to some feature of MMC not being captured by the model molecules chosen for the fitting procedure. Other than this, the force field performs similarly as it did for the molecules of the fitting set (overall MUE = 1.30 kcal/mol), and it's deemed to be appropriately accurate.

On the contrary, tests on MMC as a radical species showed a strong disagreement in the energies computed at the GROMOS_Au2 rather than B3LYP-D3 level, see Figure 3.10.

Evidently, the DFT energy of interaction of sulfur atoms involved in covalent bonds with the gold surface depends strongly on the molecular group attached to S. Thus, the benzenethiyl molecule used in the fit is not an adequate model for MMC. Figure 3.10 shows how, by using the S-Au parameter already obtained for GROMOS_Au1, GROMOS_Au2 returns to accurately reproduce the B3LYP-D3 energy curve (with, once again, a slight overestimation of the interaction energy).

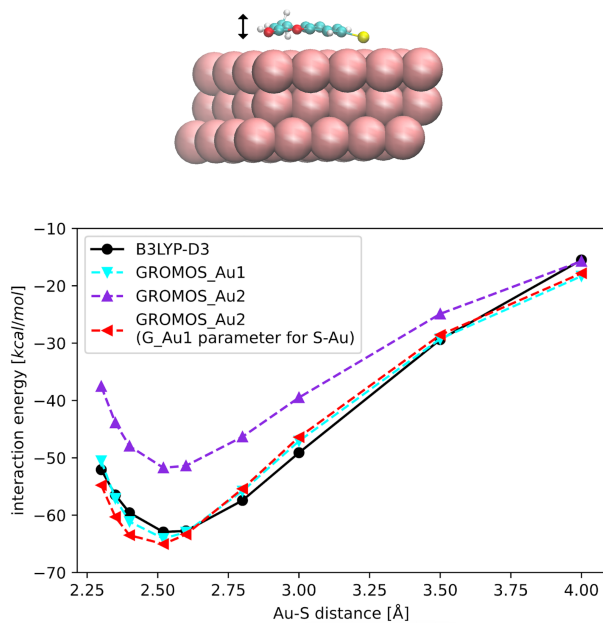


Figure 3.10: Energy scan of radical MMC on a slab of Au(111); the distance from the surface was varied incrementally, and the energy was computed at the B3LYP-D3, GROMOS_Au1, and GROMOS_Au2 levels.

The fact that S-Au parameters could be used without modifications from GROMOS_Au1 shows the ability of the force fields to correctly separate the different energy contributions in the system.

The thyl S-Au parameters in GROMOS_Au2 are thus taken directly from GROMOS_Au1; the whole set is reported below in Table 3.5.

Pair	ϵ [kcal/mol]	σ [Å]
Au - C3	0.192	3.692
Au - H	0.404	2.674
Au - CA	0.435	3.285
Au - HA	0.144	3.013
Au - OE	1.742	2.465
Au - C2	0.415	3.359
Au - OC	1.728	2.282
Au - SH	1.372	3.239
Au - HB	0.352	2.719
Au - OH	1.033	2.881
Au - SL	9.110	2.271
Au - SR	6.669	2.325
Au - SR*	13.406	2.230

Table 3.5: Lennard-Jones parameters fitted for the GROMOS_Au2 set.
 *Taken from GROMOS_Au1.

Chapter 4

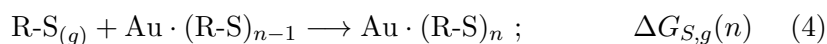
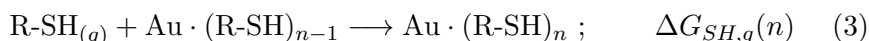
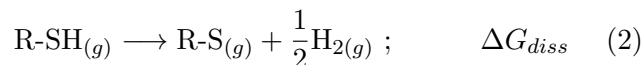
Computational study of MPA and MMC thiol monolayers on Au(111)

The force field parameters described in Chapter 3 were applied to MD simulations of monolayers composed of 7-mercapto-4-methylcoumarin (MMC) or 3-mercaptopropionic acid (MPA) molecules on a periodic Au(111) surface.

As outlined in Chapter 2, the goal of this work was to describe the thermodynamics of formation of the thiol monolayers, and in particular to compute the expected monolayer densities using free energy methods on MD trajectories.

The molecules were simulated both in their undissociated state (thiols) and in their radical (thiyl) state, as both of these kinds of gold-sulfur interface have been proposed in the literature [71–74], and it should be interesting to examine the thermodynamic differences computed from the models depending on the nature of the interface. All molecules included in the simulations are shown in Figure 4.1, while the model of the gold surface is shown in Figure 4.2.

To compute the expected monolayer densities, it's necessary to compute the chemical potential of a molecule in the monolayer as a function of the monolayer density. This is done by evaluating the free energy change associated with each of the reactions shown below:



as the chemical potentials are given by:

$$\mu_{SH}(n) = \Delta G_{desolv} + \Delta G_{SH,g}(n) \quad (4.2a)$$

$$\mu_S(n) = \Delta G_{desolv} + \Delta G_{diss} + \Delta G_{S,g}(n) \quad (4.2b)$$

First, (2) was computed at the DFT level as detailed below in Section 4.1. Then, monolayers of increasing densities were simulated for each species in Figure 4.1 through molecular dynamics (see Section 4.2) and (1), (3), and (4) were obtained from the BAR method as described in Section 4.3. The final results are reported in Section 4.4.

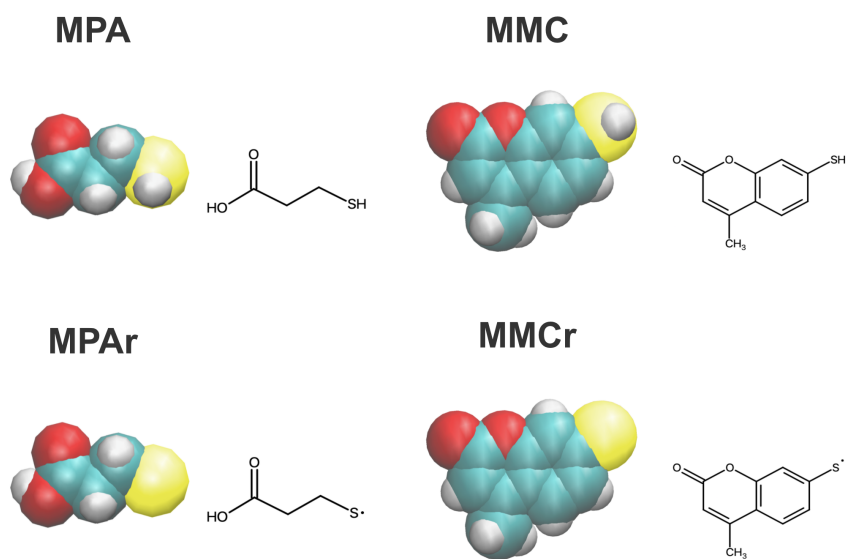


Figure 4.1: Molecules included in the simulations: 3-mercaptopropionic acid in thiol (MPA) and thiyl (MPAr) form on the left; on the right: 7-mercapto-4-methylcoumarin thiol (MMC) and thiyl (MMCr).

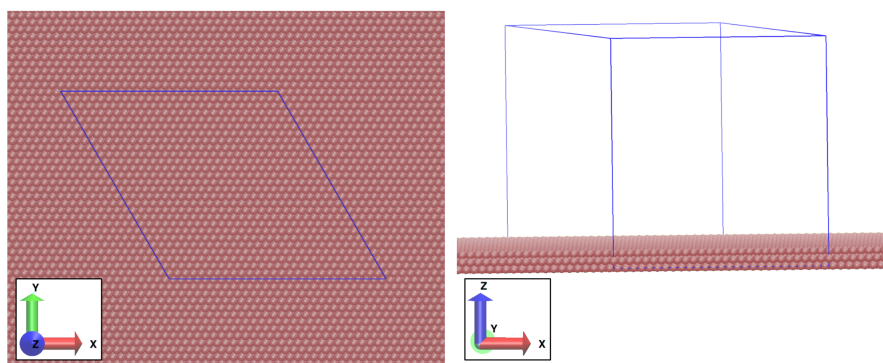






Figure 4.2: Model of the Au(111) surface used for the MD simulations; the periodic boundary is shown in blue.





4.1 DFT free energy calculations of S-H bond dissociation





Since the experimental value of ΔG_{diss} is not available for MMC and MPA, DFT calculations were used to evaluate it. The bond dissociation energies (BDE) of thiol and molecular hydrogen can be directly computed (including semiempirical corrections for dispersion energy and counterpoise corrections for the BSSE), then thermal contributions for enthalpies and entropies are added with Boltzmann averages referred to translational, rotational and vibrational motions. The last term requires the calculation of harmonic frequencies, at the same level as BDE.

Before applying this procedure to MMC and MPA, the performance of different density functionals and basis sets was tested against the experimental values of the dissociation enthalpy (ΔH_{diss}) for some simple aromatic thiols, namely benzenethiol, p-aminobenzenethiol, p-nitrobenzenethiol, and perfluorobenzenethiol, to consider both electron donating and electron withdrawing groups. This served to evaluate the basis set and functional performance for MMC, while for all aliphatic thiols, a good model for MPA, the value reported in the literature is ≈ 146 kJ/mol. We used enthalpies rather than free energies for this trial because the former are more easily available in the experimental literature; once determined the best choice of functional and basis, ΔG_{diss} was computed at the same level for MMC and MPA. The results are collected below in Table 4.1: one can see that none of the tested functionals is able to reproduce all the experimental enthalpies fairly, but the best performance is obtained by B3P86 with the largest basis set, namely cc-pvtz. Then the value of the free energy change of dissociation involving MMC and MPA was computed at the B3P86/cc-pvtz level, obtaining (kJ/mol):

$$\begin{aligned}\Delta G_{diss}^{MPA} &= 120.0 \\ \Delta G_{diss}^{MMC} &= 130.0\end{aligned}$$

B3LYP-D3				
Basis set				
6-31G(d)	90.3	69.3	103.0	97.8
6-31G(d,p)	94.9	73.5	107.4	102.3
6-31+G(d,p)	96.6	76.2	109.6	107.4
6-311G(d,p)	*	77.7	109.8	107.5
6-31G(2d,2p)	99.6	78.3	111.8	106.0
cc-pvdz	92.5	71.6	105.2	100.0
cc-pvtz	*	81.8	115.6	110.4
Exp. ^(a)	113.0	75.3	125.6	134.0

B3P86-D3				
Basis set				
6-31G(d)	98.4	76.9	111.3	104.1
6-31G(d,p)	103.1	81.1	115.8	109.0
6-31+G(d,p)	104.8	83.7	117.9	113.4
6-311G(d,p)	*	86.1	119.0	114.3
6-31G(2d,2p)	108.4	86.5	120.8	112.9
cc-pvdz	*	87.9	121.3	115.1
cc-pvtz	*	89.7	124.5	116.8
Exp. ^(a)	113.0	75.3	125.6	134.0

BLYP-D3				
Basis set				
6-31G(d)	82.0	56.7	94.6	85.2
6-31G(d,p)	86.6	61.0	*	89.9
6-31+G(d,p)	88.7	64.0	*	96.3
6-311G(d,p)	89.4	65.0	101.3	95.6
6-31G(2d,2p)	91.5	66.1	103.5	93.4
cc-pvdz	95.1	69.3	106.2	98.3
cc-pvtz	95.9	70.1	107.7	99.2
Exp. ^(a)	113.0	75.3	125.6	134.0

Basis set	B3LYP-D3	B3P86-D3
6-31G(d)	121.9	129.2
6-31G(d,p)	125.7	133.1
6-31+G(d,p)	125.8	133.3
6-31G(2d,2p)	130.2	138.1
cc-pvdz	129.3	136.3
cc-pvtz	130.0	138.0
Exp. ^(b)	146	

Table 4.1: Enthalpy change (kJ/mol) of S-H dissociation, computed with various various basis sets and functionals, including thermal contributions at 298 K and Grimme's dispersion corrections, and compared to experimental values. (A), (B), (C) Benchmark on various aromatic compounds used as a model for MMC; (D) values computed for MPA.

(a) Experimental bond dissociation enthalpies: ref. [85] for thiols, [86] for H₂. * Convergence problems. (b) Common value in the literature for BDE of aliphatic thiols.

4.2 Molecular dynamics equilibration of the monolayers

Starting geometries were obtained using the PACKMOL code [87] for monolayers of MPA, MPA_r, MMC, and MMC_r (see Figure 4.1) of increasing density. The model of the gold surface is a periodic 27x27 atoms, 3-layers thick Au(111) slab obtained as already described in Chapter 3. The equilibrium geometries and partial charges of all molecules were taken from the ATB website [88].

MD simulations were then performed using GROMACS [82]. After a routine minimization of the energy of the system, simulations were carried out in the NVT ensemble (300 K, thermostat handled by the *sd* integrator included in GROMACS) in the absence of solvent. 500000 steps were simulated with a timestep of 1 fs, for a total of 0.5 ns of simulation. The gold atoms were always kept frozen, and bonds to hydrogen were constrained using the LINCS algorithm [84].

Figure 4.3 shows some examples of the resulting equilibrated monolayers.

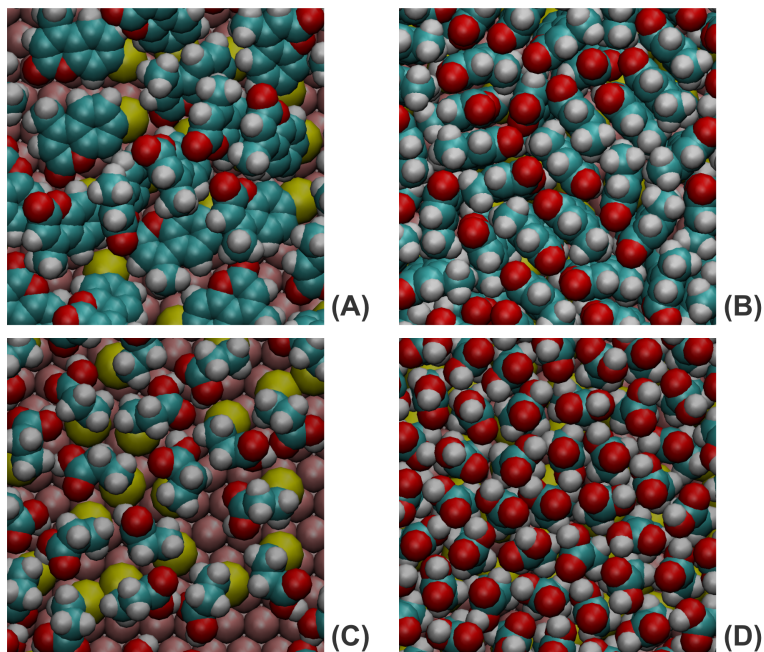


Figure 4.3: Examples of equilibrated partial ($\sim 2.5 \text{ nm}^{-2}$) and complete ($\sim 4 \text{ nm}^{-2}$) monolayers of MMC_r (A,B) and MPA_r (C,D); the images are shown along the z-axis, with a top view perpendicular to the surface.

At higher densities, some molecules leave the monolayer during the MD run to minimize the intermolecular lateral repulsions, given by the

too-crowded monolayer. The expelled molecules remain “floating” on top of the monolayer and form a second layer, as shown in Figure 4.4; note that this behavior has no physical meaning for the radicals MPA_r and MMCr, since it represents the scission of a covalent S-Au bond with the molecule leaving the interface as a free thiyl radical. However this is not relevant, since this event is observed in the simulations only at densities that are too high to be of actual interest, as will be later discussed in Section 4.4.

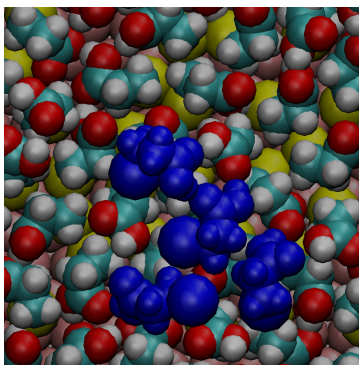


Figure 4.4: High-density MPA_r system in which some molecules (shown in blue) have escaped the monolayer during the MD simulation to form a second layer. The image is shown along the z -axis, with a top view perpendicular to the surface.

Furthermore, it was observed that thiyl molecules can be made to remain in the monolayer by applying restraints during the MD run. A “flat-bottomed” potential was applied to the sulfur atoms; the form of the restraint potential is:

$$V = \frac{1}{2}k[|z_i - Z_i| - r]^2 \theta(|z_i - Z_i| - r) \quad (4.4)$$

where k is the force constant, r dictates the span of the region of zero restraint force, and θ is the Heaviside step function which equals 0 for negative arguments and 1 for positive arguments. The restraint is applied along the z axis, as shown in Figure 4.5, so it depends only on the coordinate z_i of the sulfur atoms and the reference value Z_i .

MD simulations using restraints showed that, at high monolayer densities, the molecules that try to leave the surface do so in the very first steps of the trajectory; if they are prevented from being expelled by a restraint potential, they quickly stabilize and remain in the monolayer for the remaining of the trajectory. To further test this point, several simulations were carried out according to the following procedure: first, 0.5 ns were simulated in the presence of the restraint

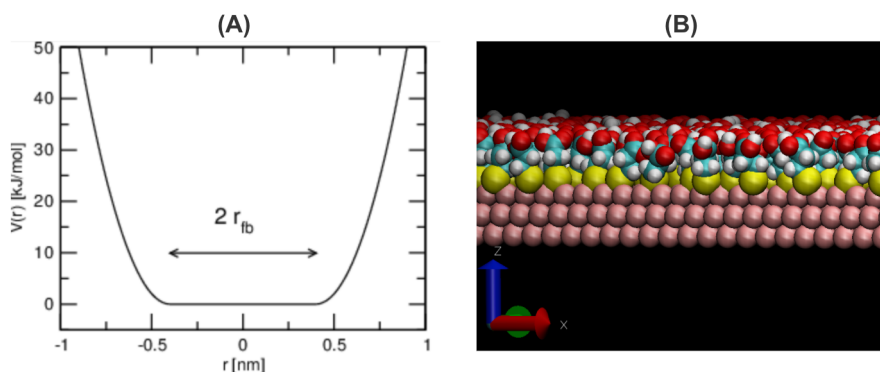


Figure 4.5: (A) shape of the flat-bottom restraint potential used. (B) side view of a MPA monolayer; the restraint is applied on the sulfur atoms along the z axis, preventing the molecules from leaving the gold interface. The values of r and the reference position Z_i were chosen by examining the maximum, minimum, and average values of the z coordinate of sulfur atoms during MD runs on (unrestrained) equilibrated monolayers.

potential; then, an additional 0.5 ns MD run was performed on the equilibrated system without any restraint. This way, it was possible to obtain very dense monolayers that are dynamically stable, meaning that no molecule leaves the monolayer during the whole MD trajectory even when the restraints are eliminated. (Note that, at higher densities still, some thiol molecules *do* leave the surface to form a second layer, but, as already stated, these unstable monolayers have densities much beyond those of experimental interest.)

In this regard, thiol molecules exhibit a different behavior compared to thiols, as highlighted in Figure 4.7.

From Figure 4.7 it can easily be seen how sulfur atoms in radicals are much closer to the gold surface, and at a very specific equilibrium distance. Thiols, on the other hand, exhibit a wider distribution of distances from the slab and are in average more distant. Furthermore, some sulfur atoms of thiol molecules are not in direct contact with the surface (shown in blue in Figure 4.7). When a detachment from the surface is observed, after equilibration, for thiol monolayers (and this happens only at very high densities), the change of configuration is permanent for the duration of the MD run. On the contrary, thiol molecules detach from the surface even at low densities (especially MMC) and can move back and forth, from the first to the second layer, during the simulation. This behavior has been labelled *pseudo*-second layer, and is attributed to the possibility for thiol molecules of min-

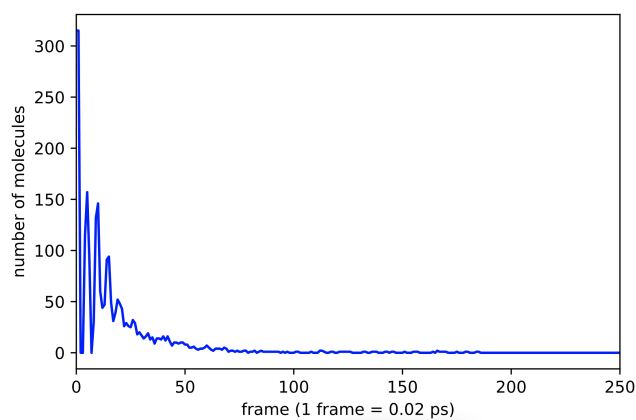


Figure 4.6: Number of molecules outside the restraint boundaries as a function of the MD step. The example considered here refers to MPA_r at the density of 6.5 molecules / nm². Only the first 5 ps of the equilibration are shown; from this point on, no molecule leaves the surface even if the restraints are lifted.

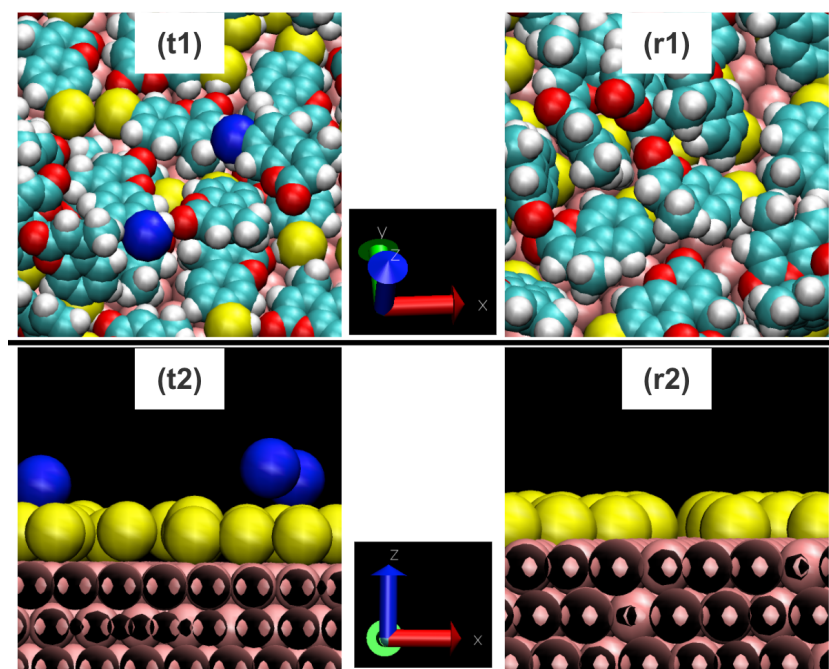


Figure 4.7: Top and side view of a thiol MMC system (t1,t2) and a radical thiol MMC_r system (r1,r2) at the same low density of roughly 2 molecules / nm². The sulfur atoms that are not in direct contact with the gold surface are shown in blue, and can be observed only for the thiols. In (t2) and (r2) only gold and sulfur atoms are shown for clarity.

imizing lateral repulsions throughout the MD run by moving slightly away from the surface when lateral repulsions prevail over the attractive surface-molecule interaction.

At higher densities, thiol molecules also begin to form a *proper* second layer as thiols do. More will be said about this in Section 4.4, where the resulting chemical potentials for all the species are presented.

4.3 BAR free energy calculations from MD trajectories

The MD simulations described in Section 4.2 served as a starting point for free energy calculations. Monolayers of increasing densities of either MPA, MPAr, MMC, or MMCr, were simulated in the NVT ensemble at 300 K, with an equilibration of 100 ps (timestep of 0.5 fs, 200000 steps) followed by a production run of 1 ns (timestep of 1.0 fs, 1000000 steps). As before, gold atoms were kept frozen, and bonds to hydrogen were constrained using the LINCS algorithm [84].

The BAR method (described in Subsection 1.4.4) was applied to the simulations, with a parameter λ altering the intermolecular non-bonded interactions so as to *decouple* one molecule from the monolayer. λ was changed in 40 steps; the first 20 steps gradually put to zero the intermolecular coulomb interactions between the molecule and the surrounding molecules, while the last 20 steps turned off the remaining LJ interactions that comprehend all van der Waals interactions between molecules and all interactions between the molecule and the gold surface.

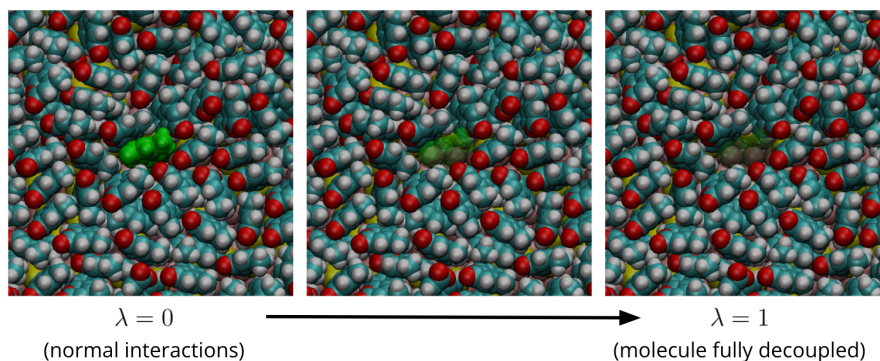


Figure 4.8: Schematic representation of the BAR procedure on a monolayer of MMCr; one molecule (shown in green) was decoupled from the surroundings by scaling the intermolecular interaction through 40 λ steps.

This way, it was possible to compute the free energy of decoupling

of a molecule (or alternatively, with a change of sign, the free energy of *insertion*) as a function of the monolayer density. Each BAR simulation was repeated 5 times with a different random choice of the decoupled molecule, so as to widen the sampling of the microstates considered in the computation. Furthermore, a check on the convergence of the results obtained from the BAR procedure was performed on MMCr, and production trajectories of length 1 ns, 2 ns, 5 ns, and 10 ns, all gave the same result within the error.

Since these free energy values were computed in the NVT ensemble, they correspond to Helmholtz free energies, while all other results in this work are reported as Gibbs free energies. The correction can likely be neglected for the monolayers, which are condensed phases; on the other hand, it can be easily computed for the thiol molecule treated as an ideal gas:

$$\Delta G_{SH/S} = \Delta F_{SH/S} - RT = \Delta F_{SH/S} - 2.5 \text{ kJ/mol} \quad (4.5)$$

The free energies thus obtained correspond to reactions (3) and (4) reported above, respectively ΔG_{SH} and ΔG_S . In Section 4.1 it was already shown how (2), ΔG_{diss} , was computed at the DFT level. To obtain the chemical potentials we are interested in, the last value that needs to be computed is (1), the free energy of desolvation of MPA and MMC in ethanol, the solvent used in experimental conditions for the preparations of the self-assembled monolayer and the washing of the product.

The free energies of desolvation were obtained from BAR simulations in which one molecule of MPA or MMC was decoupled from a periodic box of ethanol in 40 λ steps, as for the monolayers. The box was composed of 2342 molecules of EtOH in the NPT ensemble at 300 K and 1 bar (thermostat handled by the *sd* integrator, pressure coupling method of Parrinello-Rahman). The equilibration run was 200 ps long (timestep of 0.5 fs, 800000 steps), while the production run was 1 ns long (timestep of 0.5 fs, 2000000 steps.)

A correction is needed in the free energy values obtained in order for our simulations to refer to a common thermodynamic state. In the desolvation process just described, a box of 198.3 nm³ is used, so that the concentration of the decoupled thiol molecule is $\rho_1 = 8.4 \times 10^{-3}$ mol/l. For reactions (3) and (4) described above, instead, the volume of the simulation box is 386.8 nm³, so the decoupled molecule has concentration $\rho_2 = 4.3 \times 10^{-3}$ mol/l. The corrective factor can be computed as:

$$\Delta G_{gas} = RT \ln(\rho_2/\rho_1) = -1.7 \text{ kJ/mol} \quad (4.6)$$

Furthermore, it is possible to relate the simulated concentration of thiol in ethanol ($\rho_1 = 8.4 \times 10^{-3}$) with the experimental conditions during the deposition of the monolayer. The three experimental concentrations used, and the relative thermodynamic corrections at $T = 298$ K, are (for more experimental details, see Chapter 5, Chapter 6):

$$\Delta G_{conc} = RT \ln(\rho_1/\rho_{exp}) = \begin{cases} 5.3 \text{ kJ/mol,} & \text{if } \rho_{exp} = 1 \text{ mmol/l} \\ 11.0 \text{ kJ/mol,} & \text{if } \rho_{exp} = 0.1 \text{ mmol/l} \\ 16.7 \text{ kJ/mol,} & \text{if } \rho_{exp} = 0.01 \text{ mmol/l} \end{cases} \quad (4.7)$$

This way, with the assumption that gases and solutions are ideal, we can relate the computed free energies to the experimental conditions. The resulting SAM densities experimentally obtained are all equal within the error (see Chapter 5, Chapter 6), regardless of the initial thiol concentration, and the value $\rho_{exp} = 0.1$ mmol/l will be used for our calculations. The free energy values obtained for the desolvation are thus, with and without these corrective terms:

$$\begin{aligned} \Delta G_{desolv}^{MPA} &= 55.8 + \Delta G_{gas} + \Delta G_{conc} = 65.1 \\ \Delta G_{desolv}^{MMC} &= 17.9 + \Delta G_{gas} + \Delta G_{conc} = 27.2 \end{aligned}$$

4.4 Resulting chemical potentials

The chemical potentials of each species in the monolayers were computed as:

$$\mu_{SH}(n) = \Delta G_{desolv} + \Delta G_{SH}(n) \quad (4.9)$$

for the thiols, while for the thiyls:

$$\mu_S(n) = \Delta G_{desolv} + \Delta G_{diss} + \Delta G_S(n) \quad (4.10)$$

The average value of the chemical potential for each species as a function of the monolayer density is shown below in Figure 4.9

Figure 4.9(A) shows the chemical potential of MMC and MMCr, while Figure 4.9(B) shows the chemical potential of MPA and MPAr. As stated above, the BAR calculation of $\Delta G_{S/SH}$ was repeated 5 different times with a random choice of the decoupled molecule. The orange bars in Figure 4.9 show the standard error of the mean, computed as:

$$sem = \frac{\sigma_x}{\sqrt{n-1}} \quad (4.11)$$

where n is the number of repetitions and σ_x is the sample standard deviation.

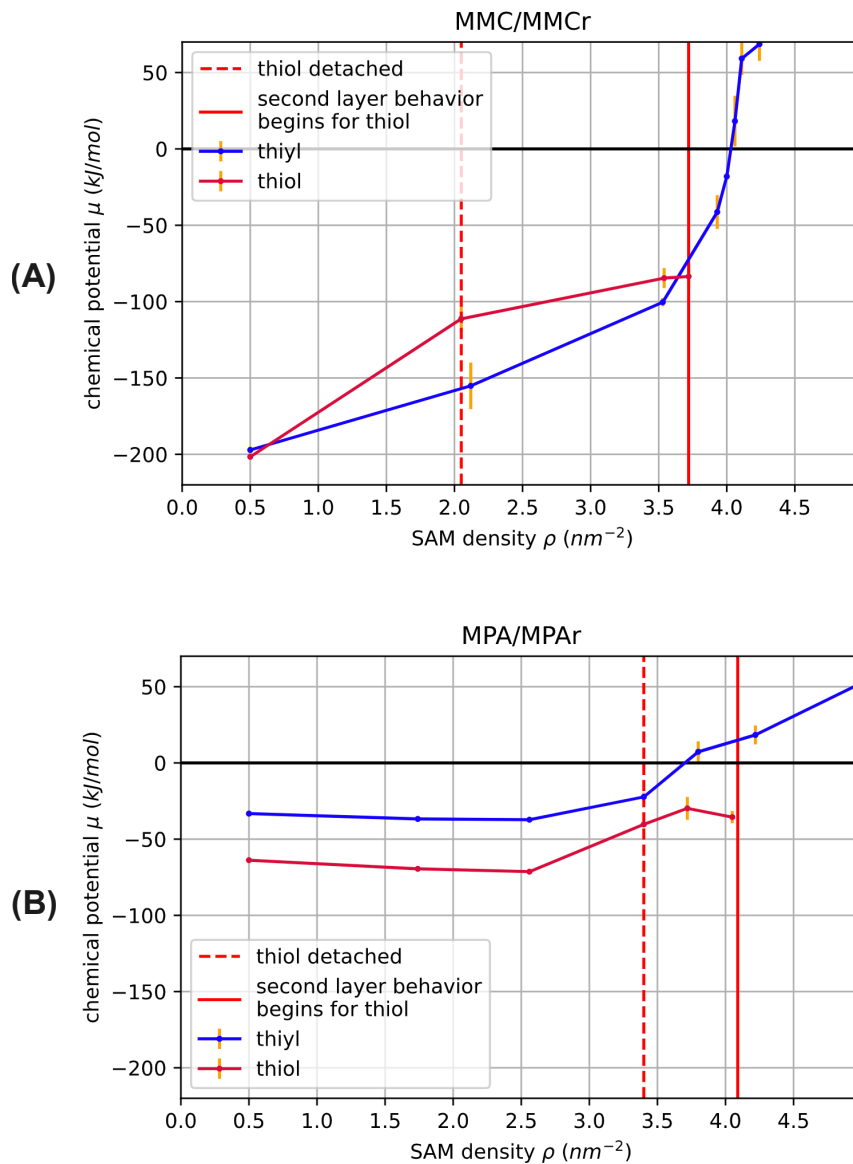


Figure 4.9: Chemical potential as a function of the monolayer density for: (A) thiol MMC and thiy l MMCr, (B) thiol MPA and thiy l MPAr.

The horizontal bold line in Figure 4.9 highlights the equilibrium value of the chemical potential (0); above this line, the monolayer is deemed thermodynamically unstable by the models. It can be seen how the chemical potential of the thiols never reaches the value of

0. This happens for the reason explained in Section 4.2: since thiol molecules can easily leave the surface when the interaction with the surrounding molecules becomes overall repulsive, it is not possible to obtain SAMs at densities as high as for thiyls. The red dotted line in Figure 4.9 shows the density above which thiol molecules begin showing the *pseudo*-second layer behavior described in Section 4.2, while the vertical red solid line demarcates the beginning of *proper* second layer behavior for thiols.

Since (A) and (B) in Figure 4.9 are reported in the same scale, the curves can be directly compared. It can be seen how, at low densities, MMC/MMCr exhibit chemical potentials that are negative and greater in absolute value compared to MPA/MPAr; this comes from the overall stronger binding energies of MMC and MMcR to gold compared to MPA/MPAr. With an increase in the density, though, the chemical potential of MMC/MMCr rises sharply towards positive values, while it remains moderately flat for MPA/MPAr, even becoming slightly more negative, before moving towards 0, with,

$$\left(\frac{d\mu}{d\rho}\right)_{MPA/MPAr} \ll \left(\frac{d\mu}{d\rho}\right)_{MMC/MMCr}$$

This is attributed to the larger number of internal degrees of motion of MPA/MPAr rather than MMC/MMCr. MPA/MPAr molecules are rather flexible, and can easily reorient during the MD run to form favorable hydrogen bonds as shown in Figure 4.10. On the other hand, MMC/MMCr molecules are rigid and can pack favorably only along the specific axis perpendicular to the plane of the aromatic rings (see Figure 4.11), and their interactions tend thus to be repulsive compared to MPA/MPAr.

The same packing dynamics are responsible for the *pseudo*-second layer behavior of some MMC thiol molecules even at low densities (Figure 4.9(A)), while this doesn't happen for MPA (Figure 4.9(B)). The results thus obtained for the computational monolayer densities are (nm^{-2}):

$$\rho_{(MMC)} = 3.7$$

$$\rho_{(MMCr)} = 4.0$$

$$\rho_{(MPA)} = 4.1$$

$$\rho_{(MPAr)} = 3.8$$

Although thiol and thiyl molecules exhibit markedly different behaviors, it is apparent that the thermodynamic monolayer density resulting from the models is quite similar for the two pairs of species,

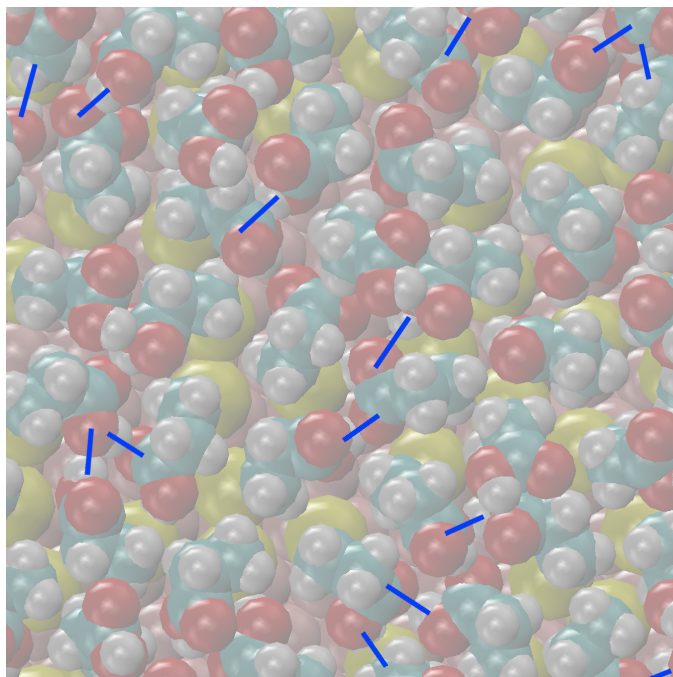


Figure 4.10: H-bonds exhibited by MPA_r at the SAM density of 3.8 molecules / nm² in a representative snapshot from the MD trajectory. The H-bonds are highlighted in blue; they can involve two or more molecules and both the -SH and -COOH functional groups.

meaning: $\rho_{(MMC)} \approx \rho_{(MMC_r)}$, and $\rho_{(MPA)} \approx \rho_{(MPA_r)}$. This is due to the big free energy “penalty” of S-H dissociation that thiol monolayers have over thiol monolayers, lowering their equilibrium density.

While an experimental value for the SAM density of 3-mercaptopropionic acid is not yet available, a comparison between the computed densities and the experimental value for 7-mercapto-4-methylcoumarin is reported here in nm⁻² (experimental procedure detailed in Chapter 6):

$$\rho_{(MMC)} = 3.7$$

$$\rho_{(MMC_r)} = 4.0$$

$$\rho_{exp} = 4.6 \pm 0.6$$

The experimental value is compatible within the uncertainty with $\rho_{(MMC_r)}$, the computed value for MMC_r, while $\rho_{(MMC)}$ is slightly too low to be compatible. However, as already stated above, the difference emerging from the models is quite small to be confidently assigned a physical interpretation.

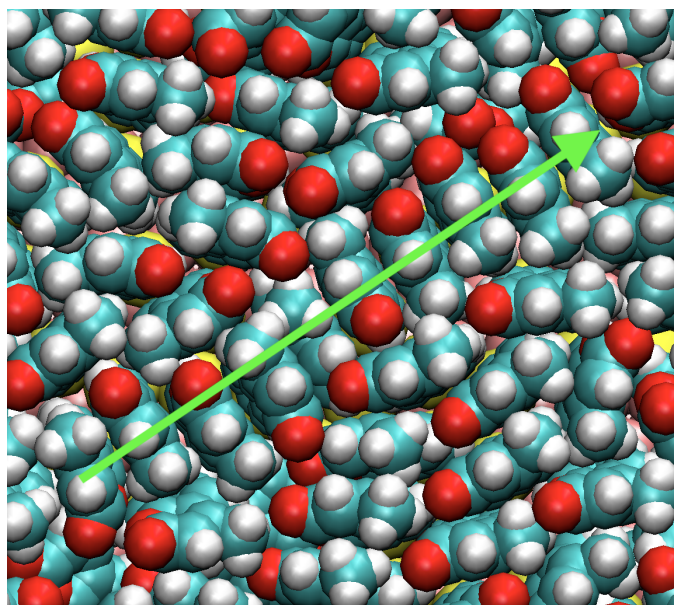


Figure 4.11: MMCr monolayer at the density of 4.84 molecules / nm²; a particularly long row of ordered molecules is highlighted by the green arrow.

Note that the orange bars in Figure 4.9, showing the standard deviation of the mean in the simulations, can be interpreted here as a measure of how dishomogeneous is the monolayer at a specific density. The *sem* would decrease by increasing the number of repetitions (here $n = 5$), approaching zero in the limit; but from the sampling already obtained it can be seen how it is in general true that at higher densities the *sem* is greater, while at lower densities it is lower. This is due to a greater variety of configurations available for molecules packed in high-density monolayers, while at low densities all the molecules tend to stay relatively isolated, each one interacting strongly with the gold substrate and to a lesser degree with the neighboring molecules. The specific differences observed between MMC/MMCr, Figure 4.9(A), and MPA/MPAr, Figure 4.9(B), in this regard are once again ascribed to the difference in internal degrees of freedom between the two molecules, with a greater *sem* observed for MMC/MMCr that is due to relatively repulsive interactions experienced by *some* molecules during the MD trajectories.

These results will be discussed further in the light of possible future calculations in Chapter 7.

Chapter 5

Towards a traceable enhancement factor in surface-enhanced Raman spectroscopy



Cite this: DOI: 10.1039/d0tc04364h

Towards a traceable enhancement factor in
surface-enhanced Raman spectroscopy†

Eleonora Cara,^a Luisa Mandrile,^a Alessio Sacco,^a Andrea M. Giovannozzi,^a Andrea M. Rossi,^a Federica Celegato,^a Natascia De Leo,^a Philipp Hönicke,^b Yves Kayser,^b Burkhard Beckhoff,^b Davide Marchi,^c Alberto Zoccante,^c Maurizio Cossi,^c Michele Laus,^c Luca Boarino^a and Federico Ferrarese Lupi^a

The enhancement factor (EF) is an essential parameter in the field of surface-enhanced Raman spectroscopy (SERS), indicating the magnification of the Raman signal of molecules interacting with the surface of plasmonic nanostructures. The calculation of EF requires a careful evaluation of both the signal intensities and the number of molecules in SERS and normal Raman conditions. The determination of the surface density of molecules adsorbed on the plasmonic substrate is a challenging task, but essential for the estimation of the number of SERS-active molecules. This paper describes the determination of EF using 7-mercapto-4-methylcoumarin (MMC) as the probe molecule on gold-coated silicon nanowires, integrating SERS and normal Raman spectroscopy with X-ray fluorescence (RF-XRF) data that provide a reference-free quantitative measurement of the molecular surface density. In addition, the surface coverage of MMC on the substrate is modelled by molecular mechanics (MM) and molecular dynamics (MD) simulations.

Received 12th September 2020,
Accepted 15th October 2020

DOI: 10.1039/d0tc04364h

rsc.li/materials-c

1 Introduction

Surface-enhanced Raman spectroscopy (SERS) is a vibrational spectroscopic technique belonging to the vast category of plasmon-enhanced molecular spectroscopies (PEMS). The intensity of the fingerprint Raman spectrum of a molecule is amplified by strong local electromagnetic fields on a plasmonic substrate and, to a minor extent, by chemical interaction with its surface.¹ Since its discovery in 1974, SERS has attracted remarkable interest concerning both the fundamental mechanism governing the amplification of Raman scattering and its applications from materials science to biomedicine. Currently, the engineering of electromagnetic hot spots at sub-nanometric spatial resolution^{2–7} allows SERS quantification capability down to the single-molecule regime.

The enhancement factor (EF) is the key parameter for the assessment of SERS substrate performances. A great effort has been addressed to the definition of EF.^{8–10} Eqn (1) reports the most commonly employed definition:

$$EF = \frac{I_{\text{SERS}}/N_{\text{SERS}}}{I_{\text{NR}}/N_{\text{NR}}} \quad (1)$$

where the signal intensities measured in SERS and normal Raman (NR) are I_{SERS} and I_{NR} , respectively. N_{SERS} and N_{NR} represent the number of probed molecules contributing to SERS and normal Raman signals. Over the years, the SERS community chased increasing enhancement capabilities and EFs with values as high as 10^{14} were reported.¹¹ The methods for the determination of I_{SERS} and I_{NR} and the estimation of N_{NR} are well-accepted.¹⁰ In contrast, N_{SERS} represents a critical parameter that depends on the adsorption behaviour of the analytes and the structural characteristics of the substrate.¹⁰ According to ref. 10, N_{SERS} contains several contributions as reported in eqn (2):

$$N_{\text{SERS}} = \mu_{\text{M}} \cdot \mu_{\text{mol}} \cdot A_{\text{M}} \cdot A_{\text{eff}} \quad (2)$$

where μ_{M} is the number of SERS-active nanostructures per unit area on the substrate, μ_{mol} is the surface density of the molecules adsorbed on the metal, A_{M} is the metallic surface area of an individual nanostructure and A_{eff} is the effective area of the confocal scattering volume of the probing laser. Among these contributions, the determination of μ_{mol} is particularly

^a Istituto Nazionale di Ricerca Metrologica (INRiM), Strada delle Cacce 91, 10135 Torino, Italy. E-mail: e.cara@inrim.it

^b Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany

^c Dipartimento di Scienze e Innovazione Tecnologica, Università del Piemonte Orientale (UPO), Via T. Michel 11, 15100 Alessandria, Italy

† Electronic supplementary information (ESI) available: Evaluation of the scattering volume, description of the geometrical considerations on the molecular surface density, determination of the force field for MM and MD simulations, description of the reference free XRF experiment, uncertainty evaluation for the metallic surface area and other uncertainty budgets presented in this manuscript. See DOI: 10.1039/d0tc04364h



challenging.¹⁰ In the especially favourable case of electroactive analytes, a direct evaluation of the molecular surface coverage can be performed through electrochemical methods.^{12,13} However, in general, the evaluation of μ_{mol} is based on geometrical assumptions with little or no information concerning the adopted model.^{14–22} This is due to the lack of suitable analytical techniques offering traceable molecular quantification thus, in turn, preventing inter-laboratory comparability that is essential to push the progress of SERS applications.^{8,9,23,24}

This work reports on the determination of EF for the model system consisting of 7-mercapto-4-methylcoumarin (MMC) as the probe molecule on gold-coated silicon nanowires, by combining SERS and normal Raman measurements with reference-free X-ray fluorescence (RF-XRF) to estimate the molecular surface density. The surface coverage of MMC on the substrate is further modelled by molecular mechanics (MM) and molecular dynamics (MD) simulations. A comparison among the EFs so calculated with those estimated using different approaches is also carried out, thus clearly highlighting the extremely high sensitivity of EF to the μ_{mol} parameter.

2 Results and discussion

2.1 Structure of the SERS substrate

A three-dimensional SERS substrate was fabricated by a combination of nanospheres lithography (NSL) and metal-assisted chemical etching (MACE) leading to high-aspect-ratio silicon nanowires ordered over large areas. Then, the nanowires were coated with gold by e-beam evaporation to form plasmonic caps on their tops as shown in Fig. 1a. The morphology of the nanowires was characterized by scanning electron microscopy

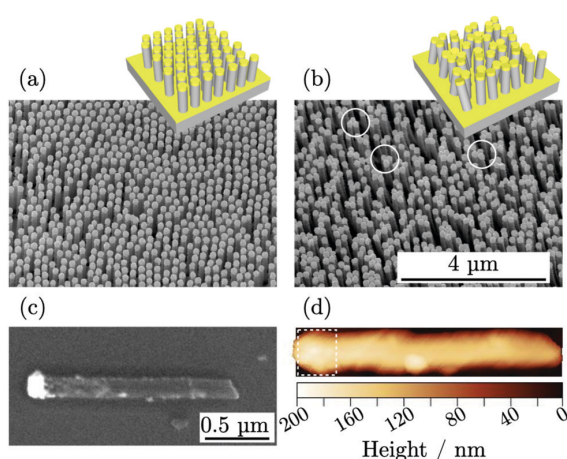


Fig. 1 (a) SEM micrograph of the SERS substrate consisting of an ordered matrix of three-dimensional gold-coated nanowires. (b) The nanowires, bending and forming bundles and hot spots at the tip-to-tip sites, are highlighted by the white circles. (c) SEM micrograph of a single nanowire lying horizontally. (d) AFM topographic map of the nanowire used to determine the gold surface area in the top cap (highlighted by the white dashed line).

(SEM), obtaining diameter of (155 ± 15) nm, lateral spacing of (250 ± 4) nm, and height of the top gold cap of (100 ± 9) nm. The top-view SEM micrographs were processed to identify the single objects and to determine the number of nanowires per unit area, $\mu_{\text{M}} = 14.8 \mu\text{m}^{-2}$, useful for N_{SERS} calculation through eqn (2). The nanowires presented an aspect ratio of 10:1 and exhibited high flexibility forming bundles and generating hot spots at the tip-to-tip sites.^{19,20,25} Fig. 1b shows a SEM micrograph of the substrate where some hot spots are highlighted with white circles.

Atomic force microscopy (AFM) was used to characterize the metallic surface area A_{M} on single nanowires spread horizontally on the solid substrate as shown in Fig. 1d. The analysis was limited to the apical part of the nanowires coated with gold, visible in the SEM image in Fig. 1c. Few metallic nanoparticles located on the length of the nanowires were observed but were not accounted for in the metallic surface area evaluation. The surface area for single nanowire, contributing to eqn (2), was estimated $A_{\text{M}} = (0.52 \pm 0.09) \mu\text{m}^2$.

2.2 SERS experiment

For the SERS experiment, thanks to the specific interaction of MMC molecules with Au surface, the substrate with standing nanowires was incubated in a solution of MMC in ethanol for 120 minutes and then the unbound MMC was removed by rinsing the substrate with ethanol. Finally, the nanowires were induced to bend by soaking the substrate in water and letting it evaporate so as to trap the molecules inside the hot spots regions.²⁵ Fig. 2a shows a graphical scheme of the bending gold-coated nanowires covered with MMC.

The substrate was probed with a 780 nm laser beam focalized on the tips of the nanowires by a $20 \times$ long working distance microscope (LWD) objective. Under these conditions, the effective area of the excitation volume,¹⁰ included in eqn (2), was calculated to be $A_{\text{eff}} = 4.4 \mu\text{m}^2$ and includes 65.1 nanowires. The vibrational peak of MMC at 1593 cm^{-1} , assigned to the conjugated -C=C- symmetric stretching, was measured to find the value of $I_{\text{SERS}} = (6.6 \pm 0.7) \times 10^4$

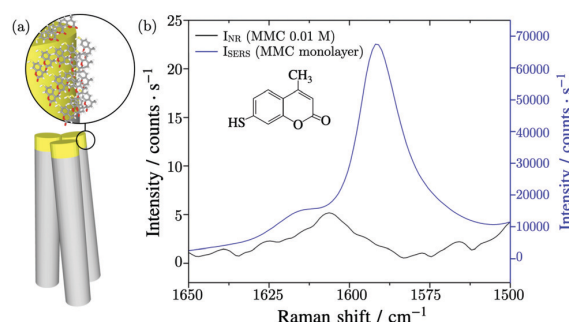


Fig. 2 (a) Graphical scheme of a bundle of nanowires where the gold-coated parts are covered with a layer of MMC probe molecules. (b) Normal Raman spectrum of MMC 0.01 M solution and SERS spectrum of MMC collected on the SERS substrate, acquired with 780 nm excitation laser, 8 mW laser power, $20 \times$ LWD objective.



counts per s, as shown in Fig. 2b. The intensity of the normal Raman signal, obtained by performing the measurements in a MMC solution with a concentration $c_{\text{NR}} = 10$ mM, was $I_{\text{NR}} = (5.2 \pm 0.3)$ counts per s, as shown in Fig. 2b. Normally, the Raman intensity of a signal is calculated as the sum of the counts under a given band in the spectrum, however, in this case the nearby signals limit this approach, because their fitting and deconvolution add variability and subjectivity in the peak analysis and determination of the peak integral. In this study, the peak height was employed instead because the ratio of the same band, varying by a proportionality factor in the two conditions (SERS and non-amplified Raman), is calculated in the EF. The number of Raman active molecules contributing to I_{NR} was calculated as $N_{\text{NR}} = c_{\text{NR}} \cdot V = (2.2 \pm 0.1) \times 10^9$, where V is the value of the laser-probe interaction volume $V = (364 \pm 20) \mu\text{m}^3$, determined experimentally through the dimension of the focal depth^{26,27} (further details in the ESI file†).

