UNIVERSITAT POLITÈCNICA DE CATALUNYA

FACULTAT DE FÍSICA, DEPARTAMENT DE FÍSICA APLICADA

PROGRAMA DE DOCTORAT DE *FÍSICA COMPUTACIONAL I APLICADA*

AB INITIO MOLECULAR DYNAMICS STUDY OF THIOLATE-PROTECTED GOLD CLUSTERS AND THEIR INTERACTION WITH BIOMOLECULES

Tesi presentada per obtenir el títol de Doctor per la Universitat Politècnica de Catalunya per part d'en

Víctor Rojas-Cervellera

realitzada al Parc Científic de Barcelona i al Departament de Química Orgànica de la Facultat de Química de la Universitat de Barcelona, sota la direcció de la

Dra. Carme Rovira Virgili

Professora d'Investigació ICREA, i comptant amb la tutoria de la

Dra. Elvira Guàrdia Manuel

Professora del Departament de Física i Enginyeria Nuclear

de la Universitat Politècnica de Catalunya.

Barcelona, Juny 2015

Abstract

Thiolate monolayer-protected gold clusters (AuMPCs) have been attracted much interest in the last years. The fact that AuMPCs can form conjugates with biomolecules allow scientist to use them as carriers of chemotherapeutic drugs, with promising applications in a broad range of diseases, including cancer and neurodegenerative diseases. However, little is known on the molecular mechanism leading to formation of AuMPCs and its reactivity towards proteins. In this thesis, we use *ab initio* molecular dynamics simulations to unravel the mechanism of formation of AuMPCs from a neutral gold cluster and thiol molecules. Afterwards, we uncover the mechanism of the ligand-exchange reaction of a peptide-protected AuMPC with an antibody. Finally, we model the enzymatic reaction of an enzyme responsible for the synthesis of oligosaccharides, α-1,3-glycosyltransferase, as a first step to study the glycosyl transfer reaction on more complex sugar-protected gold nanoparticle systems. A better understanding of the enzymatic mechanism will help the rational design of inhibitors in order to treat diseases where glycosyltransferases are important.

Keywords: thiolate monolayer-protected gold clusters, staple motif, glycosyltransferases, carbohydrates, *ab initio* molecular dynamics, density functional theory, Car-Parrinello molecular dynamics, metadynamics, HOMO-LUMO gap.

Preface

Thiolate monolayer-protected gold clusters (AuMPCs) are being used in various biological and biomedical applications due to their unique physical and chemical properties. The fact that gold-sulphur bonds are very stable enables the binding of biomolecules in the surface of gold clusters through a cysteine, an amino acid that contains a thiol group (SH). Specific AuMPCs-peptide conjugates can cross the blood-brain barrier without altering its integrity, opening the door for the treatment of pathologies related to the central nervous system, such as Alzheimer or Parkinson. Moreover, AuMPCs represent an alternative to the traditional fluorescence-based biosensors, due to their optical properties and its ability to bind specific antigens when certain AuMPCs-antibody conjugates are used. A brief introduction of these biological and biomedical applications can be found in Chapter I, as well as a description of the main synthetic methods of AuMPCs and an analysis of the AuMPC's structures.

Several synthetic approaches based on the reduction of gold salts have been proposed to synthesize AuMPCs. In 1951 Turkevich and co-workers used sodium citrate for the reduction of chloroauric acid. In 2002 a novel synthetic method was proposed, named solvated metal atom dispersion method. In this method, neutral gold atoms were mixed with alkanethiols, resulting in the formation of AuMPCs, and molecular hydrogen was detected. This finding, together with the first crystallization and X-ray structure determination of $Au_{102}(SR)_{44}$ by Jadzinsky *et.al.*, triggered a debate in the field, since the protons that were initially present in alkanethiols were not found in the AuMPC structure. One of the main goals of the present thesis is to elucidate where the alkanethiol hydrogens go during the formation of the AuMPC. To this aim, *ab initio* metadynamics have been used to unravel the molecular mechanism of the formation of AuMPCs departing from neutral gold clusters and alkanethiols (Chapter III).

Key to the usage of AuMPCs as biosensors is the better knowledge of their optical properties. The energetic difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), i.e. the HOMO-LUMO gap, is a physical parameter related with optical properties. Density Functional Theory (DFT) is extensively used to obtain a theoretical value of the HOMO-LUMO gap, although it is known to severely underestimate it with respect to the experimental values. Nevertheless, recent computational studies using DFT have reported values of the HOMO-LUMO gap of AuMPCs in a very close agreement with the experimental ones. However, a simplified model of the real system was used, raising the question whether the agreement between the theoretical and the experimental values is fortuitous. Our goal is to obtain HOMO-LUMO gap values using the whole experimental systems, *i.e.* peptides as the protecting ligands of the gold core and

water as solvent (Chapter IV) to demonstrate that only a realistic model, and not only the use of appropriate DFT functionals, can lead to values comparable to the experimental ones.

In a first step for the understanding of the reactivity of AuMPCs towards proteins, in Chapter V we modelled the binding of AuMPC to the N-terminal end of the anti-influenza N9 neuraminidase NC10 antibody. This process, known as ligand exchange reaction, is used to label proteins with gold clusters, since reducing agents cannot be used with certain biomolecules. Our results show that the neighbouring amino acids of the cysteine that binds to the gold cluster play an essential role in the ligand exchange reaction.

Finally, we investigated the mechanism of the enzymatic reaction of a glycoprotein, α-1,3-glycosyltransferase. In recent years, our group has investigated the molecular mechanism of one family of glycosyltransferases (GTs), providing its catalytic itinerary. In this thesis we extend this study to another family of GTs to elucidate whether or not a common molecular mechanism operates for GTs. This study represents one step towards the modelling of the more complex glycosyl transfer reaction on glycosyltransferases immobilized by gold nanoparticles, a promising technique for the development of automated glycosynthesis.

The theoretical methods used along this thesis are detailed in Chapter II.

Acknowledgments

Aquesta tesi no hauria estat possible sense l'ajuda de tota la gent que ha estat al meu costat durant aquests anys. Voldria agrair de manera especial:

A la *Carme*, per la seva confiança i els seus consells. Gràcies per la teva dedicació i per tot el que m'has arribat a ensenyar durant aquests anys. Sempre t'estaré agraït.

I would also thank to *Jaakko* the opportunity he gave me to visit his group in Tampere, Finland. Thanks for the great scientific ideas that you had. This thesis is in part thanks to you.

A tota la gent del laboratori: al *Fermín*, que em va introduir en el món de la Química Computacional quan encara estava acabant la carrera. A l'*Albert*, per tenir tanta paciència amb mi. He après moltíssimes coses de tu. *Xevi*, *Pietro*, *Marc M*., *Javier*, *Oriol*, *Marc W*., *Santi*, *Alba*,... guardo molt bon record de tots vosaltres, sobretot de totes les tertúlies a la hora del cafè. *Sogol*: you are the best English teacher of the world. Morageb bash. *Javi*: amb tu he compartit tots els anys de tesi; ets un gran doctor, un gran company i un gran amic. Ens queden moltes maratons per córrer junts! *Mertxe*: gràcies per ser tan bona mestra. *Lluís*: He intentat ensenyar-te tot el que he pogut, però no se qui ha après més de qui.

A mi grupo de amigos: *Javi*, *Jonathan*, *Nacho*, *Oriol*, *Raúl*, *Samuel* y *Soto*. Sin duda, sois parte de esto.

A toda mi familia, pero en especial a mis padres, que siempre han confiado en mí. Gracias por ayudarme en todo momento, vuestro apoyo ha sido esencial.

A l'*Ariadna*: amb tu he compartit moltes coses durant aquests anys, que al teu costat han estat inoblidables. Moltes gràcies per ajudar-me dia a dia, ets la millor companya que es pot tenir. Aquesta tesi al teu costat ha estat molt més fàcil.

Table of Contents

List of Abbreviations and Symbols

α3GalT	α-1,3-galactosyltransferase
Å	Angstrom
AuC	gold cluster
AuMPC	thiolate monolayer-protected gold cluster
AuMPN	thiolate monolayer-protected gold nanoparticle
a.u.	atomic units
AIMD	*ab initio* molecular dynamics
BOMD	Born-Oppenheimer molecular dynamics
CP	Car-Parrinello method
CPMD	Car-Parrinello molecular dynamics
CV	collective variable
DFT	density functional theory
D-RESP	dynamically generated restrained electrostatic potential derived charges
FES	free energy surface
fs	femtoseconds
G	Gibbs free energy
GH	glycosyl hydrolase
GT	glycosyltransferase
H	Hartree
K	Kelvin
MD	molecular dynamics
MM	molecular mechanics
MTD	metadynamics
nm	nanometer

ns	nanosecond
OtsA	trehalose 6-phosphate synthase
PBE	Perdew-Burke-Ernzerhof exchange-correlation functional
ps	picoseconds
QM	quantum mechanics
QM/MM	quantum mechanics / molecular mechanics
RMSD	root mean square deviation
Ry	Rydberg
SASA	solvent accessible surface area
TS	transition state

Chapter I - Introduction

Introduction

Thiolate Monolayer-Protected Gold Clusters and Nanoparticles

Historical Overview

Gold is historically one of the most valuable and coveted precious metals in the world because of its distinct colour, is easy handling and minting (it is malleable and ductile), easy smelting, non-corrodibility and non-reactivity towards other elements. All these unique properties make it the preferred metal in coinage, jewellery, industry or electronics. Alchemy, as many sciences such as Chemistry, Physics or Medicine were known from the Ancient Egypt until the 18th century, has always a special relationship with gold. The main goal of alchemists was to produce gold starting from other elements, such as lead, by its interaction with a legendary alchemical substance: *the philosopher's stone*. Moreover, during the *Black Death* (1346-1353), alchemist thought they could cure the sick by making them eat finely powdered gold. In particular, potions made from gold were known as the so-called *Elixir of Life*. Although they never succeeded in his attempts, alchemists promoted an interest in what can be done with metals and other substances, leading the foundation for today's chemistry.

The behaviour of metals when they are reduced to the nanoscale differs from those of the corresponding bulk metal. Michael Faraday (Newington, 1791 – London, 1867), the famed physicist and chemist, was the first to report what later came to be called *metallic nanoparticles*. In 1857, Michael Faraday gave a Bakerian lecture[i] to the Royal Society in London entitled "*Experimental relations of gold (and other metals) to light*". The main topic of his lecture was the ruby colour produced by a solution of fine particles of gold prepared by reducing a solution of gold salts: "*If the solution be weak and the phosphorus clean, part of the gold is reduced in a exceedingly fine particles, which becoming diffused, produce a beautiful ruby fluid*".[1] This is, almost certainly, the first reference to a synthesis of what today is known as *gold clusters* or *gold nanoparticles*. Nearly 100 years later, Turkevich *et. al.*[2] used microscopic investigations to reveal that the ruby-coloured colloids made by Faraday's preparative routes produce particles of gold with average sizes in the 6 ± 2 nm range.[3] Faraday's original gold colloids were very stable, only to be destroyed in the bombardment of London during the second World War.[2a]

[i] Applied to prize lectures by fellows of the Royal Society on natural history or experimental philosophy.

Synthesis of Monolayer-Protected Gold Clusters and Nanoparticles

Many approaches based on the use of gold salts and different reducing agents were proposed for the preparation of colloidal gold solutions since Faraday's work. However most of them lead to preparations with limited stability and broad particle distribution.[4] Three that succeeded to produce colloidal solutions with high stability and uniform particle size have one thing in common: the presence of a ***monolayer of a capping agent that protects the gold core***.

- ***Turkevich method (1951)***:[4] based on the reduction of chloroauric acid ($HAuCl_4$) by sodium citrate at 100 ºC. It produces a colloidal solution with excellent stability and a uniform particle size of 20 nm of diameter. The citrate acts as both a reducing agent and a capping agent, *i.e.* as a stabilizing ligand. The main drawback of this synthesis is that it is unable to make particles smaller than 14 nm. Moreover, if citrate is not desired as capping agent, ligand exchange reactions are necessary to replace citrate by the target ligand.

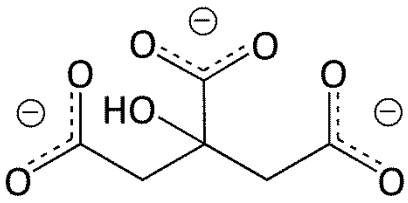

Figure I-1. Chemical representation of citrate, the reducing agent used to produce colloidal gold solutions in the Turkevich method. It also acts as a stabilizing ligand of the gold core.

- ***Brust-Schiffrin method (1994)***:[5] based on the monolayer self-assembly of alkanethiols adsorbed on gold.[6] Brust and co-workers developed a one-step method for the preparation of nanometer-sized gold particles. More precisely, $AuCl_4^-$ was transferred from an aqueous solution to toluene using tetraoctylammonium bromide (TOAB) as the phase-transfer reagent, followed by the reduction with an aqueous sodium borohydride solution ($NaBH_4$) in the presence of alkanethiols. This method produces thiolate-gold metal nanoparticles that can be handled and characterised as chemical compounds (Figure I-2). The nanoparticles are in a range of particle size of 1 to 3 nm. This method has been improved afterwards by using $AuCl_3$ instead of $AuCl_4^-$,[7] or an appropriate mixture of methanol and water (a water content from 9% to 18% (v/v) is optimal for the formation of stable an water-soluble nanoparticles).[8]

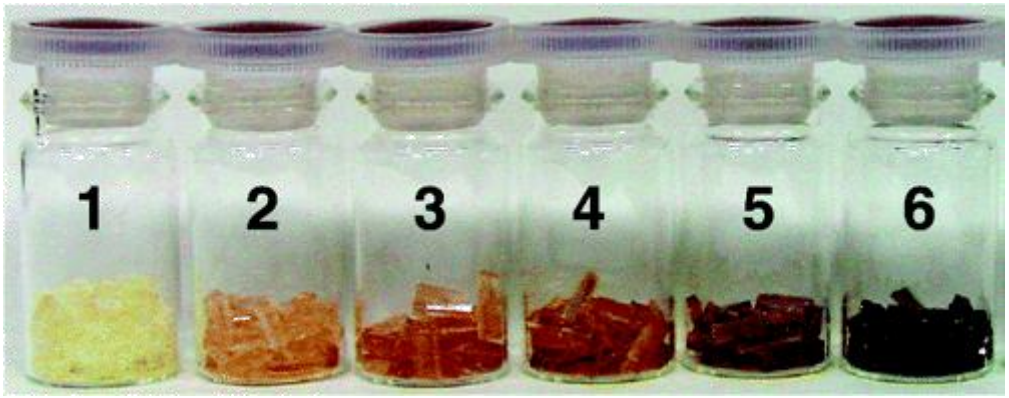

Figure I-2. Appearance of gels containing fractionated clusters of different sizes: $Au_{18}(GSH)_{11}$, $Au_{21}(GSH)_{12}$, $Au_{25}(GSH)_{14}$, $Au_{28}(GSH)_{16}$, $Au_{32}(GSH)_{18}$ and $Au_{39}(GSH)_{23}$, labelled as 1-6 respectively. Reprinted with permission from reference 9. Copyright 2004 © American Chemical Society.

- ***Solvated metal atom dispersion method, SMAD (2002):***[10] based on the vaporization of a metal under vacuum and co-deposition of the atoms with the vapours of a solvent (acetone) on the wall of a reactor cooled to 77 K. The system was subsequently warmed up and mixed with alkanethiols, resulting in the formation of molecular hydrogen[11] (H_2) and gold clusters or nanoparticles coated by alkanethiolates in a diameter range from 2 to 40 nm. The molecular mechanism of formation of gold clusters using this method is studied in detail in Chapter III.

Structure of Thiolate Monolayer-Protected Gold Clusters and Nanoparticles

Up to 2007, little was known about the structure of thiolate monolayer-protected gold clusters and nanoparticles (AuMPCs and AuMPNs). The crystallization and X-ray structure determination of a *p*-mercaptobenzoic acid (*p*-MBA)-protected gold nanoparticle, $Au_{102}(p\text{-}MBA)_{44}$, completely revolutionized the gold field.[12] Up to this date, only theoretical predictions of the structure were available,[13] with different studies predicting both structured[13a, 13c] and distorted[13b] gold cores. The above mentioned crystal structure revealed that $Au_{102}(p\text{-}MPBA)_{44}$ is formed by a central D_{5h}-symmetric Au_{79} gold core with an unstructured protective gold-thiolate layer of composition $Au_{23}(p\text{-}MBA)_{44}$ (Figure I-3).[13c] The $Au_{23}(p\text{-}MBA)_{44}$ layer is formed by 21 protecting units known as ***staple motifs***.[12a] Each staple is known as monomeric [$Au(p\text{-}MBA)_2$], dimeric [$Au_2(p\text{-}MBA)_3$],... depending if are formed by one, two,... gold atoms respectively. More precisely, $Au_{102}(p\text{-}MBA)_{44}$ is formed by 19 [$Au(p\text{-}MBA)_2$] and 2 [$Au_2(p\text{-}MBA)_3$]. The 21 staple motifs are anchored to the gold core by two S-Au bonds (Figure I-3), hence 42 anchoring points. Since only 40 gold atoms are present in the surface of the Au_{79} core, the layer is completed by two cases of double anchoring, *i.e.* one gold atom of the surface bonded to two sulphur atoms at the same time. In this respect, the $Au_{102}(p\text{-}MBA)_{44}$ system can be also described as $Au_{39+40}[Au(p\text{-}MBA)_2]_{19}[Au_2(p\text{-}MBA)_3]_2$ (Figure I-3).

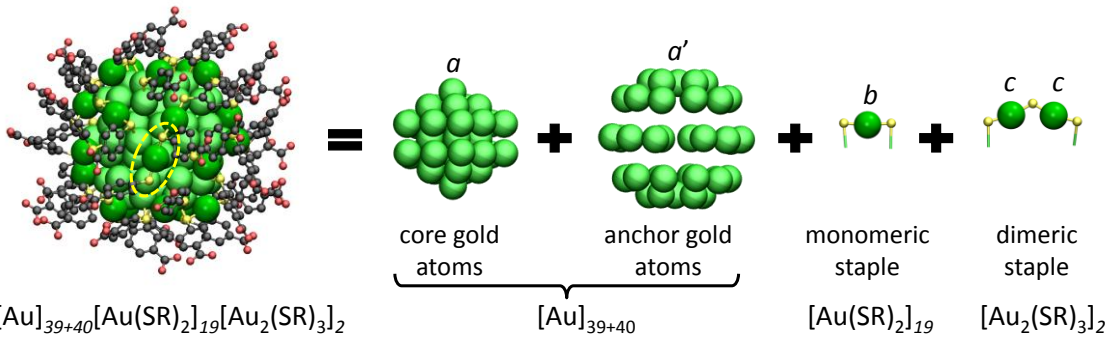

Figure I-3. Molecular structure of Au_{102}(p-MBA)$_{44}$ (p-MBA is p-mercaptobenzoic acid) and its building blocks according to the elegant "divide-and-protect" scheme.[13c] Organic groups attached to the sulphur atoms (R) in the right part of the figure not shown for clarity. Sulphur atoms are in yellow and gold atoms in green. Types of gold atoms: inner core gold atoms (*a*); anchor gold atoms (*a'*); staple gold atoms (*b* and *c*). $[Au]_{39+40}$ represents the gold core, whereas $[Au(SR)_2]_{19}$ and $[Au_2(SR)_3]_2$ are the staple motifs, respectively. The case of double anchoring is highlighted by a yellow dashed line in the full structure (left).

The divide-and-protect[13c] scheme can be applied to all thiolate monolayer-protected gold clusters synthesized so far. Hence, a thiolate-gold cluster formed by *m* gold atoms protected by *n* thiolate groups, *i.e.* $[Au_m(SR)_n]^q$ (being *q* the charge of the system and R an organic ligand),[13c, 14] can also be expressed, within the divide-and-protect scheme, as $[Au]_{a+a'}[Au(SR)_2]_b[Au_2(SR)_3]_c$. In this formula, *a* is the number of inner core gold atoms, *a'* is the number of gold atoms that constitute the anchor of the staple motifs, *b* and *c* are the number of staple motifs, which can be either monomeric or dimeric, respectively (Figure I-3). Higher order motifs have also been characterized, specially protecting small clusters.[15]

By analyzing the known AuMPN structures in the light of the divide-and-protect scheme, we can infer[16] the simple relations that the parameters *a*, *a'*, *b* and *c* must satisfy for a general structure. First, the number of gold atoms in the core ($a + a'$) must be equal to the total number of gold atoms (*m*) minus the number of gold atoms in the staple motifs ($b + 2c$, Figure I-2):

$$a + a' = m - (b + 2c) \quad [1]$$

Second, because each staple motif is anchored to the gold core by two gold atoms, a' should be twice the number of staple motifs:

$$a' = 2(b + c) \quad [2]$$

Third, the number of sulfur atoms should be twice the number of monomeric staple plus three times the number of dimeric staples:

$$n = 2b + 3c \quad [3]$$

According to a spherical jellium model, exceptional stability is associated with the following electron count ("magic numbers"), as discussed by Walter et. al. in reference:[17]

$$n^{elec} = 2,8,18,34,58,92,138, \ldots \quad [4]$$

which correspond to strong electron shell closures in an anharmonic mean-field potential (*i.e.* superatomic orbital view).[17-18] Each gold atom of the AuMPC contributes one electron to the cluster valence shell (n^{elec}). In turn, each alkanethiolate ligand withdraws one electron from the gold core, *i.e.*:

$$n^{elec} = m - n - q \quad [5]$$

Using [1] and [3]

$$n^{elec} = (a + a') - (b + c) - q \quad [6]$$

Therefore, the number of cluster valence electrons corresponds to the number of gold core atoms plus the gold anchor atoms minus the number of staple motifs (adjusted by the charge) (Table I-1).

Two conclusions can be drawn from the above relation. First, each staple motif removes one electron from the gold core, *i.e.* there are *a'*/2 oxidized gold anchor atoms (Au^+) (eq. [2]). Second, gold atoms that constitute the staple motifs (*b* + 2*c*) do not contribute to n^{elec}, thus being oxidized (Au^+).The total number of oxidized gold atoms is thus *a'*/2 + *b* + 2*c*, which, according to [2] and [3], is precisely the total number of thiolates (*n*).

The above rules allow the prediction of the structure of new AuMPCs and AuMPNs, such as $Au_{38}(SR)_{24}$ and $Au_{25}(SR)_{18}^-$. In fact, their structures were first predicted computationally[19] before their X-ray structure were obtained.[20]

The above rules were used to design the smallest thiolate monolayer-protected gold cluster, $Au_4(SCH_3)_2$, which serve as a model to simulate the molecular mechanism of the gold-thiolate cluster formation (Chapter III).

system	*m*	*n*	*q*	n^{elec}	*a*	*a'*	*b*	*c*
$Au_{102}(SR)_{44}$	104	44	0	58	39	40	19	2
$Au_{38}(SR)_{24}$	38	24	0	14	5	18	3	6
$Au_{25}(SR)_{18}^-$	25	18	-1	8	1	12	-	6

Table I-1. Atomic-electron counting for several AuMPCs and AuMPNs according to the divide-and-protect approach and relations [5] and [6].

Considering the size or the number of gold atoms that are forming the metallic core, different categories can be distinguished (Figure I-4):[21]

- ***Gold(I)-thiolate oligomers***: are the smallest systems formed by thiolates and gold atoms (from 1 to 8 Au atoms). They are the precursors of the staple motifs found capping the gold core of AuMPCs and AuMPNs. They have been widely used as therapeutic agents for the treatment of rheumatoid arthritis.[22]
- ***Self-assembled monolayers***: formed by a densely packed thiolate monolayer film on an extended gold surface. They have been used for chemical sensing, corrosion inhibition, lubrication and catalysis.
- ***Thiolate Monolayer-Protected Gold Clusters and Nanoparticles***: pseudo-spherical particles synthesized by the methods discussed in the previous section. This kind of particles lies between the two limiting systems described above. The difference between Clusters and Nanoparticles lies in the size: particles that are smaller/larger than ≈1.5 nm are known as Thiolate Monolayer-Protected Gold Clusters (**AuMPC**)/ Thiolate Monolayer-Protected Gold Nanoparticles (**AuMPN**), respectively.[21, 23] This thesis is focused on AuMPCs.

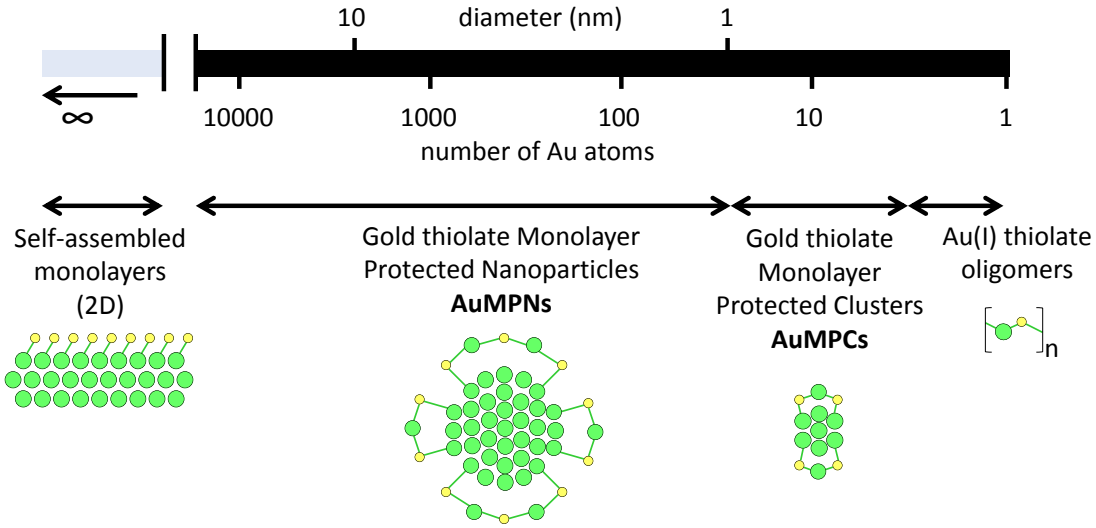

Figure I-4. Different thiolate-gold systems as a function of the number of gold atoms. Gold and sulphur are shown green and yellow, respectively. Figure adapted from reference 21.

Biological Applications of AuMPCs and AuMPNs

Gold nanoparticles are attracting strong interest in the fundamental sciences and are being extensively investigated for their potential use in various biological applications[24] due to their unique properties of suitable size, extensive thermal stability, ease synthesis and functionalization, low toxicity and ease of detection. They also exhibit exceptional optical properties making them capable of producing quantum effects suitable for imaging applications. Drug delivering, diagnosis and treatment of several pathologies related to the central nervous system (CNS) and labelling of biological molecules including proteins, enzymes, DNA, antigens and antibodies, are the most promising applications of AuMPC and AuMPN. As a small taste, two examples deserve to be highlighted here:

Molecular Surgery and D*rug Delivering*

The fact that specific gold-peptide conjugates can cross the blood-brain barrier without altering its integrity opens the door for the treatment of several pathologies, such as those related to the CNS.[25] The local heat delivered by gold nanoparticles can be used as a **molecular surgery** to safely remove toxic aggregates in the treatment of Alzheimer's disease, where unnaturally folded Aβ proteins self-assemble and deposit forming amyloid fibrils and plaques. Kogan *et. al.*[26] linked AuMPN to the peptide H-Cys-Leu-Pro-Phe-Phe-AspNH_2 (Cys-PEP), which contains the sequence H-Leu-Pro-Phe-Phe-AspNH_2 (PEP) that selectively attaches to the Aβ aggregates, forming the conjugated AuMPN-Cys-PEP.[27] When the ratio of AuMPN-Cys-PEP conjugate attached to the Aβ fibrils was adequate, AuMPN were able to onset a pro-inflammatory immune response by the bone marrow derived macrophages, which were unable to recognize Aβ alone.[28] However, if the AuMPN-Cys-PEP:Aβ ratio is too low, the fibril growth cannot be blocked. In these cases, microwave fields (0.1 W) can be applied: AuMPN-Cys-PEP conjugates attached to the fibrils absorbed the radiation and dissipated energy causing disaggregation of the amyloid deposits and aggregates (Figure I-5).[26]

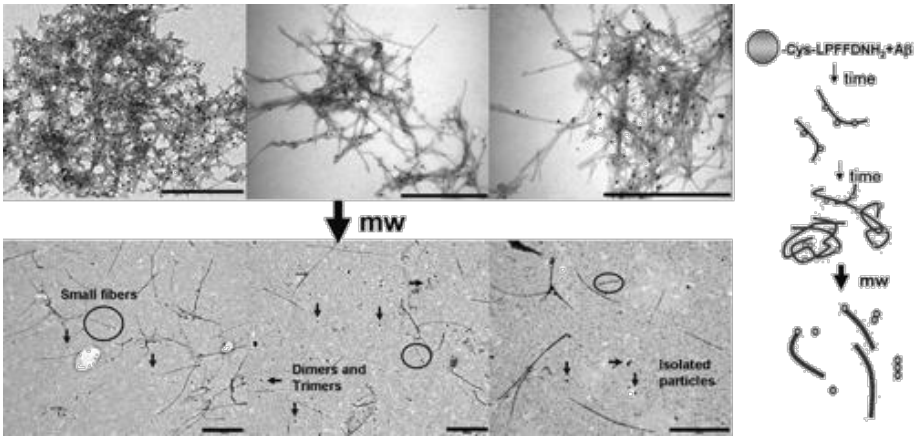

Figure I-5. Electron microscopy of the solutions before and after irradiation. Reprinted with permission from reference 26. Copyright 2006 © American Chemical Society.

Biological Labelling and *Immunoassay*

AuMPNs can be functionalized with antibodies such as single chain fragment variable antibodies (scFv). The scFvs are small heterodimers that are composed of the antibody variable heavy (V_H) and light (V_L) chains connected by a peptide linker that is used to stabilize the molecule. They represent the smallest fraction of the antibody needed for binding to the antigen. AuMPNs functionalized with engineered scFv containing cysteines in its linker region were used to develop a colorimetric immunoassay.[29] Mutated scFv fragments exposing the cysteine residues have been shown to form gold-thiolate bonds. These scFv-Cys stabilized AuMPNs have been developed for detection of a broad range of antigens. For example, they can bind to the viral neuraminidase, a ***glycoside hydrolase*** enzyme that is present in the surface of the influenza virus and is a drug target for the prevention of the spread of influenza infection (Figure I-6). These biosensors involving AuMPNs of size <60 nm were found to provide sensitivity equal to or even better than the traditional fluorescence-based biosensors.

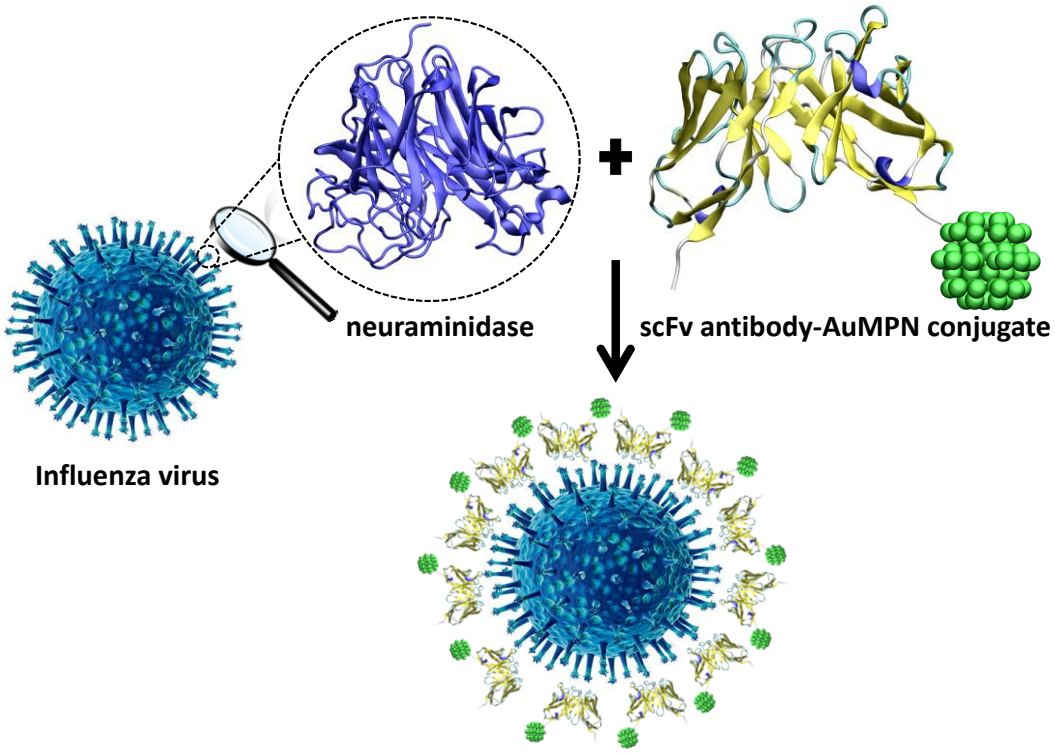

Figure I-6. Scheme of the immunoassay of a scFv antibody-AuMPN conjugate towards the neuraminidase enzyme found in the surface of the influenza virus. The structure of the proteins (neuraminidase and scFv antibody fragment) has been taken from Protein Data Bank, PDB code 1NMB. The Influenza virus image has been taken from http://www.medicalpress.es/.

The two biological applications described above have inspired our study of AuMPCs protected by biological ligands (Chapter IV) and also its interaction with an anti-neuraminidase antibody (Chapter V).

Glycosyltransferases (GTs)

As described above, AuMPCs can be functionalized with antibodies to detect specific antigens. Glycoproteins, i.e. proteins conjugated with carbohydrates, are one of the possible antigens of these antibodies. Among other activities, glycoproteins are in charge of the degradation (glycoside hydrolases, GHs) and synthesis (glycosyltransferases, GTs) of carbohydrates. These enzymes account for 1-3% of the proteins encoded by the genomes of most organisms.[30] GHs and GTs play essential roles in a wide range of biological processes such as control of protein folding, viral infections, genetically inherited disorders, inflammation processes and immune response.[31] Knowing the molecular mechanism of these enzymes is key for the design of compounds that can modulate their activity, such as inhibitors. The reaction mechanism of some GTs is not known yet, being one of the most intriguing and unanswered question in the field of glycobiology. Taking advantage of the group expertise on glycoprotein reactivity (specially GHs and GTs), the molecular mechanism of the synthesis of an oligosaccharide inside a GT is studied in detail in Chapter VI.

Carbohydrates

Carbohydrates are biomolecules that have two main biological functions: energy storage and structural support. From a structural point of view, carbohydrates can be classified in terms of size as ***monosaccharides***, ***oligosaccharides*** and ***polysaccharides*** depending on how many units are forming the biomolecule. Monosaccharides are the most basic units of carbohydrates. In general, they have the chemical formula of $C_x(H_20)_y$, being $x \geq 3$ (see Figure I-7a). Monosaccharides are the building blocks of the more complex saccharides. They exist either as open-chains or forming rings. The six-membered ring formed by five carbon atoms and one oxygen atom is known as pyranose ring. This cyclic structure is the predominant one in aqueous solution, formed from the open-chain structure of the monosaccharide (Figure I-7a). The pyranose ring originates from the intramolecular reaction between the carbon atom (C1) of the aldehyde functional group and an oxygen atom (O5) of one hydroxyl group (Figure I-7a). The formation of the new covalent bond (green bond in Figure I-7a) leads to the appearance of a new asymmetric carbon[ii] (C1), named anomeric carbon, that can has either an α or β configuration, depending on the orientation of the new hydroxyl group (red group in Figure I-7a). Both configurations are interchangeable in aqueous solution, being this process known as mutarotation process. The letter that precedes the name of the carbohydrate is assigned according to the orientation of the asymmetric carbon being further away from the carbonyl group (C1): in a Fischer projection if the hydroxyl is on the right the molecule is a D carbohydrate, otherwise it is an L.

[ii] IUPAC definition: an asymmetric carbon atom, also known as chiral carbon, is a carbon atom that is attached to four different entities (atoms or groups).

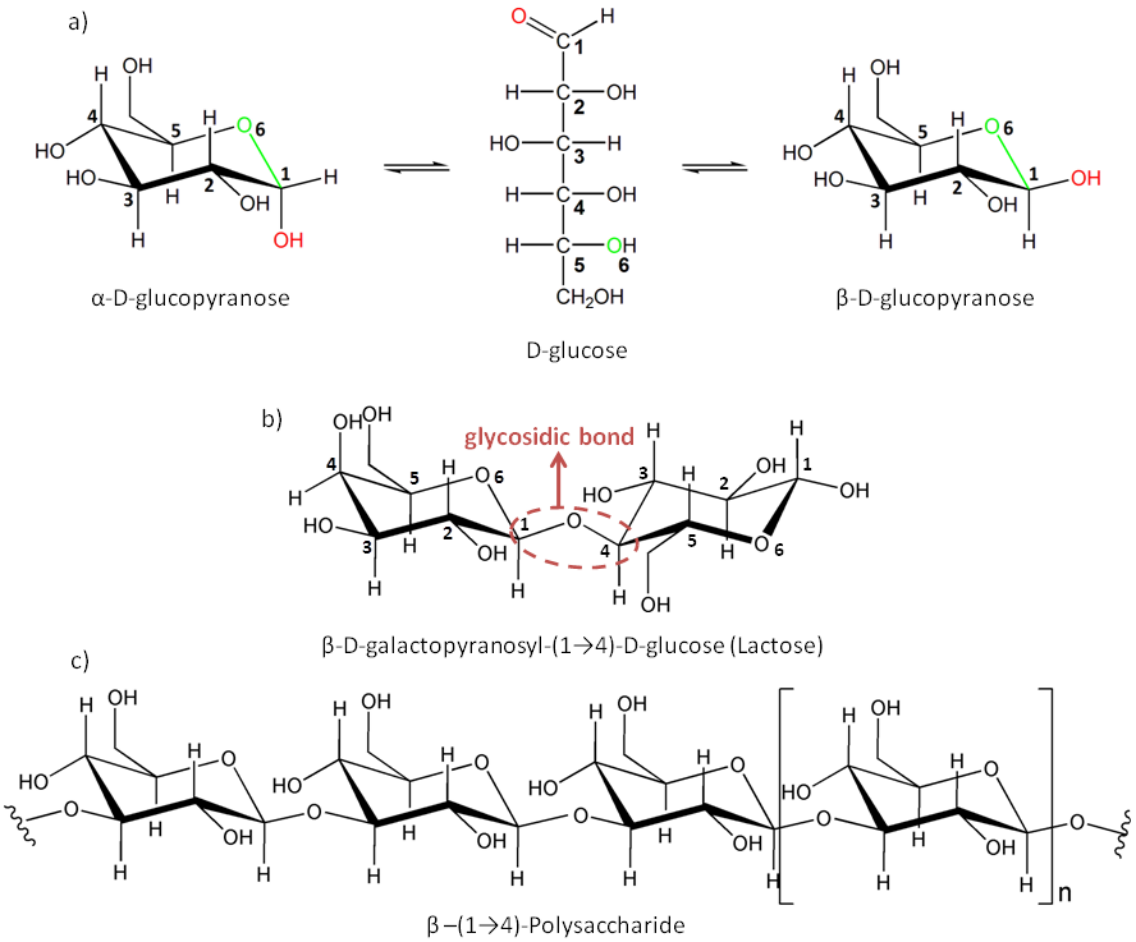

Figure I-7. Chemical representations of carbohydrates, each one corresponding to one of the three groups in terms of size: a) Mutarotation process that leads to the formation of the two cyclic forms of D-Glucose: α-D-glucopyranose and β-D-glucopyranose (monosaccharides), b) lactose (disaccharide) and c) part of a polysaccharide. Each vertex of the ring structures correspond to a carbon atom.

Oligosaccharides are short chains of monosaccharide units joined by characteristic linkages, known as ***glycosidic bonds*** (Figure I-7b). They are named according to the number of saccharide units forming the oligosaccharide: two (disaccharides), three (trisaccharides), four (tetrasaccharides) and so on. Polysaccharides are polymers containing more than 20 monosaccharide units (Figure VI-7c).

