Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

DOCTOR OF SCIENCES

Probing peptide-nanomaterial interfaces by means of molecular dynamics

De Leo, Federica

Award date:
2015

Awarding institution:
University of Namur

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. May. 2026

Thesis submitted for the achievement of *philosophiæ doctor* (Ph.D.) in chemistry

JURY

Prof. Iris Antes

Prof. Davide Bonifazi

Prof. Benoît Champagne

Prof. Alessandro De Vita

Dr. Alessandra Magistrato

Prof. Johan Wouters

PROBING PEPTIDE-NANOMATERIAL INTERFACE BY MEANS OF MOLECULAR DYNAMICS

Candidate:

Federica De Leo

Thesis supervisor:

Prof. Davide Bonifazi

Co-supervisor:

Dr. Alessandra Magistrato

JANUARY 2015

NAMUR RESEARCH COLLEGE
AND UNIVERSITY OF NAMUR

NAMUR RESEARCH COLLEGE

UNIVERSITÉ DE NAMUR

Table of Contents

List of abbreviations **VI**

Glossary **IX**

Abstract **XII**

1. Introduction **1**

1.1 Graphitic nanomaterials (GNMs) *1*

1.2 Methodological approaches towards structured materials *4*

1.3 Proteins *6*

1.4 Non-covalent interactions *7*

1.4.1 Van der Waals interactions 7

1.4.2 Hydrophobic interactions 8

1.4.3 Electrostatic interactions 8

1.4.4 π interactions 9

1.5 Hybrid fullerene *9*

1.5.1 Fullerene and amino acids 9

1.5.2 Fullerene and polypeptides 11

1.5.3 Fullerene and proteins 13

1.6 Hybrid CNT *20*

1.6.1 CNT and amino acids 20

1.6.2 CNTs and polypeptides 24

1.6.3 CNTs and proteins 30

1.7 Hybrid graphene *35*

1.7.1 Graphene and amino acids 35

1.7.2 Graphene and polypeptides 44

1.7.3 Graphene and proteins 57

1.8 Outline of the Thesis *68*

1.9 References *70*

2 Methods 80

2.1 Molecular docking 80

2.1.1 Docking 80

2.1.1.1 Genetic algorithm 80

2.1.1.2 Scoring functions 84

2.1.2 Cluster analysis 86

2.1.2.1 Hierarchical clustering 86

2.2 Molecular Dynamics 88

2.2.1 Equation of motion 90

2.2.2 Force field 91

2.2.2.1 Long range interactions 93

2.2.2.2 Ewald summation 94

2.2.2.3 Periodic boundary conditions 96

2.2.2.4 Neighbors list 97

2.2.3 MD in NPT ensemble 99

2.2.3.1 Berendsen thermostat 99

2.2.3.2 Langevin thermostat 100

2.2.3.3 Berendsen barostat 102

2.2.4 Analysis of MD trajectories 103

2.2.4.1 Root Mean Square Deviation (RMSD) 103

2.2.4.2 Root Mean Square Fluctuation (RMSF) 104

2.2.4.3 Gyration radius (Rg) 104

2.2.4.4 Radial distribution function (RDF) 105

2.2.4.5 Molecular mechanics/Generalized-Born Solvent Area (MM/GBSA) 106

2.2.4.6 Electrostatic calculations 107

2.3 Electronic structure methods 107

2.3.1 The Schrödinger equation 108

2.3.1.1 Density Functional Theory (DFT) 109

2.3.1.2 Hybrid meta exchange-correlation functional 112
2.3.1.3 Basis set approximation 115
2.4 References *118*

3. Supramolecular interactions of targeting proteins with graphitic surfaces 121

3.1 Cancer cells targeting via Antibody functionalized Fe-filled multiwalled CNTs *122*
3.1.1 Background 122
3.1.2 Experimental results 124
3.1.3 Computational challenge 128
3.1.3.1 Approximation of the model 129
3.1.3.1.1 Structure and parameters of the Cetuximab model 129
3.1.3.1.2 Structure and parameters of the CNT model 131
3.1.3.1.3 Structural properties of the CNT•Ctx conjugate 132
3.1.4 Methods sections 133
3.1.4.1 Docking studies 133
3.1.4.2 Force-field based MD simulations 134
3.1.4.3 Calculated properties 135
3.1.4.4 DFT calculations 136
3.1.5 Results and discussions 137
3.1.6 Generalization of the model 155
3.1.7 Concluding remarks 160
3.2 Design of Janus-type β-sheet for the hierarchical functionalization of graphene *160*
3.2.1 Aim of the project 160
3.2.2 Design of the interfacing targeting tool 163
3.2.2.1 Proteinic, 2D-like, Janus-type platform: the β-sheet structure 163

3.2.2.2 Graphene anchoring point 167
3.2.2.3 Targeting oriented functional groups 168
3.2.2.4 Assembling process 169
3.2.3 Computational challenge 171
3.2.4 Methods sections 175
3.2.4.1 Model systems 175
3.2.4.2 Force field based MD simulation 177
3.2.4.3 Calculated properties 177
3.2.5 Results and discussions 180
3.2.5.1 β-strand length and H-bonding pattern 180
3.2.5.2 Presence and number of pyrenyl functionalities 187
3.2.5.3 Adsorption on graphene 197
3.2.5.4 Janus-type adsorption 204
3.2.5.5 2D vs 1D: comparison with CNT 209
3.2.6 Concluding remarks and perspectives 213
3.3 References *214*

4 Structural validation of peptide-based motogenic Au surfaces by means of MD simulations 221
4.1 Introduction *221*
4.2 Aim of the project *228*
4.3 Experimental results *229*
4.4 Computational challenge *234*
4.5 Methods section *236*
4.5.1 Model approximation 236
4.5.2 Parameterization 239
4.5.3 Force field based MD simulations 239
4.5.4 Calculated properties 240
4.6 Results and discussion *241*

4.6.1 Conformational analysis of IGDQK-SH 241
4.6.2 SAMs of IGDQK-S·Au 244
4.6.3 SAM of IGDQK-S·Au / MT(PEG)4-S·Au 250
4.6.4 SAMs of IGDQK-S·Au / C8H17-S·Au 254
4.6.5 Sequence readability in the gradient SAMs 258
4.7 Concluding remarks *262*
4.8 References *264*
Conclusions and Perspectives **269**
Appendix **273**
Chapter 3 *273*
Chapter 4 *281*
Curriculum Vitae **284**

List of abbreviation

1D, 2D, 3D	Mono, Bi, Tridimensional
aa	amino acids
Ab	Antibody
AFM	Atomic Force Microscopy
Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
Au	Gold
AuNPs	Gold nanoparticles
bb	building block
C_{60}	Fullerene with 60 carbon atoms
C_8H_{17}-SH	1-octanthiol
CD	Circular Dichroism
CNTs	Carbon nanotubes
C region	Constant region
C-term	Carboxy terminal
Ctx	Cetuximab
Cys, C	Cysteine
EDS	Energy Dispersive Spectroscopy
EEL	Electrostatic Energy
EGFR	Epidermal Growth Factor Receptor
EGFR+	Model cancer cells overexpressing EGFR
EGFR-	Model cancer cells not overexpressing EGFR
FAB	Fragment Antigen Binding
Fc	Fragment crystallizable
ff	force field
GNMs	Graphitic nanomaterials
Glu, E	Glutamic acid
Gln, Q	Glutamine
Gly, G	Glycine
Hbc	Hydrogen bonding content

H-chain	Heavy chain
His, H	Histidine
Ig	Immunoglobulin
Ile, I	Isoleucine
IGDQK-SH	Isoleucine-Glycine-Aspartic Acid-Glutamine-Lysine-1-thiol-decanoic acid
L-chain	Light chain
Leu, L	Leucine
Lys, K	Lysine
mAb	Monoclonal Antibody
MD	Molecular Dynamics
Met, M	Methionine
MFH	Magnetic Fluid Hyperthermia
MRI	Magnetic Resonance Imaging
MT(PEG)$_4$-SH	α–methoxy-ω-thiol tetraethylene glycol
MWCNT	Multi Walled CNT
ns	nanosecond
N-term	Amino terminal
para	parallel
***p*-back**	pose-back
***p*-fab**	pose-fab
***p*-fc**	pose-fc
Phe, F	Phenylalanine
Pyr	pyrene
RDF	Radial Distribution Functions
Rg	Gyration Radius
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SASA	Solvent Accessible Surface Area
SEM	Scanning Electron Microscopy
Ser, S	Serine
SS	Secondary Structure
SWCNT	Single Walled CNT

TEM	Transmission Electron Microscopy
Thr, T	Threonine
Trp, W	Triptophan
Tyr, Y	Tyrosine
Val, V	Valine
vdW	van der Waals
WCA	Water Contact Angle
XRD	X-ray Diffraction
XPS	X-ray Photoelectron Spectroscopy

Glossary

Adsorption: Increase in the concentration of a substance at the interface of a condensed and a liquid or gaseous layer owing to the operation of surface forces (IUPAC definition). It is the adhesion of atoms, ions, or molecules from a gas, liquid, or dissolved solid to a surface.

Atomic force microscopy (AFM): high-resolution type of scanning probe microscopy, whose probe is the sharp tip on a cantilever used to scan the defect of a sample surface at the atomic scale. It is therefore able to image the atomic-scale topography of a surface by means of the repulsive electronic forces between the surface and the probe that moves over the sample.

Chemisorption: it is a type of adsorption that involves a chemical reaction between the surface and the adsorbate, therefore generating new chemical bonds at the adsorbant surface.

Energy-dispersive X-ray Spectroscopy (EDS): analytical technique for the elemental and chemical characterization of a sample based on the interaction between a source of X-ray excitation (focused high-energy beam of electrons) and the specimen.

Enzyme-linked immunosorbent assay (ELISA): biochemical methodology used to quantitatively measure an analyte that is immobilized on microplate surface. A specific antibody is applied over the surface and binds to the analyte. The antibody can be either linked to an enzyme that upon reaction produces a detectable signal in the substrate, or can be also biotin labeled and then upon exposition to streptavidin the resulted signal isquantified spectrophotometrically.

Magnetic fluid hyperthermia (MFH): cancer treatment that involves the injection of a fluid containing magnetic nanoparticles directly into tumors. When placed in an alternating magnetic field the nanoparticles generate heat and destroy the tumors. This minimally invasive procedure, unlike laser, microwave, and ultrasound hyperthermia,

prevents unnecessary heating of healthy tissues because only the magnetic nanoparticles absorb the magnetic field.

Magnetic resonance imaging (MRI): medical imaging technique used in radiology to investigate the anatomy and physiology of the body in both healthy and disease condition. MRI scanners use strong magnetic fields and radiowaves to image the body.

Migration: cellular migration implies the movement of cells as a consequence of directional and oriented cues that represent an external gradient. In response to chemical gradient cells undergo chemotaxis or haptotaxis, which involve the oriented movement towards increasing concentration of soluble factors and insoluble adhesive matrix components, respectively. Whereas in response to physical gradients, cell move through electrotaxis (stimulated by electric fields) and durotaxis (stimulated by mechanical forces) mechanisms.

Phase contrast confocal microscopy: optical microscopy technique that converts phase shifts in light passing through a transparent specimen to brightness changes in the image. Sample contrast comes from the interference of different path lengths of light through the sample.

Physisorption: is a process in which the electronic structure of the atom or molecule is barely perturbed upon adsorption on a surface.

Water contact angle (WCA): The contact angle is the angle, conventionally measured through a pure liquid, where a liquid/vapor interface meets a solid surface. It quantifies the wettability of a solid surface by a liquid *via* the Young equation.

Scanning electron microscopy (SEM): spectroscopic technique that produces images of a sample by scanning it with a focused beam of electrons. . The electrons interact with the atoms consisting the sample, producing signals that contain information about the sample's surface topography, composition, and other properties such as electrical conductivity.

Transmission electron microscopy (TEM): microscopy technique that operates on the same basic principles as the light microscope but uses electrons instead of light transmitted through an ultra thin specimen. TEMs use electrons as "light source" and their much lower wavelength makes it possible to get a resolution a thousand times better than with a light microscope with a resolution in the order of a few angstrom (10^{-10} m).

1 Introduction

1.1 Graphitic nanomaterials (GNMs)

After the discoveries by Kroto *et al.* in 1985,[1] Iijima in 1991,[2] and Novoloselov *et al.* in 2004,[3] the vast world of nanotechnology has not been the same anymore. By discovering, isolating and characterizing fullerenes, carbon nanotubes (CNTs) and graphene, they radically changed the story of graphitic nanomaterials (GNMs) presenting to the scientific community new allotropes of carbon with outstanding and unique chemical and physical properties. The trivalent carbon sp^2 hybridization, that these allotropes have in common results in different structures with spherical and monodimensional (0D) geometry for fullerene, tubular and one-dimensional (1D) for CNT, planar and two-dimensional (2D) for graphene (Figure 1.1a-d).

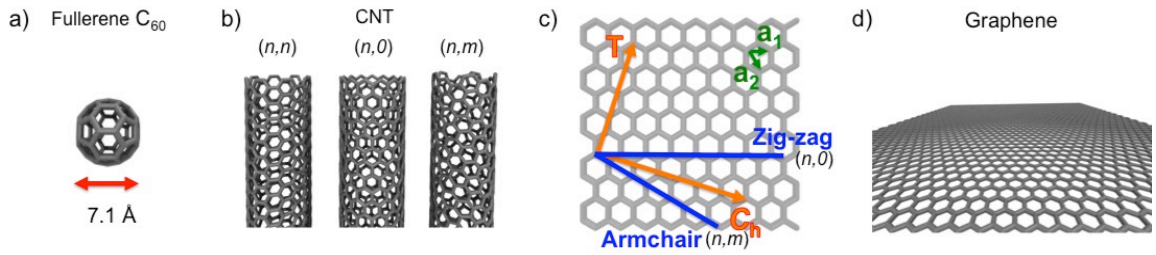


Figure 1.1. Structural representation of fullerene C_{60} (a), CNTs (b), the CNT(*n,m*) naming scheme with the vectors a_1 and a_2 and the rolling vector C_h (c). Structural representation of graphene (d).

With its 7.1 Å 0D cage (Figure 1.1a), C_{60} contains 12 pentagons and 20 hexagons that, fused in the football-like shape, conjugate all the double bonds within its geodesic structure.[4] The peculiar cage curvature is responsible of the pronounced reactivity of fullerenes, induced by the pyramidalization of the carbon atoms within the conjugated non-planar topology.[5] The established sp^2 trigonal hybridization is forced by the spherical C_{60} geometry to increase the classical angle between the axe of the π-orbitals and the three σ bonds, namely the pyramidalization angle ($\theta_p = \theta_{\sigma\pi} - 90°$), from 0 ° to 11.6 ° (Figure 1.2a-b).

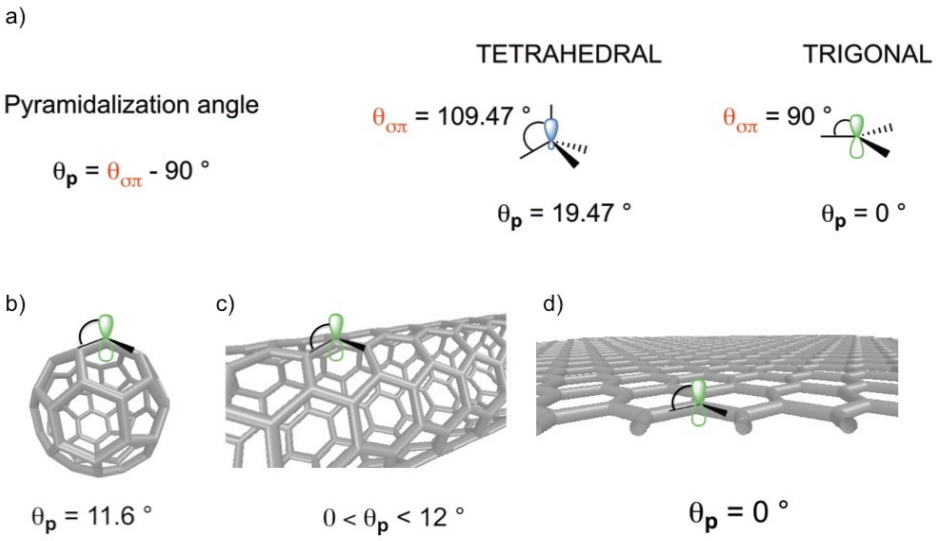


Figure 1.2. The pyramidalization angle θ_p is measured between the axe of the π-orbitals and the three σ bonds (a). Examples of θ_p values of the tetrahedral and the trigonal hybridazion are reported (a). θ_p can be a measure of the strain of the carbon atoms within C_{60} (b), CNT (c) and graphene (c) compared to the canonical angle of a trigonal hybridization.

This, together with misalignment of the π-orbitals,[5–7] explains the higher fullerene reactivity compared to the flat graphene sheet (θ_p = 0°). A low HOMO/LUMO gap, which decreases with the increase of the fullerene's cage (C_{70}, C_{72}, C_{76}, C_{80}, C_{82}, C_{84}),[8] further contributes to the fullerene reactivity. Fullerenes are therefore efficient electron-acceptor as well as radical-scavenger.[9] The presence of both five and six-membered rings introduces a more reactive site, in particular at the 6,6-junction, that represents the most reactive site for covalent derivatization[10,11] but might also represent a specific anchor point for non-covalent interactions.

For their part, CNTs can be considered as a rolled graphitic sheet and they are considered as monodimensional nanostructures due to the high length/diameter ratio, *i.e* 10^4-10^5.[12] In fact, when a single layer of graphene sheet is rolled up to form a single-walled CNT (SWCNT), its diameter can range between 0.7 to 1.4 nm, while its length can go up to several micrometres.[13,14] On the other hand, when multiple graphene sheets are simultaneously rolled up, concentric multi-walled CNTs (MWCNTs) characterized by a bigger diameter are obtained.[15] The modality in which the graphene sheet bent and close itself in a tube determines the CNT helicity, according to the orientation of the six-membered rings in the honeycomb lattice relatively to the SWCNT axe (T) (Figure 1.1c). The helicity is defined by a vector,

namely the Chiral vector (C_h), indicated by a pair of integers (n,m) and by the two unit graphene vectors (a_1, a_2) as described by the equation $C_h = na_1 + ma_2$. This equation determines the two directions in the lattice axe following which the nanomaterial can wrap in a chiral or achiral CNT. Specifically, presuming an uncapped tube, when $n = m$ the hexagon orientation will display an armchair extremity in the resulting CNT, meanwhile when $m = 0$ a zig-zag extremity is obtained for the resulting CNT; all the other cases lead to the formation of chiral CNTs (Figure 1.1b-c).[16] Armchair and zig-zag CNTs differ in their conductivity properties, being the first metallic and the second semi-conductive. Notably, by tuning methods and synthetic conditions,[17] CNTs with different chirality, diameters and length can be produced exerting networks with different chemical behaviour. Similarly to fullerenes, due to their tubular topology, CNTs present a local strain caused by the piramidalization angle and the misalignment of the π-orbitals.[18,19] However, in this case different θ_p are identified along the CNT surface. In fact, end-cupped CNTs will be characterized by a major curvature at the extremities where a resembling fullerene hemispherical geometry closes the tube and exhibit similar pyramidalization angle ($\theta_p = 12°$). The pyramidalization in that region strongly scale inversely to the diameter of the tube, therefore causing major reactivity for smaller CNTs. On the other hand, along the external sidewall the θ_p can reach the canonical value of 0 (Figure 1.2c). Here, the pronounced reactivity of the tubular sidewall is not ascribable to the pyramidalization angle, but rather to a more consistent misalignment of the π-orbitals.[20]

Finally, graphene is the planar, one atom-thick allotropic form of carbon (Figure 1.1d). Graphene is the monolayer form of the 3D multilayers graphite. It consists of a 2D sheet where the sp^2 hybridization of conjugated carbons is characterized by the established trigonal arrangement ($\theta_p = 0°$, Figure 1.2d). This makes of graphene a long-range extended pure aromatic carbon system,[21] particularly suitable for non-covalent interactions driven by van der Waals forces and π–π stacking. The lack of curvature of the planar sheet of graphene excludes the pyramidalization effect as well as the misalignment of π-orbitals, thereof their consequently strained effect. However, thermal manipulation can induce controlled curvatures and ripples formation as a source of site with an enhanced reactivity.[22] Furthermore, akin to CNTs, graphene can also exist with armchair or zig-zag edges. It has been demonstrated that the edges manifest a different reactivity of the side wall,[22] and in particular the zig-zag edges

show a more pronounced reactivity as compared to the armchair edges.[23] Additionally to the thermal, mechanical and electrical properties of the graphene sheets,[24] it is also worth mentioning the exposed double-sided surface, which can provide an extensive platform to be functionalized.

1.2 Methodological approaches towards structured materials

One of the recent and fascinating avenues to confer a new functional essence to GNMs through chemical modification is represented by the idea of merging nanotechnology and biology. By designing biomimetic nanomaterials, the so-called nano-biotechnology opens great opportunities towards the use of graphitic structures in combination with biomolecules, *i.e.* peptide, proteins, enzymes or nucleic acids.

By this way, the biomolecular modification would increase the poor solubility and low biocompatibility of these graphitic materials, leading to a resulting biohybrid that, gained the new functional property, would make available a potpourri of properties exploitable for a large variety of applications ranging from imaging[25] and bio-sensing,[26–28] nanomedicine and cancer therapy,[29] gene therapy, drug delivery[30–32] and tissue engineering.[33] Aware of the concerns about the toxicological aspects of graphitic nanomaterials, the reader is redirected to reviews that broadly tackle this topic.[34–36]

One of the main aspects to take into account in the view of using and exploring GNMs potentiality is their insolubility as well as the difficulties in separating them. Regardless of their structural and geometrical differences, due to their intrinsic hydrophobic nature, all the GNMs tend to form insoluble aggregates that stack in bundles. Therefore, by exploiting the knowledge in the chemical reactivity of the graphitic frameworks, substantial efforts have been carried out to modify their external surface in order to increase their solubility in organic solvents as well as in water,[37,38] thus implementing their practical manipulation and separation. Up to now, many reports support the successful functionalization of GNMs that can be achieved by a twofold approach: either by covalent functionalization taking advantage of the external wall reactivity or surface defections, or by a non-covalent approach relying on weak interactions between the nanomaterial and the chemical modification.[37,39–44] The covalent functionalization implies the irreversible disruption of the conjugated

aromatic network, thereby modifying the structural and conductive properties of the nanomaterial. On the other hand, the non-covalent approach can be reversible and in addition preserves the structural and electronic aspects of the GNMs.

Among the different biomolecular functionalizations (with proteins, enzymes, polysaccharides, lipids, DNA, RNA), the exponential interest in combining GNMs and proteins, as main exponents of biomolecular repertoire, requires a deep understanding of their way of interaction. In fact, the deep knowledge of the free and spontaneous tendency of GNMs to interact with proteins, intrinsic in their chemical nature, results of pivotal importance for the achievement of a contrived and conscious human manipulation towards tailored applications. The full understanding of the processes occurring at their interface is therefore the central key point to proficiently handle the two entities reaching a smart ensemble or avoiding a toxic consequence. That is, the functional aspect of the designed nanomaterial is directly linked to the weak interactions involved in the protein adsorption and hybrid formation.

While examples, potentialities, prospects and challenges about hybrid carbon-based materials in nano-biotechnology are extensively reported in several recent reviews,[45–50] here we want to focus on the understanding of the main non-covalent interactions governing the interface between GNMs and proteins. Hence, in the present Chapter, only the behaviour of non-functionalized GNMs will be considered to figure out the modalities through which they spontaneously undergo association and bind biomolecules. The provided picture of the non-covalent interactions between pristine GNMs and proteins can serve both as reference to the exploitable aspects for engineering new materials for applications in the biomedical field as well as a guide to predict parasite adsorbing processes that would hit at toxicological implications. Aimed at understanding in depth the forces ruling the adsorption of proteins onto GNMs surface, a systematic review of the experimental and theoretical studies carried out on non-functionalized fullerenes, CNTs and graphene with amino acids (aa), polypeptide and proteins will be reported separately in the following sections. This will provide with an exhaustive picture, both at the molecular (aa) and macroscopic (proteins) level, of the specific driving forces governing the GNMs/proteins adsorption process.

Herein, the reader is thus guided through a systematic overview on the most representative examples of non-covalent GNMs/proteinic biomolecules conjugates, whereas numerous reports dedicated to the covalent functionalization of the

buckyball,[39,40,51–54] CNT[55–57] and graphene[58–61] and their biological applications[31,49,58,62–69] can be found elsewhere.

1.3 Proteins

To understand in depth the propensity of GNMs to interact with proteins, a brief description of the latter is required.

As a general definition, aa are the proteinogenic building blocks that, combining through peptides bonds, form linear sequences, known as polypeptides. The 20 natural aa have all in common a central α-carbon to which is attached an amino ($-NH_3$) and a carboxylate ($-COO^-$) group, a hydrogen and a variable side chain. According to the nature of the latter, aa are classified in hydrophobic, sulfur-containing, aromatic, positively and negatively charged, polar and cyclic residues (Figure 1.3).

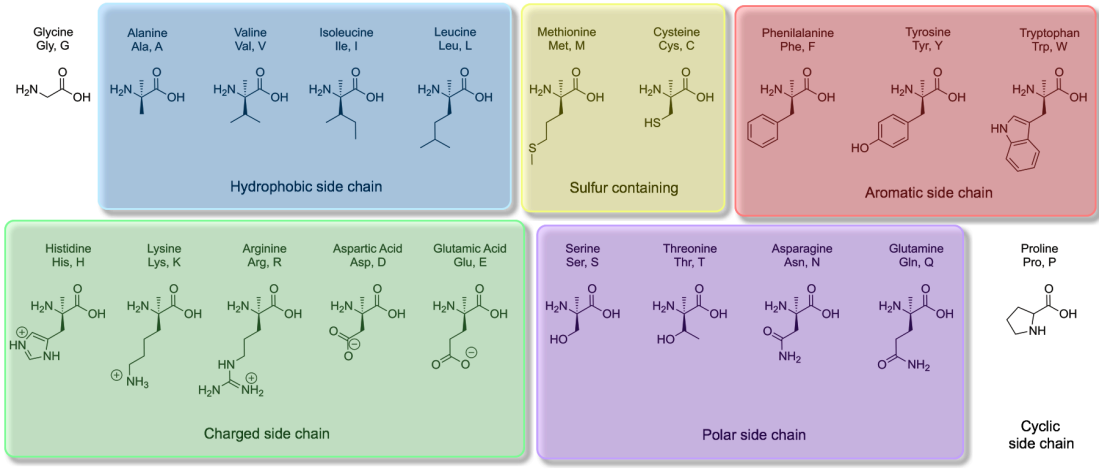


Figure 1.3. The structures of the 20 natural amino acids (aa) are also reported and classified according to the nature of the side chains, namely hydrophobic, sulfur-containing, aromatic, charged, polar and cyclic. For each aa the common nomenclature, the full name, the three and one letter abbreviation also are reported.

Polypeptide sequences, which represent the *primary structure*, can fold into specific 3D structure taken together by weak forces (H-bonds, van der Waals, ionic interactions, hydrophobic packing). In particular, the 3D folding path of the polypeptides backbone represents protein *secondary structure*, such as α-helix and β-sheet which are stabilized by non-covalent interactions occurring between both side

chains (vdW interactions) and the backbone atoms (Hydrogen-bonds, H-bonds).[70] Additionally hydrophobic effect can influence the folding process by burying in the internal core the hydrophobic residues present in the primary sequence, while exposing the hydrophilic ones.[71] By this way, water-soluble macromolecules can be produced along with hydrophobic pockets often involved in the protein activity. The intermolecular arrangements among secondary structure elements define *tertiary structure* of proteins. Finally, the 3D organization of several protein molecules represents the *quaternary structure*. Importantly, the ability of a protein to exert a biological function is directly connected to the correct folding in a specific tertiary and quaternary structure.

1.4 Non-covalent interactions

The possible non-covalent interactions that might come into play in the adsorption of proteins and GNMs are here briefly summarized.

1.4.1 Van der Waals interactions

Van der waals (vdW) interactions are the weakest electrostatic intermolecular attractive forces between molecules with a permanent or induced dipole that get close to each other with a distance equal to the sum of the vdW radii of the atom pairs.[72] The convergence towards this ideal distance (r_0) defines the attractive forces between two atoms that found in r_0 the minimum beyond which repulsive forces prevail and the energy increases. This has the form of the Lennard-Jones potential. The range of distance where vdW interactions are verified stays between 0-10 nm.[34]

The weak nature of this interatomic interaction may become considerably important when multiple atom pairs are interacting with each other and a cooperative effect comes into play. In the view of maximizing the contact surface between bigger systems, the shape complementarity, degree of conformational freedom and size of the two interactors clearly contribute in enhancing the strength of the vdW energy.

1.4.2 Hydrophobic interactions

The hydrophobic interactions refer to insolubility of non-polar substance in water that is described by two driving forces relevant to their mixing. The entropic effect induces the clustering of hydrophobic molecules to reduce the contact surface exposed to the ordered water molecules. On the other hand, mixing non-polar molecules in water is opposed by the enthalpies of interaction; ordinarily, dispersion forces leading to water-solute attractions are smaller than the strong H-bonding interaction (high cohesive energy) of water and those leading to the corresponding solute attraction.[73–75] This kind of interaction clearly involves hydrophobic surfaces and moieties, whose dynamics and features can be determined by these forces.

Concerning the GNMs, their full hydrophobic nature leads them to strongly clusterize in bundles, therefore making them highly insoluble in most of the organic solvents. On the other hand, hydrophobic forces are also involved in the folding process of proteins inducing the formation of arranged and complex tertiary structures that most likely expose externally the hydrophilic residues, allowing for a better solubility, while the active sites are often characterized by narrow and hidden hydrophobic pockets.

1.4.3 Electrostatic interactions

The electrostatic interactions involve molecules with an opposite charge that tend to attract each other.[76] For example within protein structures, the formation of salt bridges can occur between residues bearing a positive and negative charge. Similarly, the electron rich π system can interact with a cationic/anionic moiety in the so-called cation/anion···π interactions. Hence, considering the positively charged aa approaching the electron rich conjugated surface of GNMs, the aforementioned interactions can certainly play a role in stabilizing the protein structure as well as the adsorption pathway onto the graphitic surfaces.

1.4.4 π interactions

Pivotal role, especially in biological recognition, is played by π interactions.[77] The electron rich cloud of conjugated aromatic rings can entertain metal, polar, anion, cation, CH···π interactions or another π systems can lead to the formation of π-π stacking.[78–81] Concerning the latter, two main conformations in the arrangement of aromatic groups can be envisaged: parallel-displaced or T-shape (edge-to-face) orientations. A delicate balance between vdW forces and quadrupole contribution defines the major stability of one orientation rather than the other.

The average distance in the π-π stacking is generally larger than the vdW radii. Considering the protein/GNMs way of approaching, these kind of interactions can be intended as specific since involve only aromatic aa, which are expected to stack parallel to the surfaces.

1.5 Hybrid fullerene

1.5.1 Fullerene and amino acids

We start analyzing the behavior of C_{60} in complex with the 20 natural aa.

de Leon *et al.*[82] reported in 2008 a computational study on the interaction of C_{60} with the 20 natural aa performing density functional theory (DFT) calculations at the B3LYP/3-21 G* level of theory. The aa were classified according to the nature of their side chains and analyzed considering the interacting sites as well as the geometric and energetic properties. It results that all the 20 aa form favorable complexes with C_{60}, with the Leu, Arg, Trp and Glu units featuring the highest calculated dissociation energies (ΔE = -3.91, -3.36, -2.36 and -2.36 kcal mol^{-1} respectively). Despite the presence of different functional groups, the Leu, Arg, Trp and Glu units commonly feature hydrophobic alkyl chains that, thanks to their conformational flexibility, can embrace the spherical surface of C_{60} further stabilizing the complex (Figure 1.4). Considering that different alkyl chains can be introduced, one might use different chains to tune the embracing and curvature can be in principle selectively complexed. This has been further demonstrated by de Leon *et al.*, who modeled the adsorption of aa on C_{80},[83] demonstrating the formation of significantly stronger complexes compared to those formed with C_{60}. Notably, the sulfur-

containing Met and Cys aa featured higher interaction energies (ΔE = -12 and -11.45 kcal mol^{-1}, respectively). Beside vdW and C-H···π interactions (α-CH···C_{80} distance of 2.07 and 2.91 Å), it appears that the low HOMO/LUMO gap of C_{80} favor sulfur-fullerene surface (S···π) interactions. Also, Lys, Tyr and Trp residues were found to form strong complexes, with the Trp term tangentially lying on the fullerene surface. Notably, the contact between each aa and the C_{80} cage is occurring through the amino acidic α-CH proton, rather than with the β-CH, probably prevailing over the hydrophobic contribution of the alkyl side chain.

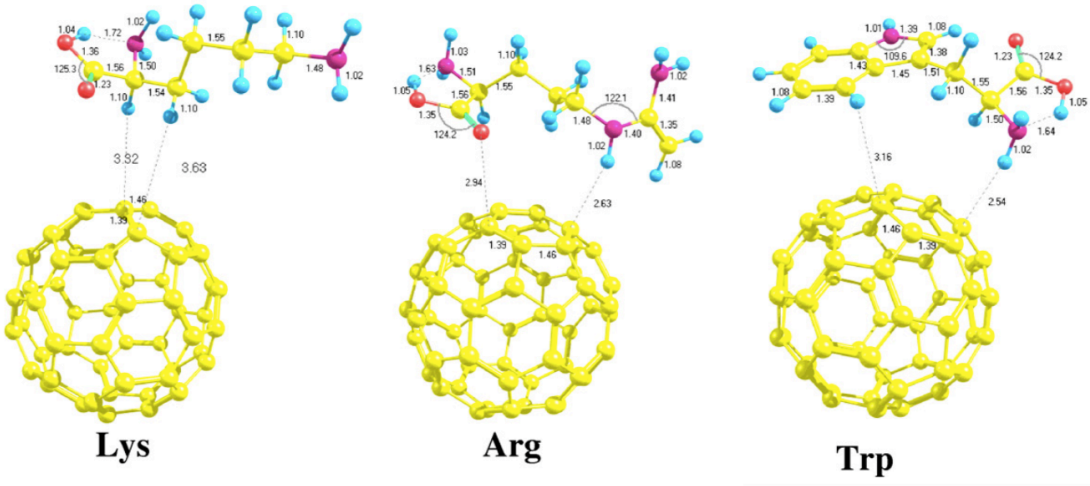


Figure 1.4. Optimized structure of the three most interacting aa with C_{60}. Also some geometrical parameters are reported, namely bond lengths (Å) and angle (°). Adapted with permission from ref. 81. Copyrights 2008, Elsevier.