Then, the determination of the number of molecules contributing to the SERS signal requires the molecular surface density μ_{mol} . This quantity was estimated by different approaches.

2.3 Determination of the molecular surface density by geometrical considerations

Various geometrical assumptions are carried out in literature to estimate the number of molecules contributing to the SERS signal, considering in general the molecules adsorbed in a monolayer or submonolayer. To test the reliability of such an approach, we evaluated the monolayer density with some simple geometrical models. First, using the tabulated molecular weight and density of MMC, we computed the molar volume $V_{\text{mol}} = (148 \pm 3) \times 10^{12} \mu\text{m}^3 \text{mol}^{-1}$ and the volume of a single molecule $V_{\text{MMC}} = V_{\text{mol}}/N_{\text{A}} = (2.45 \pm 0.05) \times 10^{-10} \mu\text{m}^3$, where N_{A} is the Avogadro constant. Considering the molecules as spherical objects, as the simplest approximation, their circular area projected on the surface was estimated $A_{\text{MMC}} = (4.7 \pm 0.1) \times 10^{-7} \mu\text{m}^2$ (Fig. S2a, ESI†) and the surface density, *i.e.*, the reciprocal of A_{MMC} , resulted in $\mu_{\text{MMC}/\text{geom1}} = (2.1 \pm 0.1) \times 10^6 \mu\text{m}^{-2}$. This is the maximum surface density that can be obtained when no specific information about the molecular shape and orientation in space is available, and a sphere is used to roughly approximate the actual molecular hindrance.

A second geometrical model considered the MMC molecule, with an end-to-end length of 0.73 nm, freely rotating around the thiol group bound to the gold surface (Fig. S2b, ESI†). In this way, a single molecule spans an area of $1.7 \times 10^{-6} \mu\text{m}^2$ leading to a density of $\mu_{\text{MMC}/\text{geom2}} = 0.6 \times 10^6 \mu\text{m}^{-2}$. An alternative description was based on an ordered arrangement of MMC molecules, as close as possible given the gold-gold distances on the Au(111) surface and a reasonable estimate of atomic van der Waals radii. This approach, illustrated in the ESI file,† leads to maximum densities of $\mu_{\text{MMC}/\text{geom3}} = 3.6 \times 10^6 \mu\text{m}^{-2}$ or $\mu_{\text{MMC}/\text{geom4}} = 5.0 \times 10^6 \mu\text{m}^{-2}$, depending on the lateral or vertical arrangements assumed by the MMC in the monolayer, respectively (see the ESI file†).

2.4 Determination of the molecular layer density by MM and MD simulations

To gain more insights into the MMC layer structure and interactions, we performed molecular mechanics (MM) and molecular dynamics (MD) theoretical simulations, based on suitable force fields, as described in the Experimental details section. To evaluate the most plausible density, several monolayers were optimized, with N MMC molecules anchored to the gold surface, with $N = 1, 9, 18, 24, 27$ (Fig. 3a–c). The nature of the S–Au chemical bond in this kind of interface is still debated. Most models assume that the thiol hydrogen is lost giving rise to a covalent sulphur-gold bond,²⁸ while others propose that the thiol group is undissociated and a donor-acceptor bond is formed between –SH and Au.²⁹ Here, we adopt the former model, so that the MMC molecules lose their thiol hydrogen homolytically before anchoring to the surface.

The MM energies, referred to the bare surface and N isolated molecules, are plotted in Fig. 3d with respect to the molecular layer density. The minimum energy is found for $\mu_{\text{MMC}/\text{simul}} = 4.3 \times 10^6 \mu\text{m}^{-2}$, corresponding to a monolayer of 23 molecules packed on the model surface. During the optimization of the monolayers with 24 or 27 MMC molecules, one or three molecules were expelled from the organic layer, respectively, returning to the most favourable density. Clearly, above the density $\mu_{\text{MMC}/\text{simul}}$, the intermolecular repulsion in the crowded monolayer prevails on the MMC/surface attractive interaction. This result was refined by adding kinetic energy contributions, with MD runs at 298 K performed on all the previously optimized monolayers, to verify whether the thermal motions could induce some other MMC to leave the surface. All the monolayers, however, were found stable at this temperature, as no molecules were detached from the surface during the 3 ns simulations. Then the most stable monolayer density is expected to be $4.3 \times 10^6 \mu\text{m}^{-2}$. It is worth noting that this value is only slightly lower than the maximum density ($\mu_{\text{MMC}/\text{geom4}}$)

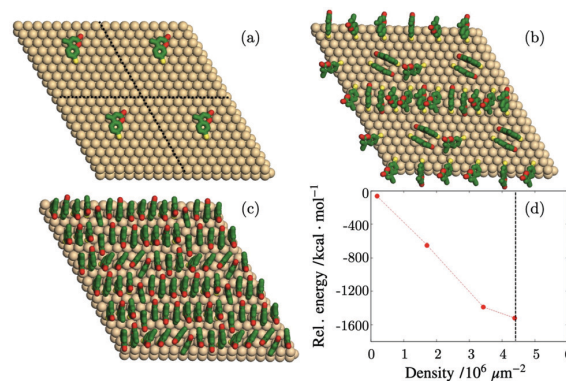


Fig. 3 Results of the MM optimizations. Structure of the monolayers with (a) 1, (b) 9 and (c) 23 MMC molecules on the gold surface unit. Four repeated units are shown for each system, and hydrogen atoms are not displayed for clarity. Yellow, S; green, C; red, O; light brown, Au. (d) Plot of the MM potential energies with respect to the bare surface and N isolated MMC molecules, at different monolayer densities.



Paper

obtained by considering only geometrical restraints in an ordered packing of MMC, as mentioned above.

2.5 Determination of the molecular surface density by X-ray fluorescence

The experimental quantification of the surface density of MMC was performed by the reference-free X-ray fluorescence to target the sulphur atom present in the molecules. RF-XRF offers a well-established tool which requires little or no sample preparation, for non-destructive and quantitative characterization of virtually any material for the determination of the elemental composition, layer thickness, mass deposition and contamination.³⁰ With the atomic fundamental parameters (FPs) method proposed by Sherman,³¹ the sulphur K fluorescence line photon count $P_{S,K}$ can be converted to its mass per unit area σ_S , expressed in g cm^{-2} . This can be obtained by using the physical modelling of the characteristic fluorescence radiation emitted by the sample while including atomic FPs^{32–34} and physically-calibrated instrumentation. The calculation for σ_S is reported in eqn (3):

$$\sigma_S = \frac{1}{k} \cdot P_{S,K} \cdot N \cdot \frac{1}{\varepsilon_{(E_{S,K})} \cdot \tau_{S,K}(E_0) \cdot \omega_{S,K}} \quad (3)$$

where the FPs are the fluorescence yield $\omega_{S,K} = 0.08038$ of the K shell of sulphur, which gives the probability of radiative de-excitation, and the partial photoionization cross-section $\tau_{S,K}(E_0) = 1901.26 \text{ cm}^2 \text{ g}^{-1}$,³⁵ which gives the probability that an incident X-ray photon of energy $E_0 = 2.5 \text{ keV}$ removes one of the K-shell electrons. Eqn (3) also includes the physically calibrated instrumental parameters which are, the efficiency of the radiometrically calibrated fluorescence detector $\varepsilon_{(E_{S,K})} = 0.988$ at the energy of the sulphur fluorescence line $E_{S,K}$ and the normalisation factor accounting for the angle of incidence, the incident photon flux ϕ_0 and the solid angle of detection $\Omega/4\pi$.³⁴ The adimensional factor $k = (10.2 \pm 0.3)$ is the ratio between the effective area offered to the molecules to be absorbed to and the flat area in the absence of a nanostructure. It rescales the mass deposition by considering that the detected fluorescence radiation originated from a three-dimensional region rather than a flat one.

The RF-XRF measurements were performed on the substrate without MMC (NW_{bg}) and on the substrate on which I_{SERS} was measured. The latter substrate was named NW_{MMC} . NW_{bg} was used to measure the background content of sulphur that has to be subtracted to the signal of NW_{MMC} . The two substrates were probed with monochromatised synchrotron radiation, as schematically represented in Fig. 4a, to excite their fluorescence radiation (Fig. 4b). The content of sulphur per unit area on NW_{MMC} is $\sigma_S(\text{NW}_{\text{MMC}}) = (3.9 \pm 0.5) \times 10^{-8} \text{ g cm}^{-2}$, while the background content evaluated on the non-functionalized substrate was $\sigma_S(\text{NW}_{\text{bg}}) = (0.60 \pm 0.08) \times 10^{-8} \text{ g cm}^{-2}$. Thus, the amount of sulphur per unit area results $\sigma_S = (3.3 \pm 0.5) \times 10^{-8} \text{ g cm}^{-2}$. Absorption effects within the Si nanowires and the thin Au layer were considered to be negligible. From σ_S , the number of sulphur atoms per unit area can be derived from the

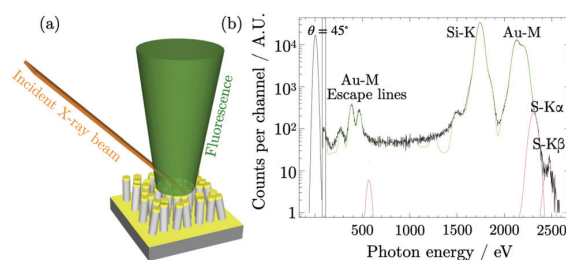


Fig. 4 (a) Schematic representation of the RF-XRF measurement where the incident X-ray beam excites the fluorescence radiation. (b) Deconvoluted fluorescence spectrum, from which the photon count for sulphur K fluorescence lines is extracted and converted to a sulphur mass per unit area, related to the presence of MMC probe molecules.

following relation $\mu_S = \sigma_S \cdot N_A / w_S$, where w_S is the atomic weight of sulphur and N_A is the Avogadro constant. Given that to each atom of sulphur there corresponds one molecule of MMC, μ_S is equal to $\mu_{\text{MMC/XRF}} = (6.1 \pm 1.0) \times 10^6 \mu\text{m}^{-2}$.

The XRF elemental specificity to target the sulphur present in MMC along the nanostructure and quantitatively measure its mass per unit area, which was related to the surface density of MMC molecules. The thiols in MMC are bound preferentially to the gold surface and it is common to assume that any unbound excess is removed from the silicon surface by rinsing the substrate. This assumption could be further verified by performing the quantification of MMC on silicon nanostructures prepared without the gold features. However, the residual MMC on silicon, if any, was neglected. The values of the molecular surface density obtained with different geometrical methods, reported in Table 1, exhibited large differences with respect to $\mu_{\text{MMC/XRF}}$, varying from 1.2 to 10 times its value. Among the geometrical calculations, it is possible to observe that the packing density of vertically arranged molecules is larger than the density of rotating molecules by a factor 8.4, marking the largest difference among the geometrical results.

The MD simulations could be used to interpret these values. Both the assumptions leading to $\mu_{\text{MMC/geom1}}$ and $\mu_{\text{MMC/geom2}}$ correspond to energetically inefficient configurations of submonolayer coverage where the molecules are isolated, respectively laying flat on the gold surface and rotating around the anchor group. On the other hand, the density $\mu_{\text{MMC/geom3}}$ and $\mu_{\text{MMC/geom4}}$ correspond to a closely packed monolayer of molecules with the same occupancy but different spatial arrangement, laterally or vertically oriented molecules, respectively. The value of $\mu_{\text{MMC/geom4}}$ corresponds to the maximum theoretical density of the molecules monolayer. The surface density $\mu_{\text{MMC/simul}}$, obtained from theoretical simulations including the lateral interaction of the molecules, corresponds to a compact self-assembled monolayer (SAM) presenting strong interactions and average orientation which is intermediate between the vertical and lateral alignments.

2.6 Calculation of the enhancement factor

The calculation of the EF requires the value of N_{SERS} as in eqn (2), for which the surface area of the gold caps was used



Table 1 Table comparing the values of the molecular surface density μ_{MMC} , the number of SERS molecules and the enhancement factor for the same substrate calculated by four different geometrical calculations, theoretical simulations and RF-XRF experimental measurement

	$\mu_{\text{MMC}} [\mu\text{m}^{-2}]$	N_{SERS}	EF
Geom. calc. 1 circular area	$(2.1 \pm 0.1) \times 10^6$	$(0.7 \pm 0.2) \times 10^8$	$(3.9 \pm 1.0) \times 10^5$
Geom. calc. 2 rotating molecule	0.6×10^6	$(0.20 \pm 0.04) \times 10^8$	$(14 \pm 3) \times 10^5$
Geom. calc. 3 lateral arrangement	3.6×10^6	$(1.2 \pm 0.3) \times 10^8$	$(2.3 \pm 0.5) \times 10^5$
Geom. calc. 4 vertical arrangement	5×10^6	$(1.7 \pm 0.4) \times 10^8$	$(1.7 \pm 0.4) \times 10^5$
Theoretical simulations	4.3×10^6	$(1.4 \pm 0.3) \times 10^8$	$(1.9 \pm 0.5) \times 10^5$
RF-XRF	$(6.1 \pm 1.0) \times 10^6$	$(2.0 \pm 0.5) \times 10^8$	$(1.4 \pm 0.4) \times 10^5$

together with the nanostructure density, effective laser excitation area and MMC surface density μ_{MMC} determined by the different approaches. The outcomes are reported in Table 1. It was then possible to evaluate the enhancement factor by combining the signal intensities and number of molecules as in the eqn (1), these values are also reported in Table 1. The comparison among the resulting EFs is independent of the surface area. Different estimations of the molecular surface density affected the calculation of N_{SERS} and, consequently, the enhancement factor resulting in significant variability of its value. The value of the EF_{geom1} and EF_{geom2} are not comparable within the uncertainty with each other, neither with the EF_{geom3} and EF_{geom4} nor with EF_{simul} and EF_{XRF} . The largest difference among the reported values is seen as a consequence of the assumptions of the sub-monolayer coverage. The RF-XRF experimental measurements led to the estimation of a compact monolayer coverage of MMC on gold and to EF_{XRF} , in agreement with EF_{simul} within the uncertainty, thus supported by molecular dynamics computational calculations. For the compact monolayer EF_{geom3} and EF_{geom4} were found comparable with each other within the uncertainty, because corresponding to the same assumption on the occupancy of the MMC molecule but different spatial arrangement. Of these two, the EF calculated in the case of vertically-arranged molecules (geom4) in a compact monolayer is comparable within the uncertainty with both EF_{simul} and EF_{XRF} . The quite large relative uncertainty associated with EF_{XRF} should be read as an inclusive estimation of all the relevant contributions to the uncertainty budget rather than as a lack of precision. A complete description of the derivation and uncertainty budget analysis for these quantities is reported in the ESI file.†

The proposed RF-XRF method constitutes a suitable solution for the experimental determination of the molecular surface density which is performed with fully calibrated instrumental setup and physical traceability to the International System of Units. This solid method fills the absence of experimental analytical techniques supporting the calculation of the SERS enhancement factor.

3 Conclusions

In summary, we proposed reference-free X-ray fluorescence as a viable experimental methodology to evaluate the surface density of molecules adsorbed on a surface. This absolute quantitative information is fundamental to estimate the number of active molecules contributing to the enhanced Raman signal on

a plasmonic substrate and, for this reason, strongly impacts the estimation of the enhancement factor. Despite relying on a specialized synchrotron radiation facility with radiometrically calibrated instrumentation, RF-XRF could be extended to commonly available XRF laboratory setup with the use of proper reference standards, so that an accurate molecular quantification could be widely implemented. This analytical methodology for the quantification of SERS-active adsorbates complies with the necessity to obtain a more reliable EF evaluation with high accuracy. It could be extended to other molecules or common analytes for SERS, independently of the presence of a thiol group. In fact, RF-XRF could be used to quantify any substance in which a convenient target element, with atomic number higher than 5, could be identified. In addition to this, we stress the importance to include a thorough uncertainty analysis in the characterization of the EF as well as the complete description of the experimental or theoretical adopted methods.³⁶ Reporting the relevant parameters systematically and adopting standardized methodologies would promote the comparison among different nanostructured systems and boosting the applicability and progress of SERS.

4 Experimental details

4.1 SERS substrate: fabrication and characterisation

The details on the fabrication protocol are reported in ref. 37. The SEM measurements, performed with FEI Inspect F field emission gun SEM (FEG-SEM), and uncertainty evaluation were carried out in comparison to a calibrated reference standard.³⁸ The AFM height micrographs were acquired with a Bruker Multimode V AFM in tapping mode with resonance frequency of 88.8 kHz. The micrographs were analysed by Gwyddion freeware.³⁹ The surface area was extracted by means of a built-in statistical tool. Additional details on the surface area characterization are reported in the ESI file.†

4.2 SERS experiment

For the preparation of the SERS experiment, the substrate was incubated in a 1 mM solution of MMC (Sigma-Aldrich 63759 7-Mercapto-4-methylcoumarin) in ethanol for 120 minutes. For normal Raman measurements, a concentration c_{NR} of 10 mM concentration was used to prepare the analyte solution. The Raman spectra were acquired using a Thermo Scientific DXR xi Raman Imaging confocal microscope, where the excitation laser source was chosen at 780 nm with a power of 8 mW. The exposure time was set to 1 s for 20 scans.



4.3 Reference-free XRF experiments

The RF-XRF characterization was performed at the laboratory of Physikalische-Technische Bundesanstalt (PTB), the German national metrology institute, with a dedicated beamline at BESSY II electron storage ring, a third-generation synchrotron radiation source. A thorough description of the experimental setup used at PTB can be found in ref. 40, while a scheme of the adopted experimental condition is visible in Fig. S6 in the ESI.† The employed beamline⁴¹ is equipped with a four crystals monochromator (FCM) which allows varying the energy of the incident beam from 1.75 keV to 10.5 keV. To quantify the amount of sulphur, the low-divergence monochromatic incident X-ray beam was set to have energy $E_0 = 2.5$ keV, larger than the X-ray absorption edge for sulphur K shell 2.472 keV. The instrumental setup, contained in an ultra-high vacuum (UHV) chamber, was equipped with both calibrated and other photodiodes used to perform the alignment procedure and to measure some characteristics of the incoming beam, including the full-width at half maximum beam width equal to (361 ± 18) μm and the photon flux $\phi_0 = (4.5 \pm 0.1) \times 10^8$ photons per s. The sample was mounted on a manipulator and moved along three translational axes and one rotational axis, varying the angle of incidence θ between the beam and the sample surface in the range from 27.5 and 45 with a step width of 0.5. The zero angle motor position θ_0 , at which the beam is parallel to the sample surface, was measured for each sample. The emitted fluorescence was detected utilizing an energy-dispersive silicon drift detector (SDD), mounted at 90 with respect to the incident beam and fully calibrated. The solid angle of detection $\Omega/4\pi$ was equal to 0.01134sr in the angular range of the measurement. The knowledge of the detector response functions and detection efficiency at different energies⁴² allowed the XRF spectral deconvolution, reported in Fig. 4b, to obtain the photon counts for the elemental fluorescence lines. The vicinity of the strong Au–M peak does not influence the sulphur peak deconvolution since it is adjusted for the peak height, rather than position and shape, reducing the degree of freedom and making it more reliable.

4.4 Theoretical simulations

A model of the gold substrate was prepared with CRYSTAL17 code by cleaving a three layers thick (111) surface out of the bulk structure and by optimizing the top layer at the DFT level with PW91 functional. During the molecular mechanics (MM) and molecular dynamics (MD) simulations an 8×8 supercell was used, comprising 91 gold atoms per layer. The force field (FF) parameters were checked, and fitted when necessary, against *ab initio* model calculations, as detailed in the ESI file.† During the MM energy minimizations, the geometry of the gold slab was kept fixed. MD calculations were performed with NVT ensemble at 298 K with 0.5 fs time step, and included 1 ns equilibration and 3 ns production runs.

Authors contributions

E. C. prepared the samples, L. M., A. S., A. M. G. conducted the SERS experiments, F. C. conducted the AFM measurements,

P. H., Y. K. and E. C. conducted the XRF experiments, D. M., A. Z. and M. C. conducted the theoretical simulations, E. C., L. M., P. H. and Y. K. performed data analysis, all authors contributed to writing the manuscript, F. F. L., M. L., M. C., A. M. R., N. D. L., B. B. and L. B. provided research infrastructure support and revised the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The project 16ENV07 AEROMET has received funding from the EMPIR programme co-financed by the Participating States and from the European Unions Horizon 2020 research and innovation programme. Part of this work has been carried out in the framework of the EMPIR Researcher Mobility Grant 16ENV07-RMG1. Part of this work has been carried out at Nanofacility Piemonte INRiM, a laboratory supported by the “Compagnia di San Paolo” Foundation.

Notes and references

- S.-Y. Ding, E.-M. You, Z.-Q. Tian and M. Moskovits, *Chem. Soc. Rev.*, 2017, **46**, 4042–4076.
- S.-Y. Ding, J. Yi, J.-F. Li, B. Ren, D.-Y. Wu, R. Panneerselvam and Z.-Q. Tian, *Nat. Rev. Mater.*, 2016, **1**, 1–16.
- C. Chen, Y. Li, S. Kerman, P. Neutens, K. Willems, S. Cornelissen, L. Lagae, T. Stakenborg and P. Van Dorpe, *Nat. Commun.*, 2018, **9**, 1–9.
- D. O. Sigle, S. Kasera, L. O. Herrmann, A. Palma, B. de Nijs, F. Benz, S. Mahajan, J. J. Baumberg and O. A. Scherman, *J. Phys. Chem. Lett.*, 2016, **7**, 704–710.
- S. Laing, L. E. Jamieson, K. Faulds and D. Graham, *Nat. Rev. Chem.*, 2017, **1**, 1–19.
- J. E. Park, N. Yonet-Tanyeri, E. Vander Ende, A.-I. Henry, B. E. Perez White, M. Mrksich and R. P. Van Duyne, *Nano Lett.*, 2019, **19**, 6862–6868.
- N. Kim, M. R. Thomas, M. S. Bergholt, I. J. Pence, H. Seong, P. Charchar, N. Todorova, A. Nagelkerke, A. Belessiotis-Richards and D. J. Payne, *et al.*, *Nat. Commun.*, 2020, **11**, 1–12.
- M. J. Natan, *Faraday Discuss.*, 2006, **132**, 321–328.
- M. D. Porter and J. H. Granger, *Faraday Discuss.*, 2017, **205**, 601–613.
- E. Le Ru, E. Blackie, M. Meyer and P. G. Etchegoin, *J. Phys. Chem. C*, 2007, **111**, 13794–13803.
- X. Wang, S.-C. Huang, S. Hu, S. Yan and B. Ren, *Nat. Rev. Phys.*, 2020, 1–19.
- M. F. Cardinal, E. Vander Ende, R. A. Hackler, M. O. McAnally, P. C. Stair, G. C. Schatz and R. P. Van Duyne, *Chem. Soc. Rev.*, 2017, **46**, 3886–3903.



- 13 A. Sivanesan, W. Adamkiewicz, G. Kalaivani, A. Kamińska, J. Waluk, R. Holyst and E. L. Izake, *Analyst*, 2014, **140**, 489–496.
- 14 S. He, J. Chua, E. K. M. Tan and J. C. Y. Kah, *RSC Adv.*, 2017, **7**, 16264–16272.
- 15 F. Xu, F. Ma, Z. Ding, L. Xiao, X. Zhang, Q. Lu, G. Lu and D. L. Kaplan, *ACS Appl. Mater. Interfaces*, 2019, **11**, 42896–42903.
- 16 P. Wang, L. Wu, Z. Lu, Q. Li, W. Yin, F. Ding and H. Han, *Anal. Chem.*, 2017, **89**, 2424–2431.
- 17 D.-K. Lim, K.-S. Jeon, J.-H. Hwang, H. Kim, S. Kwon, Y. D. Suh and J.-M. Nam, *Nat. Nanotechnol.*, 2011, **6**, 452.
- 18 Y. S. Hu, J. Jeon, T. J. Seok, S. Lee, J. H. Hafner, R. A. Drezek and H. Choo, *ACS Nano*, 2010, **4**, 5721–5730.
- 19 M. S. Schmidt, J. Hubner and A. Boisen, *Adv. Mater.*, 2012, **24**, 11–18.
- 20 M. Hu, F. S. Ou, W. Wu, I. Naumov, X. Li, A. M. Bratkovsky, R. S. Williams and Z. Li, *J. Am. Chem. Soc.*, 2010, **132**, 12820–12822.
- 21 D. Lin, Z. Wu, S. Li, W. Zhao, C. Ma, J. Wang, Z. Jiang, Z. Zhong, Y. Zheng and X. Yang, *ACS Nano*, 2017, **11**, 1478–1487.
- 22 J.-A. Huang, Y.-Q. Zhao, X.-J. Zhang, L.-F. He, T.-L. Wong, Y.-S. Chui, W.-J. Zhang and S.-T. Lee, *Nano Lett.*, 2013, **13**, 5039–5045.
- 23 J. Langer, *et al.*, *ACS Nano*, 2020, **14**(1), 28–117.
- 24 S. E. Bell, G. Charron, E. Cortés, J. Kneipp, M. L. de la Chapelle, J. Langer, M. Procházka, V. Tran and S. Schlücker, *Angew. Chem., Int. Ed.*, 2020, **59**, 5454–5462.
- 25 S. Kara, A. Keffous, A. Giovannozzi, A. Rossi, E. Cara, L. D'Ortenzi, K. Sparnacci, L. Boarino, N. Gabouze and S. Soukane, *RSC Adv.*, 2016, **6**, 93649–93659.
- 26 C. Korzeniewski, J. P. Kitt, S. Bukola, S. E. Creager, S. D. Minteer and J. M. Harris, *Anal. Chem.*, 2018, **91**, 1049–1055.
- 27 A. D'Agostino, A. M. Giovannozzi, L. Mandrile, A. Sacco, A. M. Rossi and A. Taglietti, *Talanta*, 2020, **216**, 120936.
- 28 C. Vericat, M. Vela, G. Benitez, P. Carro and R. Salvarezza, *Chem. Soc. Rev.*, 2010, **39**, 1805–1834.
- 29 M. S. Inkpen, Z.-F. Liu, H. Li, L. M. Campos, J. B. Neaton and L. Venkataraman, *Nat. Chem.*, 2019, **11**, 351–358.
- 30 B. Beckhoff, *J. Anal. At. Spectrom.*, 2008, **23**, 845–853.
- 31 J. Sherman, *Spectrochim. Acta*, 1955, **7**, 283–306.
- 32 B. Beckhoff, B. Kanngießler, N. Langhoff, R. Wedell and H. Wolff, *Handbook of practical X-ray fluorescence analysis*, Springer Science & Business Media, 2007.
- 33 B. Beckhoff, R. Fliegau, M. Kolbe, M. Müller, J. Weser and G. Ulm, *Anal. Chem.*, 2007, **79**, 7873–7882.
- 34 M. Müller, P. Hönicke, B. Detlefs and C. Fleischmann, *Materials*, 2014, **7**, 3147–3159.
- 35 T. Schoonjans, A. Brunetti, B. Golosio, M. S. del Rio, V. A. Solé, C. Ferrero and L. Vincze, *Spectrochim. Acta, Part B*, 2011, **66**, 776–784.
- 36 J. E. Millstone, W. C. Chan, C. R. Kagan, L. M. Liz-Marzán, N. A. Kotov, P. A. Mulvaney, W. J. Parak, A. L. Rogach, P. S. Weiss and R. E. Schaak, Redefining the Experimental and Methods Sections, *ACS Nano*, 2019, **13**(5), 4862–4864.
- 37 E. Cara, L. Mandrile, F. Ferrarese Lupi, A. M. Giovannozzi, M. Dialameh, C. Portesi, K. Sparnacci, N. De Leo, A. M. Rossi and L. Boarino, *Sci. Rep.*, 2018, **8**, 11305.
- 38 G. Aprile, F. Ferrarese Lupi, M. Fretto, E. Enrico, N. De Leo, L. Boarino, F. G. Volpe, G. Seguini, K. Sparnacci and V. Gianotti, *et al.*, *ACS Appl. Mater. Interfaces*, 2017, **9**, 15685–15697.
- 39 D. Necas and P. Klapetek, Gwyddion: an open-source software for SPM data analysis (<http://gwyddion.net/>), *Cent. Eur. J.*, 2012.
- 40 J. Lubeck, B. Beckhoff, R. Fliegau, I. Holfelder, P. Hönicke, M. Müller, B. Pollakowski, F. Reinhardt and J. Weser, *Rev. Sci. Instrum.*, 2013, **84**, 045106.
- 41 M. Krumrey and G. Ulm, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 2001, **467**, 1175–1178.
- 42 F. Scholze and M. Procop, *X-Ray Spectrom.*, 2009, **38**, 312–321.



Chapter 6

Structure and Stability of
7-mercapto-4-methylcoumarin SAM on
Gold: an Experimental and
Computational analysis

Structure and Stability of 7-mercapto-4-methylcoumarin SAM on Gold: an Experimental and Computational analysis.

Davide Marchi,[‡] Eleonora Cara,[¶] Federico Ferrarese Lupi,[¶] Philipp Hönicke,[§]
Yves Kayser,[§] Burkhard Beckhoff,[§] Micaela Castellino,^{||} Alberto Zoccante,[‡]
Michele Laus,[‡] and Maurizio Cossi^{*,‡}

[‡]*Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte
Orientale, via T. Michel 11, I-15121, Alessandria, Italy*

[¶]*Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce, 91, I-10135,
Torino, Italy*

[§]*Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany.*

^{||}*Department of Applied Science and Technology, Politecnico di Torino, C.so Duca degli
Abruzzi 24, 10129 Turin, Italy*

E-mail: maurizio.cossi@uniupo.it

Abstract

Self-assembled monolayers of 7-mercapto-4-methylcoumarin (MMC) on a flat gold surface were studied by Molecular Dynamics (MD) simulations, reference-free grazing incidence X-ray fluorescence (GIXRF) and X-ray pho-

toemission spectroscopy (XPS), to determine the maximum monolayer density and to investigate the nature of the molecule/surface interface. In particular, the protonation state of the sulfur atom upon adsorption was analyzed, since some recent literature presented evidences for physisorbed thiols (preserving the S-H bond), unlike the common picture of chemisorbed thiyls (losing the hydrogen). MD with a specifically tailored force field was used to simulate either thiol or thiyl monolayers with increasing number of molecules, to determine the maximum dynamically stable densities. This result was refined by computing the monolayer chemical potential as a function of the density with the Bennet Acceptance Ratio method, based again on MD simulations. The monolayer density was measured with GIXRF, which provided a quantitative estimate of the number of sulfur atoms on top of flat gold surfaces embedded in a solution of MMC, to allow the formation of a dense monolayer. The sulfur core level binding energies in the same monolayers were measured by XPS, fitting the recorded spectra with the binding energies proposed in the literature for free or adsorbed thiols and thiyls, to get insight on the nature of the molecular species present in the layer.

1 Introduction

Since their first description,¹ thiol self-assembled monolayers (SAM) on gold surfaces and nanoparticles have been widely used in a variety of technological applications,²⁻¹³ and studied with a wealth of diffraction,¹⁴⁻¹⁸ spectroscopic^{8,19-23} and other surface science techniques.²⁴ In addition, thiol SAMs have been modeled theoretically by several ab initio (mainly density functional theory, DFT)²⁵⁻³² and molecular mechanics^{27,28,33-38} studies.

Despite such a widespread interest, many key features of thiol/gold monolayers are still debated: even for some fundamental characteristics of the interface, as the

nature of sulfur-gold bond or the weight of intermolecular interactions inside the organic layer, different models have been proposed with no unique interpretation of the experimental data. For instance, the protonation state of the sulfur atom bonded to the metal surface is not unanimously accepted:^{30,32,39} though most researchers assume that the S-H bond is dissociated and de-protonated sulfur is covalently bound to gold atoms,^{2,40-42} some evidences have been presented showing that the layers can also be formed by undissociated thiol molecules.⁴³⁻⁴⁶ (The former model is also referred to as “chemisorption”, in contrast with the latter described as “physisorption” to stress the absence of a typical covalent bond between protonated sulfur and gold).

The present work contributes to this investigation, comparing the structure and stability of different SAMs of 7-mercapto-4-methylcoumarin (MMC) on gold (111) surfaces. A number of monolayers, formed either by undissociated thiol or radical thiyl MMC units (Figure 1), have been modeled theoretically, and the results compared with the absolute quantification of MMC surface density obtained by means of reference-free grazing incidence X-ray fluorescence (GIXRF) in SAMs prepared on 100 nm-thick gold layers. Additional X-ray photoelectron spectroscopy (XPS) characterization was performed to determine the nature of the thiol/thiyl S-Au chemical bond.

(A note on the terminology: since nearly all the proposed models for the dissociation of the S-H bond are based on a homolytic cleavage, eventually leading to H₂ formation, we prefer to consider the dissociated species as a radical thiyl rather than a thiolate ion, unlike many published studies. Whether the R-S unit has to be seen as a radical or an ion depends on the charge distribution in the S-Au bond, and appears as a rather unessential question in this context.)

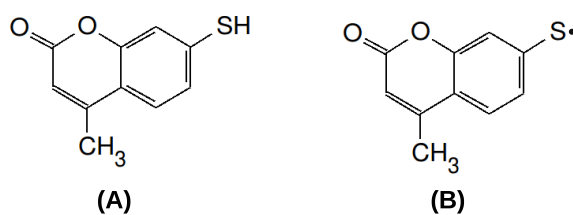


Figure 1: 7-mercapto-4-methylcoumarin (MMC) thiol molecule (A) and thiyl radical (B).

2 Methods and models

2.1 Theoretical modeling

Some models of thiol and thiyl SAM with different densities were prepared, to study their kinetic and thermodynamic stability with molecular dynamics (MD); the force fields (FF) were specifically parameterized for MMC on gold surfaces. Our goal is to determine the highest stable density and evaluate the chemical potential and the order degree for SAMs of both species: if thiol and thiyl SAM models exhibit different characteristics, the comparison with the experiments performed on the same system could shed some light on the protonation state of the sulfur atoms and in general on the structure of the SAM.

The gold (111) surface was modeled by cleaving a three layers thick slab out of the metal bulk structure: the periodic unit cell comprises 24×24 atoms in each layer and its surface area, considering an atomic radius for gold of 0.1385 nm, is 38.2 nm². MMC thiol and thiyl structures were optimized at the DFT level (with B3LYP/cc-pvDZ functional and basis), and a variable number of such units were assembled on the slab using PACKMOL package,⁴⁷ to create monolayers with the desired densities; in the initial conformations, all the organic units were in ‘vertical’ position (see below).

The FF parameters for MMC and ethanol were taken from GROMOS 54A7 set⁴⁸ as provided by the Automated Topology Builder (ATB) website; interactions of MMC molecules and radicals with the gold surface were described with pairwise non-bonding parameters fitted on DFT calculations as detailed in the Electronic Supporting Information (ESI). We decided to model both $-\text{SH} \cdots \text{Au}$ and $-\text{S} \cdots \text{Au}$ interactions with 6–12 Lennard-Jones functions, though the latter can be considered a real covalent bond, to allow thiyl units to shift on the surface and possibly also leave too crowded monolayers.

All the FF-based calculations were carried out with GROMACS2020 package.⁴⁹ After an initial energy minimization, to remove spurious close contacts, the MD simulations were performed with 2×10^5 steps of 0.5 fs for equilibration, and 2×10^6 steps of 1 fs for production runs. A 3 nm cut-off was used for the Van der Waals

interactions, while electrostatic interactions were computed with a 2 nm cut-off, and using the PME method for longer distances; during all the simulations, Au positions were kept frozen.

The MMC chemical potential in the monolayers at various densities was computed as the free energy of decoupling⁵⁰ of one thiol or thiyl from the gold slab and the rest of the layer, using the Bennet Acceptance Ratio (BAR) method⁵¹ implemented in GROMACS (gmx BAR procedure). Following this procedure, a coupling parameter λ (varying from 1 to 0 as the system shifts from real to decoupled) was defined to gradually switch off the intermolecular interactions between the target molecule and the rest of the system: first, Coulomb interactions were removed in 20 steps while vdW interactions remained unaltered, then also vdW terms were eliminated in 20 further steps, until the target unit was completely decoupled. For each λ value, a MD run was performed comprising 0.5 ns equilibration (timestep 0.5 fs) and 1 ns production (timestep 1 fs). The same procedure was adopted to compute the thiol free energy in ethanol solution.

2.2 Sample preparation

The preparation of the MMC SAM was conducted by following a standardized protocol.⁵² The molecules of 7-mercapto-4-methylcoumarin were purchased in powder form from Merck. A solution of MMC in ethanol was prepared in the volume of 20 ml per each sample. The substrate preparation required cutting and cleaning pieces of a silicon wafer in ultrasonic bath with acetone and then isopropanol. A layer of 100 nm of gold was deposited on the substrates by means of RF sputtering in argon plasma with residual pressure of $5 \cdot 10^3$ mbar and power of 100 W. The resulting gold surface is continuous and polycrystalline. The gold-coated substrates were then immersed in the MMC solution for two hours, then abundantly rinsed with EtOH to remove any excess molecules not bounded to the gold surface.

2.3 GIXRF characterization

The quantitative characterization of the areal density of MMC molecules by reference-free grazing incidence X-ray fluorescence (GIXRF)⁵³ was conducted at the four crystal monochromator (FCM) beamline for bending magnet radiation⁵⁴ at the BESSY II synchrotron radiation facility. The reference-free GIXRF experiments were performed employing in-house built instrumentation,⁵⁵ which allows for precise sample alignment and angular variations of the sample with respect to the incident photon beam.

To optimize the excitation conditions for sulfur K-shell X-ray fluorescence, while minimizing the contributions of Au M-shell X-ray fluorescence to the experimental spectra, an incident photon energy of $E_0 = 2.6$ keV was chosen. During the GIXRF scans, the angle of incidence (defined between sample surface and incident X-ray beam) was varied between 0° and 7° with varying stepsize. At each angular position, a fluorescence spectrum was recorded by means of a calibrated⁵⁶ silicon drift detector (SDD) mounted at 90° with respect to the incident beam. Additional calibrated photodiodes on a separate 2θ axis allow for both X-ray reflectometry (XRR) measurements as well as for a determination of the incident photon flux. The recorded spectra are deconvolved using detector response functions⁵⁶ for relevant fluorescence lines and for background contributions as bremsstrahlung.

A SI-traceable quantification of the mass deposition of sulfur can be performed using the deconvoluted sulfur fluorescence events as presented in references 57,58. Using Sherman's equation⁵⁹ and necessary experimental parameters, e.g., the solid angle of detection or the incident photon flux, as well as atomic fundamental parameters, the mass deposition of sulfur can be calculated in absolute terms from the sulfur K X-ray fluorescence count rate as obtained from the spectral deconvolution. The required instrumental parameters are known due to the use of the well-known physically calibrated instrumentation.⁵⁸ The relevant fundamental parameters are taken from databases.⁶⁰ Here, the mass deposition of sulfur was quantified with an overall uncertainty of 11%, deriving mainly from the fundamental parameter uncertainties.

2.4 XPS characterization

A PHI 5000 Versaprobe Scanning X-ray Photoelectron Spectrometer (Physical Electronics, Chanhassen, MN, USA) has been involved in this study to get information regarding the relative atomic concentration (at.%) of each element present on the surface of both bare and functionalized Au thin film, and also to have further evidence regarding the bonds established between the MMC molecule and the golden surface. XPS measurements have also been carried out on MMC commercial powder, to check its bare chemical composition as a reference. The X-ray source was a monochromatic Al $K\alpha$ radiation (1486.6 eV, 15 kV voltage and 1 mA anode current). All samples were subjected to a combined electron and Ar ion gun neutralizer system, to decrease the electrical charging effect during the analysis.

The semi-quantitative atomic concentration and fitting procedures were acquired using CasaXPS 2.3.23 dedicated software (Casa Software Ltd., Wilmslow, UK). All core-level peak energies were referenced to C1s peak at 284.5 eV and the background contribution in HR scans was subtracted by means of a Shirley function. A spot size of 100 μm was used to collect the photoelectron signal for both the high resolution (HR) and the survey spectra.

Different pass energy values were employed: 187.8 eV for survey spectra and 23.5 eV for HR peaks. Survey scans (from 1200 to 0 eV, energy step $\Delta E = 0.1$ eV) have been performed as the first step measurements to detect all the elements on sample surfaces. HR scans have been performed only in smaller ranges in the Binding Energy (BE) scale (energy step $\Delta E = 1$ eV) around chemical element peaks of major interest for this study, i.e. C(1s), O(1s), S(2p) and Au(4f).

Powder sample has been loaded directly on a steel mask by attaching it on a double-sided conductive tape. Au thin film deposited on Si substrates (both functionalized with MMC molecule and not) have been attached on the 2-inches sample holder surface by means of double-sided conductive tape. Working pressure, inside the main chamber, has reached a maximum value of 10^{-6} Pa.

3 Results and discussion

3.1 Molecular dynamics modeling

Several models were defined placing an increasing number of thiol or thiyyl units randomly onto the gold slab: then, after a MM minimization to remove close-contacts and spurious structures, MD simulations were run at 298 K until equilibration, and then for further 2 ns dynamics to check the monolayer stability. In the densest layers some of the organic molecules or radicals detached from the slab during the dynamics, remaining in contact with the other MMC units, in a sort of disordered double layer. We consider that such molecules, not interacting with the gold surface, would be washed away during the SAM preparation, so did not include them in the calculation of the SAM density.

The densities of the dynamically stable layers are collected in Table 1. Thiyls can be packed more closely than thiols, forming denser SAMs, as expected for their much larger interaction energy with gold, which balances the intermolecular repulsions in the crowded layers. Only when 216 thiyl units are initially placed on the surface, some of them are forced to leave the monolayer during the MD, while with thiols we find some units leaving the SAM even with 81 initial molecules. Clearly, the number of units expelled from the layer depends on the starting conformation also, but this effect becomes less important as the initial number grows and most of the surface is covered by the organic units: in fact, no attempt to start with more than 162 thiols or 216 thiyls led to densities larger than those reported in Table 1.

Table 1: Number of molecules or radicals in the SAM and corresponding density (the gold slab area being 38.2 nm^2): when some units left the surface during the MD, the initial number is indicated in parentheses.

Thiol	# starting units	27	54	(81) \rightarrow 75	(108) \rightarrow 97	(135) \rightarrow 126	(162) \rightarrow 142
	density (nm^{-2})	0.71	1.41	1.96	2.54	3.30	3.72
Thiyl	# starting units	27	81	135	155	162	(216) \rightarrow 185
	density (nm^{-2})	0.71	2.12	3.53	4.06	4.24	4.84

Besides investigating the density limit of stable monolayers, MD provides useful insights also about the SAM structure at various coverages. Apart from the units leaving the surface at high densities, mentioned above, molecules and radicals were found either in ‘horizontal’ or ‘vertical’ arrangement: in the former the units lie down on the slab maximizing the interaction of all the atoms with the surface, in the latter sulfur interacts strongly with gold, while the rest of the organic atoms are involved mainly in side intermolecular interactions, which can be overall attractive or repulsive, depending on the SAM density.

Some representative snapshots of the simulated thiol and thiyl SAMs are shown in Figures 2 and 3, respectively: clearly the position and orientation of the organic units change during the dynamics, but we found that the number of horizontal and vertical arrangements at the various densities is remarkably stable at 298 K.

The simulations show that at low coverage the organic units prefer to lie on the surface, because the interactions with the gold atoms are favored with respect to the intermolecular ones. As the SAM density increases, more and more molecules and radicals stand vertically, in agreement with the mechanism of the monolayer formation often proposed in the literature.⁶¹⁻⁶⁴ Thiol SAMs are less ordered than thiyl layers: in the former numerous molecules remain in horizontal position even at high densities competing with their vertical counterparts for the gold surface. As noted above, it is not possible to simulate an all-vertical thiol SAM, since the molecules prefer to leave a too crowded layer.

On the other hand, the vertical arrangement is easier in thiyl SAMs, since it allows a better interaction between the sulfur and gold atoms (which is markedly stronger than the analogous interaction in thiols) compensating the partial loss of stabilization when the molecule/surface interactions are substituted by side-side intermolecular ones. As a consequence, thiyl SAMs can be denser and more ordered: at high densities, almost all the radicals are in vertical position, allowing a closer packing. A picture of the ordered patterns established in a dense thiyl SAM, with several MMC phenyl rings stacked in lines is shown in the ESI.

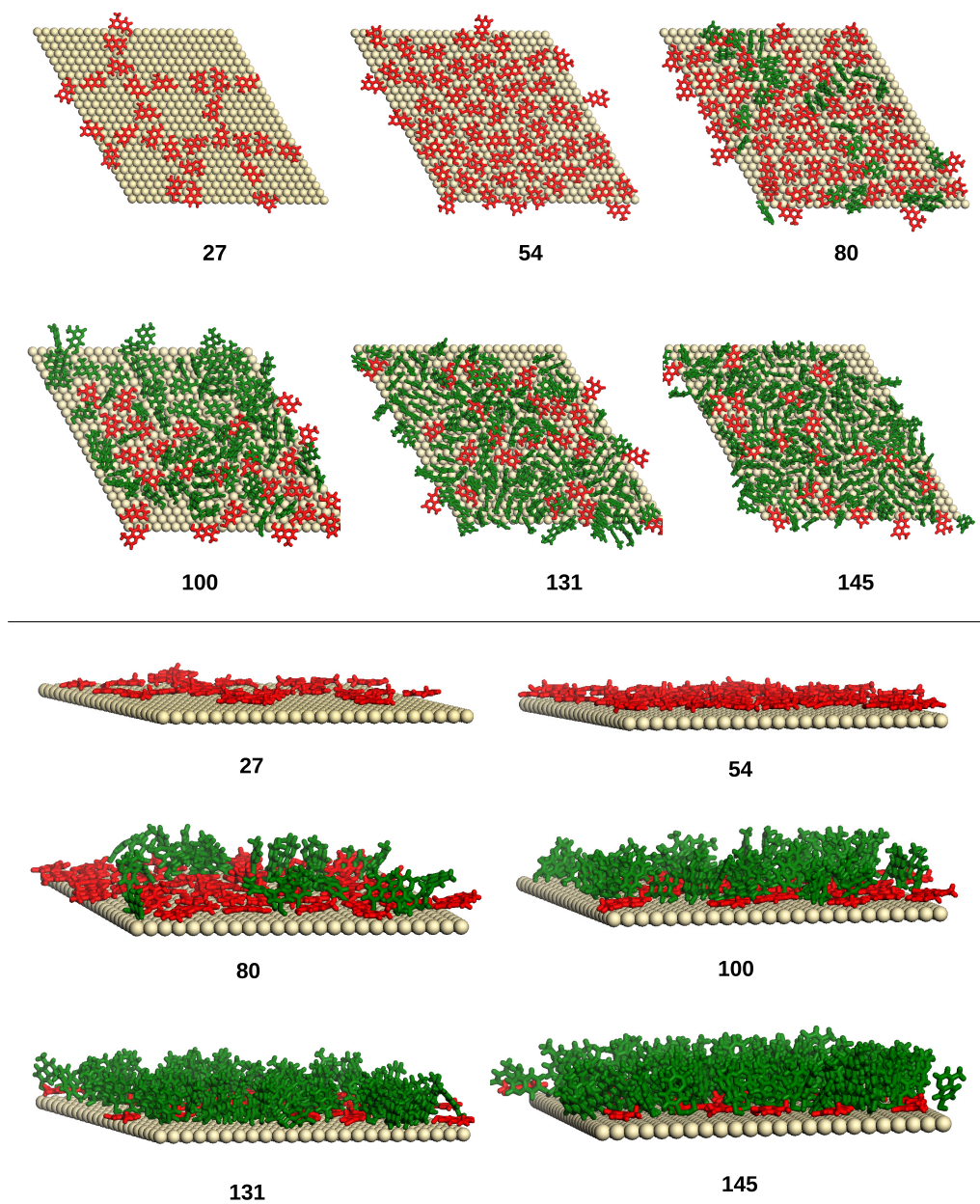


Figure 2: MD snapshots (top and side views) of thiol SAMs with different numbers of attached units (indicated below each image). In red/green molecules in horizontal/vertical arrangement.