Six-membered rings like pyranose are very flexible molecules that can adopt a wide range of conformations. All of them share one common feature: at least four atoms of the ring are on the same plane, being the remaining atoms out-of-plane. Five possible groups of conformations exist depending on which are the four co-planar atoms. These groups are named as chair (C), boat (B), skew-boat (S), half-chair (H) and envelope (E). To indicate which atoms are out-of-plane, superscript or subscript numbers are added to the label depending on whether atoms are up or down with respect to the plane, respectively:[32]

- Chair: the reference plane is formed by two parallel ring sides with the two out-of-plane atoms in opposite positions of the ring, one up and one down. In aqueous solution is the most stable conformation. There are two possible chair conformations: $^{1}C_{4}$ and $^{4}C_{1}$.
- Boat: the reference plane is again formed by two parallel ring sides with the two out-of-plane atoms in opposite positions of the ring, but the two exoplanar atoms are on the same side, both up or down. $^{1,4}B$, $B_{1,4}$, $B_{2,5}$,... are examples of this conformation.
- Skew-Boat: the reference plane includes three adjacent atoms and one non-adjacent atom. $^{1}S_{3}$, $^{3}S_{1}$, $^{2}S_{O}$,... are possible skew-boat conformations.
- Half-chair: the reference plane is defined by four adjacent atoms. $^{4}H_{3}$, $^{O}H_{1}$, $^{1}H_{O}$,... are possible half-chair conformations.
- Envelope: the reference plane is formed by five atoms. ^{3}E, E_{1}, E_{O} are posible envelope conformations.

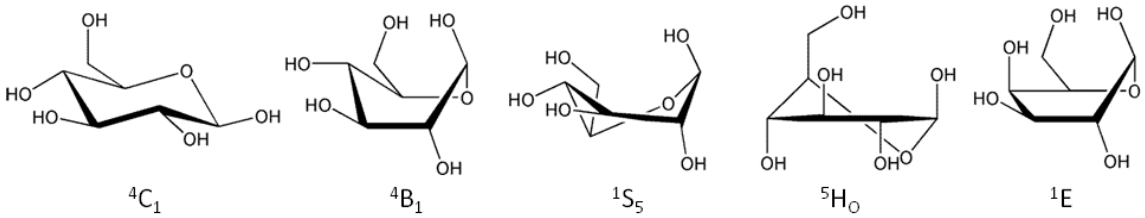

Figure I-8. Representative structures of each of the five different conformation groups of a pyranose ring and its nomenclature. Hydrogen atoms are not shown for clarity. Each vertex of the ring structures correspond to a carbon atom.

Glycosyltransferases

Glycosyltransferases (GTs) are enzymes involved in the synthesis of oligosaccharides, polysaccharides and glycoconjugates.[33] They catalyze the formation of glycosidic linkages by the transfer of a saccharide, typically a monosaccharide, from an activated donor substrate, to an acceptor substrate. The donor substrate is usually a nucleotide-sugar or a sugar-phosphate conjugate, named as Leloir and non-Leloir donors, respectively.[34]

Glycosyltransferases catalyze the glycosyl transfer reaction with either inversion or retention of the donor anomeric stereochemistry.[34] The mechanism of inverting GTs follows a bimolecular nucleophilic substitution (S_N2) reaction in a single displacement step involving a general base (aspartate or glutamate) that deprotonates the incoming acceptor hydroxyl group, increasing its nucleophilicity (Figure I-9). The phosphate group acts as leaving group.

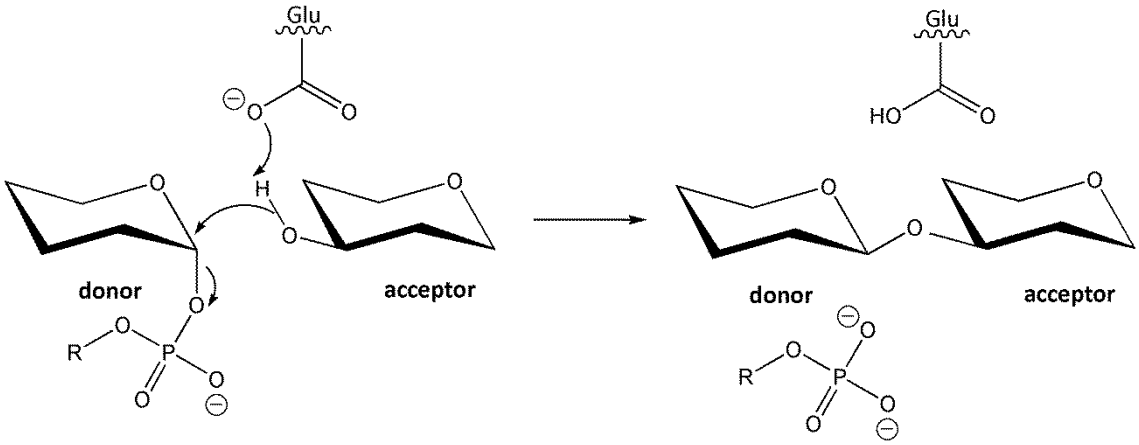

Figure I-9. Reaction mechanism in inverting glycosyltransferases. The reaction proceeds via a S_N2 single-displacement mechanism.

However, the mechanism of retaining GTs is still very controversial. A double-displacement mechanism (two consecutive S_N2 reactions) involving a covalent glycosyl-enzyme intermediate was early proposed.[35] During the first S_N2 reaction (glycosylation step), an aspartate or a glutamate sidechain in the active site makes the first nucleophilic attack on the anomeric carbon of the donor molecule, forming the glycosyl-enzyme intermediate. In the second step, the acceptor molecule performs a second nucleophilic attack (the second S_N2 reaction) on the anomeric carbon of the donor molecule to give overall retention of stereochemistry (Figure I-10).

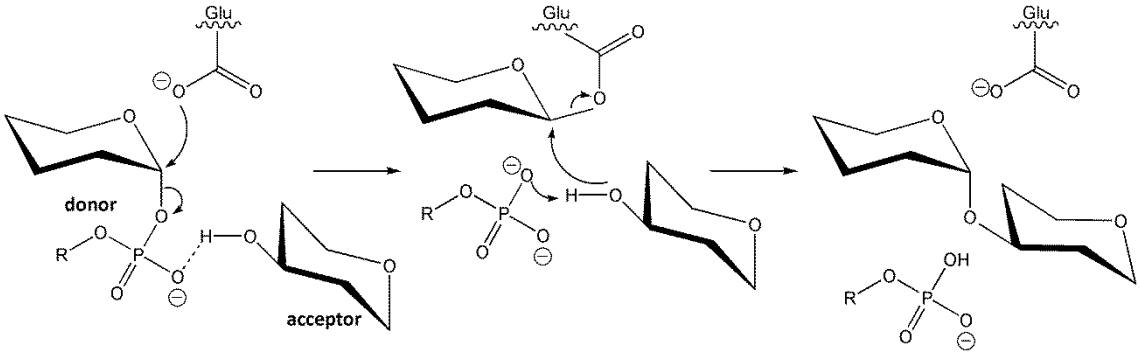

Figure I-10. Double-displacement mechanism proposed for retaining glycosyltransferases.

The observation that most GTs do not have a residue (aspartate or glutamate) properly located to act as a nucleophile in a double-displacement reaction lead to the suggestion of an alternative mechanism,[36] in which the reaction proceeds via a front-face single displacement. In this mechanism, the nucleophilic hydroxyl group of the acceptor performs a nucleophilic attack to the anomeric carbon of the donor at the same side from which the leaving group departs (Figure I-11).

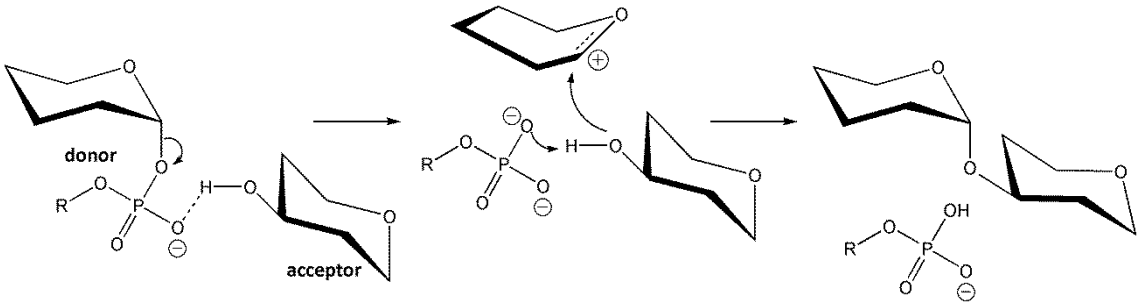

Figure I-11. Front-face reaction mechanism proposed for glycosyltransferases.

It is clear that, if the catalytic reaction follows a front-face single displacement mechanism, the nucleophilic residue is not needed.

A nucleophilic residue in the β face of the donor molecule has only been reported for two GT enzymes from family 6:[37] mammalian α-1,3-galactosyltransferase (α3GalT) and blood-group A and B α-1,3-glycosyltransferase (GTA/GTB). Mutation of the putative nucleophile (glutamate in both enzymes) to alanine inactivates the enzyme, showing the essential role of this residue. On the other hand, some authors have proposed that the glutamate residue is more likely required for proper acceptor substrate orientation.[38] Hence, it remains to be proved that a double-displacement mechanism (Figure I-11) is operative for these GTs.

Concerning the structure, GTs have been classified into more than 90 families based on their amino acid sequence similarities (CAZy database). Enzymes from the same family exhibit the same fold, but enzymes with different donor/acceptor specificities are found within the same family. Despite the large number of GT families

known, structural analysis has shown that they exhibit a limited number of fold types: GT-A, GT-B and GT-C folds.

- GT-A: the catalytic domain consists in a single domain. They display a metal binding motif (DXD motif, being D aspartate and X any amino acid) and are Mg^{+2}/Mn^{+2} dependent GTs.
- GT-B: they are formed by two distinct domains separated by a cleft that binds the acceptor. The GT-B fold is metal-ion independent and do not exhibit a DXD motif.
- GT-C: the folding is similar to the one of GT-A, but it does not contain the DXD motif.

All the above GT folding types types include enzymes that catalyze glycosyl transfer with inversion and retention of the anomeric configuration. Therefore, there is no correlation between fold type and reaction stereochemistry.

In Chapter VI, we study the molecular mechanism of α3GalT, a fold-type A GT with a nucleophilic residue properly located to act as a nucleophile, showing that it can only operate through a double-displacement mechanism, with the formation of a glycosyl-enzyme covalent intermediate.

Objectives

Objectives

The present thesis is aimed at unraveling the molecular mechanism of the formation of thiolate monolayer-protected gold clusters (AuMPCs), as well as the study of their interactions with biomolecules such as small peptides or proteins.

The following objectives have been pursued:

- **Unravel the molecular mechanism of the formation of thiolate monolayer-protected gold clusters departing from neutral gold clusters and alkanethiols.**

  Previous studies revealed that the protons of the incoming alkanethiols were not present in the structure of AuMPCs. Furthermore, molecular hydrogen was detected during the formation process. It would be interesting to elucidate the molecular mechanism of AuMPCs formation to determine where the alkanethiol hydrogens go and how they form molecular hydrogen.

- **Study the effect of the water solvent and the ligands on the structural and electronic properties of a gold cluster.**

  Recent computational studies using density functional theory (DFT) have reported values of the HOMO-LUMO gap of AuMPCs in very close agreement with the experimental ones, although DFT is known to underestimate the HOMO-LUMO gap. Our goal is to find the reasons of this surprising agreement.

- **Model the complex between a protein and an AuMPC.**

  It is known that not all proteins are able to attach to AuMPCs. Our aim is to investigate the factors that influence the binding process for the particular case of the NC10 antibody of the neuraminidase glycoprotein.

- **Uncover the molecular mechanism of the ligand exchange reaction for an AuMPC towards the NC10 antibody.**

  A recent experimental study revealed that AuMPCs can form covalent linkages with variants of the anti-influenza N9 neuraminidase NC10 antibody. We plan to elucidate the molecular mechanism of the ligand exchange reaction for a prototype AuMPC, finding out which is the preferred ligand to be substituted and the reasons behind the lack of activity with certain protein variants.

- **Unravel the molecular mechanism of the enzymatic reaction of a glycoprotein.**

  Proteins with biotechnological interest can be immobilized by AuMPCs. Previous studies on the group elucidated the molecular mechanism of glycosyltransferases (GTs). Our aim is to elucidate whether or not a common molecular mechanism operates for GTs that retain the configuration of the anomeric carbon.

Chapter II – Computational Methods

Computational Methods

Molecular dynamics

Introduction

A set of atomic coordinates, which may be obtained from experimental data, from a theoretical model or a combination of both, gives us a static view of the studied system. If one wants to capture the dynamics of that system, a successive set of atomic coordinates should be obtained. In molecular dynamics (MD) simulations, successive configurations for a given initial structure of a system are generated by integrating Newton's equations of motion. If all the static pictures of the given system are placed one after the other, the result is a trajectory that captures how the positions of the particles vary with time. An analogy with a movie can be done here: each frame of the movie correspond to one set of atomic coordinates, being a static representation. One second of a standard movie is formed by 24 of that frames, i.e. by 24 static representations, each one with a different position of the actors. In our case, atoms play the role of actors. The study of that trajectory is useful to obtain valuable amount of information concerning structural and dynamic properties, such as molecular geometries, energies, binding processes, reactivity and so on. All the programs that perform MD simulations are based in the following scheme:

1. **Initialize the system**, selecting the initial positions and velocities.
2. **Compute the forces** that are acting over all the atoms.
3. **Integrate the Newton's equations of motion** to obtain new positions and velocities. This step and the previous one make up the core of the simulation. They are repeated until we have computed the time evolution of the system for the desired length of time.

Initialize the System

The initial configuration of the system is normally obtained from experimental data, usually from X-ray experiments. For example, when proteins are simulated, the standard way to obtain an initial structure is to download it from a database, usually the Protein Data Bank[39] (PDB, www.rcsb.org). In some other cases initial coordinates can be obtained from theoretical models.

Next step is to assign initial velocities for all the atoms (N) of the system. Velocities are related with temperature through the equipartition equation over all the degrees of freedom (N_f). In particular, for the average kinetic energy per degree of freedom, we have the following relation between α component of the velocity of a given particle i and the temperature, T:

$$\langle \frac{1}{2N_f} mv^2 \rangle = \frac{1}{2} k_B T \quad \textbf{(Eq.II-1)}$$

which is used to define the instantaneous temperature, *T(t)*, in a simulation:

$$T(t) \equiv \sum_{i=1}^{N} \frac{m_i v_{\alpha,i}^2(t)}{k_B N_f} \quad \textbf{(Eq.II-2)}$$

Hence, the assignment of the initial velocities is normally achieved by selecting random velocities from a Maxwell-Boltzmann distribution at the temperature of interest. The Maxwell-Boltzmann distribution gives the probability that an atom *i* of mass m_i has a velocity $v_{\alpha,i}$ at a temperature *T*:

$$p(v_{\alpha,i}) = \left(\frac{m_i}{2\pi kT}\right)^{1/2} e^{\left(-\frac{m_i v_{\alpha,i}^2}{2kT}\right)} \quad \textbf{(Eq.II-3)}$$

which means that choosing the working temperature of the system is the same as choosing the velocities that atoms should have.

Compute the Forces

This part is the most time-consuming part of all MD simulations: the calculation of the force acting on every particle. Forces can be computed using a potential function that is different for each type of MD simulation. As an example, classical methods relay on force field equations where the energy depends only on the atomic coordinates

$$E = f(r_{\alpha,i}) \quad \textbf{(Eq.II-4)}$$

Considering Newton's equations of motion, forces are computed as the minus derivative of the energy over the atomic positions, and the acceleration that one particle experiences at each step is proportional to the force acting upon it

$$-\frac{\partial E}{\partial r_{\alpha,i}} = f_{\alpha,i} = m_i \cdot a_{\alpha,i} \quad \textbf{(Eq.II-5)}$$

Integrate the Newton's Equations of Motion

The integration algorithm can be derived from a Taylor expansion of the coordinate of a particle at time *t*,

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\dddot{r} + \mathcal{O}(\Delta t^4) \quad \textbf{(Eq.II-6)}$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\dddot{r} + \mathcal{O}(\Delta t^4) \quad \textbf{(Eq.II-7)}$$

being Δt the time step, which should be chosen to ensure that the variation in the potential energy function of the atoms within each time step is small. When simulating flexible molecules, a useful guide is that the time step should be approximately one-

tenth of the shortest period of motion. The highest frequency vibrations are due to bond stretches (compression/extension of bonds), especially those involving hydrogen atoms. Carbon-hydrogen bonds vibrate with a period of approximately 10 fs, so a time step of 1 fs is a reasonable value simulating molecules.[40]

Summing equations II-6 and Eq. II-7, we obtain

$$r(t+\Delta t) \approx 2r(t) - r(t-\Delta t) + \frac{f(t)}{2}\Delta t^2 \quad \textbf{(Eq.II-8)}$$

where the new position is estimated with an error that is of order Δt^4. The above is known as the Verlet algorithm,[41] where the velocity is not used to compute the new position. However, the velocity can be derived from knowledge of the trajectory:

$$v(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} \quad \textbf{(Eq.II-9)}$$

Variations of the Verlet algorithm have been developed. The simplest among these is the so-called leap-frog algorithm, which is currently implemented in the AMBER software. In this algorithm, first the velocities are calculated at half integer time steps (t + 1/2 Δt) and then these are used to calculate the position at time t + Δt

$$v\left(t+\frac{1}{2}\Delta t\right) = v\left(t-\frac{1}{2}\Delta t\right) + a(t)\Delta t^2 \quad \textbf{(Eq.II-10)}$$

$$r(t+\Delta t) = r(t) + v\left(t+\frac{1}{2}\Delta t\right)\Delta t \quad \textbf{(Eq.II-11)}$$

Molecular Mechanics (MM) Simulations

MM methods ignore the electronic degrees of freedom and calculate the energy of the system using functions of the atomic coordinates (Eq. II-4). Simple functions are used to describe processes such as the stretching of bonds, the bending of angles (opening and closing) and rotation of dihedral angles. Electrons are not considered, but point charges, localized on each nuclear centre, are used to reproduce the electrostatic properties of each atom. To calculate the total energy of the system for a certain atomic configuration, a force field is needed. A force field is a set of equations employed to calculate the different contributions to the energy, along with the values of structural and other adjustable parameters, such as atomic charges. These parameters are obtained by fitting a potential function with experimental data and high level quantum chemical calculations performed on small molecules. In this thesis, we have used the AMBER force field,[42] where the energy function is

$$E = E_{bonds} + E_{angles} + E_{dihedrals} + E_{VdW} + E_{electrostatic} =$$

$$= \sum_{bonds} k_b \cdot \left(r - r_{eq}\right)^2 + \sum_{angles} k_\theta \cdot \left(\theta - \theta_{eq}\right)^2 + \sum_{dih.} \frac{V_n}{2} \cdot [1 + cos(n\varphi - \alpha)] +$$

$$+ \sum_{i<j} \frac{A_{i,j}}{R_{i,j}^2} - \frac{B_{i,j}}{R_{i,j}^6} + \sum_{i<j} \frac{q_i \cdot q_j}{\varepsilon \cdot R_{i,j}} \quad \textbf{(Eq.II-12)}$$

Two different terms can be observed, whether they described interactions of bonded (E_{bonds}, E_{angles} and $E_{dihedrals}$) or non-bonded (E_{VdW} and $E_{electrostatic}$) atoms. E_{bonds} represent the energy of the vibration of a bond between two atoms. E_{angles} represent the bending of the angle formed between three bonded atoms. Both, E_{bonds} and E_{angles} are modeled by harmonic potentials where r_{eq} or θ_{eq} refer to the equilibrium bond length and angle, respectively, and k_b or k_θ are the harmonic constants. $E_{dihedrals}$ is a torsion potential that captures the energy changes because of the bonds rotation, being V_n the torsion barrier and α the phase. Interatomic interactions between atoms that are non-bonded are modeled through the Van der Waals interactions (E_{VdW}) and an electrostatic interaction ($E_{electrostatic}$). E_{VdW} is described by a Lennard-Jones potential, which contains an attractive term ($1/R^6$) and a repulsive term ($1/R^{12}$). $E_{electrostatic}$ is calculated as a sum of pair interactions between atomic point charge using Coulomb's law.

Quantum Mechanics (QM) Simulations

The way how the energy is calculated in the MM methods does not allow the study of chemical reactions, where atomic rearrangements should be taken into account. Moreover, electrons are not taken explicitly, being also a drawback. Inclusion of electrons requires a quantum mechanical description of the system. The starting point in quantum mechanics is the full, time-dependent form of the Schrödinger equation:

$$\left\{-\frac{\hbar}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + v\right\}\psi(r,t) = i\hbar\frac{\partial\psi(r,t)}{\partial t} \quad \textbf{(Eq.II-13)}$$

The equation above refers to a single electron of mass *m* moving through the space at time *t* under the influence of an external field *v*, which in our case is the electrostatic potential due to the atoms of a certain molecule. $\hbar$ is the Planck's constant divided by 2π and ψ is the wavefunction, which characterizes the particle's motion. If the external potential is independent of time, then the wavefunction can be written as the product of a spatial and a temporal part:

$$\psi(r,t) = \psi(r) \cdot T(t) \quad \textbf{(Eq.II-14)}$$

Hence, the Schrödinger equation can be written in a time-independent form:

$$\left\{-\frac{\hbar}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + v\right\}\psi(r) = E\psi(r) \quad \textbf{(Eq.II-15)}$$

where *E* is the total energy. The left-hand part of the above Schrödinger equation is the Hamiltonian operator, *H*:

$$H = -\frac{\hbar}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{Ze^2}{4\pi\varepsilon_0 r} \quad \textbf{(Eq.II-16)}$$

The Hamiltonian operator has two terms, the kinetic and the potential energy. When the system is composed of several nuclei and electrons, the Hamiltonian has an additional term, the electron-electron potential, v_{ee}, due to the electron's repulsion.

The wavefunction depends on both the electronic and the nuclear degrees of freedom, but the electrons are by orders of magnitude lighter than the nuclei. Within the Born-Oppenheimer approximation, nuclear and electronic degrees of freedom could be separated and the problem is reduced to the calculation of the energy of the system for a given nuclear configuration.

Density Functional Theory (DFT)

DFT provides a framework to obtain the total energy of a polyatomic system given their atomic coordinates. In 1964, Hohenberg and Kohn demonstrated that the ground state energy of a system of interacting electrons subject to an external potential *v* is a unique functional of the electron density:[43]

$$E^{DFT} = \min_{\rho(r)} E[\rho(r)] \quad \textbf{(Eq.II-17)}$$

where the energy functional can also be expressed in terms of single-electron Kohn-Sham (KS) orbitals $\psi_i(r)$:[44]

$$E^{DFT} = \min_{\{\psi_i\}} E^{KS}[\{\psi_i(r)\},\{R_I\}] \quad \textbf{(Eq.II-18)}$$

where orbitals are defined as

$$\psi_i(r) = \sum_j c_j \varphi_{ij}(r) \quad \textbf{(Eq.II-19)}$$

It is computationally convenient to use plane-waves (PW) as the basis set in which the KS orbitals are expanded. Hence the Kohn-Sham orbitals are expressed as

$$\psi_i(r) = \frac{1}{\sqrt{\Omega}}\sum_{G}^{G_{max}} c_{ij} exp\big(iG_j(r)\big) \quad \textbf{(Eq.II-20)}$$

where Ω is the volume of the cell, *G* is the planewave momentum, c_{ij} are the coefficients of the basis set expansion, and G_j are the reciprocal vectors (periodic boundary conditions are assumed). PW basis sets are denoted by an energy value E_{cut}, which is related to the maximum *G* value of the PW expansion (G_{max}). The number of planewaves N_{PW} can be approximated as

$$N_{PW} \sim \frac{\Omega}{6\pi^2} E_{cut}^{3/2} \quad \textbf{(Eq.II-21)}$$

Compared to the popular Gaussian functions, PWs are not centered at the atoms but extend throughout all the space, and thus they do not suffer from basis set superposition error (BSSE). However, PWs require the use of pseudopotentials to describe the effect of the core electrons.

Pseudopotentials are usually derived from all electron atomic calculations and several types have been proposed to date. Here, we used norm-conserving pseudopotentials derived with the Martins-Troullier method.[45]

Car-Parrinello Molecular Dynamics (CPMD)

During the middle 80's, Car and Parrinello (CP) developed a method that combines MD and DFT, in which the electronic and nuclear degrees of freedom evolve simultaneously according to a modified set of classical equations of motion. To do that, a fictitious electron mass, μ, is assigned to the electronic Kohn-Sham orbitals $\{\psi_i(r)\}$. The CP method is based on the description of N interacting atoms through the following Lagrangian:

$$\mathcal{L} = E_N^{kin} + E_{el}^{kin} - E^{KS} + \sum_{i,j} \Lambda_{ij} \left(\int dr \cdot \psi_i^*(r) \cdot \psi_j^*(r) - \delta_{ij} \right) \quad \textbf{(Eq.II-22)}$$

where E_N^{kin} is the classical kinetic energy of the nuclei, E_{el}^{kin} is the "fictitious" kinetic energy associated with the electronic subsystem $\{\psi_i(r)\}$ of electronic mass μ:

$$E_{el}^{kin} = \sum \mu \int \delta r |\psi_i(r)|^2 \quad \textbf{(Eq.II-23)}$$

and Λ_{ij} are the Lagrangian multipliers that impose the orthonormality constraints between orbitals along the simulation. Since the Lagrangian is expressed according to the DFT formalism, E^{KS} corresponds to the potential energy (Kohn-Sham energy). The total energy in the Car-Parrinello method is a constant of motion and is given by the following equation:

$$E_{tot}^{CP} = E_N^{kin} + E_{el}^{kin} + E^{KS} \quad \textbf{(Eq.II-24)}$$

And the equations of motion are

$$\mu \ddot{\psi}_i = \frac{\partial E^{KS}}{\partial \psi_i^*} + \sum_j \Lambda_{ij} \psi_j(r) \quad \textbf{(Eq.II-25)}$$

$$M_N \ddot{R}_N = \frac{\partial E^{KS}}{\partial R_N} \quad \textbf{(Eq.II-26)}$$

The integration of equations II-25 and II-26 provides the time evolution of not only the atomic positions $\{R_I(t)\}$, but also the KS orbitals $\psi_i(r,t)$. In practice, since orbitals are expanded in a basis set, what is obtained from the integration is the value of the expansion coefficients at each time, i.e. the c_j parameters in equation II-19.

Therefore, the CPMD methodology is given by Figure II-1, where it can be notice that the electronic energy only needs to be calculated at the beginning of the simulation, because the electrons also evolve in time, following the nuclear motion. Hence, the equations of motion have to be integrated for both nuclei and electrons at each MD step, obtaining a new set of orbital coefficients, besides the new nuclear positions.

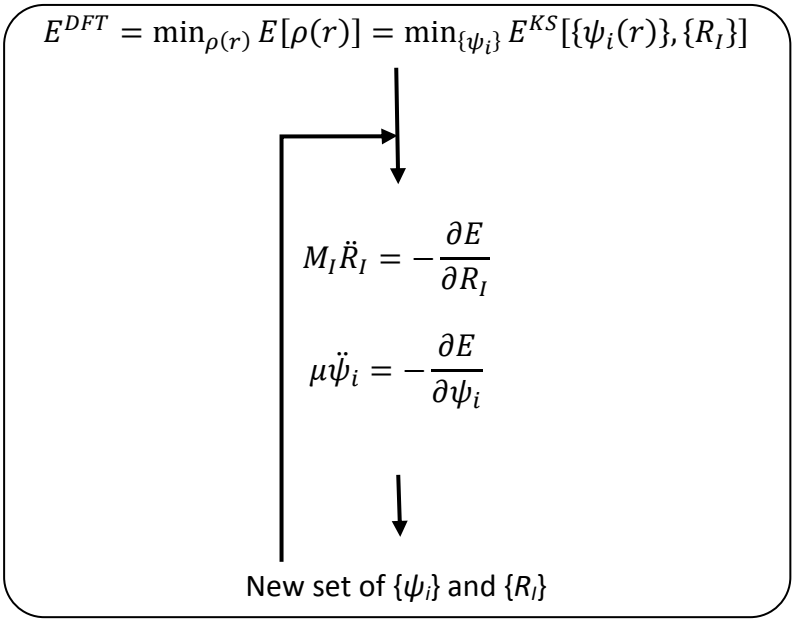

Figure II-1. Schematic diagram of a Car-Parrinello MD. Adapted with permission from reference 46.

The decoupling of the two subsystems can be achieved by a suitable choice of the fictitious electronic mass μ, such that the frequencies of the electrons are well separated from the nuclear frequencies. In that case, the electronic subsystem remains on the Born-Oppenheimer surface, oscillating around the ground state, and the forces on the atoms are practically the same as those in a Born-Oppenheimer MD (BOMD). Therefore, in a CP simulation, the electronic energy only needs to be calculated at the beginning of the simulation, and not at each step as in BOMD. By contrast, the simulation time step is shorter (0.1-0.2 fs) compared to BOMD (1 fs), in order to describe properly the faster movement of the electrons.

Hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) Methods

The interest in multiscale approaches is continuously growing and has been honoured in 2013 with the Nobel Prize in Chemistry that was jointly awarded to M. Karplus, M. Levitt and A. Warshel. The combination of Quantum Mechanics / Molecular Mechanics (QM/MM) together with the increasing computational power of modern parallel, vector-parallel and hybrid CPU-GPU platforms has lead to a real breakthrough in the simulations of biomolecular reactions.[47]

Quantum Mechanics / Molecular Mechanics (QM/MM) is a hybrid methodology which permits first-principles calculations of a small portion of a system in a classical force field environment, thus combining the advantages of the two methods.[48] Whereas classical force fields allow to treat big systems, first-principles methods are needed to study electronic reorganizations, such those that occur in certain molecule-molecule interactions and chemical reactions.

In QM/MM techniques, the overall system is divided in two regions: the QM region is treated with AIMD, whereas the rest of the system is described by classical MD. The QM part is formed by the region that should be described more precisely, as the active site of a protein or the core of an organic protected metallic cluster. The remaining part of the system as well as solvent and counter ions is treated with MM methods (MM part). The total energy of the hybrid system can be thus written as

$$E = E_{QM} + E_{MM} + E_{QM/MM} \quad \textbf{(Eq.II-27)}$$

where E_{QM} is the energy of the quantum subsystem, E_{MM} is the energy of the classical subsystem and $E_{QM/MM}$ contains the interactions between the two regions. The crucial issue regarding hybrid methods is the treatment of the interactions involving both the classical and the quantum parts of the system.

In this thesis we have used the QM/MM methodology developed by A. Laio, J. VandeVondele and U. Röthlisberger,[49] which combines Car-Parrinello Molecular Dynamics (CPMD) with the classical molecular dynamics based on the GROMOS or AMBER force fields. In this framework, E_{QM} is described by DFT and the potential created by the MM atoms, which polarizes the electronic density of the QM region, is included in the energy functional as an external potential. E_{MM} is computed using the AMBER parameters with the GROMOS force field equation. $E_{QM/MM}$ is calculated as

$$E_{QM/MM} = E_{QM/MM}^{bond} + E_{QM/MM}^{non-bond} \quad \textbf{(Eq.II-28)}$$

QM/MM boundary: bonding QM/MM interactions

When the partitioning between the QM and MM atoms occurs at a covalent bond, the valence of the QM atom needs to be saturated to have a good description of the electronic structure. In this thesis we have used Optimized Effective Core Pseudopotentials (OECP),[50] a pseudopotential constructed with a reduced valence and empirically optimized to reproduce the structural and electronic description of the corresponding atom. In particular, we have used monovalent carbon pseudopotentials optimized to reproduce Csp^3-Csp^3 bond length. Bonded, angular and dihedral terms involving MM and QM atoms are treated as in the classical MM Hamiltonian. These terms permit to keep stable the geometry at the interface.

QM/MM boundary: non-bonding QM/MM interactions

The non-bonding interactions between the QM and the MM regions are divided into electrostatic and Van der Waals interactions:

$$E_{QM/MM}^{non-bond} = \sum_{i \in MM} q_i \int \frac{\rho(r)}{|r - r_i|} dr + \sum_{\substack{i \in MM \\ j \in \in QM}} V_{VdW}(r_{ij}) \quad \textbf{(Eq.II-29)}$$

where r_i is the position of the MM atom *i*, with charge q_i, ρ is the total charge of the quantum system and $V_{VdW}(r_{ij})$ is the Van der Waals interaction between atom *i* and atom *j*. The steric non-bonded interactions due to the Pauli repulsion and the dispersion interactions are usually kept into account in a straightforward way by retaining the Van der Waals interaction as described by the classical force field.

The electrostatic interaction between the QM density and the point charges representing the charge distribution in the MM system often constitute the main environmental effect on the QM system and is technically more subtle. More precisely, calculation of the electrostatic interactions in a QM/MM scheme has two main problems, related to both its short range and its long range behaviour:

- *Electron spill-out*: it is an anomalous rearrangement of the electron density that tends to localize on the positively charged classical atoms. This effect is particularly pronounced in a planewave basis-set approach, in which the electrons are fully free to delocalize. To solve this problem in CPMD QM/MM calculations, the point charges of the classical system are replaced by a charge distribution v,[49] so that the 1/r behaviour is maintained for large r, whereas for *r* values shorter than the covalent radius of an atom i (r_{ci}) the Coulomb potential goes to a finite value:

$$E_{QM/MM}^{electrostatic} = \sum_{i \in MM} q_i \int \rho(r) v_i(|r - r_i|) dr = \sum_{i \in MM} q_i \int \rho(r) \frac{r_{ci}^4 - r^4}{r_{ci}^5 - r^5} dr \quad \textbf{(Eq.II-30)}$$

- Computational cost: the explicit calculation of the electrostatic term in a planewave-based approach requires a very high number of operations. Therefore, the computational cost is huge. To speed up the calculation, CPMD QM/MM uses a *multilayer approach* (Figure II-2).

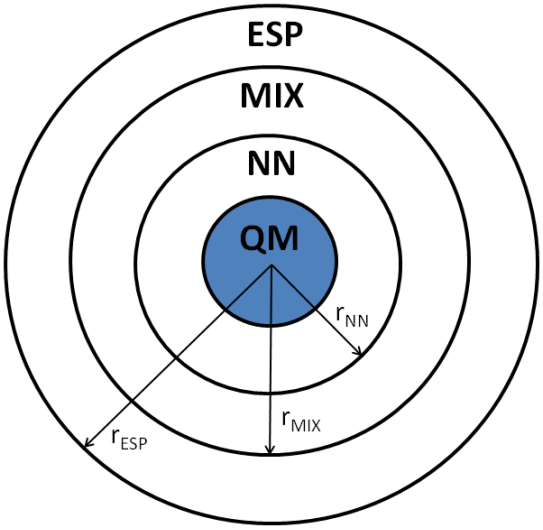

Figure II-2. Schematic representation of the NN, MIX and ESP regions of electrostatic interaction used in the multilayer approach. The QM box is depicted as a blue region.

The MM subsystem is divided in four different regions, in which the QM/MM electrostatic interactions are calculated differently. For the atoms within the first region (NN region; $r < r_{NN}$) the QM/MM electrostatic interaction is calculated explicitly as in equation II-29. In the second region (MIX region; $r_{NN} < r < r_{MIX}$) equation II-29 is only applied for those MM atoms whose charge is larger than 0.1 *e*. For the MM atoms with a charge < 0.1 e or within the ESP region ($r_{MIX} < r < r_{ESP}$) a classical Coulomb potential is used:

$$E_{QM/MM}^{electrostatic} = \sum_{\substack{i \in MM \\ j \in QM}} \frac{q_i Q_j^{RESP}}{|r_i - r_j|} \qquad \textbf{(Eq.II-31)}$$

where Q_j^{RESP} are dynamically generated RESP (D-RESP) charges assigned to the QM atoms. To obtain the RESP charges, the electrostatic potential created by the electron density $\rho(r)$ is calculated at the position of the classical atoms inside the NN region. Then, the point charges that best reproduce this potential are calculated. This fitting is performed with a harmonic constraint such that the RESP charges do not deviate from a reference value, which is taken as the Hirshfeld charges obtained from the QM electronic density.

Outside the outer layer ($r > r_{ESP}$) the QM electronic density is simply described as a multipole expansion up to quadrupolar order.

Metadynamics

Metadynamics is a novel molecular dynamics based technique aimed at enhancing the sampling of the phase space and at estimating the free energy landscape.[51] The algorithm is based on a dimension reduction, defining a set of well-chosen collective variables (CVs) which allow the description of the essential modes associated with the transition of interest.

Direct Metadynamics

Let us consider a system described by a set of coordinates *x* moving under an interatomic potential *V(x)*.[51b] In a standard molecular dynamics simulation the system is oscillating in one of the minima of the energy surface and the crossing to other minima is rather unlikely to occur (rare event) if the barrier is higher than the thermal energy fluctuations ($1/2 \cdot k_b \cdot T$) (Figure II-3a). To accelerate the exploration of the phase space we can use metadynamics. This technique is based on defining a collective variable *s* which is an explicit function of a subset of coordinates x' [s = f(x')]. For example, distances, angles, dihedrals, coordination numbers and so on are typical CVs.

The metadynamics simulation starts in one of the minima of the energy surface. At the beginning of the simulation, the force acting on the system is given by the gradient of the potential *V*:

$$f_i = -\frac{\partial V}{\partial r_i} \quad \textbf{(Eq.II-32)}$$

After a time interval *t'*, a small Gaussian-like repulsive potential term (*V'*) is added, centered at the value of *s* at time *t'* (Figure II-3b). This is repeated every time interval *t'* (deposition time) along the MD simulation, such that at time *t* the total potential is the sum of the interatomic potential *V* and the sum of the *V'* terms added up to that time, *i.e.* V_G:

$$V_T = V + V_G = V + \sum_{t'} w \cdot exp\left(-\frac{[s(t) - s(t')]^2}{2(\delta s)^2}\right) \quad \textbf{(Eq.II-33)}$$

where *w* is the height of the Gaussian potential term, δs is its width and V_G is the history-dependent potential added (*i.e.* the number of Gaussian terms added during the simulation). As a result, the force acting on the system is now the sum of two components, one coming from *V* and the other from the added Gaussian potential, V_G:

$$f_i^{V_T} = f_i^V + f_i^{V_G} = -\frac{\partial V}{\partial r_i} - \frac{\partial V_G}{\partial r_i} \quad \textbf{(Eq.II-34)}$$

When a Gaussian potential term is added to V_G, the total potential changes and hence the system needs to be equilibrated by using a thermostat. In addition, the resulting forces are also modified, and thus the metadynamics trajectory is not an equilibrium trajectory.

The added potential V_G progressively counterbalance the underlying free energy, penalizing those regions of the phase space already explored. As a result, the system tends to escape from already visited points to another energy minima, through the nearest and lowest-in-energy transition state. Hence, metadynamics can accelerate the exploration of the phase space and the observation of rare events if the CVs are properly chosen. In a metadynamics run, the system will evolve by the same pathway that would have explored in a longer standard MD.