In summary, by applying DFT methods to estimate the interaction of the 20 natural aa and different size fullerenes, the authors demonstrated a general stability of the complex formed with C_{60} mainly ruled by Lys and Arg residues that, bearing long alkyl chains, can establish shape-adaptable hydrophobic interactions.[82] At the same time, bigger the cage, lower is the fullerene HOMO/LUMO gap, thus favoring other types of interactions towards the formation of more stable complexes, in particular with aa bearing electron-donating moieties like Cys and Met.[83]

1.5.2 Fullerene and polypeptides

As discussed above, the 20 natural aa can differently adsorb onto the fullerenes' surface, thus suggesting that the interaction between a fullerene molecule and proteins' primary structure strongly depends on the polypeptide sequence. This can have significant repercussions on the proteins folding pattern, and thus on its functionality. For instance, it has been demonstrated that fullerene derivatives inhibit amyloid fibrillation through tightly binding the central hydrophobic motif (KLVFF) of the β-amyloid peptide (Ab).[84] The strong hydrophobic-driven binding of C_{60} to Ab prevents it from polymerizing and aggregating into the typical β-sheet secondary structure. Specifically, it has been found that pristine C_{60} inhibit Ab polymerization with an impressive median inhibitory constant (IC_{50}) of 9 μM, 15-fold higher than the control experiments carried out on other known potent inhibitors described in the literature.[84] Following these evidences, Xie *et al.* proposed a molecular mechanism for the C_{60}-ihnibition Ab fibrillation by reporting their replica exchange MD (REMD) simulations, whose reliability was further validated by AFM measurements.[85] They simulated the octamer of the Ab(16-22) fragment, consisting of the hydrophobic motif KLVFFAE, alone or in the presence of C_{60}, a cluster of three C_{60} ($3C_{60}$) or C_{180} molecules and measured the structural and energetic parameters quantifying the inhibition effect on the fibrillation. It results that the inhibition effect was due to the ability of the cage to intercalate into the β-barrels[86] with a concentration-dependent trend. Notably, the presence of three C_{60} molecules causes a drastic reduction of the β-sheet content (from 44.5 to 25.7% for Ab and Ab+ $3C_{60}$, respectively), consistently higher than that induced by one C_{60} (45.2 %), as also appreciable by the structural representation of the most-populated clusters shown in Figure 1.5a-b. This was nicely confirmed by AFM images that show a strong diminution of the fibrils elongation and thickness upon increasing the C_{60} concentration (Figure 1.6). Whilst bearing the same number of atoms of $3C_{60}$ with a reduced surface area, C_{180} displays higher inhibition effect as detected by a lower β-sheet content (18.1%) and increased coils formation (Figure 1.5c). An in-depth analysis revealed that the inhibition effect was ascribed to the ability of the cages to compete for π-stacking interactions normally occurring between the aromatic side chains, which are responsible of the fibril compactness. In fact the aromatic Phe moieties are now engaged in π-π staking interactions with the

hexagonal fullerenes' regions, thus losing the stabilizing effect on the fibrils that these interpeptide interactions induce.[85]

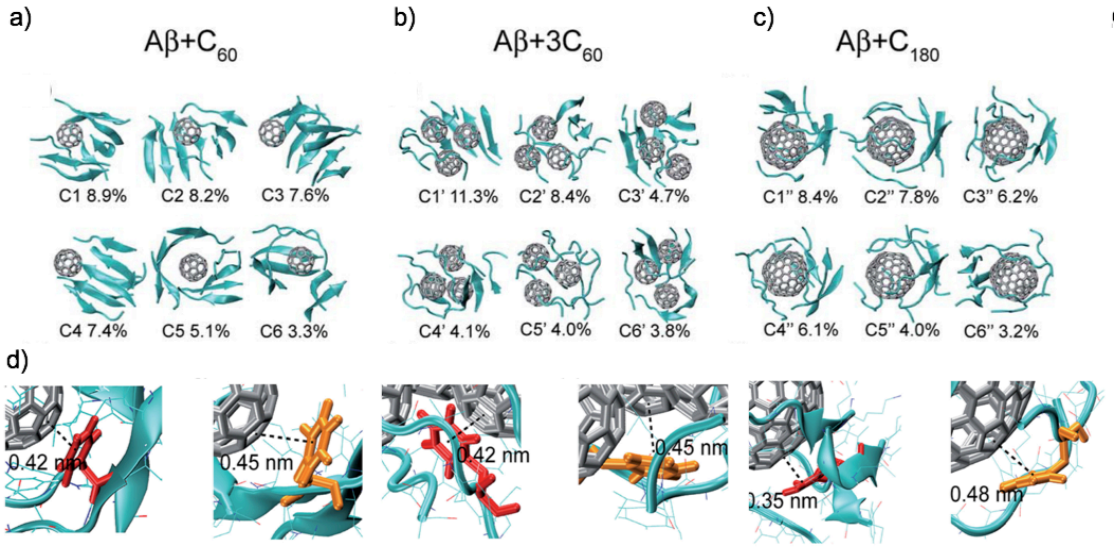


Figure 1.5. Representative structures of the top six most-populated clusters (C1-C6) and their % of occupancy for the systems Aβ+ C_{60} (a) Aβ+ $3C_{60}$ (b) Aβ+ $3C_{180}$ (c). The β-sheet secondary structure is denoted with the cyan arrows, random coils are represented otherwise. Analysis of aromatic stacking interactions is shown displaying in red the parallel-aligned hexagons-Phe interaction, in orange the pentagons-Phe, the distance between the centroids are also reported (d). Adapted with permission of ref. 84. Copyright 2014, Royal Society of Chemistry.

Consistently, this hypothesis was in line with other simulations reported in the literature.[86] Noteworthy, the smaller curvature of C_{180} allows for closer parallel-aligned aromatic-stacking interactions (Figure 1.5d) as compared to C_{60} (0.35 and 0.42 nm, respectively), which corresponds to higher binding free energies (-217 *vs* -193 kcal mol^{-1}, respectively). Conversely, for all the simulated systems the π-stacking interactions between aromatic aa and the pentagonal rings of the cages feature a staggered stacking at a distance of 0.45 nm (Figure 1.5d), demonstrating the preferential and predominant role of the fullerene hexagons in these interactions. The active research on the inhibiting effect of fullerene on the amyloid formation denotes its importance as well as the potential role of GNMs in the field. For a comprehensive account, the interested reader is redirected to the topical reviews.[46,87]

Beside the chemical nature of aa sequence, high conformational freedom characterizes polypeptides before reaching the stable folded state. In this regard, the

reader is addressed to the section 1.7.2, where some comparative studies between all the GNMs and different polypeptides are highlighted as explicative examples.

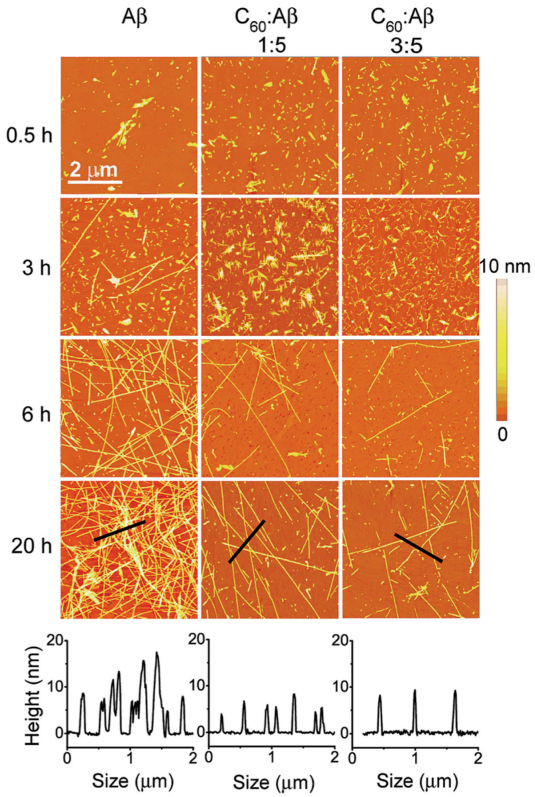


Figure 1.6. AFM images of Aβ, Aβ+C_{60} and Aβ+ $3C_{60}$ at four incubation times, and height analysis at 20h incubation time (bottom). Reprinted with permission of ref. 84. Copyright 2014, Royal Society of Chemistry.

1.5.3 Fullerene and proteins

More than three decades ago with the pioneer work of Friedman *et al.* it was found out that the peculiar C_{60} structure presents the ideal and suitable size and shape for the formation of stable complexes with proteins. They presented, aided by a molecular model and experimental binding essays, the surprising ability of a fullerene derivative to inhibit the HIV-1 protease.[88] The protein active site that, apart from the catalytic acidic residues, is characterized exclusively by hydrophobic aa, perfectly accommodates the fullerene derivative within its pocket consecrating an efficient antiviral activity of the modified nanomaterial.[54,64] Kinetic analysis characterized experimentally the binding revealing a competitive behavior with inhibition constant (K_i) of 5.3 μM,[89] and a dissociation constant (K_d) in the order approximately of 10^{-6}-

10^{-9} M. The molecular model also validated the binding mode reporting the reduced solvent-exposed surface area before and after the binding thus showing that the 92% of the desolvated hydrophobic surface was due to C_{60} binding.

In line with the aforementioned work, in 1998, the group of Erlanger identified by binding studies several anti-fullerene antibody (Ab), a subpopulation of which was also able to selectively discern C_{70} over C_{60}.[90] Aided by docking calculations, a combination of hydrophobicity matching, π-stacking and a combining site-fit between the Ab fragment antigen-binding region (Fab) and the geodesic structure was hypothesized to drive the mutual recognition. Two years later, Erlanger and co-workers reported the crystal structure of the Fab of the first selective monoclonal Ab (mAb) against C_{60},[91] confirming the above-mentioned binding hypothesis. From one side, they managed to quantify experimentally a host-guest specific recognition with a binding affinity of 22 nM, while, due to the difficulties in co-crystalizing the Fab with the C_{60} cage, a molecular model of their complex complements the experimental evidences. It clearly showed that the nanomaterial, resembling the native Ab-antigen binding, establishes π-stacking interactions with aromatic aa (Trp, Tyr, Phe), while a weak OH···π hydrogen bond between the Tyr side chain and the fullerene surface was also observed (Figure 1.7a). Additionally, Asn and Gln are also found parallel to the curved cage, framing the nanomaterial in a perfect embrace within the mAb Fab as shown in Figure 1.7a.

Similarly, Noon et al. displayed the net predominance of π-stacking of Phe, Tyr and Trp residues in the stabilization of C_{60}-Ab complex validated by MD simulations.[92] Of particular interest is the triple π-stack taking place among the ball, the indole ring of Trp and the phenol group of a Tyr residue with a preferential localization over the 6,6-junction of the C_{60} (Figure 1.7b). The predicted displacement of the aromatic stacking was indeed consistent with the reported X-ray crystal structure of benzene-solvated C_{60} where the benzene ring is clearly located over the 6,6-junction of the carbon sphere (Figure 1.7c).[93] Along the dynamics of the C_{60}-Ab complex, CH···π interactions are also observed completing a general picture of dense network of π-electron rich connections (Figure 1.7d-e).

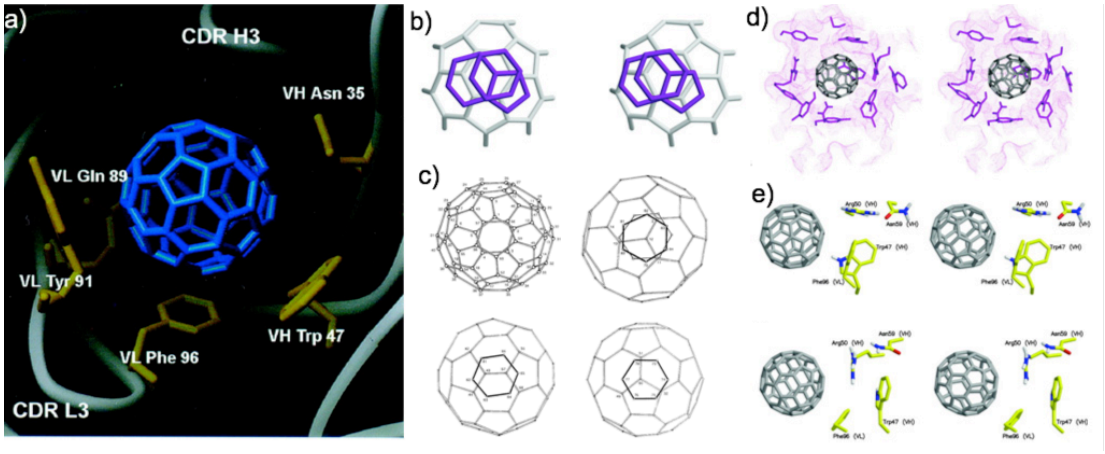


Figure 1.7. The model binding of C_{60}–Fab complex reported by Erlanger and co-workers shows the fullerene ball in blue and the residues involved in the binding in yellow (a). The triple π-stack among C_{60}, indole ring and phenol ring are shown in (b), localized in the 6,6-junction as also observed in the crystal structure of benzene solvated fullerene (c). The binding mode of C_{60}–Fab proposed by Noon *et al.* (d) and some snapshots of different configurations of the interaction where the main interacting residues are highlighted in yellow (e). Adapted with permission from refs. 90 and 91. Copyrights 2000 and 2002, National Academy of Science. Adapted with permission from ref. 92. Copyright 1969, Royal Society of Chemistry.

The aforementioned cases were the first reporting the interesting ability of C_{60} to perfectly accommodate within a protein active sites (HIV-1 protease[88] and mAb[90–92]), to tightly bind it through π-stacking interactions with aromatic aa (Trp, Tyr, Phe) and H···π contacts. By this way, the activity of the interacting proteins was demonstrated to be dependent on the binding of the fullerene ball.

Afterwards, modulating effect of C_{60} on protein activity, in particular its antiviral inhibition, was thoroughly explored as the extensive theoretical and synthetic efforts give testimony.[54,63,94–98] Among the others, many fullerene derivatives were prepared reporting their binding to several proteins, such as human and bovine serum albumin (HSA and BSA),[99,100] voltage-gated K^+ channel,[101] lysozyme,[102] glutathione S-transferase (GST)[103] and reductase (GSR),[104] and glutamate receptor.[105]

The targeting aptitude of the C_{60} towards proteins was further investigated by Zerbetto and co-workers by docking calculations, quantifying the baiting protein tendency of fullerene. In a remarkable work, they applied a reverse ligand-protein docking protocol, by which instead of testing the ability of a unique protein to bind several ligands they screened the tendency of the Protein Drug Target Database

(PDTD), counting 1200 structures, to bind to a unique ligand, *i.e.* C_{60}.[106] This was aimed at identifying new potential C_{60} binding targets by validating the protocol with the available experimental data. The docking algorithm chosen includes additional binding scores that account for all the possible weak forces, namely the vdW interactions, partial electrostatics, explicit hydrogen and disulfide bonds contribution, π-stacking and cation-π interactions. This allowed for a thorough quantitative estimation of the interaction between C_{60} and the surface of all the available proteins in the database. The top 10% of the most binding proteins was selected and only a small percentage (20%) of the known fullerene binding proteins was discarded by the program as false negative, thereby validating the predictive power of the analysis. Concerning the abovementioned proteins already known to experimentally interact with C_{60}, this protocol provides with important structural clues that give further explanation of the binding activity. At the same time, with this exhaustive docking procedure, Zerbetto and co-worker drafted a comprehensive list, equipped with structural information, of new binding proteins that might potentially interact with the carbon cage.

For instance among the know proteins, they were able to identify the preferential binding site of C_{60} on GST and GSR, ascribing to a mixture of Ser, Phe, Asp, Ala, Tyr, Thr, Leu aa, located in the active site, the ability of C_{60} to bind and exert its inhibition activity on these proteins. The previously reported noncompetitive activity of C_{60} derivatives on Acetylcholinesterase (AChE)[107] finds now an explanation in the preferential binding of fullerene to a hydrophobic cavity, other than the active site; the allosteric inhibition of cytochrome P450 by the carbon ball is unraveled by considering these docking results that find fullerene too bulky to fit in the active site; the binding to ATPase was now clarified by an allosteric behavior of fullerene that, overlooking the ATP binding pocket, finds a cavity that impedes protein rotational and conformational changes. The first ranked protein was the voltage gated K^+ channel that, according to the docking algorithm, was plugged at the level of the chamber rather than at the entry region as previously postulated.[101] This is consistent with consecutive MD studies that identified a variety of binding sites within K^+ channel that C_{60}, passing through the hydrophilic mouth of the channel, occupies reaching an hydrophobic surface formed by an interhelical domain.[108]

To mention few among the unknown binding proteins, they show that the peculiar hydrophobic structure of the steroid hormones was nicely resembled by C_{60} when tightly binds different glucocorticoid and glucocorticoid-like receptors. The fullerene cage naturally competes for the steroids binding site that, rich in Trp, Phe and Tyr aa, easily lend itself for hydrophobic interactions. Furthermore, another binding pocket, particularly prone to hydrophobic interactions, was identified by fullerene on the Nitrile oxide synthase.

Hence, Zerbetto and co-worker shed light on the incredible affinity between a vast protein database and C_{60} and presented a peculiar and significant trend: the fullerene cage displays the outstanding ability of scanning a protein surface and identify the ideal pocket to accommodate and establish, through weak interactions, strong complexes.[106] The identified pockets, non-necessarily corresponding to the active site, have all in common aromatic, hydrophobic or charged residues that can entertain π-stacking and cation-π interactions. Among the 1200 proteins, the climb of the ranking, deciphered as the major or minor affinity to the C_{60}, is therefore achieved when the aforementioned interactions are maximized and the protein-fullerene surface complementarity leads to the perfect match.

Up to now it is hard to talk about specificity, especially when the driving forces are vdW and hydrophobic interactions that can come into play whenever aromatic or hydrophobic cavities can accommodate the C_{60}. But this is not all. The energetic gain derived from the solvophobic interactions, *i.e.* the desolvation free energy released by solubilizing the GNM with biomolecules, clearly plays a determinant role.[109] The latter is further maximized when a complete wrapping around the geodesic structure is achieved.[110] Therefore the appropriate size of the hosting cavity may represent the point of specificity in the binding with fullerenes.

In a following work, Zerbetto and co-worker applied the same reverse docking protocol on the PDTD to explore the possibility of accomplish the selective and specific separation of different size cages (C_{60} and C_{70}).[111] By this way, they were able to classify the 1000 protein structures into 4 categories able to discern between C_{60} and C_{70} (p-C_{60} and p-C_{70}) interacting with them in the same (homosaccic protein) or in a different pocket (heterosaccic proteins). Clearly, the cage size and the protein cleft volume have to match perfectly for the accomplishment of the specificity.

For instance Inosine-5-Monophosphate Dehydrogenase (IMPDH) results a p-C_{60} homosaccic protein, where C_{60} snuggles excellently in the binding site while C_{70} sterically clashes inside, lowering the resulting scoring function. On the other hand, the KcsA Potassium Channel is labelled here as p-C_{70} homosaccic protein due to the major number of contacts that the bigger cage can entertain within the channel pore, while the smaller fullerene is not able to fill completely the pocket (Figure 1.8).

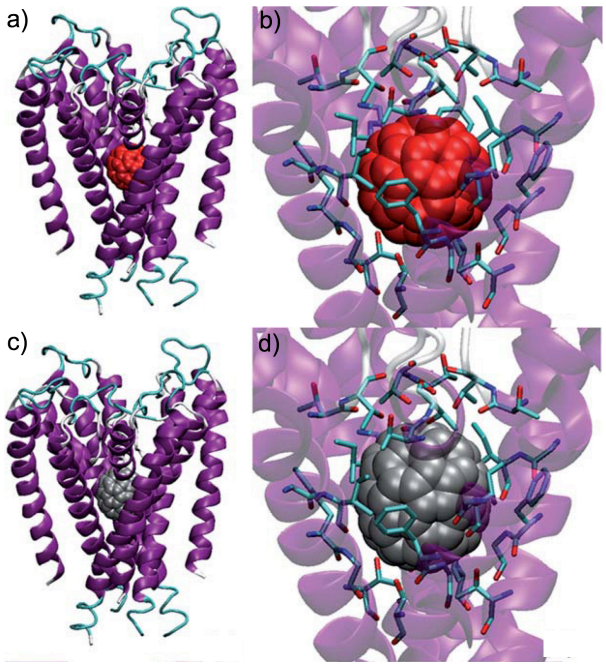


Figure 1.8. Docked complex of KcsA Potassium Channel and C_{60} (a-b) and C_{70} (c-d). Adapted from ref. 110 with permission. Copyright 2011 Royal Society of Chemistry.

From the exhaustive screenings performed by Zerbetto and co-worker, it was nicely confirmed that proteins easily accommodate fullerene cages within their hydrophobic clefts. Hydrophobic and aromatic aa play the main role in stabilizing the binding through π-π stacking interactions.[111] Furthermore, according to the nest's size, proteins are able to selectively bind the fullerene cage whose size satisfies the perfect protein-GNMs shape-complementarity. The highlighted relationship between volume fitting and binding capabilities of fullerenes interacting with proteins remind to the 55% rule, which states that the ideal binding can be expected when the ratio between the guest and host volume is in the range of 55 ± 0.9 %.[112] The peculiar hydrophobic and vdW nature of complexation that characterize fullerene-proteins

interaction are indeed the ideal conditions in which the 55% rule holds, prompting that might efficiently describe this hybrid complexation.

Noteworthy, a very recent work by Calvaresi *et al.* showed the identification and characterization at the atomic level of details of the binding pocket where a 1:1 adduct between lysozyme and C_{60} is formed.[113] They were able to distinctively determine the adsorption of fullerene onto the protein surface by UV-Vis measurements, and the imperceptible variation in the circular dichroism (CD) plot demonstrates the structural stability of the protein upon binding. NMR chemical shift perturbation analysis in combination with docking calculations localize the fullerene in one of the six subsites of the endogenous ligand, with the direct involvement of two aromatic residues (Trp108 and Trp62, Figure 1.9a) and the catalytic aa (Glu35 and Asp52). The presence of the graphitic ball in this well-defined location was further demonstrated by the quenching of the fluorescent spectra of the Trp aa (Figure 1.9b). Notably, the identified binding site, pluri-validated experimentally and theoretically, is other than the one observed for CNT,[114] suggesting a controlled sorting process of different GNMs.

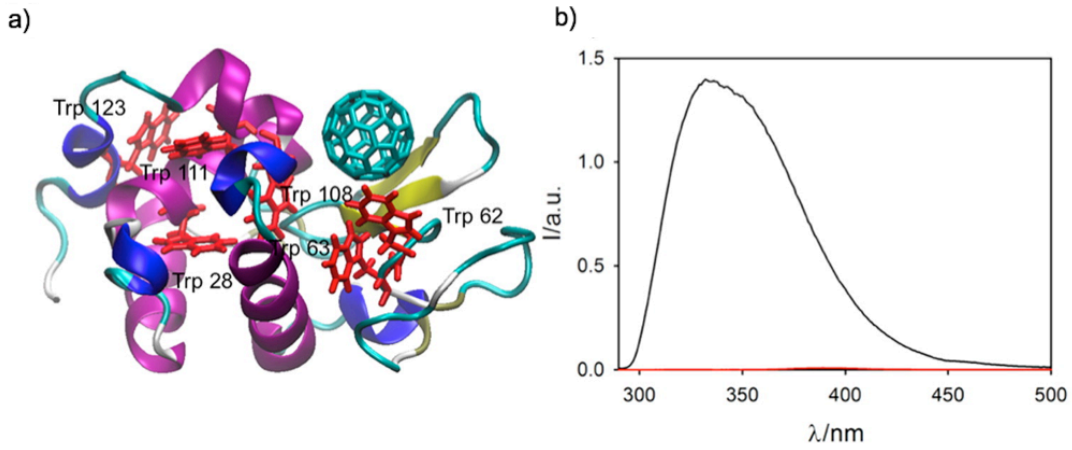


Figure 1.9. Trp residues are colored in red in the docked complex of lysozyme and C_{60} (a). The Steady-state fluorescence of monomeric lysozyme and C_{60}–lysozyme complex is plotted with black and red lines, respectively (b). Reprinted with permission from ref. 112. Copyright 2014, American Chemical Society.

In summary, the interaction between fullerenes, C_{60} in particular, and proteins is mainly driven by a shape-complementarity and the aa exposed in the protein accommodating site. In fact, the proteins active site rich in aromatic aa are

particularly prone to interact with fullerene cages thanks to ability to form π-π stacking. Additionally, charged aa, found in several binding site, are also discovered to stabilize the interaction with the buckyball.

1.6 Hybrid CNT

1.6.1 CNT and amino acids

As observed for the fullerene-aa interaction in section 1.5.1, the quantum mechanical (QM) calculations are particularly suitable for a detailed estimation of the nature and relative energy of the interaction ruling the adsorption between CNTs and aa. de Leon *et al.* extended the systematic study of the exohedral adsorption of the 20 natural aa[82,83] also on CNTs. In particular, the complex stability of the natural aa with both armchair (5,5)[115] and zig-zag (10,0) CNTs[116] treating the systems with the BLYP functional and the double-numerical plus diffusion (DND) basis set (nomenclature of the D-Mol^3 software that is comparable to the Gaussian 6-31G* basis set) has been evaluated. Their results were validated by comparison with different DFT and local approximation (LDA) methods confirming the effectiveness of the chosen protocol. Surprisingly, the same stability order as that calculated for the fullerenes was also observed for the CNTs, with the Arg, Cys and Ala residues strongly interacting with the armchair CNTs (Figure 1.10), and the Arg, Cys, and Lys with the zig-zag CNTs.

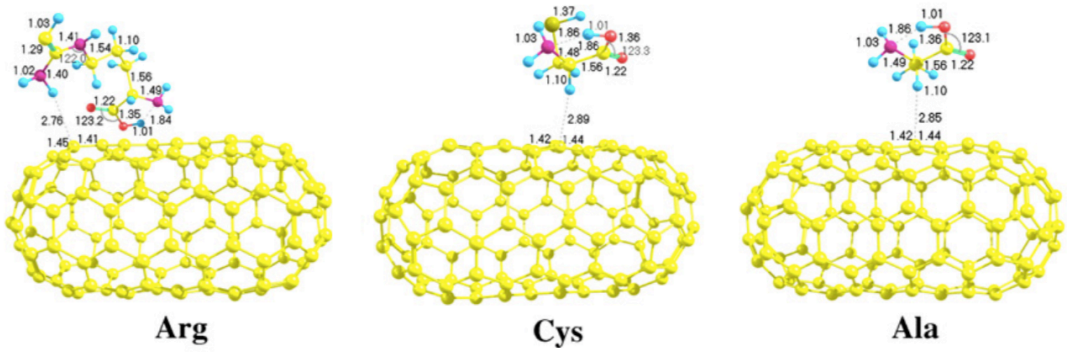


Figure 1.10. Optimized structure of the three most interacting aa with armchair CNT. Also some geometrical parameters are reported, namely bond lengths (Å) and bond angle (°). Adapted with permission from ref. 114. Copyright 2008, Elsevier.

Interestingly, the positively charged Arg displayed the highest ΔE (-10.8 and -7.6 kcal mol^{-1} for armchair and zig-zag CNTs, respectively), this resulting from the hydrophobic interactions of the alkyl chain and the polar guanidinium group, whereas a specific thiol-CNT interaction is responsible of the notable stability of the Cys interaction with the graphitic surfaces (ΔE = -6.8 and -5.9 kcal mol^{-1} for armchair and zig-zag CNTs, respectively). Finally, Ala and Lys are noteworthy residues favoring hydrophobic association. Furthermore, they generally reported a lowered complex HOMO/LUMO gap that, together with the surface contacts, contribute to the aa/CNTs stability. It also emerged a major stability of the complexes formed with armchair CNT that the author ascribed to the enhanced density of state (DOS) near to the Fermi level of the metallic more than the semi-conductive case.[115,116]

Hence the authors described a significant influence of charged and sulfur-containing aa in the interaction with both armchair and zig-zag CNTs. However, it is worth mentioning that important experimental evidences demonstrated that peptide sequences rich in Trp and His directly attach to the CNT surface showing strong adsorption, as it has been identified from a phage display analysis.[117] Indeed, the absence of the aromatic residues among the best-ranked aa in the works by de Leon *et al.* is ascribable to the level of theory used (BLYP/DND). Despite the fact that classical DFT calculations represent a good compromise in terms of computational time-accuracy, however owing to the electron density approximation, conventional DFT methods basically ignore the dispersion attractions, thus being deficient in describing vdW and π-π stacking interactions. As a consequence the BLYP/DND level of theory used to study the aa-CNT interaction might not describe properly the behavior of aromatic aa with different CNTs, while was able to capture the significant contribution of positively charged and sulfur-containing aa thanks to the use of the polarized basis set.

By means of classical MD simulations, Az'hari *et al.* studied the adsorption of the 20 natural aa evaluating the SWCNTs' structural properties, like the chirality, length and diameter.[118] They performed 5 ns MD simulations (OPLS/AA force field and TIP3P water model) in explicit water solvent modeling the natural aa on eighteen distinct CNTs, different in their chirality and length (chiral 4,1 and 5,2; zig-zag 5,0 and 7,0; armchair 3,3 and 4,4; each of them were simulated 2 and 4 nm in length) or chirality and diameter (chiral 8,2 and 10,5; zig-zag 10,0 and 14,0, armchair 5,5 and

8,8). According to their results, the aromatic aa were now found to be the most stabilizing moieties with a tangential orientation accounting for efficient π-π stacks. Following the ranking Trp, Tyr, Phe and His, these aa revealed to form the strongest complexes regardless the chirality, length and diameter of the tubular frameworks.[118] Among them, Trp and Tyr were particularly stable due to the presence of a bicyclic aromatic ring (*i.e.* indole) and an additional polar group (-OH) respectively, responsible of π-π stacking interactions and potential CH···π contribution. After aromatic aa, sulfur-containing, charged and polar residues bearing long alkyl chains were following. If the aa trend was always the same, on the other hand the chiral and zig-zag CNTs were yielding more stable complexes compared to armchairs tubes. This result was argued to be addressed to the major aromaticity of semi-conductive over the metallic CNTs.[119] Furthermore, increasing both length and diameter of the CNTs, the consequent increase in available surface induced the increase of the adsorption energies as well.

While it is well-established the effectiveness of the classical methods with fixed-charge force field to address electrostatic and hydrophobic interactions, less is certain when scientists were often questioning about the reliability of these kind of simulations to capture and efficiently reproduce π-π stacking contacts due to the lack of polarizability.[120,121] In a recent work, Yang *et al.* unraveled the issue by comparing the QM calculations with three popular fixed-charge classical force fields (OPLS/AA, AMBER and CHARMM) contextually to the aa and CNTs interactions.[122] Specifically, they have used an improved DFT method including an empirical dispersion term (DFT-D), namely the density functional tight binding (DFTB-D) method, that was previously benchmarked with other highly accurate QM methods (CCSD(T), QCISD(T), MP2). In order to verify the reliability in describing the π-π interactions of both QM and classical methods only the aromatic aa were considered in this study. They demonstrated that classical force fields, although still defective in predicting the correct aa orientation upon adsorption on CNTs, well reproduce the energetic trend displaying the experimental-validated[123] order (Trp, Tyr and Phe). As expected, classical force fields were slightly overestimating the binding strength compared to QM calculations, but, they can be considered as a valuable tool for studying the relative aa-protein/CNT interactions. The QM calculated induced charges and induction polarization energies of CNTs in the

presence of the aromatic molecules are negligible. These findings exclude a polarization effect of these residues on the nanomaterial surface, nor the need to have a polarized force field to properly describe these interactions when the classical approach is chosen.

The same aromatic trend was also verified in the work of Lin and co-workers[124] that treated Trp, Tyr and Phe adsorbed onto both armchair and zig-zag CNTs at the DFTB-D level of theory. They additionally demonstrated that the π-π interactions are ruled by weak forces that do not affect the electronic profile of the CNTs, as the HOMO/LUMO shapes are identical in the presence and in the absence of the aa (Figure 1.11).

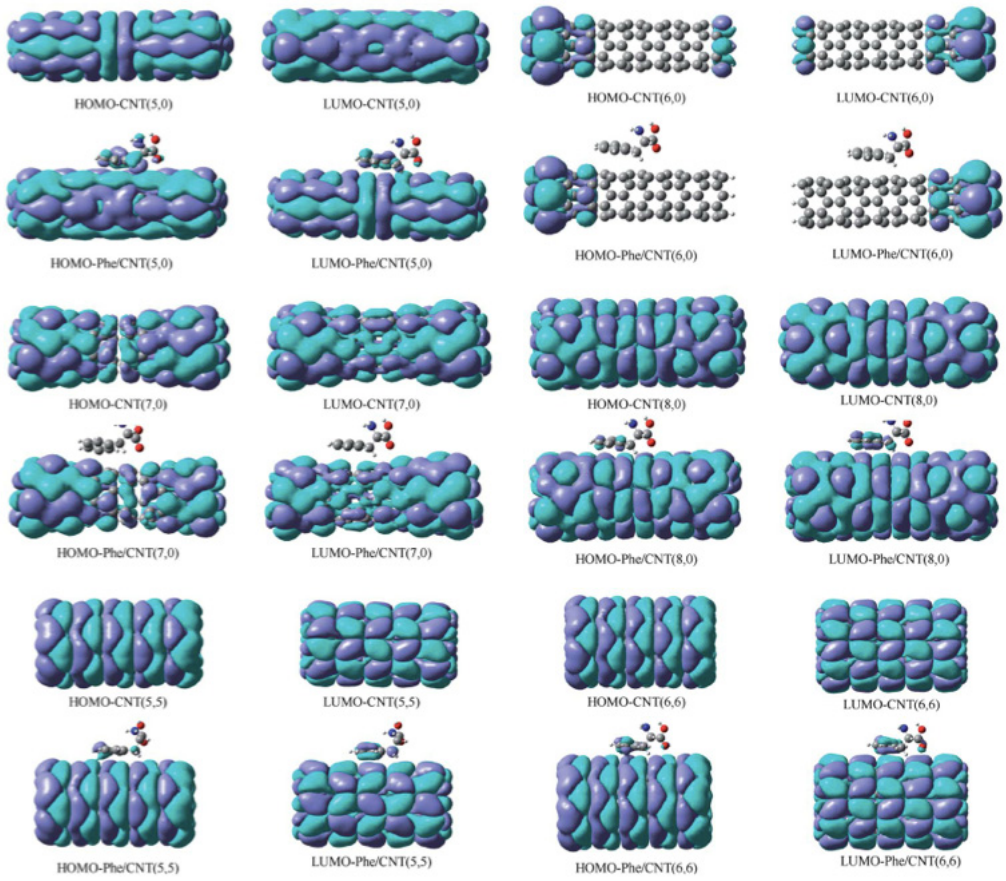


Figure 1.11. The highest occupied molecular orbitals (HOMOs) and lowest unoccupied molecular orbitals (LUMOs) of the different CNTs in the absence and in the presence of the Phe. Reprinted with permission of ref. 123. Copyright 2011, Royal Society of Chemistry.

The binding interactions of these residues generate tighter interaction when interfacing the zig-zag rather than the armchair CNTs, thus stating the preferential

adsorption of aromatic aa on the semi-conductive CNT. Finally, the DOS profile of both CNTs results modified upon adsorption, in fact the presence of the aa induces the same shift in the Fermi level energy from -4.6 eV to -4.4 eV in armachair and zig-zag CNTs.

In short, when the interaction between single aa and CNTs is under investigation, one may rely on the binding rank according to the aromatic > sulfur-containing > charged > polar > hydrophobic aa adsorb onto the graphitic material.

1.6.2 CNTs and polypeptides

As observed in the previous sections, among the essential aa the strongest interacting residues with CNTs are those bearing aromatic appendages, whose peculiar tendency to non-covalently adhere on the CNTs surface through π-π stacking was also well-demonstrated experimentally.[117,123] In this respect, Zhang and co-workers carried out QM calculations (SCC-DFTB level of theory) on complexes of NB1, B1 and B3 dodecapeptides with a SWCNT framework (Figure 1.12).[125] The different content of the aromatic aa in the three sequences clearly influences the binding properties on the graphitic surface. While NB1 peptide anchors the CNT's surface with the only aromatic residue (Tyr) giving rise to a low binding energy (E_B = -0.71 eV), B1 and B3 peptides, bearing 5 and 4 aromatic aa apiece, display stronger interaction energies (E_B = -2.85 and -2.22 eV, respectively). In fact, both peptides can adsorb onto the CNT framework wrapping their structures around the graphitic surface as favored by π-π and XH···π (where X = C and N) contacts. Furthermore, increasing the CNT diameter, an increasing binding energy was observed as a major interacting surface is exposed to the peptide. The authors also investigated the CNT chirality reporting, according to their calculations, major interaction energy for the armchair CNTs in complex with the B1 peptide.[125]

Hence, the natural tendency of polypeptide to adsorb onto CNT through aromatic aa, by establishing π-π and XH···π interactions, does not affect the conjugated π-electronic properties of the nanomaterial as demonstrated by the exploration of the molecular orbitals and the DOS. As observed in above, CNTs face singularly the 20 aa (Figure 1.11),[124] also when interacting with a longer polypeptide its electronic profile in the presence and in the absence of the biomolecules is not

perturbed (Figure 1.12), demonstrating that the governing forces are the weak stacking interactions. In confirmation of this, no charge transfer was detected.