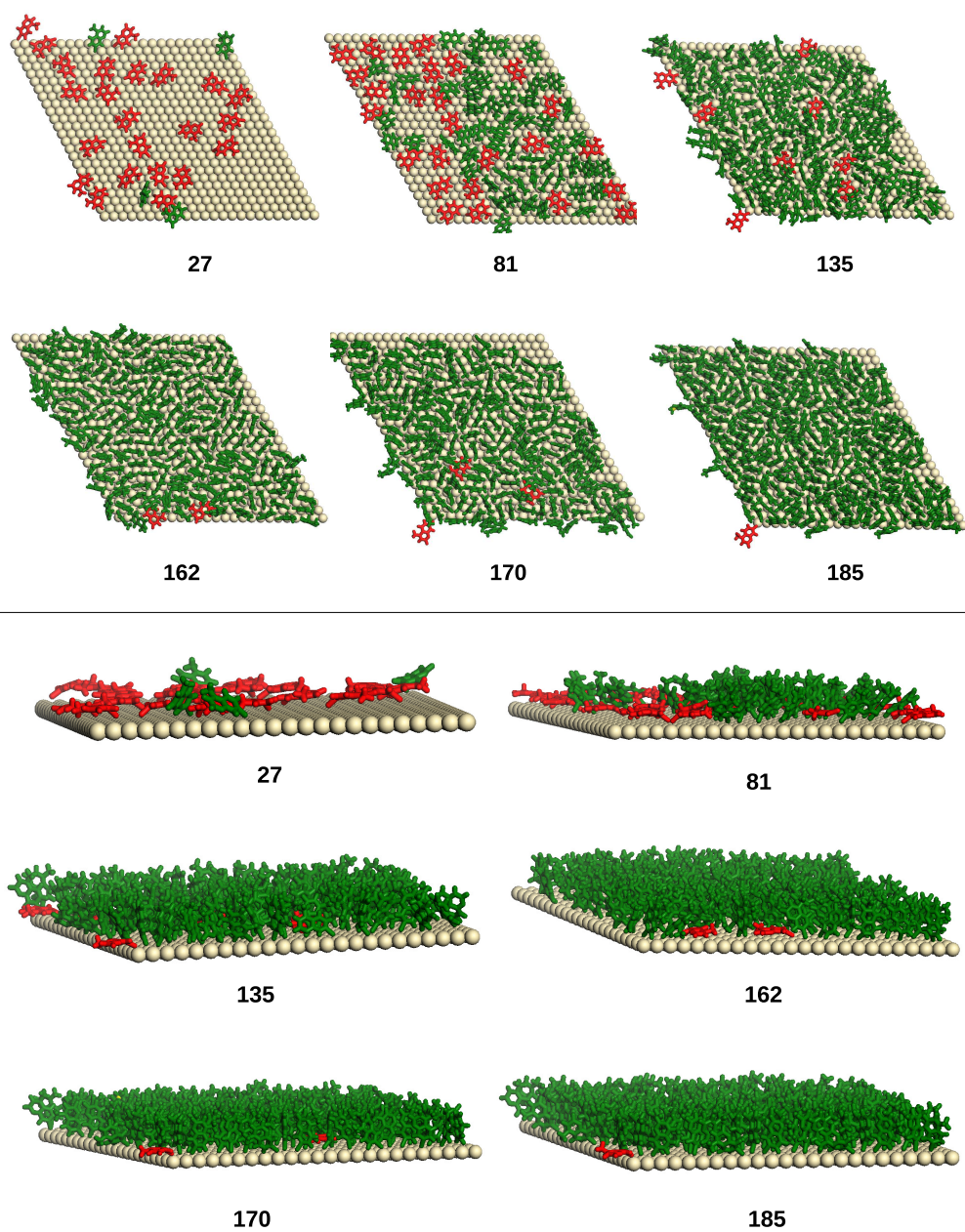
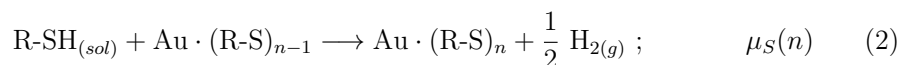
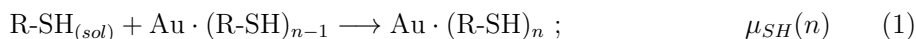


Figure 3: MD snapshots (top and side views) of thiol SAMs with different numbers of attached units (indicated below each image). In red/green molecules in horizontal/vertical arrangement.

3.2 Chemical potential calculations

The MD simulations discussed above provide useful insights about the SAM dynamical stability: however, this approach is not completely satisfactory for two reasons. First, the dynamics risk to be biased by the initial conformations, unless they can be run for a very long time and possibly with reasonable temperature annealings, to refine the exploration of the potential surfaces. A second, more severe problem is that during the MD the organic units can leave the monolayer “evaporating” into a sort of low density gas phase, unlike in the real process of SAM formation, where the equilibrium establishes between the monolayer and a liquid thiol solution. Running the MD inside a box of solvent molecules did not solve the problem, for the strong cage effect hampering the detachment of thiols from the surface: moreover, when thiol SAMs are involved, the -SH dissociation is also to be considered, as detailed below.

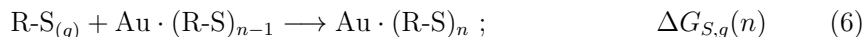
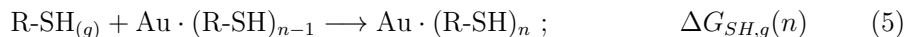
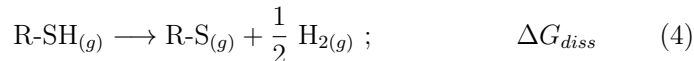
For these reasons, the information obtained with MD was complemented by the calculation of the monolayer chemical potentials as a function of the SAM density. The chemical potentials $\mu_{SH}(n)$, $\mu_S(n)$ are defined as the free energy changes associated to the following processes:



The number of organic units attached to the model slab after the reaction, from which the SAM density is computed, is n ; following the experimental conditions, the reactant thiol is dissolved in ethanol, while the surface monolayer is considered in vacuo; in the case of thiol SAM, the reaction includes the homolytic dissociation of the S-H bond and the formation of gaseous molecular hydrogen.

Then, to obtain the chemical potentials, we have to model the following elemen-

tary processes, combining the respective ΔG s:



Reactions 3 and 4 refer to the formation of molecular or radical units in the gas phase; 5 and 6 to the passage of one unit from the gas phase to thiol or thiyl monolayers, respectively. With these definitions, the chemical potentials result:

$$\mu_{SH}(n) = \Delta G_{evap} + \Delta G_{SH,g}(n) \quad (7)$$

$$\mu_S(n) = \Delta G_{evap} + \Delta G_{diss} + \Delta G_{S,g}(n) \quad (8)$$

Reaction 4 was modeled at the DFT level with Gaussian16 program, including the calculation of vibrational frequencies and the evaluation of thermal contributions to enthalpy and entropy by classical Boltzmann averages; the selection of the suitable density functional and basis set, however, required some care. Since experimental data on S-H bond dissociation in MMC are not available, we collected the experimental values for benzenethiol and some of its derivatives, to evaluate the effect of both electron-withdrawing and electron-donating substituents, and computed ΔG_{diss} for these molecules with two hybrid (B3LYP, B3P86) and one pure (BLYP) functionals and a number of different basis sets. The DFT results are reported and analyzed in detail in the ESI: in conclusion, our best estimate is $\Delta G_{diss} = 130.0$ kJ/mol.

For the other steps we used the thermodynamic integration described in the ‘‘Methods and models’’ Section, which provides ΔG values with a series of MD simulations. This procedure was applied first to the evaporation of one MMC molecule from a box of ethanol molecules, reproducing the solvent density, and the free energy associated to reaction 3 resulted $\Delta G_{evap} = 17.9$ kJ/mol.

$\Delta G_{SH,g}(n)$ and $\Delta G_{S,g}(n)$ were computed with the thermodynamic integration method as well, decoupling one organic unit from each of the SAMs previously modeled and equilibrated with MD: the values obtained for thiol and thiyl at the various densities are collected in Table 2. Combining these results with the evaporation and dissociation free energies reported above, one finally obtains the SAM chemical potentials also reported in Table 2.

Table 2: Free energy of insertion in vacuo (ΔG_g , kJ/mol) and chemical potential (μ , kJ/mol) of MMC monolayers on Au(111) surfaces; n is the number of molecules or radicals per model slab; ρ (nm^{-2}) the SAM density.

Thiol				Thiyl			
n	ρ	$\Delta G_{SH,g}(n)$	$\mu_{SH}(n)$	n	ρ	$\Delta G_{S,g}(n)$	$\mu_S(n)$
27	0.71	-228.4	-210.5	27	0.71	-371.6	-223.7
54	1.41	-206.8	-188.9	81	2.12	-316.0	-168.1
75	1.96	-116.0	-98.1	135	3.53	-244.4	-96.5
97	2.54	-117.9	-100.0	153	4.00	-182.3	-34.4
126	3.30	-101.4	-83.5	155	4.06	-144.1	+3.8
142	3.72	-88.2	-70.3	162	4.24	-85.2	+62.7
				185	4.84	+291.8	+439.7

The data show that the chemical potential of thiol SAMs remains negative also at the highest densities attainable with the MD equilibration, which is mainly due to the large negative values of $\Delta G_{SH,g}$. One could wonder why it is not possible to simulate SAMs with higher density, then: as mentioned above, when the monolayer initial density grows, an increasing number of thiols leave the gold surface during the MD and form a sort of second layer, strongly interacting with the underlying molecules still in the SAM. The free energy of these second layer thiols with respect to vacuum, computed with the same technique, falls in the range 102 – 112 kJ/mol, depending on the position, very close to the $\Delta G_{SH,g}$ of thiols in the SAMs with 100 to 145 molecules. Then the dynamical instability of denser SAMs, observed above, can derive from the competition of the second layer, favoring the shift from the crowded SAM to the spacious layer floating above it.

On the other hand, in the case of thiol layers we see that the densest SAMs, with 162 and 185 radicals on the model slab, are not thermodynamically favored, though stable during the MD. In fact, with $n = 162$ the free energy of insertion of one radical in gas phase ($\Delta G_{S,g}$) is still negative, but once considering the solvation and dissociation free energies, the chemical potential gets positive; even worse the situation for $n = 185$, where $\Delta G_{S,g}$ is already positive and is further increased by the other contributions. The system with 155 initial thiols is borderline, as the computed chemical potential turns out very close to zero, even if $\Delta G_{S,g}$ is strongly negative: the result depends mainly on the large dissociation energy, whose calculation is quite approximated, as explained in the ESI, so the thermodynamic stability of this system is uncertain.

A further comment can be done about the system with 185 thiols, where the strongly positive $\Delta G_{S,g}$ suggests a thermodynamically unstable SAM even without considering the de-solvation and bond dissociation contributions. Evidently, in this case the MD could not lead to a complete equilibration, since no radicals left the surface despite the thermodynamic advantage that could have been gained, because of the very high energy needed to break the strong S-Au bond, which “trapped” the SAM in a less favorable conformation. This is a good example of how the chemical potential calculation can refine the MD analysis.

3.3 Experimental quantification of the molecular density

The surface density of the MMC molecules in the SAMs formed on a flat reflecting gold surface was determined by performing a reference-free GIXRF experiment,⁵³ sketched in Figure 4a.

In this analytical method, the angle of incidence of the X-ray beam is varied around the critical angle for total external reflection allowing the formation of an X-ray standing wave (XSW) interference field just above the sample surface. The XSW field enhances the fluorescence emitted by the atom inside it, while reducing the fluorescence signal from the substrate and thus spectral background.⁵⁷ The quantification can be performed, without any calibration standard, through the conversion of the element-specific fluorescence photon count rate $P_{e,K}$ to the mass

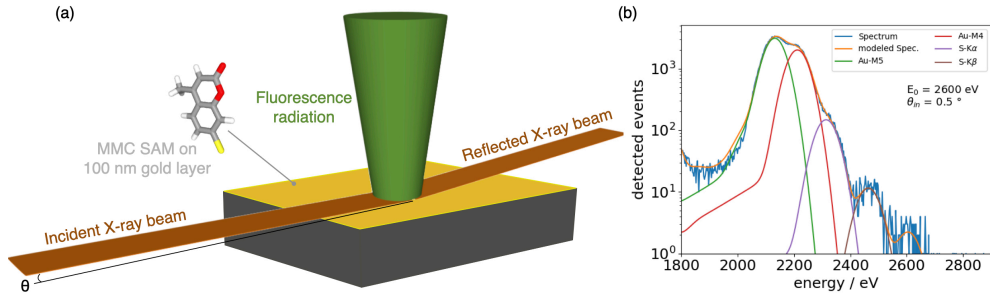


Figure 4: (a) Schematic representation of the GIXRF measurement where the X-ray beam impinges on the reflecting sample surface at θ angle and excites the fluorescence radiation detected at 90° . (b) Deconvolution of the spectrum acquired at $\theta = 0.5^\circ$.

of the element of interest per unit area σ_e by combining the atomic fundamental parameters and calibrated instrumental parameters in the following equation.⁵⁹

$$\sigma_e = P_{e,K} \cdot \frac{\sin \theta}{\Phi_0 \cdot \Omega / 4\pi} \cdot \frac{1}{\epsilon_{(E_e,K)} \cdot I_{XSW}(E_0, \theta)} \cdot \frac{1}{\tau_{e,K}(E_0) \cdot \omega_{e,K}} \quad (9)$$

where σ_e is expressed in g/cm^2 . The radiometrically calibrated instrumental parameters correct the fluorescence photon count rate $P_{e,K}$ obtained through spectral deconvolution (Figure 4b), i.e. the sine of the incidence angle θ , the incident photon flux Φ_0 , the solid angle of detection $\Omega/4\pi$. The second correction factor accounts for the SDD's detection efficiency at the photon energy of the fluorescence line K for the element e and the incident photon energy E_0 and angular dependent relative intensity of the XSW field.⁶⁵ Finally, the fundamental parameters $\tau_{e,K}(E_0)$ and $\omega_{e,K}$ are the partial photoionization cross section and fluorescence yield related to the K-shell of the target atom e , respectively. They form the production cross section for fluorescence radiation of the element of interest.

In the case of MMC, the GIXRF measurements were performed by selecting sulfur as target element. The experiment was carried out on a gold-coated substrate incubated in the MMC solution to determine the sulfur mass per unit area ascribable to the SAM. On each probed sample, the determined fluorescence photon count rate of the sulfur K fluorescence line was converted to its mass per unit area using tabulated values of the sulfur K-shell fluorescence yield $\omega_{S,K} = 0.08038$ with a relative

uncertainty of 7.5% estimated in reference 66 and of the partial photoionization cross section $\tau_{S,K}(E_0 = 2.6 \text{ keV}) = 1737 \text{ cm}^2\text{g}^{-1}$ with a relative uncertainty of 5%.⁶⁰

The mass of sulfur per unit area was found to be $\sigma_{S_{MMC}} = (24.3 \pm 4.0) \text{ ng/cm}^2$ and the same evaluation was performed on a bare gold sample where $\sigma_{S_{blank}} = (2.7 \pm 0.3) \text{ ng/cm}^2$ of sulfur were detected. The amount of sulfur ascribable to the SAM was found by $\sigma_{S_{SAM}} = \sigma_{S_{MMC}} - \sigma_{S_{blank}} = (21.6 \pm 4.0) \text{ ng/cm}^2$. The numerical density of sulfur atoms can be derived as $\sigma_{S_{SAM}} \cdot N_A / w_S$, where w_S is the atomic weight of sulfur and N_A Avogadro's number. Since there is only one sulfur atom per each molecule, its numerical density corresponds to the number of self-assembled molecules per unit area which is $(4.1 \pm 0.7) \text{ nm}^{-2}$. The total uncertainty associated to every reported value is due to the propagation of uncertainties for independent variables including statistical uncertainty on repeated measurements and the uncertainty contributions in equation 9, i.e. the uncertainties on the fundamental atomic parameters, already reported, 4% relative uncertainty on the determination of solid angle of detection, 1.5% relative uncertainty on the incident photon flux and 2% on the XSW field intensity above the sample surface.

3.4 XPS analysis of the bond chemistry

The nature of bonds between gold surfaces and organic thiols has been widely studied in the last decades. XPS is one of the most used techniques to investigate such bonds, considering in particular the difference between bound and unbound species. As reported by Castner et al.,⁶⁷ there is a sort of hierarchical displacement in the position of the S(2p_{3/2}) core-level binding energy (CLBE), which follows this general trend: unbound thiol or disulfide (164-163 eV), bound thiol or thiyl in hollow site (162 eV) and thiyl in low-coordination site (< 162 eV). Zubragel et al.⁶⁸ have also deeply studied the presence of different sulfur species in SAMs on Au and Ag, but they attribute the chemical shift at 161.8 eV to threefold bound thiols, and the shift at 163.1 eV the bound thiyl in lower coordination sites, in contrast with the work in reference 67. More recent works have tried to describe more accurately the chemical species that can be found on a Au thin film, by comparing DFT simulations with surface experimental analysis.⁶⁹⁻⁷¹ In particular, Jia et al.⁷¹ have recognized four different chemical shifts due to the interaction of sulfur species on the Au layer.

A first component at low binding energy (161.2 eV) is assigned to the thiol in a metastable site rather than atomic sulfur, as previously reported,⁶⁹ a second one (162.0 eV) to bound thiol, a third one (163.0 eV) to unbound or free -SH, and a final one at binding energy higher than 163.5 eV, is assigned to physisorbed -SH on a second layer on SAM.

Thus we have decided to check the MMC powder, which provides the reference value for unbound thiol, as well as the blank Au thin film deposited on Si, and the MMC SAM on Au thin film sample (see figure 5). MMC powder survey spectrum (not reported) has shown the presence of C(1s), O(1s) and S(2p) peaks, as expected, while the bare Au thin film and the MMC on Au samples have shown in addition the presence of Au(4f) doublet. The presence of S on the blank Au sample has been ascribed to environmental contamination, due to sulfur species present in the atmosphere. In figure 5, HR S(2p) core level spectra have been reported for the three analyzed samples, together with their deconvolution procedure.

For the MMC SAM on Au sample we identify the following four components, each made up by a doublet due to S(2p_{3/2}) and S(2p_{1/2}) spin-orbit splitting (Figure 5a): i) at 161.0 eV (8.6%), ii) at 162.0 eV (58.9%), iii) at 163.1 eV (19.9%) and iv) at 163.7 eV (12.6%). On the other hand, for the blank gold surface we find three peaks (Figure 5b): i) at 161.3 eV (41.8%), ii) at 162.1 eV (43.6%) and iii) at 163.2 eV (14.6%); the fourth component at higher binding energy is missing in this core level peak deconvolution. For the MMC powder the detected peaks are further reduced, with only two components ((Figure 5c): i) at 163.1 eV (53.9%) and ii) at 163.7 eV (46.1%).

In agreement with Jia et al. assignments,⁷¹ the components detected in the MMC powder are attributed to free -SH or mutually interacting thiols (referred to as “second layer” in the Figure caption). Passing to MMC SAM, two new, intense peaks arise at lower CLBE: most authors attribute the signals in this region to thiols chemisorbed on the Au surface (often referred to in the literature as “thiolate”, as noted above), though the possibility of physisorbed thiols is not definitely excluded. Interestingly, the peaks due to free -SH above 163 eV are also present in the SAM photoemission spectra, even if weaker than the lower energy components, suggesting the presence of unbound species, possibly in the second layer mentioned above. The presence of signals from adsorbed and, to a lower extent, free sulfur species even in

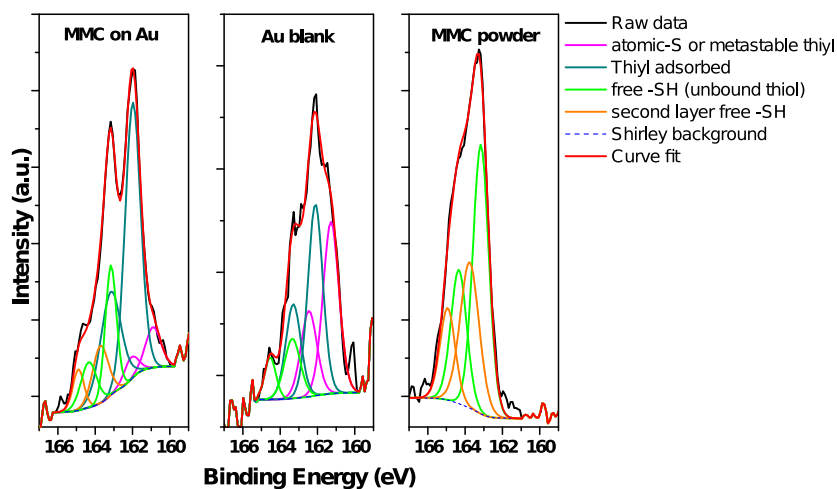


Figure 5: XPS S(2p) core level spectra for MMC SAM on Au (a), blank Au thin film (b) and MMC precursor powder (c). Deconvolution curves have been reported in each graph and their common legenda has been added aside.

the Au blank confirms the environmental contamination already detected by XRF experiments.

3.5 Comparison of experimental and theoretical results

As seen above, the MD analysis of denser and denser SAMs allows to put an upper bound to the number of thiol or thiol units that can be assembled on a flat, unreconstructed (111) gold surface: we found that the highest dynamically stable densities are 3.8 nm^{-2} for thiols, and 4.8 nm^{-2} for thiyis.

The calculation of SAM chemical potentials, obtained with MD thermodynamic integration and ab initio calculations, leads to a refinement of the previous conclusion

for the thiol monolayers: with this approach, the maximum density of thiol SAMs is confirmed at 3.8 nm^{-2} while for thiyl SAMs it results not higher than 4.1 nm^{-2} . The strong reduction of the maximum predicted density for thiyl SAMs is mainly due to the large dissociation energy of the S-H bond, which counterbalances the stronger interaction of the radicals with the surface. However, if the dissociation reaction were not completely equilibrated (for instance because the molecular hydrogen leaves the sample), the expected density for thiyl SAMs would be larger: this possibility will be examined in further studies.

The SAM density obtained from the GIXRF experiments is (4.1 ± 0.7) molecules/ nm^{-2} , thus compatible with both the theoretical predictions for thiol or thiyl SAM within the error bar. Note that this value assumes that the sulfur species contaminating the blank Au sample remain on the surface even after the MMC adsorption: if these species were substituted by MMC, totally or in part, the final SAM density would result larger, with an upper bound of (4.6 ± 0.6) molecules/ nm^{-2} (in the case that all the contaminants on the blank were substituted by MMC).

The XPS measurements performed on the same SAM allowed to probe the chemical bond at the SAM interface, and to postulate the sulfur oxidation state. Two signals have been assigned to adsorbed radicals: the more intense one (58.9%) to thiyl group and the second one (8.6%) to metastable thiyl in alternative adsorption sites. Two weaker components have been ascribed to free or physisorbed thiol species, either at the Au surface or in a second layer above the MMC SAM.

In conclusion, the comparison of theoretical and experimental densities cannot indicate the sulfur protonation state unambiguously, even if further analyses (including e.g. non-equilibrated dissociation reactions and reduced contaminant effects in the GIXRF measures) could clarify the point better. On the other hand, the XPS analysis indicates the prevalence of radical thiyls adsorbed on the gold surface, though in the presence of a lesser component of undissociated thiols.

Electronic Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:

Details on the FF parameterization, picture of high density thiol SAM, calculation of S-H dissociation free energy.

Acknowledgement

D. M., A. Za., M. L. and M. Co. acknowledge the financial support by Università del Piemonte Orientale (through the FAR-2019 funding program). E.C. F.F.L. P.H. Y.K. B.B. acknowledge the project 19ENV05 AEROMET II. The project 19ENV05 AEROMET II has received funding from the EMPIR programme co-financed by the Participating States and from the European Unions Horizon 2020 research and innovation programme.

References

- (1) Nuzzo, R. G.; Allara, D. L. Adsorption of Bifunctional Organic Disulfides on Gold Surfaces. *J. Am. Chem. Soc.* **1983**, *105*, 4481–4483.
- (2) Vericat, C.; Vela, M.; Benitez, G.; Carro, P.; Salvarezza, R. Self-assembled Monolayers of Thiols and Dithiols on Gold: New Challenges for a Well-known System. *Chem. Soc. Rev.* **2010**, *39*, 1805–1834.
- (3) Häkkinen, H. The Gold-sulfur Interface At the Nanoscale. *Nat. Chem.* **2012**, *4*, 443–455.
- (4) Laroussi, A.; Kot, M.; Flege, J.; Raouafi, N.; Mirsky, V. Self-assembled Monolayers from Symmetrical Di-thiols: Preparation, Characterization and Application for the Assembly of Electrochemically Active Films. *Appl. Surf. Sci.* **2020**, *513*, 145827.

- (5) Rovati, D.; Albini, B.; Galinetto, P.; Grisoli, P.; Bassi, B.; Pallavicini, P.; Dacarro, G.; Taglietti, A. High Stability Thiol-coated Gold Nanostars Monolayers with Photo-thermal Antibacterial Activity and Wettability Control. *Nanomaterials* **2019**, *9*, 1288.
- (6) Wang, Y.-S.; Yau, S.; Chau, L.-K.; Mohamed, A.; Huang, C.-J. Functional Biointerfaces Based on Mixed Zwitterionic Self-Assembled Monolayers for Biosensing Applications. *Langmuir* **2019**, *35*, 1652–1661.
- (7) Pengo, P.; Pasquato, L. Gold Nanoparticles Protected By Fluorinated Ligands: Syntheses, Properties and Applications. *J. Fluor. Chem.* **2015**, *177*, 2–10.
- (8) Lawton, T.; Uzarski, J.; Filocamo, S. A Multifunctional Surface That Simultaneously Balances Hydrophilic Enzyme Catalysis and Hydrophobic Water Repellency. *Chem. Eur. J.* **2016**, *22*, 12068–12073.
- (9) Bhuvana, M.; Dharuman, V. Influence of Alkane Chain Lengths and Head Groups on Tethering of Liposome-gold Nanoparticle on Gold Surface for Electrochemical DNA Sensing and Gene Delivery. *Sens. Actuators B Chem.* **2016**, *223*, 157–165.
- (10) Demirkol, D.; Yildiz, H.; Sayin, S.; Yilmaz, M. Enzyme Immobilization in Biosensor Constructions: Self-assembled Monolayers of Calixarenes Containing Thiols. *RSC Adv.* **2014**, *4*, 19900–19907.
- (11) Crudden, C.; Horton, J.; Ebralidze, I.; Zenkina, O.; McLean, A.; Drevniok, B.; She, Z.; Kraatz, H.-B.; Mosey, N.; Seki, T.; Keske, E.; Leake, J.; Rousina-Webb, A.; Wu, G. Ultra Stable Self-assembled Monolayers of N-heterocyclic Carbenes on Gold. *Nat. Chem.* **2014**, *6*, 409–414.
- (12) Valley, D.; Onstott, M.; Malyk, S.; Benderskii, A. Steric Hindrance of Photoswitching in Self-assembled Monolayers of Azobenzene and Alkane Thiols. *Langmuir* **2013**, *29*, 11623–11631.
- (13) Newton, L.; Slater, T.; Clark, N.; Vijayaraghavan, A. Self Assembled Monolayers (SAMs) on Metallic Surfaces (gold and Graphene) for Electronic Applications. *J. Mater. Chem. C* **2013**, *1*, 376–393.

- (14) Reik, M.; Calabro, M.; Griesemer, S.; Barry, E.; Bu, W.; Lin, B.; Rice, S. The Influence of Fractional Surface Coverage on the Core-core Separation in Ordered Monolayers of Thiol-ligated Au Nanoparticles. *Soft Matter* **2019**, *15*, 8800–8807.
- (15) Azzam, W.; Cyganik, P.; Witte, G.; Buck, M.; Wöll, C. Pronounced Odd-Even Changes in the Molecular Arrangement and Packing Density of Biphenyl-Based Thiol SAMs: a Combined STM and LEED Study. *Langmuir* **2003**, *19*, 8262–8270.
- (16) Heister, K.; Allara, D. L.; Bahnck, K.; Frey, S.; Zharnikov, M.; Grunze, M. Deviations from 1:1 Compositions in Self-Assembled Monolayers Formed from Adsorption of Asymmetric Dialkyl Disulfides on Gold. *Langmuir* **1999**, *15*, 5440–5443.
- (17) Gerlach, R.; Polanski, G.; Rubahn, H.-G. Growth of Ultrathin Organic Films on Au(111) Surfaces. *Thin Solid Films* **1998**, *318*, 270–272.
- (18) Whelan, C. M.; Barnes, C. J.; Walker, C. G.; Brown, N. M. Benzenethiol Adsorption on Au(111) Studied By Synchrotron ARUPS, HREELS and XPS. *Surface Science* **1999**, *425*, 195–211.
- (19) Tirotta, I.; Calloni, A.; Pigliacelli, C.; Brambilla, A.; Bussetti, G.; Duò, L.; Metrangolo, P.; Baldelli Bombelli, F. Chemical Characterization of Fluorinated/hydrogenated Mixed Monolayers Grafted on Gold Nanoparticles. *Journal of Fluorine Chemistry* **2018**, *206*, 99–107.
- (20) Pick, A.; Witte, G. Patterned Growth of Organic Semiconductors: Selective Nucleation of Perylene on Self-Assembled Monolayers. *Langmuir* **2016**, *32*, 8019–8028.
- (21) Pak, F.; Meral, K.; Altundaş, R.; Ekinici, D. Self-assembled Monolayers of Fluorene- and Nitrofluorene-terminated Thiols on Polycrystalline Gold Electrode: Electrochemical and Optical Properties. *J. Electroanal. Chem.* **2011**, *654*, 20–28.
- (22) Lewis, P. A.; Donhauser, Z. J.; Mantooth, B. A.; Smith, R. K.; Bumm, L. A.; Kelly, K. F.; Weiss, P. S. Control and Placement of Molecules Via Self-assembly. *Nanotechnology* **2001**, *12*, 231–237.

- (23) Tai, Y.; Shaporenko, A.; Eck, W.; Grunze, M.; Zharnikov, M. Depth Distribution of Irradiation-Induced Cross-Linking in Aromatic Self-Assembled Monolayers. *Langmuir* **2004**, *20*, 7166–7170.
- (24) Vericat, C.; Vela, M. E.; Benitez, G. A.; Gago, J. A. M.; Torrelles, X.; Salvarezza, R. C. Surface Characterization of Sulfur and Alkanethiol Self-assembled Monolayers on Au(111). *J. of Phys.: Condens. Matter* **2006**, *18*, R867–R900.
- (25) Mhlanga, N.; Ntho, T. A Theoretical Study of 4-Mercaptobenzoic Acid Assembled on Ag for Surface-enhanced Raman Scattering Applications. *Materials Today Communications* **2021**, *26*, 101698.
- (26) Chadha, R.; Das, A.; Kapoor, S.; Maiti, N. Surface-induced Dimerization of 2-thiazoline-2-thiol on Silver and Gold Nanoparticles: a Surface Enhanced Raman Scattering (SERS) and Density Functional Theoretical (DFT) Study. *J. Mol. Liq.* **2021**, *322*, 114536.
- (27) Engelbrekt, C.; Nazmutdinov, R.; Zinkicheva, T.; Glukhov, D.; Yan, J.; Mao, B.; Ulstrup, J.; Zhang, J. Chemistry of Cysteine Assembly on Au(100): Electrochemistry, in Situ STM and Molecular Modeling. *Nanoscale* **2019**, *11*, 17235–17251.
- (28) Roy, J.; Vasquez, E.; Pinto, H.; Kumari, S.; Walters, K.; Leszczynski, J. Computational and Experimental Approach to Understanding the Structural Interplay of Self-assembled End-terminated Alkanethiolates on Gold Surfaces. *Phys. Chem. Chem. Phys.* **2019**, *21*, 23320–23328.
- (29) Herrera, S.; Tasca, F.; Williams, F.; Calvo, E.; Carro, P.; Salvarezza, R. Surface Structure of 4-Mercaptopyridine on Au(111): a New Dense Phase. *Langmuir* **2017**, *33*, 9565–9572.
- (30) Guesmi, H.; Luque, N.; Santos, E.; Tielens, F. Does the S-H Bond Always Break After Adsorption of An Alkylthiol on Au(111)? *Chem. Eur. J.* **2017**, *23*, 1402–1408.
- (31) Peiretti, L.; Quaino, P.; Tielens, F. Competition Between Two High-density Assemblies of Poly(phenyl)thiols on Au(111). *J. Phys. Chem. C* **2016**, *120*, 25462–25472.

- (32) Bedford, E.; Humblot, V.; Méthivier, C.; Pradier, C.-M.; Gu, F.; Tielens, F.; Boujday, S. An Experimental and Theoretical Approach to Investigate the Effect of Chain Length on Amino-thiol Adsorption and Assembly on Gold. *Chem. Eur. J.* **2015**, *21*, 14555–14561.
- (33) First, J.; Webb, L. Agreement Between Experimental and Simulated Circular Dichroic Spectra of a Positively Charged Peptide in Aqueous Solution and on Self-Assembled Monolayers. *J. Phys. Chem. B* **2019**, *123*, 4512–4526.
- (34) Liu, X.; Lu, P.; Zhai, H.; Wu, Y. Temperature-dependent Surface Density of Alkylthiol Monolayers on Gold Nanocrystals. *Mater. Res. Express* **2018**, *5*, 035001.
- (35) Sridhar, D.; Gupta, R.; Rai, B. Effect of Surface Coverage and Chemistry on Self-assembly of Monolayer Protected Gold Nanoparticles: a Molecular Dynamics Simulation Study. *Phys. Chem. Chem. Phys.* **2018**, *20*, 25883–25891.
- (36) Meena, S.; Goldmann, C.; Nassoko, D.; Seydou, M.; Marchandier, T.; Moldovan, S.; Ersen, O.; Ribot, F.; Chanéac, C.; Sanchez, C.; Portehault, D.; Tielens, F.; Sulpizi, M. Nanophase Segregation of Self-Assembled Monolayers on Gold Nanoparticles. *ACS Nano* **2017**, *11*, 7371–7381.
- (37) Devi, J. Simulation Studies on Structural and Thermal Properties of Alkane Thiol Capped Gold Nanoparticles. *J. Mol. Graph. Model.* **2017**, *74*, 359–365.
- (38) Rodríguez González, M.; Carro, P.; Pensa, E.; Vericat, C.; Salvarezza, R.; Hernández Creus, A. The Role of a Double Molecular Anchor on the Mobility and Self-Assembly of Thiols on Au(111): the Case of Mercaptobenzoic Acid. *ChemPhysChem* **2017**, *18*, 804–811.
- (39) Luque, N. B.; Santos, E.; Andres, J.; Tielens, F. Effect of Coverage and Defects on the Adsorption of Propanethiol on Au(111) Surface: a Theoretical Study. *Langmuir* **2011**, *27*, 14514–14521.
- (40) Tielens, F.; Santos, E. AuS and SH Bond Formation/Breaking During the Formation of Alkanethiol SAMs on Au(111): a Theoretical Study. *J. Phys. Chem. C* **2010**, *114*, 9444–9452.

- (41) Rouhana, L. L.; Moussallem, M. D.; Schlenoff, J. B. Adsorption of Short-Chain Thiols and Disulfides Onto Gold Under Defined Mass Transport Conditions: Coverage, Kinetics, and Mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 16080–16091.
- (42) Kankate, L.; Turchanin, A.; Götzhäuser, A. On the Release of Hydrogen from the S-H Groups in the Formation of Self-Assembled Monolayers of Thiols. *Langmuir* **2009**, *25*, 10435–10438.
- (43) Inkpen, M.; Liu, Z.; Li, H.; Campos, L.; Neaton, J.; Venkataraman, L. Non-chemisorbed Gold–sulfur Binding Prevails in Self-assembled Monolayers. *Nat. Chem.* **2019**, *11*, 351–358.
- (44) Rzeźnicka, I. I.; Lee, J.; Maksymowych, P.; Yates, J. T. Nondissociative Chemisorption of Short Chain Alkanethiols on Au(111). *J. Phys. Chem. B* **2005**, *109*, 15992–15996.
- (45) Hasan, M.; Bethell, D.; Brust, M. The Fate of Sulfur-Bound Hydrogen on Formation of Self-Assembled Thiol Monolayers on Gold: ¹H NMR Spectroscopic Evidence from Solutions of Gold Clusters. *J. Am. Chem. Soc.* **2002**, *124*, 1132–1133.
- (46) Nuzzo, R. G.; Zegarski, B. R.; Dubois, L. H. Fundamental Studies of the Chemisorption of Organosulfur Compounds on Gold(111). Implications for Molecular Self-assembly on Gold Surfaces. *J. Am. Chem. Soc.* **1987**, *109*, 733–740.
- (47) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (48) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.
- (49) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

- (50) Khanna, V.; Monroe, J. I.; Doherty, M. F.; Peters, B. Performing Solvation Free Energy Calculations in LAMMPS Using the Decoupling Approach. *J. Comput. Aided Molec. Des.* **2020**, *34*, 641–646.
- (51) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (52) Love, J. C.; Estroff, L. A.; Kriebel, J. K.; Nuzzo, R. G.; Whitesides, G. M. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chemical reviews* **2005**, *105*, 1103–1170.
- (53) Hönicke, P.; Detlefs, B.; Nolot, E.; Kayser, Y.; Mühle, U.; Pollakowski, B.; Beckhoff, B. Reference-free grazing incidence x-ray fluorescence and reflectometry as a methodology for independent validation of x-ray reflectometry on ultrathin layer stacks and a depth-dependent characterization. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* **2019**, *37*, 041502.
- (54) Krumrey, M.; Ulm, G. High-accuracy detector calibration at the PTB four-crystal monochromator beamline. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **2001**, *467*, 1175–1178.
- (55) Lubeck, J.; Beckhoff, B.; Fliegau, R.; Holfelder, I.; Hönicke, P.; Müller, M.; Pollakowski, B.; Reinhardt, F.; Weser, J. A novel instrument for quantitative nanoanalytics involving complementary X-ray methodologies. *Review of Scientific Instruments* **2013**, *84*, 045106.
- (56) Scholze, F.; Procop, M. Modelling the response function of energy dispersive X-ray spectrometers with silicon detectors. *X-Ray Spectrometry: An International Journal* **2009**, *38*, 312–321.
- (57) Müller, M.; Hönicke, P.; Detlefs, B.; Fleischmann, C. Characterization of high-k nanolayers by grazing incidence X-ray spectrometry. *Materials* **2014**, *7*, 3147–3159.
- (58) Beckhoff, B. Reference-free X-ray spectrometry based on metrology using synchrotron radiation. *Journal of Analytical Atomic Spectrometry* **2008**, *23*, 845–853.

- (59) Sherman, J. The theoretical derivation of fluorescent X-ray intensities from mixtures. *Spectrochimica acta* **1955**, *7*, 283–306.
- (60) Schoonjans, T.; Brunetti, A.; Golosio, B.; del Rio, M. S.; Solé, V. A.; Ferrero, C.; Vincze, L. The xraylib library for X-ray–matter interactions. Recent developments. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2011**, *66*, 776–784.
- (61) Ravi, V.; Binz, J. M.; Rioux, R. M. Thermodynamic Profiles At the Solvated Inorganic–Organic Interface: the Case of Gold–Thiolate Monolayers. *Nano Lett.* **2013**, *13*, 4442–4448.
- (62) Jung, L. S.; Campbell, C. T. Sticking Probabilities in Adsorption of Alkanethiols from Liquid Ethanol Solution Onto Gold. *J. Phys. Chem. B* **2000**, *104*, 11168–11178.
- (63) Królikowska, A.; Kudelski, A.; Michota, A.; Bukowska, J. SERS Studies on the Structure of Thioglycolic Acid Monolayers on Silver and Gold. *Surf. Sci.* **2003**, *532-535*, 227–232.
- (64) Poirier, G. E.; Pylant, E. D. The Self-Assembly Mechanism of Alkanethiols on Au(111). *Science* **1996**, *272*, 1145–1148.
- (65) Soltwisch, V.; Hönicke, P.; Kayser, Y.; Eilbracht, J.; Probst, J.; Scholze, F.; Beckhoff, B. Element sensitive reconstruction of nanostructured surfaces with finite elements and grazing incidence soft X-ray fluorescence. *Nanoscale* **2018**, *10*, 6177–6185.
- (66) Krause, M. O. Atomic radiative and radiationless yields for K and L shells. *Journal of physical and chemical reference data* **1979**, *8*, 307–327.
- (67) Castner, D. G.; Hinds, K.; Grainger, D. W. X-ray photoelectron spectroscopy sulfur 2p study of organic thiol and disulfide binding interactions with gold surfaces. *Langmuir* **1996**, *12*, 5083–5086.
- (68) Zubrägel, C.; Deuper, C.; Schneider, F.; Neumann, M.; Grunze, M.; Schertel, A.; Wöll, C. The presence of two different sulfur species in self-assembled films of n-alkanethiols on Au and Ag surfaces. *Chemical physics letters* **1995**, *238*, 308–312.

- (69) Ishida, T.; Choi, N.; Mizutani, W.; Tokumoto, H.; Kojima, I.; Azebara, H.; Hokari, H.; Akiba, U.; Fujihira, M. High-resolution X-ray photoelectron spectra of organosulfur monolayers on Au (111): S (2p) spectral dependence on molecular species. *Langmuir* **1999**, *15*, 6799–6806.
- (70) Dufour, F.; Fresch, B.; Durupthy, O.; Chanéac, C.; Remacle, F. Ligand and solvation effects on the structural and electronic properties of small gold clusters. *The Journal of Physical Chemistry C* **2014**, *118*, 4362–4376.
- (71) Jia, J.; Kara, A.; Pasquali, L.; Bendounan, A.; Sirotti, F.; Esaulov, V. A. On sulfur core level binding energies in thiol self-assembly and alternative adsorption sites: An experimental and theoretical study. *The Journal of Chemical Physics* **2015**, *143*, 104702.

Electronic Supplementary Material (ESI)

Structure and Stability of 7-mercapto-4-methylcoumarin SAM on Gold: an Experimental and Computational analysis.

Davide Marchi¹, Eleonora Cara², Federico Ferrarese Lupi², Philipp Hönicke³, Yves Kayser³, Burkhard Beckhoff³, Micaela Castellino⁴, Alberto Zoccante¹, Michele Laus¹, and Maurizio Cossi¹

1) Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, via T. Michel 11, I-15121, Alessandria, Italy

2) Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce, 91, I-10135, Torino, Italy

3) Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany.

4) Department of Applied Science and Technology, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Turin, Italy.

Parameterization of the Force Field (FF).

The GROMOS 54A7 force field was used in all molecular dynamics (MD) simulations to model bonded interactions as well as the repulsion-dispersion interactions between the organic molecules. New Lennard-Jones (LJ) parameters were fitted to DFT calculations, as detailed below, for the Au-X pairwise interactions.

A model for the gold surface was obtained by cleaving a cluster of 60 gold atoms (5x4x3) from a periodic three-layers thick gold surface, previously optimized in CRYSTAL17 at the DFT/PW91 level with the basis set and pseudopotentials for core electrons of Hay and Wadt [1].

The molecule set used for the fitting procedure is reported in Table S1, along with the atom types and symbols used in the force field. In order to obtain a set of LJ parameters with minimal correlation, the fit was carried out incrementally, determining the optimal FF parameters for one or two atom types at each step.

Molecule	Atom types fitted	Atom types symbol
Benzene	Aromatic C, H	CA, HA
Benzenethiol	Thiol S, H	SH, HB
Methylbenzenethiol	<i>sp</i> ³ C, aliphatic H	C3, H

7-mercapto-4-methylcoumarin (MMC)	Oxygen	O
MMC thiyl radical	Thiyl S	S

Table S1. Molecule set used for the fitting procedure; the atom types and symbols fitted at each step are reported as well.

A geometry optimization was performed at the DFT/B3LYP level, using the Gaussian16 suite, for each molecule in Table S1 in close proximity with the surface of the gold cluster. Then, the distance between the molecule and the gold slab was varied incrementally, while keeping the molecular geometry fixed. At each step, the DFT energy of interaction between the molecule and the slab was computed. The basis set and pseudopotentials of Hay and Wadt [1] were used for gold atoms, while Dunning's correlation consistent cc-pVDZ [2] basis set was used for all other elements. Grimme's dispersion energy correction [3] was included in all calculations, and the basis set superposition error (BSSE) was corrected through the counterpoise method, as implemented in Gaussian16.

For benzene and benzenethiol, additional scans were performed keeping the plane of the aromatic ring perpendicular to the gold surface. This geometry is meant to mimic the preferred orientation in densely packed SAMs, a target system that the force field should describe correctly.

Finally, we optimized additional LJ parameters to describe the S-Au interactions involving MMC thiyl radicals. Though thiyls are expected to form covalent bonds with gold atoms, we used LJ functional form for this pair also, instead of the most used harmonic potential for instance, because to equilibrate the monolayers the thiyl units must be able to glide laterally on the metal surface, or even leave the surface if this lowers the energy of the system.

The fitting procedure was carried out on the model systems and configurations described above through an ad hoc python script. The script was interfaced with LAMMPS program to obtain single-point molecular mechanics energies, then the force field parameters were iteratively varied by employing the trust region reflective algorithm, up to fitting DFT energies.

All the optimized parameters are listed in Table S2; a comparison between DFT and FF energies is illustrated in Figure S1, showing the good agreement obtained with the fitting procedure.

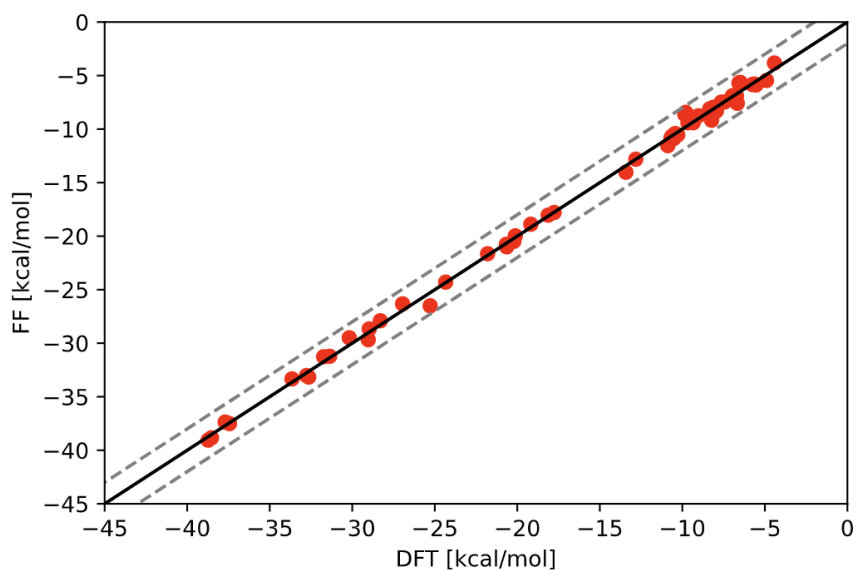


Figure S1. FF energies resulting from the fitting procedure for each atom type in the force field, compared to DFT energies. The black solid line denotes 1:1 correlation, while the grey dotted lines represent ± 2 kcal/mol mark.

Pair	ϵ (kcal/mol)	σ (Å)
Au – CA	0.571	3.228
Au – C3	0.775	2.996
Au – HA	0.020	2.975
Au – HB	0.500	2.555
Au – H	0.100	3.114
Au – O	0.146	3.435
Au – SH	1.169	3.243
A – S	13.406	2.230

Table S2. Lennard-Jones parameters for Au-X pair interactions in the FF fitted to DFT calculations.

High density thiyl SAM.

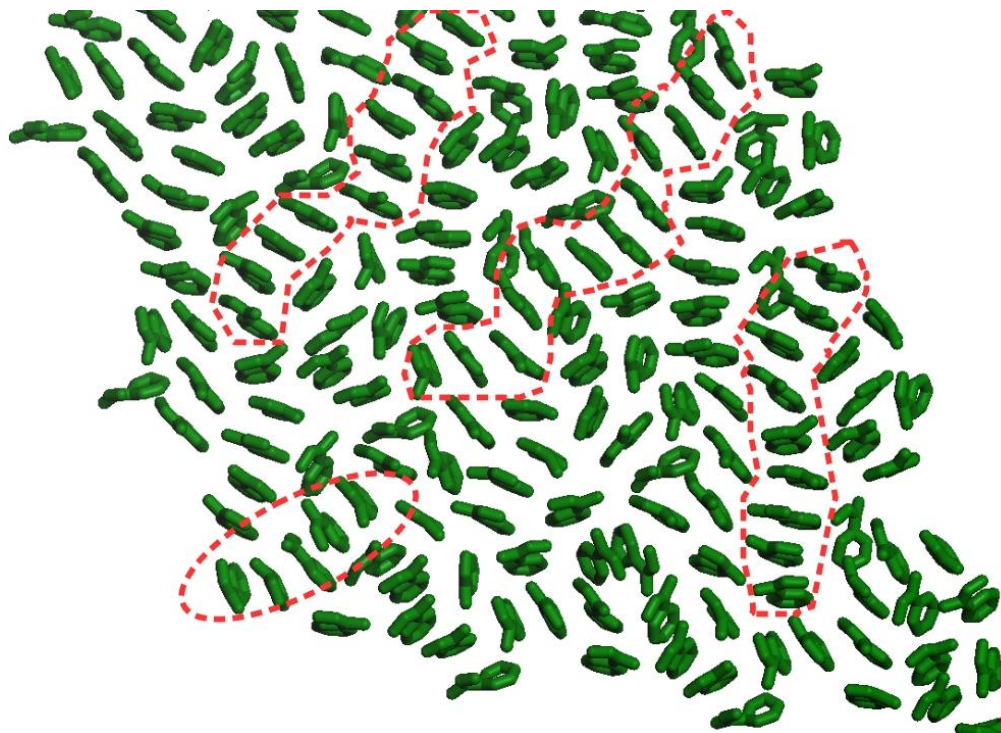


Figure S2. Some ordered stackings of MMC thiyl units in the densest monolayer (4.8 nm^{-2}).

DFT estimate of the S-H dissociation energy.

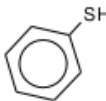
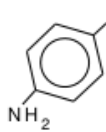
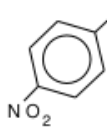
As explained in the text, the chemical potential of thiol SAMs depends on various contributions, among which the free energy change of the dissociation reaction:



Since the experimental value of ΔG_{diss} is not available for MMC, we resorted to DFT calculations to evaluate it. The bond dissociation energies (BDE) of thiol and molecular hydrogen can be directly computed (including semiempirical corrections for dispersion energy and counterpoise corrections for the BSSE), then thermal contributions for enthalpies and entropies are added with Boltzmann averages referred to translational, rotational and vibrational motions. The last term requires the calculation of harmonic frequencies, at the same level as BDE.

Before applying this procedure to MMC, the performance of different density functionals and basis sets has been tested against the experimental values of the dissociation enthalpy (ΔH_{diss}) for some simple aromatic thiols, namely benzenethiol, *p*-aminobenzenethiol and *p*-nitrobenzenethiol to consider both electron donating and electron withdrawing groups. We used enthalpies rather than free energies for this trial because the former are more easily available in the experimental literature; once determined the best choice of functional and basis, ΔG_{diss} was computed at the same level for MMC. The results are collected below in Tables S3-S5: one can see that none of the tested functionals is able to reproduce all the experimental enthalpies fairly, but the best performance is obtained by B3P86 with the largest basis set, namely cc-pvtz.

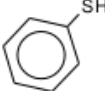
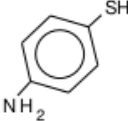
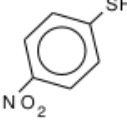
Then the value of the free energy change for reaction 1 involving MMC was computed at the B3P86/cc-pvtz level, obtaining $\Delta G_{diss} = 130.0$ kJ/mol.

B3LYP			
Basis set			
6-31G(d)	90.3	69.3	103.0
6-31G(d,p)	94.9	73.5	107.4
6-31+G(d,p)	96.6	76.2	109.6
6-311G(d,p)	*	77.7	109.8
6-31G(2d,2p)	99.6	78.3	111.8
cc-pvdz	92.5	71.6	105.2
cc-pvtz	94.0	81.8	115.6
Exp. ^(a)	113.0	75.3	125.6

(a) Bond dissociation enthalpies: ref. 1 for thiols, 2 for H₂.

* Convergence problems.

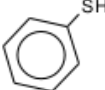
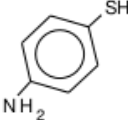
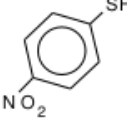
Table S3. Enthalpy change (kJ/mol) for reaction 1, computed at the B3LYP level with various basis sets, including thermal contributions at 298 K and Grimme's dispersion corrections, and compared to experimental values.

B3P86			
Basis set			
6-31G(d)	98.4	76.9	111.3
6-31G(d,p)	103.1	81.1	115.8
6-31+G(d,p)	104.8	83.7	117.9
6-311G(d,p)	*	86.1	119.0
6-31G(2d,2p)	108.4	86.5	120.8
cc-pvdz	110.2	87.9	121.3
cc-pvtz	112.0	89.7	124.5
Exp. ^(a)	113.0	75.3	125.6

(a) Bond dissociation enthalpies: ref. 1 for thiols, 2 for H₂.

* Convergence problems.