Besides, it has been demonstrated that the bias potential will converge to the negative of the free energy (Figure II-3c):[51a]

$$\lim_{t\to\infty} V_G(s,t) \approx \sum_{t'} w \cdot exp\left(-\frac{[s(t)-s(t')]^2}{2(\delta s)^2}\right) \quad \textbf{(Eq.II-35)}$$

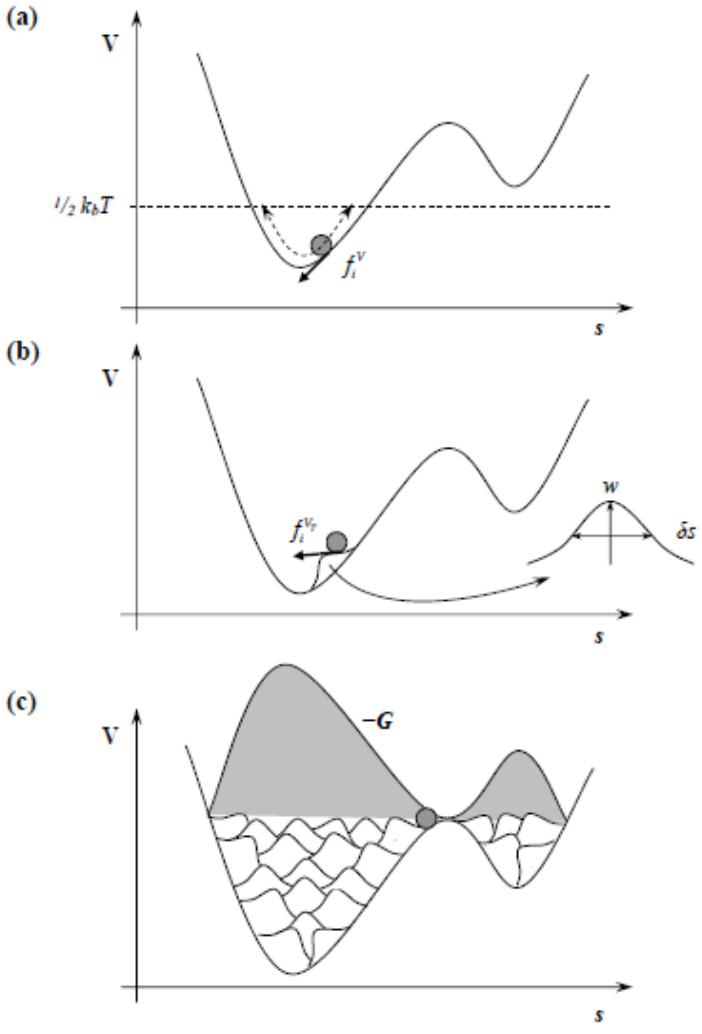

Figure II-3. a) Standard molecular dynamics simulation. The system evolves following the interatomic potential V, and thus the force acting on the system is f_i^V. **b)** Gaussian potential with height *w* and with δs added at certain *t'* time centered at the value of *s* during a metadynamics run. **c)** End of the metadynamics simulation and reconstruction of the free energy landscape (Eq.II-35). Figure reproduced with permission from reference 52.

Lagrangian Metadynamics

When metadynamics is used for simulating chemical reactions by ab initio molecular dynamics simulations, the history-dependent potential force the system to cross barriers of several tenths of kcal·mol^{-1} in a short time, usually a few picoseconds. This implies that a lot of energy has to be injected in the degrees of freedom associated with the chosen CVs. This might lead to a significant inhomogeneity in the temperature distribution of the system, and possibly to instabilities in the dynamics. To solve these problems, the extended Lagrangian metadynamics method has been introduced. In particular, here we consider the extended Lagrangian metadynamics coupled to Car-Parrinello molecular dynamics,[53] in which the Car-Parrinello Lagrangian is extended by extra terms describing the fictitious dynamics of the CVs.

The CVs in the real system (*s*) are coupled to additional fictitious particles through a harmonic potential V_H:

$$V_H = \frac{1}{2}k[s' - s(x)]^2 \quad \textbf{(Eq.II-36)}$$

and these auxiliary CVs (*s'*) are treated as additional degrees of freedom with a fictitious mass M and an associated fictitious kinetic energy $K_{s'}$ given by:

$$K_{s'} = \frac{1}{2}Mv_{s'}^2 \quad \textbf{(Eq.II-37)}$$

In this way, the dynamics of these extra degrees of freedom can be explicitly controlled using thermostats and the trajectory of *s'* will be sufficiently soft as to maintain the stability of the algorithm.[54]

Due to the introduction of an extra potential V_H, the total potential acting on the real system is modified:

$$V_T = V + V_H = V + \frac{1}{2}k[s' - s(x)]^2 \quad \textbf{(Eq.II-38)}$$

and also the resulting forces:

$$f_i^{V_T} = f_i^V + f_i^{V_H} = -\frac{\partial V}{\partial r_i} - \frac{\partial V_H}{\partial r_i} \quad \textbf{(Eq.II-39)}$$

Therefore, the dynamics of the real system in the real space (*s*) is driven by the sum of the interatomic potential *V* and the harmonic potential V_H. The mass for this fictitious particle and the force constant of the coupling potential need to be tested to ensure that the coupled particle follows naturally the value of the associated CV in the real system. The force constant has to be chosen large enough to keep *s'* close to the actual coordinates of the system *s(x)*. However, a large value of *k* increases the computational cost.[54] Within a CP scheme, another important requirement is the adiabatic separation between the auxiliary variables *s'* and the electronic degrees of

freedom. Since the extra term in the Hamiltonian introduces frequencies of the order of $\sqrt{k/M}$, *M* should be relatively large, and the CV dynamics should be adiabatically decoupled from the atomic motions. However, large values of *M* make the exploration of the free energy landscape very slow and computationally demanding in combination with CPMD. This means that a balance between minimal energy transfer between the electronic, ionic, and CV dynamics subsystem on one hand, and a workable efficiency in the free energy landscape exploration on the other hand must be reached.

The history-dependent potential V_G, instead of acting on the real system as in direct metadynamics, acts on the space of the auxiliary CV *s'*, *i.e.* the Gaussian potential terms *V'* are now centered on *s'* (Figure II-4). Therefore, the dynamics in the space of the auxiliary CVs is driven by the forces coming from the harmonic potential V_H and the history-dependent potential V_G:

$$f_i^{V_T} = f_i^{V_H} + f_i^{V_G} = -\frac{\partial V_H}{\partial r_i} - \frac{\partial V_G}{\partial r_i} \quad \textbf{(Eq.II-40)}$$

Similar to direct metadynamics, the history-dependent potential V_G will penalize visiting the regions of the s' space already explored. Moreover, because *s'* is coupled to *s* through a harmonic potential, it will also accelerate the exploration of the *s* space. Therefore, the V_G potential acting on the *s'* space will progressively counterbalance the interatomic potential *V* in the s space, such that the free energy surface can be reconstructed as the negative of the sum of the Gaussian potential terms, as in direct metadynamics:

$$\lim_{t\to\infty} V_G(s',t) \approx \sum_{t'} w \cdot exp\left(-\frac{[s'(t)-s'(t')]^2}{2(\delta s)^2}\right) \quad \textbf{(Eq.II-41)}$$

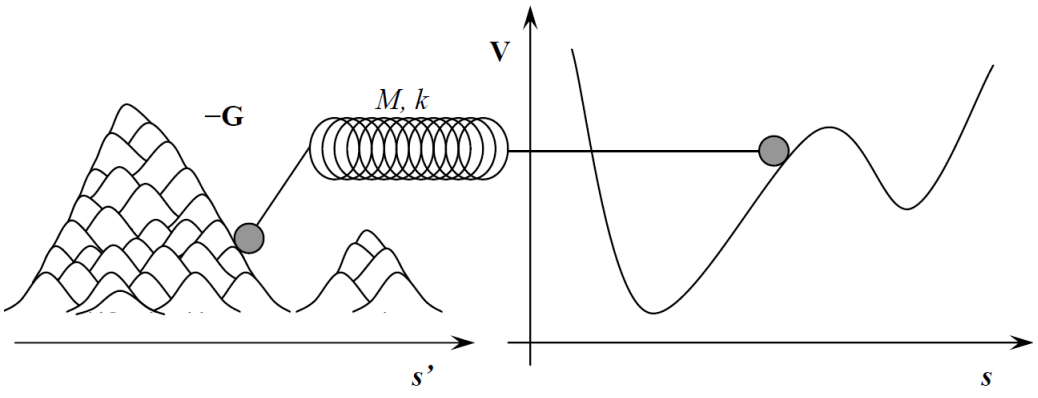

Figure II-4. Schematic representation of a completed simulation and reconstruction of the free energy surface for an extended-lagrangian metadynamics. Figure taken with permission from reference 52.

In summary, the main advantages of metadynamics are:[55]

- It accelerates the sampling of rare events by pushing the system away from local free energy minima.
- It allows the exploration of new reaction pathways as the system tends to escape the minima passing through the lowest free-energy saddle point.
- No a priori knowledge of the landscape is required.

On the other hand, it has two major drawbacks:

- In a single run, V_G does not converge to a definite value of free energy, but oscillates around it. Therefore, it is not trivial to decide when to stop a simulation metadynamics simulation. As a general rule, if metadynamics is used to find the closest saddle point, it should be stopped as soon as the system exits from the minima. For chemical reactions, it has been shown that a good estimation of the free energy barrier can be obtained after just one crossing of the TS, by restarting the simulation sometime before this event using Gaussian heights $\leq 1 k_B T$.[56] If one is interested in reconstructing a complex free energy surface, with many local minima, the simulation should be ideally stopped when the motion of the CVs becomes diffusive in the region of interest. However, this is often impracticable for *ab initio* metadynamics simulations.
- Identifying a set of appropriate CVs for describing complex processes is far from trivial.

Collective Variables (CVs)

A collective variable (CV) is a function of the microscopic coordinates of the system. To guarantee an effective application of metadynamics, the chosen CVs must respect the following guidelines:[51b, 53]

- They must be a function of the microscopic coordinates of the system and the function must have a continuous derivative.
- They should distinguish between the initial and the final states and describe all the relevant intermediates.
- They should include all the slow modes of the system. A slow mode is a variable that cannot be satisfactorily sampled in the timescale of the unbiased simulation. We also expect that the other fast variables adjust rapidly to the evolution of the slow variables. If any of the latter is not added to the CVs list, the bias potential may not converge to the FES in a reasonable simulation time.
- They should be limited in number.

The collective variables used along this thesis are described in the corresponding chapters.

Chapter III – Formation of a Thiolate Monolayer-Protected Gold Cluster: Birth of a Monomeric Staple Motif

Formation of a Thiolate Monolayer-Protected Gold Cluster: Birth of a Monomeric Staple Motif

Introduction

The most established way to synthesized thiolate monolayer-protected gold clusters is the so called Brust-Schiffrin method,[5] where tetrachloroaurate salts ($AuCl_4^-$) are reduced by sodium borohydride ($NaBH_4$) in a two-phase redox reaction in an appropriate mixture of methanol and water. However, experiments show that AuMPCs can be formed from neutral gold clusters (AuCs) and thiols.[57] In addition, recent analyses using the solvated metal atom dispersion method provide evidence of hydrogen evolution upon exposure of thiols to naked gold clusters[11], due to sulphur-hydrogen bond scission. These experiments indicate that a redox reaction occurs between the neutral gold atoms and the protons of the incoming thiol groups. Therefore, protons have been proposed as oxidants for Au^0.[58]

As it is described in page 5 of chapter I, gold atoms of AuMPCs do not form a compact gold core, but some gold atoms emerge from the surface, interacting strongly with the sulphur atoms and forming characteristic protecting units, named *staple motifs*.[12a] The presence of conserved substructures in all the AuMPCs so far characterized suggests that there is a general mechanism by which these substructures form. Key to deciphering this mechanism is to rationalize how the simplest thiolate-gold substructure, the monomeric staple motif, is formed. When the Brust-Schiffrin method is used, the first step is the formation of Au(I)-SR polymers as a result of $AuCl_4^-$ reacting with thiols. During this first step, chloride anions are proposed as the base that deprotonates the incoming thiols.[59] However, the absence of a base in the synthesis methods where naked AuCs are mixed with thiols leads to a key question: **what happen with the hydrogen of the thiols?** The fate of the protons released during thiols binding,[60] from its neutral unbound form (RSH) to the anionic bound form (RS^-) found in AuMPCs, and the atomic/electronic rearrangements upon formation of staple motifs remain controversial issues. Only thiolate-gold bonds, in the form of staple motifs, are found in X-ray structures of AuMPCs. However, a signal of a hydrogen atom was detected in NMR spectroscopic measurements,[60] suggesting the presence of intact thiols on their surface, without excluding the fact that hydrogen could be adsorbed on the gold surface, i.e. H-Au bonds can be formed. Therefore, a current lack of mechanistic insight does not allow firm conclusions regarding the molecular details of the mechanism.

We report here an *ab initio* metadynamics study showing that gold-thiol complex readily evolves in a multiple-pathway chemical reaction to form a monomeric staple motif, as is present on the surface of AuMPCs. A number of intermediate species, including those proposed in experimental studies, are found during the

chemical redox reaction. We also present theoretical evidence for the release of molecular hydrogen upon the formation of Au-S bonds in an AuMPC. The results obtained reconcile the findings of previous spectroscopic techniques (intact thiols or chemisorbed hydrogen) with those obtained by X-ray measurements (thiolate-sulphur bonds). Finally, a general mechanism to explain the protection of AuCs by thiol groups is proposed.

Model Systems

To elucidate the molecular mechanism of the redox reaction, we designed the smallest possible AuMPC bearing one monomeric staple motif, $Au_4(SCH_3)_2$. Obviously, this system must conform to the general stoichiometric relations governing the size of AuMPCs (see page 5 of Chapter I for a more detailed description):

$$[Au]_{a+a'}[Au(SR)_2]_b[Au_2(SR)_3]_c[Au_3(SR)_4]_d \ldots$$

where $[Au]_{a+a'}$ is the gold core and $[Au(SR)_2]_b$, $[Au_2(SR)_3]_c$ and $[Au_3(SR)_4]_d$ the monomeric, dimeric or trimeric staple motifs bonded to the core. *a* is the number of inner core gold atoms, *a'* the number of outer core gold atoms (named also as anchor atoms) and *b*, *c* or *d* the number of monomeric, dimeric or trimeric staple motifs, respectively. The $Au_4(SCH_3)_2$ cluster, which can also be expressed as $[Au]_{1+2}[Au(SCH_3)_2]_1$ fulfils these requirements. It contains one gold core atom (a = 1), two outer core gold atoms (a' = 2) and one monomeric staple motif (b = 1, c = 0, d = 0) (Figure III-1). Moreover, the number of valence electrons (n_{elec} = 2) is the smallest possible "magic number" (page 7 of Chapter I).[61] Other AuMPCs with two valence electrons have been proposed by Jiang *et al.*,[15a, 62] such as $Au_{10}(SR)_8$, $Au_8(SR)_6$, $Au_6(SR)_4$ or $Au_{12}(SR)_9^+$. However, we rely on $Au_4(SCH_3)_2$ because of its smaller size.

It is known that monomeric staple motifs are energetically not preferred on small AuCs. For example, in the case of $Au_{10}(SR)_8$ it is proposed that the protecting unit of the Au_4 core is two trimeric staple motifs.[15a] On another hand, monomeric staple motifs have been characterized in the solid state for an Au_4 cluster[63] and therefore they should be stable. To rationalize this issue, we can think on the series of organic cicloalkanes, with chemical formula C_nH_{2n}: although small cicloalkanes (n = 3, 4 or 5) has a large ring strain they are all stable molecules and as such they can be isolated and characterized experimentally. With AuMPCs the same can happen, although small staple motifs have a large strain when they are bond to small AuCs, they can be stable and therefore, can be isolated.

The question of the mechanism of formation of the monomeric staple motif thus reduces to: how does $Au_4(SCH_3)_2$ form from a bare AuC interacting with thiols (Figure III-1)?

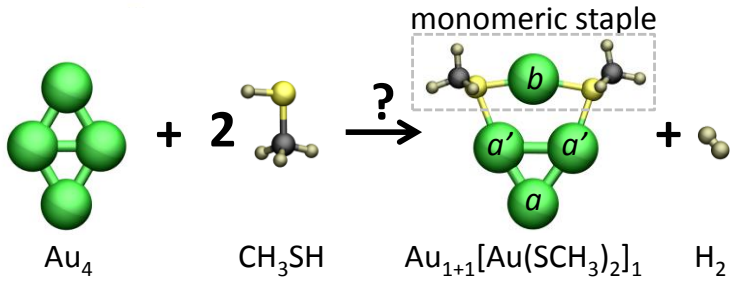

Figure III-1. Scheme of the reaction that is going to be investigated along this Chapter. Colours code: Au (green), S (yellow), C (black), O (red), N (blue) and H (grey).

Additional simulations are performed to study the mechanism in the presence of an excess of methanethiol. This scenario was modelled by considering two additional methanethiols. In principle, only two thiols should lead to thiolate-sulphur bonds; otherwise, the resulting AuMPC would not conform to the electron count rules for exceptional stability (page 7 of Chapter I).[17-18] However, any of the remaining two thiols could bind to the cluster as neutral thiols.

Computational methods

Simulation Details

Ab initio molecular dynamics simulations were performed within the Car-Parrinello approach,[64] which is based on Density Functional Theory (DFT). The Kohn-Sham orbitals were expanded in a plane wave basis set with a kinetic energy of 70 Ry. Gold atoms are represented by a Goedecker[65] type pseudopotential, leaving 11 electrons (the $5d^{10}$ and $6s^1$) in the valence shell. Carbon and sulphur are represented by a norm-conserving Troullier-Martins[45] pseudopotential type and hydrogen by a Car-von Barth type, as it was done in references 66 and 67. Calculations are made using the generalized gradient-corrected approximation of the spin-dependent density functional theory (DFT-LSD), following the prescription of Perdew, Burke and Ernzerhoff (PBE).[68] This functional has also been used in previous studies of AuMPCs (see, for instance, references 69), as well as in the study of the mechanism of CO_2 formation on an AuMPC.[70] Test calculations on AuMPCs[16] show that this choice of pseudopotentials and exchange-correlation functional provides a good description of the structures of AuMPCs. All simulations used a timestep of 0.096 fs and the fictitious electron mass of the CP Lagrangian was set to 400 a.u..

The two systems, Au_4 + 2 CH_3SH and Au_4 + 4 CH_3SH, are enclosed in a 15 Å x 15 Å x 15 Å and 20 Å x 20 Å x 20 Å simulation cubic cells, respectively. The starting

configuration for the simulations consists of a naked Au4 cluster with two or four methanethiols at ≈4 Å distance.

The metadynamics approach,[51a] in its extended Lagrangian version,[53] is used to simulate the redox reaction (see page 35 of Chapter II to a more detailed description of the method). The collective variables (CVs) used for the metadynamics simulations of the redox reaction were taken as a combination of coordination numbers of the covalent bonds being formed/broken. Exploratory simulations including the coordination numbers of all possible covalent bonds formed/broken (i.e. H-Au, H-H, S-H and S-Au) revealed that the two hydrogen atoms are transferred sequentially (i.e. the second one "waits" for the first to be transferred). Therefore, we decided to model the redox reaction in two consecutive bidimensional metadynamics simulations (hereafter labelled as MTD1 and MTD2), corresponding to the respective transfers of H' and H'' with two collective variables for each one. For both metadynamics, the first collective variable used (CV_1 or H detachment) measures the cleavage of the S-H bond and the formation of hydrogen-gold bonds or a hydrogen-hydrogen bond (formation of molecular H_2). This is done through a difference of coordination numbers:

$$\textbf{MTD1: } \Delta CN = CN_{S',H'} - CN_{H',Au'} \quad \textbf{(Eq.III-1)}$$

$$\textbf{MTD2: } \Delta CN = CN_{S'',H''} - CN_{H',H''} \quad \textbf{(Eq.III-2)}$$

where S and H are sulphur and hydrogen atoms of the corresponding methanethiolate (S' and H' or S'' and H'') and Au' the gold bonded to it (Figure III-2). $CN_{i,j}$ is computed as

$$CN_{i,j} = \frac{1 - \left(\frac{d_{i,j}}{d^0}\right)^p}{1 - \left(\frac{d_{i,j}}{d^0}\right)^{p+q}} \quad \textbf{(Eq.III-3)}$$

where d_{ij} is the internuclear distance of the atoms involved, d^0 is the threshold bonding distance, and p and q are exponents that determine the steepness of $CN_{i,j}$ decay with respect to d_{ij}. CN values range from 0 (not bonded) to 1 (bonded). These types of CVs have proven very useful for the description of chemical reactions in recent studies (see for instance refs. [71])

The second collective variable (CV_2 or H-Au coordination) captures the diffusion of hydrogens on the Au_4 by using a summation of coordination numbers:

$$CN_{i,j} = \sum_{i}^{N_A} \sum_{j}^{N_B(i)} \frac{1}{1 + e^{k[d_{i,j} - d^0(i)]}} \quad \textbf{(Eq.III-4)}$$

For MTD1 only H' is considered to be bonded to Au_4. In MTD2, both H' and H'' are bonded to Au_4. N_A and N_B are the number of atoms of species A (1 or 2 hydrogens) and B (the four gold atoms), respectively. $d_{i,j}$ is the internuclear distance of the atoms

involved, d^0 is the threshold bonding distance and k determines the steepness of the decay of the function with respect to $d_{i,j}$.

Results

Stability of $Au_4(SCH_3)_2$

Before starting the investigation of the mechanism of formation of $Au_4(SCH_3)_2$, its stability was tested by *ab initio* molecular dynamics (AIMD) using a three-step optimization procedure.[66-67, 72] First, we heated up the system up to 270-330 K and performed an AIMD equilibration. Secondly, we cool down the system to a temperature near to 0 K by scaling the nuclear velocities at each MD step using an increasing annealing factor from 0.9 to 0.999. Finally, we optimized the structure of $Au_4(SCH_3)_2$. The system was found to be stable, with an internal structure very similar to the ones available for larger AuMPCs that inspired the design of our model (Table 1).

	distances (Å)					angle (°)
	Au(a)-Au(a')	**Au(a')-Au(a')**	**Au(a')-Au(b)**	**Au(a')-S**	**Au(b)-S**	**S-Au-S**
$Au_4(SCH_3)_2$	2.61	2.89	2.95	2.39	2.37	163.0
$Au_{25}(SR)_{18}^-$	2.77	2.87	3.03	2.37	2.30	172.8
	2.80	2.99	3.24	2.40	2.34	
$Au_{38}(SR)_{24}$	2.76	2.76	3.04	-	2.30	173.5
	2.84	3.08	3.12			
$Au_{102}(SR)_{44}$	2.8	2.8	2.8	2.2	2.2	155
	3.1	3.1	3.1	2.6	2.6	175

Table III-1. Main distances (Å) and angles (degrees) computed for $Au_4(SCH_3)_2$ in comparison with literature values for $Au_{25}(SR)_{18}^-$, $Au_{38}(SR)_{24}$ and $Au_{102}(SR)_{44}$.[12a, 14, 20b]

Molecular Mechanism of the Redox Reaction in stoichiometric conditions

1. Spontaneous Thiol Chemisorption.

The first step of the formation of $Au_4(SCH_3)_2$ from a bare Au_4 gold cluster and two methanethiols is the binding of both thiols to the cluster. An AIMD simulation at room temperature shows that thiol binding occurs spontaneously in less than 2 ps, forming a Y-shaped thiol-gold complex (labelled as **I1** in Figure III-2). Chemisorption of the first thiol stabilizes the complex by 8.7 kcal/mol (Table III-2), but the energy gain for binding the second thiol is more than twice that, 19.3 kcal/mol, indicating that thiol chemisorption is initially a cooperative process.

n	ΔE (kcal/mol)
1	-8.7
2	-19.3
3	-14.9
4	>0

$$Au_4(CH_3SH)_{n-1} \xrightarrow{\Delta E} Au_4(CH_3SH)_n$$

Table III-2. Energies of the chemisorption of methanethiols to Au_4.

Interestingly, the Au_4 cluster becomes slightly anionic (by 0.257 electrons; Table AI-I of the appendix I) because of electron donation from the bonded thiols. No further structural changes were observed when the simulation was extended. For this reason, we used an enhanced sampling technique, metadynamics, to activate the redox reaction and obtain its complete free-energy landscape.

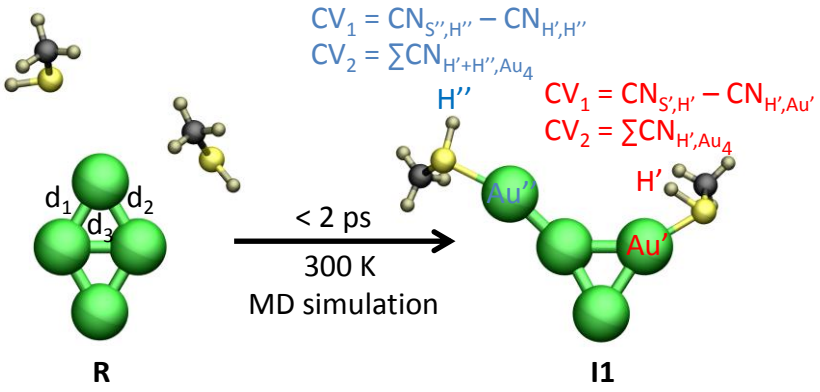

Figure III-2. Spontaneous binding of thiols to the Au_4 cluster (**R** → **I1**). The two thiols are initially located at a ≈ 4 Å distance from the cluster. The collective variables used in MTD1 (red) and MTD2 (blue) are shown.

2. First Proton Transfer.

The free-energy landscape reconstructed from the metadynamics simulations (Figure III-4) contains several local minima, labelled from **I1** to **I5**. Nevertheless, a reactive pathway connecting reactants and products can be easily drawn. The reaction begins with elongation of the S-H' bond (by ≈ 0.2 Å; Table III-3) until H' starts to interact with the gold atoms (intermediates **I2** and **I3**). This H···Au interaction has been observed for a tryptophan amino acid absorbed on a AuC as a very stable configuration, although the interaction takes place in this case via a carboxylic group instead of a thiol.[73] These intermediates, in which H' is ready to be transferred to the cluster, can be described as

"preactivated states". The reaction continues with the transfer of H' into the Y-shaped cluster. At this point, H' diffuses on the cluster, adopting various types of H-Au coordination: 2-fold as in **I4** or 3-fold as in **I5**. From an energetic point of view, the transition from the reactants to a preactivated state (I2 or I3) involves lower energetic barriers (16 kcal/mol) than proton transfer to Au_4 (28 kcal/mol).

It is interesting to analyze the evolution of the atomic charges obtained from the electron density (solid lines in Figure III-3 and Table AI-I in appendix I). The charge on H' becomes slightly positive when H' starts to interact with Au_4 (**R** to **I1**). In turn, the sulphur atoms become more negative. However, the H' charge decreases dramatically upon transfer (I1 to I4). Remarkably, the charge of the Au_4 shows the opposite trend. It receives electron density upon thiol binding (R to I1) but becomes positively charged upon H' transfer (from **I1** to **I4**). These changes indicate that H' detaches from its sulphur as a proton, but it acquires pseudohydride character once bound to the cluster.

3. Second Proton transfer and Hydrogen Release.

After the first proton (H') is transferred, the second one (H'') starts to detach from the thiol. It is first preactivated by its interaction with gold atoms, and simultaneously, one of the Au-Au bonds breaks (evolution of the main distances are shown in Table 3) and one gold atom separates from the rest (**I4** to **I6**), forming the precursor of a staple motif. This step has an energetic cost of 18 kcal/mol. Then, H'' coordinates with Au_4 (**I6** to **I7**). Both hydrogens (H' and H'') are bonded to the same gold atom in I7. Again 18 kcal/mol are needed. Finally, cleavage of the two Au-H bonds (**I7** to **P**) leads to the release of hydrogen molecule and the formation of the staple motif. Therefore, the results demonstrate that a monomeric staple motif can form in a redox reaction from a bare AuC interacting with thiols.

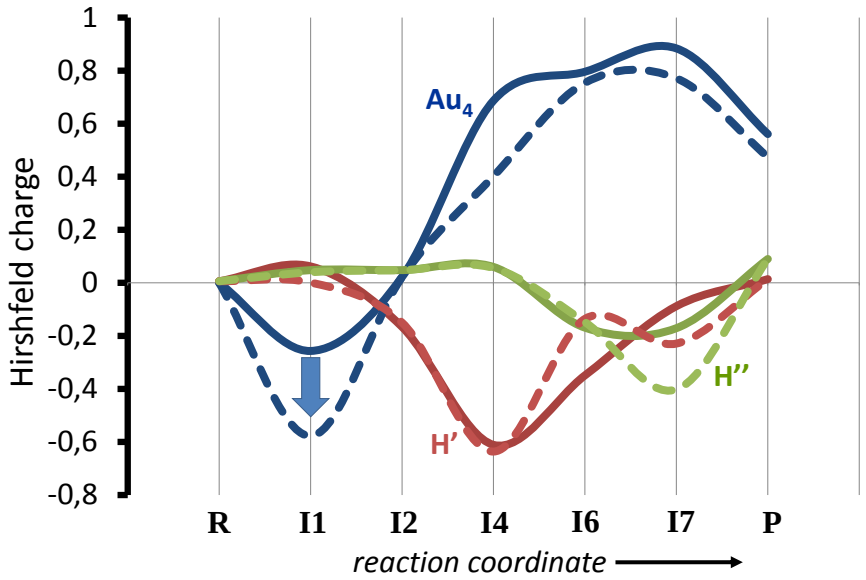

Figure III-3. Evolution of atomic charges (Hirshfeld charges) of Au_4 and the thiol hydrogen atoms along the reaction pathway. The charges are given relative to the isolated Au_4 and CH_3SH molecules. Solid line: reaction with two methanethiols. Dashed line: reaction with four methanethiols.

	distance (Å)							
	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**P**
Au-Au (d_1)	2.60	2.53	2.56	2.61	2.63	3.70	2.83	2.90
Au-Au (d_2)	4.50	4.40	4.11	4.00	3.93	2.99	3.10	2.98
Au-Au (d_3)	2.81	2.65	2.77	2.68	2.82	2.94	2.70	2.94
Au-S'	2.46	2.4	2.40	2.29	2.33	2.36 2.37	2.32 2.43	2.39 2.41
Au-S"	2.50	2.46	2.5	2.41	2.42	2.38	2.32	2.60
H'-S'	1.39	1.56	1.50					
H''-S''	1.38	1.38	1.38	1.38	1.38	1.54		
H'-Au		1.73	2.0	1.9 1.8	2.01 1.95 1.73	1.63	1.8	
H"-Au						2.5	1.85	

Table III-3. Structural parameters of each characteristic minimum along the reaction pathway for the reaction of two methanethiols with Au_4.

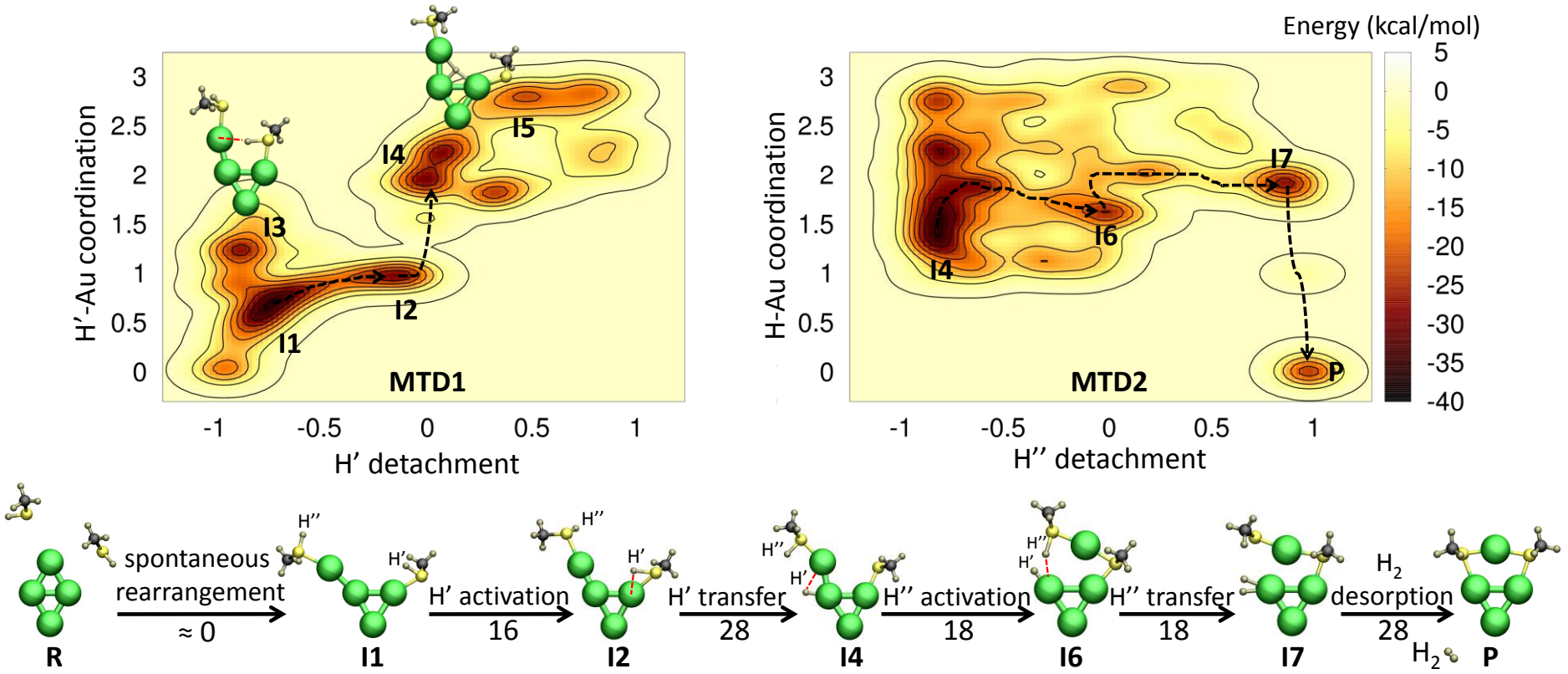

Figure III-4. Free energy surfaces reconstructed from the metadynamics simulations (MTD1 and MTD2) of the reaction of two CH_3SH molecules with Au_4. Contour lines are plotted every 2 kcal/mol. Au-Au solid bonds are drawn for distances < 2.8 Å. H-Au solid bonds are shown for distances < 1.85 Å, whereas dotted lines are used for bond distances in the range of 1.85-2.0 Å. Free energy barriers in kcal/mol are shown below the arrows.

Molecular Mechanism of the Redox Reaction in an Excess of Methanethiols

1. Spontaneous Thiol Chemisorption.

When four methanethiolates are placed close to the Au4 cluster, three of them are able to spontaneously bind to the Au_4 (Figure III-5). The shape of the cluster is not modified when three methanethiols are bind (**I1** in Figure III-5), unlike the previous case (**I1** in Figure III-4), where $Au_4(CH_3SH)_2$ adopts a Y-shaped structure. All attempts to bind the fourth methanethiol are unsuccessful. Once it approaches the cluster, another methanethiol unbinds, recovering the $Au_4(CH_3SH)_3$ configuration. This type of anticooperative effect is also reflected in the calculation of the binding energies (Table 2): thiol binding is initially favoured by the binding of another one but later becomes less favoured as the number of bonded thiols increases. Therefore, three are the maximum number of thiols that Au_4 can accommodate. Interestingly, charge transfer from the thiol groups to Au_4 upon binding increases significantly with respect to the previous scenario with only two thiols (0.6 vs 0.3 electrons for three or two methanethiols, respectively; compare solid or dashed Au_4 lines in Figure III-3). Clearly, **Au_4 becomes more basic as more thiols bind to it, and this prevents opening of the cluster.**

2. Proton Transfers and Hydrogen release.

As in the previous simulations with two thiols, the computed free energy surfaces (Figure III-5) exhibit several local minima (**I1** to **I8**). A preactivated configuration in which H' starts to interact with the cluster (**I2**) is also found (the S'-H' bond elongates by 0.12 Å with respect to the first intermediate; Table 4). Remarkably, the transfer of the first thiol proton (**I2** → **I4**) implies a much lower activation energy compared with the simulation with two thiols (18 vs 28 kcal/mol). This finding can be explained in terms of the differences in the Au4 electron density in both cases: **the greater the gold cluster electron density, the easier it is to receive a proton.** As was previously found, the detachment and subsequent transfer of H'' leads to the formation of molecular hydrogen (**I6** → **I7**).

Substantial atomic rearrangement occurs during the reactive process. Specially, one gold atom separates from the rest (**I4** → **I6**; see the evolution of the main Au-Au distances in Table III-4) and the methanethiolate group changes position to interact with two gold atoms, thus forming the precursor of a monomeric staple (**I7**). Finally, the hydrogen molecule leaves the cluster, and the monomeric staple comes into view (**P'**). As in the previous case, hydrogen transfer and formation of molecular hydrogen are the steps with the highest activation energies, with H_2 formation being rate-limiting.

During the entire process, the third methanethiol remains either unbind (I6) or weakly bonded to Au_4 (the binding energy of the S···Au interaction at **P'**, 8 kcal/mol, is significantly lower than the 15 kcal/mol value at **I1**). An attempt to transfer the proton of the third methanethiol (H''') to Au_4 failed, resulting in the complete detachment of

the methanethiol and the formation of $Au_4(SCH_3)_2$ (**P**). All of these observations suggest that, although a certain excess of thiol molecules participates in the formation of the staple motifs (thiol binding favors transfer from other thiols onto the cluster), they easily detach from the cluster.

	distance (Å)						
	I1	**I2**	**I4**	**I5**	**I6**	**I7**	**P**
Au-Au (d1)	2.67	2.70	2.8			3.27	3.1
Au-Au (d2)	2.71	3.01	2.9	2.7	2.93	2.90	2.94
Au-Au (d3)	2.65	2.59	2.71	2.60	2.8	2.66	2.8
Au-S’	2.38	2.59	2.30	2.35	2.47 2.39	2.42 2.5	2.47 2.38
Au-S"	2.50	2.4	2.6	2.5	2.44	2.48 2.44	2.60 2.44
H’-Au		1.79	1.7 1.9	1.71 1.80 2.08	1.60	1.8	
H"-Au					2.0	1.80	
Au-S’’’	2.55	2.3	2.54	2.45			2.5
H’-H"						0.85	0.75

Table III-4. Structural parameters of each minimum along the reaction pathway for the reaction of four methanethiols with Au_4.

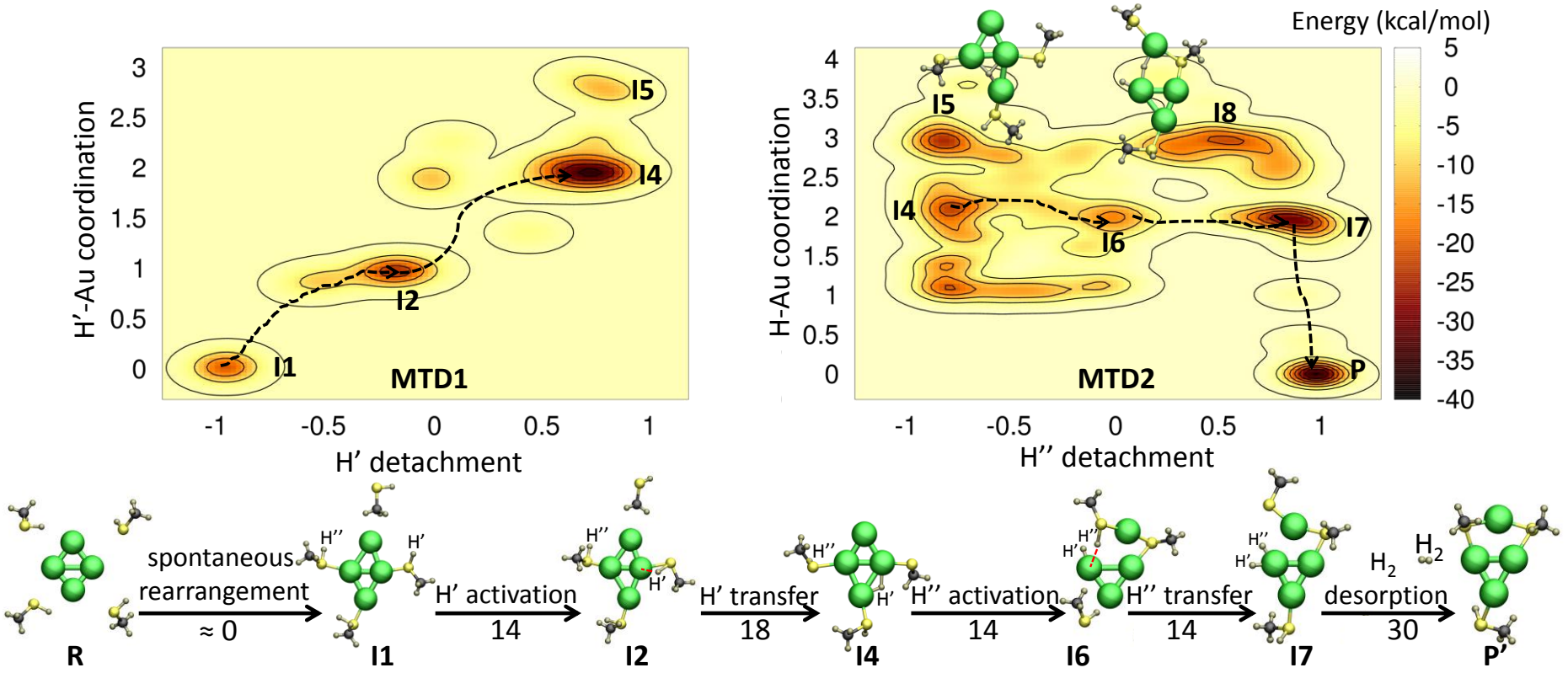

Figure III-5. Free energy surfaces reconstructed from the metadynamics simulations (MTD1 and MTD2) of the reaction of four CH_3SH molecules with Au_4. Contour lines are plotted every 2 kcal/mol. Au-Au solid bonds are drawn for distances of < 2.8 Å. H-Au solid bonds are shown for distances of < 1.85 Å, whereas dotted lines are used for bond distances in the range of 1.85-2.0 Å. Free energy barriers (in kcal/mol) are shown below the arrows.