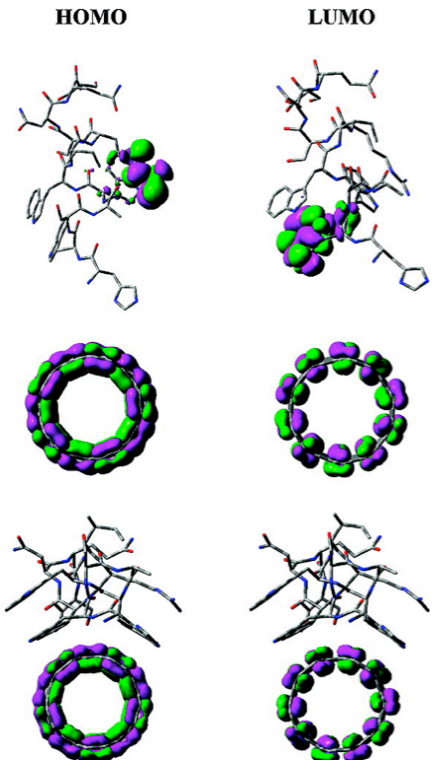


Figure 1.12. The highest occupied molecular orbitals (HOMOs) and lowest unoccupied molecular orbitals (LUMOs) of B3 peptide, (5,5) CNT and their complex. Adapted with permission from ref. 124. Copyright 2009, American Chemical Society.

Noteworthy, the aromatic interactions can be then exploited within designed structures to hierarchically control the CNT solubilization. At this regard, it is worth mentioning the nice example reported by Dieckmann *et al.* that shows the amphiphilic α-helix peptide nano-1 designed to solubilize CNTs and assemble in a macromolecular structure.[126] In fact, exploiting the amphiphilic nature of the designed nano-1, it was enabled the CNT solubilization through non-covalent interactions occurring at the nanomaterial interface with the hydrophobic (rich in Phe and Ala) side of the peptide, and the macromolecular assembly through inter-peptides interactions thanks to the hydrophilic (rich in Lys and Glu) one.

Clearly, the secondary structure preservation results fundamental for the confinement of the functional sides and the amphiphilic character of the designed peptide. So, by performing CD measurements, it was demonstrated the helical content of nano-1, increased in the presence of the CNT. Therefore, the nanomaterial assists

the peptide folding in the designed amphiphilic helix, thereby exposing the hydrophobic face, prone to π-π stacking, towards the tubular surface. Increasing the concentration of the peptide up to the saturation point, the secondary structure content increases as well, further aided by the promoted inter-peptide contacts occurring between the lateral Lys and Glu residues. In order to allow the lateral electrostatic interactions, the peptides, once adsorbed onto CNT, align parallel to each other leading to a fiber-like organization as demonstrated by both polarized Raman measurements and transmission electron microscopy (TEM).[126] Additionally, the debundle process was thoroughly verified by several techniques such as scanning electron microscopy (SEM), atomic force microscopy (AFM) and UV-Vis measurements, observing the outstanding ability of the designed peptide to isolate a single CNT.[127] Notably, the AFM length and height analysis nicely shows the efficient dispersion of CNT when in complex with nano-1 and, most importantly, the individual wrapped CNT by the helical peptide, as observed in the diameter distribution plot (Figure 1.13a).

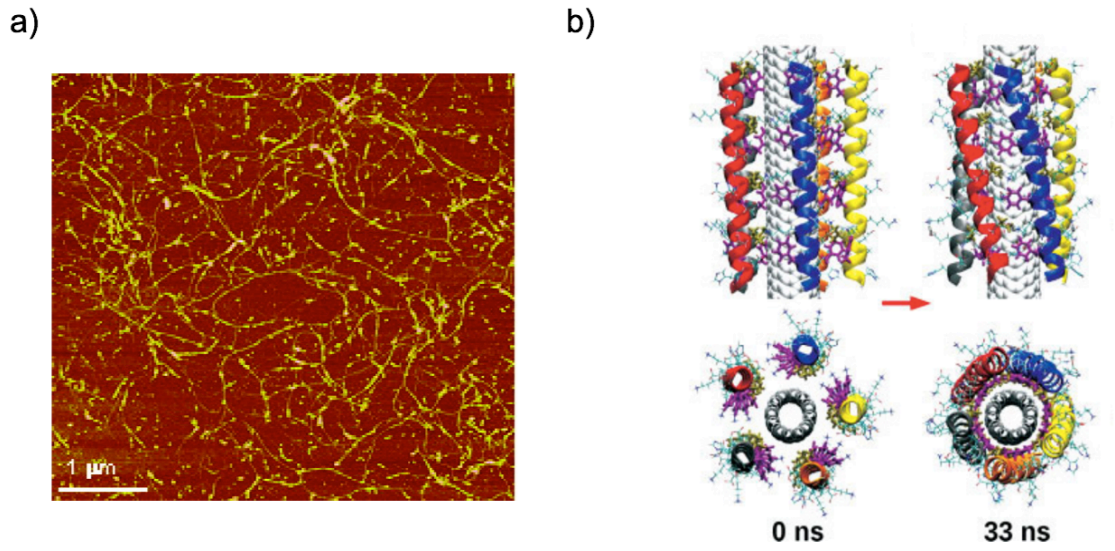


Figure 1.13. AFM image of nano1/CNT dispersion (a) and starting and last snapshot of the simulated adsorption of nano-1 pentamer onto a (6,6) CNT after 33 ns MD simulation (b) are reported. The views perpendicular (top) and parallel (bottom) to the CNT long axis are reported where the hydrophobic aa Phe and Val are colored in purple. Adapted with permission from refs. 126 and 130. Copyright 2004, American Chemical Society. Copyright 2009, Wiley Periodicals, Inc.

Indeed, the height measurements of the conjugate (~ 2 nm) match well with the diameter of the CNT (~ 1 nm) complemented with the peptide (1-3 nm) coating. Furthermore, the non-covalent interactions were characterized systematically varying the aromatic content within the nano-1 sequence, stating by this way the fundamental role of the Phe in anchoring the CNTs surface and ultimately disperse them through the π-stacking phenomena.[128]

Additionally, extensive computational studies were carried out aiming at fully comprehend the non-covalent interactions involved at the nano-1/CNT as well as at the inter-peptides interfaces towards the formation of a hierarchized fibrillar organization.[129] Notably, by comparing the simulated adsorption process and the conformational changes of one single nano-1 approaching different hydrophobic surfaces (benzene, graphite and SWCNT), the nano-1 suitability towards the specific CNT solubilization was corroborated. In fact, the peptide fulfills its function only when adsorbed onto the tubular framework where, thanks to the CNT curvature, can properly anchor through the aromatic aa, expose the hydrophilic faces and preserve the helical structure, lost otherwise.[130] Finally, the contribution of the lateral electrostatic inter-peptides H-bonds was demonstrated by MD simulations to be fundamental in strengthening the binding of nano-1 to CNT, to stabilize its secondary structure as well as the fibril formation around the nanomaterial.[131] Clearly, the adsorption onto the tubular network is driven first by the anchoring of the hydrophobic side of the peptide, then the fiber formation is settled by the lateral H-bonds between charged aa located at the hydrophilic sides (Figure 1.13b).

The masterful design of a helical peptide sequence to achieve the highly-order assembly around CNT was also reported by DeGrado and co-workers.[132] The designed surface-organized peptide functionalizes non-covalently the CNT surface through an ALA-rich side, while inter-helical interactions drive the assembly, and an external Cys represents the gold-binding site to ultimately induce the peptide-assisted gold assembly on CNT.

In general, it emerges a significant participation of hydrophobic and π-π stacking interactions when polypeptides interact with the curved surface of CNT. At this regards, the contribution and the importance of these aspects has been highlighted in the studies of the binding process between CNT and the monomeric amyloid Aβ peptide carried out by Jana *et al.*.[133,134] Performing extensive MD simulations, the

complete adsorption of Aβ peptide onto CNT was observed regardless the starting point, observing the fundamental contacts occurring at the level of the central hydrophobic core (HP1) of the peptide.[133] This domain is characterized by the peptide sequence $L_{17}VFFA_{21}$ where the aromatic Phe (F) in position 19 plays a determinant role. In fact, by simulating a single-point mutation with a non-aromatic but equally hydrophobic residue (Ile, I) or with an aromatic but less hydrophobic aa (Tyr, Y), different adsorption profiles were detected. The virtual mutation of the peptides is indeed reflected in their interaction propensity as not all the analyzed trajectories (four for each system) showed a full adsorption over the curved CNT. From one side, when the aromatic domain is substituted by the Ile, the π-π stacking cannot take place yielding a reduce interaction energy (4-simulations averaged value of -10.5 ± 5.2 kcal mol^{-1}) as compared to native peptide (-17 ± 2.6 kcal mol^{-1}) while adsorbing onto the CNT.[134] This suggests that the hydrophobic character of the HP1 still enables the adsorption on the nanomaterial, but is critically affected by the lack of π-π stacking interactions, owed to the absence of the aromatic component. On the other hand, the Tyr mutated peptide, even though able to establish strong interactions with the graphitic surface through π-π stacking, is weakened in the adsorption (-9.6 ± 9.5 kcal mol^{-1}), due to collateral solvent interactions, that represent a barrier to the complete adsorption of the HP1 domain onto the tube.

However, as stated in section 1.5.2, charged aa exert a remarkable tendency to adsorb also onto CNT mainly due to their peculiar surfactant-like structure where coexist hydrophobic and hydrophilic moieties. Based on this concept, recent experimental results demonstrate the remarkable ability of poly-L-arginine (PLA) peptide of efficiently bind and disperse CNT.[135] The authors used a tip sonicator to disperse SWCNT in the presence of PLA and poly-L-Lys (PLL) as reference polypeptide, reporting very different results between the former and the latter. In fact, the adsorption spectra of different molecular weight polypeptides wrapped around the CNTs displays a smother profile when PLL is used (blue and green lines in Figure 1.14a) compared to the PLA (red and orange lines in Figure 1.14a), as a measure of the better CNT dispersion in Arg polypeptide. The dispersed nanomaterial was quantified by measuring the concentration of the supernatant after ultracentrifugation of the CNT in complex with different molecular weight polypeptides (5-, 10- and 20-mer labeled with R for PLA and K for PLL). The high concentration of CNTs in

complex with the R rather than K oligomer reveals the major dissolving effectiveness of PLA over PLL.[135]

The authors argued for the guanidinium group as main responsible of the efficient dispersion, addressing to its positive charges the ability of inducing electrostatic repulsions among the CNTs, thereby increasing their dispersibility. In fact, by playing with the pH up to acidic condition, the charged aa were deprotonated and the CNT aggregation occurred. The adsorption spectra changed drastically at acidic pH for both the polypetides, but the lower attenuation in the case of PLA suggests its tighter binding to the CNT. Furthermore, the authors availed themselves of steered (SMD) and classical MD simulations to clarify the way of interaction of the polypeptides understudy. 50 ns MD simulations showed that both R15 and K15 adsorb onto the CNT sidewall wrapping around the graphitic framework (Figure 1.14b).

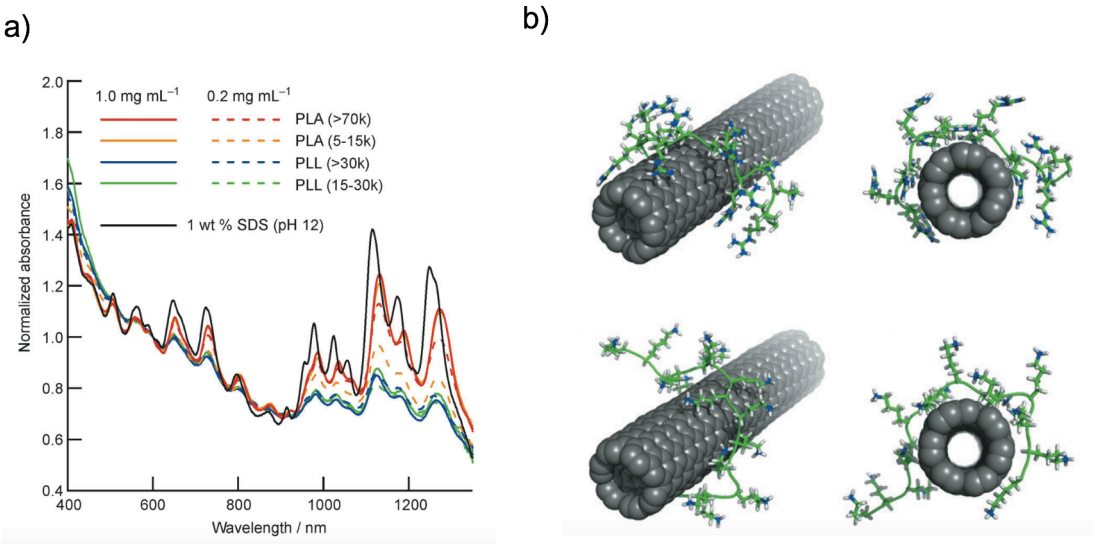


Figure 1.14. Adsorption spectra of SWCNTs dispersed by sodium dodecyl sulfate (black line), PLA > 70 k (red), PLA 5-15 k (yellow), PLL >30 k (blue) and PLL 15-30 k (green) are plotted (a). Significant snapshot of the MD simulation of PLA (b, top) and PLL (b, bottom) adsorbed onto the CNT sidewall. Adapted with permission from ref. 134. Copyright 2004, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

However, the radial distribution functions calculated between each atom of the side chain of the polypeptides and the CNT reveal that the PLA mainly interacts with

the CNT surface through the guanidinium groups, while the PLL binds the nanomaterial through the backbone atoms pointing the charged moiety towards the solvent. This is then reflected in the force values estimated by the SMD when the unbinding process of the polypeptide was simulated: the force necessary to pull out PLA was double than that needed for PLL (200-225 pN and 100-125 pN, respectively), ultimately confirming the more efficient binding and thereof solubilizing effect of PLA over PLL.[135]

In conclusion, the trend as observed for the single aa can be extended to the dynamics of absorption of polypeptides onto CNT, stating the importance of aromatic, hydrophobic and charged aa as long as they can assist for the formation of π-π stacking or hydrophobic contacts.

1.6.3 CNTs and proteins

In section 1.5.3, when the interaction between fullerenes and proteins was probed, it has been highlighted the incidence of the nature of aa within the exposed protein surface, size and shape variability in influencing the way of interfacing between the different biomolecules and the spherical GNM. On the other hand, when it comes into the evaluation of the non-covalent combination of complex and diversified proteins and CNT, also the variable size (length and diameter) of the latter has to be taken into account. In fact, SWCNT diameter can range between 0.7 and 1.4 nm, resembling at their tips approximately the fullerene size, while the length can reach even several microns. MWCNT, often used for experimental studies with biomolecules, can instead display several nm in diameter as well as in length. Reasonably, bigger CNT diameter corresponds to higher available surface and reduced degree of curvature of the external wall. As a matter of fact, none of these features can be underestimated in the study of the interaction of CNT and proteins.

If we start considering the small SWCNT size, several examples have been reported suggesting the actual insertion of the nanomaterial within hydrophobic pockets of proteins. For instance, the classical MD simulations performed by Zuo *et al.*[136] show the snuggling of SWCNTs with different diameters (5.38, 6.73 and 8.08 Å) in the binding site of YAP65WW domain. This protein is characterized by a triple stranded β-sheet with a hydrophobic core that binds the natural ligand, the proline-

rich motifs (PRMs) (Figure 1.15a). Several trajectories of 200 ns time scale were always converging towards the complementary shape fitting of the CNT within the hydrophobic pocket of WW protein, where Trp and Tyr were mainly found to stabilize the interaction (Figure 1.15b). These crucial residues normally stabilize the protein between the first and second β-sheet though the stacking between them. With the interaction of CNT, the distance between Trp and Tyr, denoting their π-stacking, increases in favor of the approach to the CNT.

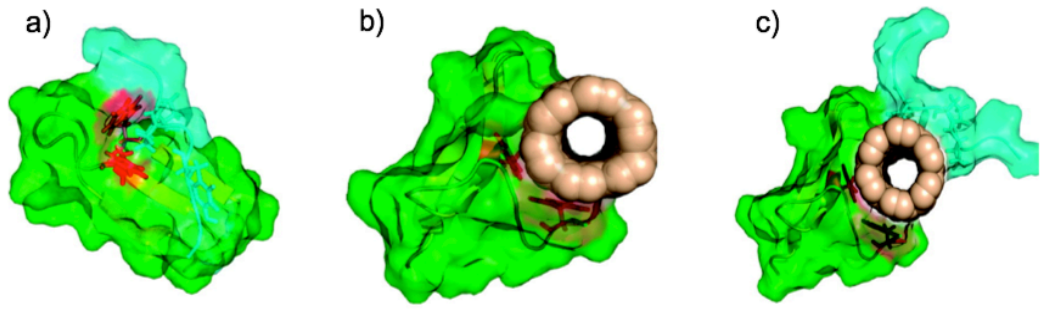


Figure 1.15. YAP64WW protein, in green, is represented in complex with the PRM in light green (a). Licorice representation of Tyr and Trp is highlighted in red. SWCNT is plugged into the WW protein (b). When CNT is bind to WW protein PRM is not able to bind anymore the active site (c). Adapted with permission from ref. 135. Copyright 2010, American Chemical Society.

Thus, the loss of the original contacts within the protein β-sheet and the formation of π-π stacking between the CNT and the aromatic aa cause the partial disruption of the active site that, occupied by the nanomaterial, thus unable to recognize the PRMs ligand, results altered in its functionality (Figure 1.15c). Similarly, the authors reported on the same plugging mode found simulating the interaction between CNT and the SH3 domain, another PRMs-binding protein.[137] MD simulations show also in this case that the PRMs fails in binding the SH3 domain loosing the competition with CNT, confirmed by the major binding energy calculated for the CNT compared to the endogenous ligand (-6.08 and -5.51 kcal mol^{-1}, respectively). As underlined above, aromatic aa, such as Trp and Tyr, and the cyclic Pro were found to stabilize the interaction of the CNT in the active site through hydrophobic interaction and π-stacking.

The CNT-protein shape complementarity driven by the interaction that a proteinic cavity can entertain with the hydrophobic nanomaterial surface directly remind to the ability of C_{60} to place in and tightly bind proteins. Notably, Sesti and co-workers demonstrated that C_{60} as well as SWCNT were suitable ion channel blockers thanks to their shape matching with the channel pore.[101] Very recently, a docking study performed by Turabekova *et al*. found that (5,5)CNT and C_{60} can stably bind to Toll-like receptors giving a plausible explanation of the cytokines and chemokines release induced by CNTs.[138] Additionally, the anti-fullerene antibody was demonstrated to similarly bind SWCNT by AFM measurements.[139]

However, these findings can be easily affected by a diameter dependent effect as analyzed by Regev and co-workers that, in a recent work, found the ability of a whey protein, the β-lactoglobulin, to solubilize CNT with two different diameter ranges.[140] They demonstrated that the solubilizing effect of milk is ascribable to its proteinic part and in particular to the β-lactoglobulin, that presents a narrow crevice (site A, Figure 1.16a-i/ii) where is able to accommodate CNT with a diameter < 4nm, while an external surface area (site B, Figure 1.16a-iii/iv) assist the solubilization of CNT with diameter > 40 nm. These outcomes were nicely verified by experimental, *i.e.* solubilization tests, UV-vis measurements (Figure 1.16b) and Cryo-TEM imaging, and in silico simulation, such as docking studies (Figure 1.16a). Both site A and B are characterized by hydrophobic and positively charged aa, found to interact with the different diameter CNTs.

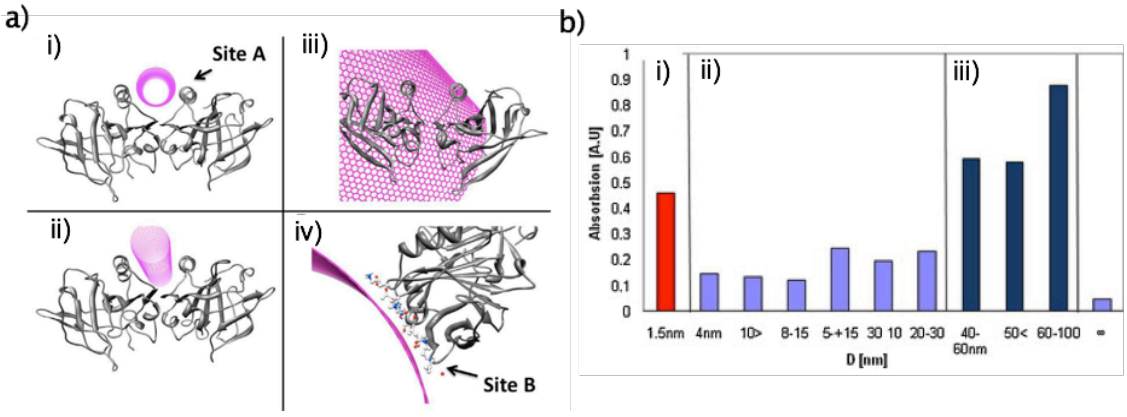


Figure 1.16. a) Docking models of the β-lactoglobulin interacting with a 1nm diameter CNT in the site A (i and ii) and with a 100 nm diameter CNT in the site B (iii and iv). b) UV-vis absorption of several diameters CNTs organized in zone i), ii) and iii) according to their dispersion efficiency. Adapted with permission from ref. 139. Copyright 2013, Elsevier.

As demonstrated above, the CNT insertion or the external surface adsorption onto proteins inevitably depends on the size and diameter of the GNM, while the orientation between the two strictly depends on the 3D structure and the amount of hydrophobic aa exposed by the interacting protein.[45] Concerning this aspect, a noteworthy confirmation came from the work of Ge *et al.* which described the different interaction processes between SWCNT and several blood proteins (fibrinogen BFG, immunoglobulin Ig, albumin BSA, transferrin Tf and ferritin) both experimentally and computationally.[141] Combining AFM and CD techniques, the authors report a non-uniform adsorption of BFG and Ig with significant changes in the secondary structure profile that did not recover over the time (Figure 1.17); conversely, a uniform adsorption was described for Tf and BSA, whose changes in secondary structure were instead recovered over the time.

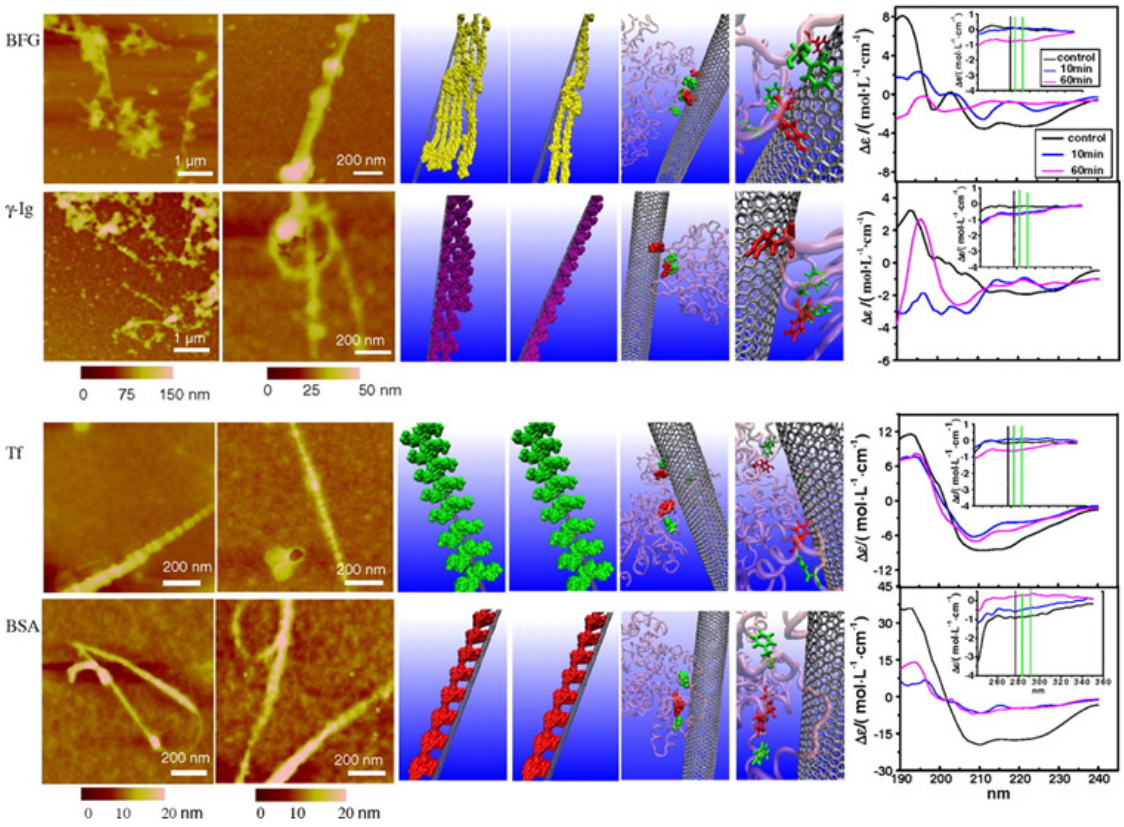


Figure 1.17. From left to right: AFM images of proteins in complex with SWCNT after 10 mins and 5h (first and second column, respectively). Molecular models of the proteins adsorbed onto the graphitic surface and the highlight of the Tyr (red) and Phe (green) aa. Far-UV CD spectra of proteins incubated with SWCNT. In the inset near-UV CD spectra are also reported. Adapted with permission from ref. 140. Copyrights 2011, National Academy of Science.

These experimental evidences were further elucidated by MD simulations, shedding lights on the structural aspects of the proteins that influence the different adsorption profiles. In fact, by simulating the complexes formation and their dynamics, it was highlighted the crucial role of the aromatic Tyr, Phe and Trp aa interacting at the CNT/proteins interfaces. Interestingly, the interacting aromatic aa content, the average contact residue number and the average contact surface area of the simulated systems lead to the rank BFG > Ig > Tf > BSA that is in perfect agreement with the experimental results.[141]

Further and more detailed studies on the interaction between CNT and BSA were reported, as a representative blood protein able to solubilize CNT.[142–144] BSA is α-helical multi-domains protein that folds in a globular tertiary structure including thereby a hydrophobic core at its interior. Both experimental and computational studies hypothesize a stepwise adsorption on CNT along which the secondary structure is conserved, but the globular information is lost exposing the hydrophobic core to the carbon surface.[142,143] The interaction was characterized at the atomistic level by MD simulations confirming the labile nature of the albumin tertiary structure upon adsorption onto the tubular network, in the attempt of exposing the hydrophobic aa. In fact, the orientations that yield a higher interaction energy bear more hydrophobic and aromatic components.[144]

Another well-know macromolecule used to disperse CNT in water, is Lysozyme (LYS) that along with the hydrophobic and π-π interactions, through which can irreversibly bound to CNT, presents a high content of basic residues that consents a pH-sensitive debundling process.[142,145,146] At this regard, the non-covalent interaction between LYS and CNT was skillfully investigated by MD simulations in a recent work by Calvaresi *et al.*.[114] Remarkably, the site in which the CNT preferentially adsorb is found far from the active site of the LYS, assuming the preservation of its functional activity. Furthermore, the secondary as the tertiary structure of the protein was not affected by the binding with the nanomaterial that lead to the formation of a stable hybrid complex. A more in depth analysis reveals that the formation of this complex was mainly driven by the presence of two helical structural motifs bearing Arg residues that act as tweezers in anchoring and solubilizing the CNT. This is in perfect agreement with the previously reported experimental works.[145,146]

In general, SWCNTs have the suitable size to enter protein structures and nestle in their hydrophobic pockets. On the other hand, bigger CNTs fail in snuggling within the internal protein pockets but exert a remarkable tendency of interacting and non-covalently bind the external surface of the protein. The adsorption process is ruled by the shape and size of the protein surface and the amount of the exposed hydrophobic, aromatic and charged residues that are demonstrated to play a determinant role, according to the different protein.

1.7 Hybrid graphene

1.7.1 Graphene and amino acids

As far as aa and graphene are concerned, their interaction might be influenced by the zwitterionic nature of the aa backbone. At the same time, the amino (N-) and carboxy (C-) *termini* might be capped with an acetyl and N-methyl groups, respectively, thereby resembling the typical peptide bond.

At this regard, in a recent work, Rubel and co-workers performed 40 ns classical MD simulations using the AMBER 03 force field modeling both in vacuum and in explicit water the 20 natural aa capped and uncapped, placed with their side chains parallel to graphene.[147] At first, comparing the graphene adsorption of capped and uncapped aa in vacuum, a higher adsorption energy of the former was observed compared to the latter (vertical lined *vs* parallel lined columns, respectively in Figure 1.18a). This behavior was ascribed to the larger number of atoms within the capped residues contributing to the adsorption energy as it is driven by vdW forces.

Afterwards, focusing on the capped aa, in order to assess the effect of the explicit treatment of water, vacuum/water simulations were compared. A drastic decrease in the adsorption energy of the water capped compared to vacuum capped aa was observed (right oblique *vs* vertical lined columns, respectively Figure 1.18b), suggesting a negative influence of water molecules in the aa binding onto the graphitic surface. In fact, aa/water interactions were now competing with the aa/graphene complexes, therefore weakening the final adsorption energy to graphene. However, from the adsorption energy plot of the water capped aa reported in Figure 1.18b, Arg > Trp > Tyr > His > Gln were identified as the most interacting residues. As observed for fullerenes and CNTs (see section 1.5.1 and 1.6.1, respectively),

positively charged and aromatic aa represent the best residues for the interaction also with graphene, favored by their bigger side chains, as compared to the other aa, able to maximize the vdW contact, and by the aromatic moieties able to stack on the flat surface.[147]

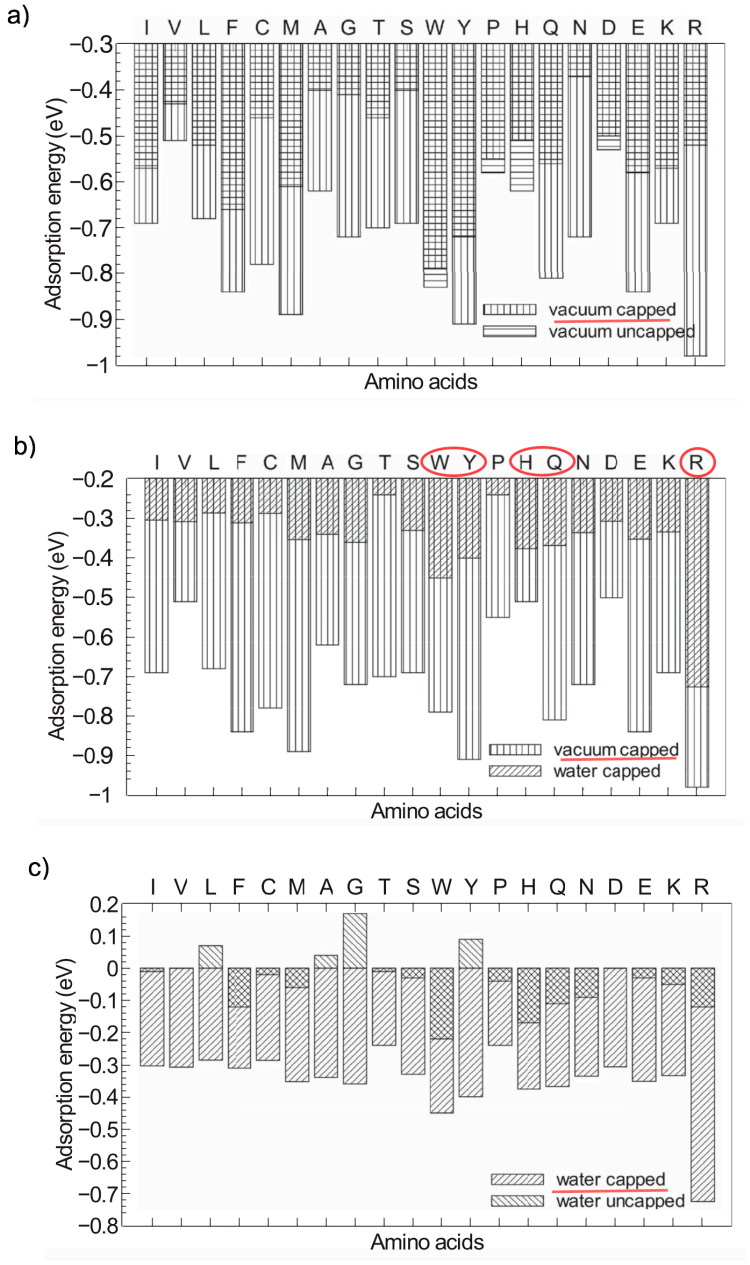


Figure 1.18. Adsorption energy onto graphene of vacuum capped and uncapped aa (a), of vacuum capped and water capped aa (b) and of water capped and uncapped aa (c). Red lines underline the most graphene interacting system within each simulated pair. Red circles highlight the most interacting water capped aa to graphene (b). The aa are labeled with the one letter abbreviation. Adapted with permission from ref. 146. Rights managed by AIP Publishing LLC.

Finally the authors distinguished between capped and uncapped aa in water, observing that the latters tend to desorb from the graphene surface, due to the better solvation of their zwitterionic state. For this reason, while a favorable interaction of water capped aa was detected (negative adsorption energies, right oblique lined columns, Figure 1.18c) lower or even positive energies were registered for water uncapped aa simulated onto graphene (left oblique lined columns, Figure 1.18c).

Consistent results were also obtained by Pandey *et al.* that, by simulating the 20 natural capped aa adsorbed onto graphene by means of classical MD simulations (AMBER ff99SB force field), found a similar trend of interaction (Trp > Tyr > Arg > Phe > His > Lys) both in vacuum and explicit water.[148]

In parallel, Camden *et al.* designed a Gly-X-Gly (GXG) tripeptide to quantify the single aa (X) contribution in the adsorption to graphene surface but inserted in a more realistic peptide chain model.[149] The binding energies were calculated from MD simulations in explicit solvent according to the equation $BE = (E^{SURF+SOLN} + E^{WAT}) - (E^{SOLN} + E^{SURF+WAT})$, where E^{WAT} is the contribution of a water box, E^{SOLN} of a peptide solution, $E^{SURF+SOLN}$ of graphene and peptide, and $E^{SURF+WAT}$ of graphene alone. Furthermore, the contribution of the lateral Gly, modeled in the zwitterionic state, was estimated correlating the binding energy of G-$(G)_n$-G systems to graphene as a function of the chain length. This procedure allows the determination of the lateral Gly energy values, which were then subtracted to the total GXG energy values, obtaining thereby the individual X contribution to the graphene binding.

Among the 20 aa spontaneously adsorbed onto the graphitic surface (negative binding energy values), Arg (-17.5 kcal mol^{-1}), Gln (-15.9 kcal mol^{-1}), Asn (-15.3 kcal mol^{-1}) and Lys (-14.9 kcal mol^{-1}) are the most interacting residues. Analyzing the hydrophaty index of aa where positive (negative) numbers refer to hydrophobic (hydrophilic) side chains,[150] this well correlates to the trend described by the binding energies. As shown in Figure 1.19, Arg, the aa with highest binding energy (-17.5 kcal mol^{-1}), has the lowest hydropathy index (-4.5 kcal mol^{-1}), as also verified for the other most interacting polar aa (Gln, Asn and Lys). The linear correlation between the hydropathy index and the binding energy confirms the role of hydrophilic (charged and polar) aa in the adsorption to graphene.