Table S4. Enthalpy change (kJ/mol) for reaction 1, computed at the B3P86 level with various basis sets, including thermal contributions at 298 K, and compared to experimental values.

BLYP			
Basis set			
6-31G(d)	82.0	56.7	94.6
6-31G(d,p)	86.6	61.0	*
6-31+G(d,p)	88.7	64.0	*
6-311G(d,p)	89.4	65.0	101.3
6-31G(2d,2p)	91.5	66.1	103.5
cc-pvdz	95.1	69.3	106.2

cc-pvtz	95.9	70.1	107.7
Exp. ^(a)	113.0	75.3	125.6

(a) Bond dissociation enthalpies: ref. 1 for thiols, 2 for H₂.

* Convergence problems.

Table S5. Enthalpy change (kJ/mol) for reaction 1, computed at the BLYP level with various basis sets, including thermal contributions at 298 K, and compared to experimental values.

Chapter 7

Discussion and future perspectives

In the course of this PhD program, I applied computational modelling methods to systems composed of molecules, often in interaction with a solid surface (Chapter 3, Chapter 4, Chapter 5, Chapter 6, Appendix B, Appendix C), to assist in the understanding and interpretation of experimental data. The bulk of the work focused on molecules containing sulfur forming self-assembled monolayers (SAMs) on a gold (111) surface. Two experimentally interesting target molecules were studied: 7-mercapto-4-methylcoumarin and 3-mercaptopropionic acid, both either in their dissociated (S-Au covalent bond) or undissociated (SH \cdots Au non-covalent interaction) states, see Figure 7.1.

The systems were studied with a multiscale approach, combining DFT calculations and force field (FF) molecular dynamics (MD) simulations. While FF parameters for intramolecular and intermolecular interactions of the organic species could be taken directly from widely-implemented, widely-applicable FFs (here either UFF or GROMOS 54A7), all gold-molecule interactions were fitted to DFT B3LYP-D3 calculations using an ad hoc python script interfacing with LAMMPS.

In general, the strategy of using simple FF Lennard-Jones (LJ) functional forms fitted to ab initio calculations has proven promising, allowing to perform MD simulations with good accuracy even in the absence of force fields specifically designed for those systems. This approach has potential applications that extend beyond the ones shown here for thiol-gold SAMs (*vide infra*), but it entails some open problems. Chemically-interesting systems in waste removal [89], catalysis [90], and many more fields are often composed of molecules interacting with some kind of surface (see also Chapter 3, Chapter 4, Chapter 5, Chapter 6, Appendix B, and Appendix C). For systems like these, dispersion and other non-covalent interactions between the surface and the molecules are of utmost importance, having the most relevant energetic contributions. As detailed in Subsection 1.5.3, non-bonded inter-

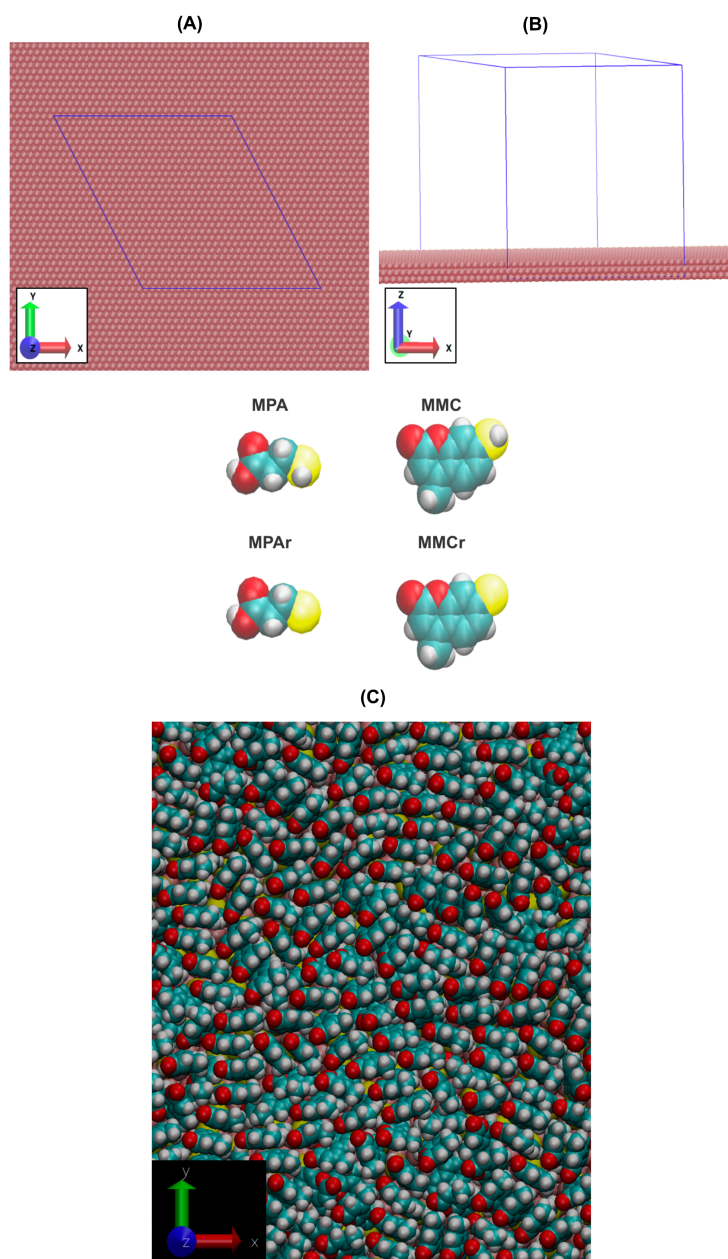
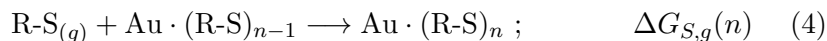
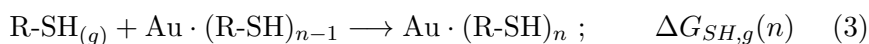
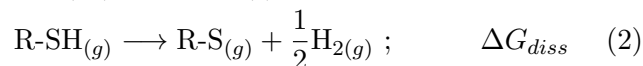


Figure 7.1: Top (A) and side (B) view of model used for the gold surface in MD simulations; the periodic box is highlighted in blue. Molecules studied in the simulations: thiol MPA, thiol MPAr, thiol MMC, thiol MMCr. (C) An equilibrated monolayer of MMCr at the computational thermodynamic density of $4.0 \text{ molecules / nm}^{-2}$.

actions are the hardest to capture computationally even with *ab initio* calculations. In fact, DFT B3LYP calculations (used here in the FF fitting procedure) require themselves a correction term, such as Grimme’s D3, to account for long-range non-covalent interactions. As detailed in subsection 1.5.3.3 many possible solutions exist to account for dispersion energies in *ab initio* calculations, but it is often problematic to check the performance of such approaches against experimental data, since the comparison between computed and measurable quantities is not straightforward. Nevertheless, once an *ab initio* method has been deemed sufficient, it is interesting to note that a single LJ curve for each Au-X pair was shown to perform fairly against B3LYP-D3 energies in configurations similar to those used in the fitting procedure, and of experimental interest for this thesis. Further work could address the question of accuracy/simplicity trade-off in designing FF parameters fitted to higher-level calculations. From the work presented in Chapter 3, Chapter 4, Chapter 5, and Chapter 6, it can be seen how, for some systems, it seems possible to capture the majority of both short-range *ab initio* and dispersion energy terms with a simple pairwise LJ potential, with an accuracy that is adequate for calculations on model systems of experimental interest; plausibly, a greater accuracy still could be achieved with a small increase in the complexity of the FF, perhaps with the inclusion of charge effects in the gold-molecule interactions.

The fitted FF parameters were applied in MD simulations of monolayers of either MPA, MPAr, MMC, or MMCr (see Figure 7.1), and Bennett acceptance ratio (BAR) calculations were used in conjunction with DFT calculations to evaluate the chemical potential of a molecule in the monolayer as a function of the SAM density. The reference reactions are:



$$\begin{aligned} \mu_{SH}(n) &= \Delta G_{desolv} + \Delta G_{SH,g}(n) \\ \mu_S(n) &= \Delta G_{desolv} + \Delta G_{diss} + \Delta G_{S,g}(n) \end{aligned}$$

The results are reported in Figure 7.2.

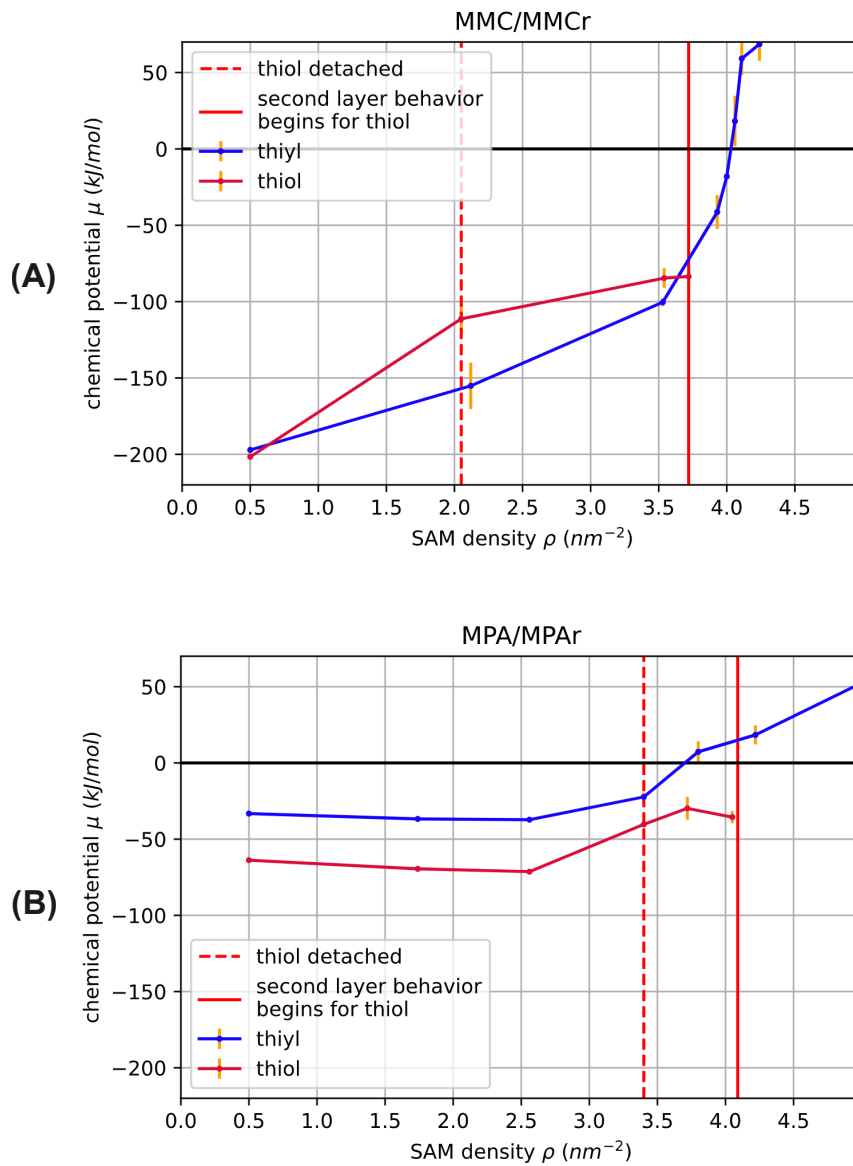


Figure 7.2: Chemical potential as a function of the monolayer density for: (A) thiol MMC and thiyol MMCr, (B) thiol MPA and thiyol MPAr.

The experimentally-determined value for the SAM density of 7-mercapto-4-methylcoumarin is given here:

$$\rho_{exp} = 4.6 \pm 0.6$$

It can be seen how the two simulated species, MMC and MMCr, give a result that is very similar, with MMCr being compatible with the experimental value within the uncertainty:

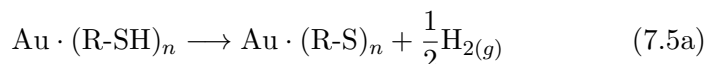
$$\rho_{(MMC)} = 3.7$$

$$\rho_{(MMCr)} = 4.0$$

while an experimental value for 3-mercaptopropionic acid is not yet available. As stated in Section 4.4, the similarity of the results obtained for thiols rather than thiy radicals is due to the free energy of S-H bond dissociation that enters the thermodynamic calculations for radicals, lowering the expected monolayer density despite the much greater molecule/gold interaction.

Since some controversy still exist in the literature about whether thiols are expected to dissociate at the interface or not, and there is reason to assume that different molecules and preparation methods could lead to S/Au interactions of different nature, further modelling investigations in this sense are proposed. The computational methodology developed here could be applied to a wider range of molecules to test for other specific behaviors in analogy to those shown by MPA and MMC (see Figure 7.3); a wider data sample could allow to evaluate whether the small differences computed by the models in distinguishing $\rho_{(SH...Au)}$ from $\rho_{(S-Au)}$ are to be considered meaningful.

Moreover, the thermodynamic model could be probed further to examine the hypothesized reaction:



at the interface, and consider the effects of a reaction of formation of the monolayer that is not at thermodynamic equilibrium. By hypothesizing *a priori* that, during the reaction of formation of the SAM, thiols do dissociate at the interface to form S-Au covalent bonds, and that hydrogen consequently leaves the gold surface in molecular form, the thermodynamic calculations presented here change considerably, and preliminary data (not presented) shows that different molecules (e.g. MPA rather than MMC) are affected in a significantly different way.

Another possible field of investigation for future simulations is to consider a more realistic model at the atomic level for the gold surface. In fact, although the approximation of a flat, unreconstructed surface

is a useful model often used in the literature [62, 64, 91], recent studies highlight how, upon thiol adsorption, gold surfaces undergo reconstruction [63]. This further complicates the description of the interface, widening the range of possible surface sites occupied by a molecule. It could be interesting to explore the modelling limits of MD simulations in this sense, and examine the magnitude of change in the computed ρ_{comp} resulting from a reconstructed, rather than flat, model of the gold surface.

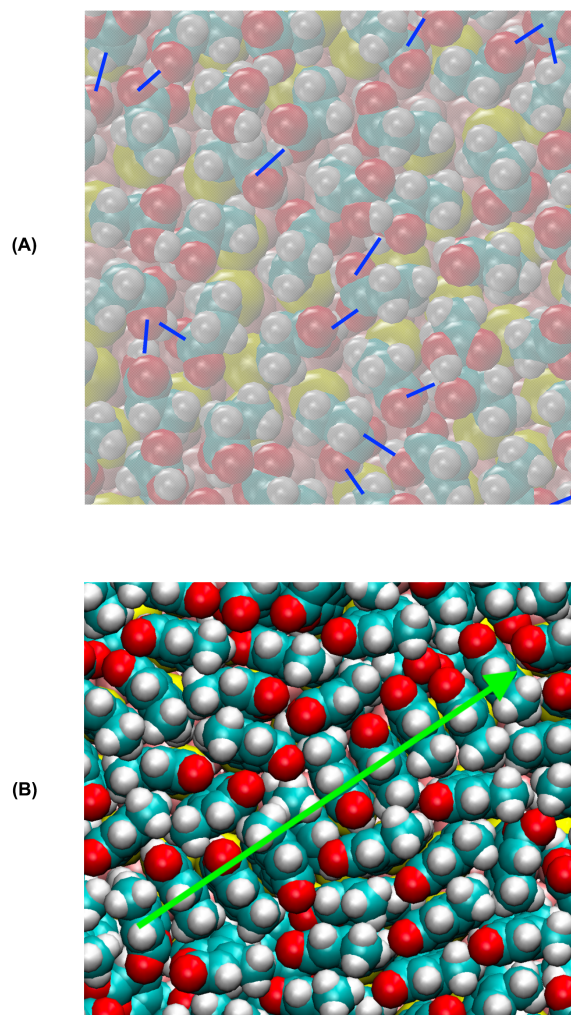


Figure 7.3: (A) H-bonds exhibited by MPA/MPAr at the SAM density of 3.8 molecules / nm² in a representative snapshot from the MD trajectory. The H-bonds are highlighted in blue; they can involve two or more molecules and both the -SH and -COOH functional groups. (B) MMCr monolayer at the density of 4.84 molecules / nm²; a particularly long row of ordered molecules is highlighted by the green arrow.

Different packing dynamics are observed between (A) and (B): MPA/MPAr (A) is flexible and capable of reorienting to form stabilizing H-bonds, while MMCr/MMCr (B) is rigid and packs favorably only along the axis perpendicular to the plane of the coumarin rings. This leads to the different shape of the chemical potential curves between MMCr/MMCr and MPA/MPAr shown in Figure 7.2.

Bibliography

- [1] David Goldsman, Richard Nance, and James Wilson. A brief history of simulation. pages 310 – 313, 01 2010.
- [2] Eric Winsberg. Computer Simulations in Science. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2019 edition, 2019.
- [3] Eric Winsberg. Simulated experiments: Methodology for a virtual world. *Philosophy of Science*, 70(1):105–125, 2003.
- [4] Till Grüne-Yanoff and Paul Weirich. The philosophy and epistemology of simulation: A review. *Simulation & Gaming*, 41(1):20–50, 2010.
- [5] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Number 1 in Computational science series. Academic Press, San Diego, 2nd ed edition, 2002.
- [6] Timothy O’Connor. Emergent Properties. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2020 edition, 2020.
- [7] Paul Humphreys. The philosophical novelty of computer simulation methods. *Synthese*, 169(3):615–626, Aug 2009.
- [8] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, 27(5):1208–1209, 1957.
- [9] W. W. Wood and J. D. Jacobson. Preliminary results from a recalculation of the monte carlo equation of state of hard spheres. *The Journal of Chemical Physics*, 27(5):1207–1208, 1957.
- [10] Wendy S. Parker. Franklin, holmes, and the epistemology of computer simulation. *International Studies in the Philosophy of Science*, 22(2):165–183, 2008.

- [11] A. Franklin. The epistemology of experiment. In *The uses of experiment*, pages 437–460.
- [12] Berend Smit, Sami Karaborni, and J. Ilja Siepmann. Computer simulations of vapor–liquid phase equilibria of n-alkanes. *The Journal of Chemical Physics*, 102(5):2126–2140, 1995.
- [13] Christopher J. Cramer. *Essentials of computational chemistry: theories and models*. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2nd ed edition, 2004.
- [14] H. Bernhard Schlegel. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *Journal of Computational Chemistry*, 24(12):1514–1527, 2003.
- [15] Chris M. Handley and Paul L. A. Popelier. Potential energy surfaces fitted by artificial neural networks. *The Journal of Physical Chemistry A*, 114(10):3371–3383, Mar 2010.
- [16] Junmei Wang and Tingjun Hou. Application of molecular dynamics simulations in molecular property prediction ii: Diffusion coefficient. *Journal of Computational Chemistry*, 32(16):3505–3519, 2011.
- [17] Pratyush Tiwary and Axel van de Walle. *A Review of Enhanced Sampling Approaches for Accelerated Molecular Dynamics*, pages 195–221. Springer International Publishing, Cham, 2016.
- [18] An-hui Wang, Zhi-chao Zhang, and Guo-hui Li. Advances in enhanced sampling molecular dynamics simulations for biomolecules. *Chinese Journal of Chemical Physics*, 32(3):277–286, 2019.
- [19] William H. Cropper. Rudolf clausius and the road to entropy. *American Journal of Physics*, 54(12):1068–1074, 1986.
- [20] Max Planck. Ueber das gesetz der energieverteilung im normal-spectrum. *Annalen der Physik*, 309(3):553–563, 1901.
- [21] Radu Iftimie, Peter Minary, and Mark E. Tuckerman. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proceedings of the National Academy of Sciences*, 102(19):6654–6659, 2005.
- [22] Andrew R. Leach. Chapter 6 - Computer Simulation Methods. In *Molecular modelling: principles and applications*. Prentice Hall, Harlow, England ; New York, 2nd ed edition, 2001.

- [23] Andrew R. Leach. Chapter 4 - Empirical Force Field Models: Molecular Mechanics. In *Molecular modelling: principles and applications*. Prentice Hall, Harlow, England ; New York, 2nd ed edition, 2001.
- [24] Thomas A Halgren and Wolfgang Damm. Polarizable force fields. *Current Opinion in Structural Biology*, 11(2):236–242, 2001.
- [25] Jesse G. McDaniel and J. R. Schmidt. First-principles many-body force fields from the gas phase to liquid: A “universal” approach. *The Journal of Physical Chemistry B*, 118(28):8042–8053, 2014. PMID: 24655231.
- [26] Jonathan Barnoud and Luca Monticelli. *Coarse-Grained Force Fields for Molecular Simulations*, pages 125–149. Springer New York, New York, NY, 2015.
- [27] Jaewoon Jung, Wataru Nishima, Marcus Daniels, Gavin Bascom, Chigusa Kobayashi, Adetokunbo Adedoyin, Michael Wall, Anna Lappala, Dominic Phillips, William Fischer, Chang-Shung Tung, Tamar Schlick, Yuji Sugita, and Karissa Y. Sanbonmatsu. Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *Journal of Computational Chemistry*, 40(21):1919–1930, 2019.
- [28] Andrew R. Leach. Chapter 7 - Molecular Dynamics Simulation Methods. In *Molecular modelling: principles and applications*. Prentice Hall, Harlow, England ; New York, 2nd ed edition, 2001.
- [29] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.
- [30] Charles H Bennett. Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976.
- [31] Paul Adrien Maurice Dirac. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 123(792):714–733, April 1929.
- [32] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part iii. term values and intensities in series in optical spectra. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(3):426–437, 1928.

- [33] J. C. Slater. Note on hartree's method. *Phys. Rev.*, 35:210–211, Jan 1930.
- [34] V. Fock. Naherungsmethode zur losung des quantenmechanischen mehrkorperproblems. *Zeitschrift Physik*, 61(1-2):126–148, January 1930.
- [35] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34:1293–1322, Nov 1929.
- [36] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23:69–89, Apr 1951.
- [37] C. F. Matta and R. J. Gillespie. Understanding and interpreting molecular electron density distributions. *Journal of Chemical Education*, 79(9):1141, 2002.
- [38] L. H. Thomas. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(5):542–548, January 1927.
- [39] Fermi Enrico. Un Metodo Statistico per la Determinazione di alcune Prioprietà dell'Atomo. *Rendiconti Accademia Nazionale del Lincei*(6):602–607, 1927.
- [40] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.
- [41] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.
- [42] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, Sep 1988.
- [43] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566–569, Aug 1980.
- [44] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 59:1200, August 1980.
- [45] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, April 1993.

- [46] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.
- [47] John P. Perdew, Adrienn Ruzsinszky, Jianmin Tao, Viktor N. Staroverov, Gustavo E. Scuseria, and Gábor I. Csonka. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *The Journal of Chemical Physics*, 123(6):062201, 2005.
- [48] Narbe Mardirossian and Martin Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics*, 115(19):2315–2372, 2017.
- [49] Anthony J. Stone. Chapter 1 - physical basis of intermolecular interactions. In Alberto Otero de la Roza and Gino A. DiLabio, editors, *Non-Covalent Interactions in Quantum Chemistry and Physics*, pages 3–26. Elsevier, 2017.
- [50] JoséM. Pérez-Jordá and A.D. Becke. A density-functional study of van der waals forces: rare gas diatomics. *Chemical Physics Letters*, 233(1):134–137, 1995.
- [51] Sándor Kristyán and Péter Pulay. Can (semi)local density functional theory account for the london dispersion forces? *Chemical Physics Letters*, 229(3):175–180, 1994.
- [52] Pavel Hobza, Jiří šponer, and Tomáš Reschel. Density functional theory and molecular clusters. *Journal of Computational Chemistry*, 16(11):1315–1325, 1995.
- [53] Jireí Seponer, Jerzy Leszczynski, and Pavel Hobza. Base stacking in cytosine dimer. a comparison of correlated ab initio calculations with three empirical potential models and density functional theory calculations. *Journal of Computational Chemistry*, 17(7):841–850, 1996.
- [54] Lars Goerigk. Chapter 6 - a comprehensive overview of the dft-d3 london-dispersion correction. In Alberto Otero de la Roza and Gino A. DiLabio, editors, *Non-Covalent Interactions in Quantum Chemistry and Physics*, pages 195–219. Elsevier, 2017.
- [55] Roberto Peverati and Donald G. Truhlar. Quest for a universal density functional: the accuracy of density functionals across a

- broad spectrum of databases in chemistry and physics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2011):20120476, 2014.
- [56] Gino A. DiLabio. Chapter 7 - atom-centered potentials for noncovalent interactions and other applications. In Alberto Otero de la Roza and Gino A. DiLabio, editors, *Non-Covalent Interactions in Quantum Chemistry and Physics*, pages 221–240. Elsevier, 2017.
- [57] Elsebeth Schröder, Valentino R. Cooper, Kristian Berland, Bengt I. Lundqvist, Per Hyldgaard, and Timo Thonhauser. Chapter 8 - the vdw-df family of nonlocal exchange-correlation functionals. In Alberto Otero de la Roza and Gino A. DiLabio, editors, *Non-Covalent Interactions in Quantum Chemistry and Physics*, pages 241–274. Elsevier, 2017.
- [58] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132(15):154104, 2010.
- [59] Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Hagen Neugebauer, Sebastian Spicher, Christoph Bannwarth, and Stefan Grimme. A generally applicable atomic-charge dependent london dispersion correction. *The Journal of Chemical Physics*, 150(15):154122, 2019.
- [60] Erin R. Johnson. Chapter 5 - the exchange-hole dipole moment dispersion model. In Alberto Otero de la Roza and Gino A. DiLabio, editors, *Non-Covalent Interactions in Quantum Chemistry and Physics*, pages 169–194. Elsevier, 2017.
- [61] Ralph G. Nuzzo and David L. Allara. Adsorption of bifunctional organic disulfides on gold surfaces. *Journal of the American Chemical Society*, 105(13):4481–4483, June 1983.
- [62] Abraham Ulman. Formation and structure of self-assembled monolayers. *Chemical Reviews*, 96(4):1533–1554, 1996. PMID: 11848802.
- [63] J. Christopher Love, Lara A. Estroff, Jennah K. Kriebel, Ralph G. Nuzzo, and George M. Whitesides. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chemical Reviews*, 105(4):1103–1170, 2005. PMID: 15826011.

- [64] Lauren Newton, Thomas Slater, Nick Clark, and Aravind Vijayaraghavan. Self assembled monolayers (sams) on metallic surfaces (gold and graphene) for electronic applications. *J. Mater. Chem. C*, 1:376–393, 2013.
- [65] C. Vericat, M. E. Vela, G. Benitez, P. Carro, and R. C. Salvarezza. Self-assembled monolayers of thiols and dithiols on gold: new challenges for a well-known system. *Chem. Soc. Rev.*, 39:1805–1834, 2010.
- [66] Héctor M Saavedra, Thomas J Mullen, Pengpeng Zhang, Daniel C Dewey, Shelley A Claridge, and Paul S Weiss. Hybrid strategies in nanolithography. *Reports on Progress in Physics*, 73(3):036501, jan 2010.
- [67] Younan Xia and George M. Whitesides. Soft lithography. *Annual Review of Materials Science*, 28(1):153–184, 1998.
- [68] Nirmalya K Chaki and K Vijayamohanan. Self-assembled monolayers as a tunable platform for biosensor applications. *Biosensors and Bioelectronics*, 17(1):1–12, 2002.
- [69] Yantian Wang, Yanxiu Zhou, Jonathon Sokolov, Basil Rigas, Kalle Levon, and Miriam Rafailovich. A potentiometric protein sensor built with surface molecular imprinting method. *Biosensors & Bioelectronics*, 24(1):162–166, September 2008.
- [70] E. C. Le Ru, E. Blackie, M. Meyer, and P. G. Etchegoin. Surface enhanced raman scattering enhancement factors: A comprehensive study. *The Journal of Physical Chemistry C*, 111(37):13794–13803, 2007.
- [71] Hazar Guesmi, Noelia B. Luque, Elizabeth Santos, and Frederik Tielens. Does the sh bond always break after adsorption of an alkylthiol on au(111)? *Chemistry – A European Journal*, 23(6):1402–1408, 2017.
- [72] Michael S. Inkpen, Zhen-Fei Liu, Haixing Li, Luis M. Campos, Jeffrey B. Neaton, and Latha Venkataraman. Non-chemisorbed gold-sulfur binding prevails in self-assembled monolayers. *Nature Chemistry*, 11(4):351–358, April 2019.
- [73] Izabela I. Rzeźnicka, Junseok Lee, Petro Maksymovych, and John T. Yates. Nondissociative chemisorption of short chain alkanethiols on au(111). *The Journal of Physical Chemistry B*, 109(33):15992–15996, 2005. PMID: 16853029.

- [74] Masihul Hasan, Donald Bethell, and Mathias Brust. The fate of sulfur-bound hydrogen on formation of self-assembled thiol monolayers on gold: 1h nmr spectroscopic evidence from solutions of gold clusters. *Journal of the American Chemical Society*, 124(7):1132–1133, 2002. PMID: 11841257.
- [75] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992.
- [76] Nathan Schmid, Andreas P. Eichenberger, Alexandra Choutko, Sereina Riniker, Moritz Winger, Alan E. Mark, and Wilfred F. van Gunsteren. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*, 40(7):843–856, July 2011.
- [77] P. Jeffrey Hay and Willard R. Wadt. Ab initio effective core potentials for molecular calculations. potentials for k to au including the outermost core orbitals. *The Journal of Chemical Physics*, 82(1):299–310, 1985.
- [78] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian~16 Revision A.03, 2016. Gaussian Inc. Wallingford CT.
- [79] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.
- [80] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G.

- Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.*, 271:108171, 2022.
- [81] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [82] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.
- [83] C. J. Casewit, K. S. Colwell, and A. K. Rappe. Application of a universal force field to organic molecules. *Journal of the American Chemical Society*, 114(25):10035–10046, 1992.
- [84] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.
- [85] Fabrice Dénès, Mark Pichowicz, Guillaume Povie, and Philippe Renaud. Thiyl Radicals in Organic Synthesis. *Chemical Reviews*, 114(5):2587–2693, March 2014.
- [86] Crc handbook of chemistry and physics, 88th ed editor-in-chief: David r. lide (national institute of standards and technology) crc press/taylor & francis group: Boca raton, fl. 2007. 2640 pp. \$139.95. isbn 0-8493-0488-1. *Journal of the American Chemical Society*, 130(1):382–382, 2008.
- [87] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. Packmol: A package for building initial configurations for molecu-

- lar dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, 2009.
- [88] Alpeshkumar K. Malde, Le Zuo, Matthew Breeze, Martin Stroet, David Poger, Pramod C. Nair, Chris Oostenbrink, and Alan E. Mark. An automated force field topology builder (atb) and repository: Version 1.0. *Journal of Chemical Theory and Computation*, 7(12):4026–4037, 2011. PMID: 26598349.
- [89] Vittoria Sacchetto, Chiara Bisio, Diana F. Olivas Olivera, Geo Paul, Giorgio Gatti, Ilaria Braschi, Gloria Berlier, Maurizio Cossi, and Leonardo Marchese. Interactions of toluene and n-hexane on high silica zeolites: An experimental and computational model study. *The Journal of Physical Chemistry C*, 119(44):24875–24886, 2015.
- [90] Varinia Bernales, Manuel A. Ortuño, Donald G. Truhlar, Christopher J. Cramer, and Laura Gagliardi. Computational design of functionalized metal–organic framework nodes for catalysis. *ACS Central Science*, 4(1):5–19, 2018. PMID: 29392172.
- [91] Daniel Käfer, Gregor Witte, Piotr Cyganik, Andreas Terfort, and Christof Wöll. A comprehensive study of self-assembled monolayers of anthracenethiol on gold: Solvent effects, structure, and stability. *Journal of the American Chemical Society*, 128(5):1723–1732, 2006. PMID: 16448148.

Appendices

Appendix A

Optimizing the relaxivity at high fields:
systematic variation of the rotational
dynamics in polynuclear Gd-complexes
based on the AAZTA ligand

RESEARCH ARTICLE

View Article Online
View Journal



Cite this: DOI: 10.1039/d1qj00904d

Optimizing the relaxivity at high fields: systematic variation of the rotational dynamics in polynuclear Gd-complexes based on the AAZTA ligand†

Lorenzo Tei,^a Giuseppe Gugliotta,^a Davide Marchi,^a Maurizio Cossi,^a Simonetta Geninatti Crich,^b and Mauro Botta^{b*}

A homogeneous series of polynuclear structures containing from 2 to 6 GdAAZTA complexes (AAZTA = 6-amino-6-methylperhydro-1, 4-diazepine tetraacetic acid) were synthesized covering a broad range of molecular weights, ca. 1200–6000 Da. A frequency and temperature dependent ¹H and ¹⁷O NMR relaxometric study on the Gd(III) polynuclear complexes clearly highlights a considerable gain of relaxivity (per Gd) passing from the monomer to the hexanuclear complex, with an enhancement of +370% at 1.5 T and 298 K (+470% at 310 K). In particular, the relaxivity of the hexamer reaches a remarkable value of 28.2 mM⁻¹ s⁻¹ at 1.5 T and 298 K (23.8 mM⁻¹ s⁻¹ at 310 K). The NMR study, supported by DFT calculations, allows analysis in detail of the dependence of the rotational dynamics on the size and molecular geometry of the multimers, considering both the global tumbling of the system and the local motions of the monomer units. In addition, T₁-weighted phantom MR-images at 1, 3 and 7 T on selected polynuclear complexes highlighted the effective signal enhancement of the new MRI probes at clinically relevant magnetic field strengths in comparison with the clinically approved contrast agent ProHance®. Finally, the results obtained enable us to highlight effective strategies for relaxation enhancement, each suitable for a well-defined range of applied magnetic field strength.

Received 20th July 2021,
Accepted 3rd September 2021

DOI: 10.1039/d1qj00904d

rscl.li/frontiers-inorganic

Introduction

Soon after its initial discovery in 1972, Magnetic Resonance Imaging (MRI) emerged as one of the prominent diagnostic clinical modalities thanks to its superb spatial and temporal resolution, the absence of ionizing radiation and the remarkable ability to visualize virtually every part of the human body.¹ As MRI relies on the magnetic properties of water protons, whose concentration is of the order of 90 mol L⁻¹ in biological tissues, the contrast in a MR image is mainly determined by differences in tissue water proton relaxation times T₁ and T₂.² However, MRI suffers from a limited inherent sensitivity; hence the use of contrast agents (CAs), *i.e.* chemicals able to alter markedly the water proton relaxation times in the tissues where they distribute, has become a widespread procedure.

CAs allow achievement of remarkable improvements in medical diagnosis in terms of higher specificity, better tissue characterization, reduction of image artefacts and functional information.^{3,4} Currently, about half of all clinical MR scans make use of contrast enhancing media.⁴ While the contrast agents currently used were developed about 30 years ago, when most of the clinical scanners operated at 0.47 T, the tendency in MRI development is towards even higher magnetic field strengths. The clinical 1.5 T scanners are gradually giving way to a new generation of 3 T scanners, which might be ultimately replaced by the 7 T scanners, already common in research environments.⁵ This trend is explained on the basis of the search for a greater signal-to-noise ratio (SNR), higher spatial resolution and/or reduced acquisition times. For experimental animal studies, even higher fields are commonly applied: from 9.4 up to 17.6 T.⁶ However, a similar technological-instrumental development has not taken place in the field of contrast agents. Currently employed CAs are small, rapidly tumbling Gd^{III} chelates with low relaxivity, r₁, that decreases smoothly with the field strength.^{3,4} Relaxivity is the enhancement of the longitudinal relaxation rate (R₁ = 1/T₁) of the water protons normalized to 1 mM concentration of paramagnetic ions. Improved systems of higher efficiency were developed for MR-angiography, exploiting the non-covalent binding to slowly

^aDipartimento di Scienze e Innovazione Tecnologica and Magnetic Resonance Platform (PRISMA-UPO), Università del Piemonte Orientale "A. Avogadro", Viale T. Michel 11, I-15121 Alessandria, Italy. E-mail: mauro.botta@uniupo.it

^bDepartment of Molecular Biotechnology and Health Sciences and Molecular Imaging Center, Università di Torino, Via Nizza 52, 10126 Torino, Italy

† Electronic supplementary information (ESI) available: Further experimental details and relaxometric data. NMR and HPLC characterization of the ligands. DFT calculated minima. See DOI: 10.1039/d1qj00904d

tumbling macromolecules (*i.e.*, Human Serum Albumin). However, this strategy is effective for achieving a large r_1 enhancement between *ca.* 0.5 and 1 T, while at larger fields relaxivity decreases abruptly and can reach values even lower than those of small chelates.⁷ Clinically used Gd-based CAs and their macromolecular derivatives become markedly less effective as the strength of the magnetic field increases.⁸ A different approach for the relaxivity enhancement of Gd-based probes is necessary for the currently used high-field systems.

Furthermore, efforts should be made in the direction of developing new T_1 agents that have excellent relaxation properties over a broad range of imaging field strengths in order to take advantage of the current 1.5–3 T transition. Multimeric Gd^{III} agents with medium molecular weight (~3–4 kDa) have been suggested as one possible solution as they show improved r_1 values at high fields.^{9–13} In fact, the Solomon–Bloembergen–Morgan (SBM) theory of paramagnetic relaxation predicts that, in the case of Gd^{III} systems, at frequencies above *ca.* 200 MHz (~4.7 T) the relaxivity increases with the inverse of the rotational correlation time τ_R , in contrast to lower frequencies, where it is roughly proportional to τ_R .¹⁴ Therefore, medium-sized polynuclear systems should exhibit improved r_1 values at frequencies above 1.5 T in comparison with the clinical CAs. These systems should be characterized by reorientational correlation times in the range of *ca.* 200 to 600 ps to maximize r_1 above 1.5 T. These estimates are derived from a series of simulations carried out by P. Caravan in order to find the best combination of molecular parameters for achieving the highest r_1 for each desired frequency value.⁹ These parameters are essentially the number of coordinated water molecules q , their exchange rate k_{ex} ($= 1/\tau_M$) and the molecular rotational correlation time.

With the aim to explore in detail the dependence of r_1 at high fields from the rotational dynamics and understand the effect of the different degree of rotational flexibility of the discrete chelates compared to the global tumbling motion of the polynuclear system, we designed and synthesised a homogeneous series of multimeric structures containing from 2 to 8 Gd^{III} complexes. These chelates were linked to a central scaffold through a spacer chosen to minimize the internal rotation. The broad range of molecular weights, *ca.* 600–6000 Da, was covered through the synthesis of polynuclear complexes obtained by adding, as a building block, an identical paramagnetic unit in a controlled fashion such as to reduce as much as possible the variations of the relaxometric parameters and thus allow to clearly highlight the changes in the rotational dynamics. As a building block, we selected the AAZTA platform (AAZTA = 6-amino-6-methylperhydro-1,4-diazepine tetraacetic acid).¹⁵ The $[\text{Gd}(\text{AAZTA})(\text{H}_2\text{O})_2]^-$ chelate exhibits several desirable properties: it possesses two water molecules in the inner coordination sphere of the metal ion with a sufficiently high exchange rate, good thermodynamic stability ($\log K_{\text{GdL}} = 20.24$) and remarkable kinetic inertia under physiological conditions ($t_{1/2}$ at pH = 7.4: 4.3×10^3 h).^{15b} In addition, the complex does not easily form ternary complexes with endogenous anions (carbonate, phosphate, lactate, *etc.*) by displacement of the water ligands and loss of relaxivity.^{15,16}

To the best of our knowledge, the only systematic study on the effect of multimerization on the relaxivity of macrocyclic Gd-chelates was published by Tweedle and co-workers in 1998.¹⁷ They described a series of multimeric derivatives, from binuclear to octanuclear species, based on DOTA-monoamide (DOTAMA, DOTA = 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) and HPDO3A (10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid) chelators linked to central polyamine or polyhydroxyl cores (Gd₄1 and Gd₈2 in Scheme 1). However, due to the difficulty in accessing time domain NMR instrumentation (fast-field cycling relaxometer) in those years, only single frequency measurements were performed which prevented a complete and thorough analysis.

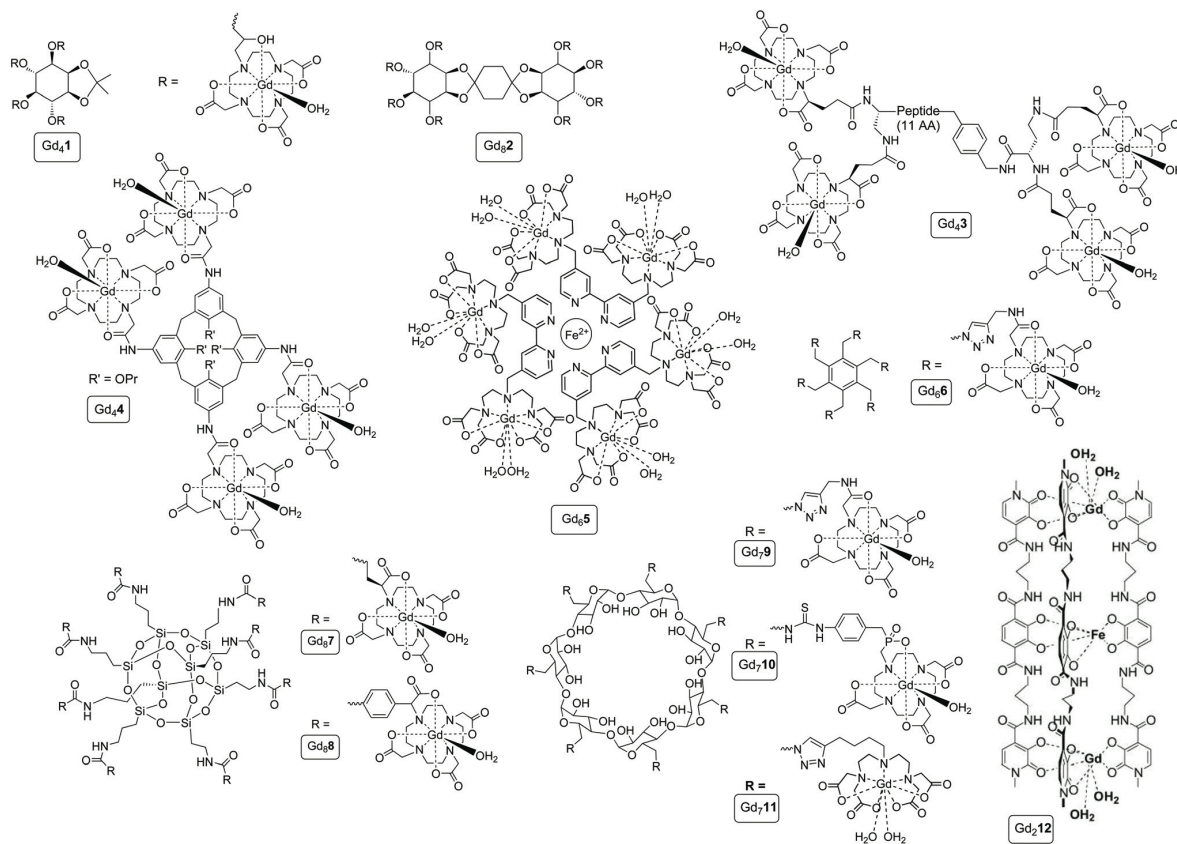
We report herein the synthesis and a frequency and temperature dependent ¹H and ¹⁷O NMR relaxometric study on the homogeneous series of Gd(III) polynuclear complexes. Information on the relaxivity changes with magnetic field strengths (from 2.3×10^{-4} to 9.4 T) and temperature (283–310 K), on the stability over a broad pH range and on values of the key molecular parameters was obtained. The dependence of the internal rotational barriers on the molecular weight was also evaluated by DFT calculations. Finally, T_1 -weighted phantom MR-images at 1, 3 and 7 T were acquired on the clinically used CA ProHance® and on selected polynuclear complexes to evaluate the effective signal enhancement of the new MRI probes at the most common magnetic field values of modern scanners.

Results and discussion

Background: design of the polynuclear complexes

The design of multimeric Gd-based systems requires the judicious choice of three different moieties that constitute the entire structure: (i) the monomeric Gd^{III} complex, (ii) the central scaffold and (iii) the linker connecting the metal complex to the central unit. One of the key objectives is to obtain rigid and compact systems that can achieve high ionic relaxivity (per Gd) combined with a high relaxivity density (r_1 per unit mass).

A large number of Gd-based multimeric systems have been reported since the first example of a tetrameric DTPA derivative, reported by Martin and co-workers for blood pool MRI applications.¹⁸ A remarkable example is represented by the development of the “metallostar” reported by E. Tóth and co-workers (Gd₆5 in Scheme 1).¹⁰ The metallostar is a polynuclear Gd-complex based on the self-assembly of dimeric (GdDTTA)₂-bipyridine complexes around a Fe^{II} complex core. It has a molecular mass of 3744 g mol⁻¹ and an unusually broad relaxivity peak centred at relatively high frequencies. As a result of the two inner-sphere water molecules, with a near-optimal exchange rate, and the restricted flexibility, the metallostar exhibits significant ¹H molar relaxivity at high fields which is confined to a small molecular space (high density of relaxivity). Further studies were carried out by changing the central unit (Scheme 1) and thus using a benzene ring (Gd₆6),¹⁹ a



Scheme 1 Multimeric Gd^{III} complexes (from tetramers to octamers) discussed in the text.

calixarene (Gd₄),²⁰ a β -cyclodextrin (Gd₇–9–11),^{19a,21,22} a silsesquioxane (Gd₇–8),²³ or a low generation PAMAM dendrimer.²⁴ However, often the conjugation of the Gd^{III} chelate to a molecular scaffold involves a fairly flexible connection that favours a high degree of local rotation of the terminal metal complexes, hence lower relaxivity enhancement.

The relevance of a rigid spacer connecting the Gd-chelate to the central core was highlighted by switching from an ethylene to a benzene linker (from Gd₇ to Gd₈), which showed an increase of r_1 of 40%.²³ Also peptides were used for the conjugation of a number of Gd^{III} units with a dual positive result to combine a high efficiency with targeting capabilities (Gd₄3).²⁵

However, it is well known that if the coordinated water exchange is not fast enough, the increase in relaxivity is significantly limited by τ_M . In the case of multimeric systems ($\tau_R < 0.5$ ns) the fast exchange regime occurs for $\tau_M < 100$ ns, approximately. In fact, polynuclear systems based on GdDOTAMA (Gd₄4, Gd₆6, Gd₇9) or GdHPDO3A (Gd₄1, Gd₈2), which have relatively long τ_M values (ca. 0.6–1 μ s), did not show significant relaxivity enhancements.^{17,19,20a,24} GdDOTA-like chelates (e.g. Gd₄3, Gd₇7, Gd₈8) are characterized by having coordinated water exchange rates dependent on the population of the diastereoisomers present in solution, thus

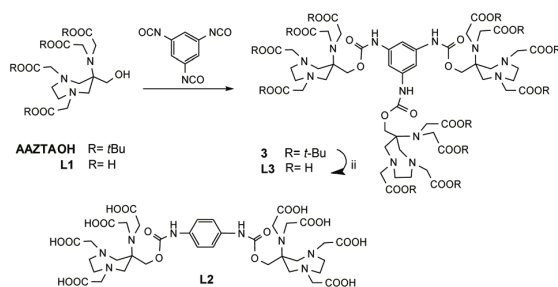
making it more difficult to carry out a systematic study and analyse the data in great detail. Therefore, even these complexes do not represent the most suitable building blocks to optimize the relaxivity of polynuclear systems.^{21,23–25}

Therefore, we considered that a complex such as [Gd(AAZTA)(H₂O)₂][−] with two inner sphere water molecules ($q = 2$) featuring a $\tau_M < 100$ ns, in addition to a good thermodynamic and kinetic stability,^{15,16} could represent a suitable choice for a proof of concept study.

Notably, di- and tetranuclear Gd-AAZTA complexes have also been recently reported that contain a propylene or an aromatic spacer between the two coordination cages,^{26,27} and a polylysine core in the case of the tetramer.²⁸

Synthesis and characterization of AAZTA multimers

A general synthetic procedure towards mono- and binuclear AAZTA bifunctional prochelators containing isothiocyanate groups for coupling to NH₂-containing (bio)molecules has been recently set up in our laboratories.²⁹ These bifunctional agents were used for the preparation, through a synthetic modular approach, of different generations (G0, G1 and G2) of ethylenediamine-cored PAMAM dendrimers.^{29,30} A similar methodology was herein employed for the synthesis of the



Scheme 2 i: CH_2Cl_2 , rt, 3 h; ii: TFA/ CH_2Cl_2 (1 : 1 v : v), rt, 18 h.

polynuclear Gd^{III} complexes covering a broad range of molecular weights (1200–6000 Da). Aromatic isocyanates were selected to react with the hydroxyl group of AAZTA-OH (Scheme 2) under mild conditions such as neutral pH and room temperature to form stable carbamates.³¹ In particular, aromatic isocyanates were readily synthesized from carboxylic acids by transformation first into acyl chlorides and then into acyl azides followed by Curtius rearrangement, as reported for 1,3,5-phenylene trisocyanate.³²

Thus, while the synthesis of the dimeric ligand **L2** (Scheme 2) was reported elsewhere,²⁷ the AAZTA trimer was synthesised by reaction of 1,3,5-phenylene trisocyanate with AAZTA-OH followed by column chromatography purification and deprotection of the *tert*-butyl esters with a 1 : 1 mixture of CH_2Cl_2 and trifluoroacetic acid (**L3**, Scheme 2).

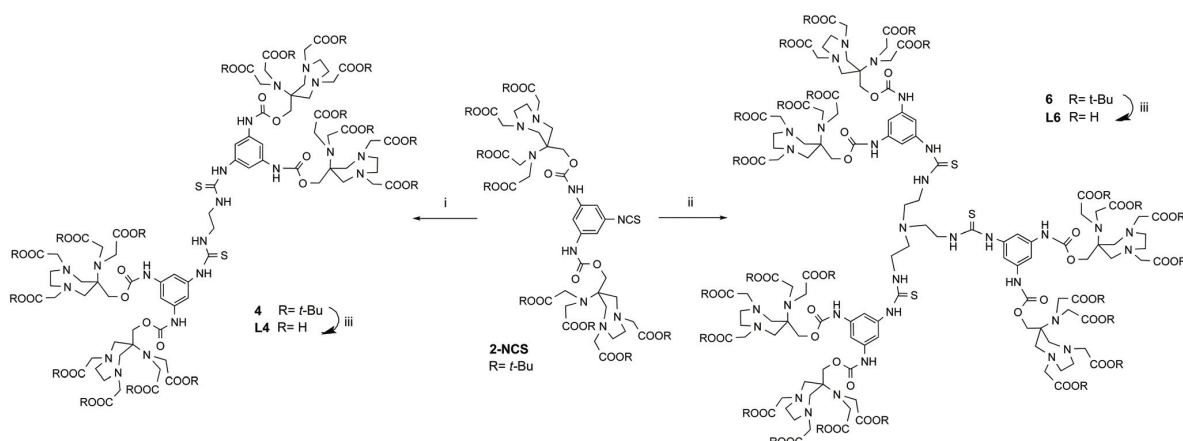
On the other hand, to synthesise higher oligomers, the isothiocyanate functionalised AAZTA dimer (2-NCS, Scheme 3) was used.²⁹ This dimeric ligand was reacted with aliphatic primary polyamines such as ethylenediamine or tris-(2-aminoethyl)amine to form in high yield the tetrameric and hexameric protected ligands after purification by column chromatography. The synthesis of the octamer starting from the G0 PAMAM dendrimer was reported earlier.³⁰ For all multimers,

the deprotected ligands were obtained after reaction with a 1 : 1 mixture of CH_2Cl_2 and trifluoroacetic acid. The final compounds were also purified and desalted by passing through a Sephadex G25 gel filtration column and characterized by reverse phase HPLC-MS. It is worth noting that, for our purposes, aliphatic polyamines represent a flexible central scaffold for the synthesis of AAZTA multimers which guarantees greater solubility of both protected and deprotected ligands in organic or aqueous media, respectively. Moreover, the use of this modular synthetic approach has several advantages: the attainment of each intermediate in high yield with simple purification conditions and the excellent versatility due to the possibility of varying the linker or the scaffold and thus obtaining the most suitable product for different needs.