Mechanism proposal

As mentioned above, a number of questions related to the formation of AuMPCs have emerged from recent studies, namely, how does the shape of the bare AuC change upon "thiol protection" and what is the fate of the thiol protons released during thiol binding (can they form molecular hydrogen as observed for thiol binding to gold surfaces? Are HS-Au and H-Au bond formed?). Because of the difficulties in trapping short-lived species along the reaction pathway, answering these questions using experimental probes is a major challenge. In fact, both thiol-gold and thiolate-gold interactions have been detected by spectroscopic and X-ray techniques, respectively.[60] Theoretical methods can be of major value in this area because they can offer the mechanistic insight needed to answer the above questions.

The picture emerging from the present *ab initio* study is as follows (see Figures III-4 and III-5). First, thiols attach to the bare AuC in an almost spontaneous process. In so doing, some gold atoms to which the thiols are attached may rearrange, decreasing the number of Au-Au interactions in which they participate. Second, transfer of protons from a fraction of the attached thiols occurs immediately, followed by migration through the AuC. Third, one gold atom with an attached thiol unbinds from the other gold atoms while making an Au-S bond with a neighbouring attached thiolate. This is the precursor of a staple motif. Fourth, the hydrogen atom of the thiol attached to this "peculiar" gold atom is transferred to the neighbouring gold with a singly bonded hydrogen. Fifth, formation of the new Au-S completes the staple, while H_2 desorbs from this gold atom.

All of these steps are expected to be easier (i.e. lower free energy barriers) for larger cluster sizes, where the number of gold atoms is higher and the system may be less tense. It should be noted that the two gold atoms anchoring the two methanethiolates of the monomeric staple are nearest neighbours in $Au_4(SCH_3)_2$ but next-nearest neighbours in larger clusters as $Au_{102}(SR)_{44}$ or $Au_{25}(SR)_{18}^-$. In these cases, the number of possible configurations would most likely increase, leading to a much richer energy landscape with a large number of alternative pathways. However, the essential steps should be the same. Thus, the full mechanism can be summarized as follows:

Thiol chemisorption	$Au_4 + 4\ CH_3SH$	→	$Au_4(SHCH_3)_3 + CH_3SH$
Transfer of one proton	$Au_4(CH_3SH)_3$	→	$CH_3S\text{-}Au_4H\text{-}(SHCH_3)_2$
Transfer of a 2nd proton	$CH_3S\text{-}Au_4H\text{-}(CH_3SH)_2$	→	$(CH_3S)_2\text{-}Au_4H_2\text{-}SHCH_3$
Release of H_2 and monomeric staple formation	$(CH_3S)_2\text{-}Au_4H_2\text{-}SHCH_3$	→	$(CH_3S)_2\text{-}Au_4\text{-}SHCH_3 + H_2$
Excess thiol release	$(CH_3S)_2\text{-}Au_4\text{-}SHCH_3$	→	$Au_4(SCH_3)_2 + CH_3SH$

An important conclusion that emerges from the calculations is that neutral bonded thiols can be present either in the reaction intermediates or as part of the final product, in agreement with the interpretations of spectroscopic experiments from solutions of AuMPCs.[60] However, thiol binding is found to be weak, and therefore these extra thiols can easily detach from the cluster (Figure III-5). Under this scenario, it is likely that the media surrounding the cluster determines the presence of neutral thiols in the final product. We believe this reconciles the fact that interact thiols on AuMPCs are detected in solution experiments but not in the crystalline phase.

Interestingly, thiol binding induces subtle but relevant electronic changes in the AuC. In particular, the cluster receives charge from the thiol groups (Figure III-3); i.e. it becomes more basic. The degree of charge transfer is more pronounced when there is an excess of bonded thiols, thereby indicating that such an excess facilitates proton transfer. In fact, our results clearly show that in this situation the energy barrier for the first proton transfer drops off considerably; i.e. an excess of bonded thiols favors the redox reaction. However, there is a maximum number of thiol groups that can bind to the cluster (three in our case); otherwise, the binding energy becomes positive. The balance between both factors will determine the number of bonded thiols. The results in Figures III-4 and III-5 also suggest that the cluster tends to form staples in such a way that appropriate “magic number” is reached (two in our case). Consequently, for larger clusters, the degree to which the initial number of valence electrons of the cluster deviates from the “magic number” may also be an important parameter in controlling the number of staples formed.

Conclusions

In conclusion, we find that one staple motif readily forms in a multiple-pathway chemical reaction when Au_4 interacts with methanethiols, releasing molecular hydrogen. Both H-Au bonds and intact Au-thiol bonds are present during the reactive process, but the former are weakly bonded and easily detach from the cluster. This explains why experimental detection of these species has been elusive and reconciles the interpretations of experimental studies in the crystalline and solution phases. Finally, we propose a mechanism in which an excess of bonded thiols triggers the redox reaction by facilitating proton-transfer events.

Chapter IV – Influence of the Water Solvent and Glutathione Coordination on the Structure and the Electronic Properties of AuMPCs

Influence of the Water Solvent and Glutathione Coordination on the Structure and Electronic Properties of AuMPCs

Introduction

Thiolate monolayer-protected gold clusters (AuMPC) have been extensively used as carriers of biological molecules, namely DNA[74], antibodies[29b] and specific proteins as glutathione-*S*-transferase.[75] The synthesis of various AuMPC of different sizes and using many kinds of ligands for protecting the gold cluster was reported recently.[12a, 20a] As a specific example, the structure of $Au_{25}(SR)_{18}^-$ (being R an alkyl substituent) was predicted computationally[19b] and verified experimentally by X-ray measurements of a crystallized sample.[20b] $Au_{25}(SR)_{18}^-$ is the smallest ligand-protected gold cluster that is stable. It is spherical and symmetric, constituted by an icosahedral Au_{13} gold core (one central atom and twelve atoms in the vertices of the icosahedra) and six $Au_2(SR)_3$ dimeric staple motifs[12a] that are protecting the gold core in an octahedral arrangement (Figure IV-1).

Small alkanethiolates such as 2-phenylethanethiolate[20b], ethanethiolate[76], or p-bromobenzenethiol[77] have been used as ligand units for $Au_{25}(SR)_{18}^-$. Furthermore, chiral ligands such as (*R* or *S*)-2-amino-2-phenylpropanethiolate ($SCH_2C^*H(NH_2)CH_2Ph$) and (*R* or *S*)-2-methyl-2-phenylthiolate ($SCH_2C^*H(CH_3)CH_2Ph$), have been used as well.[78] All these alkyl chains are known to be hydrophobic. Therefore, if these AuMPCs are used to bind biological molecules, there may be nonspecific hydrophobic interactions that interfere in the binding target. To eliminate these interactions, some functionalities need to be introduced to the AuMPCs to selectively bind the target molecules. Glutathione (GSH), a tripeptide formed by a glutamate, a cysteine and a glycine (Figure IV-2), is often used as a ligand[21, 79] because it is able to selectively bind proteins as glutathione-*S*-transferase[75] and single chain Fv antibody fragments.[29b]

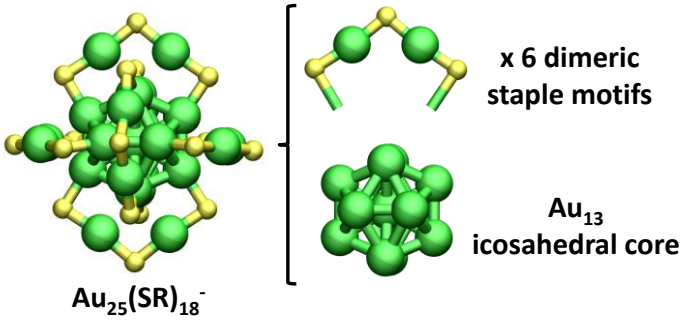

Figure IV-1. Description of the pieces that form the structure of $Au_{25}(SR)_{18}^-$. The organic ligands attached to the sulphur atoms are not shown for clarity. Six dimeric staple motifs are the protecting units that are bonded to the Au_{13} icosahedral core. Gold and sulphur are shown green and yellow, respectively.

The glutamate of GSH is bound to cysteine through a gamma peptide linkage between the amino moiety of cysteine and the terminal carboxylic acid of the glutamate side-chain. In turn, cysteine is attached to glycine by a normal peptide linkage (see Figure IV-2). Therefore, GSH can be written as γ-Glu-Cys-Gly, and it has 2 asymmetric carbons which introduce chirality. At physiological pH, both carboxylic acids are deprotonated (*i.e.* negatively charged), whereas the amino group is protonated. Therefore, GSH has a total charge of -1 in aqueous solution.

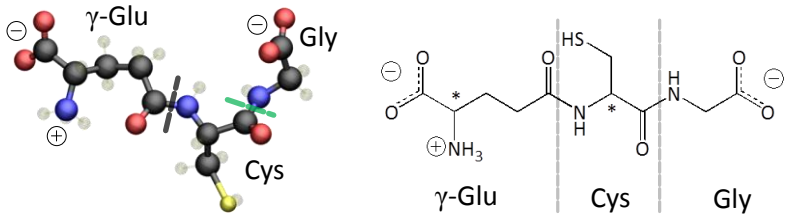

Figure IV - 2. Ball & stick (left) and chemical (right) representations of a GSH peptide. In the left view, hydrogens are shown transparent for clarity. The black dashed line shows the gamma peptide linkage between glutamate and cysteine. The green dashed line shows the normal peptide linkage between cysteine and glycine. In the chemical representation (right figure), asymmetric carbons are marked with stars.

Although it has been demonstrated that $Au_{25}(GSH)_{18}$ is very stable,[79b] its crystallographic structure has not been resolved yet. The presence of a bulky biological ligand carrying several charged groups (GSH) is expected to modify the structural and electronic properties of the AuMPC with respect to small monocharged ligands such as synthetic alkanethiol ligands. In this chapter, we assess these differences by investigating the structure and electronic properties of two clusters of formula $Au_{25}(SR)_{18}^-$ (R = glutathione or alkanethiolate) in vacuum and aqueous solution. For computational reasons, we have used methanethiolate, *i.e.* the smallest alkanethiolate, as a model of the synthetic ligands. This system already captures the essential chemistry at the Au-S interface layer.

We use *ab initio* molecular dynamics (AIMD) simulations based on the density functional theory (DFT) to study the properties of AuMPCs. One well known issue of standard DFT methods is that the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), *i.e.* the HOMO-LUMO gap, tends to be severely underestimated with respect to the experimental values, although tendencies are correct.[80] In principle, the absolute value of the HOMO-LUMO gap calculated by DFT should not be used as a reference for a direct comparison with optical measurements. Nevertheless, recent computational studies (with standard DFT) have reported values of the HOMO-LUMO gap for phosphine and thiolate protected gold (and silver) clusters in a very close agreement with the experimental ones.[69c, 81] Recently, the role of (implicit) solvent has been considered for small Au clusters,[82] and we think this aspect deserves more attention as simulation methods and computational resources are constantly improving. For that reason, we

have performed two sets of simulations for $Au_{25}(SR)_{18}^-$ in the vacuum and solvated in water. Moreover, we consider the full GSH ligand used in the experiments.[83] It is known that the use of a different ligand can induce distortions in the $Au_{25}(SR)_{18}^-$ framework via mechanochemical coupling.[12b] Therefore, we have quantified how the distortion of the AuMPC caused by the bulky biological ligand (GSH), as well as the water solvent, affects the HOMO-LUMO gap. Our results demonstrate that the previously reported agreement for the HOMO-LUMO gap is fortuitous, and that considering the real system (the full GSH ligands and explicit solvent conditions) reduces the DFT HOMO-LUMO gap by a factor of 2. This effect can be compensated by applying a hybrid DFT functional which includes an exact-exchange component.

Computational Methods

Systems Analyzed

Three different systems were prepared for the simulations:

i) **Isolated $Au_{25}(SCH_3)_{18}^-$:** the AuMPC is placed in a 20.3Å x 21.0Å x 21.3Å box. A full quantum mechanical (QM) description is used.

ii) **Solvated $Au_{25}(SCH_3)_{18}^-$:** the AuMPC is surrounded by 1465 water molecules. A sodium cation is also added to neutralize the negative charge of the AuMPC. The size of the full box is 36.9Å x 34.9Å x 35.7Å. 30 ns of classical molecular mechanics (MM) simulations are done to equilibrate the solvent and the alkane chains of the methanethiolate ligands (a graph of the RMSD during the equilibration is shown in the appendix II). For the subsequent QM/MM study, only the AuMPC is considered in the QM region, with the water molecules and the cation being part of the MM region.

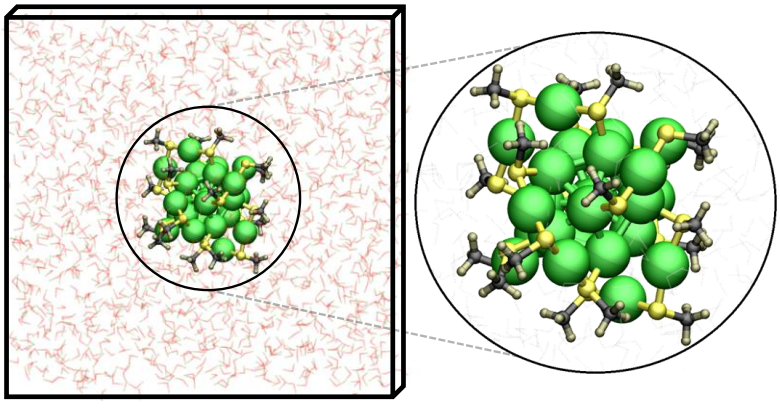

Figure IV-3. Left image: caption of the solvated $Au_{25}(SCH_3)_{18}^-$ system. **Right image:** zoom over the QM region (right). The QM region is shown in colour (Au in green, S in yellow, C in black and H in grey) whereas the MM zone is shown colourless.

iii) **Solvated $Au_{25}(GSH)_{18}^-$:** the AuMPC is surrounded by 3576 water molecules. 19 sodium cations are added, 18 to neutralize the negative charge of the GSH and 1 to neutralize the global charge of the AuMPC. The size of the full box is 53.0Å x 55.5Å x 53.7Å. 8 ns of classical MM simulation are done to equilibrate the solvent and the peptides. Solvent is equilibrated in less than 2 ns, although peptides need a bit more than 3 ns (a graph of the RMSD during the equilibration is shown in the appendix II). We have considered a *S* configuration for the asymmetric carbons (it is known that there is no influence in considering S or R configuration).[69c] For the subsequent QM/MM study, the 25 gold atoms and the 18 cysteines (only the S-CH_2-C- part of cysteine) of each GSH form the QM region (115 atoms). The rest of the cysteine, glutamate, glycine, 19 sodium cations and 3576 water molecules are took into account in the MM part.

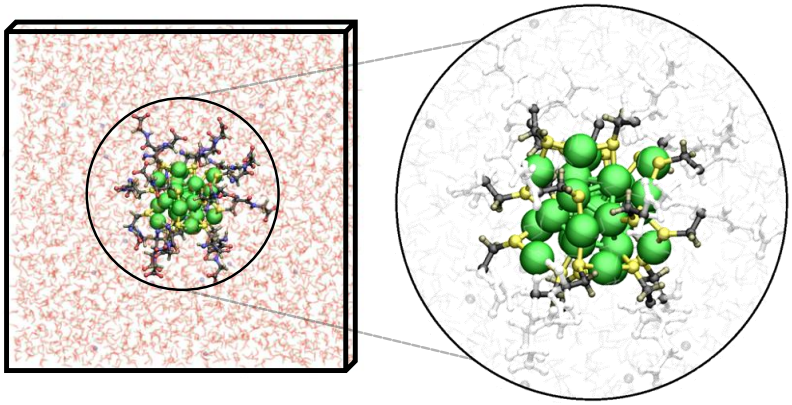

Figure IV-4. Left image: caption of the solvated $Au_{25}(GSH)_{18}^-$ system. **Right image:** zoom over the QM region (right). The QM zone is shown in colour (Au in green, S in yellow, C in black and H in grey) whereas the MM zone is shown colourless.

Simulation Details

The MM simulations with classical potentials are performed using the NAMD[84] software. Ligands are modelled with the FF99SB force field,[85] and water molecules are described with the TIP3P force field.[86] Snapshot of the MM equilibrated system is taken for the subsequent QM/MM simulations.

AIMD simulations are performed within the Car-Parrinello[64] approach, which is based on DFT, using the CPMD package.[87] A fictitious mass of 600 a.u. and an integration time step of 0.12 fs guarantee a good control of the adiabaticity for ionic and electronic equations of motion in all the systems where QM or QM/MM molecular dynamics simulations are performed. The simulation temperature is set to 300 K by coupling it to a Nosé-Hoover chain thermostat for the ionic degrees of freedom.[88] The Kohn-Sham orbitals are expanded in a plane wave (PW) basis set with a kinetic energy cutoff of 90 Ry. The exchange-correlation functional employed the parametrization by Perdew-Burke-Ernzerhoff (PBE),[68] which has been widely used to describe AuMPCs.[16, 66-67, 89] Furthermore, a hybrid functional (PBE0[90]) is used to compute HOMO-LUMO

gaps for the optimized structures in order to evaluate the effect of including a portion of the exact exchange interaction on the electronic properties. The gold atoms are represented by a Goedecker[65] type pseudopotential ($5d^{10}6s^{1}$ valence), carbon and sulphur by a Troullier-Martins[45] type, and hydrogen by a Car-von Barth type, as in references 66 and 67. For geometry optimization, an annealing factor is introduced for molecular dynamics (MD) and its value has been gradually incremented from 0.9 to 0.999 to remove kinetic energy from the ionic degrees of freedom. We have considered that the structure is optimized when the largest nuclear gradient is lower than $5.0 \cdot 10^{-4}$ a.u..

To study $Au_{25}(GSH)_{18}^{-}$ in an aqueous environment, the QM/MM interface is modelled by the use of a monovalent-carbon pseudopotential that saturates the QM region.[91] The electrostatic interactions in the interface of the two regions are described in the Chapter II of this thesis.

Structures and trajectories are analyzed using standard tools of VMD.[92] VMD is also used to compute other important parameters that provide structural information as the radius of gyration (r_{gyr}), the solvent accessible surface area (SASA) and radial distribution functions (g(r)):

i) ***Radius of gyration, r_{gyr}***: is used to describe the dimensions of a molecule. Is defined as the radius of a group of atomic positions $\{r(1), r(2), \dots, r(n)\}$ with respect to their weighted centre, $\bar{r}$:

$$r_{gyr} = \sqrt{\left(\sum_{i=1}^{n} w(i)|r(i) - \bar{r}|^2\right) \Big/ \left(\sum_{i=1}^{n} w(i)\right)} \quad \textbf{(Eq.IV-1)}$$

where $w(i)$ is the weight of each atom.

ii) ***Solvent accessible surface area***: is the area of the surface of a molecule that is accessible to a given solvent. It is calculated by tracing the centre of a probe sphere that is rolling along the Van der Waals surface of the molecule of interest. This probe sphere has a radius of 1.4 Å when the solvent is water.

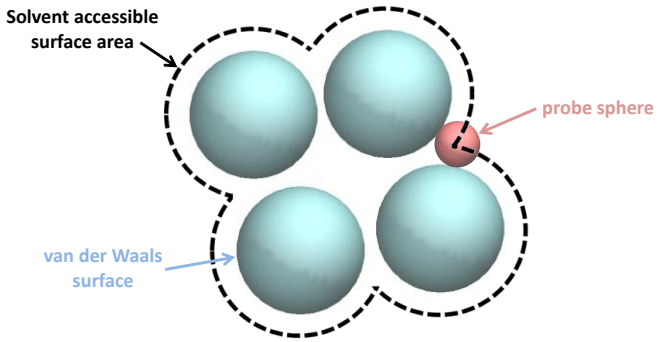

Figure IV - 5. Schematic representation of the solvent accessible surface area, depicted here by the black dashed line. The Van der Waals surface of the

atoms of the molecule of interest is rendered in blue. The probe sphere is shown red.

iii) ***Radial distribution function, g(r)***: is the probability of finding a particle at a certain distance r from a reference particle. It describes the variation of the density as a function of the distance from the reference particle. It is evaluated by determining how many particles are inside a crown of radii's r and r + dr.

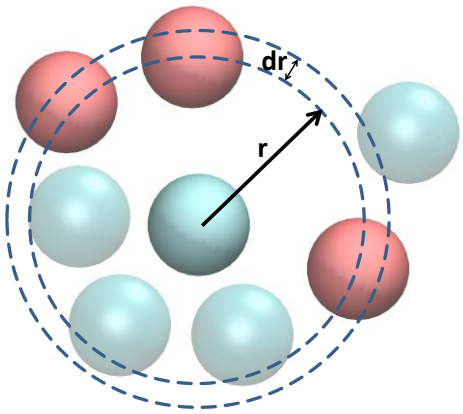

Figure IV-6. Schematic representation of the evaluation of g(r). The blue sphere in the centre is the reference particle. The three red particles are the ones that should be taken into account since they are inside the crown. The light blue particles are or too close or too far from the reference particle and they do not have to be taken into account.

Results

Structural analysis of $Au_{25}(SCH_3)_{18}^-$

For $Au_{25}(SCH_3)_{18}^-$, the effect of the water solvent was investigated by comparing the properties of the isolated and solvated systems (water were considered classically). The average structure over 7.5 ps of room temperature QM/MM MD simulation has been considered for analysis.

Different types of gold and sulphur atoms are found in AuMPCs, as it is explained in the page 5 of the chapter I. This labelling procedure is applied to $Au_{25}(SCH_3)_{18}^-$, as can be seen in the upper image of Figure IV-7. There are three types of gold atoms (a, a' and c) and two types of sulphur atoms (S and S_{ap}) in $Au_{25}(SR)_{18}^-$. Six distances are considered as the key parameters to analyze the structure (Table IV-I). To facilitate the structural analysis, radar charts are constructed, taking $Au_{25}(SCH_3)_{18}^-$ in the gas phase as the reference system. The vertices of the hexagonal chart correspond to the value of each individual distance as a percentage with respect to the reference system, $Au_{25}(SCH_3)_{18}^-$ in the gas phase:

$$\text{Value represented in the graph} = \frac{\text{computed distances for each system}}{\text{isolated } Au_{25}(SCH_3)_{18}^- \text{ computed distances}} \cdot 100$$

The background of the hexagonal plot is formed by the six values of the isolated $Au_{25}(SCH_3)_{18}^-$ (corresponding to 100%). The same procedure is used in order to represent the three key angles of the AuMPC in a triangular graph.

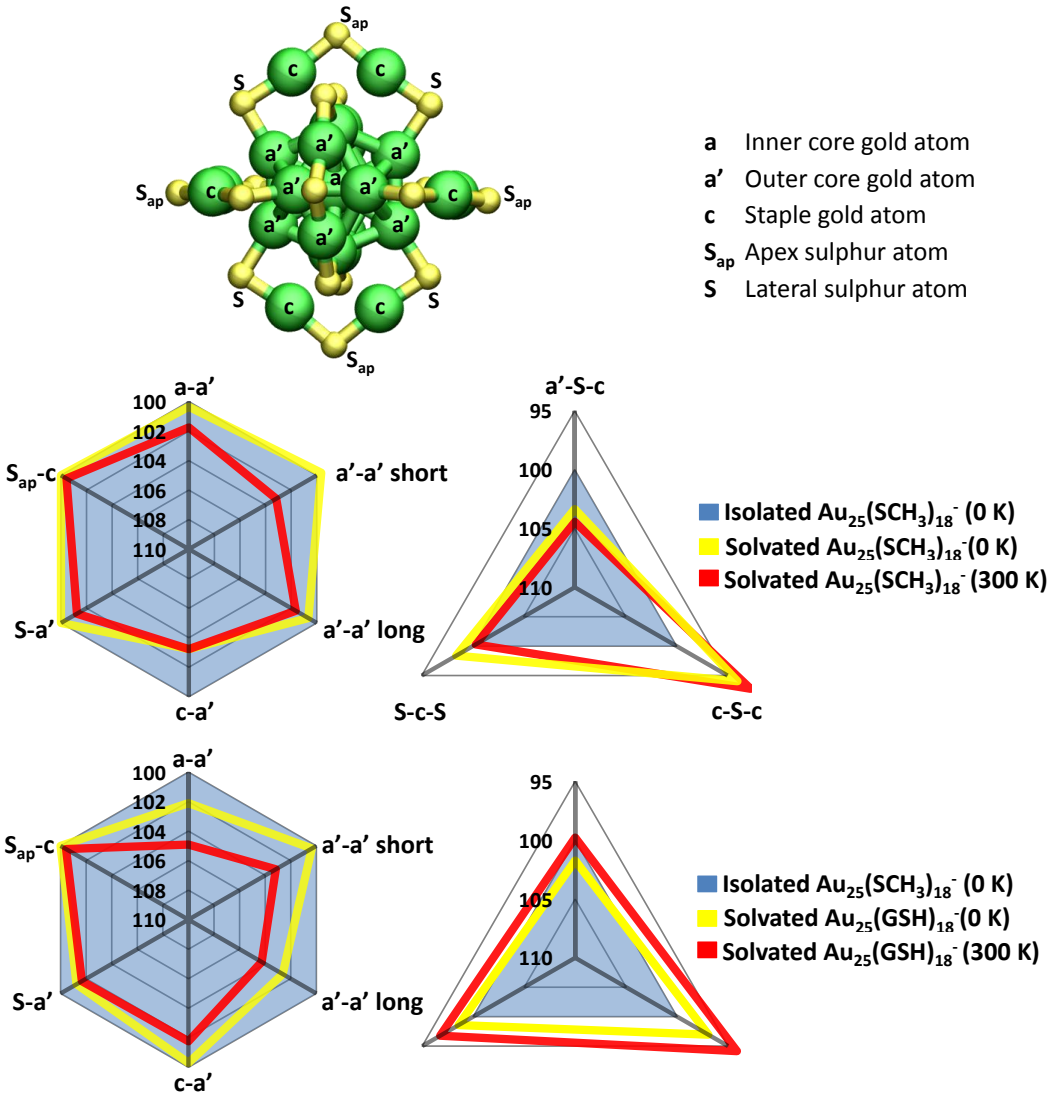

Figure IV-7. Top: $Au_{25}(SR)_{18}^-$ atom labelling used through the study. Colour codes: Au (green), S (yellow). **Center:** Representative bond distances (left) and angles (right) of $Au_{25}(SCH_3)_{18}^-$ isolated and solvated in water. **Bottom:** Representative bond distances (left) and angles (right) of $Au_{25}(GSH)_{18}^-$ solvated in water with respect to the isolated $Au_{25}(SCH_3)_{18}^-$. The labels 0 K and 300 K are used to distinguish the optimized structure with respect to the average structure at room temperature, respectively. Values in the chart axes correspond to a percentage with respect to the reference $Au_{25}(SCH_3)_{18}^-$ system in vacuum (shaded area).

The results of Table IV-1 show that the solvent does not affect the global structure of $Au_{25}(SCH_3)_{18}^-$, as the Au_{13} core and the six staple motifs are preserved. However, the cluster expands: the largest difference between the isolated and the solvated $Au_{25}(SCH_3)_{18}^-$ at 0 K correspond to Au(c)-Au(a') distance (involving outer core

gold atoms and those of the staple motifs), which increases by 0.11 Å (3.2%). As a consequence, the Au(c)-S_{ap}-Au(c) angle (involving the apex S atom and the two neighbouring gold atoms) and the Au(a')-S-Au(c) angle at the core-ligand interface, change by -7.3^{o} and 4.1^{o}, respectively. During the MD simulation (300 K), the average values of the Au-Au bonds show a further increase up to 0.10Å (for Au(a')-Au(a')), while the S-Au bonds remain almost intact (within ± 0.03 Å). The S-Au(a') bonds that link the staple motifs to the Au_{13} gold core and their variation during the 7.5 ps of QM/MM MD simulations at 300 K are illustrated in Figure IV-8.

	$Au_{25}(SCH_3)_{18}^-$			$Au_{25}(GSH)_{18}^-$	
	vacuum (0K)	**solvated (0 K)**	**solvated (300 K)**	**solvated (0 K)**	**solvated (300 K)**
Au(a)-Au(a')	2.85	2.86	2.90	2.91	2.99
Au(a')-Au(a')	2.84	2.83	2.93	2.85	2.93
Au(a')-Au(a')	3.04	3.06	3.09	3.12	3.17
Au(c)-Au(a')	3.41	3.52	3.52	3.42	3.47
S-Au(a')	2.40	2.40	2.43	2.43	2.44
S-Au(c)	2.32	2.32	2.33	2.32	2.33
Au(a')-S-Au(c)	92.5	96.6	95.60	92.3	93.9
Au(c)-S_{ap}-Au(c)	100.6	93.3	94.60	94.7	91.8
S-Au(c)-S	172.9	173.4	170.10	167.3	164.9

Table IV-1. Most relevant parameters defining the $Au_{25}(SR)_{18}^-$ systems investigated.

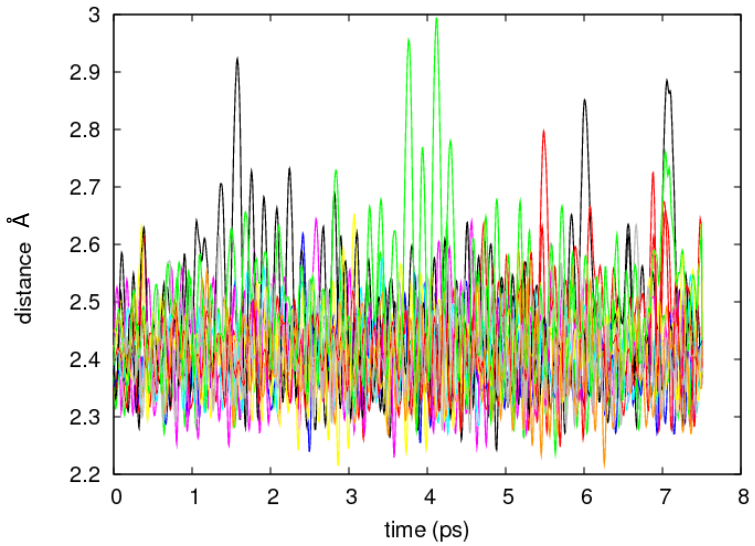

Figure IV-8. Variation of the twelve S-Au(a') bonds of the six dimeric staple motifs with the gold atoms of the icosahedral Au_{13} core along the 7.5 ps of QM/MM MD simulations at 300 K.

Structural analysis of $Au_{25}(GSH)_{18}^-$

The structure of $Au_{25}(GSH)_{18}^-$ is compared with the one obtained for isolated $Au_{25}(SCH_3)_{18}^-$ to investigate the effect of introducing a bulky biological ligand (GSH) instead of the small inorganic alkanethiolates used in the synthetic methods. To analyze the structures, a similar methodology to the one used in the previous section is used here, using again hexagonal and triangular charts to show the deviations from a reference system, the $Au_{25}(SCH_3)_{18}^-$ in vacuum. Therefore, now we evaluate the values in the bottom image of Figure IV-8 as:

$$\text{Value represented in the graph} = \frac{\text{computed distances for each system}}{\text{solvated } Au_{25}(GSH)_{18}^- \text{ computed distances}} \cdot 100$$

The GSH ligands do not change the overall structure of AuMPC, since the six protecting units and the Au_{13} core are preserved. However, significant variations are found in the distances in comparison with the methanethiolate ligands (Figure IV-7 and Table IV-1), and in particular, Au(a')-Au(a') bonds increase up to 0.08 Å. Furthermore, the increase of the Au-Au bond distances is even more pronounced at room temperature (by 0.09-0.14 Å in the core) in comparison with $Au_{25}(SCH_3)_{18}^-$ in the vacuum. This is due to the solvent effects and steric hindrance of the bulky ligands on the AuMPC surface, which modify the Au-Au bonding via mechanochemical coupling. The calculated radii of gyration confirm this finding: 4.12 and 4.10 Å for the full Au_{25} gold cluster or 1.49 and 1.44 Å for only the Au_{13} gold core, for ligands being GSH and SCH_3, respectively. The Au_{13} core slightly expands when GSH is used, in order to better accommodate the bulky ligands, and the first shell water solvent causes electrostatic effects on the electron density. Instead, the Au(a')-S-Au(c) angle decreases for GSH due to the formation of hydrogen bond interactions among the GSH ligands.

GSH contains three charged groups, namely two carboxylates ($-COO^-$) and one amino ($-NH_3^+$), besides two amide groups (-CO-NH-). The tendency of these groups to form hydrogen bonds with each other causes ligand rotation. Moreover, GSH has two chiral carbon atoms. To demonstrate the effect of the ligand, Figure IV-9 shows the change in the AuMPC symmetry from C_i for the methanethiolate ligand to C_1 for GSH. Such transformation was previously suggested in reference 69c, where N-acetyl-cysteine (NAC) was used to mimic GSH. A histogram of the main distances along the AIMD simulation of $Au_{25}(GSH)_{18}^-$ is shown in Figure IV-10. The fluctuations for the S-Au bonding are small (narrow first peaks in Figure IV-10). For the Au-Au bonds, the distributions are very broad, which demonstrates the flexibility of these bonds, termed previously as "fluxionality" for bare Au clusters.[93] Hence, the introduction of bulkier ligands can easily induce changes in the Au_{13} core.

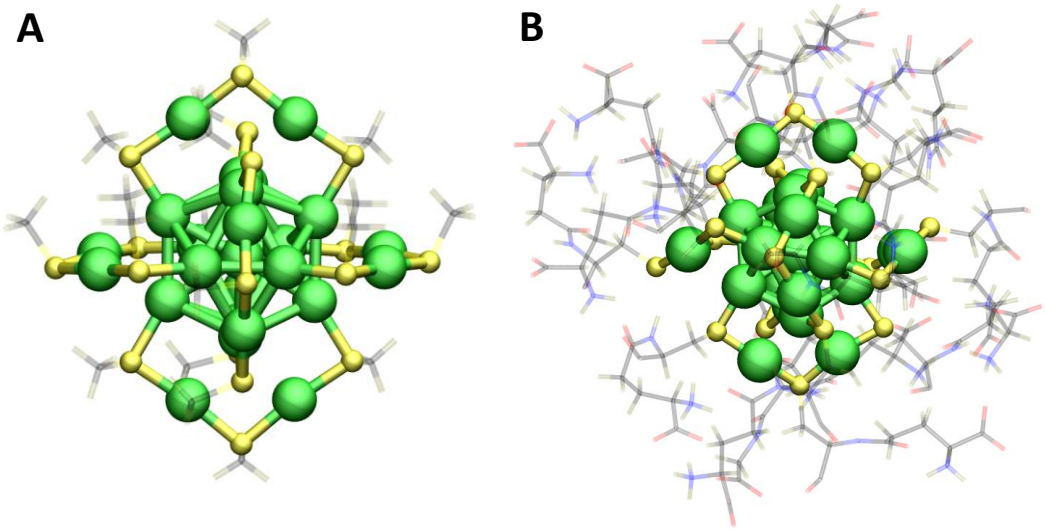

Figure IV-9. Structures of $Au_{25}(SCH_3)_{18}^-$ (A) and $Au_{25}(GSH)_{18}^-$ (B) solvated in water (the solvent is not shown for clarity). A belongs to a C_i point group while B to the C_1. Colour codes: Au (green), S (yellow), C (black), N (blue), O (red) and H (grey). Methanethiolate and glutathione are shown semitransparent.

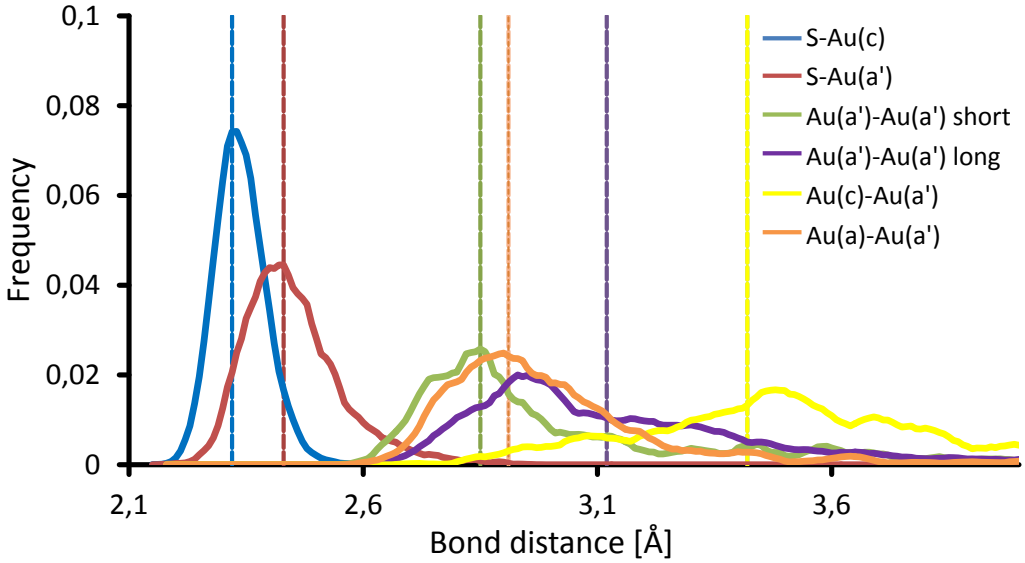

Figure IV-10. Histogram of the six main distances that define the structure of $Au_{25}(GSH)_{18}^-$ during a 7.5 ps MD QM/MM simulation. The dashed lines show the optimized values for each distance.

Solvent Accessible Surface Area

Another structural effect of the GSH ligands concerns to the modification of the water structure around the cluster. In fact, these ligands also provide better "shielding" than small alkanethiolates, which protects the Au core more effectively from external agents. To directly visualize this effect, the solvent accessible surface area (SASA) of AuMPCs and all water molecules that are within 4.5 Å from any gold atom are represented in Figure IV-11. The SASA value of the Au core is very different for the two systems: 16.5 $Å^2$ and 3.5 $Å^2$ for SCH_3 and GSH, respectively. All the gold atoms of the surface of $Au_{25}(SCH_3)_{18}^-$ are solvent-exposed (left image in Figure IV-11). On the contrary, the space surrounding Au atoms in $Au_{25}(GSH)_{18}^-$ is mostly occupied by the GSH ligands, which effectively prevent the solvent from approaching the metallic core. However, a few water molecules are still able to cross the shielding of peptides exposing some Au atoms to the solvent.