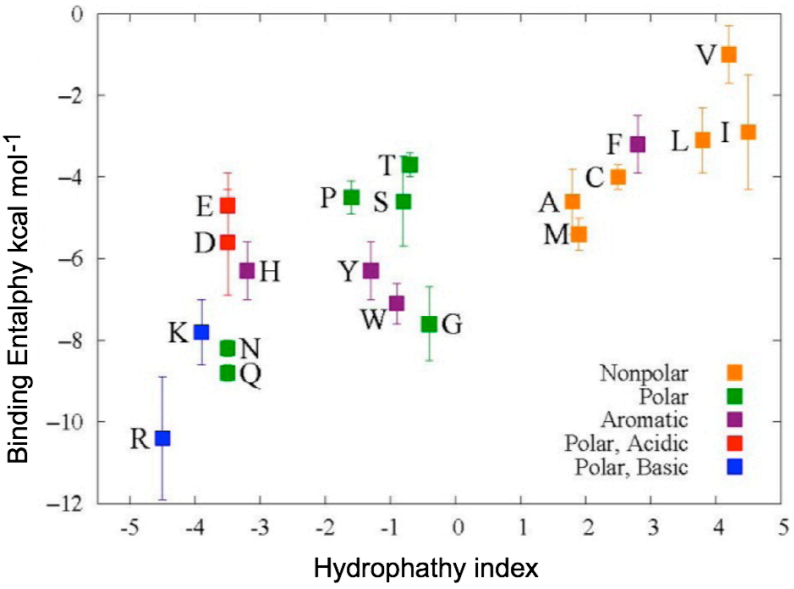


Figure 1.19. Binding energy of the interacting aa (labeled with the one letter abbreviation) plotted as a function of the hydropathy index. Adapted with permission from ref. 148. Copyright 2013, American Chemical Society.

In line with this, the highest interaction energy of hydrophilic aa (blue, red and green residues in Figure 1.19), found in the first water shell, is explained by the favorable coulombic interactions occurring with the negatively charged water oxygen. Negatively charged aa, such as Glu, similarly located in the first water shell, display instead a reduced binding energy (-11.8 kcal mol^{-1}) due to the repulsive interactions with water.[149] In summary, the prominent role of positively charged aa involved in the interaction between aa, polypeptides, proteins and GNMs was widely explored both experimentally and theoretically, and was now found significant also in the adsorption on the 2D allotropic form of carbon.

This behavior was further verified when the interaction between a well-known bioactive tripeptide Arg-Gly-Asp (RGD) was simulated adsorbed onto the graphitic sheet.[151] DFT calculations were used to characterize the RGD-graphene non-covalent binding where the zwitterionic form of the tripeptide was modeled on a 240 atoms graphitic surface by semicore pseudopotential with DNP basis set. As shown in Figure 1.20a, different RGD conformations adsorbed onto the graphitic sheet were investigated among which the planar flat disposition yields the higher adsorption energy (Figure 1.20a-i/v).

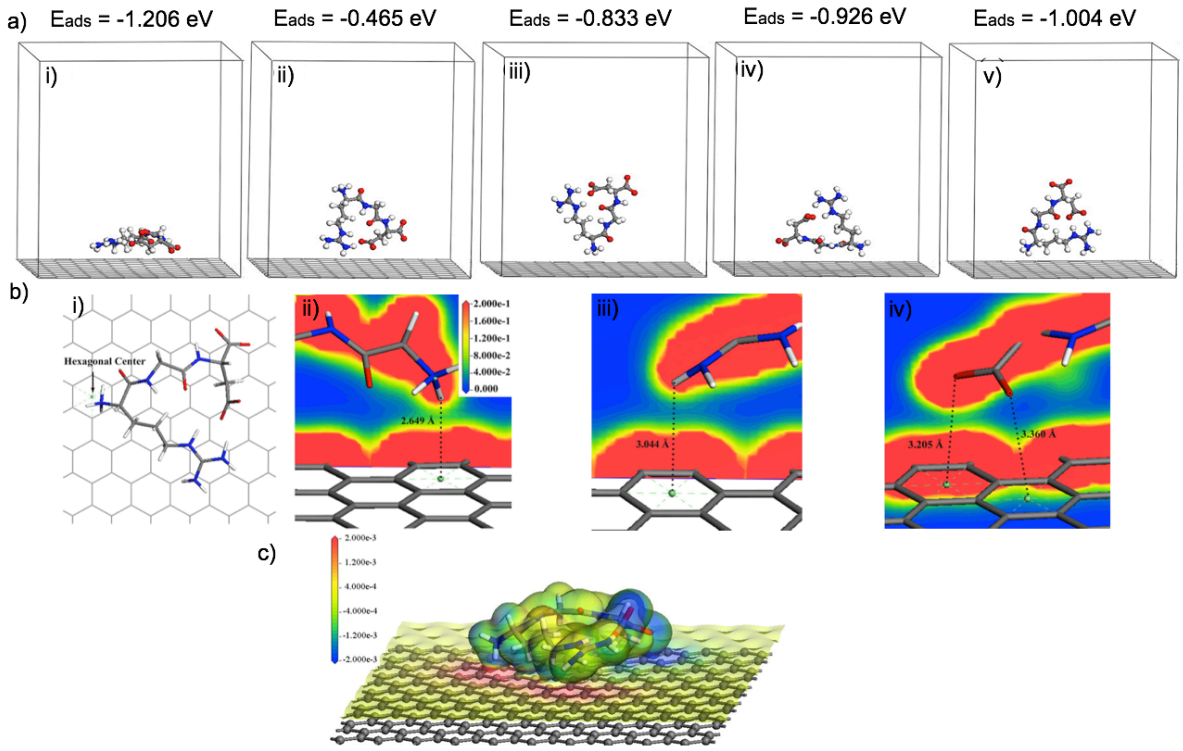


Figure 1.20. a) 5 configuration of RGD peptide on graphene (i-v). For each of them is reported the adsorption energy, E_{ads} (eV). b) Best ranked RGD conformation adsorbed onto graphene, whose hexagonal center is labeled (i). Electron density and geometrical distance between the functional groups, NH_3^+ (ii), the guanidinium-NH_2 (iii) and the COO^- (iv) and the graphitic hexagonal centers are highlighted. c) Charge accumulation (red) and depletion (blue) are highlighted in the electron density isosurface mapped with electron density difference of the RGD-graphene conformation. Adapted with permission from ref. 150. Copyright 2013, American Chemical Society.

An in depth visual inspection reveals that all the three functional groups within the RGD structure in the best ranked orientation are in contact with the graphitic surface pointing the NH_3^+, the guanidine-NH_2 and the COO^- towards the hexagonal center (Figure 1.20b-ii/iv). It is worth noting that a charge accumulation on graphene is observed in particular in the area of the amino and guanidine groups, while a charge depletion corresponds to the area of COO^- location (Figure 1.20c). The low intensity of the electron density is a clear clue of the weak non-covalent interactions coming into play in this binding geometry. This study further confirms the predominant role of positively charged or amino functional groups in the adsorption of aa onto graphene.[151]

In order to evaluate the solvent effect, the adsorption energy of the RGD/graphene systems in water was also calculated. The water molecules result to

intercalate between the NH_3^+ and COO^- groups, inducing thereby a conformational change in the tripeptide structure that is now optimized in a more opened conformation. This allows for a better backbone adhesion onto graphene and for the attainment of higher energy value (-1.713 eV) compared to that in vacuum (-1.206 eV).[151]

Besides the interest in positively charged aa interacting with graphene, the spontaneous adsorption of the hydrophobic aa L-Leu, one of the most common and bioactive protein building block, was also investigated by a combination of DFT and MD calculations.[152] The aa, simulated in the non-ionic uncapped state at the LDA/DNP level of theory, was demonstrated to preferentially adsorb onto the graphene in a flat orientation (named G3-L) pointing the hydrophobic side along with the NH_2 group towards the surface (Figure 1.21a).

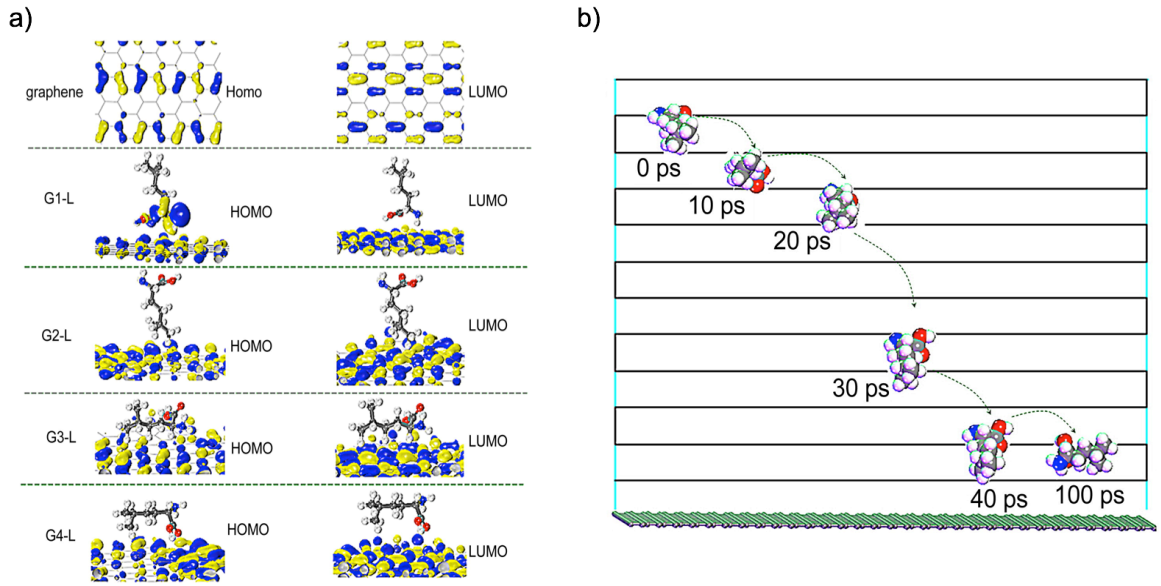


Figure 1.21. a) HOMO/LUMO mapping for graphene and the four Leu/graphene orientations understudy. b) Dynamic of the adsorption process of Leu onto graphene simulated by means of MD simulation in 100 ps time scale. Adapted with permission from ref. 151. Copyrights 2010, Elsevier.

The G3-L orientation yields a binding energy of -0.31 eV, which is in good agreement with the aforementioned cases (see Leu among uncapped aa in vacuum displayed in the Figure 1.18a and 1.20a).[147,151] The nature of this interaction was thoroughly characterized by analyzing the electronic profile of the HOMO/LUMO orbitals, the electronic density of state (DOS) and the electronic band of structures

confirming allover the physisorption process occurring between Leu and graphene (Figure 1.21a).[152]Additionally, the dynamic aspects of the adsorption were nicely described by the MD simulations that, in agreement with the DFT calculations, predict a favorable and spontaneous adsorption within a short time scale (100 ps) with a parallel flat orientation (Figure 1.21b). A self-assembled parallel Leu organization was also observed when 100 molecules were simulated adsorbing onto the graphitic sheet as driven by intermolecular interactions between the uncapped NH_2 and the COOH moieties (Figure 1.22).

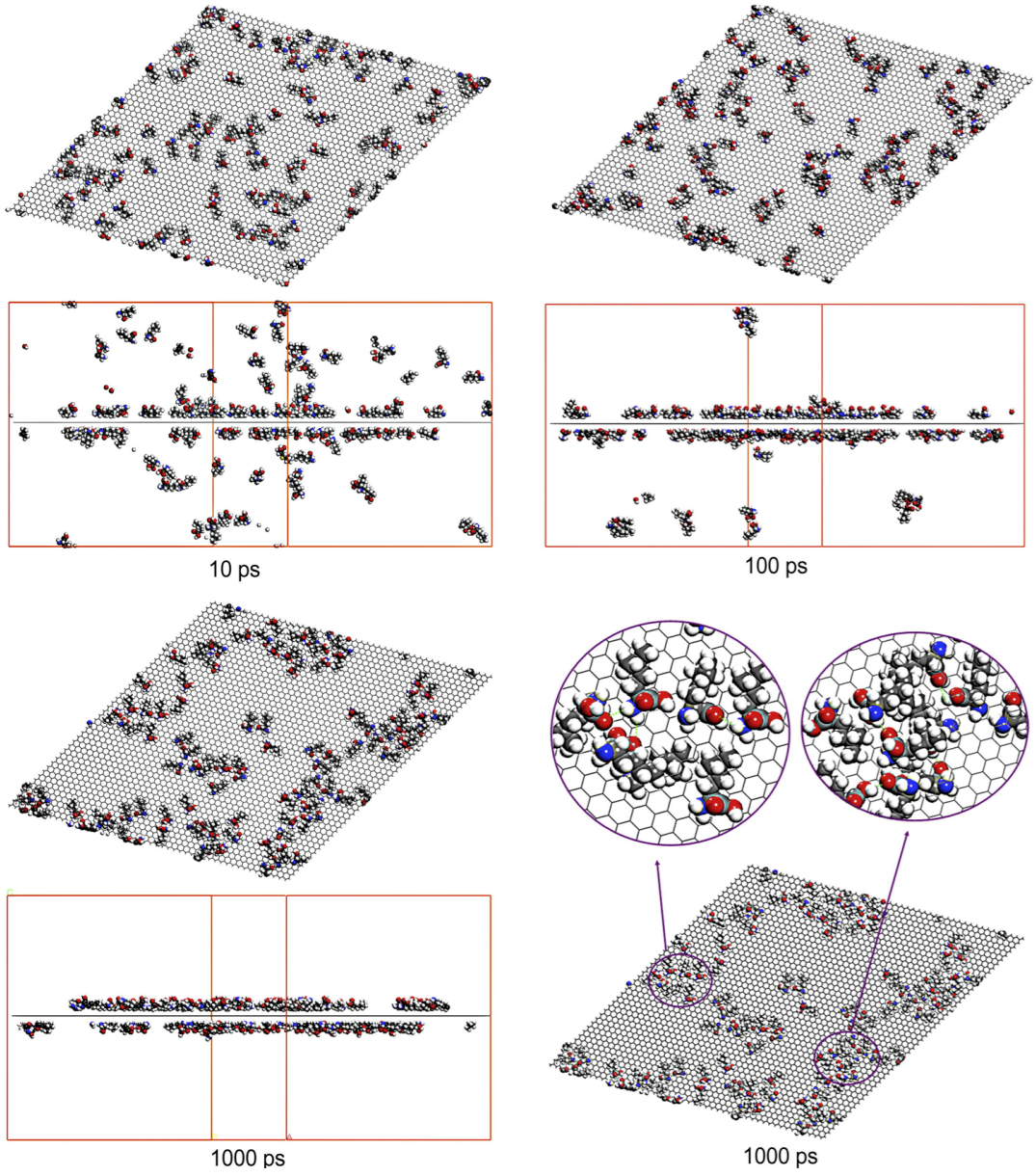


Figure 1.22. Top and side views of the adsorption process of 100 Leu molecules onto graphene, arranging in a parallel self-assembled organization (inset). Reprinted with permission from ref. 151. Copyrights 2010, Elsevier.

Consistently with the above study, the highly ordered organization of L-Met aa self-assembled onto graphene was proved by scanning tunnel microscopy (STM) and further corroborated by classical MD simulations.[153–155] The sulfur-containing aa, deposited on the graphitic surface under ambient conditions, was able to achieve a low-coverage (40%) of the surface by organizing in regular rows spaced by a 45 Å gap and 18 Å in width, while a space of 8.6 Å was found between the individual Met molecules (Figure 1.23a-b).

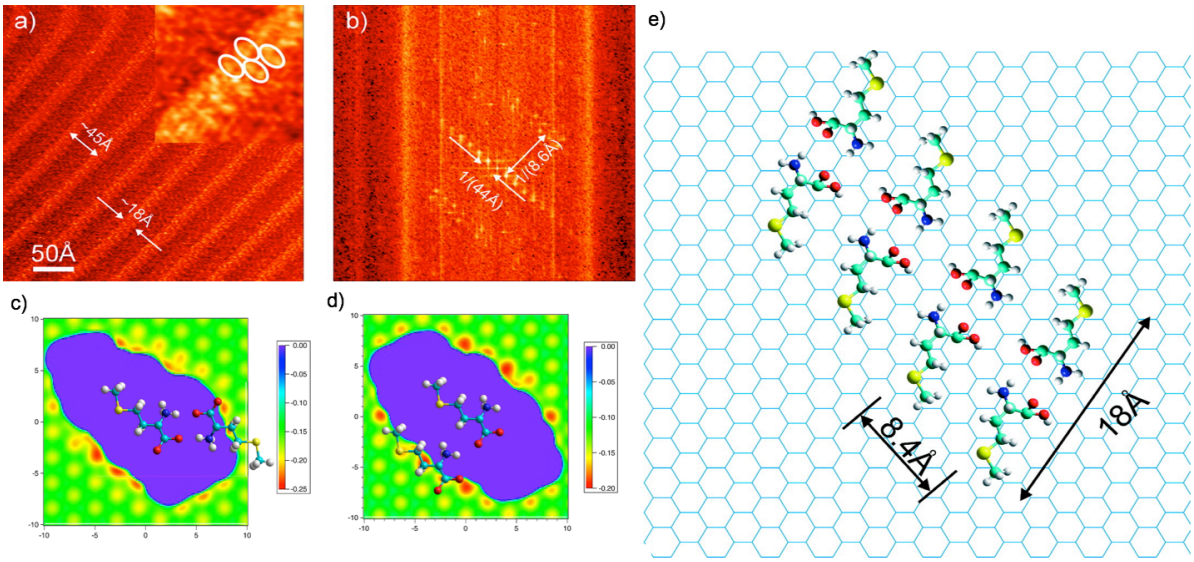


Figure 1.23. STM images show the low-coverage assembly of Met with the width rows of 18 Å (a), the spacing between the individual Met molecules of 8.6 Å and a row spacing of 45 Å (b). Antiparallel (c) and parallel (d) orientations of two Met molecules plotted over the potential energy. The proposed model for Met adsorption onto graphene surface (e). Adapted with permission from refs. 152 and 153. Copyrights 2009, American Chemical Society and 2010, Elsevier.

Theoretical studies complement the experimental evidences identifying the configurations behind the self-organized wires formation that well match the observed geometrical parameters. In fact, by simulating the orientation of a Met around another Met molecule taken fixed, it was mapped the total potential energy of the possible dimers. By looking at the energy level (lowest and higher energies are colored red and blue, respectively) was identified an antiparallel configuration of two Met molecules interacting through H-bonds between amino and carboxylic groups, the distance of which is in perfect agreement with the experimental ones (Figure 1.23c). On the other

hand, a parallel organization was also identified, well describing the intermolecular distance along the row (Figure 1.23d).

As a general trend, aa tend to spontaneously adsorb onto the graphene flat platform and in order to maximize the surface contact a flat orientation of the proteinogenic building blocks is always preferred. Additionally, positively charged, aromatic and polar aa are demonstrated to non-covalently bind more tightly than the others 20 aa thanks to the right combination of molecular weight, vdW contribution and hydropathy index within their structure.

The importance of the aromatic participation in the non-covalent interactions of aa and GNMs results slightly surpassed by the positively or amino-containing residues when the adsorption involves individually protein building blocks and graphene. Nevertheless, as observed for CNTs in section 1.6.1, the choice of the level of approximation in modeling non-covalent interactions might be crucial to the reliability of the final results. At this respect, focusing exclusively on the interaction between aromatic aa and GNMs, Rajesh *et al.* compared the behavior of Trp, Tyr, Phe and His on the flat graphene and the rolled CNT, simulating the complex stability at the MP2/6-31G* level of theory.[156] The optimized geometries of the selected aa onto the planar and tubular frameworks display a flat orientation of all of them on both the GNMs with similar, although smaller for CNT, interplanar distances found in the range of 3.3 to 3.50 Å (Figure 1.24).

This clearly alludes to π-π interactions. Interestingly, the interaction energies of these systems were calculated depicting, for both the nanomaterials, the predicting classification Trp > Tyr > Phe > His, which perfectly correlate to their calculated polarizability α (362.65, 109.46, 76.35 and 66.71 a.u. respectively). However, it is worth noting that the aromatic aa interact more strongly with the planar graphitic platform rather than with the curved CNT, as the calculated interaction energies reveal (0.84/0.72 eV for Trp on graphene and CNT, respectively).

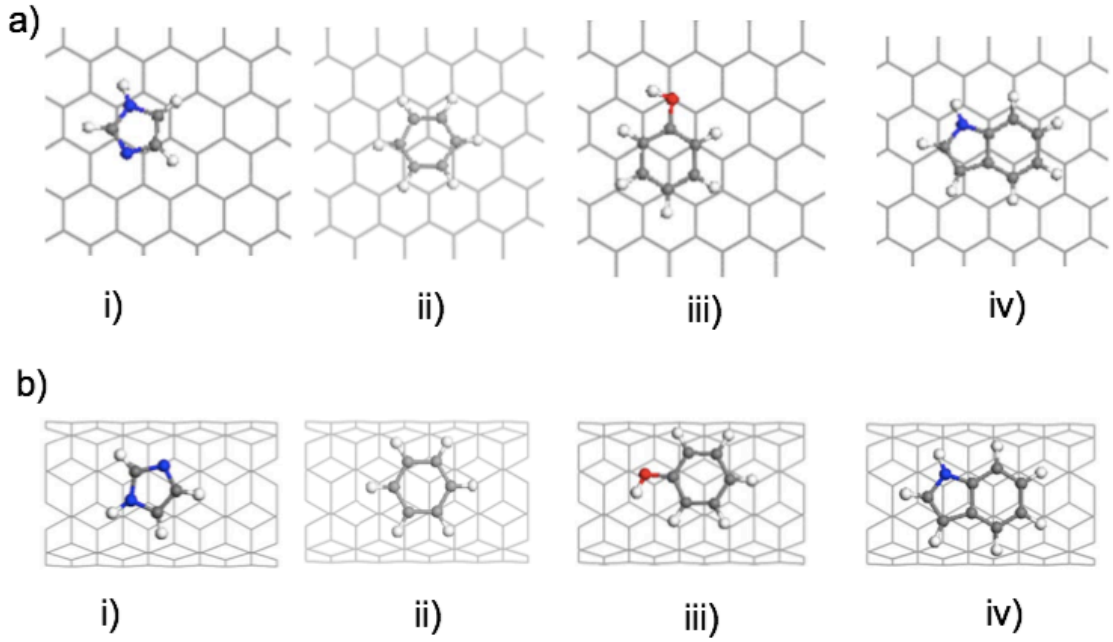


Figure 1.24. Optimized geometries of His (i), Phe (ii), Tyr (iii) and Trp (iv) on the flat surface of graphene (a) and on the curved CNT (b). Adapted with permission of ref. 155. Rights managed by AIP Publishing LLC.

It gathers that the planar geometry of graphene clearly enhances the π-π stacking occurring between the aromatic rings and the surface generating smaller interplanar distances when the aromatic aa adsorb on it rather than on CNT. Consequently, higher interaction energies are observed for aromatic aa/graphene complexes demonstrating the stability of these residues on planar graphitic surfaces and the nature of this interaction evidently influenced by the curvature of the GNM.

1.7.2 Graphene and polypeptides

Thanks to their short length and the ease in their production, polypeptides are a versatile and suitable molecules for studying the interaction of peptide biomolecules with GNMs, such as graphene, to explore potentiality, driving forces and mutual effects involved in the combination of aminoacidic chains interfacing the flat graphitic sheet.

In particular, the combinatorial phage display libraries have been widely used to identify specific graphene-binding peptides, which were thereafter investigated adsorbed onto the graphitic surface in the view of designing new biomimetic hybrids.

The structure of a dodecamer peptide, identified as graphene/graphite binding peptide (GBP) by phage display libraries, was thoroughly studied once adsorbed on these GNMs by means of AFM, Raman, Fourier transform infrared (FTIR) spectroscopy and MD simulations.[157] The authors studied GBP structure first in its powder form by IR spectroscopy demonstrating the formation of an organized α-helical secondary structure (peaks at 1660 and 1530 cm^{-1}), lost when it is immersed in aqueous medium (peak at 1673 cm^{-1}). When deposited onto the graphitic surfaces, the same identical mesh-like adsorption was observed both in graphene and graphite, as proved by the AFM images (Figure 1.25a-c) and their height analysis (thickness of 0.99 ± 0.63 and 1.10 ± 0.45 for doped graphene and graphite, respectively).

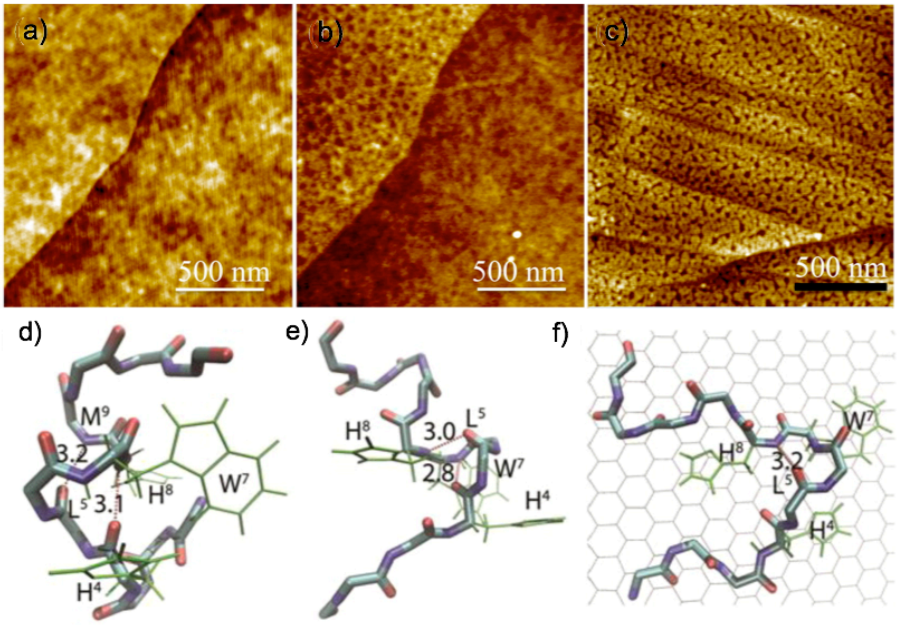


Figure 1.25. AFM images of graphene in the absence (a) and in the presence of GBP on its surface (b). GBP deposited on graphite (c) is also reported. The α-helical structure of GBP in vacuum is shown along with the intermolecular H-bonds (d), while a destabilized helix is found in water (e) and a fundamental aromatic anchoring of GBP is observed when adsorbed onto graphene (f). Adapted with permission from ref. 156. Copyrights 2012, American Chemical Society.

The adsorption nature and effect was clarified by the Raman spectra in which no damage of the graphitic lattice was observed, meaning a non-covalent interaction driven by weak forces. Additionally, peptide/nanomaterial FTIR spectra, resembling the spectrum of the powder form, clearly show the actual adsorption of GBP with a blue-shifted amide I peak (peak at 1670 cm^{-1}) as a sign of a different helical organization induced by the presence of the nanomaterial. Aiming at comprehend the

structural differences described experimentally, the authors carried out MD simulations modeling the powder form of the GBP in vacuum, the solvated form in explicit water, and the non-covalent adduct adsorbed onto graphene.[157]

In good agreement with the experimental outcomes showed by the FTIR, the peptide in vacuum adopts the organized α-helical structure (Figure 1.25d), which is lost in solution (Figure 1.25e) and differently reorganized once adsorbed on the nanomaterial (Figure 1.25f). Notably, the exposition of the aromatic Trp and His induce the structural reorganization that allows for the anchoring of peptide on the flat surface. Furthermore, their virtual mutation provokes a drastic reduction of the calculated interaction energies (126 ± 0.2 for GBP and 112 ± 0.2 kcal mol^{-1} for Ala-mutated peptide), confirming their fundamental role in the binding.[157]

Clearly aromatic aa within the GBP play the determinant role for the favorable adsorption on graphene and graphite. However, edge-specific peptides were identified revealing other non-covalent interactions occurring at their interface. In fact, a phage display identified eptapeptide (P2) was found to bind preferentially to the edges of graphene surface and was compared to another plane-binding dodecapeptide (P1).[158] The peptides adsorption on graphene was monitored by AFM images that clearly show the lateral confinement of P2, even more evident when the latter was coated with AuNPs (Figure 1.26a and c). On the other hand, P1 was clearly localized on the central position of the graphitic sheet (Figure 1.26b).

In order to fully comprehend the dynamics of the specific surface binding, MD simulations demonstrated the border placement of P2 thanks to electrostatic interaction between Glu aa and the H-terminated edges, while the preferred central binding of P1 is ascribable to π-π stacking interactions occurring between the highly conjugated graphitic surface and the HYWYF aromatic motif.

To further corroborate the electrostatic mechanism, P2 was simulated protonated as well as tested experimentally in acidic conditions, thereby exerting a different binding to the surface. Conversely, P1 behavior remained unaffected by the acidic conditions, unambiguously confirming the hypothesized nature of the interaction.

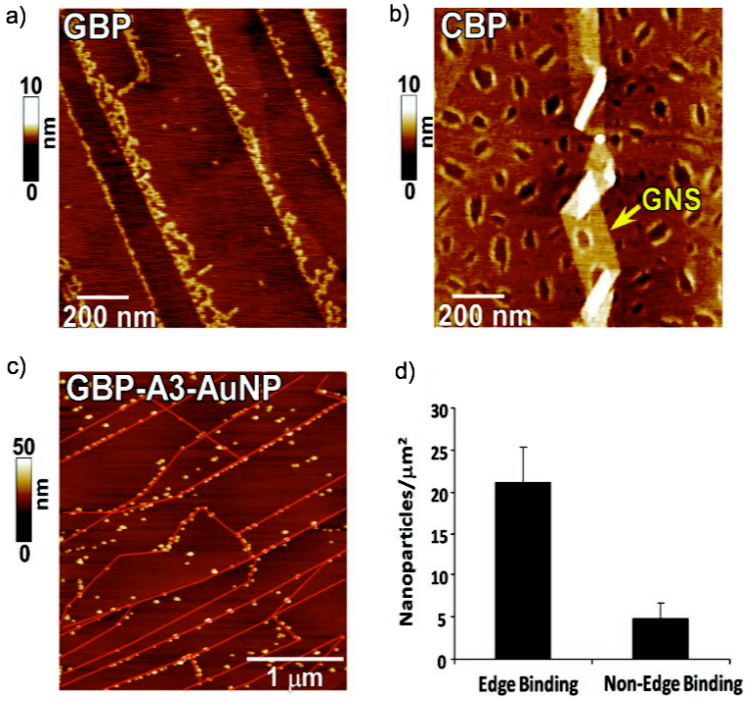


Figure 1.26. AFM images of P2 adsorbed onto graphene edges (a), P1 adsorbed onto graphene plane (b) and P2 coated with AuNPs adsorbed onto graphene edges (c). Bar plot of the number of AuNPs-coated P2 non-covalently bound to graphene edges or non-edges regions. Adapted with permission from ref. 157. Copyright 2011, American Chemical Society.

This nicely show how the structural information encoded in the aminoacidic sequence can also influence the polypeptide behavior adsorbed on GNMs, introducing an element of control in the formation of tailored biomimetic surfaces. Clearly, the different aromatic content of P2, the chain length and the vdW contribution explain the graphene plane selectivity of P1. The presence of the aromatic anchoring motif induces a stronger binding as further demonstrated by coarse grained MD simulations.[148] The shorter P2 eptapeptide, lacking of the aromatic anchoring residues, results more flexible and finds its stability at the edges of the surface where can entertain non-covalent stabilizing interactions.

Moreover, a specific aminoacidic sequence can rule and guide the self-assembly of a short peptide onto graphitic surfaces achieving well-organized biostructures exploitable for technological purposes. Noteworthy example was reported by Sarikaya and co-workers, who studied by AFM microscopy the graphite induced self-assembly of another phage display identified dodecapeptide (GrBP).[159] GrBP sequence was characterized by three distinct chemical domains constituted of

the hydrophobic domain I, the hydrophilic domain II and the aromatic domain III (Figure 1.27a). The observed uniform organization that this peptide is able to achieve adsorbed onto graphite was reached by a stepwise process that, monitored by systematic time lapse AFM images acquisition, reveals that the GrBP undergoes an initial binding, followed by an amorphous aggregation that unveils in an ordered phase (Figure 1.27b-c).

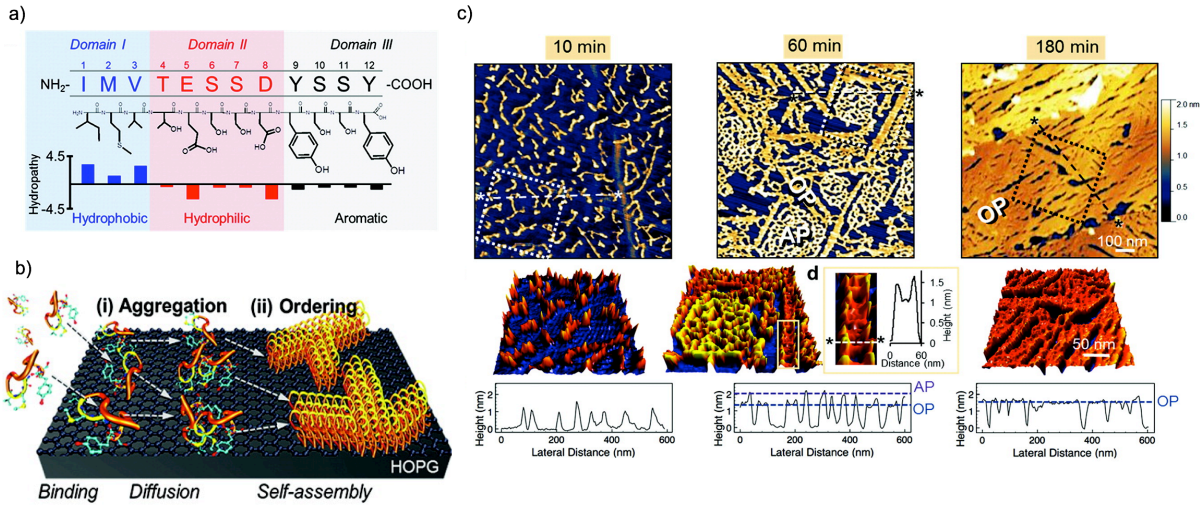


Figure 1.27. Chemical structure of GrBP characterized by hydrophobic domain I, hydrophilic domain II and aromatic domain III (a). Schematic representation of the stepwise adsorption and self-assembly of GrBP onto graphite (b). (c) Time-lapse AFM images of GrBP assembly after 10, 60 and 180 min of incubation. 3D and classical height analysis as a function of the phase formation. Adapted with permission from ref. 158. Copyrights 2012, American Chemical Society.

To understand in depth the adsorption and self-assembly process, rational mutations of the peptide chain enable the identification of the key residues governing the organization. Starting with the mutation of the aromatic domain (YSSY), the Tyr aa were replaced by an Ala in mutant 1 (M1), knocking out the aromatic contribution, which was then modulated when Tyr was instead substituted by a Trp and Phe in M2 and M3, respectively. M1 completely eliminated the binding capability of GrBP, first demonstrating the crucial role of the aromatic domain in the first phase of binding (Figure 1.28a).

On the other hand, M2 displays a porous network formation assuming a correct binding, while cluster of reduced size were obtained with M3 probably due to the inefficient self-assembly. Additionally, the aggregation rate, dynamics of coverage

and the calculated binding affinity reveal that GrBP attains the best self-assembled organization onto graphite over M2 and M3 (Figure 1.28a).