Relaxometric studies

The complete ^1H and ^{17}O NMR relaxometric characterization of the complexes in aqueous solution involves the measurement of the solvent nuclei relaxation rate as a function of pH, temperature and applied magnetic field. From all these data, much information is obtained on the stability of the complexes in aqueous solution, on the hydration state of the metal ion, on the exchange regime of the coordinated water molecules and on an accurate assessment of the molecular parameters that determine their effectiveness as relaxation agents.

pH dependency. The variation of the proton relaxivity with pH provides useful insights into the solution behaviour of paramagnetic complexes as it may highlight the occurrence of several processes: ternary complex formation with dissolved $\text{HCO}_3^-/\text{CO}_3^{2-}$, hydrolysis at high pH, hydration equilibria, protonation steps, structural variations, and stepwise decomplexation at low pH. The Gd-based commercial CAs ($q = 1$) show a behaviour that depends on the macrocyclic or acyclic nature of the ligand. Gd complexes of macrocyclic ligands combine high thermodynamic and kinetic stabilities and therefore their r_1 values are constant over a broad range of values (pH ~ 1–12).



Scheme 3 Reagents and conditions: i: ethylenediamine, CH_2Cl_2 , TEA, rt, 18 h; ii: tris-(2-aminoethyl)amine, CH_2Cl_2 , TEA, rt, 18 h; iii: TFA/ CH_2Cl_2 (1 : 1 v : v), rt, 18 h.

For the complexes with acyclic ligands r_1 assumes a constant value from 12 to *ca.* 4, and then abruptly increases signalling the progressive increase of the hydration state as the pH becomes more acidic, which eventually leads to the formation of the aqua ion. AAZTA has a pseudo-macrocyclic structure and the corresponding Gd-complex is characterized by a good thermodynamic stability and a kinetic inertness higher than GdDTPA.^{15,16} In fact, in spite of its higher hydration number ($q = 2$), the pH dependency of r_1 reproduces quite well that of GdDTPA. This implies that the coordinated water molecules are not displaced even at high pH, in the presence of a relatively high concentration of carbonate anions. The multimeric complexes Gd₂L2–Gd₄L4 show a trend that is completely similar to that of the parent complex (Fig. S1†). This indicates that the integrity of the coordination cage is maintained in the multimeric systems and that potential protonation steps involving the central scaffold do not affect the molecular parameters that govern the relaxivity (*e.g.* rotational dynamics, water exchange rate, second sphere of hydration).

Water exchange. In order to accurately investigate and analyse the dependence of r_1 on rotational dynamics, it is necessary to first verify that τ_M is short enough not to limit relaxivity. In this regard, it is worth remembering that the magnetic interaction between the protons of the coordinated water molecules and the Gd^{III} ion involves both a short-range interaction (inner sphere relaxivity, r_1^{IS}) and a long-range interaction (outer sphere relaxivity, r_1^{OS}), which are additive:³³

$$r_1 = r_1^{IS} + r_1^{OS} \quad (1)$$

All r_1 enhancement strategies are aimed at optimizing the IS contribution (r_1^{IS}), which is given by the following equation:

$$r_1^{IS} = \frac{p_M}{T_{1M} + \tau_M} \quad (2)$$

where p_M is the molar fraction of the bound water molecules ($p_M = [\text{Gd}^{\text{III}}]q/55.6$) and T_{1M} is the longitudinal nuclear magnetic relaxation time of the IS water protons. For Gd-based multimers with $q = 2$ and molecular masses covering the range from about 1200 to 6000 Da, we can safely assume that the relaxivity values at the magnetic fields of clinical interest are included in the wide range of 5–35 mM⁻¹ s⁻¹ (at 298 K). This implies, using eqn (2), that T_{1M} assumes the values from about 1 to 10 μ s and therefore that the fast exchange condition occurs for $\tau_M < 100$ ns. [Gd(AAZTA)(H₂O)₂]⁻ has a reported τ_M value of ~90 ns at 298 K which is sensibly lower at 310 K. However, we have recently found that a derivative bearing in the 6-position a 4-nitrophenylcarbamate moiety (Gd(AAZTA-Ph-NO₂)) has an exchange rate slower by a factor of 2.8.³⁴ This dependence of τ_M on the nature of the substituent does not allow us to simply assume that this parameter is nearly identical to that of the parent compound in all polynuclear complexes. Therefore, we extracted experimentally the water exchange rate and its activation parameters by measuring, for selected complexes (Gd₂L2, Gd₃L3 and Gd₆L6), the temperature dependence of ¹⁷O NMR transverse relaxation

rates of bulk water induced by the presence of the paramagnetic solution. The experimental data were collected on a high-resolution NMR spectrometer, operating at 11.7 T, on aqueous solutions of the complexes (~5 mM) at neutral pH. The data were then analysed according to the well-established set of Swift–Connick equations.³⁵ The increase of the reduced transverse relaxation rate ($1/T_{2r} = R_{2r}$) with decreasing temperature over the temperature range of 280–350 K clearly indicates the occurrence of a fast rate of water exchange, significantly faster than that of [Gd(AAZTA)(H₂O)₂]⁻ (Fig. 1).

In the fitting procedure, some parameters were fixed to standard values. The activation energy of τ_M (E_V) was set to 1.0 kJ mol⁻¹ and the distance $r_{\text{Gd-O}}$ was fixed at 2.50 Å. The scalar coupling constant, A_0/h , was let to vary around values quite typical of Gd^{III} complexes ($3\text{--}3.8 \times 10^6$ rad s⁻¹). The results of the best-fit procedure (Table 1) show for all the systems an exchange rate of the inner sphere water molecules significantly higher than that of GdAAZTA. This result, opposite to that observed for Gd(AAZTA-Ph-NO₂), confirms the sensitivity of this parameter to the electronic and/or steric effects of the substituent. Further support to the evidence that the polynuclear complexes are in a fast exchange regime derives from the temperature dependence of the proton relaxivity at the fixed frequency of 20 MHz (Fig. S2†). In all cases, r_1 increases as the temperature decreases (275–330 K), demonstrating that τ_M does not represent a limiting factor.

NMRD. The knowledge of the k_{ex} value and of the occurrence of the rapid exchange condition makes it possible to

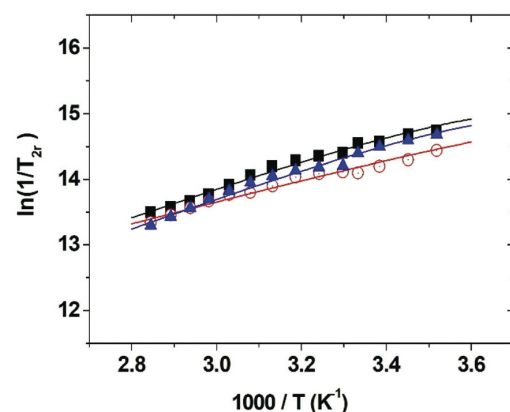


Fig. 1 Reduced transverse ¹⁷O relaxation rates measured at 11.7 T (pH 7) for Gd₂L2 (blue triangles), Gd₃L3 (black squares) and Gd₆L6 (empty red circles).

Table 1 Selected best fit parameters obtained from the analysis of the ¹⁷O NMR profiles of Gd₂L2, Gd₃L3, and Gd₆L6

Parameter	Gd ₂ L2	Gd ₃ L3	Gd ₆ L6
²⁹⁸ τ_M (ns)	40 ± 1	40 ± 2	27 ± 3
ΔH_M (kJ mol ⁻¹)	16.6 ± 0.3	15.9 ± 0.1	11.1 ± 0.3
A_0/h (10 ⁶ rad s ⁻¹)	-3.1 ± 0.2	-3.3 ± 0.1	-3.2 ± 0.1

obtain the other parameters that govern the relaxivity of the polynuclear complexes through the measurement and analysis of the r_1 dependence on the applied magnetic field strength. Then, we recorded the ^1H Nuclear Magnetic Relaxation Dispersion ($1/T_1$ NMRD) profiles of $\text{Gd}_{1-6}\text{L1-L6}$ at pH 7.0 and at 298 and 310 K in the proton Larmor frequency range of 9.97×10^{-3} to 80 MHz (Fig. 2 and Fig. S3–S6†).

The lowest NMRD profile in Fig. 2 corresponds to the monomer and has a standard shape of small $\text{Gd}(\text{III})$ complexes, with a dispersion between 3 and 10 MHz that separates two regions with almost constant r_1 . As expected, passing from the mononuclear to the binuclear complex we observe an increase in r_1 over the entire frequency range, particularly evident at the clinical fields (>20 MHz). At 1.5 T, where relaxivity is mainly controlled by τ_R , the gain in r_1 is around 90%. In the case of the trimer, we clearly notice a further increase of r_1 , albeit slightly less pronounced, which corresponds to *ca.* +30% at 1.5 T relative to the dimer. In addition, the formation of the characteristic broad hump at high fields typical of slowly tumbling systems begins to appear. These features, in an initial phase of formation in the case of the trimer, become quite evident for the tetranuclear complex. Relaxivity does not significantly increase in the 0–10 MHz region, while at higher frequencies a pronounced, albeit large, peak of r_1 takes shape with a maximum around 1–1.5 T. The r_1 enhancement with respect to the trimer corresponds to +40% at 1.5 T. Furthermore, at frequencies greater than 80–100 MHz we see a steep and marked decrease in relaxivity. Indeed, for frequencies >300 MHz we do not see a significant difference in r_1 between the dimer and the tetramer, despite the significant increase in molecular mass. In the case of $\text{Gd}_6\text{L6}$, the ionic relaxivity (for Gd) increases by a further 23% at 1.5 T and reaches a value of about $30 \text{ mM}^{-1} \text{ s}^{-1}$. However, the NMRD profile has a very similar shape to that of $\text{Gd}_4\text{L4}$ and only slightly greater amplitude. It is clear that the increase in molecular size is not reflected in a corresponding increase in r_1 . Overall, passing from the monomer to the hexanuclear

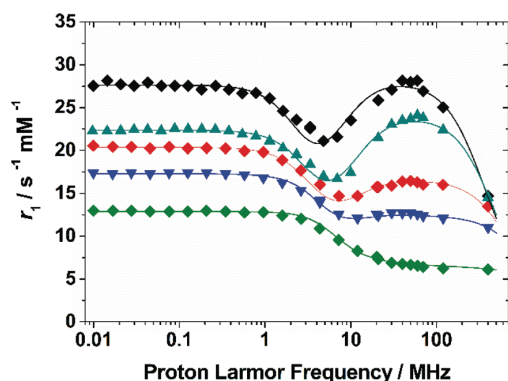


Fig. 2 NMRD profiles of $\text{Gd}_{1-6}\text{L1-L6}$ complexes at 298 K. GdL1 (green diamonds); $\text{Gd}_2\text{L2}$ (blue triangles); $\text{Gd}_3\text{L3}$ (red diamonds); $\text{Gd}_4\text{L4}$ (grey triangles); $\text{Gd}_6\text{L6}$ (black diamonds). The relaxivity, r_1 , is per Gd ion.

complex the relaxivity per Gd increases considerably, with an enhancement of +370% at 1.5 T and 298 K. On the other hand, at frequencies close to or above 300 MHz there is no appreciable difference in efficacy as a T_1 contrast agent between the dimer and hexamer. The case of $\text{Gd}_8\text{L8}$ is different, as the NMRD profile reproduces closely that of $\text{Gd}_4\text{L4}$, despite its much larger molecular size and a molecular mass more than twice higher (Fig. 3).³⁰ In this case, more than the global molecular tumbling is the local rotational flexibility the dominant factor that controls the relaxivity of the system. In the absence of stereochemical rigidity, there is no advantage in designing polynuclear systems with a number of metal centres greater than about 4–5. For all the investigated complexes, the value of r_1 decreases at a temperature of 310 K over the entire frequency range (Fig. S3–S6†), confirming that the fast exchange condition is met, as indicated by the ^{17}O NMR data.

A least-squares fit of the ^1H NMRD profiles was carried out to estimate the parameters governing the relaxivity of all complexes on the basis of the Solomon–Bloembergen–Morgan equations¹⁴ (for the inner sphere mechanism, IS) and of Freed's equation³⁶ (outer sphere model, OS). The IS contribution to relaxivity ($r_{1,\text{IS}}$) in $\text{Gd}(\text{III})$ complexes depends on the relaxation rate of inner sphere protons ($T_{1\text{M}}^{\text{H}}$), the mean residence time of a water molecule bound to $\text{Gd}(\text{III})$ ($\tau_{\text{M}} = 1/k_{\text{ex}}$) and the number of IS water molecules q :

$$r_{1,\text{IS}} = \frac{1}{1000} \times \frac{q}{55.55} \times \frac{1}{T_{1\text{M}} + \tau_{\text{M}}} \quad (3)$$

The relaxation rate of IS protons in $\text{Gd}(\text{III})$ chelates arises from the time modulation of the dipole–dipole magnetic coupling:

$$\left(\frac{1}{T_{1\text{M}}}\right)^{\text{DD}} = \frac{2}{15} \left(\frac{\mu_0}{4\pi}\right)^2 \gamma_1^2 g^2 \mu_{\text{B}}^2 \frac{S(S+1)}{r_{\text{GdH}}^6} \times \left(\frac{3\tau_{\text{d1}}}{1 + \omega_1^2 \tau_{\text{c1}}^2} + \frac{7\tau_{\text{d2}}}{1 + 4\omega_s^2 \tau_{\text{c2}}^2}\right) \quad (4)$$

In eqn (4), g is the electron g factor, r_{GdH} is the distance between the electron and nuclear spins, μ_{B} is the Bohr magneton, γ_1 is the proton gyromagnetic ratio, S is the total spin ($7/2$

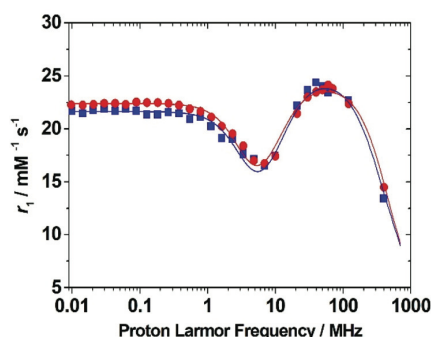


Fig. 3 ^1H NMRD profile of $\text{Gd}_8\text{L8}$ (blue squares) at 298 K compared to that of $\text{Gd}_4\text{L4}$ (red circles). The relaxivity, r_1 , is per Gd ion.

for Gd^{3+}), ω_1 is the proton resonance frequency and ω_S is the Larmor frequency of the Gd^{III} electron spin. The correlation time τ_{ci} is given by eqn (5), where τ_{R} is the rotational correlation time, and T_{ie} are the longitudinal ($i = 1$) and transverse ($i = 2$) relaxation times of the electron spin, described in terms of the parameters Δ^2 (the mean square transient ZFS energy) and its correlation time τ_{v} .

$$\frac{1}{\tau_{\text{ci}}} = \frac{1}{\tau_{\text{R}}} + \frac{1}{\tau_{\text{M}}} + \frac{1}{T_{\text{ie}}}, \quad \text{with } i = 1, 2 \quad (5)$$

The OS contribution depends on $^{298}D_{\text{GdH}}$, the relative diffusion coefficient, on a_{GdH} , the distance of closest approach of proton nuclei of bulk water molecules to the gadolinium ion and on T_{ie} .

The analysis of the data was performed by fixing some parameters to standard values: (1) the distance between the proton nuclei of coordinated water molecules and the Gd^{3+} ion was fixed to 3.0 Å; (2) the a_{GdH} value was fixed at 4.0 Å; (3) the parameter $^{298}D_{\text{GdH}}$ was fixed to the typical value of $22.4 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$; (4) for the mean residence lifetime of the IS water molecules the values estimated from the ^{17}O NMR data were used; (5) finally, the number of water molecules coordinated to the Gd^{3+} ion was fixed to $q = 2$. Furthermore, for the electron relaxation parameters those derived from the simultaneous fit of the NMRD profiles and ^{17}O NMR data of the monomeric complex were used as starting values. Then, the adjustable parameters were Δ^2 , τ_{v} and τ_{R} . The best-fit procedure reproduces very well the experimental data of GdL1 and $\text{Gd}_2\text{L2}$ with the parameters listed in Table 2.

Attempts to fit the data of the other multimeric complexes using the same approach did not provide results of comparable or acceptable quality. On the grounds of the experience gained in recent years,⁸ this result is somewhat predictable and can be attributed to the progressive decrease, as the size of the polynuclear compound increases, of the motional coupling between the global tumbling of the multimers and the local rotation of the Gd-chelate around the linker. In fact, we have just discussed an attenuation of the increase of r_1 in the NMRD profiles with the increase of the molecular size, in particular at high fields, which we can explain with the growing importance of local rotational motions involving the metal complexes superimposed on the overall molecular tumbling. Recently, this feature was discussed in detail in the case of dendrimeric derivatives.³⁰ Therefore, the NMRD

profiles of the other multimers were analysed using the Lipari–Szabo model for the description of the rotational dynamics.³⁷ This model allows separate evaluation of the relatively fast local rotation of the monomeric complexes with respect to the global motion of the polynuclear system. The local motion is described by a local rotational time τ_{RL} , while the overall molecular tumbling by the global rotational correlation time τ_{RG} . The degree of correlation between the two types of motions is described by the parameter S^2 , whose value varies from zero (completely independent motions) to one (totally correlated motions). With this procedure, the fit of the data improves markedly. From Table 2 we see that for the smallest members of the series (GdL1 and $\text{Gd}_2\text{L2}$), the parameter S^2 is equal to 1 and the motion is described by only one correlation time, which indicates fully coupled motions. For the $\text{Gd}_3\text{L3}$ – $\text{Gd}_8\text{L8}$ systems, the order parameter tends to progressively decrease and the difference between τ_{RL} and τ_{RG} widens, emphasizing the increasing degree of anisotropy of motion. A confirmation of the reliability of the results can be seen from the graph in which the values of τ_{RG} and τ_{RL} are plotted as a function of the molecular mass of the investigated complexes (Fig. 4). As can be clearly seen, while τ_{RG} grows linearly with the increase in the size of polynuclear systems (slower molecular tumbling), τ_{RL} instead tends to level out and deviate significantly from linearity. Evidently, this behaviour is associated with the progressive loss of correlation between the motion of the monomer units and the global rotation of the multimeric system.

At 1.5 T the relaxivity is largely dominated by the value of τ_{R} , as it can be seen from its almost linear dependence on the molecular mass and therefore it reaches the maximum value for the hexanuclear complex (Fig. 4). In the case of the octamer, the presence of a flexible scaffold, such as PAMAM G0 dendrimers, causes a poor coupling between local and global motion (low value of S^2) that strongly influences the relaxivity.

On the other hand, the behaviour at 3 T and 9 T is very different and we can observe a progressive and marked attenuation of the increase in relaxivity with the increase in the molecular size. Broadly speaking, these results fully confirm previous theoretical predictions derived from detailed simulations.⁹ For magnetic field strengths up to approximately 1.5 T, r_1 is limited by rotation and water exchange and thus macro-

Table 2 Selected best fit parameters obtained from the analysis of the $1/T_1$ NMRD profiles of the $\text{Gd}_{1-8}\text{L1-L8}$ complexes^a

Parameter	GdL1	Gd ₂ L2	Gd ₃ L3	Gd ₄ L4	Gd ₆ L6	Gd ₈ L8 ^b
MW	531	1293	1901	2761	4197	5918
Δ^2 (s^{-2})	3.1 ± 0.1	3.1 ± 0.2	2.6 ± 0.2	2.4 ± 0.1	1.4 ± 0.2	2.6
τ_{v} (ps)	29 ± 1	29 ± 2	30 ± 2	32 ± 2	42 ± 3	31
τ_{RG} (ps)	63 ± 2	142 ± 3	234 ± 5	390 ± 5	499 ± 7	547
τ_{RL} (ps)	—	—	190 ± 8	240 ± 11	270 ± 10	290
S^2	1	1	0.72 ± 0.02	0.69 ± 0.01	0.58 ± 0.02	0.37

^a Obtained from the analysis of the NMRD profile (298 K) by considering two inner sphere water molecules ($q = 2$) whose protons are at an average distance of 3.0 Å from Gd^{3+} . ^b From ref. 28.

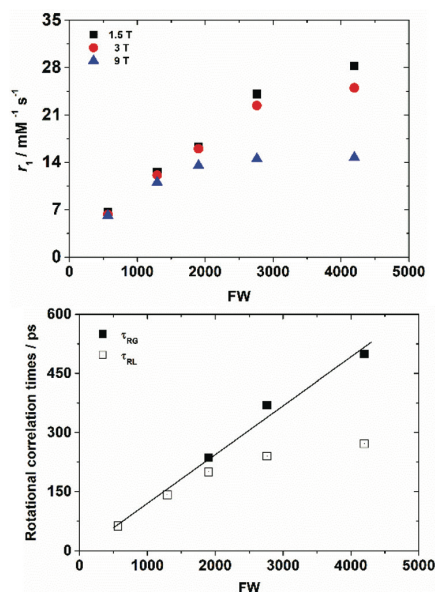


Fig. 4 Correlation between molecular weight (MW) of Gd₁₋₆L1–L6 complexes and r_1 (per Gd ion) at 1.5, 3 and 9 T (top) or τ_{RG} and τ_{RL} (bottom) ($T = 298$ K).

molecular systems are the most effective. Near 3 T and above, the problem of local flexibility or anisotropic rotation becomes increasingly important so that rigid and medium-sized polynuclear complexes (3–5 units) are endowed with the greatest relaxivity enhancement. At ultra-high fields (≥ 7 T), small and compact complexes, either monomers or dimers, represent the most effective MRI magnetic probes.

About fifteen years ago, Éva Tóth and co-workers introduced an alternative way of evaluating the effectiveness of a paramagnetic system as a relaxation agent, based on the concept of density of relaxivity. They defined the “effective” or “mass relaxivity” as the enhancement of the relaxation rate by a unit mass (g L^{-1}) of the contrast agent.¹⁰ High density of relaxivity is necessary for molecular imaging applications such as cell imaging, in which the probe is required to produce a sufficient relaxation effect in a limited mass. In Table 3 is reported a comparison of molecular weights, r_1 values at 60 MHz and 310 K (unless otherwise stated), and mass relaxivities of the multimeric agents reported herein and those shown in Scheme 1. The greatest mass relaxivity is associated with the “metallostar” (Gd₆5 in Scheme 1),¹⁰ with an outstanding value of $42.9 \text{ g L}^{-1} \text{ s}^{-1}$. Aside from that model system, Gd₄L4 and Gd₆L6 show the best results with values of 29.1 and $34.0 \text{ g L}^{-1} \text{ s}^{-1}$, respectively. All the other multimeric systems listed are far behind these values, mainly because of a lower hydration state ($q = 1$), relatively slow water exchange rate and large degree of anisotropic rotation.

A notable example that should be highlighted is Gd₂L2 (Scheme 1), which represents a Gd(III)-containing supramolecular self-assembly obtained using Fe(III) as a template.³⁸ The

Table 3 Comparison of selected parameters for the Gd_xL1–L8 and for the Gd_x1–11 multimeric complexes ($x = 1-8$)

Multimer	MW (g mol^{-1})	r_1 (per Gd), $\text{mM}^{-1} \text{ s}^{-1}$ (60 MHz, 37 °C)	Mass relaxivity ($\text{g L}^{-1} \text{ s}^{-1}$) (60 MHz, 37 °C)
GdL1	567	5.0	8.8
Gd ₂ L2	1309	7.1	10.8
Gd ₃ L3	1902	12.7	20.0
Gd ₄ L4	2761	20.1	29.1
Gd ₆ L6	4197	23.8	34.0
Gd ₈ L8	5918	18.3	24.7
Gd ₄ 1	2501	9.8 (20 MHz, 40 °C) ¹⁷	—
Gd ₈ 2	4998	13.0 (20 MHz, 40 °C) ¹⁷	—
Gd ₄ 3	4327	10.1 ²⁵	9.3
Gd ₄ 4	2889	8.2 (20 MHz, 25 °C) ²⁰	—
Gd ₆ 5	3744	26.8 ¹⁰	42.9
Gd ₆ 6	3983	11.0 ¹⁹	16.6
Gd ₈ 7	5919	9.4 ²³	12.7
Gd ₈ 8	6303	13.8 ²³	17.5
Gd ₇ 9	5480	12.2 ^{19a}	15.6
Gd ₇ 10	6327	15.9 ²¹	17.6
Gd ₇ 11	5292	6.2 (400 MHz, 37 °C) ²²	—

ligand contains a central TAM moiety (2,3-dihydroxyterephthalamide), selective for coordination of Fe, and two terminal HOPO (1-methyl-2,3-dihydropyridinone) binding groups that are highly selective for Gd coordination. Three of these ligands wrap around two Gd³⁺ and one Fe³⁺ to give a compact structure of high relaxivity: $21 \text{ mM}^{-1} \text{ s}^{-1}$ at 3 T and 298 K (per Gd). In addition, this complex presents a remarkable mass relaxivity of $17.9 \text{ g L}^{-1} \text{ s}^{-1}$, at 60 MHz and 298 K. This is probably, to the best of our knowledge, the greatest value for a binuclear Gd(III) complex, thanks to its very compact structure that minimizes local rotational motions. It is worth noting that the HOPO-based class of Gd(III) complexes share many favourable characteristics with the AAZTA-based complexes, apart from kinetic inertia, and therefore are very suitable for developing analogous highly effective polynuclear systems.³⁹

Computational models

To help the atomistic interpretation of the relaxivity trends reported in Fig. 2 and Table 2, the rotation energy profiles were computed at the Density Functional Theory (DFT) level for Gd₂L2, Gd₃L3 and Gd₄L4. The scope was to investigate the dependence of the internal rotational barriers on the molecular weight.

First, the molecular structures of the three Gd-multimers were optimized (the final geometries are illustrated in the ESI†). The central phenyl connecting unit ensures that both Gd₂L2 and Gd₃L3 assume a planar conformation with the Gd complexes well separated; on the other hand, two minima were found for the tetranuclear structure, Gd₄L4, corresponding to different dihedral angles of the flexible central chain, and leading to “closed” and “open” conformations, as shown in the ESI.† Only the latter was considered for the rotational analysis described below, because in the closed structure the molecular fragments could not rotate freely around the central chain.

Some preliminary energy scans were performed to identify the dihedral angles with the lowest rotational barriers in the dimeric and tetrameric systems (considering that the possible internal rotations in Gd_3L3 are the same as in the dimer). During these scans, all the other internal coordinates were kept rigid to reduce the computational cost. The results are shown in Fig. 5: in the dimer two rotations are expected to have a barrier lower than 20 kJ mol^{-1} (Fig. 4a), and they will be studied in detail through relaxed scans. The same dihedral angles will be also scanned in the trimer and tetramer.

In addition, some other rotations are possible in Gd_4L4 around the bonds that connect the two dimer units, and one of them appears to have a low barrier too, corresponding to the motion of a whole dimeric unit around the central C–N bond (Fig. 4b).

Then, relaxed energy scans were performed for Gd_2L2 , Gd_3L3 and Gd_4L4 around the chosen dihedral angles, with the results illustrated in Fig. 6. The energy profiles for the linker and the Gd-complex rotations (red and black lines in Fig. 4, respectively) are similar in all the studied systems, leading to rotational barriers not much higher than 20 kJ mol^{-1} . Analogously, the additional rotation in Gd_4L4 , where one dimeric unit rotates with respect to the other, exhibits a barrier of similar height.

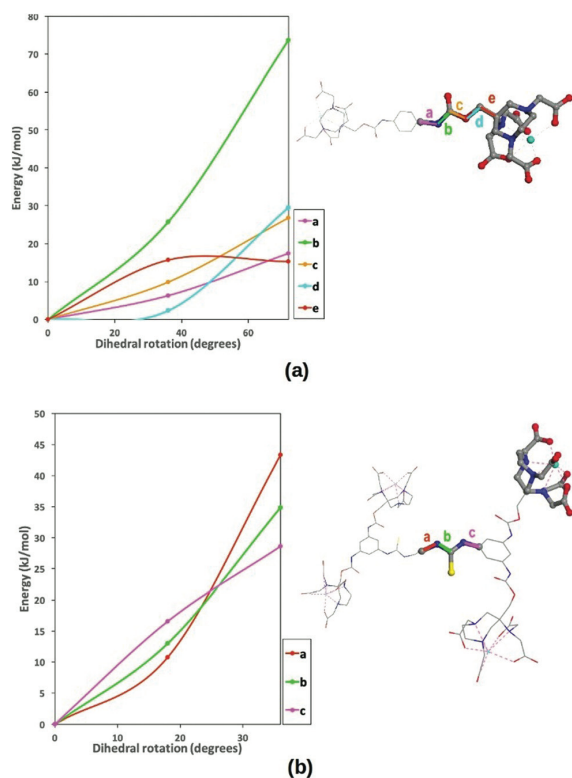


Fig. 5 Energy profiles for rigid intramolecular rotations around several bonds; (a) rotations around dimer bonds (found in all the multimers); (b) rotations around the connecting bonds in the tetramer.

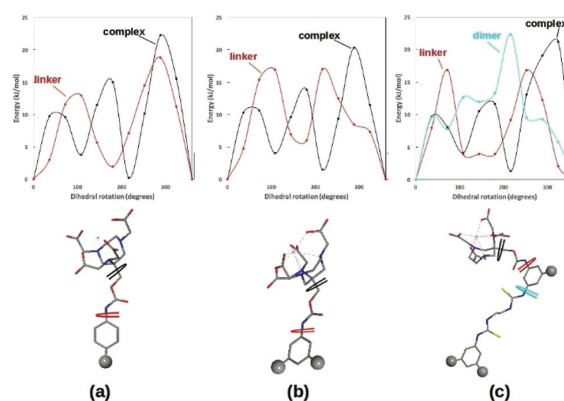


Fig. 6 Energy profiles for optimized (relaxed) rotations around the indicated bonds for Gd_2L2 (a), Gd_3L3 (b) and Gd_4L4 (c).

The calculations show that the internal rotations depend only slightly on the molecular weight, at least in the three examined systems, which exhibit similar energy profiles and barriers. This is in good agreement with the data reported in Table 2: the internal rotational correlation times (τ_{RI}) for Gd_3L3 and Gd_4L4 are very similar, while for Gd_2L2 we can only say it is nearly identical to τ_{RG} . The trend of S^2 also is coherent with the picture of internal rotation barriers which do not vary with the molecular weight, whereas the global tumbling time increases with the size of the complex, leading to the observed slow reduction of the order parameter.

MRI phantom study

To verify that the relaxation enhancement of Gd_2L2 , Gd_4L4 and Gd_6L6 translates into image contrast, phantom-imaging experiments were carried out at three different magnetic field strengths (1, 3 and 7 T).

In Fig. 7 only the images at 3 and 7 T are reported to show that even at very high magnetic field these multimeric systems can offer a superior contrast with respect to the clinical MRI contrast agent ProHance (the images at 1 T are reported in

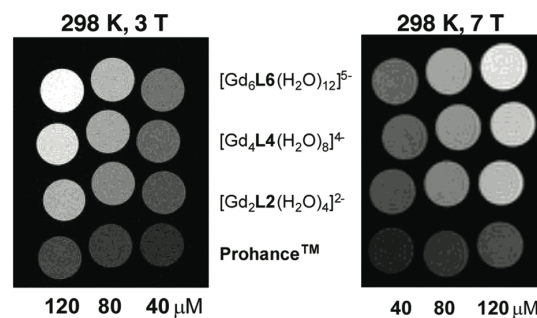


Fig. 7 T_1 weighted Multi Slice Spin Echo phantom images (TR/TE/NEX: 250/8/16) recorded at 3 T and 7 T for solutions of multimers: $[Gd_2L2(H_2O)_4]^{2-}$, $[Gd_4L4(H_2O)_8]^{4-}$ and $[Gd_6L6(H_2O)_{12}]^{5-}$ at 40, 80 and 120 μM concentrations compared to ProHance™ as a control.

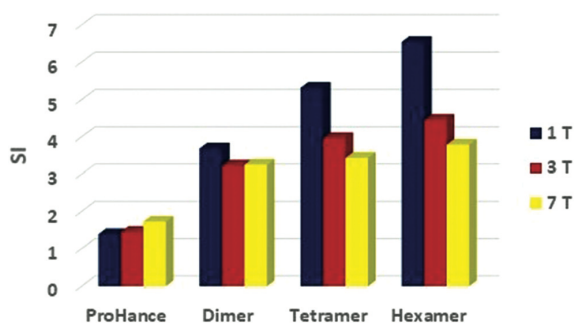


Fig. 8 Signal enhancement (298 K, 120 mM) of Gd₃L2, Gd₄L4 and Gd₆L6 compared to the commercial MRI agent ProHance at 1, 3 and 7 T.

Fig. S19†). In particular, solutions of these polynuclear complexes at three different concentrations (40, 80 and 120 mM) were imaged together with solutions of ProHance (Fig. 7). The experiment clearly showed that Gd₆L6, at concentrations equivalent to those of the other complexes, gives rise to a greater contrast even at 7 T (Fig. 8).

Experimental section

All chemicals were purchased from Sigma-Aldrich Co. and were used without purification unless otherwise stated. NMR spectra were recorded on a JEOL Eclipse Plus 400 (operating at 9.4 Tesla; ¹H at 399.968, ¹³C at 100.572 MHz) spectrometer. Chemical shifts are reported relative to TMS and were referenced using the residual proton solvent resonances. ESI mass spectra were recorded on a Waters SQD 3100. Analytical HPLC-MS was carried out on a Waters modular system equipped with a Waters 1525 binary pump, Waters 2487 UV/Vis and Waters SQD 3100 (ESCI ionization mode) detectors. Size exclusion HPLC was carried out on Sephadex G25 columns on an Amersham Akta Purifier chromatographic system equipped with two pumps, a UV-vis triple wavelength detector (set at 215, 230 and 254 nm), and on-line pH, conductivity, and temperature monitors. AAZTA-OH,³¹ 2-NCS,²⁹ L2²⁷ and L8³⁰ were prepared following reported procedures.

Synthesis of L3

A solution of AAZTA-OH (0.6 g, 1 mmol) in dry CH₂Cl₂ (10 mL) was added dropwise in 5 min to a solution of 1,3,5-phenyltriosocyanate (67 mg, 0.33 mmol) in toluene (5 mL). The reaction mixture was stirred at room temperature overnight and then the solvent was evaporated *in vacuo*. The pure protected trimer 3 was obtained after silica gel column chromatography purification (eluent petroleum ether/ethyl acetate 75/25; yield 50%). ¹H-NMR (CDCl₃) 400 MHz δ = 7.27 (s, 3H, CH), 7.07 (bs, 2H, NH), 4.20 (s, 6H, CH₂OCO), 3.71 (s, 12H, CH₂CO), 3.25 (s, 12H, CH₂CO), 3.09 (d, 6H, *J* = 14.3 Hz, CH₂C), 2.75 (d, 6H, *J* = 14.3 Hz, CH₂C), 2.70 (m, 12, CH₂CH₂), 1.43 (s, 54H, CH₃), 1.41 (s,

54H, CH₃). ¹³C-NMR (CDCl₃) 100 MHz δ = 172.9–170.9 (CO), 153.1 (NHCOO), 139.4 (C), 102.8 (CH), 80.9, 80.7 (CCH₃), 67.9 (CH₂OCO), 63.3 (C), 62.3, 62.2 (CH₂CO), 59.0, 51.5 (CH₂cyclo), 28.3, 28.2 (CH₃). ESI-MS (*m/z*): 2006.55 (M + H⁺). Calc for C₉₉H₁₆₉N₁₂O₃₀: 2006.21.

General procedure for tetra-, and hexameric ligands

A solution of 2-NCS (1.2 eq. for each amino group) in dry dichloromethane (2 mL) was gradually added to a solution of the appropriate poly primary amine (0.05–0.1 mmol) and TEA (2 eq. for each amino group) in dry dichloromethane (2 mL), cooled at *T* ≤ 10 °C with an ice-bath and kept under a N₂ atmosphere. The reaction mixture was stirred overnight at RT and then washed with water (3 × 5 mL), HCl 0.1 M (2 × 5 mL) and saturated NaHCO₃ (2 × 5 mL). The organic phase was dried over Na₂SO₄, filtered and evaporated to yield the crude product that, in case of 4 and 6, was purified on a silica gel chromatography column.

General procedure for *t*-butyl ester deprotection

Trifluoroacetic acid (1 mL) was gradually added to a solution of polyaminocarboxylate-*t*-butyl esters (2–8, 0.2 mmol) in CH₂Cl₂ (1 mL) and the mixture was stirred at room temperature overnight. The solution was then evaporated *in vacuo* and the product was recovered with excess diethyl ether, isolated by centrifugation, washed thoroughly with diethyl ether and dried *in vacuo* obtaining a pure desired product as a trifluoroacetate salt. Final products were purified from salts and low molecular weight impurities by gel filtration on a Sephadex G25 column (30 cm × 1.6 cm) (Pharmacia Biotech, Uppsala, Sweden) using H₂O as the eluent. Analytical HPLC-MS runs on final ligands (20 μ L of a 1.0 mg mL⁻¹ solution in H₂O) using a Waters XTerra RPC18 4.6/150 column with Method 1 and H₂O-TFA 0.1% (A) and CH₃OH-TFA 0.1% (B) as eluents (ESI†).

L3: HPLC-MS, method 1, retention time 13.89 min, purity 96%; ¹H-NMR (D₂O) 400 MHz δ = 7.22 (s, 3H, CH), 4.24 (s, 6H, CH₂OCO), 3.46 (s, 12H, CH₂CO), 3.17 (s, 12H, CH₂CO), 2.98 (d, 6H, *J* = 14.4 Hz, CH₂C), 2.80 (d, 6H, *J* = 14.4 Hz, CH₂C), 2.81–2.60 (m, 12H, CH₂CH₂). ¹³C-NMR (D₂O) 100 MHz δ = 181.4–179.1 (CO), 155.6 (NHCOO), 139.4 (CH), 105.7 (C), 68.3 (CH₂OCO), 62.7 (C), 64.1, 61.4 (CH₂CO), 58.2, 54.4 (CH₂cyclo). ESI-MS (*m/z*): 1333.18 (M + H⁺). Calc for C₅₁H₇₃N₁₂O₃₀: 1333.46.

4: (eluent dichloromethane/methyl alcohol 98/2 to 92/8; yield 93%, calculated on polyamine). ¹H-NMR (CDCl₃) 400 MHz δ = 7.78 (bs, 4H, CH), 7.49 (bs, 2H, CH), 7.32 (bs, 2H, NH), 7.02 (bs, 4H, NH), 6.77 (bs, 2H, NH), 4.20 (s, 8H, CH₂OCO), 3.70 (s, 16H, CH₂CO), 3.46 (bs, 4H, CH₂NHCS), 3.26 (s, 16H, CH₂CO), 3.09 (d, 8H, *J* = 13.6 Hz, CH₂C), 2.77–2.63 (m, 24H, CH₂cyclo), 1.42 (s, 144H, CH₃). ¹³C-NMR (CDCl₃) 100 MHz δ = 172.8, 170.9 (CO), 153.3 (NHCOO), 140.5 (bs, C), 108.6 (bs, CH), 81.0, 80.8 (CCH₃), 68.0 (CH₂OCO), 63.3 (C), 62.3 (CH₂CO), 58.9, 51.6 (CH₂cyclo), 51.8 (CH₂NHCS), 28.3, 28.2 (CH₃). ESI-MS (*m/z*): 1450.59 (M + 2H⁺)/2 (5%), 968.53 (M + 3H⁺)/3 (30%), 726.29 (M + 4H⁺)/4 (100%). Calc for (C₁₄₀H₂₃₆N₂₀O₄₆S₂)/2:

1450.82, (C₁₄₀H₂₃₇N₂₀O₄₀S₂)/3: 967.55, (C₁₄₀H₂₃₈N₂₀O₄₀S₂)/4: 725.91.

L4: HPLC-MS, method 1, retention time 14.94 min, purity 82%; ¹H-NMR (D₂O) 400 MHz δ = 7.43–7.18 (bs, 6H, CH), 4.30 (s, 8H, CH₂OCO), 3.92 (bs, 4H, CH₂NHCS), 3.57 (s, 16H, CH₂CO), 3.41 (s, 16H, CH₂CO), 3.15 (bs, 32H, CH₂cyclo). ¹³C-NMR (D₂O) 100 MHz δ = 180.7, 180.6 (CO), 155.4 (NHCOO), 139.6 (bs, C), 110.0 (bs, CH), 67.8 (CH₂OCO), 63.8 (C), 62.9 (CH₂CO), 60.0, 54.2 (CH₂cyclo), 55.4 (CH₂NHCS).

6: (eluent petroleum ether/ethyl acetate 8/2 to 2/8; yield 56%). ¹H-NMR (CDCl₃) 400 MHz δ = 7.7–7.0 (bs, 9H, CH), 4.18 (s, 12H, CH₂OCO), 3.71 (s, 24H, CH₂CO), 3.56 (bs, 6H, CH₂NHCS), 3.26 (s, 24H, CH₂CO), 3.11 (d, 12H, *J* = 13.6 Hz, CH₂C), 2.78–2.63 (m, 36H, CH₂cyclo), 2.07 (bs, 6H, CH₂N), 1.43 (s, 216H, CH₃). ¹³C-NMR (CDCl₃) 100 MHz δ = 172.8, 170.9 (CO), 153.1 (NHCOO), 139.8 (C), 116.1 (CH), 80.9, 80.7 (CCH₃), 68.2 (bs, CH₂OCO), 63.3 (C), 62.3 (CH₂CO + CH₂N), 58.9, 51.6 (CH₂cyclo + CH₂NHCS), 28.3, 28.2 (CH₃). ESI-MS (*m/z*): 1470.94 (M + 3H⁺)/3 (2%), 1103.74 (M + 4H⁺)/4 (5%), 882.86 (M + 5H⁺)/5 (60%), 735.74 (M + 6H⁺)/6 (100%), 630.28 (M + 7H⁺)/7 (15%). Calc for (C₂₁₃H₃₆₀N₃₁O₆₀S₃)/3: 1469.51, (C₂₁₃H₃₆₁N₃₁O₆₀S₃)/4: 1102.38, (C₂₁₃H₃₆₂N₃₁O₆₀S₃)/5: 882.10, (C₂₁₃H₃₆₃N₃₁O₆₀S₃)/6: 735.25, (C₂₁₃H₃₆₄N₃₁O₆₀S₃)/7: 630.36.

L6: HPLC-MS, method 1, retention time 14.00 min, purity 85%; ¹H-NMR (D₂O) 400 MHz δ = 7.7–7.3 (bs, 6H, CH), 4.22 (s, 12H, CH₂OCO), 4.05 (bs, 6H, CH₂NHCS), 3.89 (s, 24H, CH₂CO), 3.78 (s, 24H, CH₂CO), 3.74 (bs, 6H, CH₂N), 3.54–3.47 (bs, 48H, CH₂cyclo). ¹³C-NMR (D₂O) 100 MHz δ = 176.9, 171.3 (CO), 154.4 (NHCOO), 139.6 (C), 127.9 (CH), 66.2 (CH₂OCO), 61.8 (C), 58.9 (NCH₂) 58.8 and 58.3 (CH₂CO), 51.6 (CH₂NHCS), 53.5, 51.5 (CH₂cyclo).

Relaxometric analysis

The ¹H 1/*T*₁ NMRD profiles were obtained with a fast-field cycling Stellar SmartTracer relaxometer (Mede, Pavia, Italy) varying the magnetic-field strength from 0.00024 to 0.25 T (0.01–10 MHz range). The 1/*T*₁ values are measured with an absolute uncertainty of $\pm 1\%$. Temperature was controlled with a Stellar VTC-91 airflow heater equipped with a calibrated copper-constantan thermocouple (uncertainty of ± 0.1 K). Data at high fields (0.5–3 T, corresponding to 20–120 MHz proton Larmor frequency) were collected with a High Field Relaxometer (Stellar) equipped with a HTS-110 3T Metrology Cryogen-free Superconducting Magnet. The measurements were performed with a standard inversion recovery sequence (20 experiments, 2 scans) with a typical 90° pulse width of 3.5 μ s, and the reproducibility of the data was within $\pm 0.5\%$. Additional data points at 400 MHz were obtained using a JEOL ECP 400 spectrometer. The exact concentration of Gd^{III} was determined by measurement of bulk magnetic susceptibility shifts of a *t*BuOH signal or by inductively coupled plasma mass spectrometry (ICP-MS, Element-2, ThermoFinnigan, Rodano (MI), Italy). Sample digestion was performed with concentrated HNO₃ (70%, 2 mL) under microwave heating at 160 °C for 20 min (Milestone MicroSYNTH Microwave lab station equipped with an optical fiber temperature control and

HPR-1000/6 M six position high pressure reactor, Bergamo, Italy).

¹⁷O NMR measurements

The spectra were acquired on a Bruker Avance III spectrometer (11.7 T) using a 5 mm probe under temperature control. An aqueous solution of the complex was enriched to reach 2.0% of the ¹⁷O isotope (Cambridge Isotope). The transverse relaxation rates were measured from the signal width at half-height as a function of temperature in the 278–350 K range.

Computational modelling

All the calculations were performed at the Density Functional Level, using the hybrid Generalized Gradient Approximation (GGA) functional B3LYP and the Hay-Wadt large-core effective core potentials and basis sets. Solvent (water) effects were included through the implicit Polarizable Continuum Model (PCM) implemented in Gaussian09 program.

MRI. MR images were acquired at 1 T on an Aspect MRI System (Aspect Magnet Technologies Ltd., Netanya, Israel) consisting of a NdFeB magnet, equipped with a solenoid coil of 35 mm inner diameter, at 3 T on a Bruker BioSpec 3 T and at 7 T on a Bruker Avance300 spectrometer equipped with a Micro 2.5 microimaging probe (Bruker BioSpin, Ettlingen, Germany). MR images were acquired using a standard *T*₁-weighted multi-slice spin echo sequence, using the following parameters: TR/TE/NEX 250/8/16, FOV 3 cm, slice thickness 1 mm at both 7 and 3 T magnetic field strengths.

Conclusions

In conclusion, the set of these results confirms the predictions of previous simulations and clearly demonstrates that the strategy for relaxation enhancement varies with the strength of the magnetic field used.^{9,11} Up to about 3 T, relaxivity is essentially controlled by the rotation of the hydrate complex (for systems that satisfy the rapid exchange condition) and therefore increases proportionally with the increase in the molecular size. Of course, it must be taken into account that for τ_R values close to or greater than one nanosecond the broad relaxivity peak at high fields becomes narrower and shifts at low frequencies.⁸ Therefore, while macromolecular systems are most effective at 0.5 T, a complex such as Gd₆L6 has a maximum of *r*₁ around 1.5 T. Between 3 and 7 T, the issue of local flexibility or anisotropic rotation becomes more and more relevant. Polynuclear systems of large dimensions and, necessarily, of high complexity do not provide significantly higher relaxation values than those of medium-sized systems ($\tau_R \sim 0.5$ –1 ns), stereochemically rigid and characterized by isotropic molecular tumbling. Large improvements are observed by restricting the local motions. At ultra-high fields (>7 T), small and compact metal complexes (mono or binuclear) have an efficacy quite comparable to that of medium-sized complexes and much greater than macromolecular systems.²⁷ The Gd₃L3 complex appears to represent an excellent compromise

between the different alternatives, as it shows high r_1 values (per Gd) that vary negligibly with the increase in the applied magnetic field: 15.8, 16.3, 16.0 and 14 $\text{mM}^{-1} \text{s}^{-1}$ at 0.5, 1.5, 3 and 7 T, respectively.

Finally, it is worth mentioning that, in addition to the optimization of rotational dynamics, other mechanisms may provide additional contributions to the effectiveness of a metal-based MRI CA, such as the presence of a well-defined network of second-sphere water molecules or the prototropic exchange, associated with the presence of mobile protons near the metal centre.^{40,41}

Author contributions

LT and MB conceived and supervised the project. GG carried out the synthesis and collected the relaxometric data. DM and MC performed the computational study. SGC carried out the phantom imaging study. MB and LT analysed the relaxometric data and wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

L. T. and M. B. acknowledge the financial support from Università del Piemonte Orientale (Ricerca locale 2019). This work was carried out within the framework of the COST CA15209 Action 'European Network on NMR Relaxometry'.