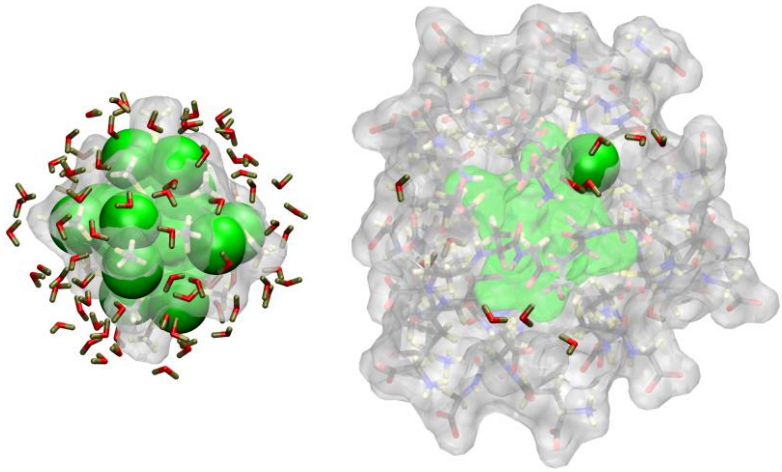

Figure IV-11. Solvent accessible surface area of $Au_{25}(SCH_3)_{18}^-$ (left) and $Au_{25}(GSH)_{18}^-$ (right). Water molecules that are up to 4.5 Å from gold atoms are also shown. The Van der Waals surface of methanethiolate or GSH is shown as a semitransparent surface to illustrate their shielding effect. Gold atoms that are more exposed to the solvent are depicted in light green. Colour codes: Au (green), S (yellow), N (blue), O (red), C (black) and H (grey).

Another way to analyze this effect is by plotting the radial distribution function of the solvent atoms with respect to the different atoms of the AuMPC, as shown in Figure IV-12. When methanethiolate is used as ligand (red curve), there are more waters at 4.5 Å, which form the first solvation sphere (see difference in intensity of the peaks at 4.5 Å in Figure IV-12). In addition, the second solvation sphere (peak at 7 Å in the red curve of Figure IV-12) is lost when GSH is the ligand, maybe due to the strong hydrogen bonding interactions of either oxygen or hydrogen of waters and the charged groups of GSH. This hydrogen bonding can be seen in the radial distribution function between oxygen of waters and GSH or SCH_3 (purple and green curves of Figure IV-12, respectively). When GSH is used, there are some water molecules that

are forming hydrogen bonds with it, being the peak at 1.8 Å a proof of that. On the other hand, when SCH_3 is used, they do not have strong interactions with waters, since the methyl group is hydrophobic (Figure IV-13). Therefore the peak is not observed at 1.8 Å in Figure IV-12.

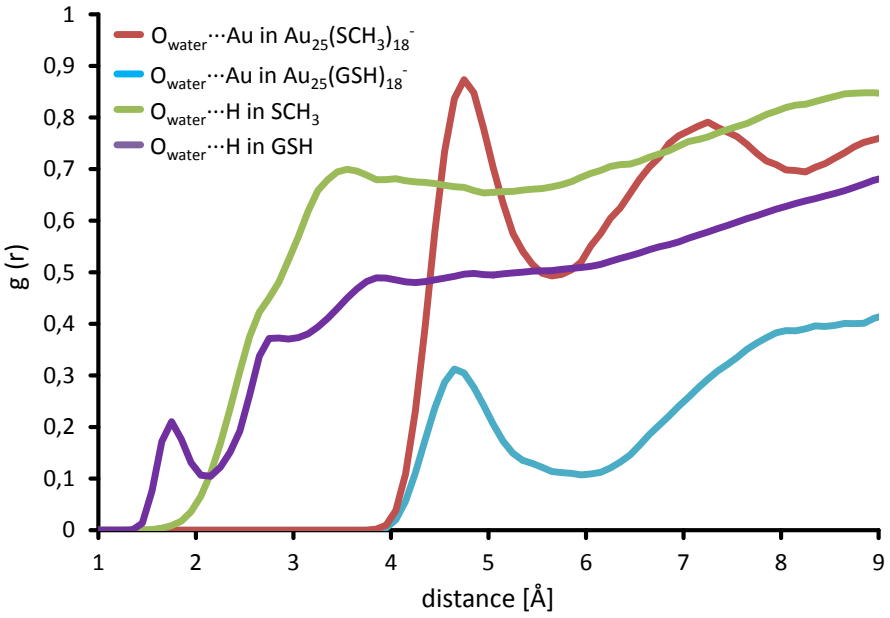

Figure IV-12. Radial distribution functions of the oxygen of water molecules towards the gold atoms (O_{wat}··· Au in $Au_{25}(SR)_{18}^-$; red and blue curves) and towards the corresponding ligands (O_{wat}···H in SR; green and purple curves).

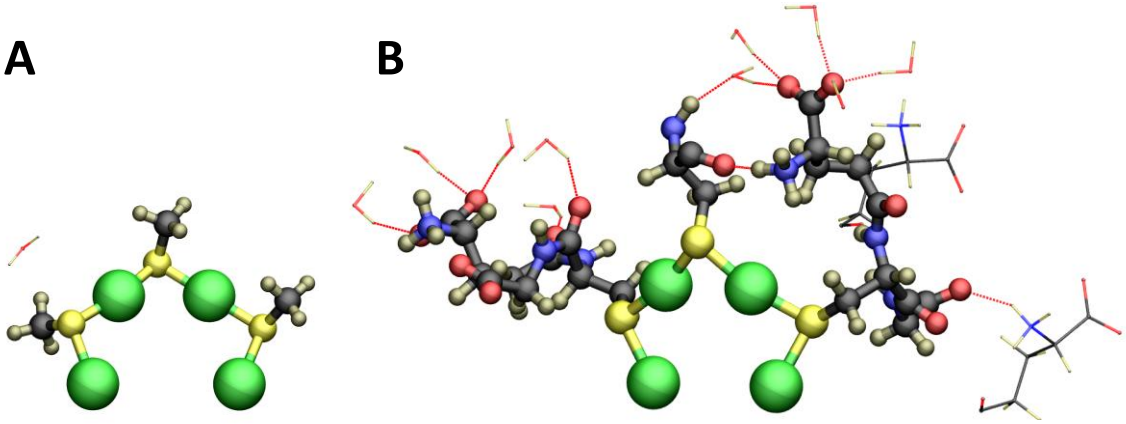

Figure IV-13. A: One dimeric staple motif in $Au_{25}(SCH_3)_{18}^-$ and the single water that is less than 2 Å from the methanethiolates. There is no hydrogen bonding between the solvent and the ligands here. **B:** Hydrogen bonding of GSH in one dimeric staple motif in $Au_{25}(GSH)_{18}^-$. Dashed red lines represent hydrogen bonds. GSH of the highlighted staple motif is represented by ball-and-stick while the other ligands or water molecules are reperented by sticks. Colour codes: Au (green), S (yellow), N (blue), O (red), C (black) and H (grey).

Charge Distribution

Atomic (Hirshfeld[94]) charges of $Au_{25}(SCH_3)_{18}^-$ and $Au_{25}(GSH)_{18}^-$ are computed to analyze the effect of the solvent and the type of ligand on the electronic structure of the AuMPC. Figure IV-14 displays the atomic charges for the different types of gold and sulphur atoms using radar charts (see Table AII-I in the appendix II for the full list of charges). Each one of the vertices of the pentagon marks an individual atomic charge normalized with respect to the reference charge (in percentage), namely the isolated $Au_{25}(SCH_3)_{18}^-$:

$$\text{Value represented in the graph} = \frac{\text{charge for each system} - \text{reference charge}}{\text{reference charge}} \cdot 100$$

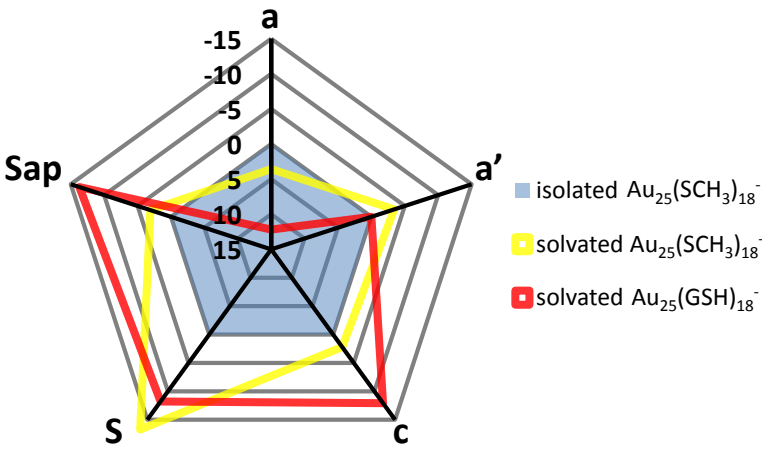

Figure IV-14. Change of the atomic Hirshfeld charges for the most representative atoms of $Au_{25}(SCH_3)_{18}^-$ and $Au_{25}(GSH)_{18}^-$ solvated in water with respect to the ones obtained for the isolated $Au_{25}(SCH_3)_{18}^-$ (blue area).

As shown in Figure IV-14, introducing the solvent makes the sulphur atoms of $Au_{25}(SCH_3)_{18}^-$ more negative (-0.263 and -0.423 in vacuum and -0.307 and -0.436 solvated in water, for S and S_{ap}, respectively), and this effect is more pronounced for S (lateral) than for S_{ap} (apex). The corresponding charge is partially transferred from the negatively charged central Au atom (a) which has an effective charge of -0.65*e*. This effect is more pronounced when the ligand is GSH, and hence the staple motifs become less electrophilic. In this case, both sulphur atoms experience a similar change but the values of the effective charges are different (-0.29*e* and -0.48*e* for S and S_{ap}, respectively). The Au atoms in the staple motifs Au(c) are positively charged (+0.40*e*) reflecting their different oxidation state. Despite the changes in the staple motifs, the gold the gold atoms charges in the Au_{13} icosahedron vertices are insensitive to the changes in environment.

HOMO-LUMO gaps

Previous analysis has shown that the electronic structure of $Au_{25}(SR)_{18}^-$ corresponds to a closed-shell configuration of an 8-electron superatom with 1S (two electrons) and 1P (six electrons, degenerate HOMO) orbitals, and that the LUMO orbitals correspond to the 1D symmetry. The shapes of the canonical Kohn-Sham orbitals (HOMO and LUMO) are depicted in Figure IV-15.

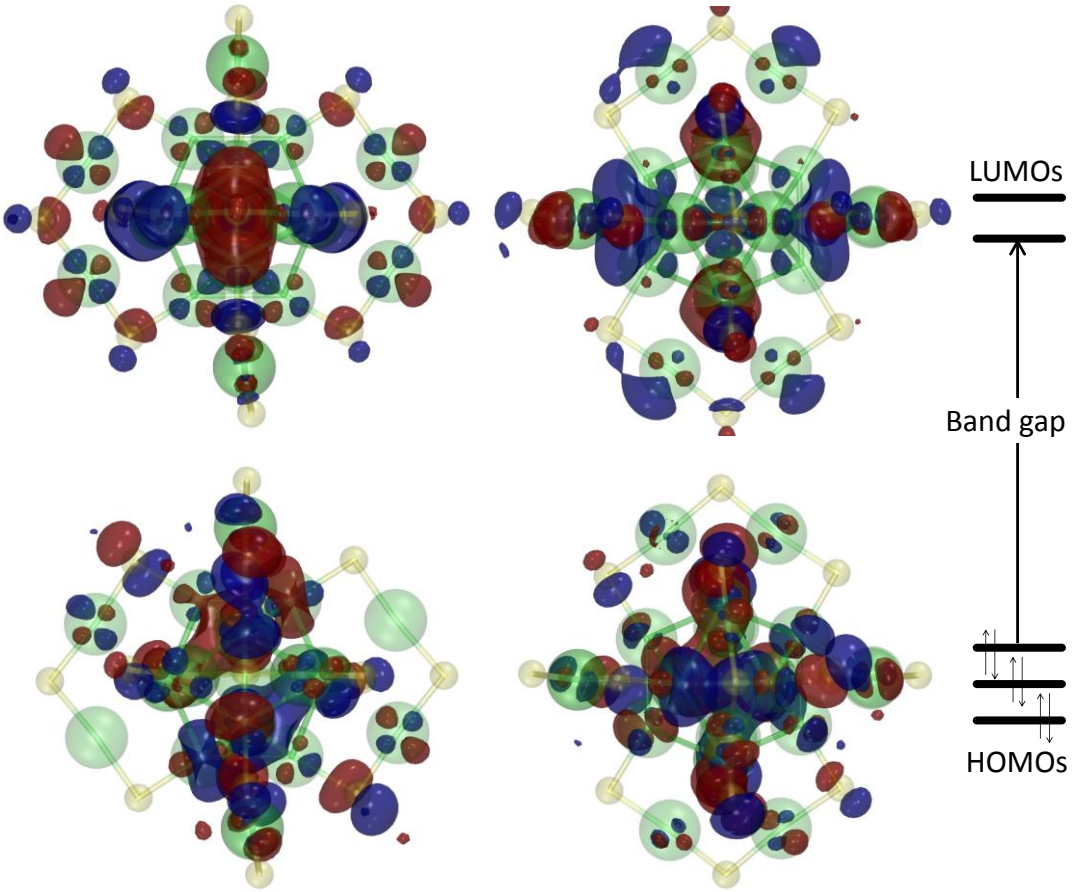

Figure IV-15. Shape of the frontier orbitals for $Au_{25}(SR)_{18}^-$. HOMO and LUMO wavefunctions are rendered from two points of view to better visualize its shape. The wavefunction is depicted blue and red to show the two signs that it can adopt. The AuMPC is shown transparent for clarity.

The calculated values of the HOMO-LUMO gap are reported in Figure IV-16. Considering that the experimental HOMO-LUMO gap is 1.3 eV,[83] it can be seen that:

- The presence of the solvent reduces the HOMO-LUMO gap by 22% for the systems with SCH_3 as ligand, with respect to the same system in vacuum (1.19 eV and 0.93 eV, respectively). Therefore, the effect of the solvent on the HOMO-LUMO gap is significant. Taking into account that the experimental gap (1.3 eV), is higher than the gas phase computed value, the reduction obtained by adding the solvent is in the wrong direction.

- The HOMO-LUMO gap is expected to depend on the ligand, as a consequence of the charge transfer effects and the variation in AuMPC symmetry (it has been reported that distortions in the AuMPC structure decrease the HOMO-LUMO gap[69c]). In fact, replacing SCH_3 by GSH distorts the AuMPC, which adopts a C_1 symmetry instead of the original C_i symmetry of $Au_{25}(SCH_3)_{18}^-$ (Figure IV-9). Most importantly, the change from SCH_3 to GSH reduces the HOMO-LUMO gap by 31% (0.93 eV and 0.65 eV for methanethiolate or glutathione, respectively), *i.e.* the resulting value is now a factor of 2 lower than the experimental one.

- It is well known that hybrid functionals such as PBE0 give more accurate results than GGA functionals for HOMO-LUMO gaps. Correspondingly, PBE0 produces a value (1.54 eV) which is close to the experimental one once the simulated system mimics the original experiment by taking into account the solvent and considering the same ligand, *i.e.* $Au_{25}(GSH)_{18}^-$ solvated in water.[82] Therefore, the previously reported poor performance of PBE0 for HOMO-LUMO gaps (severe overestimation[81a]) is related to the limited description of the simulated system itself (absence of solvent, different ligand). For example, PBE0 gives a value of 2.20 eV for $Au_{25}(SCH_3)_{18}^-$ in the gas phase, which is much larger than the experimental one, highlighting the importance of GSH ligands and solvent environment.

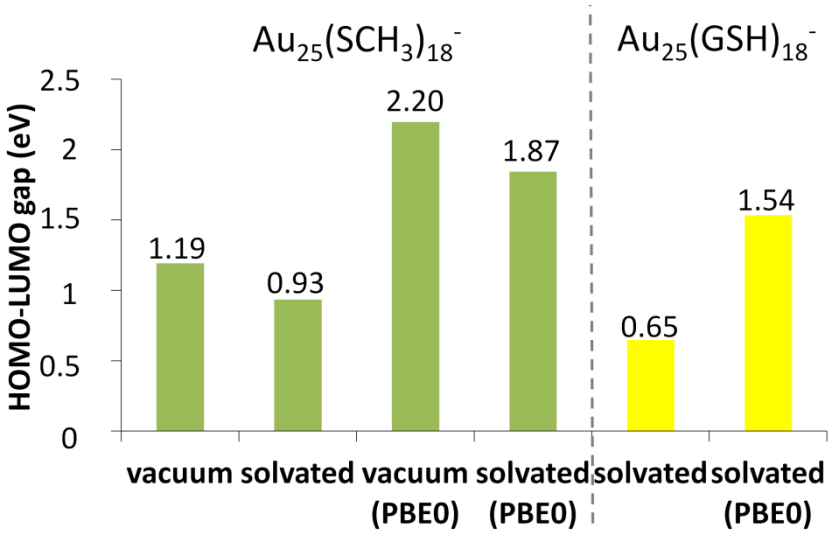

Figure IV-16. HOMO-LUMO gap for the different studied systems with PBE and PBE0 exchange-correlation functionals. Green bars correspond to $Au_{25}(SCH_3)_{18}^-$, whereas yellow bars refer to $Au_{25}(GSH)_{18}^-$. The value of the HOMO-LUMO gap is written above each bar.

Conclusions

As a summary, we have performed QM/MM simulations for the $Au_{25}(SCH_3)_{18}^-$ and $Au_{25}(GSH)_{18}^-$ clusters in the gas phase and solvated in water in order to elucidate the effect of introducing a biological ligand (GSH) and explicit water solvent. Our findings demonstrate that:

- The general structure of $Au_{25}(SR)_{18}^-$ is not modified by the fact of considering an aqueous solvent and neither by changing the ligand. The Au_{13} core and the six dimeric SR-Au-SR-Au-SR staple motifs are always preserved. However, the Au-Au bond distances and angles between S and Au experience visible changes under different conditions.

- The shielding effect of the peptides is larger than that for methanethiolates due to their larger size, and GSH protects the gold surface towards nucleophilic attacks of external agents that can lead to undesired interactions. While methanethiolate is hydrophobic, the GSH ligands actively form hydrogen bonds with water and other neighbouring peptides.

- Upon changing the ligand from SCH_3 to GSH, the changes in charge distribution affect the HOMO-LUMO gap, which is further modified by the loss of symmetry from *Ci* to C_1 due to the flexible nature of Au-Au bonds. A lower symmetry is associated with a smaller HOMO-LUMO gap.

Our main finding is that the HOMO-LUMO gap of AuMPCs depends sensitively on the ligands and solvent environment. In particular, the role of solvent has been typically neglected in the theoretical studies of monolayer-protected Au and Ag clusters, which have reported HOMO-LUMO gaps and/or optical absorption spectra in good agreement with experiments for standard DFT functionals (generalized gradient approximation, GGA).[69c, 81, 95] For example, we demonstrate here for $Au_{25}(SCH_3)_{18}^-$ that the water solvent decreases the HOMO-LUMO gap by 22% due to charge redistribution effects. Furthermore, the introduction of GSH ligands reduces the band gap more and the resulting value (0.65 eV) underestimates the experiment by a factor of 2. Therefore, the good performance of GGA functionals (such as PBE) for HOMO-LUMO gaps is due to a compensation of errors, and not because the level of theory is describing the particular system well. The calculations of AuMPCs in the vacuum give systematically larger values for the HOMO-LUMO gap than in the solvent. Hence, these larger values compensate the fact that DFT underestimates band gaps in general. The PBE0 functional which includes improvements to PBE is able to produce a computed value which is comparable with the experimental one once the real system is described accurately, *i.e.* the AuMPC composition (ligand) is the same as in experiments and the system is surrounded by explicit solvent.

Chapter V - Ligand Exchange Reaction of an Antibody Towards an AuMPC

Ligand Exchange Reaction of an Antibody Towards an AuMPC

Introduction

Binding of biomolecules to thiolate monolayer-protected gold clusters (AuMPCs) has aroused much interest in biomedicine.[22] The fact that small enough nanoparticles (≈ 10 nm) can cross the blood–brain barrier without altering its integrity[25] opens the door to possible applications for the treatment of central nervous system diseases as Alzheimer. Furthermore, AuMPCs are used to protein labelling, because of its optical properties.

AuMPCs have been largely synthesized using the so-called Brust-Schiffrin method.[5] In this direct synthesis, tetrachloroaurate salts ($AuCl_4^-$) are reduced by sodium borohydride ($NaBH_4$) in a two-phase redox reaction using an appropriate mixture of methanol and water[8] (see page 4 of Chapter I for a more detailed information about AuMPCs synthetic methods). However, many functional groups are incompatible with these reductive conditions. Therefore, another method should be used to introduce functionalities or molecules that are sensible to a reductive environment. Murray place exchange reaction[96] (hereafter named as ligand exchange reaction) has been largely used in order to introduce these new functionalities to already prepared AuMPCs. Binding of biomolecules toward gold clusters is mainly carried out through this ligand exchange reaction.

The ligand exchange reaction could be described as two S_N2-like reactions where the incoming thiol plays the role of the nucleophile and one alkanethiolate of the AuMPC acts as the leaving group and the base, simultaneously (Figure V-1). In the first step, the nucleophilic sulphydryl group of the incoming thiol (blue molecule in Figure V-1) attacks an electrophilic gold atom while the leaving alkanethiolate group (red molecule) breaks its bond with this gold. During this step, a hydrogen bond between the incoming thiol and the leaving alkanethiolate is formed. In the second step, the sulphur of the incoming thiol performs a second nucleophilic attack to another gold atom while the leaving group departs from the staple acting as a base, i.e. deprotonating the sulphydryl group. Initially, the leaving alkanethiolate group is bonded to oxidized gold atoms, forming the protecting unit known as staple motifs.[12a] Regardless of what kind of staple motif they form, alkanethiolates are always linked to two gold atoms. During the two S_N2 reactions these two bonds are broken because of the two consecutive nucleophilic attacks of the incoming thiol, which occupies the leaving alkanethiolate position.

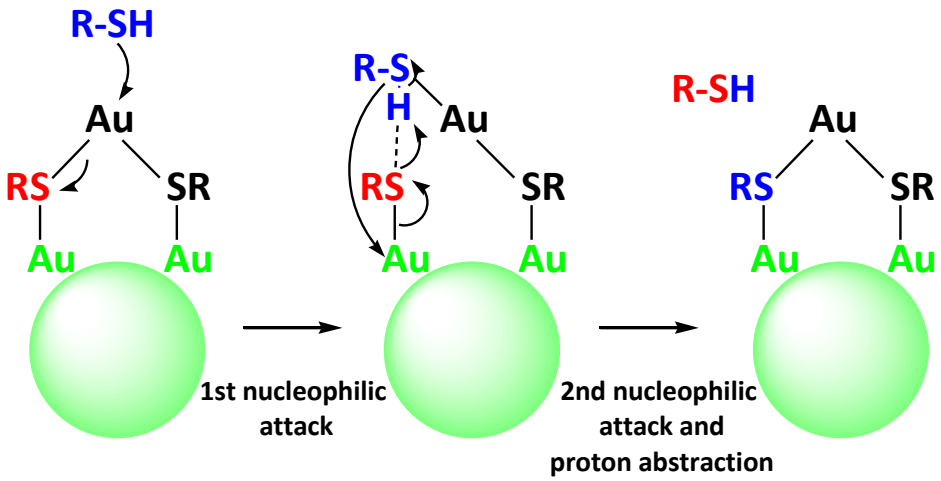

Figure V-1. Scheme of a ligand exchange reaction where the leaving alkanethiolate (red molecule) is part of a monomeric staple motif. The staple gold atom (electrophile) is shown black and the outer core gold atoms are green. The incoming thiol (nucleophile) is shown blue.

Previously, the mechanism of the ligand exchange reaction on AuMPCs with monomeric staple motifs as the protecting unit has been studied, both computationally and experimentally.[69b] However, the mechanism of the ligand exchange reaction towards an AuMPC with dimeric staple motifs has not been studied yet. The study of this mechanism would be interesting due to the different symmetry of the monomeric and the dimeric staple motifs. When the staple motif is monomeric, no matter which alkanethiolate is substituted: both are equivalent due to the symmetry of the staple (Figure V-2). In the case of dimeric staple motifs, the three alkanethiolate ligands are not equivalent. The two lateral alkanethiolates have the same bonding as the alkanethiolates in the monomeric staples (black): one bond to a staple gold atom (c) and one bond to a core gold atom (a'). The other, the apex alkanethiolate (blue), has a different bonding: both bonds are with staple gold atoms (Figure V-2).

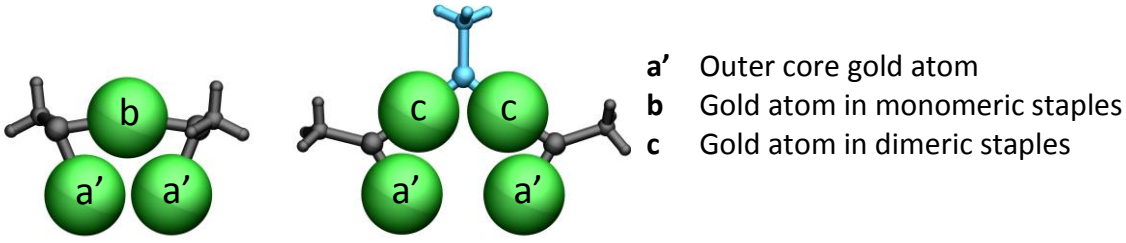

Figure V-2. Structures of monomeric (left) and dimeric (right) staple motifs. 2 types of alkanethiolates are present because of its symmetry: lateral alkanethiolates (black) and the apex alkanethiolate (blue). Gold atoms of the staples are labelled differently (*b* or *c*) to follow the labels presented in Chapter I, but both are oxidized gold atoms, i.e. Au(I).

A first question emerges when the ligand exchange reaction takes place in an AuMPC that has dimeric staple motifs: which alkanethiolate is substituted? While this work was being done, a X-ray study of a $Au_{25}(SC_2H_4Ph)_{16}(pBBT)_2$ has been presented.[77] This structure is the product obtained during a ligand exchange reaction of $Au_{25}(SC_2H_4Ph)_{18}$ with *p*-bromobenzenethiol (pBBT) as the incoming ligand. It was found that only the alkanethiolate placed in the lateral position of the dimeric staple motif is substituted during the reaction. One of the objectives of the present work is to elucidate why the apex alkanethiolate cannot be substituted. Furthermore, the AuMPC used here is the anionic $Au_{25}(SR)_{18}^-$, but in the X-ray study the neutral form was used, i.e. $Au_{25}(SR)_{18}^0$. Is there any difference in the regioselectivity of the ligand exchange reaction when the charge of the AuMPC is changed?

To give an answer to all these questions, we have chosen a process recently studied experimentally: the conjugation of a glutathione (GSH) AuMPC with a single chain Fv antibody fragment (scFv) from a tetrameric protein, the anti-influenza N9 neuraminidase NC10 antibody.[29b] The AuMPC used in experiments has a gold cluster core of 14 kDa, which correspond to an average composition of Au_{71}. Since Au_{71} does not correspond to a closed shell structure, they hypothesize that the gold core may be the closed shell Au_{75}.[29b] The goal of the experimental study was to optimize the reactivity of GSH AuMPC towards the NC10 scFv, rather than elucidating the position of substitution of the incoming thiol. Using different mutants of the NC10 scFv, the authors found that not all of them were able to bind and react with the AuMPC: the accessibility and neighbouring of the nucleophilic sulphydryl group play a crucial role in the reactivity towards the AuMPC. Here we want to give a computational insight to their findings by modeling the binding of two of the mutants of the NC10 scFv: one chemically inert and one displaying quantitative reaction towards $Au_{25}(GSH)_{18}^-$ a small gold cluster with dimeric staple motifs as protecting units. Knowledge of the mechanism of the ligand exchange reaction of the antibody towards $Au_{25}(GSH)_{18}^-$ will give an answer to the questions about the regioselectivity of the reaction described above. Furthermore, we should be able to explain why an AuMPC cannot bind to certain mutants of the scFv.

Computational Methods

Systems Analyzed

i) ***$Au_{25}(GSH)_{18}^-$:*** this system is the one used in the previous chapter. See its details in page 59 of chapter IV. 7 ps of QM/MM molecular dynamics simulations at 300 K are done to study which gold atom is the most exposed to the solvent.

ii) ***Mutants of the NC10 scFv***: the X-ray structure of the NC10 scFv has been taken from Protein Data Bank (PDB code 1NMB).[97] Only the V_H domain is considered along this study because is the responsible of the binding towards the AuMPCs.[29b] Two different mutants are considered here, which are the same as Ackerson *et. al.*

reported in reference 29b. The last residue in the X-ray structure is Ser113, and from it, two different tails are constructed using UCSF Chimera package:[98]

1. **114C, the inert mutant:** a cysteine is added through a peptide bond between free COO- of the Ser113 and the NH_2 group of the new residue, labelled from now as Cys114:

   NC10(V_H)-Cys114

2. **118C, with quantitative binding:** 6 residues are added here from the Ser113 to form a larger tail. The structure can be described as:

   NC10(V_H)-Ala114-Lys115-Lys116-Glu117-Cys118-Gly119

   Lysines are added to add positive charges to the C-terminus region. The three last amino acids of the tail are glutamate (Glu), cysteine (Cys) and glycine (Gly), which are the three amino acids that form GSH. They should replace the GSH ligands that are in the surface of the $Au_{25}(GSH)_{18}^-$ through a ligand exchange reaction.

For both mutants of NC10 scFv, classical MD simulations are performed to equilibrate its structures.

iii) Two different systems to simulate the ligand exchange reaction. To construct these two systems, the sulphydryl group of Cys118 or Cys114 is placed at ≈ 8 Å from the most exposed to the solvent gold atom of $Au_{25}(GSH)_{18}^-$:

1. **118C···$Au_{25}(GSH)_{18}^-$:** formed by the 118C mutant, $Au_{25}(GSH)_{18}^-$, 13555 water molecules together with 17 sodium cations to neutralize the negative charges. A total amount of 43253 atoms form the whole system. The size of the box is 70.5 Å x 81.0 Å x 93.5 Å.

2. **114C···$Au_{25}(GSH)_{18}^-$:** formed by the 114C mutant, $Au_{25}(GSH)_{18}^-$, 11099 water molecules together with 18 sodium cations to neutralize the negative charges. A total amount of 35810 atoms form the whole system. The size of the box is 62.0 Å x 76.7 Å x 95.9 Å.

To equilibrate the initial constructs, 15 ns of classical MD simulations are done keeping a restraint in the distance between the sulphur atom of Cys118 (for 118C) or Cys114 (for 114C) and the most exposed to the solvent gold atom. Then the restraint is removed and the simulation is extended up to 120 ns. The MM residues are not restrained or constrained during the equilibration. Graphs of the RMSD during the equilibration are shown in appendix III.

Ligand Exchange Reaction for 118C···$Au_{25}(GSH)_{18}^-$

The ligand exchange reaction is first studied in the system 118C···$Au_{25}(GSH)_{18}^-$. A equilibrated structure from the classical MD equilibration is taken for the subsequent QM/MM simulation. The QM region is formed by the 25 gold atoms and the 18 cysteines (only the -C-CH_2-S- part of cysteine) of each GSH. Furthermore, the nucleophilic cysteine (Cys118) is also included (-C-CH_2-SH). In total, there are 121 QM atoms enclosed in an orthorhombic box of 19.9 Å x 20.9 Å x 23.4 Å. The rest of the system, i.e. part of cysteines, glutamates, glycines, sodium cations and water molecules are considered in the MM part. The system is optimized by QM/MM simulations, using an annealing factor to cool down the ionic degrees of freedom up to achieve that the larger nuclear gradient is lower than $5 \cdot 10^{-4}$.

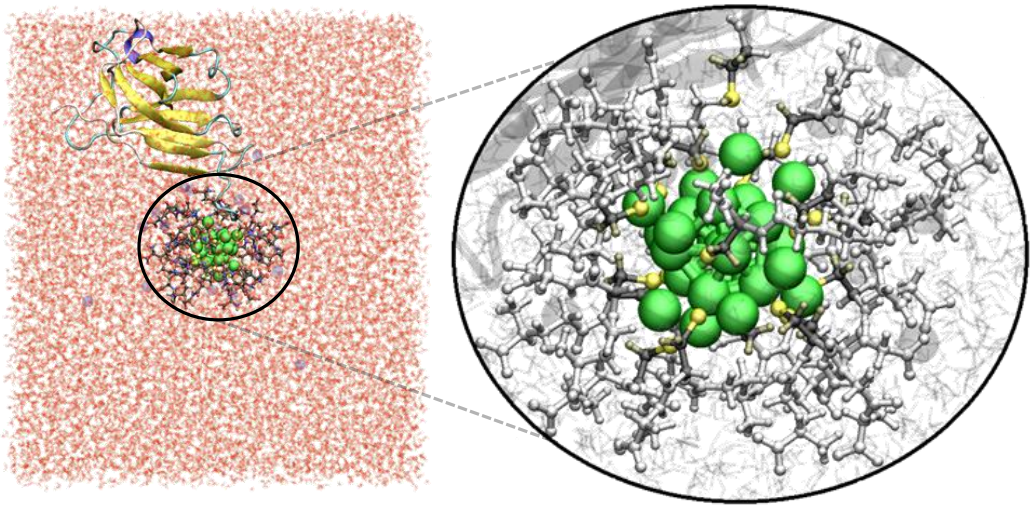

Figure V-3. Left image: the full system is shown. **Right image:** zoom over the QM region (right). The QM zone is shown in colour (Au in green, S in yellow, C in black and H in grey) whereas the MM zone is shown colourless.

For modelling the ligand exchange reaction, two different scans are done, depending on what alkanethiolate is replaced (Figure V-4). Both scans consist in a constrained structural optimization along a reaction coordinate, where the initial configuration of each point is taken from the optimized structure of the previous one. As a reaction coordinates, a difference of two distances is used:

$$d(S_{Cys118}\text{-}Au) - d(Au\text{-}S_{LG})$$

where S_{Cys118} is the nucleophilic sulphur, S_{LG} is the sulphur atom of the GSH leaving group and Au is the chosen gold atom to simulate each path: Au2 for the first path and Au22 for the second path. This distance difference is gradually varied from one point to the following and is then fixed to optimize the rest of degrees of freedom.

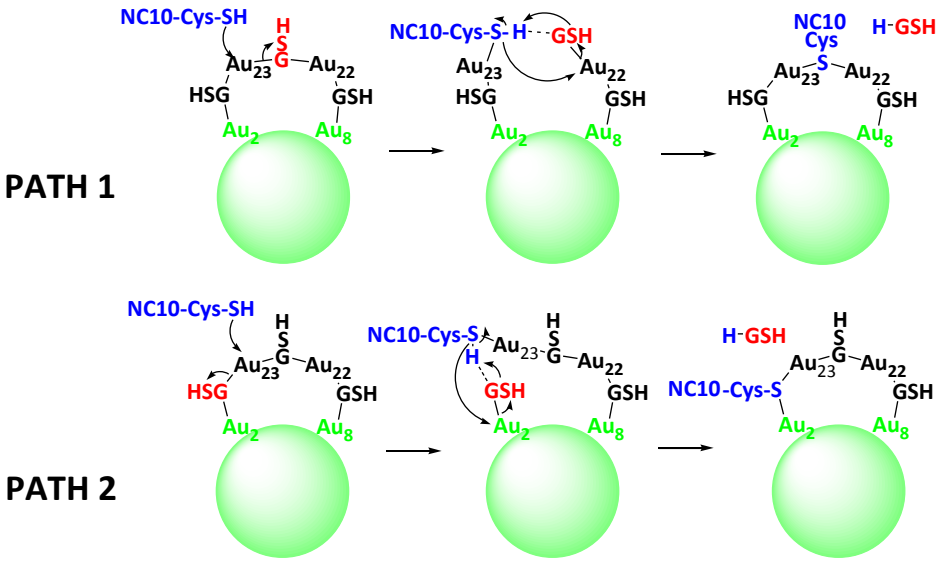

Figure V-4. Scheme of the two possible paths of the ligand exchange reaction of NC10 scFv antibody towards a dimeric staple motif of an AuMPC. The nucleophilic cysteine is shown blue, whereas the GSH leaving group is shown red. The gold cluster is schematized as a green ball. Only one staple motif is depicted for clarity. In the first path, the leaving group is the apex GSH, whereas in the second path it is one of the two equivalent lateral GSHs.

Ligand Exchange Reaction for 114C···$Au_{25}(GSH)_{18}^-$

To save computational time, only the best path of the two possibilities described previously is considered here. Furthermore, a different strategy is considered to obtain the potential energy profile of the ligand exchange reaction. We hypothesize that the mechanism of the reaction is not affected by protein mutations. Therefore, only the important points (reactants, two transition states, one intermediate and products) of the potential energy profile obtained for 118C···$Au_{25}(GSH)_{18}^-$ are used as templates to obtain the energetic profile when 114C is used. The procedure is as follows:

i) The equilibrated structure of 114C is superposed with each of the five important structures found along the potential energy profile of 118C···$Au_{25}(GSH)_{18}^-$. Our goal here is to replace 118C by 114C in each of the important points along the reaction coordinate, taking care of that the nucleophilic cysteine of 114C occupies the same position in 114C and 118C. As a result, we obtain non-equilibrated structures of 114C··· $Au_{25}(GSH)_{18}^-$ along the potential energy profile.

ii) The obtained structures of 114C···$Au_{25}(GSH)_{18}^-$ are then solvated in a water box with the corresponding number of counterions. To equilibrate the structures, classical MD simulations are performed, keeping the atoms involved in the ligand exchange reaction frozen.

iii) A frame of the equilibrated structure is used for the subsequent QM/MM optimization, where only the two distances involved in the reaction coordinate are constrained. To optimize the structures, an increasing annealing factor from 0.9 to 0.999 is added to cool down the ionic degrees of freedom.

Simulation Details

Classical molecular dynamics (MD) simulations are performed using NAMD[84] software. The NC10 scFv mutants and the GSH peptides are modelled with the FF99SB force-field.[85] Water molecules are described with the TIP3P force-field.[86] The coordinates of gold atoms are constrained due to the lack of parameters during all the classical simulations.

QM/MM simulations are done using the method developed by Laio *et. al.* [49] combining Car-Parrinello MD,[64] based on DFT theory,[44] with a force-field MD methodology as implemented in the CPMD code.[87] For the QM fragment, a fictitious mass of 600 a.u. and an integration time step of 0.12 fs guarantee a good control of the adiabaticity. For the molecular dynamics simulations, the temperature is set to 300 K by coupling it to a Nosé-Hoover chain thermostat on the ionic degrees of freedom.[88] The Kohn-Sham orbitals are expanded in a plane wave (PW) basis set with a kinetic energy cutoff of 90 Ry. The exchange-correlation functional used here is the Perdew-Burke-Ernzerhoff (PBE),[68] shown to be the best functional to describe AuMPCs.[16, 66-67, 72, 89] Gold is represented by a Goedecker[65] type pseudopotential, carbon and sulphur by a Troullier-Martins[45] type and hydrogen by a Car-von Barth type, as it was done in reference 66 and 67. Furthermore, single-point calculations using a hybrid PBE0[90] functional are done for the important points (reactants, two transition states, one intermediate and products) of the potential energy profile obtained for 118C···$Au_{25}(GSH)_{18}^-$.

The QM/MM interface is modelled by the use of a link-atom pseudopotential that saturates the QM region.[91] The electrostatic interactions in the interface of the QM and the MM regions are described in the Chapter II of this thesis. Neither restraints nor constraints are considered on the MM region, so the MM atoms can move freely.