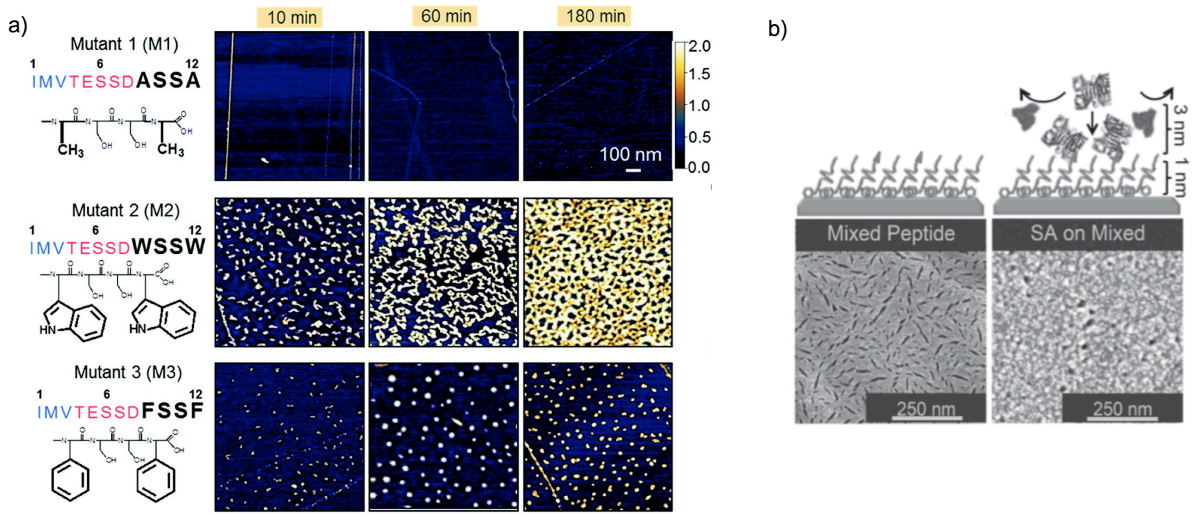


Figure 1.28. (a) Time lapse AFM images of mutant M1, M2 and M3. (b) Schematic representation of mixed self-assembled monolayer of GrBPs on graphene (top) and AFM images (bottom). On the right panel of image c) the self-assembled monolayer was tested for the detection of streptavidin in the presence of albumin buffer. Adapted with permission from refs. 158 and 159. Copyrights 2012, American Chemical Society and 2014, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

In a second avenue, when the hydrophobic domain was knocked out and replaced by hydrophilic residues, M4 results in a fully hydrophobic peptide that maintains its ability to bind graphite but it was unable to reach the high coverage and the correct folding. This meant that the amphiphilic character, gained by the combination of domain I and II, results crucial for the ordering onto the surface. As a countercheck, a similar amphiphilic mutant was tested (M5) unveiling comparable assembly on the graphitic surface.[159]

Based on this concept, the authors designed an upgrade version of GrBP that, organized in dense ordered self-assembly onto graphene, can work as a biosensor exposing a molecular probe able to selectively target a specific protein (bio-GrBP).[160] Furthermore, by producing a mixed self-assembly of bio-GrBP and SS-GrBP mutant with proved anti-fouling properties, they were able to detect specifically streptavidin protein (SA) in a bovine serum albumin buffer (Figure 1.28b).

From the analysis sieved so far, it was found out that polypeptide structures can interact with GNMs, such as fullerene, CNTs and graphene trough hydrophobic or

π-π stacking fashioned interactions occurring between aromatic, hydrophobic or positively charged aa and the different aromatic graphitic surfaces. Although the identified trend looks similar in the three cases, comparative studies have highlighted some differences in the dynamic behavior of the same polypeptide chain facing the different shapes characterizing the GNMs understudy herein. As it has been clarified by the abovementioned examples, the aminoacidic sequence as well as the GNMs shape and size can reasonably influence the approach to each other. Moreover, the intrinsic conformational freedom that short proteins have before reaching a stable and favorable folded state deserves special attention in this context.

In 2003, in the pioneering work by Wang *et al.* CNTs selective polypeptide sequences were identified (B1-B4),[117] which arise the interest in investigating further their binding ability both on the tubular framework and on the flat graphene platform. A combination of docking studies and explicit solvent MD simulations (OPLS-AA force field) were carried out by Gianese *et al.*[161] aimed at investigating the adhesion process of the B3 phage display-identified polypeptide[117] on both CNT and graphene surfaces. The conformation of the free peptide in water solution reveals a folded state stabilized by intramolecular H-bonds with a free energy of 3.32 kcal mol^{-1} more stable as compared to the unfolded state. B3 was then docked onto both the graphitic surfaces individualizing the best-scored complexes, subsequently characterized by MD simulations. While interacting with the CNT, B3 is shown to readapt its conformation according to the geometry of the tubular framework. Notably, a considerable increase of the contact area with the nanomaterial is observed as compared to the starting point (from 298.7 to 433.2 $Å^2$) corresponding in parallel to an increase of the free energy of binding (from -8.41 to -9.02 kcal mol^{-1}). The stabilization of the complex is accompanied by a further increase of the number of intramolecular H-bonds (from 6 to 11) which, enhancing the structure stability, improves the binding affinity as well. Interestingly, His and Trp were found the most interacting residues from the calculated free energy, even though resulted from a visual inspection non-parallel to the surface. The interaction of B3 and the CNT was clearly characterized by a reorganization and refolding process of the polypeptide over the tubular network, indicating that the strength of this biomolecule stays in its capability to adapt to the surface that interfaces.

Remarkably, the same trend is evidenced when B3 is simulated adsorbed onto the flat surface of graphene: increase of the contact area (from 346.3 to 410.2 $Å^2$) is

coincidental with a considerable increase of the binding energy (from -7.32 to -9.34 kcal mol^{-1}) and the enhancement of the number of H-bonds (from 6 to 12). Once again, the major contribution comes from the aromatic His and Trp aa, this time found lying parallel to the graphitic surface. The main difference concerns the lack of conformational rearrangement on graphene, the simulation of which shows a more stable protein backbone Root Mean Square Deviation (RMSD) compared to the simulation with the CNT (0.76 and 2.07 Å, respectively). The comparable results obtained with the two GNMs unveil the versatility of this polypeptide exerting a similar binding affinity for the nanomaterials understudy but with a different dynamic of adsorption. The importance of the peptide flexibility, which supports multiple strong binding conformations, was further stressed.

Similarly, Walsh and co-worker applied their validated implementation of the AMOEBAPRO force field to study the interaction of the strong binders polypeptides B1 and B3[117] both adsorbed on CNT and graphene.[162] The aromatic Trp, peculiarity of the two peptides and responsible of the binding to CNT, was systematically mutated with Tyr (B1Y and B3Y) and Phe (B1F and B3F) and simulated adsorbed on the two nanomaterials. The authors proved that the chosen polarized force field was able to successfully describe non-covalent interactions such that they decided to treat implicitly the solvent contribution.

It clearly emerged that the mutation of Trp with either Tyr or Phe exponentially decreases the normalized interaction energy (interaction energy between the peptide and surface divided by the number of the atoms in the peptide) of the mutants with both CNT and graphene. Indeed, monitoring the radial distribution functions (RDFs) of each residue within the simulated systems, it was found a high distribution of Trp close to the nanomaterials, whereas the distribution of Tyr and Phe mutants was found abundant far from the surface. Additionally, while B1 and B3 display aromatic residues preferentially parallel to the surfaces as the ring-tilt angle demonstrates, all mutants show less ring orientation stability. The aromatic aa ring-tilt angle could indeed vary from 0° (parallel) to 90° (perpendicular to the surface), assuming a continuous rotation of the peptide aromatic rings at the interface (Figure 1.29a-c).

It is worth noting that all the aromatic residues of the peptides understudy hold increased orientation stability on graphene in comparison with the tubular framework (Figure 1.29a-c). However, the aromatic orientation is not the only criterion

influencing the stabilization of the peptide/nanomaterial complex. In fact, the decrease of binding energy is observed even when the aromatic mutants are displayed parallel to the surface. As example, the authors provided pictures clearly showing a weaker-bound B1Y conformation that, though anchored to graphene through the aromatic Tyr, looses at the same time the contact of other non-aromatic aa and their plausible contribution to the binding energy (Figure 1.29d).

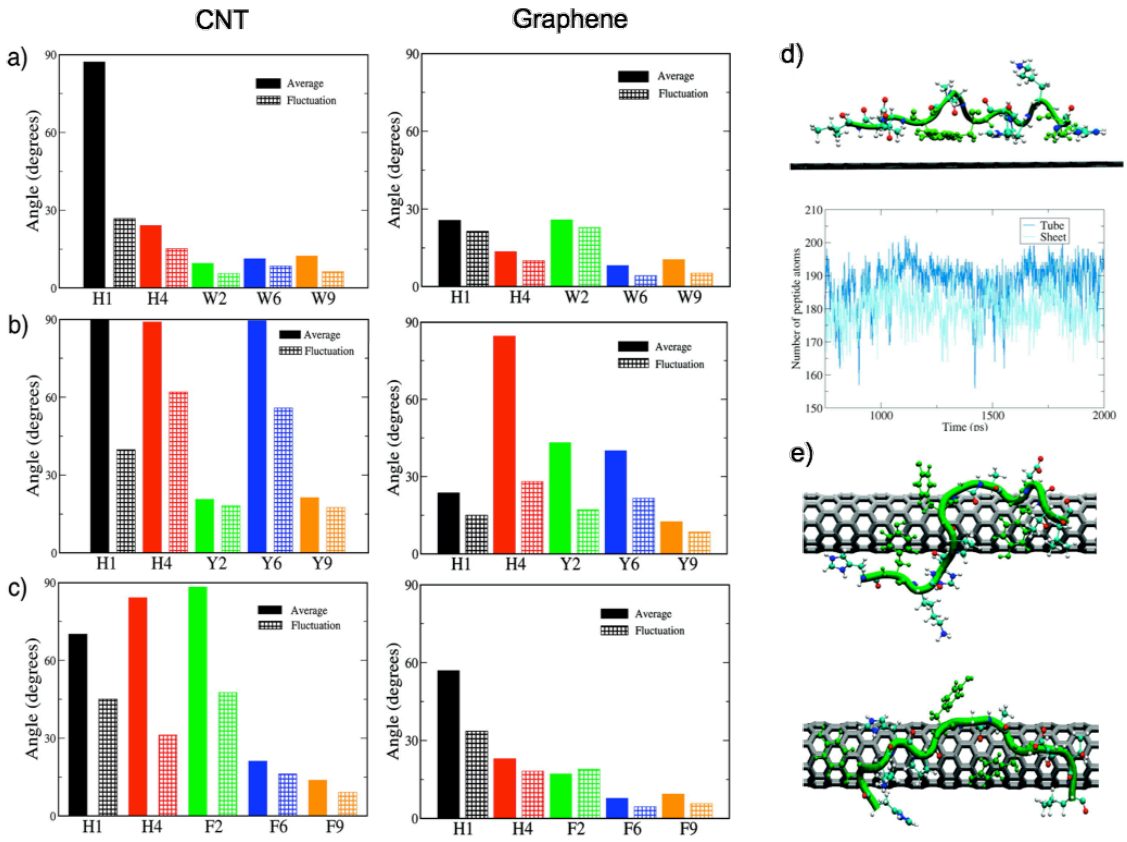


Figure 1.29. Average and fluctuation of the ring-tilt angle of all the aromatic aa within B1 (a), B1Y (b) and B1F (c) interfacing CNT and graphene. Structural representation of B1Y simulated adsorbed onto graphene (d). Number of atoms within B1 in contact with CNT (sky-blue) and graphene (turquoise) plotted as a function of simulated time (d, bottom). Snapshot of conformations of B1Y interacting with CNT from two different simulations (e). Adapted with permission from ref. 161. Copyright 2009, American Chemical Society.

Looking at the total number of peptide atoms in contact with the graphitic surface, a major number was found at the interface with CNT rather than with the flat graphene (Figure 1.29d). This is due to the higher flexibility of the B1Y, for instance, adsorbed onto the CNT whose different adopted conformations represent all possible and strong-binding orientations (Figure 1.29e).

In line with what stated by Gianese *et al.*,[161] peptide affinity can be partially ascribed to its ability to assume a variety of conformations at different interfaces. When the interface is represented by GNMs, no matter whether tubular or planar, the aromatic contribution governs the interaction ranked as follow Trp > Tyr > Phe.

As observed above, the protein flexibility can induce the adsorption of the same polypeptide chain to CNT and graphene with comparable binding affinity but with a different dynamic process. At the same time, recently Yarovsky and co-workers demonstrated how the same peptide can adsorb on different GNMs with different dynamics and binding affinity due to both the protein flexibility and the surface' size and shape.[163] The authors investigated the interaction between GNMs and an amyloidogenic peptide apoC-II(60-70) by means of classical MD simulations and DFT calculations to explore the relationship of carbon surfaces with amyloid fibrillation.[163] The binding affinity of this polypeptide was demonstrated to inversely scale with the increasing curvature of the nanomaterial, which from graphene to C_{60} has drastic reduction in the exposed surface prone to protein adsorption. During the 100 ns MD simulation of the 79 aa polypeptide onto each graphitic surface, the aromatic aa Tyr and Phe were found to act as anchor points regardless of the type of GNMs (Figure 1.30a).

Aside from this preliminary similarity, a different profile of the protein contact stabilities at the interface with the GNMs was figured out: clearly, apoC-II(60-70) was stably adsorbed onto graphene more than on CNT and C_{60}. For instance, when the polypeptide is adsorbed onto the smaller spherical GNM, it hardly anchors the C_{60} due to the small and curved surface available, whereas it undergoes intense conformational changes alternating between a β-hairpin arrangement where the aromatic aa can stack between them or with the carbon cage (Figure 1.30b). Conversely, Tyr and Phe allow for an elongated and extended conformation upon adsorption onto CNT and graphene (Figure 1.30b).

As clarified by DFT calculation employing the linear-scaling code ONETEP, C_{60}/peptide and CNT/peptide complexes are 125 and 41 kcal mol^{-1} less stable than the graphene/peptide adduct, respectively.

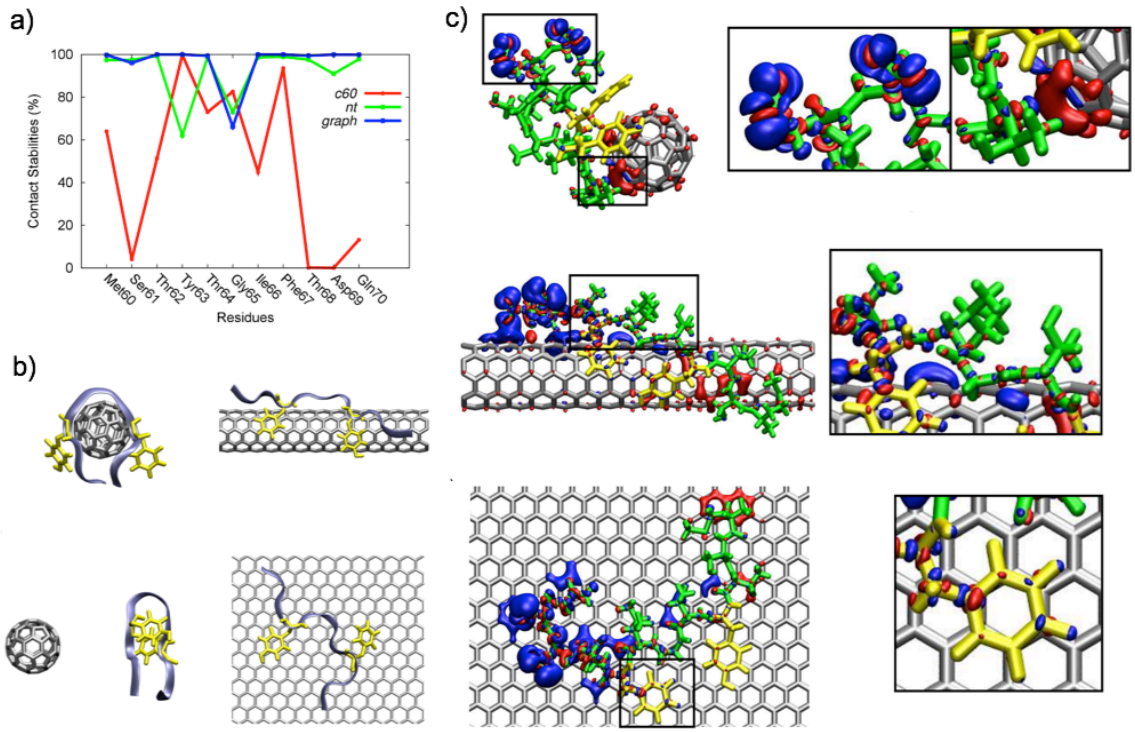


Figure 1.30. a) Contact stability of the residues within the apoC-II(60-70) structure interacting with C_{60} (red line), CNT (green line) and graphene (blue line). b) Favorable peptide conformation adsorbed onto the nanomaterial or desorbed as for C_{60}. c) The electron density difference maps shows the charge accumulation (red) and the charge depletion (blue) mapped on the complexes. Adapted from ref. 162.

In line with this, reducing the degree of curvature it increased the aromatic contact area calculated for each complex (0.92, 1.19 and 1.72 nm^2 for C_{60} CNT and graphene, respectively). Additionally, by mapping the electron density displaying in red the charge accumulation and in blue the charge depletion, intra-peptide electrostatic interactions influence the peptide's secondary structure and, thus, its binding affinity to the nanomaterial (Figure 1.30c).

Hence, the combination of the proteinic conformational freedom, such that of the polypeptide apoC-II(60-70), and the nanomaterial's curvature and size is demonstrated to yield a radically different way of interaction where the maximization of the contact area and the aromatic stacking are favored onto the GNMs.

However, polypeptide behavior can drastically change if another aminoacidic sequence is taken into account. For instance, Zuo *et al.* reported on the comparative study of the interaction between a small subdomain (35 aa) of the Villin Headpiece (HP) protein, HP35, and graphene, CNT and C_{60} by means of extensive MD

simulations.[164] During the 500 ns simulated time scale, the interaction with graphene was mainly driven by the aromatic residues within the HP35 sequence, namely the Trp and Phe, that were inducing the peptide conformation towards their better exposition and the attainment of π-π stacking with the planar graphitic surface. This was nicely supported by the contact probability as a function of the interaction energy calculated for each HP35 residues facing the three nanomaterials: all the residues are favorably interacting (negative energy values as reported in the plot in Figure 1.31) with graphene and in particular the aromatic Phe35, Trp23 and Phe10 aa exhibit more than 80% of contact probability (blue colored residues in Figure 1.31).

On the other hand, the tubular surface of CNT maintains a favorable contact with the aromatic residues (Phe35 and Trp23), whilst reduced in energy (less negative) and % (among 60 and 80 %, cyan colored residues) of contact probability (Figure 1.31), accompanied by the significant contribution of Lys30 and Gln26, which follow the curved geometry with their alkyl chains. Finally, the interaction energy of the residues adsorbed onto the curved C_{60} surface display hardly negative values and poor stacking ability of the aromatic aa with their 20% of contact probability.

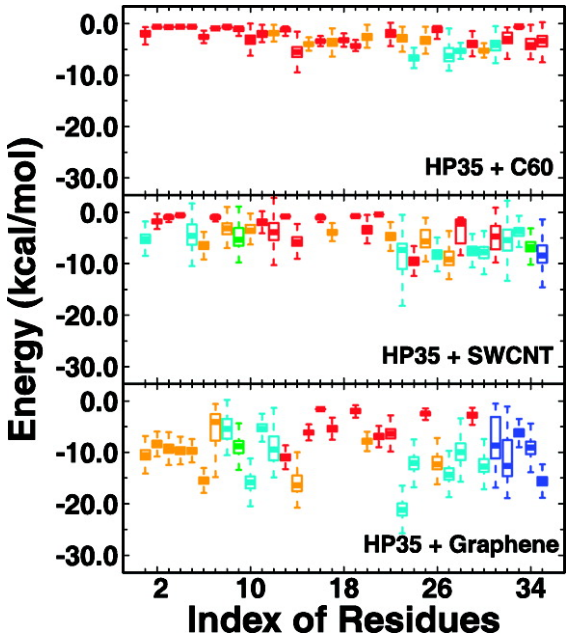


Figure 1.31. Interaction energy of each HP35 residue and C_{60}, CNT and graphene. The residues are colored according to the contact probability: 0-20% (red), 20-40% (yellow), 40-60% (green), 60-80% (cyan), and 80-100% (blue). Reprinted with permission from ref. 163. Copyright 2011, American Chemical Society.

Following the same comparative approach, recently Yao and co-workers studied the fibrillation process of the Islet amyloid polypeptide fragment 22-28 ($IAPP_{22-28}$) on graphene, CNT and C_{60} by means of extensive MD simulations.[165] This oligomer is known to aggregate in β-sheets like fibrils driven by inter-peptide hydrophobic interactions within its "NFGAIL" sequence. Hence, in order to reproduce the fibrillation process an increased concentration of $IAPP_{22-28}$ (4 and 8 peptides) was simulated alone and in proximity of the GNMs (Figure 1.32a-b).

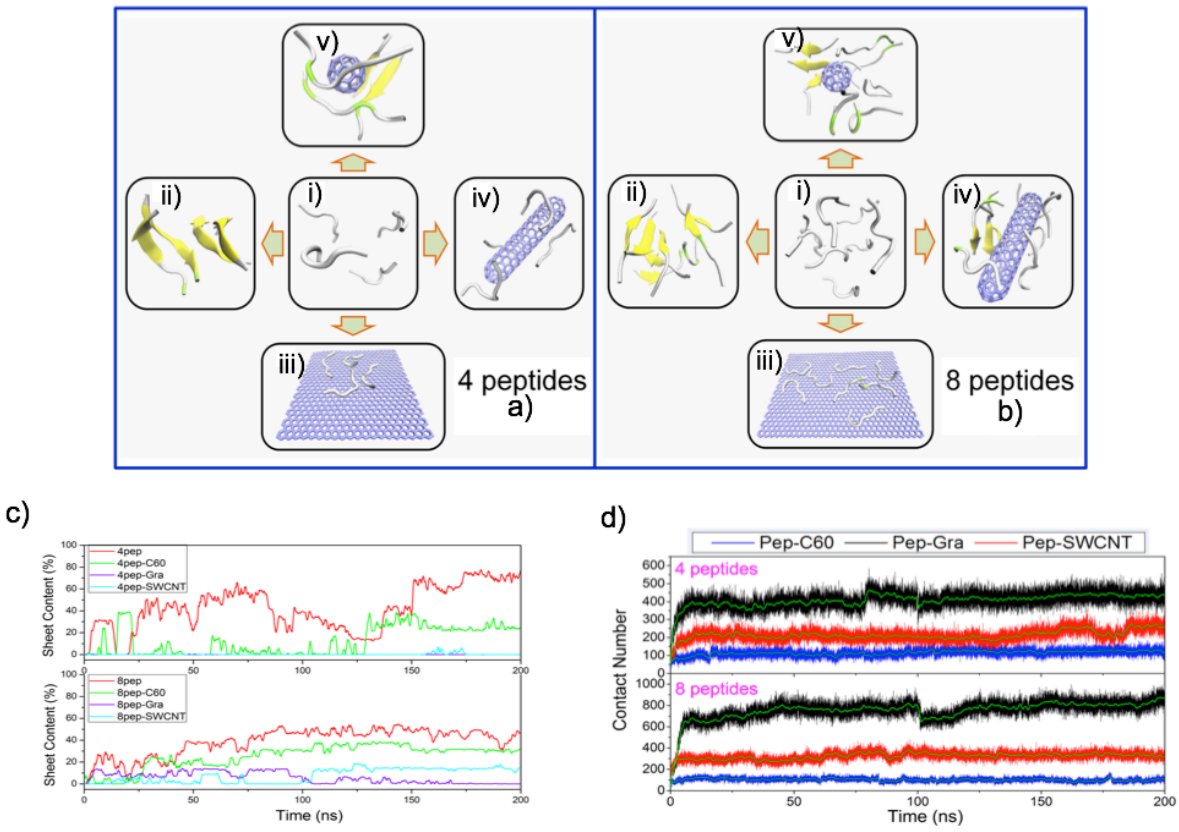


Figure 1.32. Schematic representation of the starting structures of 4 (a) and 8 (b) peptides alone (i). The conformations after 200 ns simulation of the peptides alone (ii) and in complex with graphene (iii), CNT (iv) and C_{60} (v) are reported. c) β-sheet content as a function of the simulated time is plotted for all the systems. d) Contact number as a function of the simulated time between the peptides and graphene (black lines), CNT (red lines) and C_{60} (blue lines) are reported. Adapted from ref. 164.

Notably, to the formation of fibrils corresponds the increase of β-sheets, whose content was therefore monitored along the 200 ns simulation time. Interestingly, a β-sheets content of 0% was found when 4 $IAPP_{22-28}$ peptides were simulated in the presence of the flat graphene and the tubular CNT (respectively violet and turquoise lines in Figure 1.32c). The same β-sheets content was found for

the 8-peptide system on graphene, slightly higher on CNT (20%). On the other hand, in the presence of C_{60} two peptides out of four were able to form β-sheets with a 74%, six out of eight with 81%. These behaviors find explanation in the major contacts that the $IAPP_{22-28}$ polypeptides are able to entertain on both graphene and CNT with respect to C_{60} rather than with themselves (respectively black, red and blues lines in Figure 1.32d). Furthermore, it is worth noting that increasing the number of peptide on C_{60} the number of contacts remains constant, stably oscillating around 100 (Figure 1.32d). Analyzing more in depth the nature of these contacts, it results that the aromatic (Phe) and hydrophobic (Ile and Leu) residues of $IAPP_{22-28}$, responsible of the inter-peptide interactions that induce the fibrillation, are now occupied in strong contact with graphene and CNT. Conversely, this is not verified with the spherical cage.

It is evident that the increased available surface of graphene and CNT favors hydrophobic and π-π stacking interactions with $IAPP_{22-28}$ that block the fibril formation, whereas C_{60}, unable to entertain interactions as much as the others GNMs, has a smaller inhibition effect. Additionally, the behavior of a single peptide (examples listed before) can vary as a function of its concentration.

Hence, to exhaustively describe the way of interaction between polypeptides and GNMs several aspects have to be taken into account, among the others the protein conformational flexibility, the aminoacidic sequence, the aromatic content and their exposition, the peptide concentration and of course the size and shape of the nanostructure surface.

1.7.3 Graphene and proteins

In the view of evaluating the bioapplicability and biosafety of graphene, several studies have been reported investigating the interaction of the flat 2D allotrope of carbon with numerous proteins, often chosen considering their importance, functionality and abundance in the human blood and body. As already stressed before, when it comes into studying the non-covalent approaching of proteins to GNMs, the tertiary structure of the formers certainly influences the dynamic of adsorption especially on the planar graphitic surface, where the available contact surface is much higher. The multiplicity of the protein repertory displays a variety of globular,

multidomains proteins made up of a different combination of secondary structure elements. Needless to say that proteins characterized by only α-helix rather than β-sheets or a combination of the two might interact differently with the graphitic surface. Furthermore, the distinction of the so-called "hard" proteins, those with peculiar internal structure stability, and "soft" proteins, those with low internal stability, reasonably might depict a different scenario in the protein/nanomaterial interaction.[166]

At this regards, Raffaini and Ganazzoli tackle the issue by performing several MD simulations comparing the adsorption onto graphite sheets of albumin,[167,168] as the most abundant blood protein, fibronectin,[169] as important protein involved in the blood-clotting cascade, and lysozyme,[170] an essential enzyme involved in hydrolysis of bacterial cell walls. The proteins understudies are not only relevant biomolecules involved in biological and pathological processes, but are also representative of diversified structural features that will allow to draw a picture of the interaction of globular proteins and nanomaterial. In fact, albumin is a globular protein predominantly characterized by α-helixes and the subdomains A and E chosen for the simulations counts 3 α-helixes each, with an amphiphilic character the former and a mean hydropathy value the latter; fibronectin type I module is another globular protein but consisting exclusively of antiparallel β-sheets, while lysozyme comprises both α-helixes and β-sheets in its globular structure.

The validated protocol proposed by the authors consists of a first minimization step of the starting structures adopting different orientations in the surface proximity; the best and the worst ranked orientations were then subjected to MD simulation by using consistent valence force field CVFF in implicit solvent, which allows to achieve reasonable results (as compared to explicit solvent simulations) in a very short time scale (up to 7 ns).[167,169,170]

According to their outcomes, during the minimization step, all the biomolecules tend to approach closer to the graphitic surface regardless of the type of secondary structure or of the starting orientations, revealing, at a visual inspection, partial loss of the secondary structure of the protein section in contact with the surface (Figure 1.33a). The worst interaction mode is often the one oriented with the long axe of the protein perpendicular to the surface (fibronectin and lysozyme Figure 1.33a bottom).

Interestingly, in the three cases the number of contact residues linearly correlates with the interaction energy, $E_{int} = (E_{free\ prot} + E_{CNT}) - E_{tot}$), and the protein strain energy ($E_{strain} = E_{adsor} - E_{free}$) calculated for the minimized geometries (Figure 1.33b). Additionally, the best-fit line through the origin of the E_{int}, energy required to desorb the protein from the surface, and the E_{strain}, measure of the protein deformation compared to the native state, were also derived.

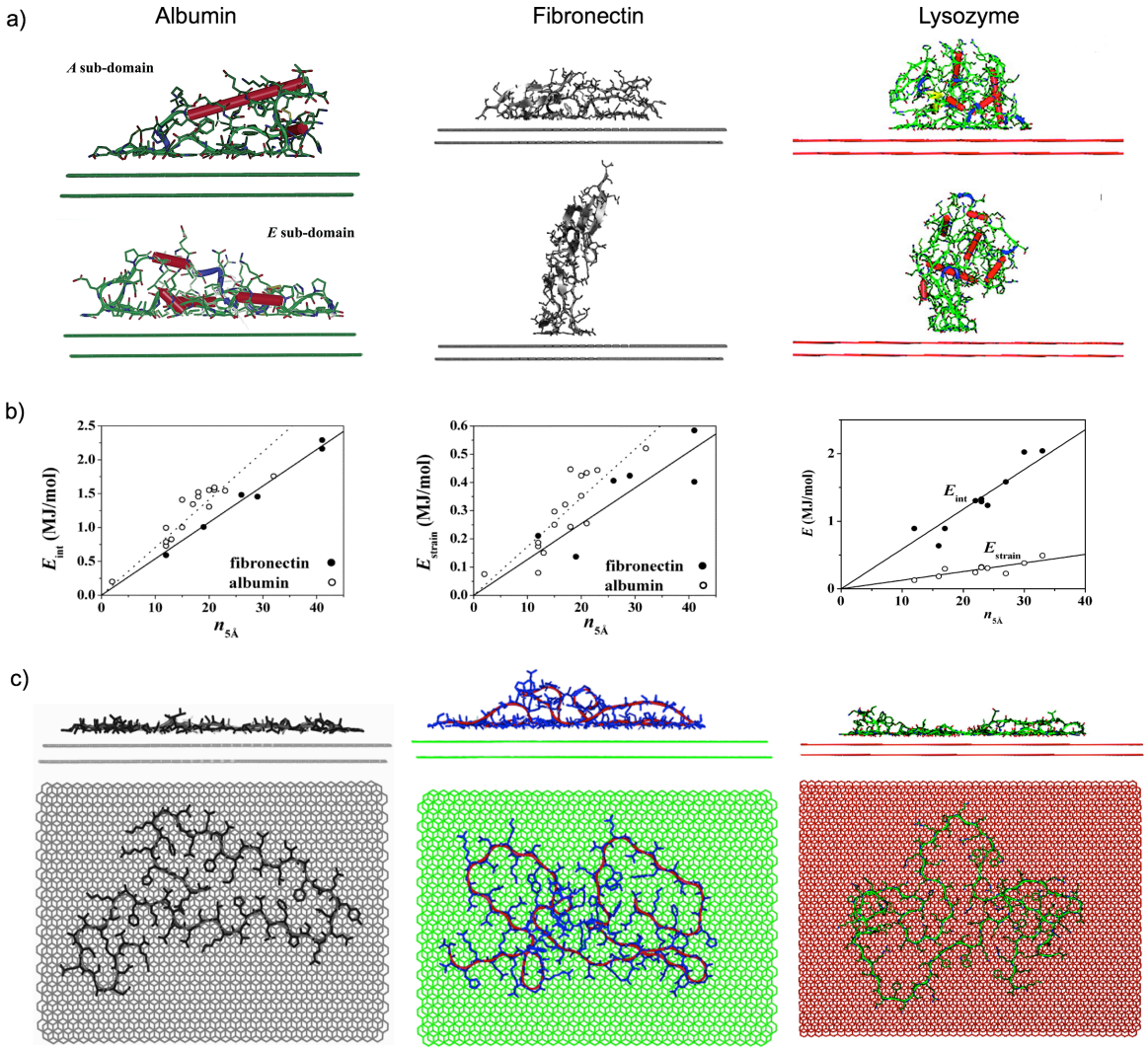


Figure 1.33. a) From left to right: minimized orientations of two subdomains of albumin, the best (top) and the worst orientations (bottom) of fibronectin and lysozyme. b) Linear correlation between the E_{int} and E_{strain} and the number of contact aa ($n_{5Å}$). c) Side and top view of the final adsorbed proteins after 1ns MD trajectory. Adapted with permission from refs. 167, 168 and 169. Copyrights 2003, 2004 and 2010, American Chemical Society.

E_{int} of 71, 54 and 59 kJ mol^{-1} were obtained for albumin, fibronectin and lysozyme respectively, suggesting a major adsorption of the α-helical globular

albumin followed by the mixed α-β lysozyme and the β-sheets fibronectin. On the other hand, E_{strain} of 17, 13 and 12.7 kJ mol^{-1} were obtained for albumin, fibronectin and lysozyme respectively, implying larger structural rearrangements occurred for the α-helical globular protein.

Investigating further the dynamic of adsorption by performing 1 ns MD simulation, the globular proteins proceed towards the complete adhesion on the planar network, inducing intense structural rearrangements to maximize the contact points (Figure 1.33c). The number of albumin aa in contact with the surface significantly increases passing from a maximum of 23/60 in one of the minimized orientations, to the complete aa contact resulting in a monolayer of aa coating the surface (59/60 aa at less than 5 Å far from graphite).[167]

Similarly, fibronectin from an initial maximum contact of 41/93 aa, adsorbs 68/93 aa onto the surface and no β-sheets are retained (Figure 1.33c). Nonetheless, a unique monolayer was not observed in the adsorption of fibronectin, which presents a net of disulfide bonds that holds together the structure preventing it from the complete spreading.[169] This divergent behavior can also refer to the different nature of albumin and fibronectin, considered as a soft and hard proteins,[166] correspondingly.

On the other hand, the lysozyme adsorption was always occurring through the β-sheets side which, probably holding higher affinity to the flat geometry, is the first to adsorb and lose its secondary structure information.[170] In this case, the adhesion process was followed by the radius of gyration (Rg) of the whole molecule and by the diagonalization of the Rg tensor in its component x, y, z which correspond to the parallel (Rg_x and Rg_y) and perpendicular (Rg_z) directions of the protein to the surface. Remarkably, while Rg_x and Rg_y were increasing during the simulations of the two orientations of lysozyme (from 7.13 and 6.75 Å to 21.67 and 14.56 Å, respectively), the Rg_z was visibly reduced (from 10.18 Å to 2.05 Å considering the least favorable selected orientation), meaning a full parallel spreading of lysozyme as well at the end of the simulation (Figure 1.33c).[170]

As a general trend, all the globular proteins analyzed by Raffaini *et al.* tend to favorably adsorb onto the graphitic surface with a two-steps process: a first adsorption phase where one side of the protein approaches the nanomaterial with local rearrangement of the secondary structure, followed by a complete surface coverage with the formation, in some cases, of one aa monolayer lying on the graphitic network with no secondary structure retention.