Notes and references

- (a) D. B. Plewes and W. Kucharczyk, Physics of MRI: A Primer, *J. Magn. Reson. Imaging*, 2012, **35**, 1038; (b) M. Bottrill, L. Kwok and N. J. Long, Lanthanides in magnetic resonance imaging, *Chem. Soc. Rev.*, 2006, **35**, 557.
- S. H. Koenig, Molecular Basis of Magnetic Relaxation of Water Protons of Tissue, *Acad. Radiol.*, 1996, **3**, 597.
- (a) A. Merbach, L. Helm and É. Tóth, *The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging*, John Wiley & Sons, New York, 2nd edn, 2013; (b) L. Helm, J. R. Morrow, C. J. Bond, F. Carniato, M. Botta, M. Braun, Z. Baranyai, R. Pujales-Paradel, M. Regueiro-Figuero, D. Esteban-Gómez, C. Platas-Iglesias and T. J. Scholl, in *Contrast Agents for MRI: Experimental Methods*, ed. V. C. Pierre and M. J. Allen, The Royal Society of Chemistry, 2017, ch. 2, pp. 121–242.
- J. Wahsner, E. M. Gale, A. Rodríguez-Rodríguez and P. Caravan, Chemistry of MRI contrast agents: current challenges and new frontiers, *Chem. Rev.*, 2019, **119**, 957.
- (a) O. Kraff, A. Fischer, A. M. Nagel, C. Mönninghoff and M. E. Ladd, MRI at 7 Tesla and above: demonstrated and potential capabilities, *J. Magn. Reson. Imaging*, 2015, **41**, 13; (b) E. Moser, F. Stahlberg, M. E. Ladd and S. Trattnig, 7-T MR- from research to clinical applications?, *NMR Biomed.*, 2012, **25**, 695.
- M. A. Schwarz, M. Pham, X. Helluy, A. Doerfler and T. Engelhorn, MRI assessment of experimental gliomas using 17.6 T, *Neuroradiology*, 2013, **55**, 709.
- M. Botta and L. Tei, Relaxivity Enhancement in Macromolecular and Nanosized Gd^{III}-Based MRI Contrast Agents, *Eur. J. Inorg. Chem.*, 2012, 1945.
- S. Aime, M. Botta, D. Esteban-Gómez and C. Platas-Iglesias, Characterisation of magnetic resonance imaging (MRI) contrast agents using NMR relaxometry, *Mol. Phys.*, 2018, **117**, 898.
- P. Caravan, C. T. Farrara, L. Frullano and U. Ritika, Influence of molecular parameters and increasing magnetic field strength on relaxivity of gadolinium- and manganese-based T₁ contrast agents, *Contrast Media Mol. Imaging*, 2009, **4**, 89.
- J. B. Livramento, E. Tóth, A. Sour, A. Borel, A. E. Merbach and R. Ruloff, High relaxivity confined to a small molecular space: a metallostar-based, potential MRI contrast agent, *Angew. Chem., Int. Ed.*, 2005, **44**, 1480.
- L. Helm, Optimization of gadolinium-based MRI contrast agents for high magnetic-field applications, *Future Med. Chem.*, 2010, **2**, 385.
- A. N. W. Kuda-Wedagedara and M. J. Allen, Enhancing Magnetic Resonance Imaging with Contrast Agents for Ultra-High Field Strengths, *Analyst*, 2014, **139**, 4401.
- P. Fries, J. N. Morelli, F. Lux, O. Tillement, G. Schneider and A. Buecker, The issues and tentative solutions for contrast-enhanced magnetic resonance imaging at ultra-high field strength, *WIREs Nanomed. Nanobiotechnol.*, 2014, **6**, 559.
- L. Banci, I. Bertini and C. Luchinat, *Nuclear and Electron Relaxation. The Magnetic Nucleus-Unpaired Electron Coupling in Solution*, VCH, Weinheim, Germany, 1991.
- (a) S. Aime, L. Calabi, C. Cavallotti, E. Gianolio, G. B. Giovenzana, P. Losi, A. Maiocchi, G. Palmisano and M. Sisti, [Gd-AAZTA]⁺: a new structural entry for an improved generation of MRI contrast agents, *Inorg. Chem.*, 2004, **43**, 7588; (b) Z. Baranyai, F. Uggeri, G. B. Giovenzana, A. Benyei, E. Brucher and S. Aime, Equilibrium and kinetic properties of the lanthanoids(III) and various divalent metal complexes of the heptadentate ligand AAZTA, *Chem. – Eur. J.*, 2009, **15**, 1696.
- A. Vagner, E. Gianolio, S. Aime, A. Maiocchi, I. Tóth, Z. Baranyai and L. Tei, High kinetic inertness of a bis-hydrated Gd-complex with a constrained AAZTA-like ligand, *Chem. Commun.*, 2016, **52**, 11235.
- R. S. Ranganathan, M. E. Fernandez, S. I. Kang, A. D. Nunn, P. C. Ratsep, K. M. R. Pillai, X. Zhang and M. F. Tweedle, New multimeric magnetic resonance imaging agents, *Invest. Radiol.*, 1998, **33**, 779.
- V. V. Martin, W. H. Ralston, M. R. Hynes and J. F. W. Keana, Gadolinium(III) Di- and Tetrachelates

- Designed for in Vivo Noncovalent Complexation with Plasma Proteins: A Novel Molecular Design for Blood Pool MRI Contrast Enhancing Agents, *Bioconjugate Chem.*, 1995, **6**, 616.
- 19 (a) Y. Song, E. K. Kohlmeier and T. J. Meade, Synthesis of multimeric MR contrast agents for cellular imaging, *J. Am. Chem. Soc.*, 2008, **130**, 6662; (b) J. B. Livramento, L. Helm, A. Sour, C. O'Neil, A. E. Merbach and E. Tóth, A benzene-core trinuclear Gd^{III} complex: towards the optimization of relaxivity for MRI contrast agent applications at high magnetic field, *Dalton Trans.*, 2008, 1195; (c) D. J. Mastarone, V. S. R. Harrison, A. L. Eckermann, G. Parigi, C. Luchinat and T. J. Meade, A Modular System for the Synthesis of Multiplexed Magnetic Resonance Probes, *J. Am. Chem. Soc.*, 2011, **133**, 5329.
- 20 (a) D. T. Schuhle, J. Schatz, S. Laurent, L. Vander Elst, R. N. Muller, M. C. A. Stuart and J. A. Peters, Calix[4]arenes as molecular platforms for magnetic resonance imaging (MRI) contrast agents, *Chem. – Eur. J.*, 2009, **15**, 3290; (b) D. T. Schuhle, M. Polasek, I. Lukes, T. Chauvin, E. Tóth, J. Schatz, U. Hanefeld, M. C. A. Stuart and J. A. Peters, Densely packed Gd(iii)-chelates with fast water exchange on a calix[4]arene scaffold: a potential MRI contrast agent, *Dalton Trans.*, 2010, **39**, 185.
- 21 Z. Kotkova, L. Helm, J. Kotek, P. Hermann and I. Lukes, Gadolinium complexes of monophosphinic acid DOTA derivatives conjugated to cyclodextrin scaffolds: efficient MRI contrast agents for higher magnetic fields, *Dalton Trans.*, 2012, **41**, 13509.
- 22 J. M. Bryson, W.-J. Chu, J.-H. Lee and T. M. Reineke, A β -Cyclodextrin “Click Cluster” Decorated with Seven Paramagnetic Chelates Containing Two Water Exchange Sites, *Bioconjugate Chem.*, 2008, **19**, 1505.
- 23 J. Henig, E. Tóth, J. Engelmann, S. Gottschalk and H. A. Mayer, Macrocyclic Gd³⁺ chelates attached to a silsesquioxane core as potential magnetic resonance imaging contrast agents: synthesis, physicochemical characterization, and stability studies, *Inorg. Chem.*, 2010, **49**, 6124.
- 24 M. Ndiaye, V. Malyskyi, T. Vangijzegem, F. Sauvage, M. Wels, C. Cadiou, J. Moreau, C. Henoumont, S. Boutry, R. N. Muller, D. Harakat, S. De Smedt, S. Laurent and F. Chuburu, Comparison of MRI Properties between Multimeric DOTAGA and DO3A Gadolinium-Dendron Conjugates, *Inorg. Chem.*, 2019, **58**, 12798.
- 25 (a) K. Overoye-Chan, S. Koerner, R. J. Looby, A. F. Kolodziej, S. G. Zech, Q. Deng, J. M. Chasse, T. J. McMurry and P. Caravan, EP-2104R: a fibrin-specific gadolinium-Based MRI contrast agent for detection of thrombus, *J. Am. Chem. Soc.*, 2008, **130**, 6025; (b) S. A. Nair, A. F. Kolodziej, G. Bhole, M. T. Greenfield, T. J. McMurry and P. Caravan, Monovalent and bivalent fibrin-specific MRI contrast agents for detection of thrombus, *Angew. Chem., Int. Ed.*, 2008, **47**, 4918.
- 26 (a) E. Gianolio, K. Ramalingam, B. Song, F. Kalman, S. Aime and R. Swenson, Improving the relaxivity by dimerizing Gd-AAZTA: Insights for enhancing the sensitivity of MRI contrast agents, *Inorg. Chem. Commun.*, 2010, **13**, 663; (b) G. Gambino, S. De Pinto, L. Tei, C. Cassino, F. Arena, E. Gianolio and M. Botta, A new ditopic Gd^{III} complex functionalized with an adamantyl moiety as a versatile building block for the preparation of supramolecular assemblies, *J. Biol. Inorg. Chem.*, 2014, **19**, 133.
- 27 L. Leone, L. Guarnieri, J. Martinelli, M. Sisti, A. Penoni, M. Botta and L. Tei, Rigid and Compact Binuclear Bis-hydrated Gd-complexes as High Relaxivity MRI Agents, *Chem. – Eur. J.*, 2021, **27**, 11811.
- 28 M. Tripepi, F. Capuana, E. Gianolio, F. V. C. Kock, A. Pagoto, R. Stefania, G. Digilio and S. Aime, Synthesis of High Relaxivity Gadolinium AAZTA Tetramers as Building Blocks for Bioconjugation, *Bioconjugate Chem.*, 2018, **18**, 1428.
- 29 G. Gugliotta, M. Botta and L. Tei, AAZTA-based bifunctional chelating agents for the synthesis of multimeric/dendritic MRI contrast agents, *Org. Biomol. Chem.*, 2010, **8**, 4569.
- 30 L. Tei, G. Gugliotta, G. Gambino, M. Fekete and M. Botta, Developing High Field MRI Contrast Agents by Tuning the Rotational Dynamics: Bisoqua GdAAZTA-based Dendrimers, *Isr. J. Chem.*, 2017, **57**, 887–895.
- 31 G. Gugliotta, M. Botta, G. B. Giovenzana and L. Tei, Fast and easy access to efficient bifunctional chelators for MRI applications, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 3442.
- 32 S. J. Atkinson, V.-J. Ellis, S. E. Boyd and C. L. Brown, Synthesis and characterisation of peripherally functionalised dendritic molecules, *New J. Chem.*, 2007, **31**, 155.
- 33 (a) S. H. Koenig and R. D. Brown III, Field-cycling relaxometry of protein solutions and tissue: Implications for MRI, *Prog. NMR Spectrosc.*, 1990, **22**, 487–567; (b) R. B. Lauffer, Paramagnetic metal complexes as water proton relaxation agents for NMR imaging: theory and design, *Chem. Rev.*, 1987, **87**, 901–927.
- 34 S. Karimi, L. Tei, M. Botta and L. Helm, Evaluation of water exchange kinetics on [Ln(AAZTAPh-NO₂)(H₂O)_q]_k complexes using proton Nuclear Magnetic Resonance, *Inorg. Chem.*, 2016, **55**, 6300.
- 35 T. J. Swift and R. E. J. Connick, NMR-relaxation mechanisms of ¹⁷O in aqueous solutions of paramagnetic cations and the lifetime of water molecules in the first coordination sphere, *J. Chem. Phys.*, 1962, **37**, 307.
- 36 J. P. Korb, M. Ahadi, G. P. Zientara and J. H. Freed, Dynamic effects of pair correlation functions on spin relaxation by translational diffusion in two-dimensional fluids, *J. Chem. Phys.*, 1987, **86**, 1125.
- 37 (a) G. Lipari and S. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc.*, 1982, **104**, 4546; (b) G. Lipari and S. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, *J. Am. Chem. Soc.*, 1982, **104**, 4559.
- 38 V. C. Pierre, M. Botta, S. Aime and K. N. Raymond, Fe(III) Templated Gd(III) Self-Assemblies—a New Route Toward

Research Article

Inorganic Chemistry Frontiers

- Macromolecular MRI Contrast Agents, *J. Am. Chem. Soc.*, 2006, **128**, 2972–2973.
- 39 K. N. Raymond and V. C. Pierre, Next generation, high relaxivity gadolinium MRI agents, *Bioconjugate Chem.*, 2005, **16**, 3–8.
- 40 M. Botta, Second Coordination Sphere Water Molecules and Relaxivity of Gadolinium(III) Complexes: Implications for MRI Contrast Agents, *Eur. J. Inorg. Chem.*, 2000, 399.
- 41 (a) L. Leone, M. Boccalon, G. Ferrauto, I. Fabian, Z. Baranyai and L. Tei, Acid-catalyzed proton exchange as a novel approach for relaxivity enhancement in GdHPDO3A-like complexes, *Chem. Sci.*, 2020, **11**, 7829; (b) S. Aime, S. Baroni, D. Delli Castelli, E. Brücher, I. Fábíán, S. Colombo Serra, A. Fringuello Mingo, R. Napolitano, L. Lattuada, F. Tedoldi and Z. Baranyai, *Inorg. Chem.*, 2018, **57**, 5567.

Appendix B

Hyper-Cross-Linked Polymers for the Capture of Aromatic Volatile Compounds

Hyper-Cross-Linked Polymers for the Capture of Aromatic Volatile Compounds

G. Paul,[†] F. Begni,[†] A. Melicchio,[‡] G. Golemme,[‡] C. Bisio,^{*,†,§} D. Marchi,[†] M. Cossi,[†] L. Marchese,[†] and G. Gatti[†]

[†]Department of Sciences and Technological Innovation and Interdisciplinary Nano-SiSTeMI Center, University of Eastern Piedmont A. Avogadro, via Teresa Michel 11, Alessandria, Italy

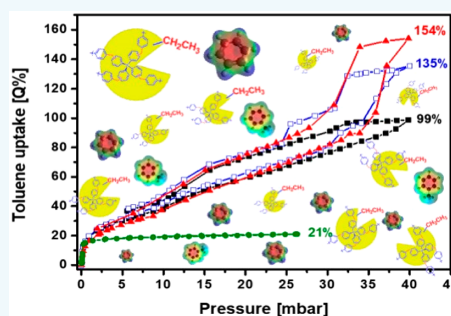
[‡]Department of Environmental and Chemical Engineering, University of Calabria, via P. Bucci 45A, Rende 87036, Italy

[§]CNR-SCITEC Institute of Science and Molecular Technologies “G. Natta”, via C. Golgi 19, Milano 20133, Italy

Supporting Information

ABSTRACT: Hyper-cross-linked porous polymers (HCPs) are proposed as sorbents for the removal of aromatic volatile pollutants by using toluene as a representative of the BTX family. The hierarchical (micro and meso) porous architecture of the HCPs has been established by N₂ physisorption at 77 K while the toluene adsorption capacities were determined by volumetric adsorption at 308 K. The HCPs display very high toluene uptakes, reaching adsorption capacities as high as 154% in weight for the polymer obtained with a tetraphenylmethane (TPM) and a formaldehyde dimethyl acetal (FDA) ratio of 1/16, whereas only very low uptakes were observed for aliphatic molecules such as *n*-hexane. HCP materials experience swelling effects evaluated by comparing the volume assessed via N₂ physisorption with the volume occupied by toluene molecules in volumetric adsorption experiments. A multispectroscopic approach involving FT-IR and solid-state NMR techniques gave direct proof of the close spatial proximity between the polymeric host framework and guest BTX molecules. Solid-state ¹H and ¹³C NMR spectroscopies have unambiguously identified the presence of CH/ π interactions between the guest molecules and the porous framework of the hyper-cross-linked polymers.

KEYWORDS: hyper-cross-linked polymers, aromatic pollutants adsorption, noncovalent interactions, volumetric adsorption, ss-NMR spectroscopy



INTRODUCTION

Hyper-cross-linked polymer (HCP)-based microporous materials, obtained through Friedel–Crafts alkylation reaction of aromatic monomers, possess an extremely high specific surface area.¹ Such materials perform a variety of functions, including gas storage and separation.^{1,2} The efficiency, selectivity, and gas storage strengths are critically dependent on the pore architecture as well as the noncovalent host–guest interactions. The key question in designing and synthesizing the host porous organic framework involves the understanding of the nature and role of potential noncovalent interactions existing during gas storage and separation processes. The demand for a fast and efficient removal of volatile organic pollutants received a great deal of attention due to their high abundance and associated impact on human health and environments. In particular, volatile pollutants coming from oil and chemical industries represent a major threat to the environment. Among the remedies to overcome this problem,^{3,4} the use of solid adsorbents is highly recommended due to its low cost and ease of operation.^{4–7}

In this work, HCP materials with excellent physicochemical and thermal stability, high selectivity, and large adsorption capacity⁸ are used as solid sorbents. There are reports in the literature about the use of HCPs for CO₂ capture purposes,^{9–12} whereas inorganic materials^{6,7,13} and porous carbons¹⁴ have been mostly proposed for the adsorption of organic pollutants. So far, the implementation of hyper-cross-linked polymers for removal of hydrocarbon pollutants is not extensively covered in the literature.^{15,16} Investigating adsorption processes involving hyper-cross-linked polymers and hydrocarbon pollutants demands an understanding of the host–guest interactions at play. In particular, it is expected for such a system that noncovalent interactions could play a significant role in guiding adsorption processes.

In recent years, it is becoming clear that noncovalent interactions are responsible for a large number of phenomena including the packing of molecules as well as the three-

Received: October 21, 2019

Accepted: December 16, 2019

Published: December 16, 2019

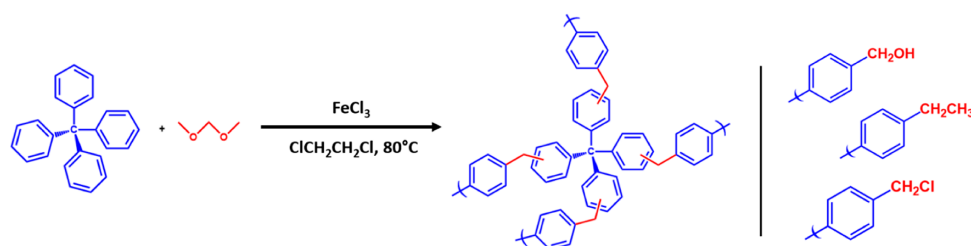


Figure 1. Scheme of reaction for the synthesis of mPAF-1/*n* materials. Possible side reaction groups are presented on the right. A qualitative estimation showed that at least one side-chain (either $-\text{CH}_2\text{OH}$, $-\text{CH}_2\text{Cl}$, $-\text{CH}_2\text{CH}_3$, or $-\text{CH}_3$) per aromatic ring exists in the mPAF-1/16 material.

dimensional structures and functions of a variety of chemical, material, and biological systems.^{17–20} Among them, cation/ π , CH/ π , and π/π interactions, although considered weaker, constitute one of the most important classes of noncovalent interactions and are important for the binding affinities of host–guest complex systems. CH/ π interactions, the family of hydrogen-bond-like interactions in which a delocalized system of sp^2 -hybridized covalent bonds can act as an acceptor group, have long been proposed to contribute to biomolecular structure and function.^{20–22} It is generally recognized that between CH bonds and π -systems there are attractive noncovalent forces between the protons of an alkyl group and the π face of an aromatic ring.^{23,24} There is a large number of articles in the literature that suggest the existence of CH/ π interactions in crystalline materials highlighting their importance in molecular recognition and crystal packing.^{25,26} Excellent reviews have appeared recently, detailing the evidence, nature, characteristics, and consequences of the CH/ π interactions in chemical and biological systems.^{27–30}

In this work, the adsorption capacities of mPAF-1/*n* materials, as in microporous porous aromatic frameworks (already studied by our group for CO_2 and methane adsorption³¹), are exploited for the removal of vapor phase organic pollutants (namely benzene, toluene, and xylene as representative of aromatic hydrocarbons) and compared to those of well-established materials such as all-silica zeolite HSZ-Y.^{6,13,32} Moreover, to the best of our knowledge, a careful study of the adsorption process taking place between a hyper-cross-linked polymeric structure and an aromatic molecule is missing from the scientific literature. This work aims to fill this gap. We have employed an integrated spectroscopic approach involving solid-state nuclear magnetic resonance (ss-NMR) and FT-IR to determine the intimate spatial relationship and interactions between host polymeric porous framework and guest toluene molecules. Combined spectroscopic approaches involving FT-IR and ss-NMR have distinct advantages in the studies of weak interactions in structurally complex systems. While ss-NMR spectroscopy provides, through the observation of multiple nuclei, different spectral insights related to host–guest interactions and to the mobility of the guest molecule inside the porous host, FT-IR spectroscopy can deliver fundamental information about surface features of materials and on both host–guest and guest–guest interactions.³³ Furthermore, the experimental data were validated by employing *ab initio* calculations performed at the density functional theory (DFT) level using the Gaussian16 program.

MATERIALS AND METHODS

Synthesis of Materials. The synthetic procedure for the synthesis of the hyper-cross-linked porous polymer material mPAF was already established;^{9,31} hence, only a brief description is given here. The standard procedure requires the suspension of the precursor monomer tetraphenylmethane (TPM), delivered by Capot Chemical Company (97%), and the catalyst, iron(III) chloride (Sigma-Aldrich, 97%), in 1,2-dichloroethane (Sigma-Aldrich, 99%). Formaldehyde dimethyl acetal (FDA), delivered by Sigma-Aldrich (99%), is then added dropwise. The resulting mixture is stirred at room temperature and then heated at 353 K under reflux overnight. A brown precipitate, which is then washed with ethanol and water, is formed after polymerization. The material is then dried in an oven. Three different TPM/FDA (1/*n*) molar ratios of 1/9, 1/16, and 1/30 were used, and the products, named mPAF-1/9, mPAF-1/16, and mPAF-1/30, respectively,³¹ were obtained. The reaction scheme is presented in Figure 1.

Characterization Techniques. Toluene adsorption isotherms were obtained at 308 K on HSZ-Y and mPAF-1/*n* samples by employing a volumetric analysis of vapor sorption in an Autosorb iQ MP-XR equipped with a cryocooler (Quantachrome Instruments). To remove possible adsorbed species, prior to the adsorption measurements, all samples were outgassed for 30 min at 323 K, 30 min at 353 K, 2 h at 393 K, and 2 h at 423 K. For the mPAF-1/*n* a final treatment at 12 h at 493 K under high-vacuum conditions (final pressure 7×10^{-4} mbar) was adopted.

The specific surface area (SSA) was measured by means of nitrogen adsorption at liquid nitrogen temperature (77 K) in the pressure range of 1×10^{-6} Torr to 1 P/P_0 by using an Autosorb-1-MP (Quantachrome Instruments). Prior to adsorption, the samples were outgassed for 16 h at 423 K, (final pressure lower than 10^{-6} Torr). The SSA of the samples was determined by the Brunauer–Emmett–Teller (BET) equation, in a pressure range (0.05–0.15 and 0.005–0.01 P/P_0 range for mPAF-1/*n* and HSZ-Y zeolite, respectively) selected to maximize the correlation coefficient of the fitted linear equation. The pore size distribution was calculated by applying the nonlocal density functional theory (NLDFT) method for cylindrical pores.

Infrared spectra were collected on a Thermo Electron Corporation FT Nicolet 5700 spectrometer (resolution 4 cm^{-1}). Pellets were prepared by mixing the prepared materials with KBr (1:10 weight ratio). Pellets were placed into an IR cell with KBr windows permanently connected to a vacuum line (residual pressure: 1.33×10^{-4} Pa, 1 Pa = 0.01 mbar), allowing all treatments and adsorption–desorption experiments to be performed *in situ*. Before the gas adsorption, mPAF-1/*n* samples were outgassed at 423 K with a heating ramp of 5 K/min for 2 h, using an oil-free apparatus and grease-free vacuum line.

Solid-state NMR spectra were acquired on a Bruker Avance III 500 spectrometer and a wide bore 11.7 T magnet with operational frequencies for ^1H and ^{13}C of 500.13 and 125.77 MHz, respectively. A 4 mm triple-resonance probe with MAS was employed in all experiments. 19 mbar of toluene- d_6 , toluene, benzene, xylene, and *n*-hexane were adsorbed directly on a dehydrated powdered sample

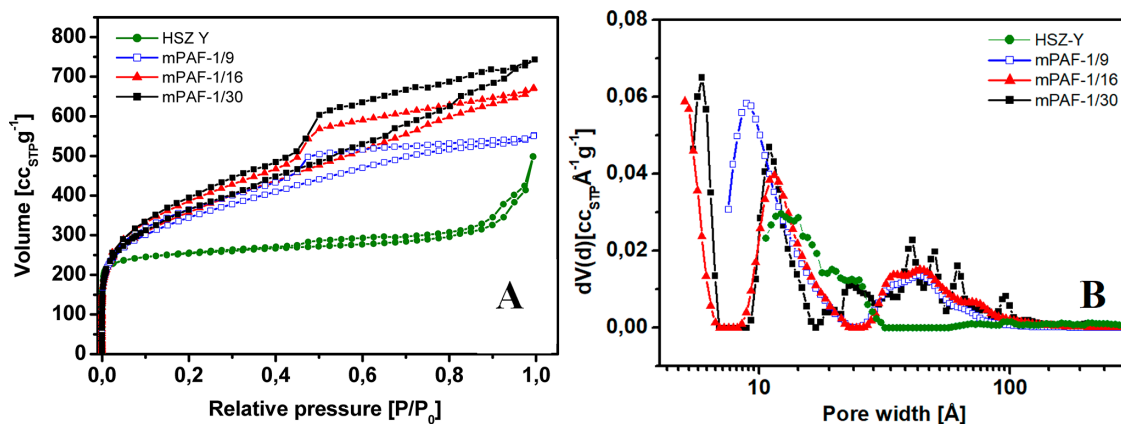


Figure 2. N_2 physisorption isotherms at 77 K (A) and pore size distribution (B) of (blue \square) mPAF-1/9, (red \blacktriangle) mPAF-1/16, (black \blacksquare) mPAF-1/30, and (green \bullet) HSZ-Y. Pore size analysis performed with the NLDFT method for silica cylindrical pores in the adsorption branch.

Table 1. Main Textural Properties of mPAF-1/*n* Materials and HSZ-Y Zeolite Obtained from N_2 Physisorption at 77 K

sample	SSA _{BET} [m ² /g]	V _{Tot} [cm ³ /g]	V _{micro} [cm ³ /g]			V _{meso} [cm ³ /g]
			total	<7 Å	7 < Å < 20	20 < Å < 300
HSZ-Y	991	0.68	0.28		0.28	0.40
mPAF-1/9	1236	0.78	0.33		0.33	0.45
mPAF-1/16	1289	0.96	0.30	0.09	0.21	0.66
mPAF-1/30	1318	1.06	0.23	0.07	0.16	0.83

that was previously packed on a zirconia rotor and inserted in a homemade cell.³³ After waiting for the equilibrium adsorption, we closed the rotor with a zirconia cap using a piston attached to the cell, and later the rotor was extracted from the cell and submitted for solid-state NMR experiments. The zirconia rotor was spun at a MAS rate between 10 and 15 kHz. The magnitude of radio-frequency field was 100 kHz for ¹H MAS NMR, and the relaxation delay, d_1 , between accumulations was 2 s. For the ¹³C cross-polarization (CP) magic angle spinning (MAS) experiments, the proton radio frequencies (RF) of 55 and 28 kHz were used for initial excitation and decoupling, respectively. During the CP period the ¹H RF field was ramped by using 100 increments, whereas the ¹³C RF field was maintained at a constant level. During the acquisition, the protons were decoupled from the carbons by using a TPPM decoupling scheme. A moderate ramped RF field of 62 kHz was used for spin locking, while the carbon RF field was matched to obtain optimal signal and CP contact times of 2–10 ms were used. The rotor synchronized spin echo sequence ($\pi/2-\tau-\pi-\tau$ -acquisition) was also applied to record the ¹H NMR spectra with τ delay time of 4000 μ s. All chemical shifts are reported by using the δ scale and are externally referenced to TMS at 0 ppm.

All the *ab initio* calculations were performed at the density functional theory (DFT) level using the Gaussian16 program.³⁴ The hybrid functional B3LYP³⁵ was employed along with the cc-pVDZ basis set.³⁶ Dispersion energies were included in the model through the semiempirical method and parameters proposed by Grimme,³⁷ and the NMR chemical shifts were calculated by using the gauge-independent atomic orbital (GIAO) method.³⁸

The molecular dynamics (MD) calculations were performed in the NPT ensemble at 300 K and 1 atm using the LAMMPS program.³⁹ Temperature and pressure were controlled through the Nosé–Hoover algorithm; the bond lengths, angles, and dihedrals were kept fixed for the whole run, and long-range nonbonded interactions were computed by using the universal force field (UFF) pairwise parameters.⁴⁰ A periodic cubic box containing 3000 toluene molecules was used to simulate a liquid-like environment, as expected inside the material for high adsorbate concentrations; the model of mPAF-1/*n* material described in the text was then inserted in the box, deleting

the overlapping toluene molecules. The equilibration and production times were 10 and 50 ps, respectively.

RESULTS AND DISCUSSION

Textural Properties of mPAF-1/*n* Materials. N_2 physisorption isotherms of mPAF-1/*n* solids are shown in Figure 2, together with the isotherm of the commercial HSZ-Y zeolite, that was already studied for environmental purposes.^{13,6,7}

The main textural properties, assessed through N_2 physisorption analysis at 77 K, of the analyzed materials, namely mPAF-1/9, mPAF-1/16, mPAF-1/30, and HSZ-Y, are summarized in Table 1.

mPAFs isotherms can be classified as type I with H_2 hysteresis loops indicating the presence of both micro- and mesopores.³¹ The hysteresis in the N_2 sorption–desorption isotherms extends up to very low (0.01) relative pressures, probably due to the swelling of the flexible mPAFs.⁴¹ The shape of the hysteresis and the reversibility of the N_2 sorption at lower pressures indicate that the pores of mPAFs form continuous networks, which allow the complete desorption of N_2 up to the equilibrium values, on the time scale of the experiments. High surface area and pore volumes are found for the porous polymers (Table 1), especially for those with TPM/FDA ratio of 1/16 and 1/30, displaying specific surface area of 1289 and 1318 m²/g, respectively. The total pore volume gradually increases from mPAF-1/9 to mPAF-1/30 because of the increasing fraction of mesopores (pore diameter in the range from 20 to 300 Å) from 0.45 to 0.83 cm³/g. By contrast, micropores between 7 and 20 Å are reduced from 0.33 to 0.16 cm³/g in the same series, probably as a consequence of occlusion effects caused by side reaction groups.

Commercial dealuminated HSZ-Y is characterized by a lower SSA and lower total pore volume with respect to porous polymers. For HSZ-Y the isotherm presents a hysteresis loop which is consistent with the presence of micro- and mesopores, the latter being generated during the dealumination process.¹³

Volumetric Adsorption of Toluene on mPAF-1/n Materials. Volumetric isotherms of toluene adsorbed on mPAF-1/n materials are reported in Figure 3, in comparison

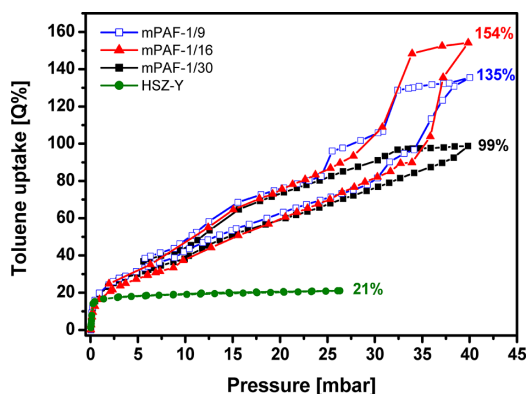


Figure 3. Toluene volumetric adsorption isotherms at 303 K of (blue □) mPAF-1/9, (red ▲) mPAF-1/16, (black ■) mPAF-1/30, and (green ●) HSZ-Y.

with data obtained on HSZ-Y zeolite. Weights of adsorbed toluene were calculated from volumetric data by dividing the volume of adsorbed toluene by the molar volume of an ideal gas at STP. The obtained moles of adsorbed gas were multiplied by the molar mass of the toluene molecule. Toluene uptakes ($Q\%$) were calculated via eq 1

$$Q [\%] = \frac{W_{\text{tol ads}}}{W_{\text{sample}}} \times 100 \quad (1)$$

where $W_{\text{tol ads}}$ is the weight of the adsorbed toluene and W_{sample} is the weight of the bare sample.

As a general feature, adsorption isotherms for all materials are very steep up to 1 mbar, indicating a rapid filling of micropores. For the HSZ-Y zeolite, there is a rapid decrease in the isotherm slope when a pressure of around 2 mbar is reached, and then the curve gradually approaches a plateau. This indicates high affinity of the HSZ-Y surface toward toluene.¹³ HSZ-Y shows a maximum toluene uptake of 21 Q [%].

In the case of mPAF-1/n materials the toluene uptakes gradually increase until a pressure of 40 mbar: the maximum uptakes for mPAF materials are 135, 154, and 99 Q [%] for mPAF-1/9, mPAF-1/16, and mPAF-1/30, respectively. All HCP materials present hysteresis loops, thus suggesting the formation of a metastable liquid-like state inside the mesopores.

Adsorption/desorption isotherms show a higher toluene uptake for less cross-linked mPAF materials with steps at 30 and 34 mbar for both mPAF-1/16 and mPAF-1/9. The isotherm desorption branch of mPAF-1/9 shows two steps at 32 and 25 mbar while the desorption branch for mPAF-1/16 shows one single step below 33 mbar. It is interesting to compare the pore volume measured via N_2 physisorption at 77

K with the volume occupied by toluene at 40 mbar in mPAFs. Table 2 reports on the toluene volumes (C_7) in mPAF materials calculated by dividing the adsorbed toluene mass (per gram of sample) by the toluene density (0.853 g cm^{-3}) at 308 K.

Table 2. Comparison of Pore Volume (per gram of Sample, cm^3/g) of mPAF-1/n Materials and HSZ-Y Zeolite Obtained from N_2 Physisorption at 77 K and from Toluene Sorption at 308 K

sample	SSA_{BET} [m^2/g]	$V_{\text{Tot}}(N_2)$ [cm^3/g]	$V_{\text{Tot}}(C_7)$ [cm^3/g]	$V(C_7)/V(N_2)$
HSZ-Y	991	0.68	0.25	0.36
mPAF-1/9	1236	0.78	1.58	2.03
mPAF-1/16	1289	0.96	1.81	1.88
mPAF-1/30	1318	1.06	1.16	1.09

For HSZ-Y, only 36% of the total pore volume is occupied by toluene, probably as a consequence of inefficient stacking of toluene in a rigid and confined environment. On the contrary, the flexible nature of the mPAF's polymeric network, in addition to its high affinity for organic molecules such as toluene, allows a more efficient stacking of molecules. This results in higher toluene uptakes. For mPAF materials the ratio $V(C_7)/V(N_2)$ decreases by increasing the degree of cross-linking of the polymeric network. Specifically, data associated with samples of mPAF-1/9 and -1/16 show that the volume occupied by toluene at 308 K is 2 times higher with respect to the volume occupied by N_2 at 77 K. The two steps, at 30 and 34 mbar, in the toluene isotherms adsorption branch, which are not present in N_2 physisorption isotherms, can be attributed to a reversible swelling effect of the polymeric chains. This process is inhibited for more cross-linked polymeric frameworks, hence the lower swelling effect observed for the mPAF-1/30.

However, the three mPAFs isotherms resemble each other up to ~ 30 mbar of toluene pressure in adsorption and down from 25 mbar in the desorption branch, thus indicating a similar adsorption mechanism below these pressures.

Reported data indicated that mPAF-1/16 is an optimal candidate for toluene adsorption, thus suggesting that high specific surface area and the cross-linking degree are key factors in improving toluene adsorption capacity.

Monitoring the Adsorption of Toluene on mPAF-1/n Materials by IR Spectroscopy. The adsorption properties of mPAF solids toward gas phase toluene have been studied by IR spectroscopy. Deuterated toluene (toluene- d_8) was used as a probe to avoid the superposition with bands of CH_n groups (i.e., CH_2 and CH_3 species), typical of porous HCPs.³¹

IR spectra of toluene- d_8 adsorbed on mPAF-1/16 material are reported in Figure 4. IR spectra of toluene- d_8 in both liquid and gas phase are reported in Figure S1.

Toluene- d_8 was first admitted at a pressure of 27 mbar and then gradually desorbed. However, for the sake of clarity, spectra are discussed starting from low to high toluene pressure.

The IR spectrum of mPAF-1/16 sample (Figure 4, curve a) presents different absorptions bands between 3060 and 2850 cm^{-1} and between 860 and 780 cm^{-1} , as detailed in Table 3.

The admission of 0.2 mbar of toluene- d_8 on mPAF-1/16 results in the formation of a series of bands in the region between 2300 and 2000 cm^{-1} (Figure 4A, curve b): bands at

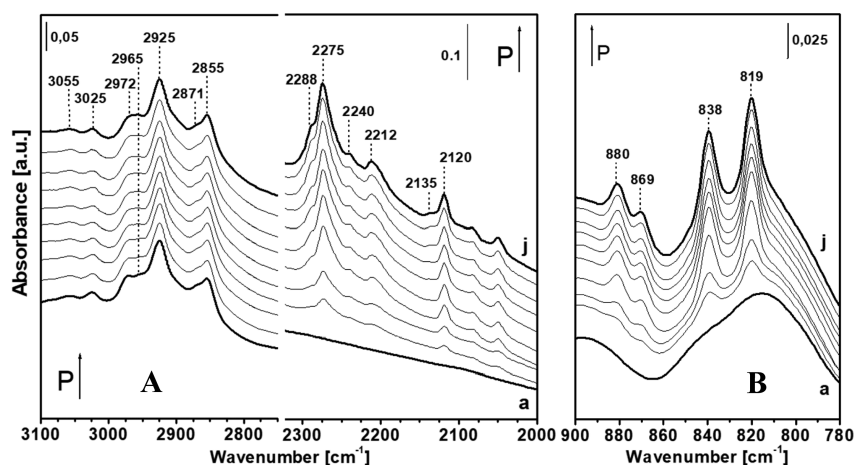


Figure 4. IR spectra of toluene- d_8 adsorbed on mPAF-1/16 at beam temperature (bt). (a) is recorded after outgassing the sample at 423 K for 2 h. Spectra from (b) to (j) are associated with the admission of 0.2 (b), 0.9 (c), 2.1 (d), 6.1 (e), 10.1 (f), 14.1 (g), 22.1 (h), 26.1 (i), and 27.1 mbar (j) of toluene- d_8 .

Table 3. Assignments of the Main IR Vibrations of the mPAF Materials

bands position [cm ⁻¹]	assignments ^{31,42}
3055	ν_{As} aromatic C-H
3025	ν_S aromatic C-H
2972	ν_{As} aliphatic C-H (-CH ₃)
2925	ν_{As} aliphatic C-H (-CH ₂ -)
2871	ν_S aliphatic C-H (-CH ₃)
2855	ν_S aliphatic C-H (-CH ₂ -)
860-780	collective vibrations of polysubstituted benzene rings

2288, 2275, and 2240 cm⁻¹ are due to C-D stretching modes of the aromatic ring, whereas bands at 2212 and 2120 cm⁻¹ are assigned to the stretching modes of the C-D₃ group.^{43,44} Upon toluene- d_8 adsorption a shoulder at 2965 cm⁻¹ is also formed (Figure 4A). This band could be indicative of interactions of the methyl C-H bonds with guest molecules (see below the ss-NMR part): when the methyl groups interact with the toluene- d_8 molecule, a small red-shift of the 2972 cm⁻¹ frequency is found.

Toluene- d_8 bands are also found in the low-frequency region between 900 and 650 cm⁻¹ where the bending modes of both aromatic and aliphatic C-D groups⁴⁴ fall (Figure 4B, curves b-j): the signal at 880 cm⁻¹ is due to scissoring modes of the aromatic C-D group, and the band at 869 cm⁻¹ is due to asymmetric rocking modes. In addition, also the band at 869 cm⁻¹ is only seen in spectra of liquid phase toluene- d_8 .

The bands at 838 and 819 cm⁻¹, assigned to bending modes associated with the C-D aromatic group,⁴⁴ formed upon toluene- d_8 adsorption, are found in a position very similar to those of the molecule in liquid phase (see Figure S1), and this suggests that even at low pressure the confinement effect due to the porous structure leads to the formation of a metastable liquid-like state.

When the toluene pressure is increased (Figure 4A, curves c-j), all bands related to the adsorbed toluene increase in intensity. In particular, the bands at 2288 and 2135 cm⁻¹ become more visible from 2 and 10 mbar of toluene- d_8 (spectra "d" and "f" of Figure 4A). Because these bands are

typical of the toluene- d_8 in the gas phase (see Figure S1), this suggests that the gas phase is simultaneously present with the adsorbed (liquid-like) phase.

In Figure 5, selected spectra collected after toluene- d_8 adsorption on all synthesized mPAF-1/*n* materials are reported.

The admission of toluene- d_8 on mPAF samples prepared with TPM/FDA molar ratio of 1/9 and 1/30 leads to the formation of the same IR bands already discussed for the mPAF-1/16 sample (Figure 5). Nevertheless, it is worth noting that at high coverage the relative intensity between the signal centered at 2288 cm⁻¹ and the one at 2275 cm⁻¹, associated

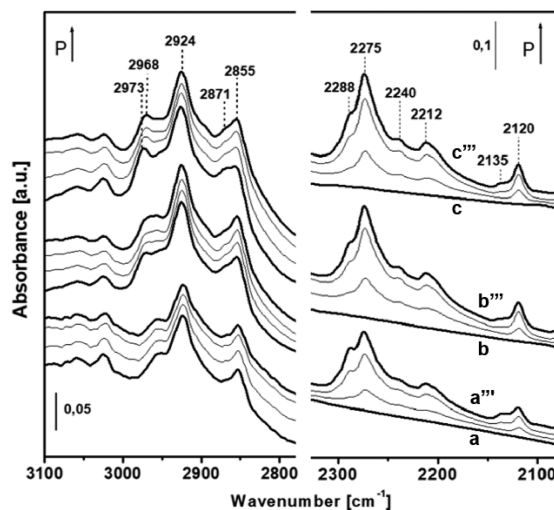


Figure 5. Comparison of selected IR spectra of toluene- d_8 adsorbed on mPAF-1/9 (a), mPAF-1/16 (b), and mPAF-1/30 (c). Spectra (a), (b), and (c) were recorded at rt after outgassing the samples at room temperature for 2 h. Spectra (a'), (b'), and (c') were recorded upon contact with 1.1 mbar, (a''), (b''), and (c'') with 15.1 mbar and (a'''), (b'''), and (c''') with 27.1 mbar of toluene.

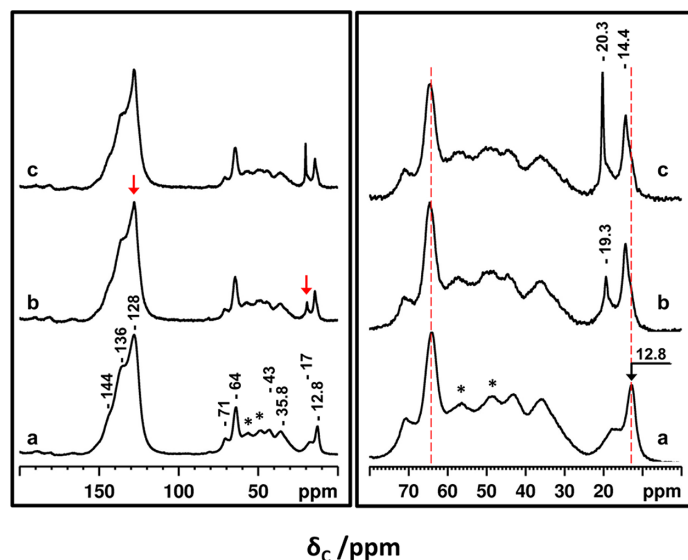


Figure 6. ^{13}C CPMAS NMR spectra of mPAF-1/16 before (a) and after adsorption of 19 mbar of toluene deuterated (b) and adsorption of 19 mbar of nondeuterated toluene (c). Right panel shows the zoomed version of the spectra recorded using a cross-polarization contact time of 10 ms and a MAS rate of 10 kHz. The red arrow marks highlight ^{13}C resonances from adsorbed toluene- d_8 molecules. Spinning sidebands are marked with asterisks.

with asymmetric and symmetric C–D stretching vibrations of the aromatic rings, decreases when TPM/FDA ratio increases (Figure 5, spectra a", b", and c"). A similar behavior is observed for the bands centered at 2135 and 2120 cm^{-1} associated with C–D₃ symmetric and asymmetric stretching vibrations. This is especially evident comparing from the mPAF-1/9 to the mPAF-1/16 sample. By comparison of the FT-IR spectra of liquid and gas-phase toluene- d_8 (see Figure S1), it is evident that a higher amount of liquid-like toluene is formed in samples with high TPM/FDA molar ratio as a result of the progressive decrease of the micropore volume and, in parallel, to the increase of the mesopore volume (*vide supra*).

It is worth noting that the intensity of the bands at 2973 and 2871 cm^{-1} , associated with $-\text{CH}_3$ asymmetric and symmetric stretching vibrations, is very low in the spectra of the mPAF-1/9, and this is likely due to the fact that a smaller FDA quantity during the synthesis limits the formation of side reaction products (i.e., lower number of methyl groups in the material framework). However, when a higher number of methyl groups are present (mPAF-1/16 and mPAF-1/30 samples), upon admission of 27.1 mbar of toluene- d_8 a low-frequency shift of 7–5 cm^{-1} is clearly observed. The small magnitude of the observed red-shift is a possible indication of the presence of weak host–guest interactions between the material's methyl groups and the toluene- d_8 molecules. Similar results were found in the context of an infrared study of the interactions between methane and porous aromatic framework (PAF) materials.⁴⁵

Monitoring the Adsorption of BTX on mPAF-1/*n* Materials by ss-NMR Spectroscopy. Structural characterization of the polymeric material was performed by using solid-state NMR spectroscopy. The ^{13}C CPMAS NMR spectrum of the mPAF-1/16 recorded by using a cross-polarization contact time of 10 ms and a MAS rate of 10 kHz is shown in Figure 6a (left panel). Aromatic ^{13}C peaks in the range 120–150 ppm

are assigned as following: 128 ppm for the aromatic C–H, 136 ppm for the methylene-substituted aromatic carbons, and 144 ppm for the carbons linked to the central quaternary carbon atom.³¹ The resonances at 64 and 35.8 ppm are assigned to the quaternary carbon atom in the tetrahedral building units and the CH_2 (benzylic carbon) bringing linker, respectively. Additional resonances at 43 and 71 ppm can be assigned to CH_2Cl and CH_2OH groups, respectively, substituted on the aromatic rings.³¹ Moreover, peaks at 17 and 12.8 ppm are also observed due to a CH_3 group corresponding to the methylation and/or ethylation of aromatic rings.⁴⁶ As the amount of FDA is very high (16 equivalence to TPM) in the reaction medium, methylation and/or ethylation of the aromatic ring with acetal are expected. Qualitative analysis based on ^{13}C CPMAS NMR data revealed the amounts of side chain groups ($-\text{CH}_2\text{Cl}/-\text{CH}_2\text{OH}/-\text{CH}_3/-\text{CH}_2\text{CH}_3$) to be ~50% of the amount of benzylic carbons. In fact, this estimation confirms the total methylene and methyl carbon content corresponds to at least one side chain per aromatic ring in the mPAF-1/16 material.

To obtain deeper insights into the nature of toluene adsorption and its interaction with the polymeric framework, several ^{13}C and ^1H MAS NMR spectra were recorded. ^{13}C CPMAS NMR spectra of toluene (nondeuterated or deuterated, toluene- d_8) adsorbed on mPAF-1/16 are shown in Figure 6. When nondeuterated toluene was adsorbed, narrow resonances due to physisorbed molecules appeared in the ^{13}C CPMAS NMR spectrum (Figure 6c) besides the resonances due to the bare material (Figure 6a). Similar resonances were also detected (Figure 6b) when toluene- d_8 was adsorbed, albeit in lower intensity.

The basis of a CPMAS NMR experiment is the through-space magnetization transfer between ^1H and ^{13}C via heteronuclear dipolar couplings. Consequently, ^{13}C species in close proximities to proton nuclei are visible in the CPMAS

spectra. In general, short cross-polarization contact time (CT) favors the detection of rigid species while at longer CT mobile units are detected.⁴⁷ However, deuterated molecule such as toluene-*d*₈ will not exhibit any signal (as there are no protons bonded to the carbon atoms) in the ¹³C CPMAS spectrum, unless they are in very close proximity to host polymeric framework protons. The presence of resonances due to methyl as well as aromatic carbons associated with toluene-*d*₈ detected in the ¹³C CPMAS spectrum (highlighted by arrows in Figure 6b, left panel) is a strong suggestion of the latter situation.

When nondeuterated toluene was adsorbed, ¹³C resonances due to methyl carbons of the toluene molecules appeared as a sharp peak at 20.3 ppm (Figure 6c, right panel). On the contrary, this resonance was upfield shifted to 19.3 ppm when toluene-*d*₈ was adsorbed (Figure 6b, right panel). This shift can be attributed to the fact that methyl carbons in toluene-*d*₈ are bonded to deuterium nuclei instead of protons as in the former case. Furthermore, it is worth noting that the resonances due to CH₃ carbons associated with HCP framework that appeared at 12.8 ppm as a broad peak were downfield shifted to 14.4 ppm (also narrower) after toluene adsorption (Figure 6, right panel).

As a matter of fact, the detection of toluene-*d*₈ carbons in a ¹³C CPMAS experiment provides a very strong indication of host–guest interactions. CPMAS experiment is a powerful tool to govern the ¹³C and ¹H through-space proximities as its magnitude is determined by the distance between two nuclei. A short cross-polarization contact time of 2 ms in the CPMAS experiment has also detected signals from guest toluene-*d*₈ carbons (data not shown for the sake of brevity), confirming the intimacy between toluene and polymeric framework. Specific interactions between host polymeric matrix and guest toluene molecules are essential to drive the molecular recognized adsorption process.

The sharp resonances associated with the guest toluene molecules (nondeuterated) in the ¹³C CPMAS NMR spectra also confirm the fact that they are in a highly mobile environment. There are two factors that contribute to the mobile nature of guest molecules, first and foremost being methyl rotation and aromatic ring flipping for adsorbed molecules. On the other hand, as also suggested by IR data, once adsorbed on porous polymers toluene experiences a liquid-like state (*vide supra*). Such observations have been earlier reported for volatile organic compounds adsorbed on porous systems such as zeolites.⁶ Likewise, the ¹³C CPMAS NMR spectra also confirm the adsorbed state of guest toluene molecules (deuterated, toluene-*d*₈) by giving rise to their ¹³C signals. Therefore, it can be suggested that guest molecules exhibit an adsorbed and liquid-like states in the micropores as well as mesopores of the HCP material. On the other hand, there exists equilibrium between the bound state and free state for guest molecules, with fast exchange between them at room temperature. The physical state of the guest molecules is directly reliant on the size of the confined space they are occupying in this hierarchically porous material.

For further evaluation of the adsorption capacity of HCP material, other volatile organic compounds such as benzene, xylene, and *n*-hexane were adsorbed (see Figure S2). The ¹³C CPMAS NMR spectra of benzene and xylene adsorbed material showed a similar pattern as that of toluene, while the *n*-hexane adsorption behavior appeared different. High amounts of benzene, toluene, and xylene uptakes were detected during the adsorption process and are reflected in

their corresponding ¹³C NMR spectra. In addition, the spectral pattern around the chemical shift ranges of methyl and methylene groups (10–20 ppm) appeared identical, suggesting the similar environments for those framework carbons upon adsorption.