Results

Ligand Exchange Reaction in 118C

For the ligand exchange reaction to take place, a nucleophilic cysteine should be close to an electrophilic gold atom. The gold surface of $Au_{25}(GSH)_{18}^-$ is not much exposed to the solvent due to the bulky GSH ligands, which has a big shielding effect (page 67 of Chapter IV). Along the 7 ps of QM/MM MD simulated for $Au_{25}(GSH)_{18}^-$, Au23 is the gold atom that shows the highest exposure to the solvent, with a solvent accessible area (SASA) of 3.5 $Å^2$ (calculated with a probe sphere of 1.4 Å, as corresponds to the water

solvent). Other surface gold atoms exhibit a lower exposure to the solvent, thus these were not considered (Figure V-5). The different conformation of the GSHs that are surrounding the gold cluster determines which gold atom is more or less exposed to the solvent.

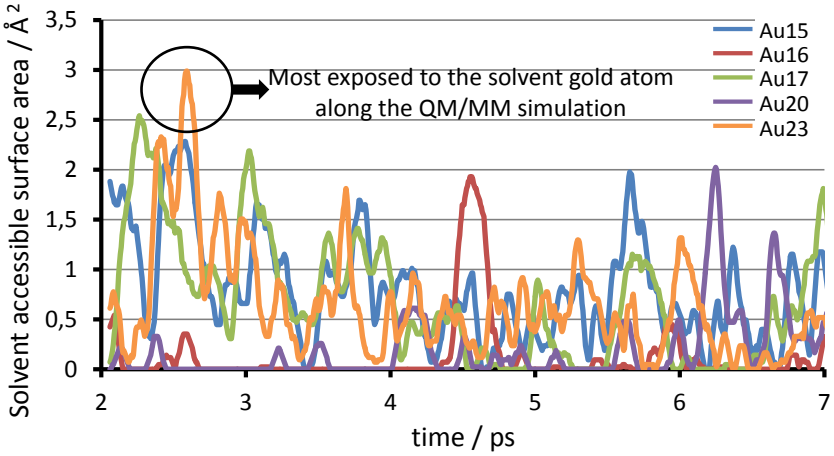

Figure V-5. Solvent accessible surface area (SASA) of gold atoms in $Au_{25}(GSH)_{18}^-$ along the 7 ps of QM/MM MD simulation. A running average over 10 data points has been used. The highest SASA is 3.5 Å, corresponding to Au23 (value shifted to 3 Å due to the running average).

The system was equilibrated by classical MD simulations, following the prescription described in the computational methods. Initially the sulphydryl group of Cys118 is placed in the vicinity of Au23 (at ≈ 8 Å). During the equilibration, the S_{Cys118}-Au23 distance oscillates around 6.5 Å. Interestingly, this distance decreases up to 4.46 Å upon QM/MM optimization. The optimized structure is shown in the left panel of Figure V-6, where the incoming thiol of Cys118 is hydrogen bonded to one carbonyl oxygen of the apex alkanethiolate. From this structure, two different scans were done depending on which ligand is being substituted: the lateral or the apex alkanethiolate. The adsorption reaction starts with the adsorption of the incoming thiol (S_{Cys118}) at Au23 (right panels in Figure V-6) and it is exothermic in both cases, -1.8 kcal/mol for the first path and -0.4 kcal/mol for the second one. This is in concordance with previous experimental observations of associative mechanisms.[96a, 99] Hereafter, the adsorbed states will be labeled as reactants, since they correspond to the most stable ones (Figure V-6, right panels).

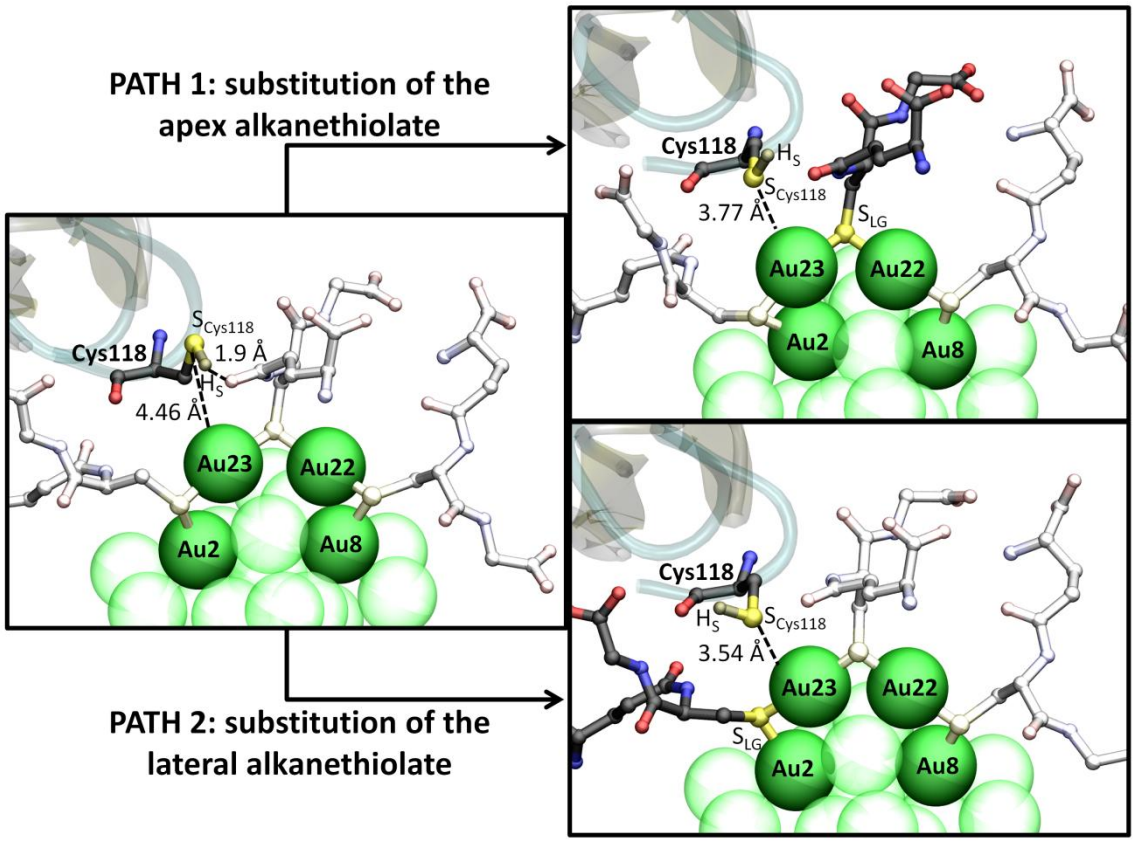

Figure V-6. Adsorption of the incoming Cys118 at Au23. For simplicity, only the three GSH ligands of the dimeric staple, where the reaction takes place, are shown. The two GSH that do not participate for the two modelled paths in the reaction are colourless. Gold atoms that are far from the dimeric staple are shown as semitransparent. Part of the secondary structure of the protein is displayed, together with the nucleophilic cysteine (Cys118). Water molecules are not shown for clarity. Colours code: Au (green), S (yellow), C (black), O (red), N (blue) and H (grey).

Path 1: Substitution of the Apex GSH Ligand

The adsorbed state shows a S_{Cys118}-Au23 distance of 3.77 Å. This value is similar to previously reported Au-SHR distances for a ligand exchange reaction in the gas-phase.[69b] Following adsorption, a two step reaction is observed (Figures V-7 and V-8). In the first step, the incoming thiol approaches Au23 from 3.77 Å to 2.27 Å while the leaving GSH breaks its bond with Au23, increasing the S_{LG}-Au23 distance from 2.3 Å to 3.42 Å. This leads to an intermediate state where the staple motif has been opened (the Au23-Au22 distance increases from 3.68 Å in reactants up to 5.08 Å in the intermediate) to accommodate the incoming thiol (Figure 7, intermediate of the top panel). In this intermediate, the thiol hydrogen of Cys118 is hydrogen bonded to the oxygen of a carbonyl group of the leaving GSH (1.84 Å). This hydrogen is still far from S_{LG} (4.03 Å). Besides, the Au23-Au2 distance increases up to 4.17 Å. The activation barrier for this reaction step is 13.4 kcal/mol and the energy of this intermediate is 7.8 kcal/mol higher than the reactants, i.e. the first reaction step is endothermic. To

complete the reaction, the incoming thiol approaches Au22 while desorption of the leaving GSH is completed. The abstraction of the thiol of Cys118 by GSH occurs spontaneously during this step. Finally Au23 recovers its original position in the dimeric staple motif. The activation barrier for the second reaction step is 36.6 kcal/mol and the energy of the products is 34.6 kcal/mol higher than that of the reactants, thus the reaction is very endothermic.

Path 2: Substitution of the Lateral GSH Ligand

The adsorbed state shows a S_{Cys118}-Au23 distance of 3.54 Å, *i.e.* 0.23 Å shorter than in the previous case. Again a two step reaction is observed (Figures V-7 and V-8). During the first step, the incoming thiol approaches the most solvent-exposed gold atom (Au23) up to 2.36 Å. Simultaneously breakage of the contiguous Au23-S_{LG} occurs (Au23-S_{LG} = 3.78 Å in the intermediate structure, see bottom panel of Figure V-8). At this point, the hydrogen of R-SH (Cys118) is hydrogen bonded to S_{LG} (H- S_{LG} = 2.28 Å). The position of Au23 does not change during this process, contrary to what happens for the first path (*i.e.* the staple does not open). The activation barrier of the first reaction step amounts to 13.9 kcal/mol, being very similar to the one obtained for path 1. The energy difference between the intermediate and the reactants is now -6.1 kcal/mol, thus being an exothermic step. From the intermediate, the thiol group performs a second nucleophilic attack, now to one of the Au atoms of the Au_{13} core (Au2). During this second attack, the leaving GSH breaks its bond with Au2, and is the ligand that deprotonates the nucleophilic sulphydryl. This second step shows an activation barrier of 16.5 kcal/mol. The energy of products is 13.6 kcal/mol more stable than reactants, i.e. the global reaction is exothermic.

The potential energy profile of both paths is shown in Figure V-7. The substitution of the lateral alkanethiolate (second path) is favored over the apex one. The reaction is stepwise with the rate-limiting step being the second one. This is similar to the substitution reaction obtained for a ligand exchange reaction in a monomeric staple motif. Therefore, we can conclude that the mechanism of the ligand exchange reaction does not change if it happens in a monomeric or in a dimeric staple motif. Interestingly, a recent X-ray study of Ackerson *et. al.*[77] revealed that only the lateral alkanethiolate bonded to the most solvent-exposed gold atom is substituted when $Au_{25}(SC_2H_4Ph)_{18}$ was exposed to a 5-fold molar excess of pBBT for 7 min. The fastest ligand exchange reaction should correspond to the one with the lowest energetic barrier, *i.e.* the substitution of the lateral alkanethiolate in a dimeric staple motif. Our results are in excellent agreement and thus give support to this experimental investigation. Furthermore, our results also show that the charge state of the AuMPC does not affect to the energetic or the regioselectivity of the ligand exchange reaction.

Structural and Electronic Analysis of Paths 1 and 2

The energetic difference between the paths can be explained in terms of structural and electronic properties. In particular, it can be traced back to the different behavior of two key atoms: the hydrogen of the sulphydryl group of Cys118 (H_S) and Au23, together with the different charge state of the atoms involved in the reaction (Table V-I). In the second path, H_S is near Au2 before its transfer to the alkanethiolate. This is reflected on the Au23 Hirshfeld charge, -0,176*e*, *i.e.* the gold atom is not fully oxidized and still has some basic character. In path 1, H_S is always in the vicinity of Au23 and Au22, which are gold atoms being oxidized. The Hirshfeld charge of Au23 (or Au22) is 0.4*e*. Therefore, there are no H_S···Au interactions that can assist the transfer of H_S from the thiol to the alkanethiolate. These interactions become crucial for the ligand-exchange reaction. It was already reported in chapter III that covalent interactions between hydrogen and gold in AuMPCs and self-assembled monolayers (SAMs) are important.[16, 60, 100] A second factor that increases the activation energy of the second step of path 1 is that Au23 should recover its position in the dimeric staple and this atomic reorganization has a larger energetic cost compared to path 2. In the second path there are no gold rearrangements during the reaction, leading to a lower energetic process.

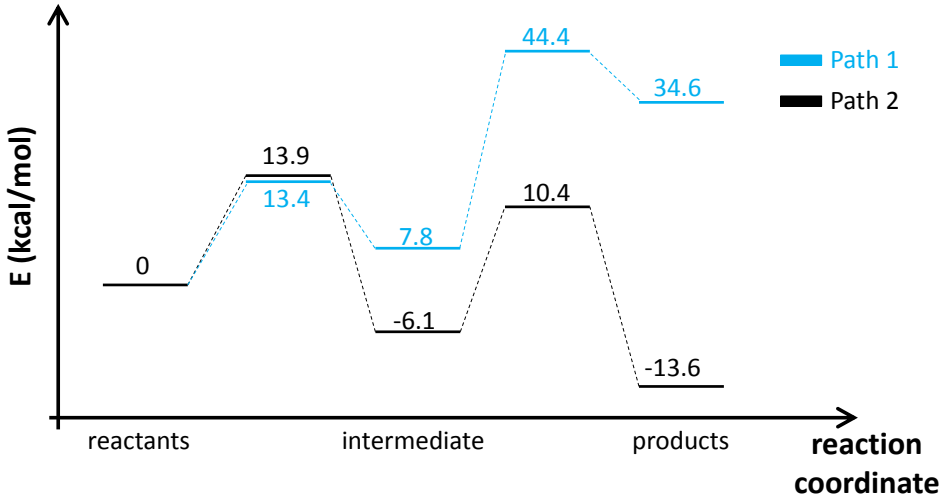

Figure V-7. Potential energy profiles of the ligand exchange reaction of 118C towards $Au_{25}(GSH)_{18}^-$ for the substitution of the apex alkanethiolate (blue) or the lateral alkanethiolate (black).

The evolution of the most relevant distances for the best path (path 2) is shown in appendix III.

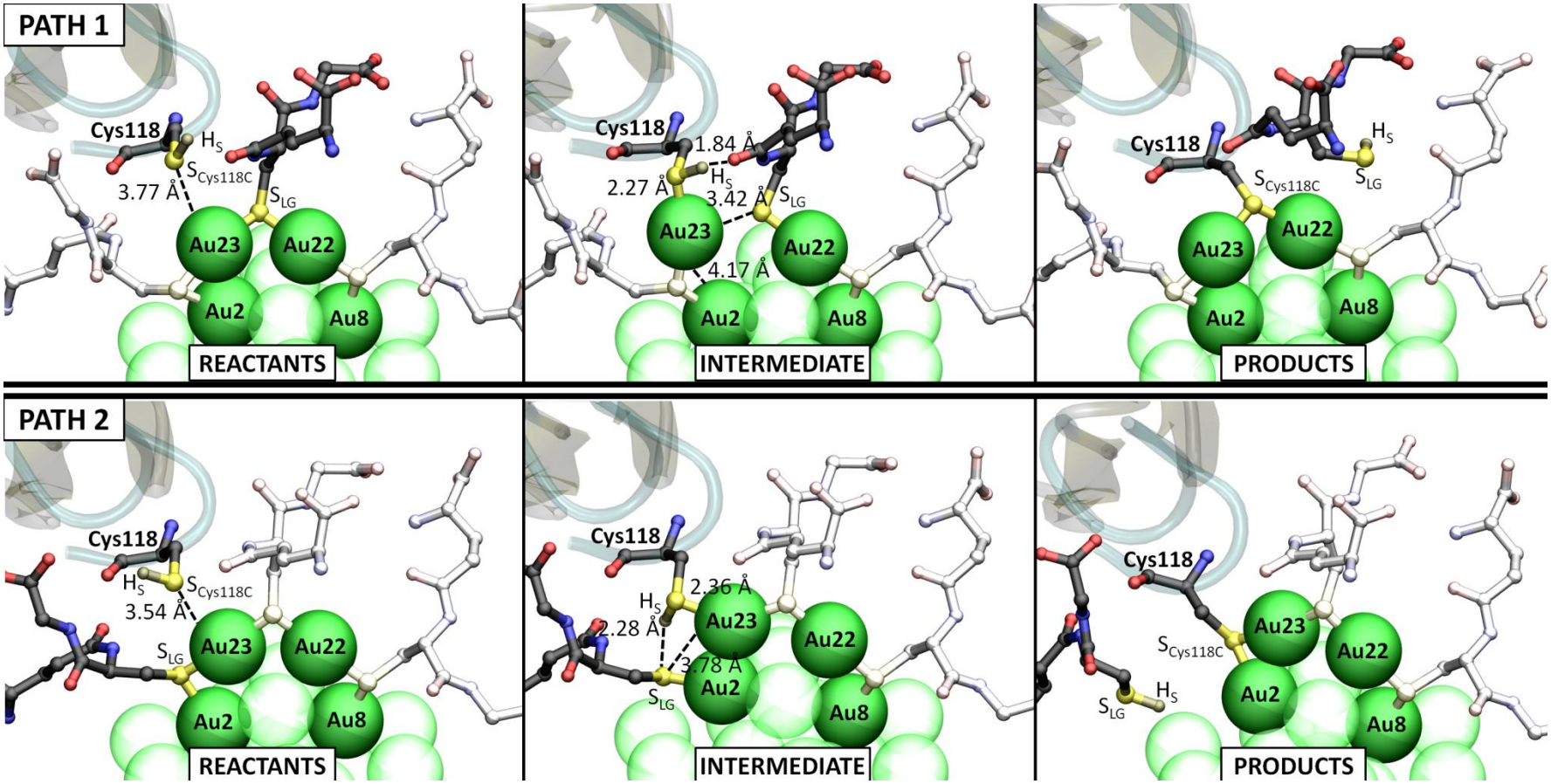

Figure V-8. Structures found along the potential energy profile of the ligand exchange reaction for the two paths studied. Only the three GSH ligands of the dimeric staple motif where the reaction takes place are shown. The two GSH that do not participate are colourless. Gold atoms that are far from this staple motif are shown as semitransparent. Part of the secondary structure of the protein, together with the nucleophilic cysteine (Cys118) is shown in the left top corner of each panel. Water molecules are not shown for clarity. Colours code: Au (green), S (yellow), C (black), O (red), N (blue) and H (grey).

Ligand Exchange Reaction in 114C

The local environment of the nucleophilic cysteine in 114C is very different than in 118C:

114C: NC10(V_H)-Leu109-Thr110-Val111-Ser112-Ser113-**Cys114**
118C: NC10(V_H)-Ala114-Lys115-Lys116-Glu117-**Cys118**-Gly119

114C displays both hydrophobic (leucine and valine) and polar uncharged (threonine and cysteine) amino acids. In contrast, 118C exhibits hydrophobic (alanine and gycine), positively charged (lysine) and negatively charged (glutamate) amino acids. The presence of positively charged groups in the vicinity of the reactive cysteine was suggested to be the main reason of the enhance reactivity of 118C towards the negatively charged AuMPC.[29b] To evaluate the consequences of changing the local environment of the nucleophilic cysteine, the sulphydryl group of Cys114 was placed at 8 Å from Au23 (as it was done for 118C). The system was equilibrated by classical MD simulations according with the prescription described in the computational methods section. The S_{Cys114}-Au23 distance was monitored along the 120 ns of classical MD simulations, and compared with the one obtained for the previous system, S_{Cys118}-Au23 (Figure V-9).

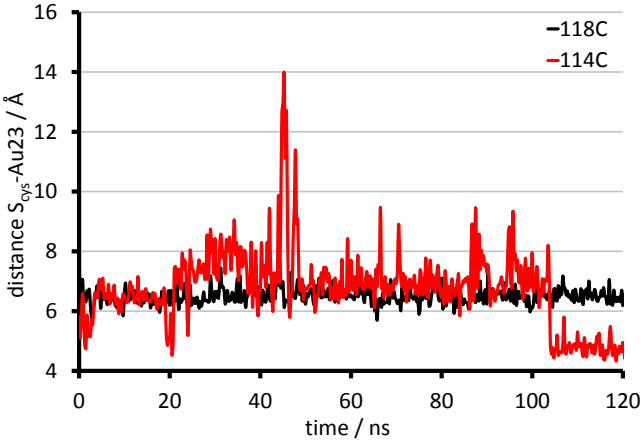

Figure V-9. S_{Cys}-Au23 distance in the two NC10 scFv mutants studied, 114C (red) and 118C (black), along the classical MD simulation.

It can be observed that fluctuations of S_{Cys}-Au23 are significantly larger for 114C are larger than for 118C. The average value of S_{Cys118}-Au23 for the 118C mutant is 6.5 Å, with just small deviations from this value during the course of the simulation. Since we know that 118C is able to react with AuMPCs, we can consider that this the optimum distance from where cysteine is able to interact with AuMPCs. For 114C, the average value of S_{Cys114}-Au23 is 7.1 Å during the first 100 ns, slightly larger than the optimum 118C value, and its oscillations are much larger. However, a conformational change takes place at ≈ 105 ns and the cysteine is able to approach Au23 up to a 4.8 Å, just 0.3 Å longer than the one obtained for the QM/MM optimized structure of *118C···$Au_{25}(GSH)_{18}^-$*. This suggests that 114C is also able to interact with the AuMPC

through its cysteine (Cys114). Therefore, we decided to model the ligand exchange reaction for 114C to see the difference with respect to 118C. To save computer time, we used a method based on single-point optimizations using as templates the five important points (minimums and maximums) found along the path 2 of 118C (see methods). When the procedure was tested using 118C as the superposed mutant, very similar results were obtained with respect to the ones obtained previously (Figure V-10), validating the method used here.

When 114C is used for the ligand exchange reaction, it is found that the first activation barrier is 31.7 kcal/mol (Figure V-11). This value is 17.8 kcal/mol larger than the one obtained for 118C, indicating that 114C cannot even be bonded to the most solvent-exposed gold atom. If this first step was feasible, the second step would have also a larger activation barrier, 24.2 kcal/mol.

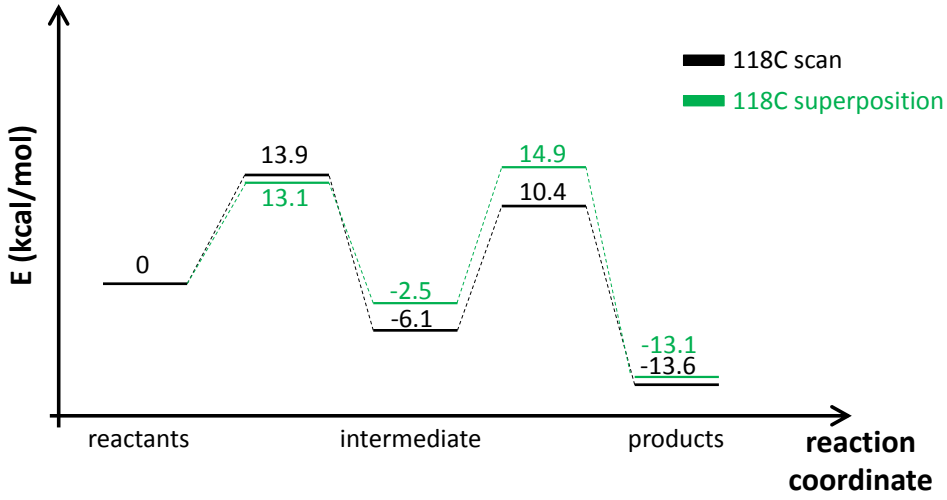

Figure V-10. Potential energy profiles of the ligand exchange reaction of 118C towards Au25(GSH)18- using two different procedures: a scan on the chosen reaction coordinate (black) or single-point optimizations of the important structures obtained by superposition (green).

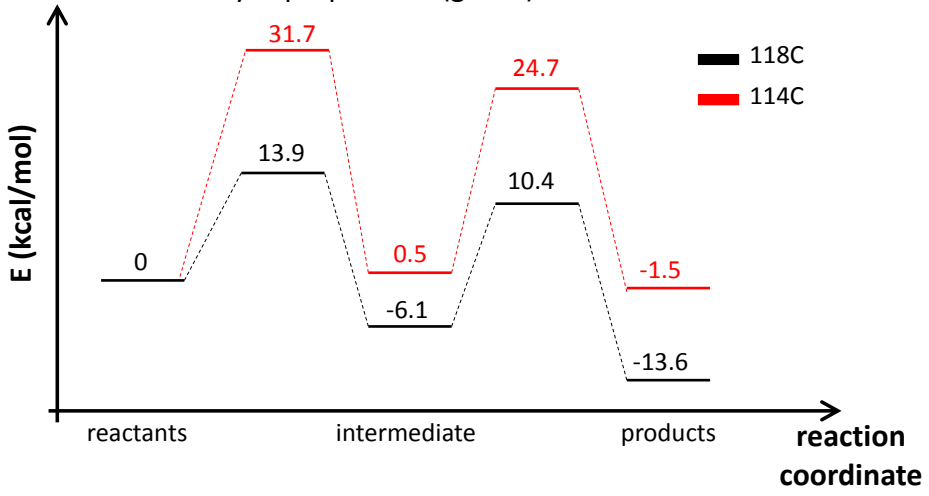

Figure V-11. Potential energy profiles of the ligand exchange reaction of 114C and 118C towards $Au_{25}(GSH)_{18}^-$.

Although the Cys114 of 114C can interact with the exposed gold atom, it is clear that the binding between $Au_{25}(GSH)_{18}^-$ and 114C is weaker than in the case of 118C (Figure V-9). The larger fluctuations in the S_{Cys}-Au23 distance for the 114C mutant show that both entities have fewer interactions than in 118C. In the latter, the S_{Cys}-Au23 distance is quite stable due to the strong interactions of the neighboring positively charged groups towards the negatively charged AuMPC. This is further confirmed by analysis of the structures obtained along the ligand exchange reaction for both NC10 scFv mutants. For all 118C structures, Lys115 and Lys116 are interacting with the negatively charged carboxylate group of glycine or glutamate of GSH. The fact that these interactions are absent in 114C leads to a much loose interaction between the antibody and the AuMPC (Figure V-11). While 118C exhibits three lysine-carboxylate interactions, 114C only has one (Lys13, which is far from the nucleophilic cysteine). This interaction is after present in both mutants. Apart from Lys13 the only amino acid that helps in the interaction with the AuMPC in 114C is Ser112, but a hydroxyl-carboxylate interaction is much weaker than a lysine-carboxylate one. Therefore, we conclude that the higher energy barriers observed for the ligand-exchange reaction in 114C is due to the weaker Scys-Au23 adsorption (S-Au distance increases with respect to the one in 118C) and because of the loose of peptide-AuMPC interactions.

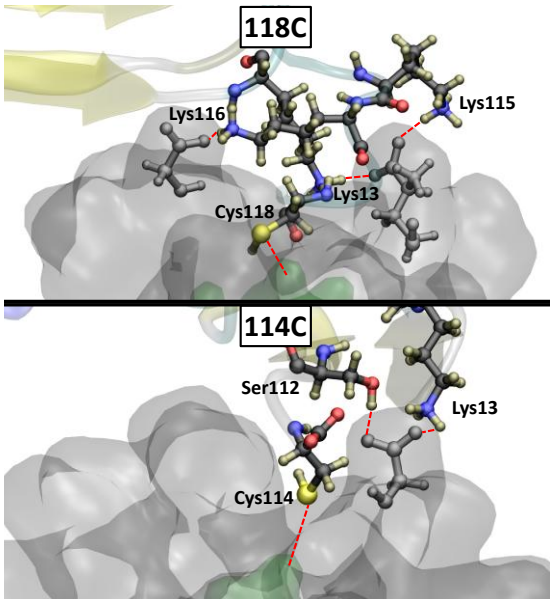

Figure V-12. Interactions in the interface of the antibody (shown by its secondary structure) and $Au_{25}(GSH)_{18}^-$ (shown by its Van der Waals surface). Glycine and glutamate amino acids of the AuMPC is shown grey. Gold and the amino acids of the antibody are coloured following: Au (green), S (yellow), C (black), O (red), N (blue) and H (grey).

Conclusions

The mechanism of the ligand exchange reaction of the anti-influenza N9 neuraminidase NC10 antibody fragment towards $Au_{25}(GSH)_{18}^-$ is solved by means of QM/MM simulations. It is found that the ligand exchange reaction follows an associative double S_N2-like mechanism with low activation barriers (13.9 and 16.5 kcal/mol) when the lateral ligand is substituted. We found that the fact of having monomeric or dimeric staple motifs does not affect the mechanism of the ligand exchange reaction. When the substituted ligand is the one that occupies the apex position of the dimeric staple motif, the activation barrier of the second step of the reaction is 20.1 kcal/mol larger than for the substitution of the lateral ligand. Therefore, the preferred ligand to be substituted is the lateral one.

Our results are in excellent agreement with a recent X-ray study showing that only one type of ligands is substituted. We also found that the charge of the AuMPC does not change the energetic nor the regioselectivity of the ligand exchange reaction. When the vicinity of the incoming thiol is neutral (114C) and not positively charged as in 118C, the activation barriers of both reaction steps increase (31.7 and 24.2 kcal/mol) due to the lack of strong interactions between the antibody and the AuMPC. This result opens the door to future computational and experimental investigations design mutants to increase the reactivity of biomolecules towards a desired AuMPC.

Chapter VI – Formation of a Covalent Glycosyl-Enzyme Species in a Retaining Glycosyltransferase

Formation of a Covalent Glycosyl-Enzyme Species in a Retaining Glycosyltransferase

Introduction

As shown through this thesis, AuMPCs are gaining considerable attention in biomedicine. In Chapter V, we showed that they can be used to bind an antibody, and the corresponding antibody-AuMPC conjugate has been developed for binding to the viral neuraminidase, a glycoside hydrolase enzyme that is present in the surface of the influenza virus (see page 14 of Chapter I for more details). ***Glycoside hydrolases*** (GHs) are group of a type of proteins named as ***glycoproteins,*** which are proteins that can be conjugated with ***carbohydrates*** (page 10 of Chapter I for a detailed introduction of carbohydrates). Glycoproteins are important components of cell surfaces and the extracellular environment and they participate in cell-cell recognition and adhesion processes. The function of GHs is to degrade large carbohydrates into shorter ones. Another important group of glycoproteins are the ones responsible for the synthesis of larger carbohydrates departing from short ones: ***glycosyltransferases*** (GTs). In this chapter we focus in the mechanistic insights of GTs. They are responsible for protein glycosylation, the most recurrent post-translational modification occurring in nature for cell recognition and signaling. GT alterations cause several diseases, such as infection, inflammation, and either normal or abnormal cellular developments.[101]

GTs catalyze the formation of glycosidic linkages by the transfer of a saccharide, typically a monosaccharide, from a sugar nucleotide donor to an acceptor substrate[102] (Figure VI-1). One of the major challenges for the rational design of specific and potent drugs/inhibitors for this class of enzymes[102-103] is unraveling their detailed reaction mechanisms, which still escape direct experimental probes.

The transfer of the glycosyl group can take place either with inversion or retention of the anomeric stereochemistry with respect to the donor sugar.[34] The mechanism of inverting GTs is clearly established as a single displacement S_N2 reaction with a general base catalyst that increases the nucleophility of the attacking group.[34] This is analogous to the well-characterized mechanistic strategies used by inverting glycoside hydrolases (GHs) to catalyze the cleavage of glycosidic bonds.[104] However, the mechanism of retaining GTs remains a controversial issue in the field of glycobiology.[34] By comparison to retaining GHs, retaining GTs were first suggested to follow a double-displacement reaction involving a covalently bound glycosyl–enzyme intermediate (Figure VI-1, top). This requires the existence of an appropriately positioned nucleophile within the active site residues (aspartate or glutamate).

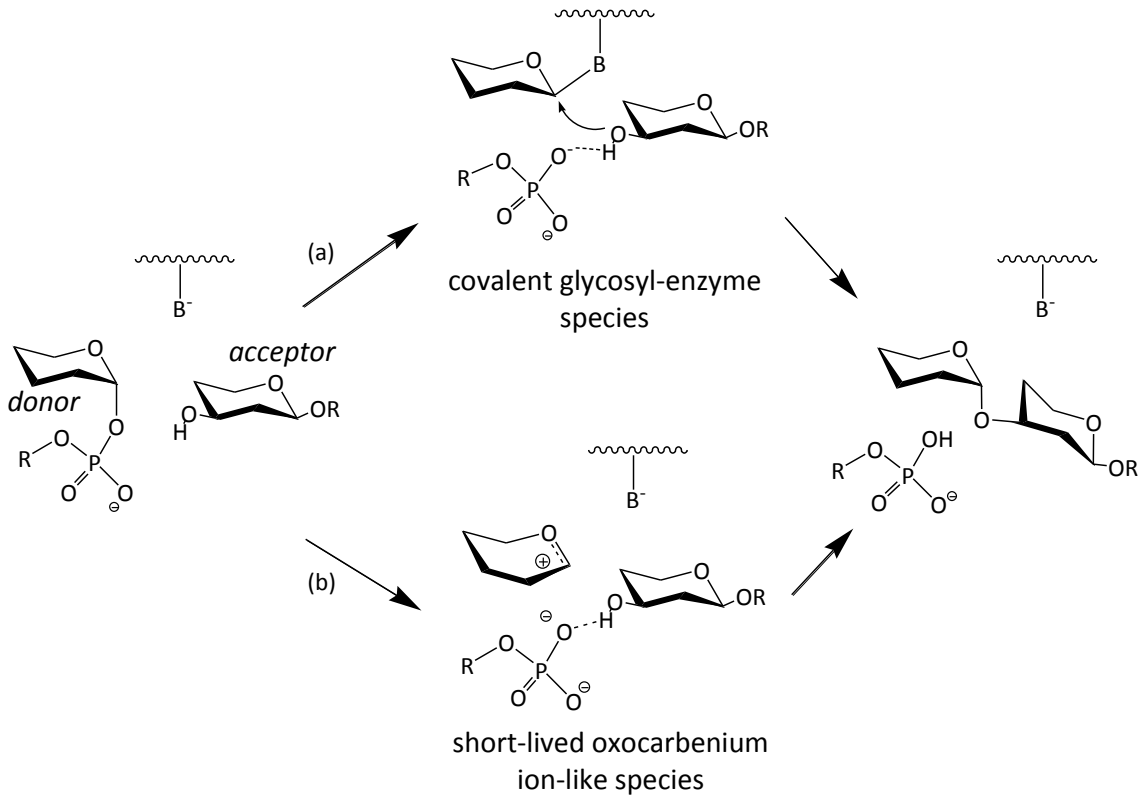

Figure VI-1. Possible reaction mechanisms for enzymatic glycosyl transfer with retention of configuration proposed in the literature. a) Double-displacement mechanism. b) Front-face mechanism. The donor and acceptor molecules are those of α3GalT. Hydroxymethyl and hydroxyl substituents of Gal and Glc are omitted for clarity.

However, such a requirement is so far only fulfilled by two GT enzymes from family 6:[37] mammalian α-1,3-galactosyltransferase (α3GalT) and blood-group A and B α-1,3-glycosyltransferase (GTA/GTB). Both enzymes share a common type of fold (A). Mutation of the putative nucleophile (Glu317 in α3GalT and Glu303 in GTA/GTB) to alanine inactivates the enzyme (k_{cat} decreases at least 10^4-fold[105]), and this allows to infer that these residues play an essential role in catalysis. Formation of a covalent adduct (Figure VI-1, top) has been highlighted by mass spectrometry and chemical rescue experiments on mutants of GTA/GTB[106] and α3GalT,[105a] respectively. However, no clear evidence has been reported to date for the formation of such adduct for the wild type enzyme, raising the question whether or not the mutation could affect the molecular mechanism. On the other hand, some authors have proposed that Glu317 is more likely required for proper acceptor substrate orientation.[38] Hence, it remains to be proved that a double-displacement mechanism (Figure VI-1, top) is operative for these GTs.

A different scenario is provided by those retaining GTs that lack a putative nucleophile in the active site, such as trehalose-6-phosphate synthase, mannosylglycerate synthase, glycogen synthase or glycogenin.[34] It has been proposed that these GTs operate via a front side single displacement, in which the nucleophilic

hydroxyl group of the acceptor attacks the anomeric carbon from the same side from which the leaving group departs (Figure VI-1, bottom), thus leading to retention of stereochemistry.[107] This hypothesis, which generated a vigorous debate for several years (see discussion in reference 34) has been recently verified by several independent studies of retaining glycosyl transfer on enzymatic[71f, 108] and non-enzymatic[109] environments. Of particular interest is the recent kinetic isotope effect (KIE) study that demonstrates that a front-face mechanism is operative for trehalose-6-phospate synthase.[108b] Consistently, QM/MM metadynamics simulations[110] provided evidence for the formation of a short-lived oxocarbenium ion-like species (Figure 1, bottom).

The recent insight provided by both experiment and theory raises the question whether the front-face type of mechanism could be common for retaining GTs. In other words, do α3galT and blood GTs, which unlike most GTs exhibit a well-positioned active site residue, also operate via a front-face mechanism? Indeed, early theoretical studies[111] on small gas-phase models predicted the double-displacement to be feasible for a "nucleophile-containing" active site. However, two recent static QM/MM studies by the same group[112] conclude that both mechanisms are feasible, differing only in which mechanism is favored over the other. It should be noted that these pioneering studies did not take into account the dynamics of the active site and protein atoms during the chemical reaction. It was also argued that a 2D free-energy surface would be required to draw definitive conclusions regarding the fine details of the front-face mechanism.

Unlike GHs, GTs are very flexible. They undergo significant conformational changes upon substrate binding, as evidenced by structural studies on binary complexes.[113] It has been found that flexible loop rearrangements occur upon donor binding and that the nucleotide participates in acceptor binding, with the enzyme changing from an open to a closed conformation.[114] For this reason, a fair description of dynamic effects during the chemical reaction seems to be crucial to elucidate the molecular mechanism of glycosyl transfer. Here we use classical MD and hybrid QM/MM metadynamics simulations to investigate the glycosyl transfer reaction in α3GalT. Using this approach, we demonstrate that a GT with a putative nucleophile residue, such as α3GalT, can only operate through a double-displacement mechanism, with the formation of a glycosyl-enzyme covalent intermediate.

Computational methods

Preparation of the System: Building the Ternary Complex

A major obstacle to the model study of α3GalT (and GTs in general) is the limited structural data for the ternary complex between the enzyme (α3GalT), the donor (UDP-Gal), and the acceptor (lactose, Lac). However, binary complexes of the enzyme with either the nucleotide sugar donor (UDP-2F-Gal, PDB code 2VFZ)[115] or both the nucleotide and the acceptor (UDP + Lac, PDB code 1GWV)[114] have been reported (Figure VI-2). While both structures are very similar in the donor region, only the last one displays an ordered structure of the highly flexible C-terminal α-helix that closes the acceptor binding site. Therefore, we took the enzyme and the donor molecule of the former complex (the 2-F modification of UDP-2-F-Gal was reverted to 2-OH) and inserted in the acceptor molecule and the C-terminal α-helix from the latter. In the resulting complex (Figure VI-2a), the anomeric carbon of the donor sugar (C1) is located close to the oxygen atom of the acceptor molecule (C1···O3' = 2.56 Å) with whom the glycosidic bond is expected to form. In addition, the side chain of Glu317 is also close to the anomeric carbon of the sugar donor (COO^-···C1 = 4.09 Å) and well oriented for nucleophilic attack (a similar structure results from the inverse superposition, i.e. taking as a basis the complex with the acceptor, see Figure VI-2b).