In line with the aforementioned reports, another comparative study was recently carried out by Guo *et al.*, which performed extensive MD simulations (AMBER ff03 force field) in explicit water to model the adsorption of WW domain (full β-sheets), BBA protein (mixed α-helix and β-sheets) and λ-repressor (α-helix) on the graphene surface.[171] Placed 10 Å far for the graphitic platform, after 100 ns the three proteins, for which two possible orientations were simulated each, were tightly adhered to the surface (Figure 1.34).

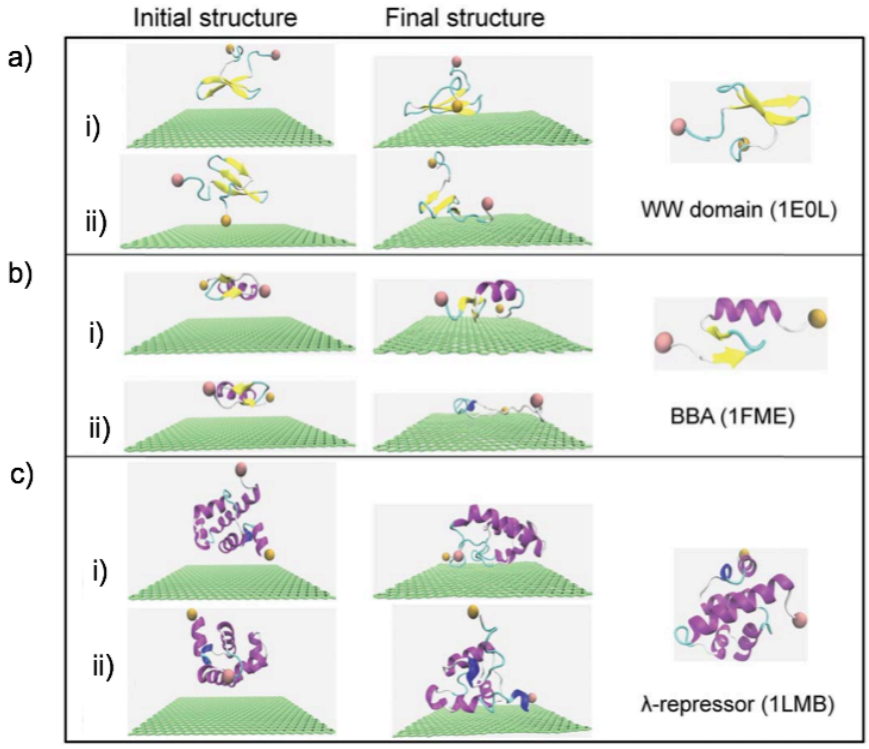


Figure 1.34. Starting and final geometries of two orientation (i and ii) of WW domain (a), BBA (b) and λ-repressor (c) after 100 ns MD simulations. Adapted from ref. 170. Copyright 2014, Royal Society of Chemistry.

As a matter of fact, all the three proteins similarly display an increasing contact area along the simulation time as a sign of their adsorption and stabilization at the interface with the nanomaterial.

Nevertheless, a more in depth analysis reveal some structural and dynamical differences. In fact, the energetic profile reveals that the proteins interacting with the α-helix side, namely BBA in the ii) orientation (Figure 1.34b-ii), and both orientations of λ-suppressor, bind with higher binding energy (-151.65, 136.77 and -119.10 kcal

mol^{-1}, respectively) compared to the full β-sheet WW domain and the β-sheets orientation of BBA protein, the energies of which are indeed lower and similar among them (-86.56, -79.36 and -86.49 kcal mol^{-1}).

Moreover, by looking at the energetic decomposition, a major vdW contribution can be appreciated produced by the actual interacting residues, whose different location within the six simulated complexes depends on the related orientation (magenta squares in Figure 1.35).

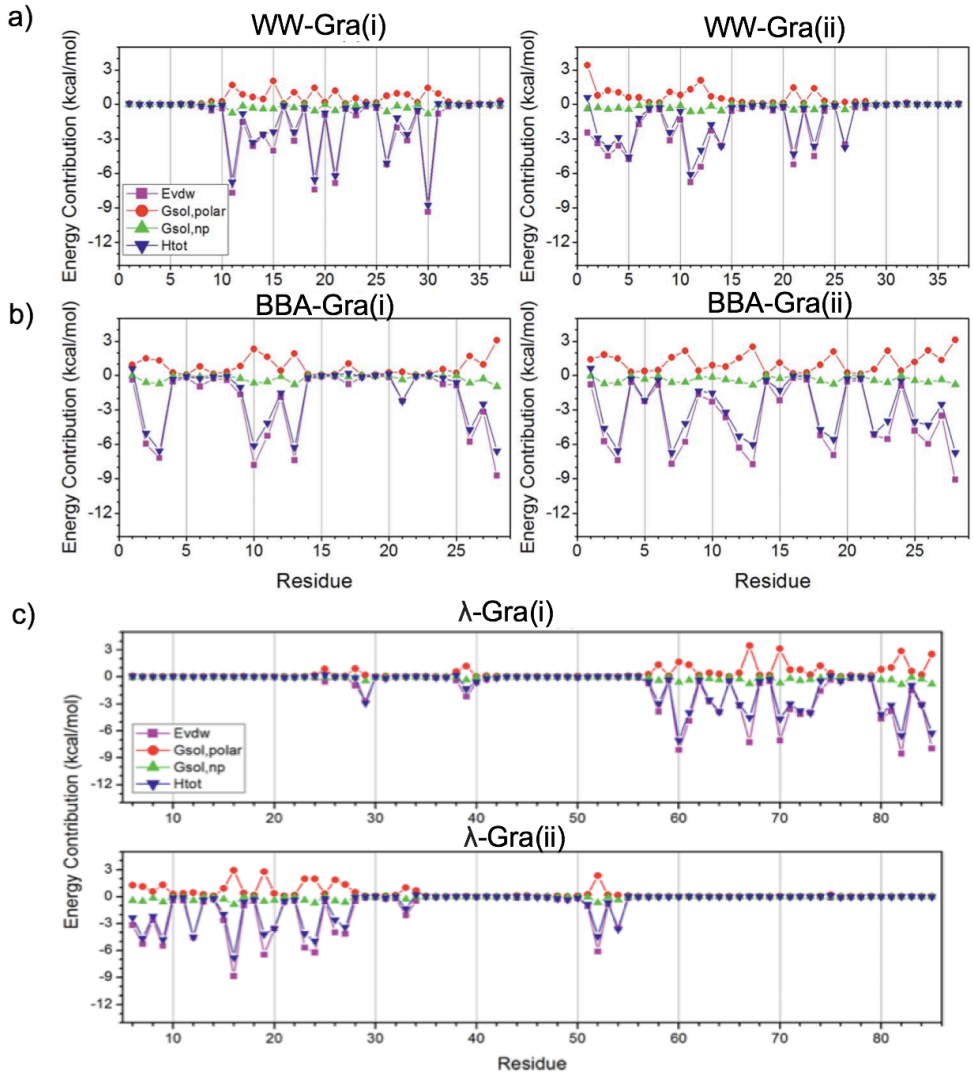


Figure 1.35. Interaction energies plotted as a function of the simulated time for WW domain (a), BBA (b) and λ-repressor (c) in the two orientation i) and ii). Adapted from ref. 170. Copyright 2014, Royal Society of Chemistry.

For instance, while only the residues at the two extremities are shown to contribute the most for BBA-gra(i) complex, a major number of residues distributed allover the protein in BBA-gra(ii) orientation are found to actively interact with the surface, thereby explaining the higher interaction energy displayed by this orientation (Figure 1.35b).

Interestingly the maximization of the contact points is accompanied by different secondary structure dynamics: when the full β-sheets WW domain is adsorbed onto graphene no secondary structure loss is detected for both the orientations (sheet content is 32.43% at the beginning and 29.73 and 35.14% for the two final steps, respectively) as also observed when the mixed α-β BBA is interacting with graphene throughout its β-sheets side (β-sheets content is 14.29% for both initial and final conformations, α-helix 35.71% and 32.14% at the beginning and at the end, respectively). Conversely, BBA adsorbed with the α-helix side considerably loses its secondary structure (β-sheets content drops from 14.29% to 0 and the α-helix content from 35.71% to 10.71%), alike the λ-suppressor behavior, which partially loses its α-helix content (from 67.50% to 52.50% and 50% for the two orientations, respectively). Therefore, the orientations interacting more strongly to the graphitic surface are the ones adsorbed through the α-helix portion of their structure, which more easily undergoes structural rearrangements to maximize the contact points and stabilize the interaction.

Hence, in a scenario in which globular proteins spontaneously interact and tightly adsorb on the flat graphitic surface, it results a major stability of the β-sheets elements over the α-helix.[171] In the adsorption process of globular proteins, considerable structural rearrangements are envisaged causing an increase of the molecular fingerprint along with a loss of the secondary structure folding.[172]

Additionally, it has been demonstrated that adsorption of globular proteins induces a flat coating over graphene with a particular parallel orientation among the strands,[173] assuming an active role of the graphitic template in the amyloid fibrillation.[46]

In parallel, a notable work, aimed at exploit the β-sheets stabilization and organized assembly directed by the flat graphitic surface, was reported first by Brown *et al.* whom experimentally demonstrated the patterned adsorption and arrangement of a *de novo* designed protein onto graphite.[174] AFM images clearly proved that the

deposition of an amphiphilic β-strands on highly ordered pyrolitic graphite (HOPG) leads to the nucleation of fibers with 120° orientated symmetry induced by the hexagonal graphitic template. FTIR probed the antiparallel β-sheets formation with the typical peak at 1621 cm^{-1}.

Furthermore, Wang and co-workers reported on the patterned adsorption of peptides organized in β-sheets upon deposition on the flat HOPG surface with a particular focus on the peptide sequences.[175] In particular, they studied the behavior of two synthetic model peptides, namely $R_4G_4H_8$ and $F_4G_4H_8$, adsorbed onto HOPG by STM images and rationalized the experimental evidences with theoretical calculations.[175] The two peptides differ in their sequences by the presence of Arg (R) and Phe (F) aa, which are both followed by four Gly (G) residues and eight aromatic His (H). Interestingly, the STM images revealed not only the formation of extended aligned lamellae (Figure 1.36a-b), but gave also significant insights at the atomistic level.

The authors found that each peptide strand, aligned in the β-sheets organization, is characterized by three bands with different brightness and lamellae width that nicely account for the corresponding sequence (Figure 1.36a-ii and b-ii): in the $R_4G_4H_8$ sample, the first band displays a width of 1.2 nm for the 24.5% of the full-length peptide, the second band a width of 1.0 nm for a 25% and the third a width of 2.4 nm for the 49% of the full-length peptide.

The width values (1.2:1:2.4) and the % proportion (24.5:25:49) perfectly match with the $R_4G_4H_8$ sequence. Similarly, a proportion of 1.3:1:2.4 nm corresponding to the 28.6:22.4:53% of the total length peptide was observed for $F_4G_4H_8$ peptide. Furthermore, from the STM images was clearly noted that each strand, organized in a parallel orientation as demonstrated by the FTIR, are not perfectly aligned but shifted probably due to inter-peptides interactions. By normalizing the brightness intensity to the reference residue H_8, it was possible to rank the four aa following the order Phe > His > Arg > Gly.

For a deep comprehension of the conformational and dynamics of adsorption, MD simulations aided the experimental observations, ultimately confirming the atomistic details drawn experimentally. 100 ns MD trajectories were therefore carried out simulating ten strands of $R_4G_4H_8$ and $F_4G_4H_8$ organized in parallel β-sheets onto graphite surface.

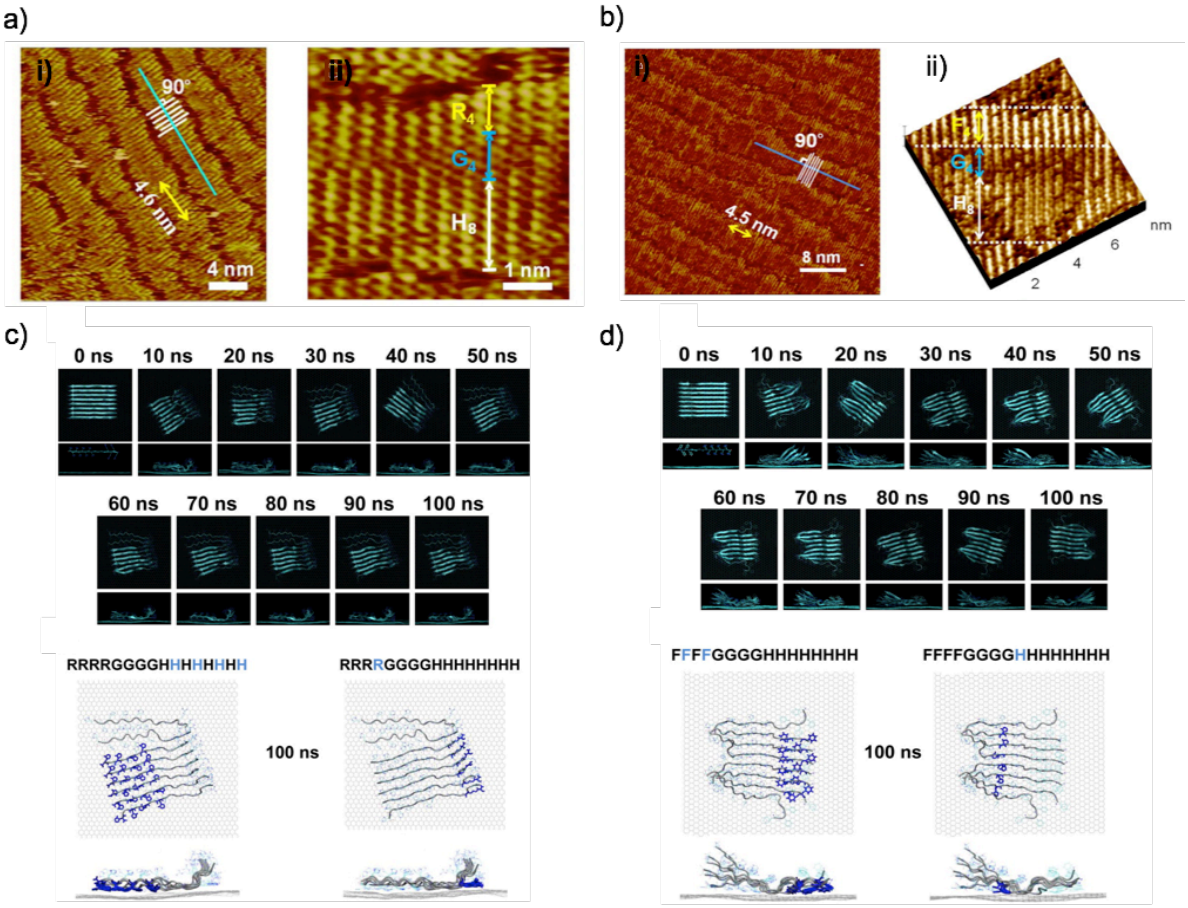


Figure 1.36. STM images of the $R_4G_4H_8$ (a) and $F_4G_4H_8$ (b) aligned on HOPG surfaces. The three bands with different brightness and length can be distinguished corresponding to the relative sequences (ii). MD simulation snapshots of the $R_4G_4H_8$ (c) and $F_4G_4H_8$ (d). Adapted with permission from ref. 174. Copyright 2013, American Chemical Society.

The β-sheets were found to stably interact with the graphitic surface with a slightly shifted alignment showing the aromatic residues anchoring the platform in a π-π stack, while the positively charged Arg interact through their alkyl side chain pointing towards the solvent the guanidinium group (Figure 1.36c-d). Additionally, the interaction energy perfectly correlate with the brightness intensity detected by the STM images following the trend Phe > His > Arg > Gly of the most interacting residues.

A practical application, exploiting the high affinity of protein for the planar graphitic surface, was reported in the work by Laanksonen *et al.* that employed the amphiphilic microbial adhesion protein HFBI, known to strongly adsorb onto hydrophobic surfaces, to non-covalently bind graphite sheets and promote their exfoliation upon sonication.[176] The method efficiently produces single graphene

sheet, as demonstrated by TEM, SEM and Raman measurements. Moreover, to prove the retained activity of the adsorbed protein, variations of HFBI protein were produced and employed in the validated exfoliation procedure. The exfoliated flakes with the modified protein were then labeled through AuNPs able to bind the modified HFBI, as TEM and AFM images nicely confirmed.

So far we highlighted the stable and spontaneous interaction of globular as well as planar proteins with the flat graphitic surface undergoing structural rearrangements that, involving both the tertiary and secondary structures, are directed to the maximization of the whole molecule contacts with the surface. However, having stated the general affinity of proteins to non-covalently attach CNTs as well (see section 1.6.3), a comparison is now required to evaluate the effect of the curvature and topography of the nanomaterial on the interaction strength and on the dynamics of adsorption of proteins on the GNMs surfaces.

At this regard, Raffaini *et al.* carried out a comparative theoretical study evaluating the behavior of the same model proteins probed on graphite,[167,169] namely albumin and fibronectin, now adsorbed on different diameter CNTs and graphene,[177] by applying the same protocol as the aforementioned works. After the minimization step, both albumin and fibronectin adhered to (8,8) (Figure 1.37a), (10,10)CNTs (Figure 1.37b) and graphene (only albumin Figure 1.37c) with local structural rearrangements at the interface with the graphitic surfaces.

The same linear correlation was found between the E_{int}/E_{strain} and the number of contact residues, while the best-fit line through the origin describes significant differences among all. Since the authors did not simulated fibronectin on graphene, in a view of a comprehensive comparison only the energetic values of albumin are now analyzed. Interestingly, the adsorption of albumin onto the one sheet graphene yields similar E_{int}/E_{strain} compared to the simulations on graphite (54.7/11.3 and 57/17 kJ mol^{-1}, respectively), whereas lower values were found for the albumin adsorbed on CNT with different diameter (30.7/9.2 and 33.8/9.2 kJ mol^{-1} on (8,8) and (10,10)CNTs, respectively). After MD simulations, an increased coverage was achieved by both proteins on (8,8) (Figure 1.37d), (10,10)CNTs (Figure 1.37e) and on graphene (only albumin Figure 1.37f) with evident loss of the secondary structure information of the biomolecules spread on the surfaces.

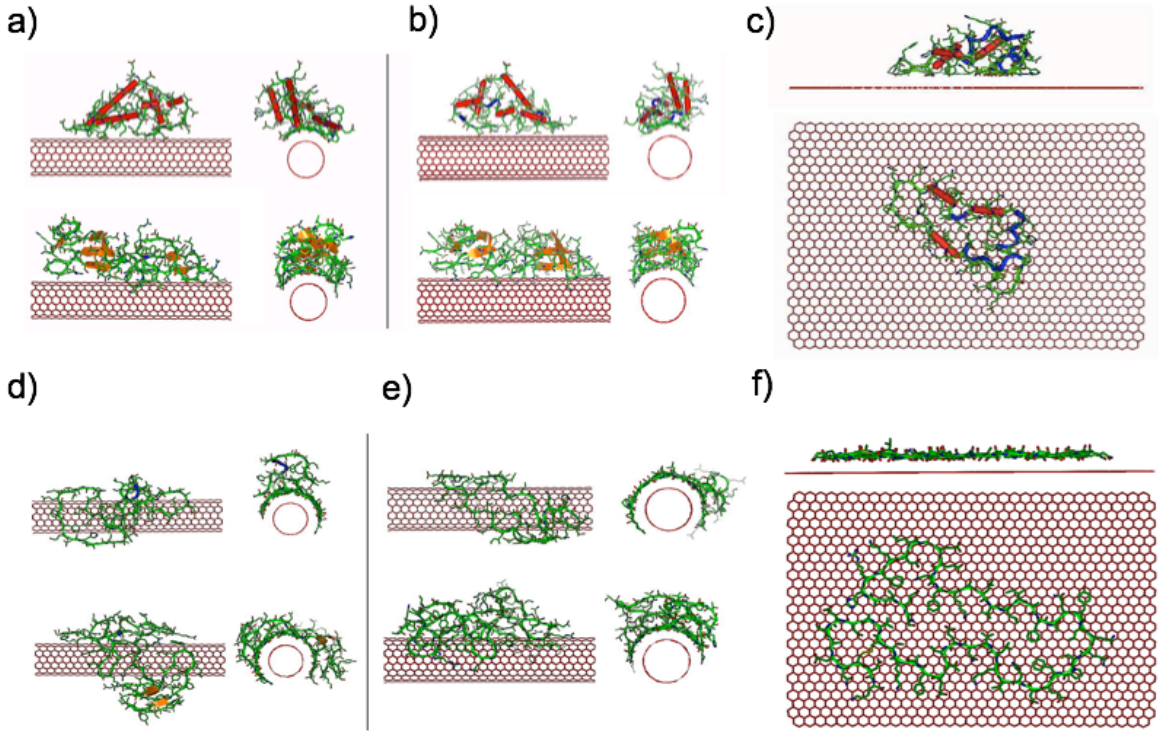


Figure 1.37. Minimized orientations of albumin subdomain and fibronectin adsorbed on (8,8) (a), (10,10)CNTs (b) and graphene (only albumin, c). The most stable orientations of albumin subdomain and fibronectin adsorbed on (8,8) (d), (10,10)CNTs (e) and graphene (only albumin, f) after MD simulations. Adapted with permission from ref. 176. Copyright 2013, American Chemical Society.

Remarkably, the energy values of E_{int} are hardly changed for the albumin/graphene complex (intrinsic E_{int} 53.3 kJ mol^{-1}) compared to the initial adsorbed geometry, while are significantly increased for albumin on (8,8) and (10,10)CNTs (intrinsic E_{int} 44.8 and 43.9 kJ mol^{-1}) as a measure of the increased interacting points. It is worth noting that very similar results are obtained for the albumin subdomain on either the CNTs, certainly due to the small difference between the diameters taken into account, but especially to the 'soft' nature of this globular protein whose low internal stability allows it for easier rearrangements and adaptations to the different graphitic surfaces. Conversely the 'hard' fibronectin more visibly differs in the intrinsic E_{int} values that vary from 38.8 to 47.0 kJ mol^{-1} once adsorbed on (8,8) or (10,10)CNTs, correspondingly.

The comparable values obtained for graphene and graphite confirm that without changing the topography of the nanomaterial similar interaction behaviors are yielded, while for the interaction of the same proteins on tubular frameworks, such as

CNTs, lower interaction strength is observed. Additionally, different degree of curvature clearly displays different adsorption pathways according to the intrinsic stability of the protein considered.

1.8 Outline of the Thesis

In this chapter, a detailed overview of the theoretical and experimental studies carried out to unravel the non-covalent interactions between graphitic nanomaterials (GNMs, such as fullerene, CNTs and graphene), and proteinic biomolecules has been given. By analyzing in a systematic way the modality through which amino acids, polypeptides and folded proteins interface the abovementioned GNMs surfaces, the mutual effects of both protein structural complexity and nanomaterial topography were considered. Going through the most recent and representative examples in the field, an exhaustive picture of the chemico-physical and dynamical aspects ruling the spontaneous tendency of such biomolecules to adhere to the graphitic networks has been provided. The importance of the deep understanding of the protein/GNMs interface originates from the increasing interest in exploiting the outstanding nanomaterial properties in combination with biomolecules as promising innovative tools in nano-biotechnology. As a matter of fact, the merging of biomolecules and nanomaterials sounds interesting as alarming at the same time: the combination of GNMs properties with molecules able to confer them functionality and biocompatibility clearly stimulates the scientific fantasy towards the ideation of new materials for biological applications, while toxicological aspects emerge as reasonable concerns. Hence, the comprehension of such protein/GNMs interactions would attain their contrive and conscious manipulation allowing to engineer new bioactive hybrid material avoiding toxicological consequences.

The aim of this thesis is therefore to master protein/nanomaterial interfaces by means of computational techniques (thoroughly described in Chapter 2) and, taking advantage of the concepts reported so far, present new bio-hybrids towards biological applications. We provide three main examples, reported in Chapter 3 and 4 of this thesis, where we apply theoretical calculations to study the interactions occurring within bio-hybrids systems, which are aimed at investigating and controlling the targeting and migration of cancer cells. Particularly, Chapter 3 is divided in two main

sections, entitled *"Cancer cells targeting via Antibody functionalized Fe-filled multiwalled CNTs"* and *"Design of Janus-type β-sheet for the hierarchical functionalization of graphene"*, which tackle the non-covalent interaction between protein and GNMs with different approaches.

The former deals with the computational study of the interactions occurring at the CNT-Antibody interface within a newly designed magnetic nanomaterial used as cancer cells targeting agents. The experimentally tested bioactivity of the conjugate is here validated by extensive classical and QM calculations, which clarify the structure-activity aspects of such bio-hybrid material. Furthermore, the way of interaction between big proteins and the tubular CNT structure was generally addressed.

The latter section of Chapter 3 focuses on the theoretical engineering of a Janus-type β-sheet, designed to attain the controlled, directional and non-covalent functionalization of graphene towards targeting applications. By simulating several β-sheet candidates adsorbed onto graphene, different interacting profiles were observed and the best candidate was selected. In a reverse approach, the theoretical design is here guiding the production of the β-sheet protein, that is now under experimental validation.

Finally, in Chapter 4, entitled *"Structural validation of peptide-based motogenic Au surfaces by means of MD simulations"*, we extended the biohybrid construction approach to investigate the cellular migration of cancer cells. A biomimetic model surface made of self-assembled motogenic peptide immobilized on gold (Au) surfaces was designed and tested experimentally in our group. Different organization and biological responses of the biomimetic surfaces were observed and are here rationalized by classical MD simulations.

1.9 References

[1] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, R. E. Smalley, *Nature* **1985**, *318*, 162–163.

[2] S. Iijima, *Nature* **1991**, *354*, 56–58.

[3] K. S. Novoselov, A. K. Geim, S. V Morozov, D. Jiang, Y. Zhang, S. V Dubonos, I. V Grigorieva, A. A. Firsov, *Science (80-.).* **2004**, *306*, 666–669.

[4] W. Krätschmer, L. D. Lamb, K. Fostiropoulos, D. R. Huffman, *Nature* **1990**, *347*, 354–358.

[5] R. C. Haddon, *Science* **1993**, *261*, 1545–50.

[6] K. M. Kadish, R. S. Ruoff, Eds. , *Buckminsterfullerenes*, John Wiley & Sons, **1993**.

[7] B. R. Weedon, R. C. Haddon, H. P. Spielmann, M. S. Meier, *J. Am. Chem. Soc.* **1999**, *121*, 335–340.

[8] B. L. Zhang, C. Z. Wang, K. M. Ho, C. H. Xu, C. T. Chan, *J. Chem. Phys.* **1993**, *98*, 3095–3102.

[9] P. J. Krusic, E. Wasserman, P. N. Keizer, J. R. Morton, K. F. Preston, *Science* **1991**, *254*, 1183–1185.

[10] M. Maggini, G. Scorrano, M. Prato, *J. Am. Chem. Soc.* **1993**, *115*, 9798–9799.

[11] W. E. Billups, A. M. Ciufolini, Eds. , *Fullerenes: Chemistry, Physics, and Technology*, John Wiley & Sons, **2000**.

[12] X. Wang, Q. Li, J. Xie, Z. Jin, J. Wang, Y. Li, K. Jiang, S. Fan, *Nano Lett.* **2009**, *9*, 3137–3141.

[13] J. W. G. Wilder, L. C. Venema, A. G. Rinzler, R. E. Smalley, C. Dekker, *Nature* **1998**, *391*, 59–62.

[14] T. W. Odom, J.-L. Huang, P. Kim, C. M. Lieber, *J. Phys. Chem. B* **2000**, *104*, 2794–2809.

[15] J. M. Stephanie Reich, Christian Thomsen, *Carbon Nanotubes: Basic Concepts and Physical Properties*, Wiley-VCH Verlag GmbH & Co. KGaA, **2004**.

[16] C. N. R. Rao, R. Voggu, A. Govindaraj, *Nanoscale* **2009**, *1*, 96–105.

[17] C. N. Rao, B. C. Satishkumar, A. Govindaraj, M. Nath, *Chemphyschem* **2001**, *2*, 78–105.

[18] Y. Chen, R. C. Haddon, S. Fang, A. M. Rao, P. C. Eklund, W. H. Lee, E. C. Dickey, E. A. Grulke, J. C. Pendergrass, A. Chavan, B. E. Haley, R. E. Smalley, *J. Mater. Res.* **1998**, *13*, 2423–2431.

[19] M. A. Hamon, M. E. Itkis, S. Niyogi, T. Alvaraez, C. Kuper, M. Menon, R. C. Haddon, *J. Am. Chem. Soc.* **2001**, *123*, 11292–11293.

[20] S. Niyogi, M. A. Hamon, H. Hu, B. Zhao, P. Bhowmik, R. Sen, M. E. Itkis, R. C. Haddon, *Acc. Chem. Res.* **2002**, *35*, 1105–1113.

[21] J. Wu, W. Pisula, K. Müllen, *Chem. Rev.* **2007**, *107*, 718–747.

[22] K. P. Loh, Q. Bao, P. K. Ang, J. Yang, *J. Mater. Chem.* **2010**, *20*, 2277–2289.

[23] D. Jiang, B. G. Sumpter, S. Dai, *J. Chem. Phys.* **2007**, *126*, 134701–134707.

[24] M. J. Allen, V. C. Tung, R. B. Kaner, *Chem. Rev.* **2010**, *110*, 132–145.

[25] H. Hong, T. Gao, W. Cai, *Nano Today* **2009**, *4*, 252–261.

[26] V. A. Zamolo, G. Valenti, E. Venturelli, O. Chaloin, M. Marcaccio, S. Boscolo, V. Castagnola, S. Sosa, F. Berti, G. Fontanive, M. Poli, A. Tubaro, A. Bianco, F. Paolucci, M. Prato, *ACS Nano* **2012**, *6*, 7989–7997.

[27] M. B. Lerner, J. D'Souza, T. Pazina, J. Dailey, B. R. Goldsmith, M. K. Robinson, A. T. C. Johnson, *ACS Nano* **2012**, *6*, 5143–5149.

[28] S. Okuda, S. Okamoto, Y. Ohno, K. Maehashi, K. Inoue, K. Matsumoto, *J. Phys. Chem. C* **2012**, *116*, 19490–19495.

[29] C. Samorì, H. Ali-Boucetta, R. Sainz, C. Guo, F. M. Toma, C. Fabbro, T. Da Ros, M. Prato, K. Kostarelos, A. Bianco, *Chem. Commun.* **2010**, *46*, 1494–1496.

[30] W. Wu, S. Wieckowski, G. Pastorin, M. Benincasa, C. Klumpp, J.-P. Briand, R. Gennaro, M. Prato, A. Bianco, *Angew. Chem. Int. Ed.* **2005**, *44*, 6358–6362.

[31] C. Ménard-Moyon, E. Venturelli, C. Fabbro, C. Samorì, T. Da Ros, K. Kostarelos, M. Prato, A. Bianco, *Expert Opini. Drug Discov.* **2010**, *5*, 691–707.

[32] G. Pastorin, W. Wu, S. Wieckowski, J.-P. Briand, K. Kostarelos, M. Prato, A. Bianco, *Chem. Commun.* **2006**, 1182–1184.

[33] S. L. Edwards, J. A. Werkmeister, J. A. M. Ramshaw, *Expert Rev. Med. Dev.* **2009**, *6*, 499–505.

[34] S.-T. Yang, Y. Liu, Y.-W. Wang, A. Cao, *Small* **2013**, *9*, 1635–1653.

[35] N. Yanamala, V. E. Kagan, A. A. Shvedova, *Adv. Drug Deliv. Rev.* **2013**, *65*, 2070–2077.

[36] G. Zuo, S.-G. Kang, P. Xiu, Y. Zhao, R. Zhou, *Small* **2012**, *9*, 1546–1556.

[37] A. Hirsch, *Angew. Chem. Int. Ed.* **2002**, *41*, 1853–1859.

[38] J. L. Bahr, E. T. Mickelson, M. J. Bronikowski, R. E. Smalley, J. M. Tour, *Chem. Commun.* **2001**, 193–194.

[39] F. Diederich, C. Thilgen, *Science* **1996**, *271*, 317–324.

[40] M. Prato, *J. Mater. Chem.* **1997**, *7*, 1097–1109.

[41] D. Tasis, N. Tagmatarchis, A. Bianco, M. Prato, *Chem. Rev.* **2006**, *106*, 1105–1136.

[42] P. Singh, S. Campidelli, S. Giordani, D. Bonifazi, A. Bianco, M. Prato, *Chem. Soc. Rev.* **2009**, *38*, 2214–2230.

[43] C. A. Dyke, J. M. Tour, *J. Phys. Chem. A* **2004**, *108*, 11151–11159.

[44] V. Georgakilas, M. Otyepka, A. B. Bourlinos, V. Chandra, N. Kim, K. C. Kemp, P. Hobza, R. Zboril, K. S. Kim, *Chem. Rev.* **2012**, *112*, 6156–6214.

[45] M. Calvaresi, F. Zerbetto, V. F. Selmi, C. G. Ciamician, *Acc. Chem. Res.* **2013**, *46*, 2454–2463.

[46] C. Li, R. Mezzenga, *Nanoscale* **2013**, *5*, 6207–6218.

[47] A. J. Makarucha, N. Todorova, I. Yarovsky, *Eur. Biophys. J.* **2011**, *40*, 103–115 .

[48] D. Bitounis, H. Ali-Boucetta, B. H. Hong, D.-H. Min, K. Kostarelos, *Adv. Mater.* **2013**, *25*, 2258–2268.

[49] H. Y. Mao, S. Laurent, W. Chen, O. Akhavan, M. Imani, A. A. Ashkarran, M. Mahmoudi, *Chem. Rev.* **2013**, *113*, 3407–3424.

[50] M. Mahmoudi, I. Lynch, M. R. Ejtehadi, M. P. Monopoli, F. B. Bombelli, S. Laurent, *Chem. Rev.* **2011**, *111*, 5610–5637.

[51] A. Bianco, T. Da Ros, M. Prato, C. Toniolo, *J. Pept. Sci.* **2001**, *7*, 208–219.

[52] D. Pantarotto, N. Tagmatarchis, A. Bianco, M. Prato, *Mini Rev. Med. Chem.* **2004**, *4*, 805–814.

[53] J. G. Rouse, J. Yang, A. R. Barron, N. A. Monteiro-Riviere, *Toxicol. In Vitro* **2006**, *20*, 1313–1320.

[54] T. Da Ros, M. Prato, *Chem. Commun.* **1999**, 663–669.

[55] C. Ménard-Moyon, K. Kostarelos, M. Prato, A. Bianco, *Chem. Biol.* **2010**, *17*, 107–115.

[56] C. Klumpp, K. Kostarelos, M. Prato, A. Bianco, *Biochim. Biophys. Acta* **2006**, *1758*, 404–412.

[57] A. Bianco, K. Kostarelos, M. Prato, *Expert Opin. Drug Deliv.* **2008**, *5*, 331–342.

[58] Y. Yang, A. M. Asiri, Z. Tang, D. Du, Y. Lin, *Mater. Today* **2013**, *16*, 365–373.

[59] V. C. Sanchez, A. Jachak, R. H. Hurt, A. B. Kane, *Chem. Res. Toxicol.* **2012**, *25*, 15–34 .

[60] V. Singh, D. Joung, L. Zhai, S. Das, S. I. Khondaker, S. Seal, *Prog. Mater. Sci.* **2011**, *56*, 1178–1271.

[61] S. Guo, S. Dong, *Chem. Soc. Rev.* **2011**, *40*, 2644–2672.