As has been pointed out earlier, upon adsorption of benzene, toluene, and xylene (BTX), resonances due to CH₃ carbons associated with HCP framework were downfield shifted to 14.4 ppm. Interestingly, methylene carbons that appeared at 17 ppm in the pristine HCP sample were also downfield shifted to 19.2 ppm after benzene adsorption. Similar features could also be present in toluene and xylene adsorbed samples. However, they are masked by the resonances due to CH₃ groups from the guest molecules. As the hyper-cross-linked porous aromatic framework-based systems are considered as breathing solids, the strong modifications observed in these materials can be related to the swelling of the organic framework during the BTX adsorption. Importantly, *n*-hexane was not adsorbed much in these polyaromatic framework-based materials, as indicated by the fact that no downfield shifts for the resonances of methyl and methylene carbons of the framework were detected in *n*-hexane adsorbed material. This result should be associated with the aliphatic nature of the molecule, not allowing specific interactions with the surface. Moreover, the architecture of the pollutant molecules (and the possible orientations) can also influence the uptake capacity. Solubility parameters should also be considered. Solubility parameters describe the attractive strengths between molecules of a certain material, and the similarity of the solubility parameters of two different substances indicates that the two substances are mutually soluble. From the point of view of the three-dimensional Hansen solubility parameters the larger swelling capacity of toluene is a direct consequence of the chemical similarity with HCP polymers, made of alkylated aromatic rings.⁴⁸ Dispersion interactions (15.3 MPa^{1/2}) are the only nonzero contributions to the overall solubility parameter of *n*-hexane, whereas the more polarizable toluene displays not only stronger dispersion interactions (18 MPa^{1/2}) but also non-negligible permanent dipolar attractions (1.4 MPa^{1/2}) and other association bonds (e.g., hydrogen bond, permanent dipole–induced dipole, etc.) (2 MPa^{1/2}).

To summarize this section, the environments of methyl and methylene carbons (around 10–20 ppm) of the framework backbone were affected due to the adsorption of BTX, which gives a first insight into the local packing of guest molecules in the confined spaces. On the other hand, π – π packing of aromatic moieties cannot be excluded here, although direct evidence for its existence was absent due to lack of resolution in the aromatic region of the ¹³C CPMAS NMR spectra. Similarly, the impact of guest molecules on bridging benzylic units and other methylene carbons (–CH₂Cl and –CH₂OH) of the polymer backbone is negligible as their ¹³C chemical shifts (between 35 and 71 ppm) were not influenced after BTX adsorption. Furthermore, bridging benzylic carbons could be inaccessible for the guest molecules due to the spatial constraints generated from the hyper cross-linking of framework aromatic units.

¹H chemical shifts are enormously sensitive to confined environments and can be used to probe local intermolecular interactions. Figure 7 inset shows the ¹H MAS NMR spectra of mPAF-1/16 material, before and after toluene adsorption, recorded with a MAS rate of 15 kHz. Aliphatic and aromatic protons can be easily distinguished from the spectrum;

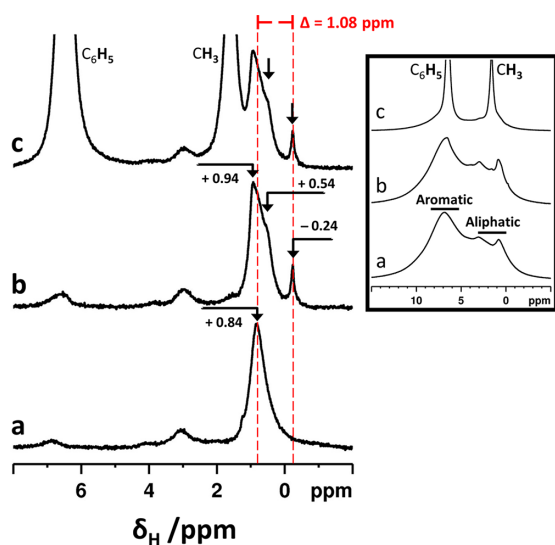


Figure 7. ^1H spin-echo NMR spectra of mPAF-1/16 before (a) and after admission of 19 mbar of toluene deuterated (b) and admission of 19 mbar of nondeuterated toluene (c). Inset shows the ^1H MAS NMR spectra recorded by using a MAS rate of 15 kHz.

however, the poor resolution, partially due to proton background signal of the probe and due to strongly dipolar coupled proton spins, restricted the extraction of further details. To suppress such effects, the rotor synchronized spin-echo pulse sequence was employed to record the ^1H NMR spectra with a τ delay time of 4 ms. The delay time was chosen as an optimized compromise between the signal decay owing to relaxation and the resolution gain owing to longer delay times.⁴⁹

The ^1H echo spectrum in Figure 7a exhibits a far superior baseline, where the resonances of the relevant host polymer protons are well resolved. A broad ^1H peak, centered at 0.84 ppm, for host HCP material is clearly detected and is due to the framework methyl groups. Upon adsorption of guest toluene molecules (either nondeuterated or deuterated), the above resonance is split into two resonances at 0.94 and 0.54 ppm (Figure 7, curves b and c). In addition, a new sharper signal at -0.24 ppm appeared as well for toluene adsorbed on the polymeric material. ^1H resonances from guest toluene molecules were clearly detected in the nondeuterated toluene adsorbed system as evident in Figure 7c (marked as C_6H_5 and CH_3); however, respective signals were absent in the toluene- d_8 system (Figure 7b) due to the absence of protons. Consequently, the new upfield resonances appeared at 0.54 and -0.24 ppm can be safely attributed to host polymeric framework.

Guest-induced upfield shifts of proton resonances on the host sites as a result of ring-current effects of aromatic rings of guest species are well reported in the literature.⁴⁷ ^1H NMR spectroscopy is an excellent tool for studying the geometrical features as well as molecular interactions and is capable of detecting the weaker noncovalent CH/π interactions.⁵⁰ Provided that a CH/π interaction takes place, the chemical shifts of the protons involved in the interaction should be modified according to their spatial positions in the field created by the aromatic ring current, resulting in an upfield shift for the

^1H NMR signals of the interacting CH group. There are some articles recently appeared in the literature that highlighted the application of ^1H MAS NMR spectroscopy in detecting CH/π interactions in solid-state samples.⁵¹

To test the hypothesis of the guest induced upfield shifts of proton resonances on the host sites, ^1H NMR spectra were recorded on benzene, xylene, and *n*-hexane adsorbed systems (see Figure S3). Interestingly, the same spectral features that were detected for toluene adsorbed system were detected in benzene and xylene systems. More importantly, no such features were detected in *n*-hexane adsorbed material, thus confirming that they are due to aromatic moieties. From this study, it is evident that hyper-cross-linked porous polymer system favors the adsorption of aromatic ring-based compounds (BTX) and that the guest–host CH/π interactions could play a relevant role.

On the other hand, BET measurements (Table 1 and Figure 2) have revealed the hierarchical nature (micropores as well as the wide range of mesopores) of the pores present in mPAF-1/16 material. When toluene molecules are adsorbed, they may occupy both micropores and mesopores. Moreover, FT-IR data confirmed the liquid-like state for the guest molecules in these materials, indicating the adsorption in mesopores. Therefore, strong confinement effects may be present when adsorption occurs in the micropores due to restricted space availability. In essence, the methyl protons of the polymer framework units experience strong shielding effects due to a relative orientation that places these protons above or below the aromatic system of the guest molecules in hierarchical pores, hence the upfield shift of the ^1H resonances. A difference of $\delta = 1.08$ ppm ($0.84 + 0.24$) for selective methyl protons of the framework units is due to strong intermolecular CH/π interactions probably occurring in micropores. More importantly, not all host methyl groups display significant resonance changes due to the presence of the guest toluene molecules. For example, the ^1H resonance at 0.54 ppm might be due to host methyl protons that are present in the mesopores where liquefaction of the guest molecule occurs. Moreover, a significant portion of host methyl protons were not influenced directly by the presence of guest species, however, experiencing a different environment than in the parent material (probably due to the expansion of the polymer framework upon guest adsorption) as evidenced by their chemical shift at 0.94 ppm.

The apparent lack of a D/H isotope effect between deuterated and nondeuterated toluene validates the directional nature as well as the unique geometry and environment of CH/π interactions. The proton chemical shifts detected in ^1H NMR and the ^1H – ^{13}C through-space couplings (especially in toluene- d_8 adsorbed system) detected in the CPMAS NMR experiments present a compelling case for CH/π interactions. The ^{13}C CPMAS experimental results on the toluene- d_8 adsorbed system uphold the existence of very close proximities between guest toluene aromatic rings, and a significant fraction of framework methyl groups as well elucidate the power of guest–host noncovalent interactions in influencing the strong and selective adsorption of aromatic molecules.

Computational Studies. To provide some atomistic insights into the host–guest molecular interactions and support the interpretation of the spectroscopic characterization, *ab initio* and MD calculations were performed on a small model of the mPAF-1/*n* hyper-cross-linked polymer, adding the necessary toluene or *n*-hexane molecules. The

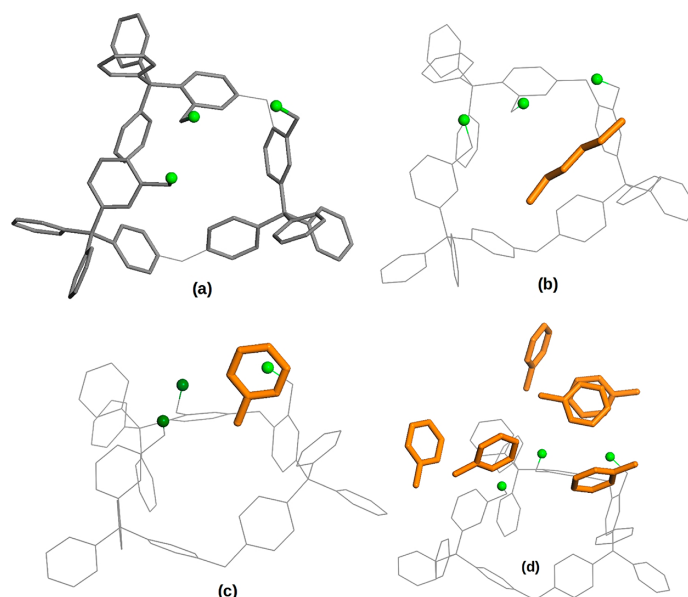


Figure 8. (a–c) DFT optimized structures of (a) mPAF-1/*n* polymer model, (b) polymer model with *n*-hexane guest molecule, and (c) polymer model with toluene guest molecule: guest molecules in orange, methyl carbon atoms in green; in (c) the carbon atom involved in the CH/ π host-guest interaction (upshifted in ^{13}C NMR) is in light green, and the adjacent carbons (downshifted or left unaltered) are in dark green. (d) An example of the clusters extracted from the MD snapshots, including the polymer model and the closest toluene molecules.

mPAF-1/*n* model is depicted in Figure 8a: it comprises three tetraphenylmethane units linked to each other by a methylene bridge; according to the analysis discussed above, each unit bears one $-\text{CH}_2\text{CH}_3$ side chain, bonded to one of the phenyl rings.

First, one adsorbate molecule (either toluene or *n*-hexane) was added to the mPAF-1/*n* model, and the geometries were optimized at the DFT level, freezing the coordinates of the phenyl rings but allowing the side chains to adjust, with the results shown in Figure 8b,c. With *n*-hexane, loose interactions largely dominated by the long-range dispersive forces were computed; on the other hand, toluene was found to interact strongly with one of the dangling alkyl chains, with the methyl group directed toward the aromatic ring, at a relatively short distance (i.e., 3.93 Å between the methyl carbon and the phenyl ring geometric center). Such CH/ π interaction was proposed above on the basis of solid-state ^1H and ^{13}C NMR spectra.

To assess the reliability of these structures further, we computed the ^{13}C NMR isotropic shielding of the mPAF-1/*n* model before and after the addition of the adsorbate: with *n*-hexane no significant changes were found, as expected for such a weakly interacting system, and in agreement with the experimental data. With toluene, however, the addition of one molecule caused a 2.1 ppm upfield shift in the interacting carbon and a 2.3 ppm downfield shift in the adjacent carbon, while the third one was barely affected (Figure 8c). This is in contrast with the experiments, where only an 1.6 ppm downfield shift (14.4–12.8 ppm, see Figure 6 for details) was measured for the $-\text{CH}_3$ carbons after toluene adsorption.

The disagreement might be due to the high number of guest molecules interacting with the polymeric framework in the experimental conditions, so that the shielding effect of toluene on the close-interacting methyl groups could be overwhelmed

by the deshielding due to the surrounding molecules. Thus, the calculations should be repeated with a higher number of toluene molecules added to the model: however, optimizing a structure with many molecules interacting among them and with the framework could be biased by the arbitrary choice of the initial geometries. To avoid this problem and obtain a sound average value of the adsorbate effects, we resorted to a mixed MD/*ab initio* procedure: first, a MD run was performed on a cubic periodic box containing the polymer model and 3000 toluene molecules, at constant temperature and pressure (300 K, 1 atm). After equilibration, 10 snapshots, separated by 5 ps, were selected from the MD file, and from each frame a model was extracted, including the polymer and the two toluene molecules closest to each side chain methyl group. Then ^{13}C isotropic shieldings were computed at the DFT-GIAO level and compared to the values of the bare polymer model.

The resulting shifts after toluene addition are reported in Table S1 for the three methyl carbons in all the snapshots extracted from MD: the average effect is a downfield shift of 1.7 ppm, in good agreement with the experimental results. Interestingly, in some cases one or two carbon atoms experienced an upfield shift, depending on the positions of the toluene molecules in the various snapshots, but all carbons were on average downshifted (i.e., deshielded by the global effect of the adsorbate).

The modeling, then, reproduces the NMR findings provided a sufficient number of adsorbate molecules are considered.

CONCLUSIONS

Hyper-cross-linked polymers, with increasing cross-linking agent/monomer ratios, specifically 30:1, 16:1, and 9:1, were tested for adsorption of aromatic volatile compounds. The

toluene molecule was chosen as representative of the BTX family.

Data obtained via volumetric analysis showed significantly higher toluene adsorption capacities associated with mPAF materials (up to 154 Q [%]) with respect to those displayed by a well-established silica-based adsorbent, namely HSZ-Y zeolite (21 Q [%]). The presence of a swelling effect of the polymeric network was evidenced by comparing data obtained by volumetric adsorption of both N₂ and toluene. The effect is reduced for more cross-linked polymeric frameworks, namely for mPAF-1/30, thus suggesting an inverse relationship between the cross-linking degree of the porous network and the swelling effects experienced by it.

FT-IR and ss-NMR spectroscopy were used to investigate the interactions occurring between the host material and guest molecules. Toluene-*d*₈ allowed a more detailed investigation of the adsorption process because of the absence of IR and NMR signals overlapping between the material and the pollutant. FT-IR spectra, collected upon adsorption of small doses of toluene-*d*₈ (up to 2 mbar) on mPAF-1/16, already showed the appearance of the spectral profile of liquid-like phase toluene-*d*₈. Host-guest interactions, specifically due to weak forces between -CH₃ groups of the HCPs framework and the toluene-*d*₈ molecules, become more detectable by increasing toluene-*d*₈ pressure above 14 mbar.

ss-NMR spectra were crucial to determine with higher degree of precision the nature of the noncovalent interactions involved in toluene-*d*₈/toluene/benzene/xylene/mPAF-16 systems. In particular, from ¹³C CPMAS NMR spectra, evidence was found of close proximity, during the adsorption process, between the guest molecules and framework methyl groups. In addition, ¹H spin-echo NMR spectra allowed to gain more information about the spatial position and orientation between the framework -CH₃ groups and the toluene-*d*₈/toluene/benzene/xylene molecule. BTX-induced upfield shifts of proton resonances on the host sites, as a result of ring-current effects of aromatic rings of guest species, indicate the presence of methyl protons pointing toward aromatic rings, hence suggesting the presence of CH/π interactions.

Ab initio and MD calculations were used to gain additional insights with regard to guest molecules interacting with the material's framework. Good agreement was found between the calculations and the experimental data, thus in support of the picture of CH/π interactions present between alkyl groups of the material's framework and aromatic guest molecules, upon adsorption. In conclusion, the organic and swellable nature of the aromatic framework displayed by HCPs and their high surface area and pore volume are key properties to significantly improve aromatic pollutants adsorption capacities.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsapm.9b01000>.

Experimental details; (1) IR spectra of *d*₈-toluene in gas and liquid phase, (2) ¹³C CPMAS NMR spectra (expansion of the aliphatic region) of mPAF-1/16 before and after adsorption of BTX, (3) ¹H spin-echo NMR spectra of mPAF-1/16 before and after adsorption of BTX, and (4) table reporting the variation of the methyl ¹³C isotropic shielding (ppm) with respect to the

bare polymer model, after addition of some toluene molecules (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail chiara.bisio@uniupo.it.

ORCID

G. Golemme: 0000-0001-5675-4416

C. Bisio: 0000-0002-7591-5104

M. Cossi: 0000-0002-1817-9008

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge Dr. D. F. Olivas Oliveira for the help in performing the IR and volumetric adsorption measurements. Alessandro Melicchio and Giovanni Golemme thank the European Regional Development Fund for supporting part of this work through Project PONA3_00370 MaTeRiA.

■ REFERENCES

- (1) Xu, S.; Luo, Y.; Tan, B. Recent Development of Hypercrosslinked Microporous Organic Polymers. *Macromol. Rapid Commun.* **2013**, *34* (6), 471–484.
- (2) Dawson, R.; Stevens, L. A.; Drage, T. C.; Snape, C. E.; Smith, M. W.; Adams, D. J.; Cooper, A. I. Impact of Water Co-adsorption for Carbon Dioxide Capture in Microporous Polymer Sorbents. *J. Am. Chem. Soc.* **2012**, *134* (11), 10741–10744.
- (3) Vignola, R.; Bagatin, R.; De Folly D'Auris, A.; Flego, C.; Nalli, M.; Ghisletti, D.; Millini, R.; Sisto, R. Zeolites in a permeable reactive barrier (PRB): One year of field experience in a refinery groundwater. Part 1: The performances. *Chem. Eng. J.* **2011**, *178*, 204–209.
- (4) Vignola, R.; Bagatin, R.; De Folly D'Auris, A.; Flego, C.; Nalli, M.; Ghisletti, D.; Millini, R.; Sisto, R. Zeolites in a permeable reactive barrier (PRB): One-year of field experience in a refinery groundwater. Part 2: Zeolite characterization. *Chem. Eng. J.* **2011**, *178*, 204–209.
- (5) Ahmaruzzaman, M. Adsorption of Phenolic Compounds on Low-Cost Adsorbents: A review. *Adv. Colloid Interface Sci.* **2008**, *143*, 48–67.
- (6) Sacchetto, V.; Olivas Olivera, D. F.; Paul, G.; Gatti, G.; Braschi, I.; Marchese, L.; Bisio, C. On the Adsorption of Gaseous Mixtures of Hydrocarbons on High Silica Zeolites. *J. Phys. Chem. C* **2017**, *121* (11), 6081–6089.
- (7) Sacchetto, V.; Gatti, G.; Paul, G.; Braschi, I.; Berlier, G.; Cossi, M.; Marchese, L.; Bagatin, R.; Bisio, C. The Interactions of Methyl Tert-butyl Ether on High Silica Zeolites: A Combined Experimental and Computational Study. *Phys. Chem. Chem. Phys.* **2013**, *15* (32), 13275.
- (8) Huang, J.; Turner, S. R. Hypercrosslinked Polymers: A Review. *Polym. Rev.* **2018**, *58* (1), 1–41.
- (9) Dawson, R.; Stöckel, E.; Holst, J. R.; Adams, D. J.; Cooper, A. I. Microporous Organic Polymers for Carbon Dioxide Capture. *Energy Environ. Sci.* **2011**, *4* (10), 4239–4245.
- (10) Martin, C. F.; Stöckel, E.; Clowes, R.; Adams, D. J.; Cooper, A. I.; Pis, J. J.; Rubiera, F.; Pevida, C. Hypercrosslinked Organic Polymer Networks as Potential Adsorbents for Pre-Combustion CO₂ Capture. *J. Mater. Chem.* **2011**, *21* (14), 5475.
- (11) Yao, S.; Yang, X.; Yu, M.; Zhang, Y.; Jiang, J. X. High Surface Area Hypercrosslinked Microporous Organic Polymer Networks Based on Tetraphenylethylene for CO₂ capture. *J. Mater. Chem. A* **2014**, *2* (21), 8054.
- (12) Puthiaraj, P.; Ahn, W. S. CO₂ Capture by Porous Hyper-Cross-Linked Aromatic Polymers Using Tetrahedral Precursors. *Ind. Eng. Chem. Res.* **2016**, *55* (29), 7917–7923.
- (13) Sacchetto, V.; Bisio, C.; Olivas Olivera, D. F.; Paul, G.; Gatti, G.; Braschi, I.; Berlier, G.; Cossi, M.; Marchese, L. Interactions of

- Toluene and *n*-Hexane on High Silica Zeolites: An Experimental and Computational Model Study. *J. Phys. Chem. C* **2015**, *119* (44), 24875–24886.
- (14) Sawant, S. Y.; Munusamy, K.; Somani, R. J.; John, M.; Newalkar, B. L.; Bajaj, H. C. Precursor Suitability and Pilot Scale Production of Super Activated Carbon for Greenhouse Gas Adsorption and Fuel Gas Storage. *Chem. Eng. J.* **2017**, *315*, 415–425.
- (15) Liu, S.; Chen, D.; Zheng, J.; Zeng, L.; Jiang, J.; Jiang, R.; Zhu, F.; Shen, Y.; Wu, D.; Ouyang, G. The Sensitive and Selective Adsorption of Aromatic Compounds With Highly Crosslinked Polymer Nanoparticles. *Nanoscale* **2015**, *7* (40), 16943.
- (16) Ju, H.; Zhou, X.; Shi, B.; Kong, X.; Xing, H.; Huang, F. A Pillararene-Based Hydrogel Adsorbent in Aqueous Environments for Organic Micropollutants. *Polym. Chem.* **2019**, *10*, 5821–5828.
- (17) Sherrill, C. D. Energy Component Analysis of π Interactions. *Acc. Chem. Res.* **2013**, *46* (4), 1020–1028.
- (18) Raynal, M.; Ballester, P.; Vidal-Ferran, A.; van Leeuwen, P. W. N. M. Supramolecular catalysis. Part 1: Non-Covalent Interactions as a Tool for Building and Modifying Homogeneous Catalysts. *Chem. Soc. Rev.* **2014**, *43* (5), 1660–1733.
- (19) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, *543*, 637–646.
- (20) Plevin, M. J.; Bryce, D. L.; Boisbouvier, J. Direct Detection of CH/ π Interactions in Proteins. *Nat. Chem.* **2010**, *2* (6), 466–471.
- (21) Desiraju, G.; Steiner, T. *The Weak Hydrogen Bond*; Oxford University Press: 1999.
- (22) Suzuki, S.; Green, P. G.; Bumgarner, R. E.; Dasgupta, S.; Goddard, W. A., III; Blake, G. A. Benzene Forms Hydrogen Bonds with Water. *Science* **1992**, *257* (5072), 942–945.
- (23) Nishio, M.; Umezawa, Y.; Honda, K.; Tsuboyama, S.; Suezawa, H. CH/ π Hydrogen Bonds in Organic and Organometallic Chemistry. *CrystEngComm* **2009**, *11* (9), 1757–1788.
- (24) Zhao, C.; Parrish, R. M.; Smith, M. D.; Pellechia, P. J.; Sherrill, C. D.; Shimizu, K. D. Do Deuteriums Form Stronger CH– π Interactions? *J. Am. Chem. Soc.* **2012**, *134* (35), 14306–14309.
- (25) Suezawa, H.; Yoshida, T.; Hirota, M.; Takahashi, H.; Umezawa, Y.; Honda, K.; Tsuboyama, S.; Nishio, M. The CH– π Interaction as an Important Factor in the Crystal Packing and in Determining the Structure of Clathrates. *J. Chem. Soc.* **2001**, *2* (11), 2053–2058.
- (26) Jie, K.; Zhou, Y.; Li, E.; Huang, H. Nonporous Adaptive Crystals of Pillararenes. *Acc. Chem. Res.* **2018**, *51* (9), 2064–2072.
- (27) Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M. S.; Chakrabarti, P. CH– π Hydrogen Bonds in Biological Macromolecules. *Phys. Chem. Chem. Phys.* **2014**, *16* (25), 12648–12683.
- (28) Nishio, M. The CH/ π Hydrogen Bond in Chemistry. Conformation, Supramolecules, Optical Resolution and Interactions Involving Carbohydrates. *Phys. Chem. Chem. Phys.* **2011**, *13* (31), 13873–13900.
- (29) Tsuzuki, S.; Honda, K.; Uchimar, T.; Mikami, M.; Fujii, A. Magnitude and Directionality of the Interaction Energy of the Aliphatic CH/ π Interaction: Significant Difference from Hydrogen Bond. *J. Phys. Chem. A* **2006**, *110* (33), 10163–10168.
- (30) Takahashi, O.; Kohno, Y.; Nishio, M. Relevance of Weak Hydrogen Bonds in the Conformation of Organic Compounds and Bioconjugates: Evidence from Recent Experimental Data and High-Level ab Initio MO Calculations. *Chem. Rev.* **2010**, *110* (10), 6049–6076.
- (31) Errahali, M.; Gatti, G.; Tei, L.; Paul, G.; Rolla, G. A.; Canti, L.; Fraccarollo, A.; Cossi, M.; Comotti, A.; Sozzani, P.; Marchese, L. Microporous Hyper-Cross-Linked Aromatic Polymers Designed for Methane and Carbon Dioxide Adsorption. *J. Phys. Chem. C* **2014**, *118* (49), 28699–28710.
- (32) Navalon, S.; Alvaro, M.; Garcia, H. Highly Dealuminated Y Zeolite as Efficient Adsorbent for the Hydrophobic Fraction from Wastewater Treatment Plants Effluents. *J. Hazard. Mater.* **2009**, *166* (1), 553–560.
- (33) Paul, G.; Bisio, C.; Braschi, I.; Cossi, M.; Gatti, G.; Gianotti, E.; Marchese, L. Combined Solid-State NMR, FT-IR and Computational Studies on Layered and Porous Materials. *Chem. Soc. Rev.* **2018**, *47* (15), 5684–5739.
- (34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasingh, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, Liang, G. W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Zhrussell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision A.03; Gaussian, Inc.: Wallingford, CT, 2016.
- (35) Becke, A. D. Density Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (36) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007–23.
- (37) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H–Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.
- (38) Wolinski, K.; Hinton, J. F.; Pulay, P. Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations. *J. Am. Chem. Soc.* **1990**, *112* (23), 8251–8260.
- (39) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
- (40) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. Uff, A Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035.
- (41) Thommes, M.; Kaneko, K.; Neimark, A. V.; Olivier, J. P.; Rodriguez-Reinoso, F.; Rouquerol, J.; Sing, K. S. W. Physisorption of Gases, with Special Reference to the Evaluation of Surface Area and Pore Size Distribution (IUPAC Technical Report). *Pure Appl. Chem.* **2015**, *87* (9–10), 1051–1069.
- (42) Coates, J. Interpretation of Infrared Spectra, A Practical Approach. *Encyclopedia of Analytical Chemistry*; John Wiley & Sons Ltd.: 2006.
- (43) Knaanie, R.; Šebek, J.; Tsuge, M.; Myllys, N.; Khriachtchev, L.; Rasänen, M.; Albee, B.; Potma, E. O.; Gerber, R. B. Infrared Spectrum of Toluene: Comparison of Anharmonic Isolated-Molecule Calculations and Experiments in Liquid Phase and in a Ne Matrix. *J. Phys. Chem. A* **2016**, *120* (19), 3380–3389.
- (44) Fuson, N.; Garrigou-Lagrange, C.; Josien, M. L. Spectre infrarouge et attribution des vibrations des toluènes C₆H₅CH₃, C₆H₅CD₃ et C₆D₅CD₃. *Spectrochim. Acta* **1960**, *16* (1–2), 106–127.
- (45) Errahali, M.; Gatti, G.; Tei, L.; Canti, L.; Fraccarollo, A.; Cossi, M.; Marchese, L. Understanding Methane Adsorption in Porous Aromatic Frameworks: An FTIR, Raman, and Theoretical Combined Study. *J. Phys. Chem. C* **2014**, *118* (19), 10053–10060.
- (46) Woodward, R. T.; Stevens, L. A.; Dawson, R.; Vijayaraghavan, M.; Hasell, T.; Silverwood, I. P.; Ewing, A. V.; Ratvijitvech, T.; Exley, J. D.; Chong, S. Y.; Blanc, F.; Adams, D. J.; Kazarian, S. G.; Snape, C. E.; Drage, T. C.; Cooper, A. I. Swellable, Water- and Acid-Tolerant Polymer Sponges for Chemoselective Carbon Dioxide Capture. *J. Am. Chem. Soc.* **2014**, *136* (25), 9028–9035.
- (47) Paul, G.; Steuernagel, S.; Koller, H. Non-Covalent Interactions of a Drug Molecule Encapsulated in a Hybrid Silica Gel. *Chem. Commun.* **2007**, *48*, 5194–5196.
- (48) Grulke, E. A. In *Polymer Handbook*, 4th ed.; Brandrup, J.; Immergut, E. H.; Grulke, E. A.; Abe, A.; Bloch, D. R., Eds.; John Wiley & Sons Inc.: 1999; Vol II, pp 675–709.

(49) Paul, G.; Musso, G. E.; Bottinelli, E.; Cossi, M.; Marchese, L.; Berlier, G. Investigating the Interaction of Water Vapour with Aminopropyl Groups on the Surface of Mesoporous Silica Nanoparticles. *ChemPhysChem* **2017**, *18* (7), 839–849.

(50) Kobayashi, K.; Asakawa, Y.; Kikuchi, Y.; Toi, H.; Aoyama, Y. CH- π Interaction as an Important Driving Force of Host-Guest Complexation in Apolar Organic Media. Binding of Monools and Acetylated Compounds to Resorcinol Cyclic Tetramer as Studied by NMR and Circular Dichroism Spectroscopy. *J. Am. Chem. Soc.* **1993**, *115* (7), 2648–2654.

(51) Brown, S. P. Applications of High-Resolution ^1H Solid-State NMR. *Solid State Nucl. Magn. Reson.* **2012**, *41*, 1–27.

Appendix C

Predicting the Conformation of
Organic Catalysts Grafted on Silica
Surfaces with Different Numbers of
Tethering Chains: The Silicopodality
Concept

Predicting the Conformation of Organic Catalysts Grafted on Silica Surfaces with Different Numbers of Tethering Chains: The Silicopodality Concept

Ivana Miletto, Chiara Ivaldi, Enrica Gianotti, Geo Paul, Fabio Travagin, Giovanni Battista Giovenzana, Alberto Fraccarollo, Davide Marchi, Leonardo Marchese, and Maurizio Cossi*

Cite This: *J. Phys. Chem. C* 2021, 125, 21199–21210

Read Online

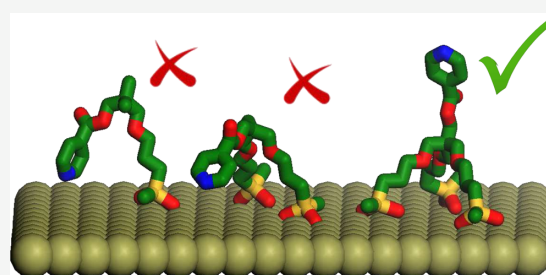
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Hybrid catalysts are attracting much attention, since they combine the versatility and efficiency of homogeneous organic catalysis with the robustness and thermal stability of solid materials, for example, mesoporous silica; in addition, they can be used in cascade reactions, for exploring both organic and inorganic catalysis at the same time. Despite the importance of the organic/inorganic interface in these materials, the effect of the grafting architecture on the final conformation of the organic layer (and hence its reactivity) is still largely unexplored. Here, we investigate a series of organosiloxanes comprising a pyridine ring (the catalyst model) and different numbers of alkylsiloxane chains used to anchor it to the MCM-41 surface. The hybrid interfaces are characterized with X-ray powder diffraction, thermogravimetric analyses, Fourier-transform infrared spectroscopy, nuclear magnetic resonance techniques and are modeled theoretically through molecular dynamics (MD) simulations, to determine the relationship between the number of chains and the average position of the pyridine group; MD simulations also provide some insights about temperature and solvent effects.



INTRODUCTION

Hybrid catalysts, formed by organic moieties anchored onto inorganic matrices, are interesting for their capacity to combine enormous varieties of organic functional groups with a solid support with superior mechanical and thermal stabilities.^{1–11} Such materials offer the advantages of homogeneous catalysis, as high yields, selectivities and possible stereoselectivity, along with those of heterogeneous catalysis, with easily separable products, reduced volume of solvents and so on.^{12–17} Even more interesting is the possibility of using active inorganic surfaces, for instance silica with suitable concentrations of silanol groups, or silica–alumina surfaces with Brønsted and Lewis acid sites, which can in turn act as cocatalysts in cascade reactions.^{18–22}

Despite the great potential of hybrid heterogeneous catalysts, very little is known so far about the preferred conformations of the organic moieties grafted to the surface, and their influence on the catalytic activity. In particular, the active organic groups are likely more available for interactions with the substrates when they spend most of the time as far as possible from the surface, increasing the efficiency, while the tendency to lie down on the surface is expected to have the opposite effect.

Here, we consider an organic catalyst (a pyridine moiety) grafted to the silica surface through alkylsiloxane chains

terminated by $-\text{Si}(\text{OR})_n\text{R}'_{3-n}$ with $n = 1-3$, $\text{R}, \text{R}' =$ methyl and ethyl.^{23–25} The terminal $-\text{OR}$ groups can condensate with surface silanols to form siloxane bridges binding the organosiloxane to amorphous or mesoporous silica or silica–alumina: the number of bridges formed by each chain depends on the number of alkoxy groups (n above) and on the surface silanol concentration.

We investigate how the design of the tethering unit affects the interface conformation, adopting a combined experimental/computational approach, which is particularly suited to deal with hybrid materials.²⁶ Indeed, there are two structural parameters which can be easily varied during the organosiloxane synthesis, illustrated in Figure 1. The first parameter is the number of siloxane bridges formed by each chain: in ref 27, we proposed to call this parameter *silicodactyly* (from dactyls, the Greek word for fingers) and studied its effect on the conformation of the hybrid structures with a combination of

Received: July 9, 2021

Published: September 17, 2021



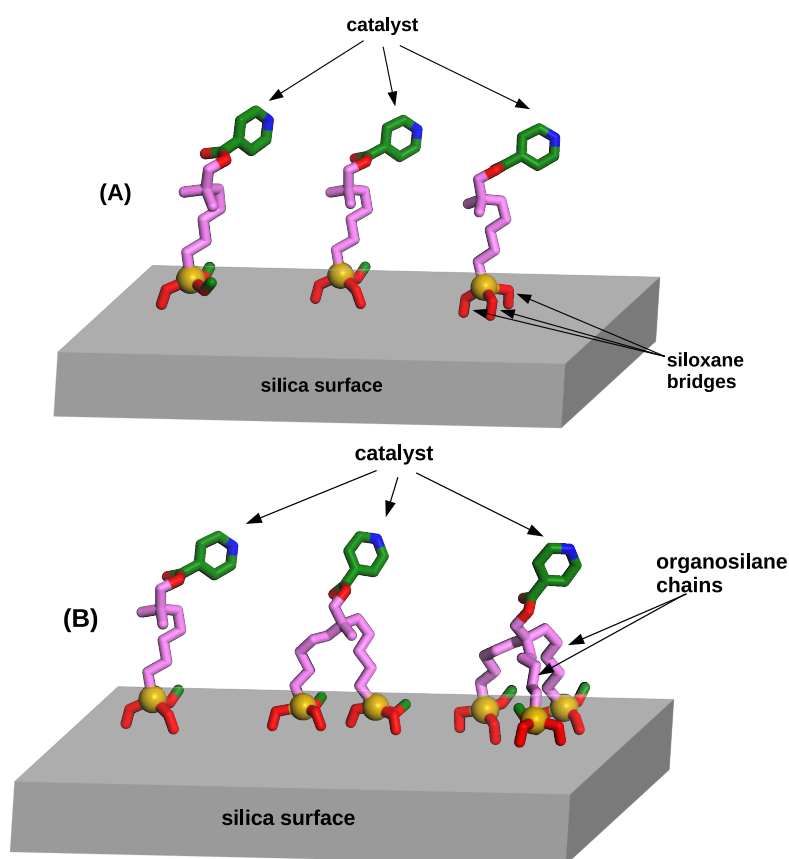


Figure 1. Scheme of the possible tethering architectures: (A), changing the number of siloxane bridges (dactyl); (B), changing the number of siloxane chains (podality).

experimental and computational techniques. The conclusion was that silicodactylity has a very little effect on the conformation of grafted organic chains, which tend to lie close to the silica most of the time, irrespective of the number of siloxane “fingers” used to grab the surface.

The second structural element that can be adjusted is the number of siloxane chains linked to the same organic catalyst, holding the active group on the silica surface. This number will be referred to as *silicopodality*, and in the following we study how it affects the conformation of the organic/silica interface, using a combination of theoretical modeling and experimental characterization.

This study concerns a series of pyridine-substituted derivatives, with different numbers of alkylsiloxane chains, illustrated in Figure 2. Here, the pyridine moiety represents the organic catalyst (e.g., in acylation reactions), and it can also interact with silanols, revealing how close to the surface the active center lies.

Notably, all the tethering chains in the considered systems bear two –OMe residues, along with an unreactive methyl group, so that each chain can form two siloxane bridges with the surface (a structure called didactyl in ref 27). We did not examine other tethering schemes, because the risk of monodactyl grafting is too low, while tri-dactyl structures are statistically unlikely, unless a very hydrophilic silica is used:

anyway, as mentioned above, the effect of silicodactylity on the conformation of the organic/silica interface is very limited.²⁷

Three pyridine derivatives, namely MP-Py, DP-Py, and TP-Py, bearing one, two, or three alkylsiloxane chains, respectively, were synthesized; MP-Py and TP-Py were grafted to MCM-41 ordered mesoporous silica, and characterized by Fourier-transform infrared (FT-IR) spectroscopy, X-ray powder diffraction (XRPD), thermogravimetric analysis (TGA), and solid-state nuclear magnetic resonance (SS-NMR) analyses. In addition, all the hybrid systems (MP-Py-MCM41, DP-Py-MCM41, and TP-Py-MCM41) were modeled theoretically with molecular dynamics (MD) simulations, comparing the results of the simulations with the experimental results.

■ MATERIALS AND METHODS

Synthesis of the Multipodal Pyridine Derivatives. The pyridine derivatives, bearing one to three alkylsiloxane chains for grafting to the silica surface, were prepared following the scheme shown in Figure 3.

The commercially available polyols **1** and **6** were used as the starting materials for the monopodal and dipodal derivatives. Mono and dialylation of these two polyols were accomplished through a classic Williamson synthesis, by treating the polyol with sodium hydride in tetrahydrofuran (THF), followed by reaction with allyl bromide. Separation of the mono (**2**) or

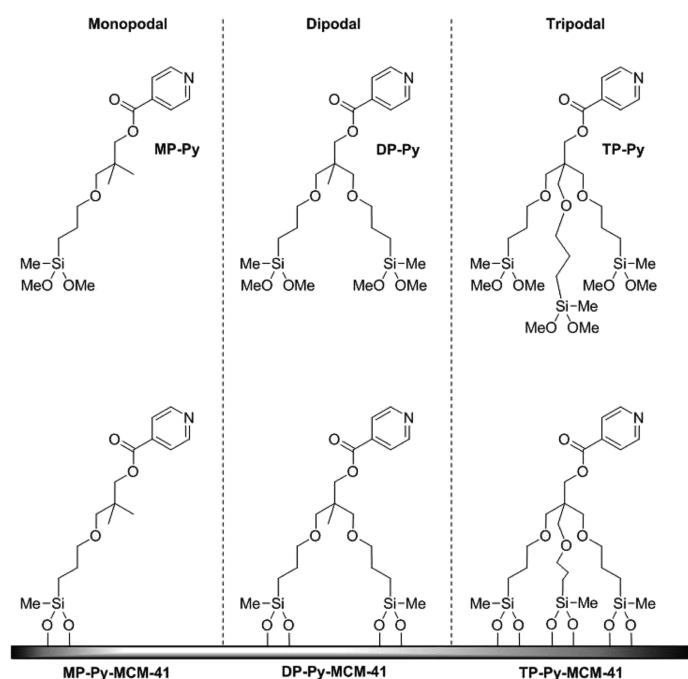


Figure 2. Chemical structure and acronym of the multipodal 4-pyridine derivatives isolated and grafted to the mesoporous silica.

diallyl derivative (7) from various allylated derivatives was achieved by column chromatography.

Compound 7 is the starting material for the preparation of the tripodal derivative. It is commercially available as a mixture of allylated pentaerythritols, enriched in the triallyl derivative (assay 70%), from which the pure pentaerythritol triallyl ether (7) may be obtained by simple displacement chromatography.²⁸ The pyridine ring was introduced in the structure through an esterification step, taking advantage of the purposefully residual alcoholic group in each of the allylated derivatives (2, 7, and 9).

Isonicotinic acid was activated by conversion into the corresponding acid chloride 3, and isolated as the hygroscopic crystalline hydrochloride. Reaction of 3 with the allylated derivatives 2, 7, and 9 provided the corresponding esters 4, 8, and 10, respectively. Then, the mono-, di-, and tripodal esters were hydrosilylated with methyltrimethoxysilane (5). The latter reacts with alkenes in the presence of a platinum catalyst (Karstedt catalyst),²⁹ adding to the double bond with anti-Markovnikov selectivity, and leading to the introduction of the didactyl reactive silicon-based functional group and the desired monopodal, dipodal, and tripodal didactyl derivatives.

Solvents, MCM-41, and starting materials were purchased from Merck or TCI and used without further purification. Pentaerythritol triallyl ether (9) was purchased from Sigma-Aldrich at a 70% technical grade and it was purified by chromatography through a silica column (Pet/EtOAc 5:1). Sodium hydride (60% in mineral oil) was suspended and stirred in petroleum ether and then the supernatant was poured. This procedure was repeated three times to remove the mineral oil. Isonicotinoyl chloride hydrochloride (3) was prepared according to the literature procedure.³⁰

¹H and ¹³C NMR spectra were recorded at 300 MHz on a JEOL Eclipse ECP300 spectrometer or at 400 MHz on a

Bruker AVANCE Neo 400 instrument. Chemical shifts are reported in ppm with the protic impurities of the deuterated solvent as the internal reference. Mass spectra were obtained with a Thermo Finnigan LCQ-Deca XP-PLUS ion trap spectrometer equipped with an electrospray source. TLC was performed with silica gel (MN Kieselgel 60F254) and visualized by UV or sprayed with Dragendorff reagent or alkaline KMnO₄. Column chromatography was carried out on Macherey-Nagel silica gel 60 (0.063–0.200 mm).

The details of the synthesis reactions and of the spectroscopic characterization of products are reported in the [Supporting Information](#).

Synthesis of the Hybrid Materials. MCM-41 (0.3 g), dried overnight at 100 °C, was suspended in dry toluene (30 mL) and heated at 120 °C under stirring. Pyridine derivatives (0.0579 mmol) were added dropwise and the mixture was refluxed for 18 h; the reaction solution was filtered and washed with toluene. The white solid obtained was dried at 80 °C overnight.

Physicochemical Characterization. X-ray powder diffraction (XRPD) patterns were recorded using an ARL XTRA48 diffractometer with Cu K α radiation ($\lambda = 1.54062$ Å). Diffractograms were recorded at room temperature in the high-angle ($2\theta = 5$ – 55°) and low-angle ($2\theta = 1$ – 10°) range with a rate of $1.0^\circ \text{ min}^{-1}$. The X-ray profiles at low angles were collected with narrower slits.

Thermogravimetric analyses (TGA) were carried out on a SETSYS Evolution TGA-DTA/DSC thermobalance, under argon flow at a gas flow rate of 100 mL min^{-1} . The samples were heated from 30 to 1000 °C at a heating rate of $5^\circ \text{ C min}^{-1}$.

Before recording the FT-IR and SS-NMR spectra, all the samples were outgassed at 150 °C for 1 h to remove physisorbed water. FT-IR analyses of the self-supporting pellets

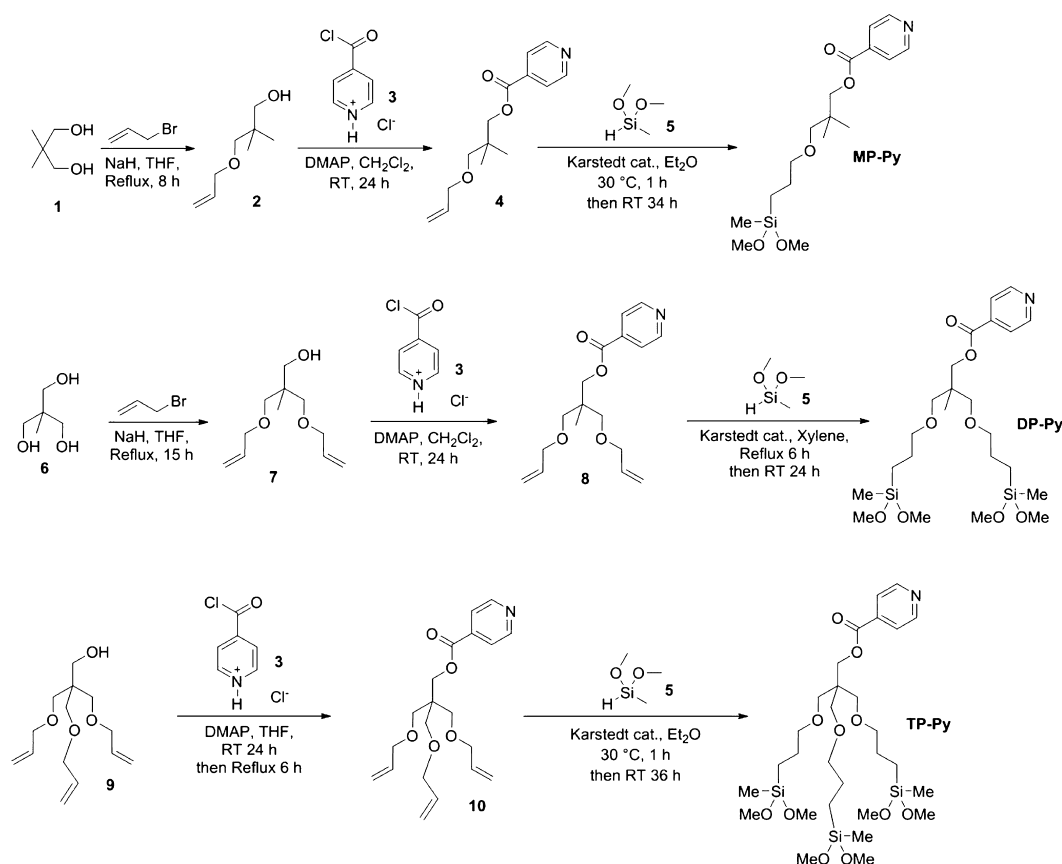


Figure 3. Synthesis of the multipodal pyridine-substituted derivatives.

were performed under vacuum conditions (residual pressure $<10^{-4}$ mbar) using a Bruker Equinox 55 spectrometer equipped with a pyroelectric detector (DTGS type) with a resolution of 4 cm^{-1} . FT-IR spectra were normalized with respect to the pellet weight. Variable temperature FT-IR measurements were performed in the $30\text{--}500\text{ }^{\circ}\text{C}$ temperature range, using a specifically designed cell permanently connected to the vacuum line.

Solid-state NMR spectra were acquired on a Bruker AVANCE III 500 spectrometer and using a wide bore 11.7 T magnet with operational frequencies 500.13, 99.35, and 125.77 MHz for ^1H , ^{29}Si , and ^{13}C , respectively. A 4 mm triple resonance probe with magic angle spinning (MAS) was employed in all the experiments and the samples were packed on a Zirconia rotor and spun at a MAS rate between 10 and 15 kHz. The magnitude of radio frequency (RF) fields was 100 and 42 kHz for ^1H and ^{29}Si , respectively. For the ^{13}C and ^{29}Si cross polarization (CP) MAS experiments, the RF fields of 55 and 28 kHz were used for initial proton excitation and decoupling, respectively. During the CP period, the ^1H RF field was ramped using 100 increments, whereas the $^{13}\text{C}/^{29}\text{Si}$ RF fields were maintained at a constant level. During the acquisition, the protons are decoupled from the carbons/silicons by using a two-pulse phase-modulated decoupling scheme. The relaxation delay, d_1 , between accumulations was 5 s for ^1H MAS and $^{13}\text{C}/^{29}\text{Si}$ CPMAS NMR. All chemical shifts

were reported by using the δ scale and are externally referred to TMS. ^1H MAS NMR spectra were deconvoluted for quantitative interpretation of overlapping peaks.³¹

Computational Modeling. MD simulations were performed in the canonical (n, V, T) ensemble at 298 and 353 K (using a Langevin thermostat to maintain a constant temperature),³² either in vacuum or in solution of dimethylformamide (DMF) or THF; for the simulations in liquid phase, a box of solvent was previously prepared and equilibrated to the experimental density, then a suitable number of solvent molecules were deleted to accommodate the hybrid systems. In all the MD calculations, an equilibration step of 0.5 ns was performed, followed by a 1 ns production run, using the LAMMPS simulation package.³³ The Conjugate Gradients algorithm was used, with an energy tolerance of 10^{-3} kcal/mol, and a force tolerance of 0.5 kcal/(mol Å). Atom-atom parameters were taken from the universal force field (UFF),³⁴ and atomic partial charges were generated by the QEq equilibration method, with a 10^{-6} e convergence. Coulombic interactions were computed with standard Ewald summation with a 10^{-6} kcal/mol accuracy, and van der Waals interactions were calculated with 6-12 Lennard-Jones function, using a 20 Å cutoff and parameters extracted from UFF.

Three models of pyridine-substituted derivatives were defined, as specified in the Introduction, and grafted on a silica slab with a thickness of 13.96 Å and a silanol surface

density of 2.4 nm^{-2} . Periodic boundary conditions were applied, with a $23.34 \text{ \AA} \times 26.55 \text{ \AA} \times 50.00 \text{ \AA}$ simulation box, large enough to exclude image interactions; a picture of the simulation box and the coordinates of the silica slab are provided in the Supporting Information.

RESULTS AND DISCUSSION

XRPD and TGA. To confirm that the grafting procedure does not alter the inorganic support structure, X-ray powder diffraction was performed on hybrid materials as well as on MCM-41. Both MP-Py-MCM-41 and TP-Py-MCM-41 exhibit all the characteristic reflections of hexagonally ordered MCM-41,³⁵ though less intense for the anchoring procedure lowering the structural order (patterns are shown in the Supporting Information).

TGA and differential thermogravimetry (DTG) provided insights on the thermal stability and hydrophilicity of the hybrid materials: as reported in Table 1 and illustrated in the

Table 1. Weight Losses (%) in plain MCM-41, Monopodal and Tripodal Hybrids

<i>T</i> (°C)	MCM-41	MP-Py-MCM-41	TP-Py-MCM-41
30–100	3.1	3.1	3.5
350–550		6.4	5.5

Supporting Information, over the whole temperature range, the weight loss is slightly lower for TP-Py-MCM-41, indicating that a smaller amount of pyridine derivatives is bound to the surface, possibly because the tripodal arrangement is statistically less favorite.

Both hybrids and plain MCM-41 undergo a first weight loss in the range 30–100 °C, due to physisorbed water evaporation: the total loss for $T < 100 \text{ °C}$ is slightly higher for TP-Py-MCM-41, revealing a higher hydrophilicity for this material, in agreement with the previous observation of a smaller organic coverage. A second weight loss is detected around 350–550 °C for both the hybrids, attributed to the decomposition of grafted pyridine derivatives: in this case, MP-Py-MCM-41 loses more organics than its tripodal counterpart (around 6.4 and 5.5%, respectively).