The protonation states and hydrogen atom positions of all other ionizable amino acid residues were selected based on their hydrogen bond network. Four histidine residues were modeled in their neutral states. All the crystallographic water molecules were retained and extra water molecules were added to form a box of 73.8 x 81.9 x 77.3 Å. Six chloride ions were added to neutralize the enzyme charge. The system was prepared using the XLeap module of AMBER.[116]

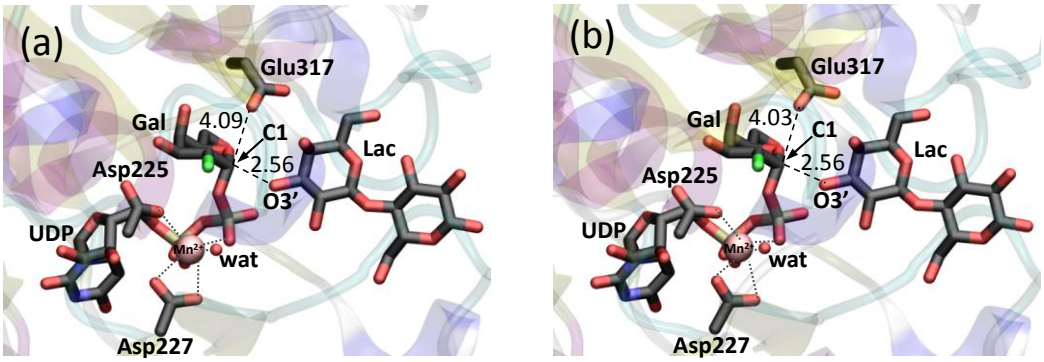

Figure VI-2. Active site structures of the ternary complexes obtained upon superposition of: (a) Enzyme complex with UDP-2F-Gal (PDB code 2VFZ) and lactose acceptor and C-terminal α-helix from the enzyme complex with UDP and Lac (PDB code 1GWV). (b) Enzyme complex with UDP and Lac (PDB code 1GWV), taking the donor from the enzyme complex with UDP-2F-Gal (PDB code 2VFZ). Enzyme subunits B have been used for the superposition. Similar structures are obtained when considering subunits A (C1···OGlu317 = 4.37/4.04 and C1···O3' = 2.45 Å).

Simulation Details

Details of the Classical Molecular Dynamics Simulations

A classical molecular dynamics (MD) using NAMD[84] software was initially performed. The protein was modeled with the FF99SB force field.[85] All carbohydrates were modeled with the GLYCAM06 force field[117] and water molecules were described with the TIP3P force field.[86] The MD simulation was carried out in several steps. First, the system was minimized, holding the protein and substrate fixed. Then, the entire system was allowed to relax. To gradually reach the desired temperature of 300 K in the MD simulation, the system was coupled to a thermostatic bath for 130 ps of MD simulation at constant volume, followed by 50 ps at constant pressure. The MD was extended to 25 ns using a restraint on the C1···O3' distance. Once the system had reached equilibrium, the simulation was extended to 32 ns without any restraint. The overall structure did not change significantly from the initial structure of the ternary complex (the root mean square displacement (RMSD) of the backbone atoms was ≈ 1.9 Å).

The main scope of the classical MD simulation is to relax the initial enzyme·substrate complex, in order to remove residual spurious forces and small tensions coming from the superposition procedure adopted to build up the complex. In this respect, the length of MD run is long enough to obtain a stable initial structure. This is ensured by the RMSD of the backbone atoms, which is found to be always < 2 Å, hence compatible with the experimental resolution of the structure; such a check of the structure is a crucial step to safely switch to subsequent QM/MM simulations. Moreover, we remark that no significant conformational changes (those that could possibly occur on the millisecond time-scale) were expected, since the structure of the C-terminal α-helix, i.e. the only flexible part of the protein, was taken already folded from the crystal structure of the enzyme in complex with the acceptor molecule (accession code of the Protein Data Bank: 1GWV). Analysis of the trajectories was carried out using standard tools of AMBER[116] and VMD.[92] A snapshot of the MD-equilibrated system was taken for the subsequent QM/MM and metadynamics simulations.

Details of the QM/MM Molecular Dynamics Simulations

QM/MM calculations were performed using the QM/MM CPMD code,[49] which combines Car–Parrinello MD,[64] based on DFT, with the AMBER force-field (see details in page 26 of Chapter II). The QM region included the acceptor lactose, the sugar and phosphate groups of the UDP-Gal donor, the Mn^{2+} cation and a close water molecule, as well as the side chain of Glu317, Asp63 and Asp65 (84 QM atoms, 37600 MM atoms). The side chains of Glu317, Asp63, Asp65, were capped at its Cβ (glutamate) or Cα (aspartates) atoms with a link-atom pseudopotential. The QM region (84 atoms) was enclosed in an isolated 20.3 x 22.3 x 13.1 Å supercell. Kohn–Sham orbitals were

expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry. Norm-conserving Troullier–Martins ab initio pseudopotentials[45] were used for all elements. The calculations were performed using the Perdew, Burke and Ernzerhoff generalized gradient-corrected approximation (PBE).[68] This functional form has been proven to give a good performance in the description of hydrogen bonds[118] and was already used with success in previous works on glycoside hydrolases and other glycosyltransferases in our group.[71d, 110, 119] The simulations were performed in a spin unrestricted scheme to allow the unpaired electron to be localized on the metal cation and to allow bonds to break and reform. The fictitious mass for the electronic degrees of freedom of the CP Lagrangian was set at 400 a.u., and the simulation time step at 0.096 fs. The system was equilibrated at 300 K for 3 ps before starting the metadynamics run.

Metadynamics Simulation of the Glycosyl Transfer Reaction

QM/MM metadynamics simulations (see a detailed description of the method at page 33 in Chapter II), aimed at breaking the sugar-phosphate glycosidic bond and forming the glycosidic bond between the donor and the acceptor, were performed on the extended Lagrangian formalism.[53] Two collective variables (CVs) were used. The first one (CV_1) was taken as the coordination number (CN) between the anomeric carbon of the Gal donor (C1) and the phosphate oxygen atom (O_P) ($CV_1 = CN_{C1,OP}$, Figure VI-3). The CN is defined using the following equation:

$$CN_{i,j} = \frac{1 - \left(\frac{d_{i,j}}{d^0}\right)^p}{1 - \left(\frac{d_{i,j}}{d^0}\right)^{p+q}} \quad \textbf{(Eq.VI-1)}$$

where d_{ij} is the internuclear distance of the atoms involved, d^0 is the threshold distance for bonding, and p and q are exponents that determine the steepness of CN_{ij} decay with respect to d_{ij}. CN values range from 0 (atoms not bonded) to 1 (atoms bonded). The parameters used are p = 15, q = 3 and d^0 = 3.21 a.u. (1.7 Å). This variable measures the degree of formation of the sugar-phosphate bond.

The second collective variable (CV_2) was taken as the distance between the anomeric carbon of the Gal donor (C1) and the oxygen atom of the Gal acceptor (O3') with whom the glycosidic bond is to be formed ($CV_2 = d_{C1\text{-}O3'}$, Figure VI-3). The use of a distance as collective variable will allow us to capture the existence of more than one minimum as the acceptor enters the active site. CV_2 measures the formation of the glycosidic bond. It is important to stress that the selected CVs do not self-select any specific reaction pathway (either a double displacement reaction (Figure VI-1a) or a front-face reaction (Figure VI-1b) could be obtained). The simulation does not implicitly assume any of these situations, i.e. CVs do not force (nor restrict) Glu317 to bind to the sugar donor.

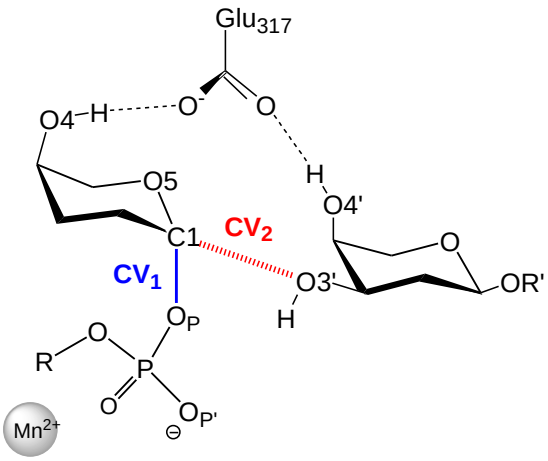

Figure VI-3. Collective variables (CVs) used in the QM/MM metadynamics simulation. CV_1 is defined as a coordination number and CV_2 as a distance.

The selected mass values of the fictitious particles (μ) were 200 a.u. for both collective variables, while those of the force constant (k) were 15 a.u. The height of the Gaussian used was of 1.255 kcal·mol^{-1}. The width of the Gaussian terms (0.10 coordination number units) was selected from the oscillations of the CVs in a free Car-Parrinello QM/MM simulation. A new Gaussian-like potential was added every 150 MD steps and the simulation was stopped after 15 ps of simulation once the intermediate and the products wells were sampled. A total of 1033 Gaussian hills had been deposited. The values of the two collective variables at the reactants (CV1 ≈ 1, CV2 ≈ 3) and product states (CV1 ≈ 0, CV2 ≈ 1.6) are different enough to ensure that the two states will appear in different regions of the free energy surface, a necessary condition for a suitable characterization of the reaction path in a metadynamics simulation.[54, 120]

Results

The constructed ternary complex was submitted to a molecular dynamics simulation (32 ns of classical MD, followed by 3 ps of QM/MM MD; see the Computational Details above) to accommodate all the actors (donor, Glu317, and acceptor) optimally in the binding site and to remove any bias due to the initial superposition. The final structure of the ternary complex (Figure VI-4) shows that the side chain of Glu317 (the putative nucleophile) bridges both the donor and the acceptor Gal moieties (specifically, the carboxylate side chain interacts with both the O4H group of the Gal donor and the O4'H of the Gal acceptor ring). As a consequence, Glu317 is closer to the anomeric carbon (3.45 Å) than it was in the complex obtained by structural alignment (4.09 Å, see Figure VI-2).

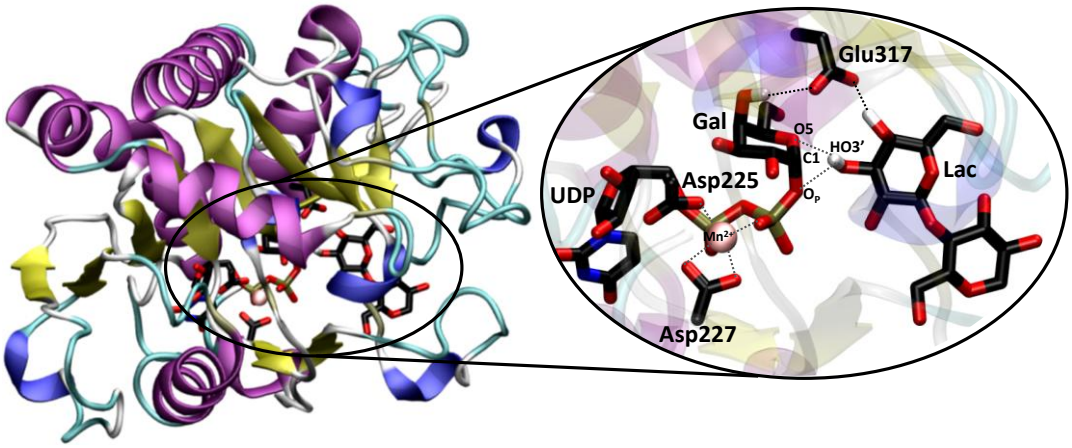

Figure VI-4. Ternary complex of α3GalT with UDP-Gal and Lac obtained from the QM/MM MD simulation (average structure over the last 200 fs). The atoms of the QM region are represented in licorice representation (water molecules and hydrogen atoms have been omitted for clarity, except the acceptor O3'H and O4'H and the donor O4H).

As shown in Figure VI-4, Glu317 is well positioned for nucleophilic attack on the donor anomeric carbon. Therefore, a S_N2-type of reaction (Figure VI-1a) seems to be feasible. However, a front-face type of reaction (Figure VI-1b), cannot be ruled out, as the O3'H of the acceptor interacts with the UDP-Gal donor (specifically, the HO3' proton interacts either with O_P atom or the O5 of the galactose, oscillating among the two on a timescale of a few femtoseconds). In fact, the relative position between the donor and the acceptor is similar to what we previously found for trehalose-6-phosphate (OtsA),[110] a GT that lacks the putative nucleophile and operates by means of a front-face mechanism.[108c] Therefore, the structure of the ternary complex alone does not give any clue on the type of mechanistic transformation.

To inspect the glycosyl transfer reaction in α3galT, we used QM/MM metadynamics with two CVs corresponding to the main bonds undergoing breaking and formation (see Figure VI-3). The evolution of the system during the metadynamics simulation shows that once the C1-O_P bond breaks, the donor sugar collapses with Glu317, forming a glycosyl-enzyme species. Afterwards, the C1-O_{Glu317} bond breaks and the galactose sugar moves towards the acceptor, forming the glycosidic bond. As mentioned before, the choice of the CVs does not impose that C1 binds to Glu317. Hence, the fact that it does is a clear indication that the system prefers to form a glycosyl-enzyme species rather than migrating directly towards the acceptor, as it would be the case in a front-face type of mechanism. Additional simulations using more sophisticated collective variables (e.g., including the transfer of the HO3' proton as a CV, see next section for further details) also led to C1-OGlu317 bond formation, thus providing further confirmation. Altogether, these results strongly indicate that the double-displacement is the only possible mechanism in α3GalT.

The free-energy surface (FES) obtained from the metadynamics simulation is shown in Figure VI-5, top image. Several minima can be identified. Those in the reactants region (R'', R', and R) correspond to stable positions of the acceptor in the active site. The minimum on the lower-right corner (P) corresponds to the Galα3Galβ4Glc trisaccharide, that is, the products of the reaction. The remaining minima correspond to glycosyl-enzyme species. A detailed description of the reaction mechanism can be obtained by following the minimum free-energy pathway connecting reactants and products. The bottom images of Figure VI-5 show representative snap-shots of relevant configurations along this pathway and Figure VI-6 shows the evolution of the most relevant distances (see the absolute values of the distances in the appendix IV, page 120). The reaction begins with the approach of the acceptor towards the donor (from R'' to R), tightening the interactions between Glu317 and both donor and acceptor. At this point, the sugar-phosphate bond starts to break, the anomeric carbon moves up towards Glu317 and the O3'-HO3' bond rotates to establish a strong interaction with the phosphate leaving group. Formation of the glycosyl-enzyme complex (R→1→2) is a dissociative process, in which the cleavage of the C1-O_P bond precedes the formation of the C1-O_{Glu317} bond (Figures VI-5 and VI-6) and involves an energy barrier of 23 kcal·mol^{-1}. This value is similar to the approximate values obtained from the measurements of reaction rate constants of retaining GTs (17-20 kcal·mol^{-1}).[105a, 106, 108c, 121] Most importantly, the energy barrier is identical to the one obtained for OtsA, a GT lacking a putative nucleophile that, unlike α3GalT, operates by means of a front-face type of reaction.[110]

As shown in Figure VI-5, the conformation of the donor sugar changes from being a chair ($^4C_1$) to a boat ($^{1,4}B$) on the glycosyl–enzyme intermediate (**2**) (see page 13 of Chapter I for a description of the different conformations of carbohydrates). This conformational change, triggered by the nucleophilic substitution, facilitates the subsequent step of the reaction (deglycosylation) by placing the O3' atom in a suitable position to attack the anomeric carbon. It thus follows a conformational itinerary that favors catalysis, as found in glycosyl hydrolases.[122] To complete the reaction, the C1-O_{Glu317} bond breaks as the C1-O3' glycosidic bond forms, in concert with the transfer of the HO3' proton to the leaving phosphate group. This step involves a lower energy barrier (13 kcal·mol^{-1}) than the cleavage of the C1-O_P bond, supporting previous experimental predictions.[34]

We have performed additional calculations comparing the performance of several functionals for α3GalT. In particular, energy differences between the reactants (**R**) and the transition state (**1**) of the first step of the reaction (Figure VI-5) were computed, using the structures obtained from the QM/MM metadynamics simulations. For each structure, a QM/MM optimization was performed and the potential energy difference was computed (in the case of the TS structure, a constraint on CV-space was added to ensure that the system does not evolve towards reactants

of the covalent intermediate). The energy barrier was computed in this way considering four different functionals (PBE, HCTH, BP86 and BLYP) (hybrid functionals, which are computationally prohibitive for plane-wave calculations, were not tested). Except HCTH, for which the energy barrier rises by 3.6 kcal·mol^{-1}, all functionals tested gave essentially the same results (± 0.5 kcal·mol^{-1}). In view of these results and the previous work in the literature, we consider the PBE functional good enough to describe the glycosyl transfer reaction. The relative error of the metadynamics simulations is expected to be smaller than the error due to the functional employed.

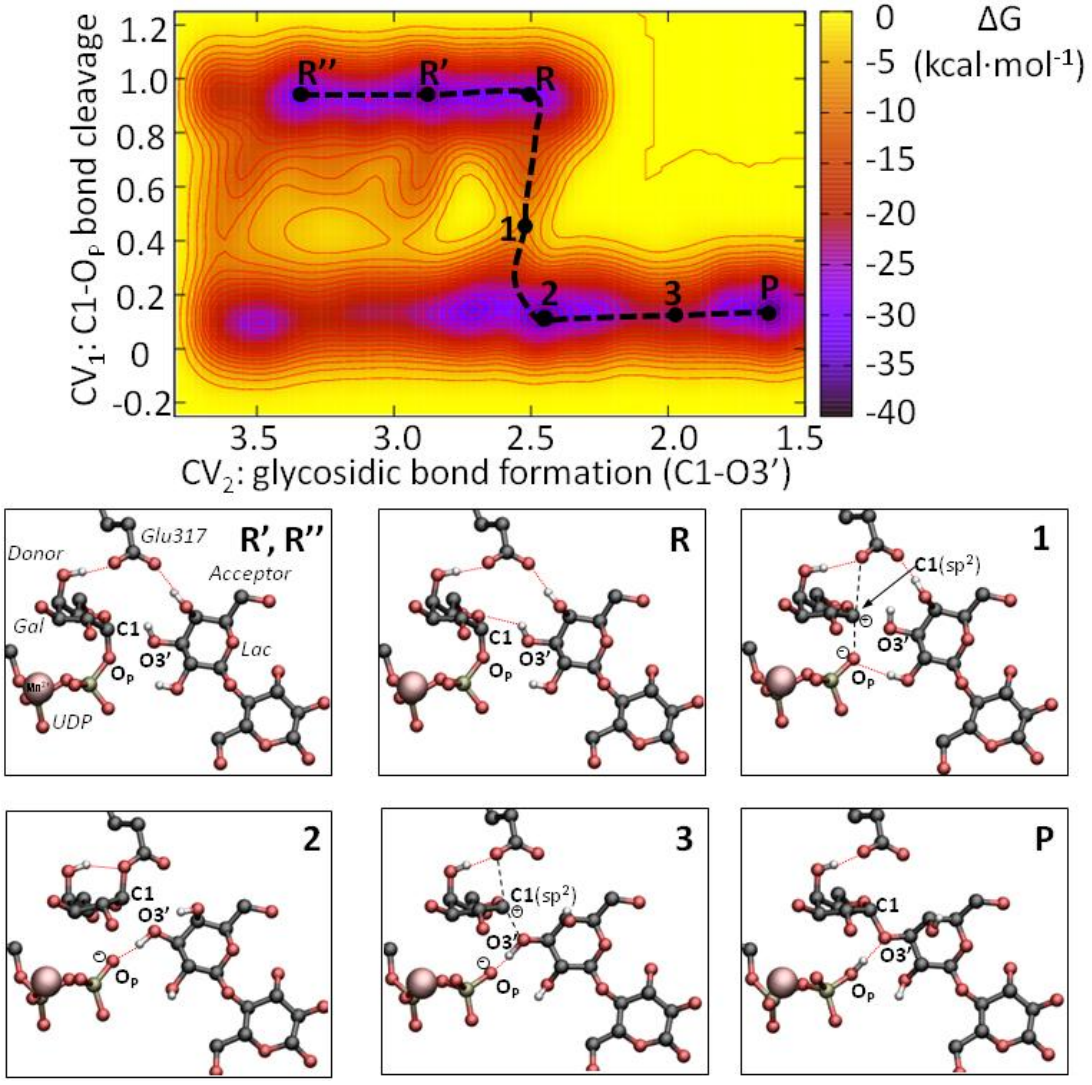

Figure VI-5. Top image: Computed free-energy landscape with respect to the two chosen CVs. Each contour line corresponds o 2 kcal·mol^{-1}. Bottom images: Atomic rearrangement along the reaction pathway. Bonds being broken/formed are represented by a dashed black line (configurations **1** and **3**), whereas relevant hydrogen bonds are represented by a dotted red line. Hydrogen atoms have been omitted for clarity, except HO2', HO3' and HO4'.

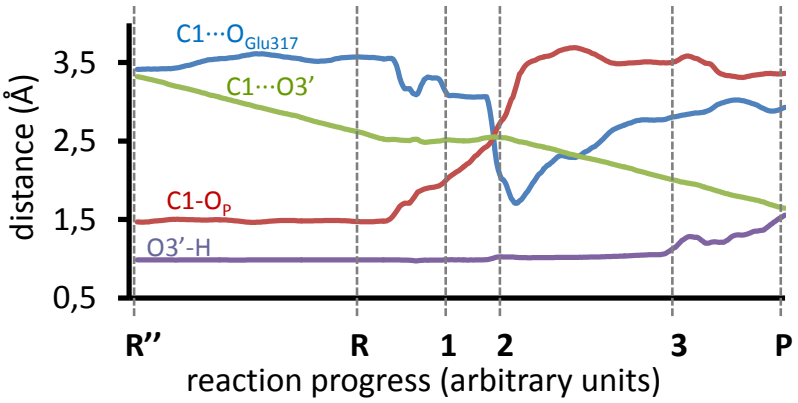

Figure VI-6. Evolution of the most relevant distances involving the donor and the acceptor (see atom labeling in Figure VI-3) along the reaction pathway (Figure VI-5, top). Each distance is an average from all configurations falling into a small region (± 0.05 in terms of the CVs) around the corresponding point of the FES.

To quantify the role of Glu317 in the reaction mechanism, additional metadynamics simulations were performed by using a model that knocks out the reactivity of this residue, either by preventing that it binds to the sugar donor or eliminating its electrostatic interaction. The first metadynamics simulation used the same CVs as in the previous reaction but kept the carboxylate side chain of Glu317 in the MM region. The free energy surface is shown in Figure VI-7. Because the carboxylate side chain is not allowed to bind to the anomeric carbon of the Gal donor fragment, a double displacement mechanism is precluded in this case. Analysis of the reconstructed FES (Figure VI-7) shows that the glycosyl transfer reaction takes place via a front-face mechanism, with the formation of a very stable oxocarbenium ion intermediate. However, the global energy barrier of the reaction increases substantially with respect to the "normal" simulation (from 23 to 30 kcal/mol), indicating that the front-face reaction is not a competitive reaction pathway in α3GalT.

Another simulation was performed by, not only including Glu317 in the MM region (as in the previous case), but also setting its atomic charges to zero, thus cancelling the electrostatic interactions of the Glu317 side chain with the donor and acceptor moieties. The filling of the reactants well reached energies of 35 kcal/mol without the reaction taking place, which indicates that the energy barrier is higher than this value. For this reason, the simulation was not pursued further.

In both cases, the energy barrier of the reaction increases, highlighting the crucial role of Glu317 in the reactive process. These results are consistent with the loss of enzyme activity for Glu317 mutants.[105a]

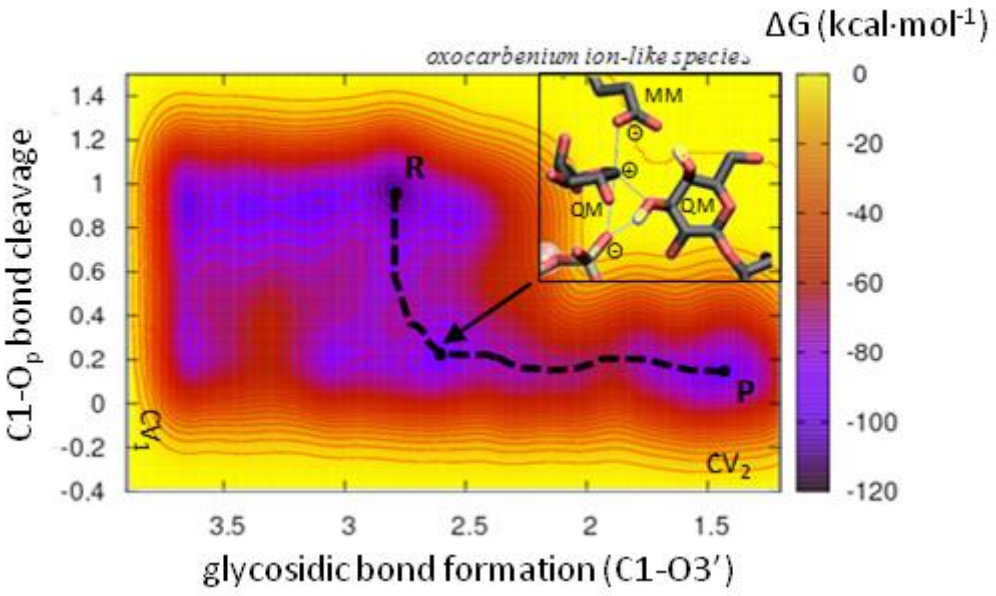

Figure VI-7. Computed free energy surface of the glycosyl transfer reaction when the side chain of Glu317 is not included in the QM region. Each contour line corresponds to 5 kcal·mol^{-1}.

Conclusions

In summary, using classical MD simulations and ab initio QM/MM metadynamics, we unambiguously demonstrate that a GT with a putative nucleophile residue, such as α3GalT and GTA/GTB, operates by means of a double-displacement mechanism, with the formation of a glycosyl-enzyme covalent intermediate as the rate-limiting step. Therefore, two mechanistic outcomes operate in GTs, depending on the presence or absence of a properly positioned nucleophile residue. This is consistent with very early work by Lewis and Boozer on non-enzymatic systems which showed that the mechanistic landscape is finely poised and that relatively subtle environmental changes may lead to quite different mechanisms.[123]

The present calculations provide a detailed picture of the atomic rearrangement during the complete reaction pathway, confirming the analyses of recent experimental studies.[105a, 106] They also explain why hydrolysis of UDP-Gal takes place with retention of configuration.[106, 124] However, an intriguing question is raised- why has nature placed a carboxylate residue in the active site of a few GTs so that they undergo a different type of reaction (double-displacement) than most GTs, which presumably follow a front-face mechanism? In our opinion, the two types of mechanisms for GTs are not opposed. Their main difference is the way the enzyme stabilizes the oxocarbenium-ion-like species that forms upon cleavage of the donor sugar–phosphate bond. In OtsA (and related GTs), the electrostatic potential at the active site is such that it can stabilize the oxocarbenium-ionlike intermediate for a very short time (picoseconds), but long enough for the active site to reorganize and the

oxocarbenium-ion species and the acceptor to move one towards the other.[110] In the case of family 6 GTs (α3GalT and blood group galactosyltransferases), the oxocarbenium-ion-like transition state is stabilized by the formation of a covalent bond with a carboxylate residue. Therefore, both modes of operation can be considered as variations of a common mechanism, a two steps reaction via oxocarbenium-ion-like transition states that flank an intermediate, either an oxocarbenium ion or covalent glycosyl–enzyme depending of the fine structure of the active site.

Chapter VII – Summary and Conclusions

Summary and Conclusions

In this thesis, *ab initio* molecular dynamics simulations have been used to study the formation of a thiolate monolayer-protected gold cluster (AuMPCs), using metadynamics as a sampling method, with the aim to determine how neutral gold clusters and thiol molecules can form the monolayer protecting units that are observed in all the AuMPCs so far characterized experimentally. Additionally, hybrid quantum mechanics / molecular mechanics simulations have been performed to study the structural and electronic properties of an AuMPC protected by an organic ligand (the glutathione tripeptide). Furthermore, its interaction with an antibody has been modelled with the aim to determine which amino acids are key to the binding of the AuMPC to the antibody, as well as elucidating the ligand exchange reaction mechanism. Finally, the molecular mechanism of a glycosyltransferase is studied to determine whether a glycosyl-enzyme bond is needed or not for the reaction to take place, a question of current interest in glycobiology. The main conclusions reached in this dissertation are the following:

- The interaction between neutral gold clusters and alkanethiol molecules leads to the formation of a thiolate monolayer-protected gold cluster (AuMPC) through a reduction-oxidation reaction. More precisely, a monomeric staple motif (*i.e.* the smallest protecting unit found in AuMPCs) is formed when Au_4 interacts with methanethiols, releasing molecular hydrogen.

- H-Au and Au-thiol bonds are formed during the reactive process, as previously predicted by experimental studies. In addition, we demonstrate that they are crucial to stabilize intermediate structures along the reaction pathway.

- The general structure of AuMPCs is not affected either by water solvation or by changes in the identity of the ligands.

- The HOMO-LUMO gap of AuMPCs depends significantly on the type of ligands and the (explicit) solvent environment. The good performance of GGA functionals (such as PBE) for HOMO-LUMO gaps previously reported in theoretical studies is due to a compensation of errors. Calculations of isolated AuMPCs give systematically larger values for the HOMO-LUMO gap than solvated AuMPCs. Only an accurate description of the real system, *i.e.* the AuMPC solvated in water, as used in experiments, in combination with the use of a functional that includes corrections/improvements to GGA can guarantee that the computed value is comparable with the experimental one for the right reasons.

- The ligand exchange reaction follows an associative double S_N2-like mechanism with low activation barriers (13.9 and 16.5 kcal/mol) when the lateral alkanethiolate of a dimeric staple motif is substituted.

- The energy of the binding process increases when the vicinity of the incoming thiol is neutral and not positively charged, due to the lack of strong interactions between the antibody and the AuMPC. This result is in agreement with previously reported experimental studies on a similar system, opening the door to future computational investigations to find the best mutations to do in a certain biomolecule in order to increase the reactivity towards a desired AuMPC.

- A glycosyltransferase with a putative nucleophile residue, such as α3GalT, operates by means of a double-displacement mechanism, with the formation of a glycosyl-enzyme covalent intermediate as the rate-limiting step.

Publications

and

Presentations in Congresses

Publications

The present thesis has given rise to the following publications:

- Rojas-Cervellera V, Giralt E, Rovira C. *Staple Motifs, Initial Steps in the Formation of Thiolate-Protected Gold Nanoparticles: How Do They Form?* **Inorganic Chemistry** (2012), 51: 11422-11429.

- Rojas-Cervellera V, Ardèvol A, Boero M, Planas A, Rovira C. *Formation of a Covalent Glycosyl-Enzyme Species in a Retaining Glycosyltransferase.* **Chemistry – A European Journal** (2013), 19: 14018-14023.

- Rojas-Cervellera V, Rovira C, Akola J. *Study of a peptide protected gold cluster. Submitted.*

- Rojas-Cervellera V, Rovira C, Akola J. *Conjugation between an antibody and a peptide protected gold cluster: studying the specificity of the interactions. Submitted*

Besides, the work during the PhD training has resulted in other publications not included in this thesis:

- Cazorla C, Rojas-Cervellera V, Rovira C. *Calcium-based functionalization of carbon nanostructures for peptide immobilization in aqueous media.* **Journals of Materials Chemistry** (2012), 22: 19684-19693.

- López-Gallego F, Gorostiza P, Rojas-Cervellera V, Rovira C. *Optical Control of Enzyme Enantioselectivity in Solid Phase.* **ACS Catalysis** (2014), 4: 1004-1009.

- Bagherzadeh K, Rojas-Cervellera V, Ganjali M R, Saboury A A, Rovira C, Amanlou M. *A look into tyrosinase lag phase by QM/MM MD simulations. To be submitted.*

- Thompson A, Rojas-Cervellera V, Rovira C, Williams S, Davies G. *Is an epoxide a possible intermediate of a glycoside hydrolase? To be submitted*

Presentations in Congresses

The work of this thesis has been presented in the following congresses:

- CECAM tutorial "Understanding Molecular Simulations, MolSim" (January 2012). University of Amsterdam (Netherlands). **Poster presentation**.
- Stabilized Noble Metal Nanoparticles, recent advances and present challenges (June 2013). Aalto University (Helsinki, Finland). **Poster presentation.**
- CPMD Meeting 2013: Matter, life, light from ab initio molecular dynamics simulations (September 2013). University of Leipzig (Germany). **Poster presentation.**
- Frontiers in dynamics simulations of biological molecules (November 2013). Institute for Research in Biomedicine (Barcelona, Spain). **Poster presentation.**

Appendix

Appendix I

	$Au_4 + 2CH_3SH$				$Au_4 + 4CH_3SH$			
	AuC	**H'**	**H''**	**S'**	**AuC**	**H'**	**H''**	**H'''**
R	0	0.006	0.006	-0.017	0	0.006	0.006	0.006
I1	-0.257	0.062	0.048	-0.031	-0.577	0.001	0.04	0.037
I2	0.022	-0.166	0.048	-0.033	0.022	-0.15	0.048	0.06
I4	0.687	-0.611	0.06	-0.289	0.402	-0.636	0.056	0.07
I6	0.796	-0.349	-0.169	-0.430	0.754	-0.136	-0.148	0.063
I7	0.884	-0.088	-0.171	-0.474	0.771	-0.229	-0.402	0.06
P	0.561	0.014	0.09	-0.411	0.474	0.014	0.09	0.06

Table AI-I. Hirshfeld charges computed in all relevant configurations of the free energy surfaces of the reaction of two and four CH_3SH molecules with Au_4 (Figure III-3).

Appendix II

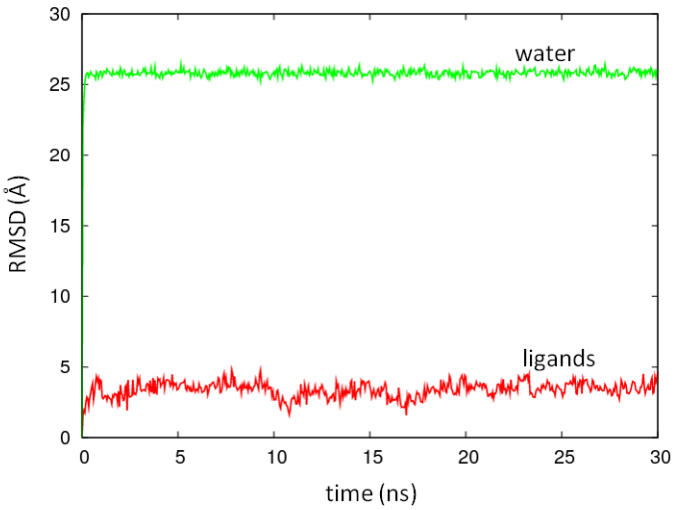

Figure AII-1. Root mean square displacement (RMSD) of $Au_{25}(SCH_3)_{18}^-$ during the classical molecular mechanics simulation. The red and green lines are the RMSD of the methanethiolate ligands and the water molecules, respectively.

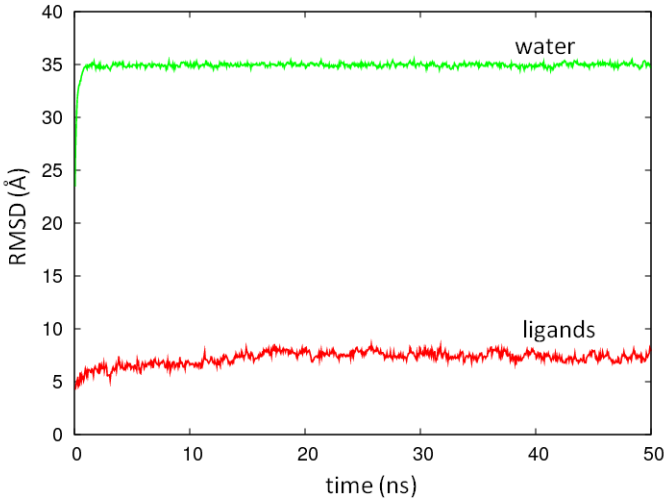

Figure AII-2. Root mean square displacement (RMSD) of $Au_{25}(GSH)_{18}^-$ during the classical molecular mechanics simulation. The red and green lines are the RMSD of the glutathione ligands and the water molecules, respectively.

System	a	a'	c	S	S_{ap}
Isolated $Au_{25}(SCH_3)_{18}^-$	-0.673	-0.176	0.455	-0.263	-0.423
Solvated $Au_{25}(SCH_3)_{18}^-$	-0.649	-0.182	0.445	-0.307	-0.436
Solvated $Au_{25}(GSH)_{18}^-$	-0.591	-0.176	0.400	-0.294	-0.481

Table AII-1. Hirshfeld charges of the five key atoms of $Au_{25}(SR)_{18}^-$. The labelling of the atoms is the same as in Figure 1.

Appendix III

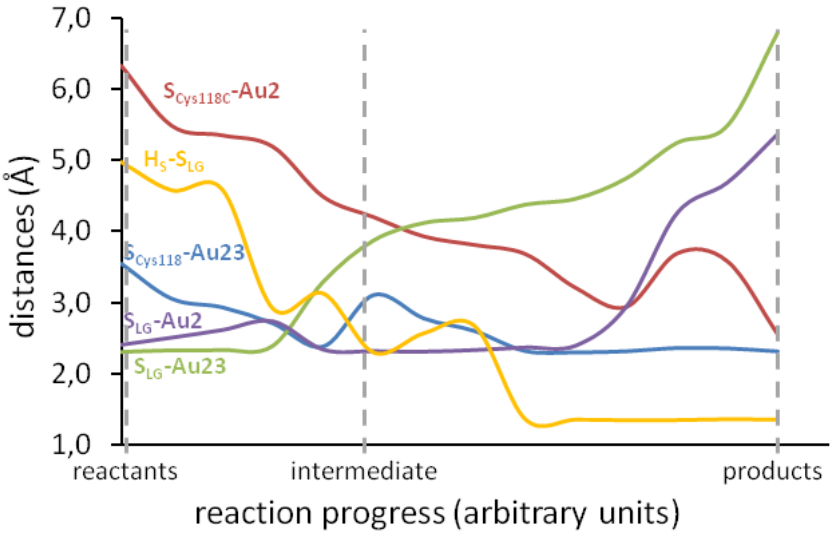

Figure AIII-1. Evolution of the most relevant distances involved in the ligand exchange reaction.

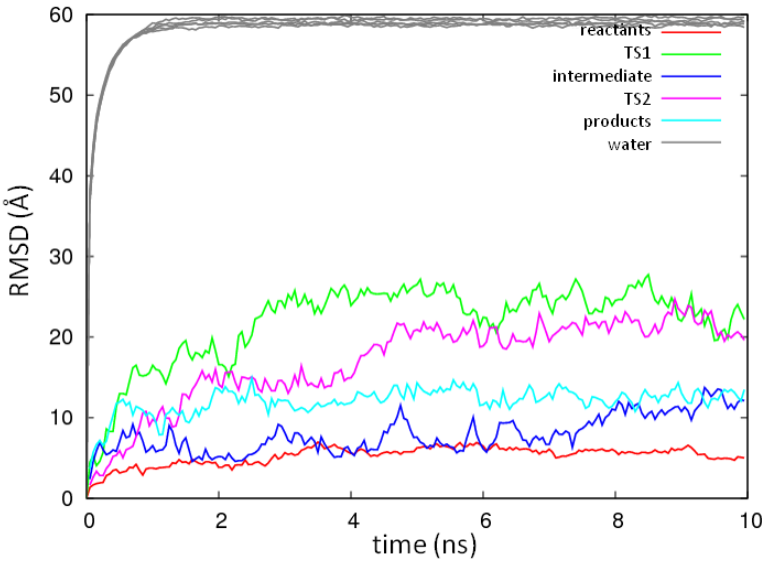

Figure AIII-2. Root mean square displacement (RMSD) of the five points studied along the potential energy surface of 114C. The colour lines correspond to the RMSD of glutathione and grey lines are the RMSD of water.

Appendix IV

	C1 - O_p	C1 - O3’	C1 - O5	C1 - $Oglu_{317}$	O3’-H
R’’	1.47 (0.05)	3.32 (0.02)	1.41 (0.03)	3.4 (0.1)	0.99 (0.01)
R	1.48 (0.04)	2.54 (0.02)	1.41 (0.03)	3.6 (0.1)	0.99 (0.01)
1	2.22 (0.1)	2.50 (0.02)	1.31 (0.01)	3.06 (0.01)	0.99 (0.01)
2	3.3 (0.1)	2.49 (0.03)	1.37 (0.03)	1.7 (0.1)	1.0 (0.02)
3	3.5 (0.2)	2.00 (0.02)	1.32 (0.02)	2.8 (0.1)	1.1 (0.1)
P	3.4 (0.1)	1.63 (0.02)	1.39 (0.02)	2.9 (0.1)	1.6 (0.2)

Table AIV-1. Relevant distances along the minimum energy reaction path. Distances are in angstroms (standard deviations are given in parenthesis).