[62] L. Lin, A. Liu, Z. Peng, H.-J. L. Lin, P.-K. Li, C. Li, J. Lin, *Cancer Res.* **2011**, *71*, 7226–7237.

[63] S. Bosi, T. Da Ros, G. Spalluto, M. Prato, *Eur. J. Med. Chem.* **2003**, *38*, 913–923.

[64] A. W. Jensen, S. R. Wilson, D. I. Schuster, *Bioorg. Med. Chem.* **1996**, *4*, 767–779.

[65] A. Bianco, K. Kostarelos, M. Prato, *Curr. Opin. Chem. Biol.* **2005**, *9*, 674–679.

[66] L. Lacerda, A. Bianco, M. Prato, K. Kostarelos, *Adv. Drug Deliv. Rev.* **2006**, *58*, 1460–1470.

[67] A. Bianco, K. Kostarelos, C. D. Partidos, M. Prato, *Chem. Commun.* **2005**, 571–577.

[68] X. Sun, Z. Liu, K. Welsher, J. T. Robinson, A. Goodwin, S. Zaric, H. Dai, *Nano Res.* **2008**, *1*, 203–212.

[69] L. Feng, Z. Liu, *Nanomedicine* **2011**, *6*, 317–324.

[70] G. D. Rose, P. J. Fleming, J. R. Banavar, A. Maritan, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16623–16633.

[71] C. N. Pace, B. A. Shirley, M. McNutt, K. Gajiwala, *FASEB J.* **1996**, *10*, 75–83.

[72] V. A. Parsegian, *Van Der Waals Forces Handbook Biologists Chemists Engineers And Physicists*, Cambridge University Press, **2006**.

[73] R. Chang, *Physical Chemistry for the Biosciences*, Edwards Brothers, Inc., Sausalito CA, **2005**.

[74] J. de P. Peter Atkins, *Physical Chemistry for the Life Sciences*, Oxford University Press, Oxford, UK, **2006**.

[75] T. Lazaridis, *Acc. Chem. Res.* **2001**, *34*, 931–937.

[76] L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, Cornell University Press, **1960**.

[77] E. A. Meyer, R. K. Castellano, F. Diederich, *Angew. Chem. Int. Ed.* **2003**, *42*, 1210–1250.

[78] A. Frontera, P. Gamez, M. Mascal, T. J. Mooibroek, J. Reedijk, *Angew. Chem. Int. Ed.* **2011**, *50*, 9564–9583.

[79] B. L. Schottel, H. T. Chifotides, K. R. Dunbar, *Chem. Soc. Rev.* **2008**, *37*, 68–83.

[80] D. A. Dougherty, *Science* **1996**, *271*, 163–168.

[81] F. J. M. Hoeben, P. Jonkheijm, E. W. Meijer, A. P. H. J. Schenning, *Chem. Rev.* **2005**, *105*, 1491–1546.

[82] A. de Leon, A. F. Jalbout, V. A. Basiuk, *Chem. Phys. Lett.* **2008**, *452*, 306–314.

[83] A. de Leon, A. F. Jalbout, V. A. Basiuk, *Int. J. Quantum. Chem.* **2009**, *110*, 953–959.

[84] J. E. Kim, M. Lee, *Biochem. Biophys. Res. Commun.* **2003**, *303*, 576–579.

[85] L. Xie, Y. Luo, D. Lin, W. Xi, X. Yang, G. Wei, *Nanoscale* **2014**, *6*, 9752–9762.

[86] S. A. Andujar, F. Lugli, S. Höfinger, R. D. Enriz, F. Zerbetto, *Phys. Chem. Chem. Phys.* **2012**, *14*, 8599–8607.

[87] M. Mahmoudi, H. R. Kalhor, S. Laurent, I. Lynch, *Nanoscale* **2013**, *5*, 2570–2588.

[88] S. H. Friedman, D. L. DeCamp, R. P. Sijbesma, G. Srdanov, F. Wudl, G. L. Kenyon, *J. Am. Chem. Soc.* **1993**, *115*, 6506–6509.

[89] R. Sijbesma, G. Srdanov, F. Wudl, J. A. Castoro, C. Wilkins, S. H. Friedman, D. L. DeCamp, G. L. Kenyon, *J. Am. Chem. Soc.* **1993**, *115*, 6510–6512.

[90] B.-X. Chen, S. R. Wilson, M. Das, D. J. Coughlin, B. F. Erlanger, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10809–10813.

[91] B. C. Braden, F. A. Goldbaum, B. X. Chen, A. N. Kirschner, S. R. Wilson, B. F. Erlanger, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12193–12197.

[92] W. H. Noon, Y. Kong, J. Ma, *Proc. Natl. Acad. Sci. USA* **2002**, *99 Suppl 2*, 6466–6470.

[93] M. F. Meidine, P. B. Hitchcock, H. W. Kroto, R. Taylor, D. R. M. Walton, *J. Am. Chem. Soc., Chem. Commun.* **1992**, 1534–1537.

[94] T. Mashino, K. Shimotohno, N. Ikegami, D. Nishikawa, K. Okuda, K. Takahashi, S. Nakamura, M. Mochizuki, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1107–1109.

[95] J. Shang, T. A. Ratnikova, S. Anttalainen, E. Salonen, P. C. Ke, H. T. Knap, *Nanotechnology* **2009**, *20*, 415101–415109.

[96] S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. K. Kumar, A. R. Barron, T. Mavromoustakos, M. G. Papadopoulos, *J. Chem. Inf. Model.* **2009**, *49*, 1139–1143.

[97] P. Nedumpully Govindan, L. Monticelli, E. Salonen, *J. Phys. Chem. B* **2012**, *116*, 10676–11683.

[98] X. Meng, B. Li, Z. Chen, L. Yao, D. Zhao, X. Yang, M. He, Q. Yu, *J. Enzyme Inhib. Med. Chem.* **2007**, *22*, 293–296.

[99] X. Bingshe, L. Xuguang, Y. Xiaoqin, Q. Jinli, J. Weijun, *MRS Proc.* **2011**, *675*, W7.4.1.

[100] B. Belgorodsky, L. Fadeev, V. Ittah, H. Benyamini, S. Zelner, D. Huppert, A. B. Kotlyar, M. Gozin, *Bioconjug. Chem.* **2005**, *16*, 1058–1062.

[101] K. H. Park, M. Chhowalla, Z. Iqbal, F. Sesti, *J. Biol. Chem.* **2003**, *278*, 50212–50216.

[102] S.-T. Yang, H. Wang, L. Guo, Y. Gao, Y. Liu, A. Cao, *Nanotechnology* **2008**, *19*, 395101–39518.

[103] N. Iwata, T. Mukai, Y. N. Yamakoshi, S. Haraa, T. Yanase, M. Shoji, T. Endo, N. Miyata, *Fullerenes, Nanotub. Carbon Nanostructures* **1998**, *6*, 213–226.

[104] T. Mashino, K. Okuda, T. Hirota, M. Hirobe, T. Nagano, M. Mochizuki, *Fuller. Sci. Technol.* **2001**, *9*, 191–196.

[105] H. Jin, W. Q. Chen, X. W. Tang, L. Y. Chiang, C. Y. Yang, J. V Schloss, J. Y. Wu, *J. Neurosci. Res.* **2000**, *62*, 600–607.

[106] M. Calvaresi, F. Zerbetto, *ACS Nano* **2010**, *4*, 2283–2299.

[107] G. Pastorin, S. Marchesan, J. Hoebeke, T. Da Ros, L. Ehret-Sabatier, J.-P. Briand, M. Prato, A. Bianco, *Org. Biomol. Chem.* **2006**, *4*, 2556–2562.

[108] S. Kraszewski, M. Tarek, W. Treptow, C. Ramseyer, *ACS Nano* **2010**, *4*, 4158–4164.

[109] E. M. Pérez, N. Martín, *Chem. Soc. Rev.* **2008**, *37*, 1512–1519.

[110] D. Canevet, E. M. Pérez, N. Martín, *Angew. Chem. Int. Ed.* **2011**, *50*, 9248–9259.

[111] M. Calvaresi, F. Zerbetto, *Nanoscale* **2011**, *3*, 2873–2881.

[112] S. Mecozzi, J. Rebek, Jr., *Chem. - A Eur. J.* **1998**, *4*, 1016–1022.

[113] M. Calvaresi, F. Arnesano, S. Bonacchi, A. Bottoni, V. Calò, S. Conte, G. Falini, S. Fermani, M. Losacco, M. Montalti, G. Natile, L. Prodi, F. Sparla, F. Zerbetto, *ACS Nano* **2014**, *8*, 1871–1877.

[114] M. Calvaresi, S. Hoefinger, F. Zerbetto, *Chem. - A Eur. J.* **2012**, *18*, 4308–4313.

[115] A. de Leon, A. F. Jalbout, V. A. Basiuk, *Chem. Phys. Lett.* **2008**, *457*, 185–190.

[116] A. de Leon, A. F. Jalbout, V. A. Basiuk, *Comput. Mater. Sci.* **2008**, *44*, 310–315.

[117] S. Wang, E. S. Humphreys, S.-Y. Chung, D. F. Delduco, S. R. Lustig, H. Wang, K. N. Parker, N. W. Rizzo, S. Subramoney, Y.-M. Chiang, A. Jagota, *Nat. Mater.* **2003**, *2*, 196–200 .

[118] S. Az'hari, Y. Ghayeb, *Mol. Simul.* **2014**, *40*, 392–398.

[119] J. Aihara, *J. Phys. Chem.* **1994**, *98*, 9773–9776.

[120] C. D. Sherrill, B. G. Sumpter, M. O. Sinnokrot, M. S. Marshall, E. G. Hohenstein, R. C. Walker, I. R. Gould, *J. Comput. Chem.* **2009**, *30*, 2187–2193.

[121] C. D. Sherrill, *Acc. Chem. Res.* **2013**, *46*, 1020–1028.

[122] Z. Yang, Z. Wang, X. Tian, P. Xiu, R. Zhou, *J. Chem. Phys.* **2012**, *136*, 025103–025113.

[123] H. Xie, E. J. Becraft, R. H. Baughman, A. B. Dalton, G. R. Dieckmann, *J. Pept. Sci.* **2008**, *14*, 139–151.

[124] C. Wang, S. Li, R. Zhang, Z. Lin, *Nanoscale* **2012**, *4*, 1146–1153.

[125] W. Fan, J. Zeng, R. Zhang, *J. Chem. Theory Comput.* **2009**, *5*, 2879–2885.

[126] G. R. Dieckmann, A. B. Dalton, P. a Johnson, J. Razal, J. Chen, G. M. Giordano, E. Muñoz, I. H. Musselman, R. H. Baughman, R. K. Draper, *J. Am. Chem. Soc.* **2003**, *125*, 1770–1777.

[127] V. Zorbas, A. Ortiz-Acevedo, A. B. Dalton, M. M. Yoshida, G. R. Dieckmann, R. K. Draper, R. H. Baughman, M. Jose-Yacaman, I. H. Musselman, *J. Am. Chem. Soc.* **2004**, *126*, 7222–7227.

[128] V. Zorbas, A. L. Smith, H. Xie, A. Ortiz-Acevedo, A. B. Dalton, G. R. Dieckmann, R. K. Draper, R. H. Baughman, I. H. Musselman, *J. Am. Chem. Soc.* **2005**, *127*, 12323–12328.

[129] E. J. Wallace, R. S. G. D'Rozario, B. M. Sanchez, M. S. P. Sansom, *Nanoscale* **2010**, *2*, 967–975.

[130] C.-C. Chiu, G. R. Dieckmann, S. O. Nielsen, *J. Phys. Chem. B* **2008**, *112*, 16326–16333.

[131] C.-C. Chiu, G. R. Dieckmann, S. O. Nielsen, *Biopolymers* **2009**, *92*, 156–163.

[132] G. Grigoryan, Y. H. Kim, R. Acharya, K. Axelrod, R. M. Jain, L. Willis, M. Drndic, J. M. Kikkawa, W. F. DeGrado, *Science* **2011**, *332*, 1071–1076.

[133] A. K. Jana, N. Sengupta, *Biophys. J.* **2012**, *102*, 1889–1896.

[134] A. K. Jana, J. C. Jose, N. Sengupta, *Phys. Chem. Chem. Phys.* **2013**, *15*, 837–844.

[135] A. Hirano, T. Tanaka, H. Kataura, T. Kameda, *Chem. - A Eur. J.* **2014**, *20*, 4922–4930.

[136] G. Zuo, Q. Huang, G. Wei, R. Zhou, H. Fang, *ACS Nano* **2010**, *4*, 7508–7514.

[137] G. Zuo, W. Gu, H. Fang, R. Zhou, *J. Phys. Chem. C* **2011**, *115*, 12322–12328.

[138] M. Turabekova, B. Rasulev, M. Theodore, J. Jackman, D. Leszczynska, J. Leszczynski, *Nanoscale* **2014**, *6*, 3488–3495.

[139] B. F. Erlanger, B.-X. Chen, M. Zhu, L. Brus, *Nano Lett.* **2001**, *1*, 465–467.

[140] F. Karchemsky, E. Drug, E. Mashiach-Farkash, L. Fadeev, H. J. Wolfson, M. Gozin, O. Regev, *Colloid. Surface. B* **2013**, *112*, 16–22.

[141] C. Ge, J. Du, L. Zhao, L. Wang, Y. Liu, D. Li, Y. Yang, R. Zhou, Y. Zhao, Z. Chai, C. Chen, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16968–16973.

[142] K. Matsuura, T. Saito, T. Okazaki, S. Ohshima, M. Yumura, S. Iijima, *Chem. Phys. Lett.* **2006**, *429*, 497–502.

[143] P. Goldberg-Oppenheimer, O. Regev, *Small* **2007**, *3*, 1894–1899.

[144] J.-W. Shen, T. Wu, Q. Wang, Y. Kang, *Biomaterials* **2008**, *29*, 3847–3855.

[145] D. Nepal, K. E. Geckeler, *Small* **2007**, *3*, 1259–1265.

[146] D. Nepal, K. E. Geckeler, *Small* **2006**, *2*, 406–412.

[147] N. Dragneva, W. B. Floriano, D. Stauffer, R. C. Mawhinney, G. Fanchini, O. Rubel, *J. Chem. Phys.* **2013**, *139*, 174711–174717.

[148] R. B. Pandey, Z. Kuang, B. L. Farmer, S. S. Kim, R. R. Naik, *Soft Matter* **2012**, *8*, 9101–9110.

[149] A. N. Camden, S. A. Barr, R. J. Berry, *J. Phys. Chem. B* **2013**, *117*, 10691–10697.

[150] J. Kyte, R. F. Doolittle, *J. Mol. Biol.* **1982**, *157*, 105–132.

[151] Y. Guo, X. Lu, J. Weng, Y. Leng, *J. Phys. Chem. C* **2013**, *117*, 5708–5717.

[152] W. Qin, X. Li, W.-W. Bian, X.-J. Fan, J.-Y. Qi, *Biomaterials* **2010**, *31*, 1007–1016.

[153] A. Riemann, B. Nelson, *Langmuir* **2009**, *25*, 4522–4525.

[154] A. Riemann, B. E. Owens, *Surf. Sci.* **2010**, *604*, 2084–2090.

[155] B. E. Owens, A. Riemann, *Surf. Sci.* **2014**, *624*, 118–129.

[156] C. Rajesh, C. Majumder, H. Mizuseki, Y. Kawazoe, *J. Chem. Phys.* **2009**, *130*, 124911–124917.

[157] J. Katoch, S. N. Kim, Z. Kuang, B. L. Farmer, R. R. Naik, S. A. Tatulian, M. Ishigami, *Nano Lett.* **2012**, *12*, 2342–2346.

[158] S. N. Kim, Z. Kuang, J. M. Slocik, S. E. Jones, Y. Cui, B. L. Farmer, M. C. McAlpine, R. R. Naik, *J. Am. Chem. Soc.* **2011**, *133*, 14480–14483.

[159] C. R. So, Y. Hayamizu, H. Yazici, C. Gresswell, D. Khatayevich, C. Tamerler, M. Sarikaya, *ACS Nano* **2012**, *6*, 1648–1656.

[160] D. Khatayevich, T. Page, C. Gresswell, Y. Hayamizu, W. Grady, M. Sarikaya, *Small* **2014**, *10*, 1505–1513.

[161] G. Gianese, V. Rosato, F. Cleri, M. Celino, P. Morales, *J. Phys. Chem. B* **2009**, *113*, 12105–12112.

[162] S. M. Tomásio, T. R. Walsh, *J. Phys. Chem. C* **2009**, *113*, 8778–8785.

[163] N. Todorova, A. J. Makarucha, N. D. M. Hine, A. A. Mostofi, I. Yarovsky, *PLoS Comput. Biol.* **2013**, *9*, 1003360–1003373.

[164] G. Zuo, X. Zhou, Q. Huang, H. Fang, R. Zhou, *J. Phys. Chem. C* **2011**, *115*, 23323–23328.

[165] J. Guo, J. Li, Y. Zhang, X. Jin, H. Liu, X. Yao, *PLoS One* **2013**, *8*, 65579–65589.

[166] K. Nakanishi, T. Sakiyama, K. Imamura, *J. Biosci. Bioeng.* **2001**, *91*, 233–244.

[167] G. Raffaini, F. Ganazzoli, *Langmuir* **2003**, *19*, 3403–3412.

[168] G. Raffaini, F. Ganazzoli, *J. Biomed. Mater. Res.* **2006**, *76*, 638–645.

[169] G. Raffaini, F. Ganazzoli, *Langmuir* **2004**, *20*, 3371–3378.

[170] G. Raffaini, F. Ganazzoli, *Langmuir* **2010**, *26*, 5679–5689.

[171] J. Guo, X. Yao, L. Ning, Q. Wang, H. Liu, *RSC Adv.* **2014**, *4*, 9953–9962.

[172] G. Raffaini, F. Ganazzoli, *Macromol. Biosci.* **2007**, *7*, 552–566.

[173] G. Raffaini, F. Ganazzoli, *J. Phys. Chem. B* **2004**, *108*, 13850–13854.

[174] C. L. Brown, I. A. Aksay, D. A. Saville, M. H. Hecht, *J. Am. Chem. Soc.* **2002**, *124*, 6846–6848.

[175] X. Mao, Y. Guo, Y. Luo, L. Niu, L. Liu, X. Ma, H. Wang, Y. Yang, G. Wei, C. Wang, *J. Am. Chem. Soc.* **2013**, *135*, 2181–2187.

[176] P. Laaksonen, M. Kainlauri, T. Laaksonen, A. Shchepetov, H. Jiang, J. Ahopelto, M. B. Linder, *Angew. Chem. Int. Ed.* **2010**, *49*, 4946–4949.

[177] G. Raffaini, F. Ganazzoli, *Langmuir* **2013**, *29*, 4883–4893.

2 Methods

The present chapter will deal with the computational methods used in this thesis. It is therefore divided into three main sections. The first describes the molecular docking and cluster analysis methodologies; the second discusses the basic concept of Molecular Dynamics (MD) simulations including the trajectories analysis procedure used herein. Finally the last section presents a brief description of the electronic structure methods applied in Chapter 3.

2.1 Molecular docking

2.1.1 Docking

Molecular docking is a powerful and commonly used methodology to predict the 3D orientations preferentially adopted by ligand and/or protein in the ligand-protein or protein-protein interaction. It employs scoring functions to quantify the strength of interaction or binding affinity between two molecules. Due to the emerging interest in studying the biomolecules/nanomaterial interface, molecular docking found recently applications also in the prediction of the way of interaction between fullerenes[1–4] or carbon nanotubes[5–8] and proteins. At this respect, we choose AutoDock 4.0[9,10] as one of the most cited docking program also used to study the interaction between proteins and GNMs.[1,2,4,11] The description of the conformational search algorithms and the scoring functions implemented in the program will be followed.

2.1.1.1 Genetic algorithm

The genetic algorithm (GA), the simulated annealing search methods and the local search (LS) are some of the available conformational space algorithms in the AutoDock package. Recently a new hybrid search algorithm, the Lamarckian Genetic

Algorithm (LGA), has been implemented in AutoDock 4.0,[12] obtained from the combination of the GA method and the adaptive LS method program.

In the LGA of AutoDock the initial creation of random population is obtained through a canonical GA algorithm. GA uses ideas based on the language of natural genetics and biological evolution. In the case of molecular docking, the *genotype* is represented by the ligand's state, which is derived by a set of variables describing the translation, orientation and conformation of the ligand with respect to protein (the ligand gene). The *phenotype* is instead defined by the atomic coordinates, while the *fitness* is the total energy obtained through the energy functions. Pairs of individuals are mated according to a c*rossover* process, thereby defining the *offspring* that can undergo random *mutations*. The *selection* of the *offspring* is based on the *fitness* of the individuals. Normally the GA employs *binary crossover* and *binary mutations* to generate new individuals.

The genetic algorithm begins by creating a random population of individuals, where the user defines the number of individuals in the population. For each random individual in the initial population, each of the three translation genes for x, y, and z is given a uniformly distributed random value between the minimum and maximum x, y, and z extents of the grid maps, respectively; the four genes defining the orientation are given by a random quaternion, consisting of a random unit vector and a random rotation angle between -180° and +180°; and the torsion angle genes, if any, are given by random values between -180° and +180°.

In the LGA algorithm, the GA steps are then followed by loop over generations and local search with a defined number of maximum generation and/or energy evaluations. The LS algorithm is based on the method proposed by Solis and Wets[13] that enables the torsional space search since it does not need the information about the gradient of the local energy landscape. It consists of minimization cycles considered adaptive since the user can define the step size based on the recent energy history: a user-defined number of consecutive failures, related to the increase of the energy, makes the step size to increase, whereas the user-defined number of consecutive successes makes the step size to reduce.

The creation of populations and their selection in the LGA algorithm counts five main steps: *mapping* and *fitness evaluation*, *selection*, *crossover*, *mutation* and *elitist selection*. The *mapping* allows the translation of the genotype information of each individual in its corresponding phenotype and the evaluation of the fitness. The

individuals with the best average fitness pass the *selection* step according to the equation:

$$n_0 = \frac{f_w - f_i}{f_w - \langle f \rangle} \quad f_w \neq \langle f \rangle \qquad Eq.: 2.1$$

where the n_0 is the number of offspring for individual, f_i is the fitness of the individual, f_w is the fitness of the worst individual, *i.e*. the highest energy in the last user-defined N generation; $\langle f \rangle$ is the mean fitness of the population.

Random members of the so-selected intermediate population undergo *crossover* and *mutation*. The crossover is preformed first with breaks occurring between genes, never within a gene. Consequently, the chromosomes will be divided in three pieces at the same gene position, each one containing one or more genes (for example ABC and abc). The replacement operated by the crossover will guarantee the maintenance of the population size obtaining for instance AbC and aBc populations. The mutation adds a random real number with a Cauchy distribution as follows:

$$C(\alpha, \beta, x) = \frac{\beta}{\pi(\beta^2 + (x-\alpha)^2} \qquad \alpha \geq 0, \beta > 0, -\infty < x > \infty \qquad Eq.: 2.2$$

where α and β are parameters of the mean and the spread of the distribution.
Finally, the *elitism selection* operates through a user-defined integer parameter, which determines the number of the top survived individuals into the next generation. The elitism selects according to the fitness of the population resulted from the previous processes of selection, crossover and mutation.

In the LGA, each generation is followed by a local search used to update the fitness associated to each individual selected in the GA steps. Conversely to when applied independently, the LS operator within the LGA searches through the genotypic space rather than in the phenotypic space (Figure 2.1). Doing so, it defines each genotype first and, through a mutation process, its corresponding phenotype. This implies that the developmental search in LGA does not require the inverse mapping as usually operated in the canonical LS (Figure 2.1), playing a more explorative role.

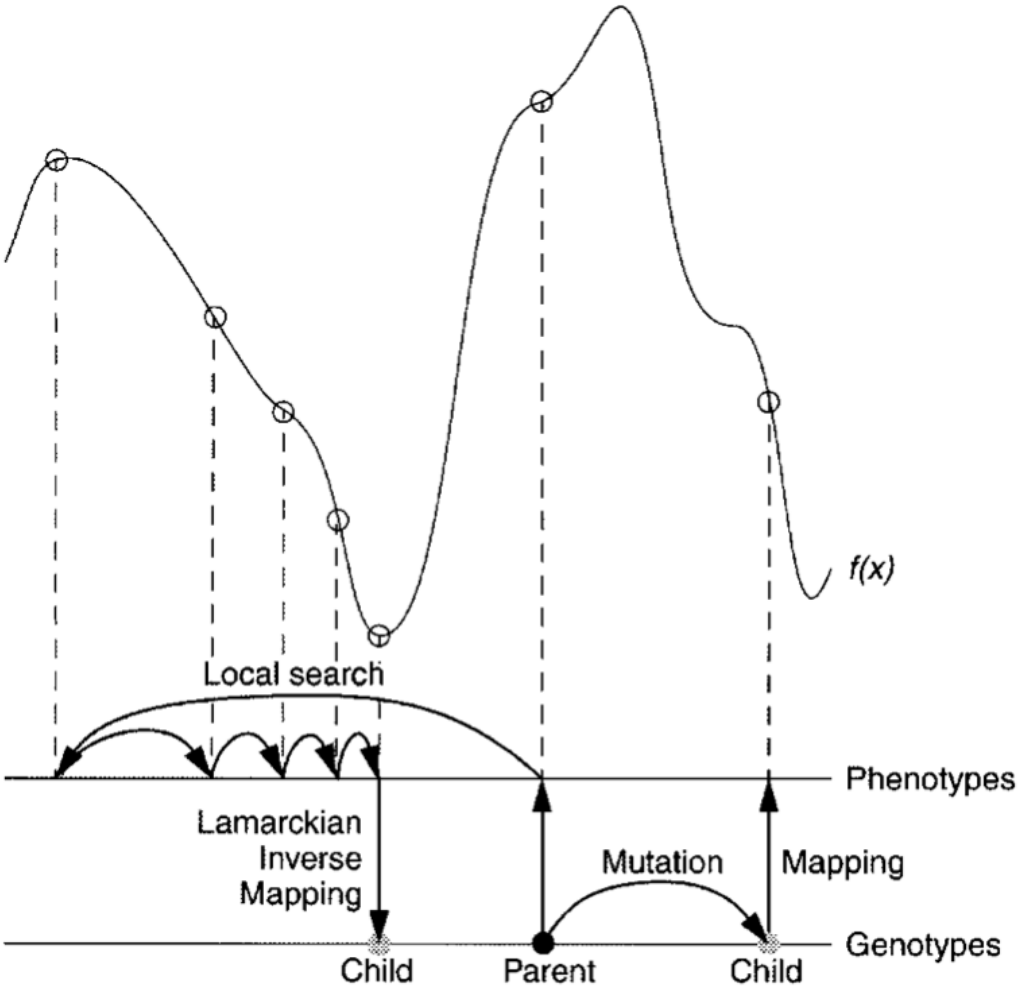


Figure 2.1. The genotypic and phenotypic search, and contrasts Darwinian and Lamarckian search are shown. The space of the genotypes is represented by the lower horizontal line, while the space of the phenotypes is represented by the upper horizontal line. Genotypes are mapped to phenotypes by a developmental mapping function. *f(x)* represents the fitness function. The result of applying the genotypic mutation operator to the parent's genotype is shown on the right-hand side of the diagram, which bears a corresponding phenotype found through the mutation process. Local search is shown on the left-hand side. It is normally performed in phenotypic space and employs information about the fitness landscape. Sufficient iterations of the local search arrive at a local minimum, and an inverse mapping function is used to convert from its phenotype to its corresponding genotype. However, in molecular docking, local search is performed by continuously converting from the genotype to the phenotype, so inverse mapping is not required. The genotype of the parent is replaced by the resulting genotype, however, in accordance with Lamarckian principles. Reprinted with permission from Ref.12.

2.1.1.2 Scoring functions

The AutoDock approach to calculate the energetics of a simulated binding process is schematized in Figure 2.2 and estimates the binding free energy (ΔG) considering the difference between the energy of ligand and protein in the unbound state and the energy of the ligand-protein complex.[14]

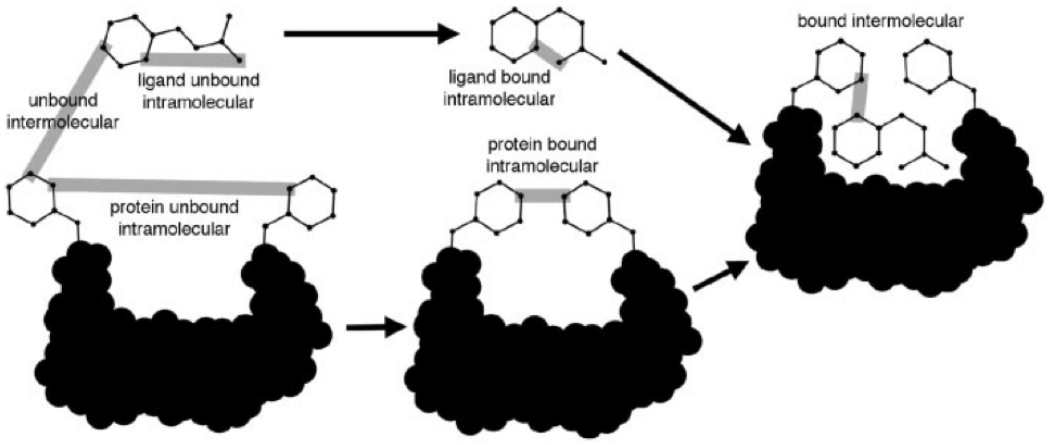


Figure 2.2. The ligand-protein binding process is schematized in two steps. First the ligand and the protein are in the unbound state (left). In the first step the intramolecular interactions occurring in the transition from the unbound state to the conformation that both the ligand and the protein will adopt in the bound state (middle) are evaluated. The second step evaluates the intermolecular energetics of ligand and protein in their bound state (right). Reprinted with permission from Ref. 14.

Interestingly, this improved thermodynamic model allows the inclusion of the intermolecular interactions as an additional term in the estimation of the ΔG. This implies that also the conformational change, occurring for both the ligand and protein side chain, is taken into account. The free energy is therefore estimated including six pair-wise evaluations (V) and an estimation of the conformational entropy lost upon binding (ΔS_{conf}):

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \qquad Eq.: 2.3$$

where L refers to the ligand, P to the protein in the ligand-protein complex. The first two terms are intramolecular energies for the bound and unbound states of the ligand, and the following terms are intramolecular energies for the bound and unbound states

of the protein. In the third parenthesis the variation in intermolecular energy between the bound and unbound states is included. Normally, the two molecules are expected to be sufficiently distant from each other in the unbound state that V^{P-L} is zero. The term of the loss of torsional entropy following the binding (ΔS_{conf}) is directly proportional to the number of rotatable bonds in the molecule (N_{tors}):

$$\Delta S_{conf} = W_{conf} N_{tors} \qquad Eq.: 2.4$$

The pair-wise evaluation of the binding free energy in AutoDock uses a semiempirical force-field that accounts for the dispersion-repulsion, hydrogen bonding, electrostatic and desolvation as:

$$V = W_{vdW} \sum_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{hbound} \sum_{ij} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{ij} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{ij} (S_i V_j + S_j V_i) e^{\left(-r_{ij}^2 / 2\sigma^2\right)} \qquad Eq.: 2.5$$

The summations are all performed over all pairs of ligand atoms i, and protein atoms j, in addition to all pairs of atoms in the ligand separated by three and more bonds. W are the coefficients empirically determined to adjust the empirical free energy as calibrated to a set of experimental complexes. The first term is a typical 6/12 potential for dispersion/repulsion interactions and parameters A and B are taken from the Amber force field.[15] The second term is a directional H-bond term based on a 10/12 potential.[16] Electrostatic interactions (third term) are evaluated with a screened Coulomb potential.[17] The final term is a desolvation potential based on the volume-based approach.[17] Fragmental volumes of protein atoms (V_j) surrounding each ligand atoms are weighted by a solvation parameter (S_i) and an exponential term based on the distance,[17] which is set at 3.5 Å. The volumes are then summed, thereby evaluating the percentage of volume around the ligand atom that is occupied by the protein atoms. The desolvation term is now implemented calculating the atomic solvation parameter (S_i) as a function of the partial charge (q_i):

$$S_i = (ASP_i + QASP \times |q_i|) \qquad Eq.: 2.6$$

where ASP_i is the atomic solvation parameter for atom i and $QASP$ is the charged-based atomic solvation parameter. The degree of shielding of a typical protein was estimated using the $QASP = 0.01097$, $std\ error = 0.0063$.

2.1.2 Cluster analysis

Cluster analysis is an important post-processing statistical approach to partition a set of objects into groups, where the objects belonging to a group are more similar to each other than objects of another group. It is a common technique used for data analysis in statistics and finds important application in bioinformatics, in particular for the rational organization of data obtained from conformational sampling analysis, docking and virtual screening studies. Many algorithms have been developed such as k-means [18] and the hierarchical clustering.[19] Specifically k-means clustering groups *n* objects in *k* cluster where each object belongs to the cluster with the nearest mean, which serves as a cluster center. Hierarchical clustering is based on the union between two nearest clusters. Two main algorithms of hierarchical clustering, used in this thesis for the analysis of docking outcomes and MD trajectories analysis, will be outlined herein: the single and average linkage algorithm[20] as implemented in AutoDock.

2.1.2.1 Hierarchical clustering

The initial condition of the hierarchical clustering considers every structure as a single cluster at lower level, while after a few iterations of the algorithm more clusters are unified in bigger clusters at higher level. From this concept derives the annotation of “hierarchical” clustering. In AutoDock, at the initial step all the conformations are ranked by energy from the lowest to the highest. Root-Mean Square Deviation (RMSD) is then calculated for each conformation and taken as a measure of conformation-to-conformation distance. Therefore, the clustering algorithm starts with n unary clusters and then stepwise merges the two closest clusters within a user-

defined threshold value. The so-called linkage rule represents the way to define the inter-cluster distance. In single-linkage clustering, the link between two clusters is made by a single element pair, namely those two elements (one in each cluster) that are closest to each other. The shortest of these links that remains at any step causes the fusion of the two clusters whose elements are involved. The linkage function is described as follow:

$$\Delta_{M,Q} = min_{m\in\{1,\dots,\chi_M\},q\in\{1,\dots,\chi_Q\}}(d_{m,q}) \qquad Eq.: 2.7$$

where M and Q are two clusters, χ is the cardinality of the clusters and $(d_{m,q})$ is the distance between the two elements m and q. A well-known drawback of single linkage rule is the so-called ‘‘chaining’’ phenomenon: first clusters naturally tend to incorporate the nearby conformations, therefore forming a ‘‘chain’’; consequently, the first clusters will result particularly populated and heterogeneous than others. On the other hand, in the average linkage method, the distance between all the pair of conformations is averaged as follow:

$$\Delta_{M,Q} = \frac{1}{\mathrm{X}_M \mathrm{X}_Q} \sum_{m=1}^{\mathrm{X}_M} \sum_{q=1}^{\mathrm{X}_Q} d_{m,q} \qquad Eq.: 2.8$$

where cluster M contains X_M members and cluster Q contains X_Q members. $(d_{m,q})$ is the RMSD between the two members m and q of clusters M and Q respectively. The so-obtained clustering results can be visualized as a dendrogram, which shows the sequence of cluster fusion and the distance at which each fusion took place. At this point it must be defined the level of clustering more suitable for the conformational space understudy. Therefore in order to define the tradeoff between the number of the clusters and the diversity of the conformations obtained, at each step of single cluster merging, the *spread* (S) of each cluster is calculated along with the *average spread* ($AvSp$). The spread follows the equation:

$$S_M = \frac{2}{\mathrm{X}_M(\mathrm{X}_M - 1)} \sum_{m=1}^{\mathrm{X}_M} \sum_{i=1, i<m}^{\mathrm{X}_M} d_{m,i} \qquad Eq.: 2.9$$

where m and i are members of cluster M. Meanwhile the $AvSp$ is then given by:

$$AvSp_w = \frac{1}{w}\sum_{M=1}^{w} S_M \qquad Eq.: 2.10$$

where w is the number of cluster at stage w of the clustering. Once the $AvSp$ values are collected, they are normalized with a penalty function equal to:

$$P_w = \frac{(n-2)\left[AvSp_w - min_{v\in\{1,\dots,n\}}(AvSp_v)\right]}{Max_{v\in\{1,\dots,n\}}(AvSp_v) - min_{v\in\{1,\dots,n\}}(AvSp_v)} + w + 1 \qquad Eq.: 2.11$$

where Max and min are maximum and minimum values respectively of average spread in the set, w is the total number of clusters at step v.