Moreover, in the latter temperature range, DTG shows two contributions for MP-Py-MCM-41, at 380 and 465 °C, while for TP-Py-MCM-41 a single loss at 465 °C is present. This can be interpreted by observing that the weight ratio of pyridine rings over organosiloxane chains is larger in the monopodal than in the tripodal hybrids: in MP-Py-MCM-41, the first loss at 380 °C can be attributed to the decomposition of the pyridine group, and the second to the loss of the remaining organic chains, while in TP-Py-MCM-41, the pyridine contribution is less appreciated and a single loss is detected for the whole organic layer.

FT-IR. The FT-IR spectra of MP-Py-MCM-41 and TP-Py-MCM-41 are reported in Figure 4; in addition, the spectra of MP-Py and TP-Py simply adsorbed in MCM-41 and KBr are compared to those of hybrids in Figures 5 and 6.

In the high-frequency region (Figure 4A) both hybrid materials show a weak absorption at 3745 cm^{-1} and a broad band between 3700 and 2500 cm^{-1} , assigned to the O–H stretching modes of isolated and hydrogen-bonded silanols, respectively. Aromatic and aliphatic C–H stretching modes are detected in $3100\text{--}3000$ and $3000\text{--}2800 \text{ cm}^{-1}$ ranges, respectively, the former with a lower intensity, as expected.

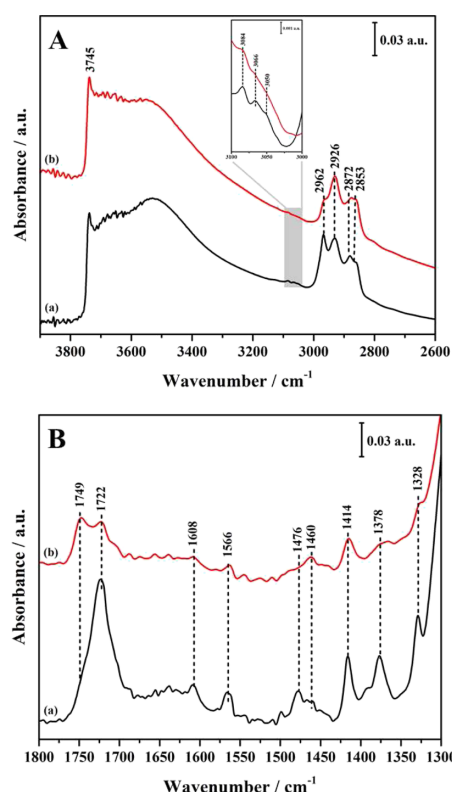


Figure 4. FT-IR spectra of MP-Py-MCM-41 (a, black) and TP-Py-MCM-41 (b, red).

The signal due to the C=O stretching mode, associated with the absorptions in the $1722\text{--}1750 \text{ cm}^{-1}$ range, is particularly interesting for our purposes. As shown in the spectra in Figures 5 and 6, when either MP-Py or TP-Py is adsorbed in the solid matrix, that is, without proceeding to the chemical grafting, this mode produces a single band at 1733 cm^{-1} ; on the other hand, after the grafting, this band is split into two signals (see Figure 4B), at 1749 and 1722 cm^{-1} . In the monopodal hybrid, the band at 1722 cm^{-1} is largely dominant, with just a weak shoulder at 1749 cm^{-1} , while in the tripodal system, the two components have the same intensity.

The splitting stems from the interaction with the surface in the hybrid materials: the absorption bands at 1749 and 1722 cm^{-1} can be associated with free and H-bonded carbonyl groups, respectively, the latter indicating a possible interaction with surface silanols. The different aspect of the FT-IR spectra, then, shows that in TP-Py-MCM-41 at least a part of the carbonyl groups remains far from the surface.

The pyridine ring also can interact with silanols if the organic molecules bend toward the surface: the characteristic aromatic C–C stretching modes, in the $1600\text{--}1400 \text{ cm}^{-1}$ region,³⁶ are actually upshifted after the grafting in both MP-Py-MCM-41 and TP-Py-MCM-41. In particular, as seen by comparing the spectra of grafted (Figure 4B) and adsorbed (Figures 5 and 6) pyridine derivatives, the bands at 1409 , 1562 , and 1597 cm^{-1} move to 1414 , 1566 , and 1608 cm^{-1} , respectively. However, such a shift to higher frequencies is observed for all MP-Py and TP-Py modes after the grafting,

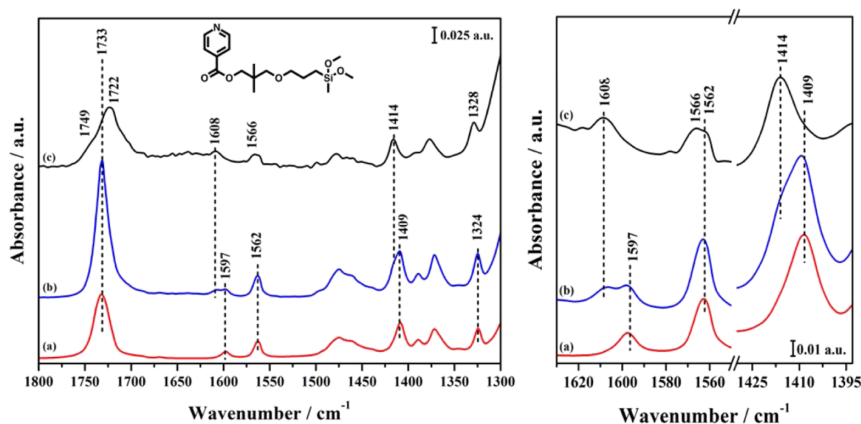


Figure 5. FT-IR spectra of MP-Py in different environments: adsorbed in KBr powder (a, red); adsorbed on MCM-41 (b, blue); grafted on MCM-41, that is, MP-Py-MCM-41 (c, black). Right panel: magnified C–C ring stretching region.

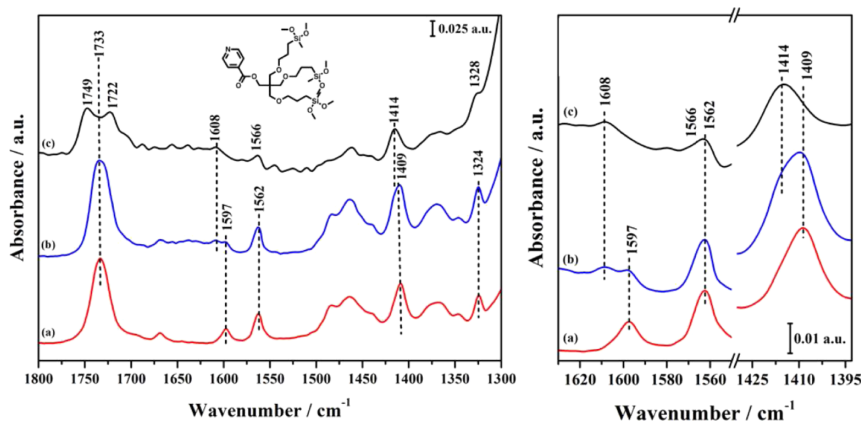


Figure 6. FT-IR spectra of TP-Py in different environments: adsorbed in KBr powder (a, red); adsorbed on MCM-41 (b, blue); grafted on MCM-41, that is, TP-Py-MCM-41 (c, black). Right panel: magnified C–C ring stretching region.

likely due to the change in the chemical environment rather than to direct interactions with silanols. Then, the shift of pyridine C–C stretching frequencies is not conclusive about the existence or entity of such interactions.

Temperature-dependent FT-IR spectra, reported in Figure 7, show a gradual disappearance of the broad band between 3700 and 2500 cm^{-1} when the samples are heated from 30 to 500 $^{\circ}\text{C}$, due to the loss of physisorbed water and, later, the weakening of H-bonds between silanols. In agreement with TGA, the gradual decomposition of the organic layer starts around 300 $^{\circ}\text{C}$, as witnessed by the decreasing intensity of aliphatic C–H stretching (3000–2800 cm^{-1}), carbonyl stretching (1750–1700 cm^{-1}), pyridine ring C–C stretching (1610–1400 cm^{-1}). Interestingly, as the temperature increases, the component of C=O stretching at 1722 cm^{-1} gradually lowers and eventually disappears, while the band at 1749 cm^{-1} becomes predominant in both the hybrid materials. This is consistent with the hypothesis that the lower frequency is associated with carbonyl groups engaged in H-bonding with the surface, which becomes weaker as the molecules extend farther from the silica at higher temperatures.

Solid-State NMR. The local environment of silicon atoms in the hybrid materials was investigated with ^{29}Si CPMAS NMR, as shown in Figure 8A. The spectra of both MP-Py-MCM-41 and TP-Py-MCM-41 show Q^4 , Q^3 , and Q^2 resonance peaks (due to tetrahedrally coordinated silicon atoms of the MCM-41 support) at -110 , -101 , and -91 ppm, respectively, while D^1 and D^2 signals (coming from the grafted chain silicon) are found at -9 and -16 ppm, respectively.³⁷ In both the spectra, D^2 is more intense than the D^1 peak, indicating that didactyl arrangement is preferred (i.e., each organic molecule tends to bind with two siloxane bridges); in the tripodal system, also a low D^0 signal is detected at -3 ppm, revealing the presence of a small amount of unattached chains, coming either from TP-Py molecules, or from TP-Py-MCM-41 with only one or two siloxane chains grafted to the surface.

The integrity of pyridine derivatives after the grafting was assessed by ^{13}C CPMAS NMR (Figure 8B); and the resonance peaks were assigned by comparison with the liquid-state ^{13}C spectrum. In both Py-MP-MCM-41 and Py-TP-MCM-41, ^{13}C signals can be clearly attributed to the organosiloxane groups, demonstrating the organic chain integrity. A close examination of the spectra reveals that two C nuclei in the hybrids display

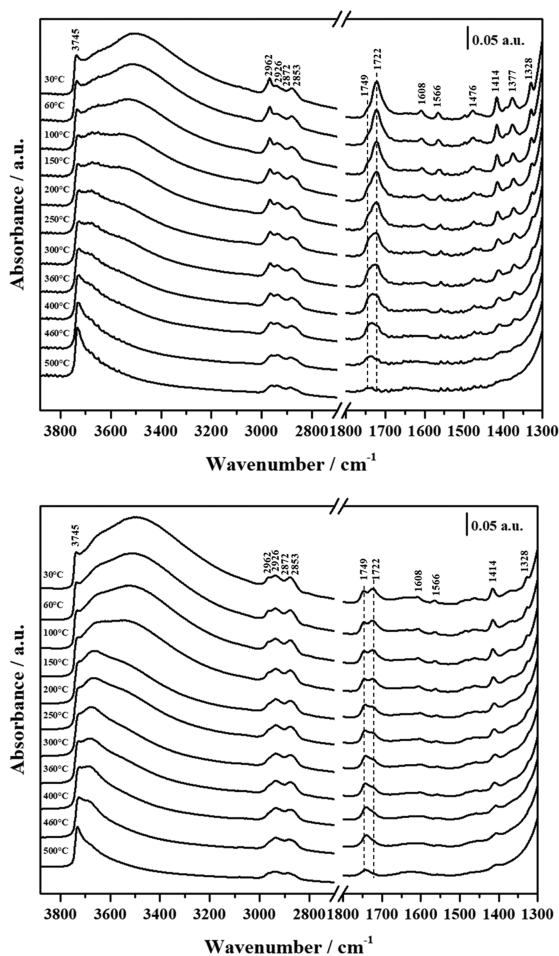


Figure 7. Variable temperature FT-IR spectra of MP-Py-MCM-41 (upper panel) and TP-Py-MCM-41 (lower panel).

multiple peaks, associated with different environments, namely carbon in position 12 in MP-Py-MCM-41 (curve a) and carbon in position *h* in TP-Py-MCM-41 (curve b).

^1H NMR resonance spectra are reported in Figure 9 for both hybrids, as prepared and after degassing to remove physisorbed water; in the same figure, the spectral deconvolution is presented to separate organosiloxane and silica contributions. In a separate experiment, the ^1H NMR spectra were recorded after exchange with D_2O at r.t., in order to remove the signals due to exchangeable protons (i.e., from water and silanols) avoiding the thermal treatment, which could damage the organics as well; besides, it is known that H-bond-accepting groups such as pyridine can form clusters with silanols and water molecules able to resist to mild thermal treatments.³⁸

Before the vacuum treatment, the most intense band, around 3.5–5 ppm, is attributed to physisorbed water (marked 1 in the deconvolution curves), H-bonded silanols (marked 2), and also $-\text{OCH}_2$ from organosiloxane. These components are strongly reduced after degassing, when most of the physisorbed water is removed: at the same time, a sharp peak due to isolated silanols appears around 1.8 ppm.

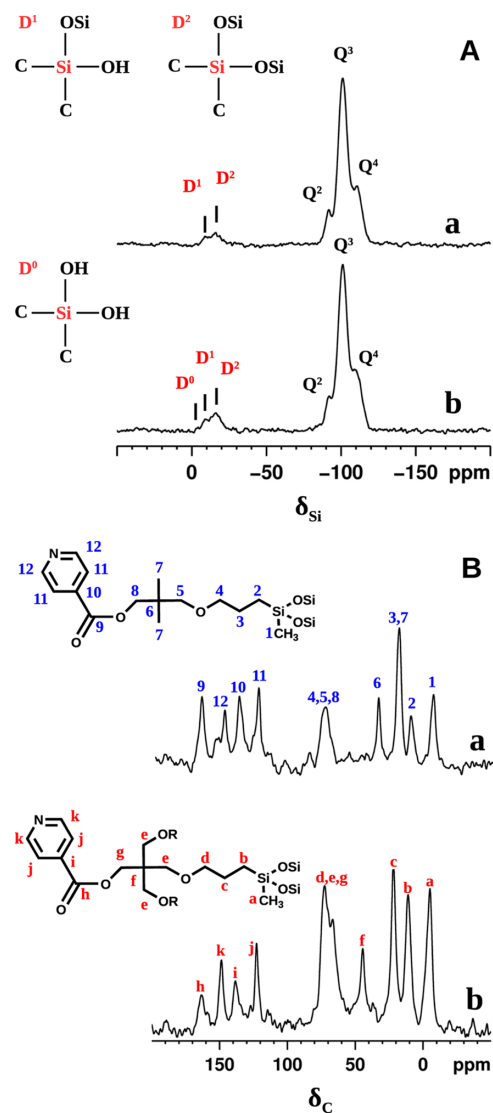


Figure 8. ^{19}Si (A) and ^{13}C (B) CPMAS NMR spectra of MP-Py-MCM-41 (a) and TP-Py-MCM-41 (b) hybrids.

In the spectra of the D_2O -exchanged samples, reported in the Supporting Information, the broad subresonance band at 6–8 ppm disappears, confirming that it is related to water and silanol groups: such downfield shifts in the δ values are indicative of strong H-bond interactions, likely due to clusters formed with pyridine nitrogen.

Other contributions from the organic chains are visible at 0 ppm (attributed to a methyl group directly bound to organosilicon), in the aliphatic region (around 1 ppm), and the sharp peaks at 7.8 and 8.5 ppm, due to the pyridine ring. An additional signal at 7.5 ppm (marked 5) is not straightforwardly assignable, but could be due to silanol or water hydrogens interacting with the carbonyl group, a motif already discussed in the FT-IR study.

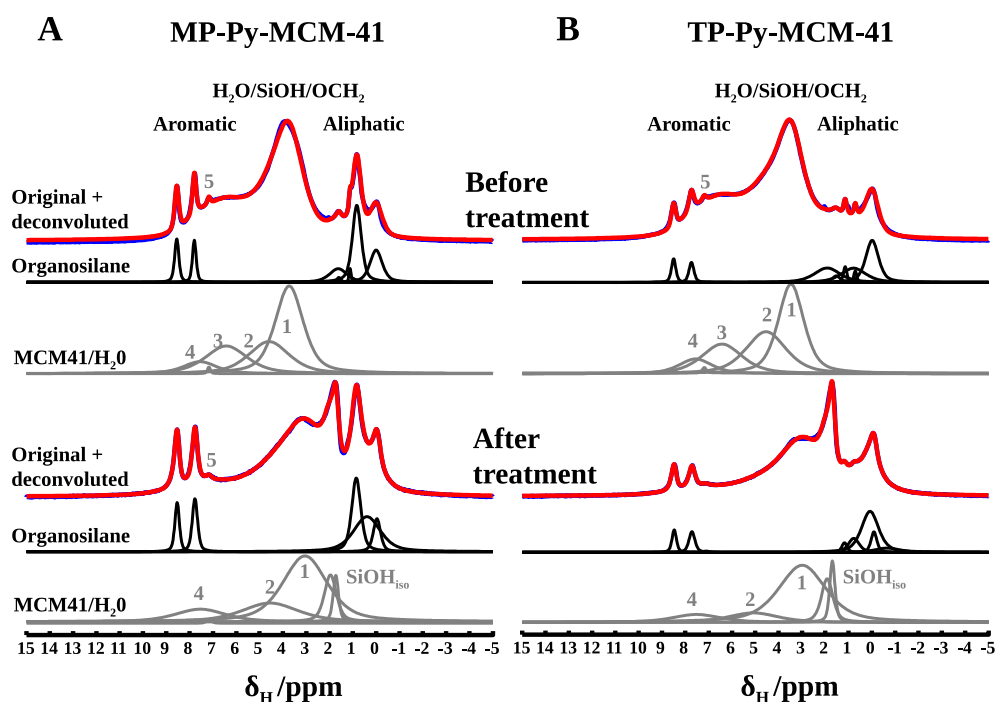


Figure 9. ^1H NMR resonance spectra (red), deconvolution spectra (blue), organic (black), and inorganic (grey) components of MP-Py-MCM-41 (A) and TP-Py-MCM-41 (B) hybrids, as prepared (upper panel) and after mild thermal treatment for degassing (lower panel).

In the deconvolution curves before the vacuum treatment, two components marked 3 and 4 are related to water and silanols interacting with the pyridine nitrogen atom: these signals, along with the contribution 5 mentioned above, show that the catalytic center of the organosiloxane interacts with the silica surface, either directly or with the mediation of physisorbed water.

After the thermal treatment for degassing, the component 3 disappears, while 4 is reduced but not eliminated, showing that the former is likely due to water and the latter to silanols, and that the silanol–nitrogen interactions are strong enough to survive after the treatment. Even more interesting for our purpose, the intensity of the band marked 4 reduces more in TP-Py-MCM-41 than in MP-Py-MCM-41 after degassing, a further indication that the tripodal organosiloxane prefers the extended conformation, keeping the pyridine group far from the surface.

Molecular Dynamics Simulations. Three periodic models of the silica surface were prepared with mono-, di-, and tripodal pyridine derivatives grafted, as described above; all the alkoxy silane chains were anchored to the surface through a didactyl bond, by condensating the organosilicon methoxy groups with silanols placed in suitable positions on the surface.

The first set of simulations was performed in vacuum at 298 K, in agreement with the conditions of the physicochemical characterization after degassing, as discussed above: after the equilibration step, the dynamical evolution of the systems was followed for 1 ns, monitoring the distance of the pyridine nitrogen from the closest surface silicon atom. This parameter varied from ca. 4 Å, when the organic chain lies as close as possible to the surface, to ca. 11 Å, when it extends almost perpendicular to the surface: in Figure 10, we show the

evolution of the nitrogen–silicon distance during the simulations.

The difference among the various grafting schemes is striking: with one or two tethering chains, during the simulation the pyridine spends almost all the time very close to the surface, while in the tripodal system, the catalyst is forced to stay at a much larger distance, always in an extended conformation. To help visualizing the different interface conformations, two representative frames extracted from the MD runs of MP-Py-MCM-41 and TP-Py-MCM-41 are shown in Figure 11. This result is in fair agreement with the FT-IR and NMR findings discussed above, which show the different behavior of monopodal and tripodal hybrids.

The results in Figure 10 demonstrate that tripodal systems only guarantee that the organic group remains far from the surface at 298 K in vacuum: one can wonder if higher temperatures or the presence of solvents could favor an extended conformation also for the other hybrids. To verify this point, more MD runs were performed on mono- and dipodal systems, either increasing the temperature to 353 K in vacuum, or adding DMF or THF at 298 K, with the results reported in Figure 12.

Even at a substantially higher temperature, thermal motions are not able to surpass the dispersion and H-bond interactions, which pull the organic chain close to the surface, neither in monopodal nor in dipodal systems. On the other hand, the tested solvents act as passivating agents on the surface, competing for the interactions with the organosilica chains: as a result, the dipodal hybrid is removed from its position and forced to extend farther from the surface; the monopodal chain, however, is flexible enough to remain close to the silica even in the presence of a solvent.

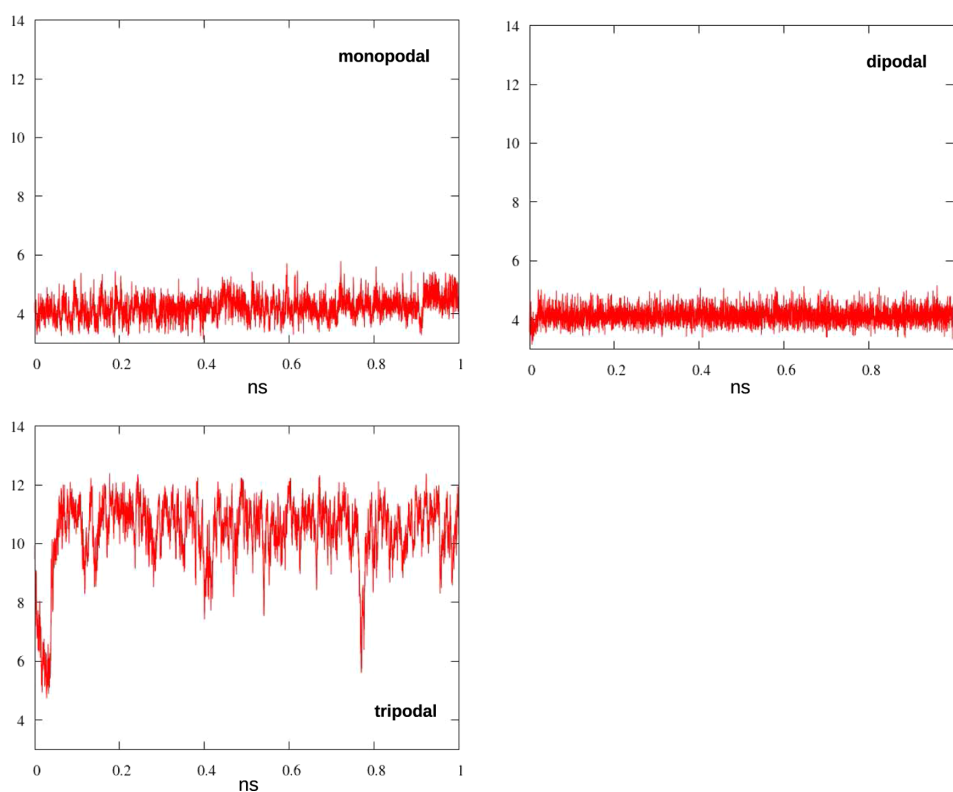


Figure 10. Distance (Å) between pyridine N and the closest Si on the surface along the MD run for mono-, di-, and tripodal hybrid models.

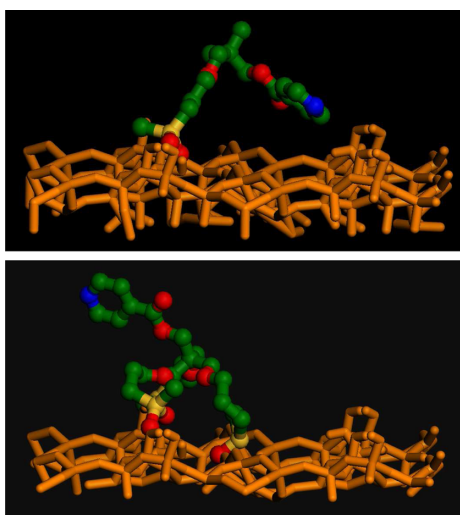


Figure 11. MD frames for MP-Py-MCM-41 (upper) and TP-Py-MCM-41 (lower): lower part of the silica slab and organic hydrogens not shown for clarity; SiO₂: orange, Si: yellow, C: green, O: red, and N: blue.

We have collected a number of pieces of evidence, to show that the organosilica podality does indeed affect the conformation of the organic/inorganic interface. Only with

three grafting points, we are guaranteed that the active center keeps far from the surface, free of interactions, which could hamper its catalytic activity: this is confirmed by the FT-IR and NMR results, as well as by the MD simulations. From the MD results, moreover, we predict that temperature has a little effect in moving monopodal and dipodal systems far from the surface, while the presence of passivating solvents could force dipodal hybrids to leave the surface.

CONCLUSIONS

In this work, we carry out our investigation on the relation between the grafting architecture in organosilica hybrids and the structure of the interface, focusing in particular on the average distance of the catalyst from the inorganic surface.

We had already demonstrated that *silicodactyly* (i.e., the number of siloxane bridges formed by a single organic chain grafted to silica) has a very little effect on the conformation of the organic fragment, which tends to lie down on the surface in all the cases, driven by dispersion and H-bond interactions. Conversely, here we find that *silicopodality* (the number of alkylsiloxane chains used by a single catalytic group to anchor to the surface) can effectively influence the structure of the interface, possibly increasing the catalytic efficiency as the organic part stays farther from the surface, more available to interact with the substrates.

A series of pyridine-substituted derivatives, with one to three alkylsiloxane chains as grafting chains, were used as probes. Mono-, di-, and tripodal systems were synthesized, grafted to MCM-41 mesoporous silica, and characterized: FT-IR and SS-

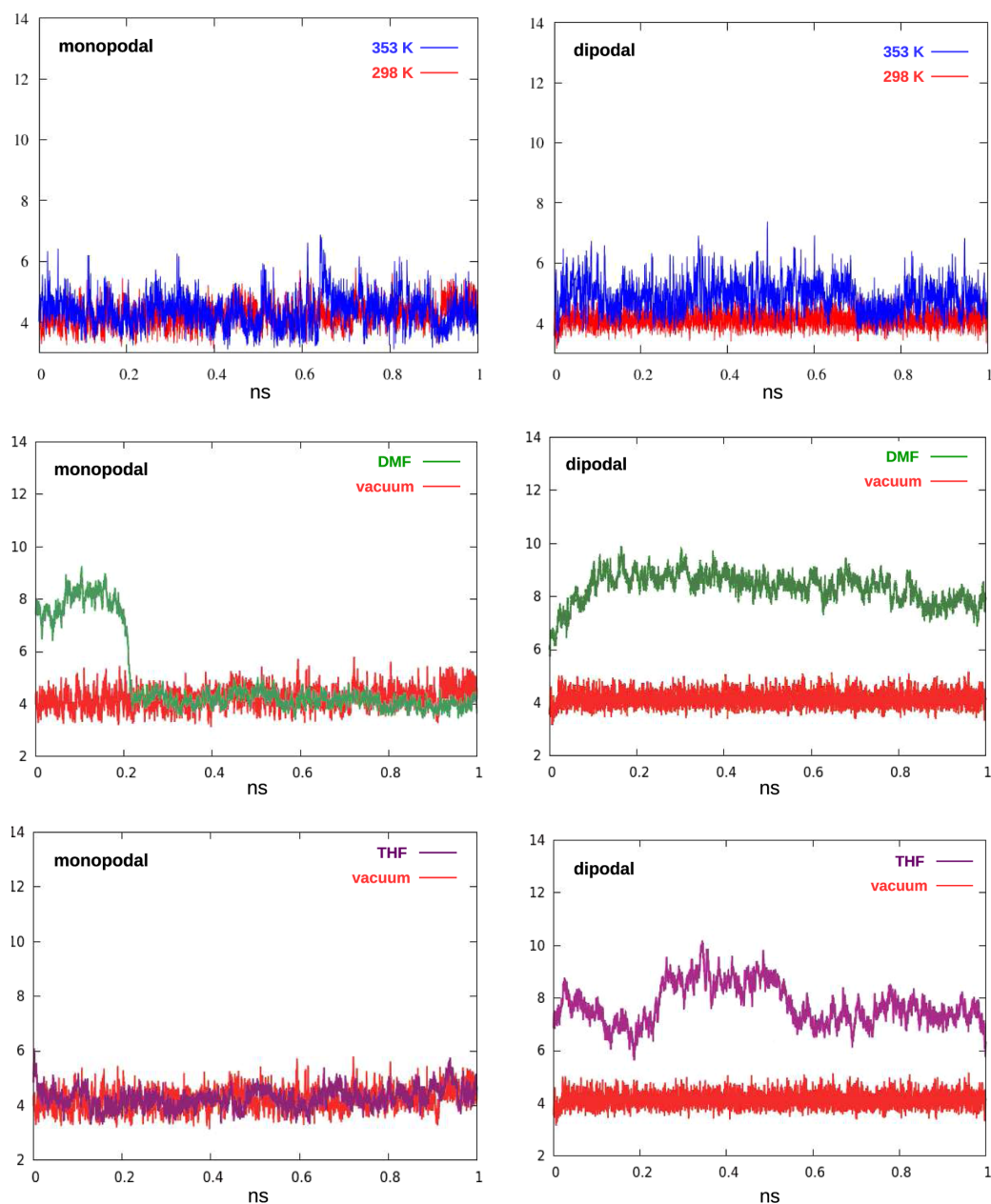


Figure 12. Distance (Å) between pyridine N and the closest Si on the surface along the MD run for monopodal and dipodal hybrids under different conditions: vacuum, 298 K (red); vacuum, 353 K (blue); in DMF, 298 K (green); and in THF, 298 K (purple).

NMR unambiguously show that the pyridine heterocyclic ring in the monopodal derivative interacts with the silica surface, while in the tripodal hybrid, such interactions reduce largely, proving that the pyridine is kept farther from the silica.

This finding is confirmed by several MD simulations, involving mono-, di-, and tripodal systems grafted to a silica model: the calculations reveal that only in the tripodal hybrid, the pyridine ring spends almost all the time far from the silica, while in the other systems, it tends to stick on the surface. MD

simulations also predict that monopodal and dipodal hybrids lie down on the surface even upon increasing the temperature to 353 K, but adding a solvent (either DMF or THF) that competes for the interactions with the surface, the conformation of the dipodal hybrid changes and the pyridine ring distances from the silica.

These results provide useful information about the relation of grafting and conformation in this kind of interfaces, which

should be taken into account when designing hybrid organic/inorganic catalysts.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.1c06150>.

Pictures of the silica slab model; XRPD patterns; thermogravimetric curves; ^1H NMR after D_2O exchange; details of the organic syntheses; and NMR spectra of new compounds (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Maurizio Cossi – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy; orcid.org/0000-0002-1817-9008; Email: maurizio.cossi@uniupo.it

Authors

Ivana Miletto – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy

Chiara Ivaldi – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy

Enrica Gianotti – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy; orcid.org/0000-0002-4383-3526

Geo Paul – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy; orcid.org/0000-0002-0944-0016

Fabio Travagin – Dipartimento di Scienze del Farmaco (DSF), Università del Piemonte Orientale, I-28100 Novara, Italy

Giovanni Battista Giovenzana – Dipartimento di Scienze del Farmaco (DSF), Università del Piemonte Orientale, I-28100 Novara, Italy; CAGE Chemicals srl, I-28100 Novara, Italy; orcid.org/0000-0001-7117-0347

Alberto Fraccarollo – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy

Davide Marchi – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy

Leonardo Marchese – Dipartimento di Scienze e Innovazione Tecnologica (DISIT), Università del Piemonte Orientale, I-15121 Alessandria, Italy; orcid.org/0000-0001-9191-1237

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcc.1c06150>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by EU in the framework of the HORIZON2020 program, through the project MULTIH-CAT (GA-720783), by the Italian Ministry of Education, University, and Research (PRIN-2010A2FSS9) and by

Università del Piemonte Orientale (Fondi di Ateneo per la Ricerca anno 2019 - FAR2019-“Multidisciplinary Study on Hybrid Organic-Inorganic Catalysts. Effect of the Anchoring Geometry on the Catalytic Properties”).

■ REFERENCES

- (1) Wight, A. P.; Davis, M. E. Design and Preparation of Organic/Inorganic Hybrid Catalysts. *Chem. Rev.* **2002**, *102*, 3589–3614.
- (2) Sanchez, C.; Julián, B.; Belleville, P.; Popall, M. Applications Of Hybrid Organic–inorganic Nanocomposites. *J. Mater. Chem.* **2005**, *15*, 3559–3592.
- (3) Hoffmann, F.; Cornelius, M.; Morell, J.; Fröba, M. Silica-Based Mesoporous Organic–Inorganic Hybrid Materials. *Angew. Chem., Int. Ed.* **2006**, *45*, 3216–3251.
- (4) Férey, G. Hybrid Porous Solids: Past, Present, Future. *Chem. Soc. Rev.* **2008**, *37*, 191–214.
- (5) Diaz, U.; Corma, A. Organic-Inorganic Hybrid Materials: Multi-Functional Solids for Multi-Step Reaction Processes. *Chem.—Eur. J.* **2018**, *24*, 3944–3958.
- (6) Mir, S. H.; Nagahara, L. A.; Thundat, T.; Mokarian-Tabari, P.; Furukawa, H.; Khosla, A. Organic-Inorganic Hybrid Functional Materials: An Integrated Platform for Applied Technologies. *J. Electrochem. Soc.* **2018**, *165*, B3137–B3156.
- (7) Gianotti, E.; Miletto, I.; Ivaldi, C.; Paul, G.; Marchese, L.; Meazza, M.; Rios, R.; Raja, R. Hybrid Catalysts Based on N-heterocyclic Carbene Anchored on Hierarchical Zeolites. *RSC Adv.* **2019**, *9*, 35336–35344.
- (8) Ricardi, N. C.; Arenas, L. T.; Benvenuti, E. V.; Hinrichs, R.; Flores, E. E. E.; Hertz, P. F.; Costa, T. M. H. High Performance Biocatalyst Based on β -D-galactosidase Immobilized on Mesoporous Silica/titania/chitosan Material. *Food Chem.* **2021**, *359*, 129890.
- (9) Bonura, G.; Todaro, S.; Frusteri, L.; Majchrzak-Kuceba, I.; Wawrzynczak, D.; Pászti, Z.; Tálás, E.; Tompos, E.; Ferenc, L.; Solt, H.; et al. Inside the Reaction Mechanism of Direct CO₂ Conversion to DME Over Zeolite-based Hybrid Catalysts. *Appl. Catal., B* **2021**, *294*, 120255.
- (10) Lei, Q.; Guo, J.; Noureddine, A.; Wang, A.; Wuttke, S.; Brinker, C. J.; Zhu, W. Sol–Gel-Based Advanced Porous Silica Materials for Biomedical Applications. *Adv. Funct. Mater.* **2020**, *30*, 1909539.
- (11) Erigoni, A.; Diaz, U. Porous Silica-Based Organic-Inorganic Hybrid Catalysts: A Review. *Catalysts* **2021**, *11*, 79.
- (12) Judeinstein, P.; Sanchez, C. Hybrid Organic–inorganic Materials: A Land of Multidisciplinary. *J. Mater. Chem.* **1996**, *6*, 511–525.
- (13) Park, J.-W.; Park, Y. J.; Jun, C.-H. Post-grafting Of Silica Surfaces with Pre-functionalized Organosilanes: New Synthetic Equivalents of Conventional Trialkoxysilanes. *Chem. Commun.* **2011**, *47*, 4860–4871.
- (14) Diaz, U.; Brunel, D.; Corma, A. Catalysis Using Multifunctional Organosiliceous Hybrid Materials. *Chem. Soc. Rev.* **2013**, *42*, 4083–4097.
- (15) Guo, S.; Asset, T.; Atanassov, P. Catalytic Hybrid Electrocatalytic/biocatalytic Cascades for Carbon Dioxide Reduction and Valorization. *ACS Catal.* **2021**, *11*, 5172–5188.
- (16) Mota, N.; Millán Ordoñez, E.; Pawelec, B.; Fierro, J. L. G.; Navarro, R. M. Direct Synthesis of Dimethyl Ether from CO₂: Recent Advances in Bifunctional/hybrid Catalytic Systems. *Catalysts* **2021**, *11*, 411.
- (17) MS, R.; Shanmuga Priya, S.; Freudenberg, N.; Sudhakar, K.; Tahir, M. Metal-organic Framework-based Photocatalysts for Carbon Dioxide Reduction to Methanol: a Review on Progress and Application. *J. CO₂ Util.* **2021**, *43*, 101374.
- (18) Martínez-Edo, G.; Balmori, A.; Pontón, I.; Martí del Río, A.; Sánchez-García, D. Functionalized Ordered Mesoporous Silicas (MCM-41): Synthesis and Applications in Catalysis. *Catalysts* **2018**, *8*, 617.

- (19) Xue, X.-X.; Ma, C.-L.; Di, J.-H.; Huo, X.-Y.; He, Y.-C. One-pot Chemo-enzymatic Conversion of D-xylose to Furfuralcohol By Sequential Dehydration with Oxalic Acid Plus Tin-based Solid Acid and Bioreduction with Whole-cells. *Bioresour. Technol.* **2018**, *268*, 292–299.
- (20) Li, Y.-Q.; Wang, P.; Liu, H.; Lu, Y.; Zhao, X.-L.; Liu, Y. Co-catalysis of a Bi-functional Ligand Containing Phosphine and Lewis Acidic Phosphonium for Hydroformylation-acetalization of Olefins. *Green Chem.* **2016**, *18*, 1798–1806.
- (21) Khassenova, G.; García Mancheño, O. Lewis Base-Brønsted Acid Co-catalyzed Morita-Baylis-Hillman Reaction of Cyclic Sulfamidate Imines. *Eur. J. Org. Chem.* **2021**, 2752–2755.
- (22) Akai, S.; Tsuchimochi, I.; Hori, S.; Takeuchi, Y.; Egi, M.; Satoh, T.-O.; Kanomata, K.; Ikawa, T. Four-Step One-Pot Catalytic Asymmetric Synthesis of Polysubstituted Tricyclic Compounds: Lipase-Catalyzed Dynamic Kinetic Resolution Followed By An Intramolecular Diels-Alder Reaction. *Synlett* **2021**, *32*, 822–828.
- (23) Yokoi, T.; Yoshitake, H.; Tatsumi, T. Synthesis Of Amino-functionalized MCM-41 Via Direct Co-condensation and Post-synthesis Grafting Methods Using Mono-, Di- and Tri-amino-organoalkoxysilanes. *J. Mater. Chem.* **2004**, *14*, 951–957.
- (24) Brühwiler, D. Postsynthetic Functionalization of Mesoporous Silica. *Nanoscale* **2010**, *2*, 887–892.
- (25) Zhang, T.; Li, B.; Zhang, X.; Wang, J.; Wei, L.; Zhao, B.; Li, B. 4-Dimethylaminopyridine Grafted on MCM-41 as An Efficient and Highly Stable Catalyst for the Production of -tocopherol Acetate. *J. Porous Mater.* **2020**, *27*, 1639–1648.
- (26) Paul, G.; Bisio, C.; Braschi, I.; Cossi, M.; Gatti, G.; Gianotti, E.; Marchese, L. Combined Solid-state NMR, FT-IR and Computational Studies on Layered and Porous Materials. *Chem. Soc. Rev.* **2018**, *47*, 5684–5739.
- (27) Ivaldi, C.; Miletto, I.; Paul, G.; Giovenzana, G.; Fraccarollo, A.; Cossi, M.; Marchese, L.; Gianotti, E. Influence of Silicodactyly in the Preparation of Hybrid Materials. *Molecules* **2019**, *24*, 848.
- (28) Turgis, R.; Billault, I.; Acherar, S.; Augé, J.; Scherrmann, M.-C. Total Synthesis of High Loading Capacity PEG-based Supports: Evaluation and Improvement of the Process By Use of Ultrafiltration and PEG as a Solvent. *Green Chem.* **2013**, *15*, 1016–1029.
- (29) Nakajima, Y.; Shimada, S. Hydrosilylation Reaction of Olefins: Recent Advances and Perspectives. *RSC Adv.* **2015**, *5*, 20603–20616.
- (30) van Houtem, M. H. C. J.; Martín-Rapún, R.; Vekemans, J.; Meijer, E. Desymmetrization of 3,3-Bis(acylamino)-2,2-bipyridine-Based Discotics: the High Fidelity of their Self-Assembly Behavior in the Liquid-Crystalline State and in Solution. *Chem.—Eur. J.* **2010**, *16*, 2258–2271.
- (31) Massiot, D.; Fayon, F.; Capron, M.; King, L.; Le Calvé, S.; Alonso, B.; Durand, J.-O.; Bujoli, B.; Gan, Z.; Hoatson, G. Modelling One- and Two-dimensional Solid-state NMR Spectra. *Magn. Reson. Chem.* **2002**, *40*, 70–76.
- (32) Schneider, T.; Stoll, E. Molecular-dynamics Study of a Three-dimensional One-component Model for Distortive Phase Transitions. *Phys. Rev. B: Solid State* **1978**, *17*, 1302–1322.
- (33) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (34) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (35) Gianotti, E.; Diaz, U.; Vely, A.; Corma, A. Designing Bifunctional Acid–base Mesoporous Hybrid Catalysts for Cascade Reactions. *Catal. Sci. Technol.* **2013**, *3*, 2677–2688.
- (36) Ureña, F. P.; Gómez, M. F.; González, J. J. L.; Torres, E. M. A New Insight Into the Vibrational Analysis of Pyridine. *Spectrochim. Acta, Part A* **2003**, *59*, 2815–2839.
- (37) Paul, G.; Steuernagel, S.; Koller, H. Non-covalent interactions of a drug molecule encapsulated in a hybrid silica gel. *Chem. Commun.* **2007**, 5194–5196.
- (38) Paul, G.; Musso, G. E.; Bottinelli, E.; Cossi, M.; Marchese, L.; Berlier, G. Investigating the Interaction of Water Vapour with

Appendix D

FFfitpy script

The following python script was used to perform the force field fit described in Chapter 3, interfacing with the LAMMPS code to compute single-point energies and iteratively updating the parameters to reproduce the target energy for each configuration (computed in this thesis at the DFT B3LYP-D3 level).

A scheme that shows the way the script works is reported below in Figure D.1, while the code itself is attached in the following pages.

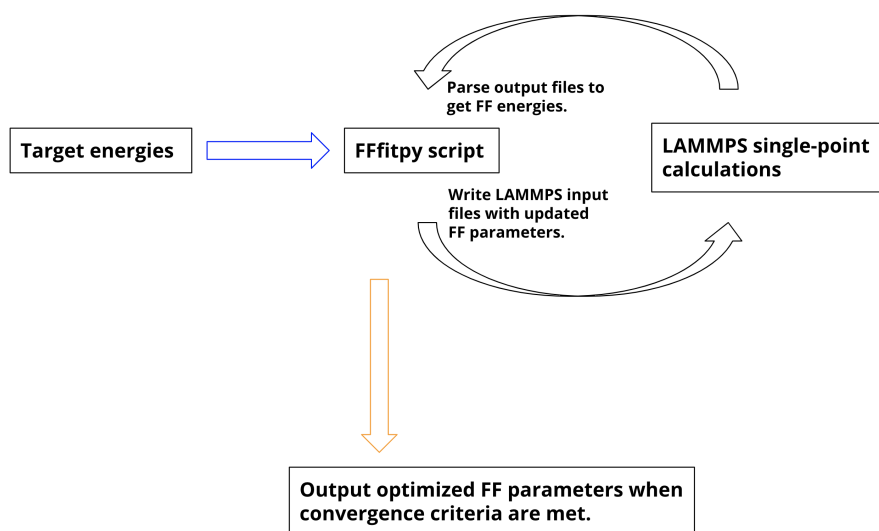


Figure D.1: Schematic representation of the fitting procedure using FFfitpy.

```

1  #!/usr/bin/env python
2
3  import matplotlib.pyplot as plt
4  from scipy.optimize import curve_fit
5  import subprocess
6  import numpy as np
7  import os
8  import time
9  import sys
10
11 # read files containing the reference (ab initio) data for
    the fit. two files are expected: one containing the 'x'
        column (ids), and the other containing the 'y' column
            (energies). the files are expected to be named: x.tofit
                , y.tofit
12 with open('x.tofit', 'r') as inx:
13     quantum_x_strings = inx.read().splitlines()
14 quantum_x = [float(i) for i in quantum_x_strings]
15 quantum_x = np.array(quantum_x)
16 with open('y.tofit', 'r') as iny:
17     quantum_y_strings = iny.read().splitlines()
18 quantum_y = [float(i) for i in quantum_y_strings]
19 quantum_y = np.array(quantum_y)
20
21 # initial guess for the force field parameters. to be read
    from the file 'initial.guess'
22 # XXX the order of the list is important!
23 with open('initial.guess', 'r') as inguess:
24     initial_guess_strings = inguess.read().splitlines()
25 initial_guess = []
26 for i in initial_guess_strings:
27     initial_guess.append(float(i.split('#')[0].strip()))
28
29 # it is assumed that all LAMMPS files needed for the
    fitting are in the same directory
30 directory = os.getcwd()
31 file_ids = [f.split('.')[0] for f in os.listdir(directory)
    if (os.path.isfile(f) and f.endswith('.in'))] # names
        of input files with no extension
32 file_ids.sort() # XXX attention: files should be
    lexicographically sortable
33
34 # function to parse LAMMPS outputs and get the energies.
    the energies are returned as an ordered list from the
        first geometry to the last (provided that the LAMMPS
            input files are correctly named to be lexicographically
                sortable)
35 def parse_outputs():
36
37     energies = []
38
39     output_files = [f for f in os.listdir(directory) if (os
        .path.isfile(f) and f.endswith('.log'))]

```

```

40 output_files.sort()
41
42 for FILE in output_files:
43     with open(FILE, 'r') as f:
44         file_list = f.read().splitlines()
45         # additional check: if last line of a LAMMPS .log
46         # file shows an error termination, exit gracefully by
47         # writing an error message to a file
48         if not file_list[-1].startswith('Total wall time'):
49             with open('python.err', 'w') as err:
50                 err.write('the script stopped because one
51                 or more of the LAMMPS outputs did not end correctly.')
52                 sys.exit()
53         for counter,i in enumerate(file_list):
54             if i.strip().startswith('TotEng'):
55                 energy = file_list[counter+1].strip()
56                 energies.append(float(energy))
57                 break
58
59 return energies
60
61 # function to update force field parameters in .in files to
62 # generate new LAMMPS inputs. all the .in files in the
63 # directory are parsed and overwritten with the updated
64 # force field parameters.
65
66 #
67 # -- ids to specify the parameter line in the LAMMPS .in
68 # file that must be updated
69 par_to_fit = 'Au-S'
70 par_fit_identifier = '#-*- '
71 # --
72 def update_parameters(parameters_list):
73
74     in_files = [f for f in os.listdir(directory) if (os.
75     path.isfile(f) and f.endswith('.in'))]
76
77     for FILE in in_files:
78         with open(FILE, 'r') as f:
79             file_list = f.read().splitlines()
80
81             new_file = []
82             for line in file_list:
83                 if not par_fit_identifier in line:
84                     new_file.append(line)
85                 else:
86                     if line.split(par_fit_identifier)[-1].strip
87                     () == par_to_fit:
88                         new_line = line.split()[0] + ' ' + line
89                         .split()[1] + ' ' + line.split()[2] + ' ' + str(
90                         parameters_list[0]) + ' ' + str(parameters_list[1]) + '
91                         ' + par_fit_identifier + ' ' + line.split(
92                         par_fit_identifier)[-1].strip()
93                         new_file.append(new_line)

```

```

80
81     # overwrite old file
82     with open(FILE, 'w') as o:
83         for i in new_file:
84             o.write(i)
85             o.write('\n')
86
87 # function to be used as model function for the fit.
88 # it takes a set of force field parameters as input and
    performs LAMMPS single point calculations using those
    parameters. outputs the energy for every output file
    produced in the current directory.
89 def func(file_ids, *params):
90
91     # generate a list of parameters to be passed to the
    update_parameters() function
92     par_list = []
93     for par in params:
94         par_list.append(par)
95
96     # write a log of the parameters optimization. the first
    entry will be the initial guess, while the last will
    be the optimized parameters.
97     with open('params.out', 'a') as prm:
98         prm.write(str(par_list))
99         prm.write('\n')
100
101     # update force field parameters in input files.
102     # (during the first iteration it will simply overwrite
    with an identical copy.)
103     update_parameters(par_list)
104
105     # launch LAMMPS jobs
106     input_files = [f for f in os.listdir(directory) if (os.
    path.isfile(f) and f.endswith('.in'))]
107     for i in input_files:
108         to_launch = i.split(".")[0]
109         # XXX modify this command according to the lammps
    script syntax
110         command = "d_lammps_amd" + " " + to_launch + " " + "1
    "
111         subprocess.call(command, shell=True)
112
113     # keep checking if the LAMMPS calculations are finished
    , until they are. how it works: it writes the 'qstat'
    output to a file, then checks for a unique_string in
    all the jobs running. the unique string identifies the
    input files used for the fitting procedure.
114     unique_string = 'dfit' # XXX
115     while True:
116         jobs_still_running = 0
117         subprocess.call('qstat -f -1 > qstat.check', shell=
    True)

```

```

118     with open('qstat.check', 'r') as check:
119         check_list = check.read().splitlines()
120     for line in check_list:
121         if line.strip().startswith('Job_Name'):
122             job_name = line.split('=')[-1].strip()
123             if unique_string in job_name:
124                 jobs_still_running += 1
125     if jobs_still_running == 0:
126         break
127
128     # return output energies as list. also log the energies
129     # in a file.
130     energy_list = parse_outputs()
131     energies_out = []
132     print('')
133     print('energies:')
134     for en in energy_list:
135         print(en)
136         energies_out.append(en)
137     print('')
138     with open('energies.out', 'a') as en_out:
139         for en in energies_out:
140             en_out.write(str(en))
141             en_out.write('\n')
142         en_out.write('-----\n')
143     return energies_out
144
145 # -----
146
147
148 # perform the fit and write optimized parameters to screen.
149 popt, pcov = curve_fit(func, quantum_x, quantum_y, p0 =
150     initial_guess, method='trf', diff_step=0.01, bounds
151     =((0.01, 2.0), (np.inf,np.inf)))
152
153 print('the optimized parameters are', popt)
154 print('the covariance is', pcov)
155
156 subprocess.call('rm qstat.check', shell=True)
157
158 print('output files written: params.out, energies.out')

```


List of publications

- Towards a traceable enhancement factor in surface-enhanced Raman spectroscopy (DOI: 10.1039/d0tc04364h)
- Predicting the Conformation of Organic Catalysts Grafted on Silica Surfaces with Different Numbers of Tethering Chains: The Silicopodality Concept (DOI: 10.1021/acs.jpcc.1c06150)
- Hyper-Cross-Linked Polymers for the Capture of Aromatic Volatile Compounds (DOI: 10.1021/acsapm.9b01000)
- Optimizing the relaxivity at high fields: systematic variation of the rotational dynamics in polynuclear Gd-complexes based on the AAZTA ligand (DOI: 10.1039/d1qi00904d)
- Structure and Stability of 7-mercapto-4-methylcoumarin SAM on Gold: an Experimental and Computational analysis (*In preparation for submission.*)