Bibliography

Bibliography

1. Faraday, M., The Bakerian Lecture: Experimental Relations of Gold (and Other Metals) to Light. *Philos. Trans. R. Soc. London* **1857,** *147*, 145-181.
2. (a) Turkevich, J., Colloidal Gold. Part I. *Gold Bull.* **1985,** *18* (3), 86-91; (b) Turkevich, J., Colloidal Gold. Part II. *Gold Bull.* **1985,** *18* (4), 125-131.
3. Thompson, D., Michael Faraday's Recognition of Ruby Gold: the Birth of Modern Nanotechnology. *Gold Bull.* **2007,** *40* (4), 267-269.
4. Turkevich, J.; Stevenson, P. C.; Hillier, J., A Study of the Nucleation and Growth Processes in the Synthesis of Colloidal Gold. *Discuss. Faraday Soc.* **1951,** *11* (0), 55-75.
5. Brust, M.; Walker, M.; Bethell, D.; Schiffrin, D. J.; Whyman, R., Synthesis of Thiol-Derivatized Gold Nanoparticles in a 2-Phase Liquid-Liquid System. *J. Chem. Soc.-Chem. Commun.* **1994,** (7), 801-802.
6. Porter, M. D.; Bright, T. B.; Allara, D. L.; Chidsey, C. E. D., Spontaneously Organized Molecular Assemblies. Structural Characterization of n-Alkyl Thiol Monolayers on Gold by Optical Ellipsometry, Infrared Spectroscopy, and Electrochemistry. *J. Am. Chem. Soc.* **1987,** *109* (12), 3559-3568.
7. Jana, N. R.; Peng, X., Single-Phase and Gram-Scale Routes toward Nearly Monodisperse Au and Other Noble Metal Nanocrystals. *J. Am. Chem. Soc.* **2003,** *125* (47), 14280-14281.
8. Zheng, M.; Davidson, F.; Huang, X., Ethylene Glycol Monolayer Protected Nanoparticles for Eliminating Nonspecific Binding with Biological Molecules. *J. Am. Chem. Soc.* **2003,** *125* (26), 7790-7791.
9. Negishi, Y.; Takasugi, Y.; Sato, S.; Yao, H.; Kimura, K.; Tsukuda, T., Magic-Numbered Aun Clusters Protected by Glutathione Monolayers (n = 18, 21, 25, 28, 32, 39): Isolation and Spectroscopic Characterization. *J. Am. Chem. Soc.* **2004,** *126* (21), 6518-6519.
10. Stoeva, S.; Klabunde, K. J.; Sorensen, C. M.; Dragieva, I., Gram-Scale Synthesis of Monodisperse Gold Colloids by the Solvated Metal Atom Dispersion Method and Digestive Ripening and Their Organization into Two- and Three-Dimensional Structures. *J. Am. Chem. Soc.* **2002,** *124* (10), 2305-2311.
11. Matthiesen, J. E.; Jose, D.; Sorensen, C. M.; Klabunde, K. J., Loss of Hydrogen upon Exposure of Thiol to Gold Clusters at Low Temperature. *J. Am. Chem. Soc.* **2012**.
12. (a) Jadzinsky, P. D.; Calero, G.; Ackerson, C. J.; Bushnell, D. A.; Kornberg, R. D., Structure of a Thiol Monolayer-Protected Gold Nanoparticle at 1.1 Å Resolution. *Science* **2007,** *318* (5849), 430-3; (b) Garzon, I. L.; Rovira, C.; Michaelian, K.; Beltran, M. R.; Ordejon, P.; Junquera, J.; Sanchez-Portal, D.; Artacho, E.; Soler, J. M., Do Thiols Merely Passivate Gold Nanoclusters? *Phys. Rev. Lett.* **2000,** *85* (24), 5250-1.
13. (a) Aiken III, J. D.; Finke, R. G., A Review of Modern Transition-Metal Nanoclusters: their Synthesis, Characterization, and Applications in Catalysis. *J. Mol. Cat. A: Chem.* **1999,** *145* (1–2), 1-44; (b) Garzón, I. L.; Reyes-Nava, J. A.; Rodríguez-Hernández, J. I.; Sigal, I.; Beltrán, M. R.; Michaelian, K., Chirality in Bare and Passivated Gold Nanoclusters. *Phys. Rev. B* **2002,** *66* (7), 073403; (c) Häkkinen, H.; Walter, M.; Grönbeck, H., Divide and Protect: Capping Gold Nanoclusters with Molecular Gold−Thiolate Rings. *J. Phys. Chem. B* **2006,** *110* (20), 9927-9931.
14. Pei, Y.; Gao, Y.; Zeng, X. C., Structural Prediction of Thiolate-Protected Au38: a Face-Fused Bi-Icosahedral Au Core. *J. Am. Chem. Soc.* **2008,** *130* (25), 7830-2.
15. (a) Jiang, D. E.; Chen, W.; Whetten, R. L.; Chen, Z. F., What Protects the Core When the Thiolated Au Cluster is Extremely Small? *J. Phys. Chem. C* **2009,** *113* (39), 16983-16987; (b) Zeng, C.; Liu, C.; Chen, Y.; Rosi, N. L.; Jin, R., Gold–Thiolate Ring as a Protecting Motif in the Au20(SR)16 Nanocluster and Implications. *J. Am. Chem. Soc.* **2014,** *136* (34), 11922-11925.

16. Rojas-Cervellera, V.; Giralt, E.; Rovira, C., Staple Motifs, Initial Steps in the Formation of Thiolate-Protected Gold Nanoparticles: How Do They Form? *Inorg. Chem.* **2012,** *51* (21), 11422-11429.
17. Walter, M.; Akola, J.; Lopez-Acevedo, O.; Jadzinsky, P. D.; Calero, G.; Ackerson, C. J.; Whetten, R. L.; Gronbeck, H.; Hakkinen, H., A Unified View of Ligand-Protected Gold Clusters as Superatom Complexes. *Proc. Natl. Acad. Sci.* **2008,** *105* (27), 9157-62.
18. (a) Knight, W. D.; Clemenger, K.; Deheer, W. A.; Saunders, W. A.; Chou, M. Y.; Cohen, M. L., Electronic Shell Structure and Abundances of Sodium Clusters. *Phys. Rev. Lett.* **1984,** *52* (24), 2141-2143; (b) Deheer, W. A., The Physics of Simple Metal-Clusters. Experimental Aspects and Simple-Models. *Rev. Mod. Phys.* **1993,** *65* (3), 611-676.
19. (a) Lopez-Acevedo, O.; Tsunoyama, H.; Tsukuda, T.; Hakkinen, H.; Aikens, C. M., Chirality and Electronic Structure of the Thiolate-Protected Au38 Nanocluster. *J. Am. Chem. Soc.* **2010,** *132* (23), 8210-8; (b) Akola, J.; Walter, M.; Whetten, R. L.; Hakkinen, H.; Gronbeck, H., On the Structure of Thiolate-Protected Au25. *J. Am. Chem. Soc.* **2008,** *130* (12), 3756-7.
20. (a) Qian, H.; Eckenhoff, W. T.; Zhu, Y.; Pintauer, T.; Jin, R., Total Structure Determination of Thiolate-Protected Au38 Nanoparticles. *J. Am. Chem. Soc.* **2010,** *132* (24), 8280-1; (b) Heaven, M. W.; Dass, A.; White, P. S.; Holt, K. M.; Murray, R. W., Crystal Structure of the Gold Nanoparticle $[N(C_8H_{17})_4][Au_{25}(SCH_2CH_2Ph)_{18}]$. *J. Am. Chem. Soc.* **2008,** *130* (12), 3754-5.
21. Negishi, Y.; Nobusada, K.; Tsukuda, T., Glutathione-Protected Gold Clusters Revisited: Bridging the Gap Between Gold(I)-Thiolate Complexes and Thiolate-Protected Gold Nanocrystals. *J. Am. Chem. Soc.* **2005,** *127* (14), 5261-70.
22. Shaw, C. F., Gold-Based Therapeutic Agents. *Chem. Rev.* **1999,** *99* (9), 2589-2600.
23. Templeton, A. C.; Wuelfing, W. P.; Murray, R. W., Monolayer-Protected Cluster Molecules. *Acc. Chem. Res.* **2000,** *33* (1), 27-36.
24. Tiwari, P.; Vig, K.; Dennis, V.; Singh, S., Functionalized Gold Nanoparticles and Their Biomedical Applications. *Nanomaterials* **2011,** *1* (1), 31-63.
25. Guerrero, S.; Araya, E.; Fiedler, J. L.; Arias, J. I.; Adura, C.; Albericio, F.; Giralt, E.; Arias, J. L.; Fernandez, M. S.; Kogan, M. J., Improving the Brain Delivery of Gold Nanoparticles by Conjugation with an Amphipathic Peptide. *Nanomedicine (Lond.)* **2010,** *5* (6), 897-913.
26. Kogan, M. J.; Bastus, N. G.; Amigo, R.; Grillo-Bosch, D.; Araya, E.; Turiel, A.; Labarta, A.; Giralt, E.; Puntes, V. F., Nanoparticle-Mediated Local and Remote Manipulation of Protein Aggregation. *Nano Lett.* **2006,** *6* (1), 110-5.
27. Soto, C.; Sigurdsson, E. M.; Morelli, L.; Kumar, R. A.; Castano, E. M.; Frangione, B., Beta-sheet Breaker Peptides Inhibit Fibrillogenesis in a rat Brain Model of Amyloidosis: Implications for Alzheimer's Therapy. *Nat. Med.* **1998,** *4* (7), 822-6.
28. Bastús, N. G.; Sánchez-Tilló, E.; Pujals, S.; Farrera, C.; Kogan, M. J.; Giralt, E.; Celada, A.; Lloberas, J.; Puntes, V., Peptides Conjugated to Gold Nanoparticles Induce Macrophage Activation. *Mol. Immun.* **2009,** *46* (4), 743-748.
29. (a) Liu, Y.; Liu, Y.; Raymond, R. L.; Zeng, X., Single Chain Fragment Variable Recombinant Antibody Functionalized Gold Nanoparticles for a Highly Sensitive Colorimetric Immunoassay. *Biosensors & bioelectronics* **2009,** *24* (9), 2853-2857; (b) Ackerson, C. J.; Jadzinsky, P. D.; Jensen, G. J.; Kornberg, R. D., Rigid, Specific, and Discrete Gold Nanoparticle/Antibody Conjugates. *J. Am. Chem. Soc.* **2006,** *128* (8), 2635-40.
30. Davies, G. J.; Gloster, T. M.; Henrissat, B., Recent Structural Insights into the Expanding World of Carbohydrate-Active Enzymes. *Curr. Op. Struc. Biol.* **2005,** *15* (6), 637-645.
31. (a) Becker, D. J.; Lowe, J. B., Fucose: biosynthesis and biological function in mammals. *Glycobiology* **2003,** *13* (7), 41R-53R; (b) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and the Immune System. *Science* **2001,** *291* (5512), 2370-2376; (c) Hurtley, S.; Service, R.; Szuromi, P., Cinderella's Coach Is Ready. *Science* **2001,** *291* (5512), 2337; (d) Zachara, N. E.; Hart, G. W., The Emerging Significance of O-GlcNAc in Cellular Regulation. *Chem. Rev.* **2002,** *102* (2), 431-438.

32. IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and Nomenclature Committee of IUBMB (NC-IUBMB) Newsletter 1996. *Glycoconj. J.* **1998,** *15* (7), 637-647.
33. Taniguchi, N. H., K.; Fukuda, M., *Handbook of Glycosyl-transferase and Related Genes.* Springer: Tokyo, 2002.
34. Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G., Glycosyltransferases: Structures, Functions, and Mechanisms. *An. Rev. of Biochem.* **2008,** *77* (1), 521-555.
35. Gastinel, L. N.; Bignon, C.; Misra, A. K.; Hindsgaul, O.; Shaper, J. H.; Joziasse, D. H., *Bovine α1,3-Galactosyltransferase Catalytic Domain Structure and its Relationship with ABO Histo-Blood Group and Glycosphingolipid Glycosyltransferases*. 2001; Vol. 20, p 638-649.
36. Persson, K.; Ly, H. D.; Dieckelmann, M.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. J., Crystal Structure of the Retaining Galactosyltransferase LgtC From Neisseria Meningitidis in Complex with Donor and Acceptor Sugar Analogs. *Nat. Struct. Mol..Biol.* **2001,** *8* (2), 166-175.
37. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active Enzymes Database (CAZy): an Expert Resource for Glycogenomics. *Nuc. Ac. Res.* **2009,** *37* (suppl 1), D233-D238.
38. Zhang, Y.; Swaminathan, G. J.; Deshpande, A.; Boix, E.; Natesh, R.; Xie, Z.; Acharya, K. R.; Brew, K., Roles of Individual Enzyme−Substrate Interactions by α-1,3-Galactosyltransferase in Catalysis and Specificity†,‡. *Biochemistry* **2003,** *42* (46), 13512-13521.
39. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nuc. Ac. Res.* **2000,** *28* (1), 235-242.
40. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer: 2000.
41. Verlet, L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967,** *159* (1), 98-103.
42. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.
43. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Phys. Rev.* **1964,** *136* (3B), B864-B871.
44. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965,** *140* (4A), A1133-A1138.
45. Troullier, N.; Martins, J. L., Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B. Condens. Matter* **1991,** *43* (3), 1993-2006.
46. Nienhaus, G. U., *Protein-ligand Interactions: Methods and Applications.* . Humana Press: 2005.
47. Senn, H. M.; Thiel, W., QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009,** *48* (7), 1198-1229.
48. Warshel, A.; Levitt, M., Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976,** *103* (2), 227-249.
49. Laio, A.; VandeVondele, J.; Rothlisberger, U., A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car–Parrinello Molecular Dynamics Simulations. *J. Chem. Phys.* **2002,** *116* (16), 6941-6947.
50. von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D., Variational Optimization of Effective Atom Centered Potentials for Molecular Properties. *J. Chem. Phys.* **2005,** *122* (1), 014113.
51. (a) Laio, A.; Parrinello, M., Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002,** *99* (20), 12562–12566; (b) Laio, A.; Gervasio, F. L., *Rep. Progr. Phys.* **2008,** *71*, 126601.

52. Biarnés, X. Mecanismes de Preactivació de Substrat en 1,3-1,4-B-glucanasa. Modelització Mitjançant Dinàmica Molecular de Primers Principis. Universitat de Barcelona, Barcelona, 2007.
53. Iannuzzi, M.; Laio, A.; Parrinello, M., Efficient Exploration of Reactive Potential Energy Surfaces using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2003,** *90* (23), 238302.
54. Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L., A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions†. *J. Phys. Chem. B* **2005,** *109* (14), 6676-6687.
55. Barducci, A.; Bonomi, M.; Parrinello, M., Metadynamics. *Comput. Mol. Sci.* **2011,** *1* (5), 826-843.
56. Nair, N. N.; Schreiner, E.; Marx, D., Peptide Synthesis in Aqueous Environments: The Role of Extreme Conditions on Amino Acid Activation. *J. Am. Chem. Soc.* **2008,** *130* (43), 14148-14160.
57. Li, G. T.; Lauer, M.; Schulz, A.; Boettcher, C.; Li, F. T.; Fuhrhop, J. H., Spherical and Planar Gold(0) Nanoparticles with a Rigid Gold(I)-Anion or a Fluid Gold (0)-Acetone Surface. *Langmuir* **2003,** *19* (16), 6483-6491.
58. (a) Letardi, S.; Cleri, F., Interaction of Benzene Thiol and Thiolate with Small Gold Clusters. *J. Chem. Phys.* **2004,** *120* (21), 10062-8; (b) Askerka, M.; Pichugina, D.; Kuz'menko, N.; Shestakov, A., Theoretical Prediction of S-H Bond Rupture in Methanethiol upon Interaction with Gold. *J. Phys. Chem. A* **2012**.
59. Barngrover, B. M.; Aikens, C. M., The Golden Pathway to Thiolate-Stabilized Nanoparticles: Following the Formation of Gold(I) Thiolate from Gold(III) Chloride. *J. Am. Chem. Soc.* **2012,** *134* (30), 12590-5.
60. Hasan, M.; Bethell, D.; Brust, M., The Fate of Sulfur-Bound Hydrogen on Formation of Self-Assembled Thiol Monolayers on Gold: ^{1}H NMR spectroscopic evidence from solutions of gold clusters. *J. Am. Chem. Soc.* **2002,** *124* (7), 1132-3.
61. (a) Walter, M.; Akola, J.; Lopez-Acevedo, O.; Jadzinsky, P. D.; Calero, G.; Ackerson, C. J.; Whetten, R. L.; Gronbeck, H.; Hakkinen, H., A Unified View of Ligand-Protected Gold Clusters as Superatom Complexes. *Proc. Natl. Acad. Sci.* **2008,** *105* (27), 9157-62; (b) Knight, W. D.; Clemenger, K.; Deheer, W. A.; Saunders, W. A.; Chou, M. Y.; Cohen, M. L., Electronic Shell Structure and Abundances of Sodium Clusters. *Phys. Rev. Lett.* **1984,** *52* (24), 2141-2143; (c) Deheer, W. A., The Physics of Simple Metal-Clusters - Experimental Aspects and Simple-Models. *Rev. Mod. Phys.* **1993,** *65* (3), 611-676.
62. Jiang, D. E.; Whetten, R. L.; Luo, W. D.; Dai, S., The Smallest Thiolated Gold Superatom Complexes. *J. Phys. Chem. C* **2009,** *113* (40), 17291-17295.
63. Bonasia, P. J.; Gindelberger, D. E.; Arnold, J., Synthesis and Characterization of Gold(I) Thiolates, Selenolates, and Tellurolates: X-ray Crystal Structures of $Au_4[TeC(SiMe_3)_3]_4$, $Au_4[SC(SiMe_3)_3]_4$, and $Ph_3PAu[TeC(SiMe_3)_3]$. *Inorg. Chem.* **1993,** *32* (23), 5126-5131.
64. Car, R.; Parrinello, M., Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985,** *55* (22), 2471-2474.
65. (a) Goedecker, S.; Teter, M.; Hutter, J., Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B Condens. Matter* **1996,** *54* (3), 1703-1710; (b) Hartwigsen, C.; Goedecker, S.; Hutter, J., Relativistic Separable Dual-Space Gaussian Pseudopotentials from H to Rn. *Phys. Rev. B* **1998,** *58* (7), 3641-3662.
66. Kruger, D.; Fuchs, H.; Rousseau, R.; Marx, D.; Parrinello, M., Interaction of Short-Chain Alkane Thiols and Thiolates with Small Gold Clusters: Adsorption Structures and Energetics. *J. Chem. Phys.* **2001,** *115* (10), 4776-4786.
67. Rousseau, R.; Marx, D., The Interaction of Gold Clusters with Methanol Molecules: Ab initio Molecular Dynamics of $Au_n^+CH_3OH$ and Au_nCH_3OH. *J. Chem. Phys.* **2000,** *112* (2), 761-769.
68. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996,** *77* (18), 3865-3868.

69. (a) Tlahuice, A.; Garzon, I. L., On the Structure of the $Au_{18}(SR)_{14}$ cluster. *Phys. Chem. Chem. Phys.* **2012,** *14* (11), 3737-40; (b) Heinecke, C. L.; Ni, T. W.; Malola, S.; Mäkinen, V.; Wong, O. A.; Häkkinen, H.; Ackerson, C. J., Structural and Theoretical Basis for Ligand Exchange on Thiolate Monolayer Protected Gold Nanoclusters. *J. Am. Chem. Soc.* **2012,** *134* (32), 13316-13322; (c) Tlahuice-Flores, A.; Whetten, R. L.; Jose-Yacaman, M., Ligand Effects on the Structure and the Electronic Optical Properties of Anionic Au25(SR)18 Clusters. *The Journal of Phys. Chem. C* **2013,** *117* (40), 20867-20875.

70. Lopez-Acevedo, O.; Kacprzak, K. A.; Akola, J.; Hakkinen, H., Quantum Size Effects in Ambient CO Oxidation Catalysed by Ligand-Protected Gold Clusters. *Nat. Chem.* **2010,** *2* (4), 329-34.

71. (a) Comas-Vives, A.; Stirling, A.; Lledos, A.; Ujaque, G., The Wacker Process: Inner- or Outer-Sphere Nucleophilic Addition? New Insights From ab initio Molecular Dynamics. *Chem. Eur. J.* **2010,** *16* (29), 8738-47; (b) Stirling, A.; Papai, I., H_2CO_3 Forms via HCO_3^- in water. *J. Phys. Chem. B* **2010,** *114* (50), 16854-9; (c) Alfonso-Prieto, M.; Biarnes, X.; Vidossich, P.; Rovira, C., The Molecular Mechanism of the Catalase Reaction. *J. Am. Chem. Soc.* **2009,** *131* (33), 11751-61; (d) Petersen, L.; Ardevol, A.; Rovira, C.; Reilly, P. J., Molecular Mechanism of the Glycosylation Step Catalyzed by Golgi Alpha-Mannosidase II: a QM/MM Metadynamics Investigation. *J. Am. Chem. Soc.* **2010,** *132* (24), 8291-300; (e) Biarnes, X.; Ardevol, A.; Iglesias-Fernandez, J.; Planas, A.; Rovira, C., Catalytic Itinerary in 1,3-1,4-β-Glucanase Unraveled by QM/MM Metadynamics. Charge is Not Yet Fully Developed at the Oxocarbenium Ion-Like Transition State. *J. Am. Chem. Soc.* **2011,** *133* (50), 20301-9; (f) Ardevol, A.; Rovira, C., The Molecular Mechanism of Enzymatic Glycosyl transfer with Retention of Configuration: Evidence for a Short-Lived Oxocarbenium-Like Species. *Angew. Chem. Int. Ed. Engl.* **2011,** *50* (46), 10897-901; (g) Nair, N. N.; Schreiner, E.; Marx, D., Peptide Synthesis in Aqueous Environments: the Role of Extreme Conditions on Amino Acid Activation. *J. Am. Chem. Soc.* **2008,** *130* (43), 14148-60.

72. Rousseau, R.; Dietrich, G.; Kruckeberg, S.; Lutzenkirchen, K.; Marx, D.; Schweikhard, L.; Walther, C., Probing Cluster Structures with Sensor Molecules: Methanol Adsorbed onto Gold Clusters. *Chem. Phys. Lett.* **1998,** *295* (1-2), 41-46.

73. Joshi, P.; Shewale, V.; Pandey, R., Tryptophan–Gold Nanoparticle Interaction: A First-Principles Quantum Mechanical Study. *J. Phys. Chem. C* **2011,** *115* (46), 22818-22826.

74. Ackerson, C. J.; Sykes, M. T.; Kornberg, R. D., Defined DNA/Nanoparticle Conjugates. *Proc. Natl. Acad. Sci.* **2005,** *102* (38), 13383-5.

75. Zheng, M.; Huang, X., Nanoparticles Comprising a Mixed Monolayer for Specific Bindings with Biomolecules. *J. Am. Chem. Soc.* **2004,** *126* (38), 12047-54.

76. Dainese, T.; Antonello, S.; Gascon, J. A.; Pan, F.; Perera, N. V.; Ruzzi, M.; Venzo, A.; Zoleo, A.; Rissanen, K.; Maran, F., $Au_{25}(SEt)_{18}$, a nearly naked Thiolate-Protected Au_{25} cluster: Structural Analysis by Single Crystal X-ray Crystallography and Electron Nuclear Double Resonance. *ACS Nano* **2014,** *8* (4), 3904-12.

77. Ni, T. W.; Tofanelli, M. A.; Phillips, B. D.; Ackerson, C. J., Structural basis for ligand exchange on $Au_{25}(SR)_{18}$. *Inorg. Chem.* **2014,** *53* (13), 6500-2.

78. Cao, T.; Jin, S.; Wang, S.; Zhang, D.; Meng, X.; Zhu, M., A Comparison of the Chiral Counterion, Solvent, and Ligand used to Induce a Chiroptical Response from Au_{25}-Nanoclusters. *Nanoscale* **2013,** *5* (16), 7589-7595.

79. (a) Shichibu, Y.; Negishi, Y.; Tsukuda, T.; Teranishi, T., Large-scale Synthesis of Thiolated Au_{25} Clusters via Ligand Exchange Reactions of Phosphine-Stabilized Au_{11} Clusters. *J. Am. Chem. Soc.* **2005,** *127* (39), 13464-5; (b) Shichibu, Y.; Negishi, Y.; Tsunoyama, H.; Kanehara, M.; Teranishi, T.; Tsukuda, T., Extremely High Stability of Glutathionate-protected Au_{25} Clusters Against Core Etching. *Small* **2007,** *3* (5), 835-9.

80. Perdew, J. P., Density Functional Theory and the Band Gap Problem. *Int. J. Quantum Chem.* **1985,** *28* (S19), 497-523.

81. (a) Muniz-Miranda, F.; Menziani, M. C.; Pedone, A., Assessment of Exchange-Correlation Functionals in Reproducing the Structure and Optical Gap of Organic-Protected Gold Nanoclusters. *J. Phys. Chem. C* **2014,** *118* (14), 7532-7544; (b) Yang, H.; Wang, Y.; Huang, H.; Gell, L.; Lehtovaara, L.; Malola, S.; Häkkinen, H.; Zheng, N., All-Thiol-Stabilized Ag44 and $Au_{12}Ag_{32}$ nanoparticles with single-crystal structures. *Nat. Commun.* **2013,** *4*; (c) Aikens, C. M., Origin of Discrete Optical Absorption Spectra of $M_{25}(SH)_{18}^-$ Nanoparticles (M = Au, Ag). *J. Phys. Chem. C* **2008,** *112* (50), 19797-19800.

82. Dufour, F.; Fresch, B.; Durupthy, O.; Chaneac, C.; Remacle, F., Ligand and Solvation Effects on the Structural and Electronic Properties of Small Gold Clusters. *J. Phys. Chem. C* **2014,** *118* (8), 4362-4376.

83. Link, S.; Beeby, A.; FitzGerald, S.; El-Sayed, M. A.; Schaaff, T. G.; Whetten, R. L., Visible to Infrared Luminescence from a 28-Atom Gold Cluster. *J. Phys. Chem. B* **2002,** *106* (13), 3410-3415.

84. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-802.

85. Hornak, V.; Simmerling, C., Generation of Accurate Protein Loop Conformations Through Low-Barrier Molecular Dynamics. *Proteins* **2003,** *51* (4), 577-90.

86. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

87. CPMD *Copyright IBM Corp. 1990-2001, Copyright MPI für Festkörperforscung*, Stuttgart, 1997-2004.

88. (a) Nosé, S., A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984,** *52* (2), 255-268; (b) Nosé, S., A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984,** *81* (1), 511-519; (c) Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985,** *31* (3), 1695-1697.

89. Konopka, M.; Rousseau, R.; Stich, I.; Marx, D., Detaching Thiolates from Copper and Gold Clusters: Which Bonds to Break? *J. Am. Chem. Soc.* **2004,** *126* (38), 12103-11.

90. Adamo, C.; Barone, V., Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 model. *J. Chem. Phys.* **1999,** *110* (13), 6158-6170.

91. Zhang, Y.; Lee, T.-S.; Yang, W., A Pseudobond Approach to Combining Quantum Mechanical and Molecular Mechanical Methods. *J. Chem. Phys.* **1999,** *110* (1), 46-54.

92. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-8, 27-8.

93. (a) Vargas, A.; Santarossa, G.; Iannuzzi, M.; Baiker, A., Fluxionality of Gold Nanoparticles Investigated by Born-Oppenheimer Molecular Dynamics. *Phys. Rev. B* **2009,** *80* (19), 195421; (b) Kacprzak, K. A.; Akola, J.; Hakkinen, H., First-principles Simulations of Hydrogen Peroxide Formation Catalyzed by Small Neutral Gold Clusters. *Phys. Chem. Chem. Phys.* **2009,** *11* (30), 6359-6364.

94. Hirshfeld, F. L., Bonded-atom Fragments for Describing Molecular Charge Densities. *Theoret. Chim. Acta* **1977,** *44* (2), 129-138.

95. (a) Goh, J.-Q.; Malola, S.; Häkkinen, H.; Akola, J., Role of the Central Gold Atom in Ligand-Protected Biicosahedral Au24 and Au25 Clusters. *J. Phys. Chem. C* **2013,** *117* (42), 22079-22086; (b) Jiang, D.-e.; Overbury, S. H.; Dai, S., Structure of $Au_{15}(SR)_{13}$ and Its Implication for the Origin of the Nucleus in Thiolated Gold Nanoclusters. *J. Am. Chem. Soc.* **2013,** *135* (24), 8786-8789; (c) Mustalahti, S.; Myllyperkiö, P.; Malola, S.; Lahtinen, T.; Salorinne, K.; Koivisto, J.; Häkkinen, H.; Pettersson, M., Molecule-like Photodynamics of $Au_{102}(pMBA)_{44}$ Nanocluster. *ACS Nano* **2015,** *9* (3), 2328-2335.

96. (a) Hostetler, M. J.; Templeton, A. C.; Murray, R. W., Dynamics of Place-Exchange Reactions on Monolayer-Protected Gold Cluster Molecules. *Langmuir* **1999,** *15* (11), 3782-

3789; (b) Templeton, A. C.; Hostetler, M. J.; Kraft, C. T.; Murray, R. W., Reactivity of Monolayer-Protected Gold Cluster Molecules: Steric Effects. *J. Am. Chem. Soc.* **1998,** *120* (8), 1906-1911.
97. Malby, R. L.; Tulip, W. R.; Harley, V. R.; McKimm-Breschkin, J. L.; Laver, W. G.; Webster, R. G.; Colman, P. M., The Structure of a Complex Between the NC10 Antibody and Influenza Virus Neuraminidase and Comparison with the Overlapping Binding Site of the NC41 Antibody. *Structure* **1994,** *2* (8), 733-46.
98. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25* (13), 1605-12.
99. Song, Y.; Murray, R. W., Dynamics and Extent of Ligand Exchange Depend on Electronic Charge of Metal Nanoparticles. *J. Am. Chem. Soc.* **2002,** *124* (24), 7096-102.
100. (a) Gronbeck, H.; Curioni, A.; Andreoni, W., Thiols and Disulfides on the Au(111) Surface: The Headgroup-Gold Interaction. *J. Am. Chem. Soc.* **2000,** *122* (16), 3839-3842; (b) Andreoni, W.; Curioni, A.; Grönbeck, H., Density Functional Theory Approach to Thiols and Disulfides on Gold: Au(111) Surface and Clusters. *Int. J. Quantum Chem.* **2000,** *80* (4-5), 598-608.
101. (a) Griffith, B. R.; Langenhan, J. M.; Thorson, J. S., 'Sweetening' Natural Products via Glycorandomization. *Curr. Opin. Biotech.* **2005,** *16* (6), 622-630; (b) Zhang, C.; Griffith, B. R.; Fu, Q.; Albermann, C.; Fu, X.; Lee, I.-K.; Li, L.; Thorson, J. S., Exploiting the Reversibility of Natural Product Glycosyltransferase-Catalyzed Reactions. *Science* **2006,** *313* (5791), 1291-1294.
102. Breton, C.; Fournel-Gigleux, S.; Palcic, M. M., Recent Structures, Evolution and Mechanisms of Glycosyltransferases. *Curr. Opin. Struc. Biol.* **2012,** *22* (5), 540-549.
103. (a) Palcic, M. M., Glycosyltransferases as Biocatalysts. *Curr. Opin. Chem. Biol.* **2011,** *15* (2), 226-233; (b) Jørgensen, R.; Grimm, L. L.; Sindhuwinata, N.; Peters, T.; Palcic, M. M., A Glycosyltransferase Inhibitor from a Molecular Fragment Library Simultaneously Interferes with Metal Ion and Substrate Binding. *Angew. Chem.* **2012,** *124* (17), 4247-4251; (c) Jørgensen, R.; Grimm, L. L.; Sindhuwinata, N.; Peters, T.; Palcic, M. M., A Glycosyltransferase Inhibitor from a Molecular Fragment Library Simultaneously Interferes with Metal Ion and Substrate Binding. *Angew. Chem. Int. Ed.* **2012,** *51* (17), 4171-4175.
104. Vocadlo, D. J.; Davies, G. J., Mechanistic Insights into Glycosidase Chemistry. *Curr. Opin. Chem. Biol.* **2008,** *12* (5), 539-555.
105. (a) Monegal, A.; Planas, A., Chemical Rescue of α3-Galactosyltransferase. Implications in the Mechanism of Retaining Glycosyltransferases. *J. Am. Chem. Soc.* **2006,** *128* (50), 16030-16031; (b) Patenaude, S. I.; Seto, N. O. L.; Borisova, S. N.; Szpacenko, A.; Marcus, S. L.; Palcic, M. M.; Evans, S. V., The Structural Basis for Specificity in Human ABO(H) Blood Group Biosynthesis. *Nat. Struct. Mol. Biol.* **2002,** *9* (9), 685-690.
106. Soya, N.; Fang, Y.; Palcic, M. M.; Klassen, J. S., Trapping and Characterization of Covalent Intermediates of Mutant Retaining Glycosyltransferases. *Glycobiology* **2011,** *21* (5), 547-552.
107. Sinnott, M. L., Catalytic Mechanism of Enzymic Glycosyl Transfer. *Chem. Rev.* **1990,** *90* (7), 1171-1202.
108. (a) Errey, J. C.; Lee, S. S.; Gibson, R. P.; Martinez Fleites, C.; Barry, C. S.; Jung, P. M. J.; O'Sullivan, A. C.; Davis, B. G.; Davies, G. J., Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. *Angew. Chem.* **2010,** *122* (7), 1256-1259; (b) Errey, J. C.; Lee, S. S.; Gibson, R. P.; Martinez Fleites, C.; Barry, C. S.; Jung, P. M. J.; O'Sullivan, A. C.; Davis, B. G.; Davies, G. J., Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. *Angew. Chem. Int. Ed.* **2010,** *49* (7), 1234-1237; (c) Lee, S. S.; Hong, S. Y.; Errey, J. C.; Izumi, A.; Davies, G. J.; Davis, B. G., Mechanistic Evidence for a Front-Side, SNi-type Reaction in a Retaining Glycosyltransferase. *Nat. Chem. Biol.* **2011,** *7* (9), 631-638; (d) Gómez, H.; Polyak, I.; Thiel, W.; Lluch, J. M.; Masgrau, L., Retaining Glycosyltransferase Mechanism Studied by

QM/MM Methods: Lipopolysaccharyl-α-1,4-galactosyltransferase C Transfers α-Galactose via an Oxocarbenium Ion-like Transition State. *J. Am. Chem. Soc.* **2012,** *134* (10), 4743-4752.
109. Chan, J.; Tang, A.; Bennet, A. J., A Stepwise Solvent-Promoted SNi Reaction of α-d-Glucopyranosyl Fluoride: Mechanistic Implications for Retaining Glycosyltransferases. *J. Am. Chem. Soc.* **2012,** *134* (2), 1212-1220.
110. Ardèvol, A.; Rovira, C., The Molecular Mechanism of Enzymatic Glycosyl Transfer with Retention of Configuration: Evidence for a Short-Lived Oxocarbenium-Like Species. *Angew. Chem. Int. Ed.* **2011,** *50* (46), 10897-10901.
111. André, I.; Tvaroška, I.; Carver, J. P., On the Reaction Pathways and Determination of Transition-State Structures for Retaining α-galactosyltransferases. *Carb. Res.* **2003,** *338* (9), 865-877.
112. (a) Gómez, H.; Lluch, J. M.; Masgrau, L., Essential Role of Glutamate 317 in Galactosyl Transfer by α3GalT: a Computational Study. *Carb. Res.* **2012,** *356* (0), 204-208; (b) Gómez, H.; Lluch, J. M.; Masgrau, L., Substrate-Assisted and Nucleophilically Assisted Catalysis in Bovine α1,3-Galactosyltransferase. Mechanistic Implications for Retaining Glycosyltransferases. *J. Am. Chem. Soc.* **2013,** *135* (18), 7053-7063.
113. Urresti, S.; Albesa-Jové, D.; Schaeffer, F.; Pham, H. T.; Kaur, D.; Gest, P.; van der Woerd, M. J.; Carreras-González, A.; López-Fernández, S.; Alzari, P. M.; Brennan, P. J.; Jackson, M.; Guerin, M. E., Mechanistic Insights into the Retaining Glucosyl-3-phosphoglycerate Synthase from Mycobacteria. *J. Biol. Chemi.* **2012,** *287* (29), 24649-24661.
114. Boix, E.; Zhang, Y.; Swaminathan, G. J.; Brew, K.; Acharya, K. R., Structural Basis of Ordered Binding of Donor and Acceptor Substrates to the Retaining Glycosyltransferase: α-1,3 galactosyltransferase. *J. Biol. Chem.* **2002**.
115. Jamaluddin, H.; Tumbale, P.; Withers, S. G.; Acharya, K. R.; Brew, K., Conformational Changes Induced by Binding UDP-2F-galactose to α-1,3 Galactosyltransferase- Implications for Catalysis. *J. Mol. Biol.* **2007,** *369* (5), 1270-1281.
116. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham Iii, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comp. Phys. Comm.* **1995,** *91* (1–3), 1-41.
117. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comp. Chem.* **2008,** *29* (4), 622-655.
118. Ireta, J.; Neugebauer, J.; Scheffler, M., On the Accuracy of DFT for Describing Hydrogen Bonds: Dependence on the Bond Directionality. *J. Phys. Chem. A* **2004,** *108* (26), 5692-5698.
119. (a) Biarnés, X.; Nieto, J.; Planas, A.; Rovira, C., Substrate Distortion in the Michaelis Complex of Bacillus 1,3–1,4-β-Glucanase: Insight from First Principles Molecular Dynamics Simulations. *J. Biol. Chem.* **2006,** *281* (3), 1432-1441; (b) Biarnes, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M., The Conformational Free Energy Landscape of β-D-glucopyranose. Implications for Substrate Preactivation in β-glucoside Hydrolases. *J. Am. Chem. Soc.* **2007,** *129* (35), 10686-93; (c) Petersen, L.; Ardèvol, A.; Rovira, C.; Reilly, P. J., Mechanism of Cellulose Hydrolysis by Inverting GH8 Endoglucanases: A QM/MM Metadynamics Study. *J. Phys. Chem. B* **2009,** *113* (20), 7331-7339; (d) Lira-Navarrete, E.; Iglesias-Fernández, J.; Zandberg, W. F.; Compañón, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R., Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide N-Acetylgalactosaminyltransferase 2. *Angew. Chem. Int. Ed.* **2014,** *53* (31), 8206-8210.
120. (a) Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L., Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Acc. Chem.Res.* **2006,** *39* (2), 73-81;

(b) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M., Assessing the Accuracy of Metadynamics. *J. Phys. Chem. B* **2005,** *109* (14), 6714-6721.
121. Seo, H. S.; Koo, Y. J.; Lim, J. Y.; Song, J. T.; Kim, C. H.; Kim, J. K.; Lee, J. S.; Choi, Y. D., Characterization of a Bifunctional Enzyme Fusion of Trehalose-6-Phosphate Synthetase and Trehalose-6-Phosphate Phosphatase of Escherichia coli. *Appl.Env. Microb.* **2000,** *66* (6), 2484-2490.
122. Davies, G. J.; Planas, A.; Rovira, C., Conformational Analyses of the Reaction Coordinate of Glycosidases. *Acc. Chem. Res.* **2012,** *45* (2), 308-316.
123. Lewis, E. S.; Boozer, C. E., The Kinetics and Stereochemistry of the Decomposition of Secondary Alkyl Chlorosulfites1. *J. Am. Chem. Soc.* **1952,** *74* (2), 308-311.
124. Sindhuwinata, N.; Munoz, E.; Munoz, F. J.; Palcic, M. M.; Peters, H.; Peters, T., Binding of an Acceptor Substrate Analog Enhances the Enzymatic Activity of Human Blood Group B Galactosyltransferase. *Glycobiology* **2010,** *20* (6), 718-723.