When clustering is finished, by averaging the coordinates of all the conformations belonging to a cluster the so-called centroid can be obtained. The determination of the centroid will allow the identification of representative conformer for a cluster, which will exhibit the highest similarity to the centroid conformation.

2.2 Molecular Dynamics

The structure and motion of a molecular system is known to be accurately described by the laws of quantum mechanics, which treat both nuclei and electrons of a system. It is also well known that the computational costs required to solve the Schrödinger equation significantly limit the simulation of sizeable complex systems. Hence in parallel, implementations of laws of classical mechanics allow the development of Molecular Dynamics (MD) as a powerful method to simulate the time-evolution of atomistic or molecular systems. MD is based on the following assumption:

1. The Born-Oppnheimer approximation holds: this approximation simplifies the solution of the Schrödinger equation by separating nuclear and electronic

motions. This is justified by the fact that nuclei have greater mass, thus lower motion, than electrons.

2. Nuclei can be approximated as classical particles.
3. The electronic variable can be integrated out.

MD simulations solve the Netwon's equation of motion to describe the time evolution of the dynamic of a system:

$$\vec{F}_i(t) = \frac{d^2\vec{r}_i(t)}{dt^2} m_i = -\frac{dE_{tot}(t)}{d\vec{r}_i} \qquad Eq.: 2.12$$

where $\vec{F_i}(t)$ is the force acting on the i^{th} atom of mass m_i. $\frac{d^2\vec{r}_i(t)}{dt^2}$ is the acceleration that is given as the derivative of the total energy of the system, E_{tot}, with respect to the position $\vec{r_i}$. Hence, starting from the initial structure of a given system, derived from experiments or molecular modeling, and a set of velocities consistent with the simulated temperature, the time-evolution of the system can be provided. The trajectory of the system of interacting particles is thus given by numerically solving the Newton's equation of motion collecting new positions, velocities and energies in a repetitive, stepwise process. The steps count:

1. Setting of the initial conditions, potential interactions as a function of position, $\vec{r_i}$, and velocities, $\vec{v}$, of all atoms.
2. Calculating forces and acceleration acting on each atom.
3. Solving the Newton's equation of motion, updating the configuration of the system.
4. Write new positions, velocities and energies.
5. Repeating the cycle starting from point 2.

To develop these steps an integrator is required which propagates particle positions and velocities from time t to $t + \delta t$. The integration time step, Δt, must be accurately chosen at the beginning of the simulation and remains unchanged along all the run. One must also select the potential energy function, also known as the *force field (ff)*, which consists of:

- A set of equations defining the relationship between the potential energy and the position of each component within a molecular system.
- A series of *atom type* defining the characteristics of each element depending upon its environment.
- *Parameter sets* that fit the equations and atom type to experimental data. Those parameter define force constants, which are values used in the equation to relate atomic characteristics to energy components and structural data (such as bond lengths and angles).

Finally, a statistical ensemble has to be set so that thermodynamic quantities (such as pressure, temperature and number of particles) are under control.
Following are, detailed information about the integrator, *ff* and statistical ensembles used herein is given.

2.2.1 Equation of motion

In the simulations run in this thesis the AMBER11[21] and 12[22] and NAMD 2.9[23] packages have been used to integrate the equation of motion. The specific programs employ the velocity-Verlet algorithm[24] by default, as it is the most common used algorithm. The velocity-Verlet is the implementation of the basic Verlet algorithm,[25] from which differs by explicitly including velocities, similarly to the so called leapfrog algorithm.[26] Positions of each atom are expressed by Taylor expansions as follow:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \vec{v}_i(t+\Delta t/2)\Delta t$$
$$\vec{v}_i(t+\Delta t/2) = \vec{v}_i(t+\Delta t/2) + \frac{\vec{F}_i(t)}{m_i}\Delta t \qquad Eq.: 2.13$$

where positions $\vec{r}_i$ and forces are written at one times step $(t+\Delta t)$ while velocities at each half time step $(t+\Delta t/2)$, thereby alternating the update of position and velocities leaping each other. Due to the fact that velocities are not calculated at the same time as the positions, kinetic and potential energies cannot directly compute the

total energy at time t, but only at the approximate value of velocities at time t according to the following equation:

$$\vec{v}_i(t) = \frac{\vec{v}_i(t - \Delta t/2) + \vec{v}_i(t + \Delta t/2)}{2} \qquad Eq.: 2.14$$

2.2.2 Force field

The *ffs* used in the simulations performed in this thesis are both the general amber *ff* (gaff)[27] and the parm99SB[28] available in AMBER. As previously mentioned, the numerical solution of the Newton's equation of motion (Eq. 2.12) needs the calculation of the forces acting on each atom which derives from the potential energies as a function of the atomic position described by 3N atomic coordinates. The potential energy equation has the following form:

$$E_{tot} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdW} + E_{elect} \qquad Eq.: 2.15$$

which results from the summation of the intramolecular bonding interactions, namely the first three terms of Eq. 2.15, and the non-bonded interaction between atoms, namely the last two terms. The former involves pairs of atoms connected through chemical bonds with a certain stretching distance E_{bonds}, bending angle E_{angles} and torsional angle $E_{dihedrals}$. The non-bonded terms are characterized by van der Waals potential E_{vdW} and electronic potential E_{elect}. Each term is calculated as:

$$E_{bonds} = \sum_{bonds} \frac{1}{2} k_r (r - r_{eq})^2$$

$$E_{angles} = \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_{eq})^2$$

$$E_{dihedrals} = \sum_{dihedrals} \frac{1}{2} V_n (1 + \cos(n\varphi) - \gamma) \qquad \text{Eq.: 2.16}$$

$$E_{vdW} = \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \qquad E_{elect} = \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Among the bonding interactions, the bond stretching and the angle bending have the form of harmonic energy functions where r_{eq} and θ_{eq} are the equilibrium bond lengths and angles, k_r and k_θ are the vibrational constants. The former are derived from structural database, the latter from experimental data of infrared spectroscopy. Examples of the stretching bond distance and bending angle parameters as calculated in Amber are reported in Figure 2.3.

a)

```
BOND
C1-CT  310.0    1.526      JCC,7,(1986),230; AA, SUGARS
C2-CT  310.0    1.526      JCC,7,(1986),230; AA, SUGARS
C1-C3  310.0    1.526      JCC,7,(1986),230; AA, SUGARS
C2-C1  310.0    1.526      JCC,7,(1986),230; AA, SUGARS
C2-C3  310.0    1.526      JCC,7,(1986),230; AA, SUGARS
CT-C3  310.0    1.526      JCC,7,(1986),230; AA, SUGARS
C2-OH  320.0    1.410      JCC,7,(1986),230; SUGARS
C1-OH  320.0    1.410      JCC,7,(1986),230; SUGARS
CA-C2  317.0    1.510      JCC,7,(1986),230; PHE,TYR
C*-C2  317.0    1.495      JCC,7,(1986),230; TRP
C2-C2  310.0    1.526      JCC,7,(1986),230; AA, SUGARS
C2-S   227.0    1.810      changed from 222.0 based on dimethylS nmodes
C3-S   227.0    1.810      changed from 222.0 based on dimethylS nmodes
C -C2  317.0    1.522      JCC,7,(1986),230; AA
C2-SH  237.0    1.810      changed from 222.0 based on methanethiol nmodes
C2-N3  367.0    1.471      JCC,7,(1986),230; LYS
C2-N2  337.0    1.463      JCC,7,(1986),230; ARG
CC-C2  317.0    1.504      JCC,7,(1986),230; HIS
C2-N   337.0    1.449      JCC,7,(1986),230; AA
C3-N   337.0    1.449      JCC,7,(1986),230; AA
C -C3  317.0    1.522      JCC,7,(1986),230; AA
```

b)

```
ANGLE
CT-C1-C3    40.0      109.50
CT-C2-C1    40.0      109.50
CT-C1-C2    40.0      109.50
C3-C1-C3    40.0      109.50
C2-C1-C3    40.0      109.50
C1-CT-H1    50.0      109.50    changed based on NMA nmodes
C2-CT-H1    50.0      109.50    changed based on NMA nmodes
C -CT-C1    63.0      111.10    AA general
C -CT-C2    63.0      111.10    AA general
C1-CT-N     80.0      109.70    AA ala, general    (JACS 94, 2657)
C2-CT-N     80.0      109.70    AA ala, general    (JACS 94, 2657)
CT-C2-OH    50.0      109.50
C2-OH-HO    55.0      108.50
CT-C2-CA    63.0      114.00    AA phe tyr          (SCH JPC  79,2379)
CA-CA-C2    70.0      120.00
C*-C2-CT    63.0      115.60    AA trp
C2-C*-CW    70.0      125.00    AA trp
CB-C*-C2    70.0      128.60    AA trp
CT-C2-C2    40.0      109.50
C2-C2-S     50.0      114.70    AA cyx           (SCHERAGA  JPC 79,1428)
C2-S -C3    62.0       98.90    AA met
```

Figure 2.3. Examples of parameters of stretching bond (a) and bending angle (b) as calculated in Amber *ff*. In both cases the first column reports the atom symbols for bonds (a) and angles (b), the second column the harmonic force constants in kcal mol^{-1} Å^{-2} (a) and kcal mol^{-1} rad^{-2} (b), the third column reports the equilibrium bond (a) and angle (b) distances in Å and degrees °, respectively.

In the dihedral energy $E_{dihedrals}$, V_n represents the torsional barrier height divided by a factor of *2*, and n the integer, which determines the periodicity of the potential (number of minima/maxima) in the interval $[0,2\pi[$. If n is 0.0 then the torsional potential is assumed to have more than one term, and the values of the rest of the terms are read from the next cards until a positive n is encountered. The negative value of n is used only for identifying the existence of the next term and only the absolute value of n is kept (Figure 2.4).

Concerning the non-bonded interaction energies, the vdW potential accounts for both repulsions of atoms, i and j, at small separation distance (r_{ij}^{12}) and weak attractions at larger distance (r_{ij}^{6}). This is described by a Lennard-Jones potential, where A_{ij} and B_{ij} are the coefficients that can be determined by experimental methods, such as non-bonding distances in crystals or gas-phase scattering measurements.

```
DIHE
X -CT-C1-X    6    1.40          0.0           3.
X -CT-C2-X    3    1.40          0.0           3.
C1-CT-N -C    1    0.2270      180.0          -3.
C1-CT-N -C    1    0.8836      180.0          -2.
C1-CT-N -C    1    0.3537      180.0           1.
C2-CT-N -C    1    0.2270      180.0          -3.
C2-CT-N -C    1    0.8836      180.0          -2.
C2-CT-N -C    1    0.3537      180.0           1.
C3-CT-N -C    1    0.2270      180.0          -3.
C3-CT-N -C    1    0.8836      180.0          -2.
C3-CT-N -C    1    0.3537      180.0           1.
C1-CT-C -N    1    0.0560        0.0          -3.
C1-CT-C -N    1    0.0657      180.0          -2.
C1-CT-C -N    1    0.7784      180.0           1.
C2-CT-C -N    1    0.0560        0.0          -3.
C2-CT-C -N    1    0.0657      180.0          -2.
C2-CT-C -N    1    0.7784      180.0           1.
C3-CT-C -N    1    0.0560        0.0          -3.
C3-CT-C -N    1    0.0657      180.0          -2.
C3-CT-C -N    1    0.7784      180.0           1.
CT-C2-C1-C3   1    0.55          0.0          -3.
```

Figure 2.4. Examples of parameters of dihedral angles as calculated in Amber *ff*. The first column reports the atom symbols for the atoms involved in the dihedral angles. When the lateral atoms are labeled as X, a general dihedral angle is defined representative of any dihedrals in the system involving the central atoms, to which are assigned the same parameters. The second column is the factor by which the torsional barrier is divided. The third column is the torsional barrier height, the fourth column is the phase shift angle γ, and the fifth column is the periodicity of the torsional barrier.

The electrostatic or Coulomb potential describes the interactions involving two atoms, i and j, with partial charges q_i and q_j, where $\varepsilon_0 r_{ij}$ is a distance dependent dielectric function. If solvent is treated explicitly in the simulations, the dielectric properties of the medium is taken into account automatically by explicitly computing all electrostatic interactions, therefore $\varepsilon_0 = 1$.

vdW and electrostatic interactions are calculated between atom pairs belonging to different molecules or for atom pairs within the same molecule separated by a distance (at least three chemical bonds) within a cutoff value. Regarding the vdW potential, this truncation introduces only a small error in the energy. Conversely, the electrostatic potential decays much less rapidly with the distance, thus the cutoff truncation might introduce a considerable error in the energy estimation. Therefore, this term is not truncated but, by employing the periodic boundary conditions, it can be treated by the Particle Mesh Ewald summation,[29] which approximates the exact result to an acceptable error similar to the vdW potential.

2.2.2.1 Long range interactions

The calculation of all non-bonded pair interaction within a simulated biological system might be extremely demanding since the computational cost would be proportional to the square of the number of atoms. A difference between short and

long ranged interactions is therefore set according to the spatial extent of the potential. Short ranged are the interaction the potential of which drops off to zero faster than r^d, where r is the distance between two particles and d is the dimension of the problem. Potential reduction slower than r^d refers to long-range interactions instead. Coulomb interactions are considered long-range interaction when dealing with 3D systems. The use of periodic boundary conditions (see below), by treating the system as an infinite number of particles, might create a problem for the electrostatic term thereby requiring a special technique to treat the situation.

2.2.2.2 Ewald summation

The Ewald summation is a method to calculate long-range interactions in periodic systems that was originally developed for calculating the electrostatic energy for ionic crystals. It is now commonly used to calculate the long-range interactions in biological systems modeled by computational simualtions. The advantage of this method is that the calculation of the long-range interaction is divided into two parts: a short-range contribution calculated in the real space, and a long-range contribution calculated as a Fourier transform. This method provides rapid convergence of the energy with high accuracy and reasonably fast computational time.

Starting over from the potential energy, it can be written as:

$$V = \frac{1}{2}\sum_{n}{}' \left(\frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{q_i q_j}{r_{ij}, n} \right) \qquad Eq.: 2.17$$

where $q_i q_j$ are the charges of atoms i and j, n is the box index, r_{ij} is the distance between the charges and the prime means that we omit the case where $i = j$ for $n = 0$. For long-range potential this sum is conditionally convergent and thus extremely slow. For a faster converging, the Ewald summation introduces a counter charge of equal magnitude and different sign, $\varrho -$, around every point charge and spreading out from it. This distribution has the following Gaussian form:

$$\varrho_i^G(r) = q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp(-\alpha|r_i + nL|^2) \qquad Eq.: 2.18$$

where α is an arbitrary parameter related to the width of the Gaussian distribution used to optimized the convergence rate. By this way an efficient screening is performed, the interactions rapidly go to 0 and direct summation can be performed. The distribution of the screened interactions become therefore short ranged and the potential of this term can be calculated summing all the particles. The addition of the counter charge is then compensated by the further addition of a charge distribution with the same sign as the original charge, same shape as the $\varrho_i^G(r)$. The resulting cancelling charge distribution, $\varrho +$, is periodic and can be represented as a rapidly converging Fourier transform, summing the distribution in the reciprocal space. This term will be long ranged. Therefore the separation in short and long-range interactions emerges to be the efficiency gain of this method, that calculates the former as direct particle-particle contributions in the real space, the latter as Fourier transform in the reciprocal space. The total charge distribution can be rewritten as:

$$\rho(r) = \sum_i q_i \delta\,(r - r_i + nL) \qquad Eq.: 2.19$$

$$\rho(r) = \sum_i \left(q_i \delta\,(r - r_i + nL) - \rho_i^G(r)\right) + \sum_i \rho_i^G(r)$$

where the first sum is the one calculating the short ranged potentials in the real space, the second calculates the long ranged potentials in the reciprocal space. It thus results:

$$V = \left(\varepsilon_{dir} - \varepsilon_{self} + \varepsilon_{rec}\right)$$

$$\varepsilon_{dir} = \frac{1}{2} \sum_n{}' \sum_{ij}^N \frac{q_i q_j}{\left|r_{ij} + nL\right|} erfc\left(\sqrt{\alpha}\left|r_{ij} + nL\right|\right)$$

$$\varepsilon_{self} = \sqrt{\frac{\alpha}{\pi}} \sum_{i=1}^N q_1^2 \qquad Eq.: 2.20$$

$$\varepsilon_{rec} = \frac{2\pi}{V} \sum_{k \neq 0} \sum_{ij} \frac{\exp\left(\frac{k^2}{4\alpha}\right)}{k^2} q_i q_j \exp\left(-ik\left(r - r_j\right)\right)$$

where $erfc = \pi^{-\frac{1}{2}} \int_r^{\infty} \exp(-r^2)\, dx$ is the complementary error function.

ε_{dir} is very similar to Eq. 2.17 with the difference that the long ranged term $\frac{1}{r}$ is now substituted by the short ranged $\frac{erfc(r)}{r}$: the cutoff value of the short range interaction is therefore $\alpha^{-\frac{1}{2}}$ and for every i and j the interaction can be approximated by only one periodic range image term. Typically, α is accurately chosen to keep the truncation error in the order of $10^{-5}/10^{-6}$ of ε_{dir}.

ε_{self} is the interaction of the Gaussian charge distribution which is subtracted from the total as it is contained in the reciprocal space of ε_{rec}, albeit it is a constant number not depending on the atomic configuration.

ε_{rec} is the sum of the long-range interactions where the factor $\frac{k^2}{4\alpha}$ ensures a fast convergence in the reciprocal space. This represents the most expensive term of the Ewald scheme.

2.2.2.3 Periodic boundary conditions

Periodic boundary conditions (PBC) is a method usually employed in MD simulations to minimize the boundary effects and mimic the presence of a 'bulk' environment approximating a large infinite system to a small part of it, called the unit cell. Specifically, the system is build from the translational images of the unit cell in the 3D space (Figure 2.5). The resulting boxes are periodic so that the structural characteristics of the system are not compromised along the simulation. The movement of each particle is periodically imaged in the 3D space (Figure 2.5). When a particle leaves the unit cell by one side, it is reintroduced by the opposite site without filling the cell boundary.

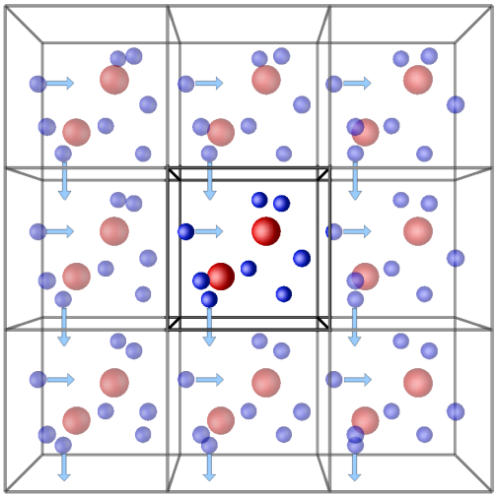


Figure 2.4. Scheme of the PBC. As a particle moves out the simulation box, an image particle comes in to replace it. The calculation of particle interaction within the cutoff range includes both real and image neighbor. Picture taken from http://isaacs.sourceforge.net/phys/pbc.html.

PBC are taken into consideration only in the calculation of the non-bonded interactions between atoms belonging to different molecules within a cutoff radius, which is half of the shortest box vector. This implies that each atom interacts with the nearest atom or image in the periodic array.

The size of the simulation box must be accurately set in order to be large enough to prevent periodic artifact from occurring, such as a macromolecule interacting with its own image in the neighbor box due to the reduced box size. It is therefore recommended to set the box length bigger than the length of the macromolecule plus twice the cutoff radius. On the other hand, the addition of solvent layer around the macromolecule can substantially increase the size of the simulated system and thus the computational costs. Commonly, a layer of solvent molecules of at least 1 nm along each orthogonal direction of the macromolecule has been reported to be a good compromise.[30]

2.2.2.4 Neighbors list

Computing the non-bonded contribution to the interatomic forces in an MD simulation involves, in principle, a large number of pairwise calculations. Let us assume that the interaction potentials are of short-range equal to 0 if $r_{ij} > r_{cut}$ (Figure 2.5a). In this case, the program skips the force calculation, avoiding expensive

calculations, and considers the next candidate j. Nonetheless, looking for atoms within the cutoff distance is quite time-consuming in MD since the time to examine all pair separations is proportional to the number of distinct pairs. However, since atoms move within a time step only < 0.2 Å, the local neighbors of a given atom remain the same for many time steps. The neighbor list method creates lists, which contain for each atom i, the indices of all atoms j whose distance from atom i is smaller than some chosen distance r_m. The potential cutoff sphere, of radius r_{cut}, around a reference atom is thus surrounded by a 'skin', to give a larger sphere of radius r_m as shown in Figure 2.5b. At the first step in a simulation, the neighbors list is thus constructed around each atom, with pair separation within r_m. Over the next few MD time steps, only the atoms inside the neighbor list are taken into account in calculations of the short-range forces. The list will be then updated soon enough to prevent the atoms coming from outside r_m from interacting with the central atom i, because according to the neighbor list they are not in short range contact with the reference particle.

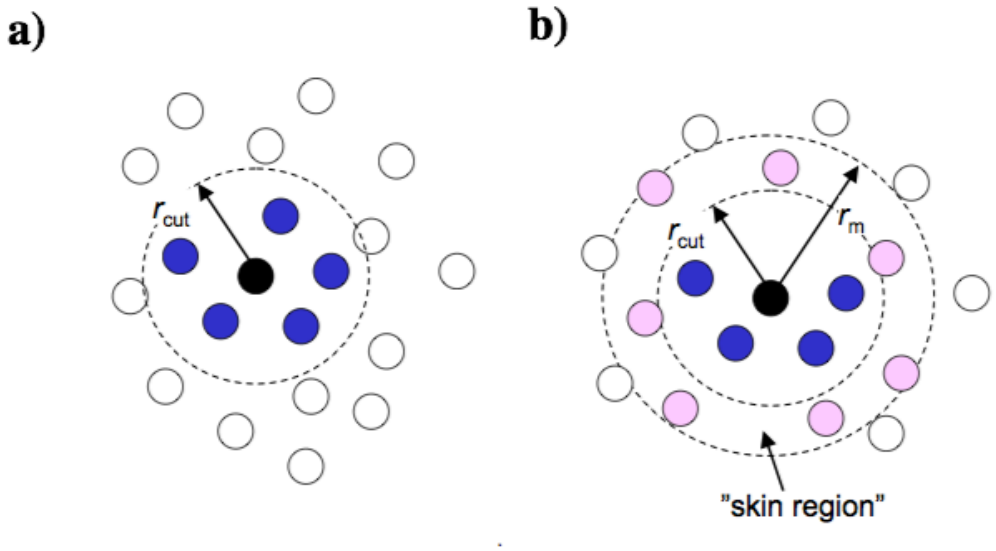


Figure 2.5. Scheme of the pairwise interactions between a given particle (black ball) and the other particles (blue ball) within the potential cutoff range r_{cut} (a). The potential cutoff range is complemented by the list range of neighbors (lilac ball) at the chosen distance r_m (b)

2.2.3 MD in NPT ensemble

In MD simulations it is possible to realize different types of thermodynamics ensembles, which are characterized by the control of certain thermodynamic quantities. For instance, the microcanonical ensemble (NVE) is obtained from the derivation of the equation of motion. However the conditions of constant volume V, number of particles N and total energy E are not representative of the conditions under which most experiments are carried out. The evolution of systems is better described by conditions of constant volume and temperature (NVT, canonical ensemble), or constant pressure and temperature (NPT, isothermal-isobaric ensemble). The simulation of systems in such ensembles requires the use of a thermostat and barostat to control temperature and pressure along the simulation. The description of the algorithms used in the simulations to regulate pressure and temperature in NPT and NVT ensembles, as performed in this thesis, will follow.

2.2.3.1 Berendsen thermostat

The instantaneous value of the temperature $T(t)$ of a system with N_{df} degrees of freedom is related to the kinetic energy E_{kin} via the particles velocities as follows:

$$E_{kin}(t) = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2(t) = \frac{1}{2} N_{df} k_B T(t) \qquad Eq.: 2.21$$

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{N_{df} k_B}$$

by scaling the velocities of a factor λ the temperature can be varied as follow:

$$\Delta T = \sum_{i=1}^{N} \frac{m_i (\lambda v_i)^2}{N_{df} k_B} - \frac{m_i v_i^2}{N_{df} k_B} \qquad Eq.: 2.22$$

$$\Delta T = (\lambda^2 - 1) T(t)$$

If T_0 is the desired temperature and $T(t)$ is the current temperature, hence the variation of the current temperature along the time is defined as:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau T} \qquad Eq.: 2.23$$

$$\Delta T = \frac{\partial t}{\tau_T}(T_0 - T)$$

It is then easy to demonstrate that:

$$\lambda(t) = \sqrt{1 + \frac{\Delta T}{\tau T}\left(\frac{T_0}{T(t)} - 1\right)} \qquad Eq.: 2.24$$

The factor λ is used to scale the velocities at each time-step, in order to relax the temperature toward the desired temperature value T_0. The relaxation rate is controlled by the time coupling constant τT, which should be small to achieve the required temperature but large enough to avoid disturbance of the physical properties of the system by coupling to the bath. Even though the Berendsen weak coupling algorithm is efficient for relaxing the system to the reference temperature, it generally does not allow for a proper sampling of the canonical ensemble. Canonical ensemble simulation is enabled by employing the Langevin thermostat.

2.2.3.2 Langevin thermostat

In the Langevin thermostat, all the particles move as immersed in a continuum of smaller particles. The Langevin equation has the form:

$$\ddot{x} = \nabla\phi - \gamma\dot{x} + \sigma\xi$$

or

$$\frac{dq_i}{dt} = \frac{p_i}{m_i}, \quad \frac{dp_i}{dt} = -\frac{\partial\phi(q)}{\partial q_i} - \gamma p_i + \sigma\xi_i \qquad Eq.: 2.25$$

where the smaller particles create a damping friction to the momenta, $-\gamma p_i$, and give random kicks to the large particles. σ and γ are connected by a fluctuation-dissipation relation:

$$\sigma^2 = 2\gamma m_i kT \qquad Eq.: 2.26$$

that allows the canonical ensemble distribution. The Langevin thermostat is the way to incorporate the solvent effect influencing the dynamic of the solute *via* random collisions, and the effect induced by imposing a frictional drag force on the motion of the solute in the solvent.

At each time step Δt, the Langevin thermostat modifies the equation of motion of the particle i:

$$\Delta p_i = \left(\frac{\partial \phi(q)}{\partial q_i} - \gamma p_i + \delta p\right)\Delta t \qquad Eq.: 2.27$$

where γp_i dumps the momenta and δp is the Gaussian distributed random number with probability:

$$\rho(\delta p) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{|\delta p|^2}{2\sigma^2}\right) \qquad Eq.: 2.28$$

with standard deviation as Eq. 2.26. The random fluctuating force represents the thermal kicks from the small particles. The damping factor and the random force combine to give the correct canonical ensemble.

Typical advantage for Langevin thermostat is that fewer computations per time step are needed since many atoms are eliminated, and are implicitly included by stochastic terms. Besides, a relatively large time step Δt can be chosen due to dissipative term, because damping term stabilizes the equations of motion. Furthermore, since Langevin thermostat replaces the fastest frequency motions in the real system by stochastic terms, Δt is now chosen to resolve the slower degrees of freedom, and thus Δt is several hundred times larger than in the non-stochastic MD. Conversely, momentum transfer is destroyed, it is thus unadvisable to use Langevin thermostats for runs in which you wish to compute diffusion coefficients.

2.2.3.3 Berendsen barostat

The Berendsen algorithm to control the pressure of the system is similar to the Berendsen thermostat.[31] In this case, the algorithm scales the coordinates and the box vectors. So the effect is a first-order kinetic relaxation of the instantaneous pressure $P(t)$ toward a reference pressure P_0, with a time constant ${}^{\tau}P$ that can be specified as an input parameter. The equations for the Berendsen algorithm barostat are:

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{{}^{\tau}P} \qquad Eq.: 2.29$$

the pressure is given by:

$$P = \frac{2}{3V}(E_k - \Theta) \qquad Eq.: 2.30$$

where Θ is the internal virial for pair-additive potentials:

$$\Theta = -\frac{1}{2}\sum_{i<j} r_{ij}F_{ij} \qquad Eq.: 2.31$$

$$r_{ij} = r_i - r_j$$

and F_{ij} is the force on particle i due to particle j. Since intramolecular contributions to the pressure vanish, in molecular systems Eqs. 2.30 and 2.31 can be evaluated using only the center of mass coordinates and velocities and forces acting on the centers of mass. A pressure change can be accomplished by changing the virial through scaling the interparticle distances. A simple proportional coordinate scaling, concomitant with volume scaling, minimizes local disturbances. So, an extra term in the equation of $\dot{x} = v$ is added, proportional to $x(\dot{x} = v + \alpha x)$, while the volume changes accordingly:

$$\dot{V} = 3\alpha V$$

$$\alpha = -\frac{P_0 - P}{{}^{\tau}P}\frac{\beta}{3}x \qquad Eq.: 2.32$$

Therefore the modified equation is:

$$\dot{x} = v - \frac{P_0 - P}{\tau P}\frac{\beta}{3}x \qquad Eq.: 2.33$$

This represents a proportional scaling of coordinates and box length l (assuming an isotropic system in a cubic box) per time step from x to μx and l to μl with the scaling factor μ:

$$\mu = 1 - \frac{\beta \Delta t}{3_{\tau_P}}(P_0 - P) \qquad Eq.: 2.34$$

The compressibility, that may not be accurately known, occurs in the same expression for the scaling factor μ. Since the accuracy in β only influences the accuracy of the noncritical time constant $^{\tau}P$, the imprecision of β has no consequence for the dynamics. The equation 2.34 applies to an isotropic system. However, it is possible to modify the equations for anisotropic systems.

2.2.4 Analysis of MD trajectories

The structural and dynamical properties derived from the MD trajectories performed in this thesis were obtained from the analysis procedures, described as follows.

2.2.4.1 Root Mean Square Deviation (RMSD)

The Root Mean Square Deviation (RMSD) is the measure of the average distance between a set of atoms N (usually the backbone atoms when calculated for the time-evolution of a protein) at time t with respect to a reference structure (usually the initial conformation):

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^{N}(r_i(t_2) - r_i(t_1))^2}{N}} \qquad Eq.: 2.35$$

where $(r_i(t_2) - r_i(t_1)$ is the displacement of the i^{th} atom at time t_2 from the reference position $r_i(t_1)$. An increase in the RMSD implies a deviation of the conformation of the protein with respect to the reference structure suggesting an ongoing conformational change along the trajectory. It can be taken as a measure of the equilibrium of a structure.

2.2.4.2 Root Mean Square Fluctuation (RMSF)

The Root Mean Square Fluctuation (RMSF) is calculated similarly to the RMSD:

$$RMSF = \sqrt{\frac{\sum_{t_j=1}^{T}(r_i(t_j) - \tilde{r}_i)^2}{T}} \qquad Eq.: 2.36$$

where T is the time along which the average is calculated, $\tilde{r}_i$ is the reference position of particle i. The reference position is usually chosen as the time-average position of the same particle. While the RMSD is averaged over the particles, the RMSF is averaged over the time, thereby giving a value for each particle i. This analysis gives insights into the most mobile region within a protein along the trajectory.

2.2.4.3 Gyration radius (Rg)

The gyration radius (Rg) is a calculated property to parameterize the "size" of a system, such the structure or conformation of a protein. It is a scalar quantity with units of length, defined as:

$$R = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\vec{r_i} - \vec{r}_m)^2} \qquad Eq.: 2.37$$

where $\vec{r_i}$ is the vector to each particle, and $\vec{r}_m$ is the center of mass of the particle. The Rg gives insights into the level of compactness of a protein, namely the folding degree of a polypeptide chain.

2.2.4.4 Radial distribution function (RDF)

The Radial Distribution function (RDF) describes the probability of finding a particle at a distance r away from a given reference particle i, and is defined as:

$$g(r)dr = \frac{1}{N}\sum_{i=1}^{N} g_i(r)dr \qquad Eq.: 2.38$$

The general algorithm determines how many particles are within a distance r and $r + dr$ away from a particle, as depicted in Figure 2.6 where the sky blue dot is the reference particle, the red dots are the particles within the circular shell, colored in orange.

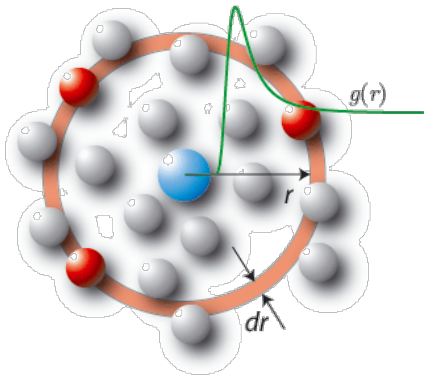


Figure 2.6. The RDF identifies the particles within a distance *r* and *r+dr* (red dots in the orange shell) with respect to a reference particle (blue dot).

The RDF calculates the distances between all particle pairs and bins them into a histogram where the height of each peak is proportional to the number of atoms at distance r from any given atom.

2.2.4.5 Molecular mechanics/Generalized-Born Solvent Area (MM/GBSA)

MM/GBSA is a post-processing trajectory method to evaluate the free energy difference between two states which often represent the bound and unbound state of two solvated molecules (ligand and receptor) or alternatively to compare the free energy of two different solvated conformations of the same molecule. We herein applied this method to calculate the binding free energy of the designed β-sheet adsorbed onto graphene, thereby having a more accurate Free energy differences are calculated by combining the gas phase energy contributions that are independent of the solvent model as well as solvation free energy components (both polar and non-polar) calculated from an implicit solvent model for each species.

$$\Delta G^0_{bind,solv} = \Delta G^0_{bind,vacuum} + \Delta G^0_{solv,complex} - \left(\Delta G^0_{solv,ligand} + \Delta G^0_{solv,receptor}\right) \quad Eq.: 2.39$$

The nonpolar contribution to the solvation free energy has been determined with solvent-accessible-surface-area dependent terms.

Individual snapshot structures of all trajectories were analyzed with program *ptraj*.[22] MM-GBSA[32] analysis was carried out to estimate the binding free energy of the β-sheet when adsorbed onto graphene. Continuum solvation contributions were estimated by using the GB module. The molecular mechanics energies are determined with the SANDER program from AMBER and represent the internal energy (bond, angle and dihedral), and van der Waals and electrostatic interactions. An infinite cutoff for all interactions is used. The electrostatic contribution to the solvation free energy is calculated with a numerical solver for the generalized Born (GB) methods implemented in SANDER.

2.2.4.6 Electrostatic calculations

The electrostatic potentials have been calculated by using the Poisson-Boltzman equation[33] as implemented in the APBS program.[34] The approach, which considers the solvent as continuum, determines the electrostatic potential $\Psi(r)$ by solving the following equation:

$$\vec{\nabla}\left(\varepsilon(\vec{r})\vec{\nabla}\Psi(\vec{r})\right) = -4\pi\rho^{f}(\vec{r}) - 4\pi\sum_{i} c_i^{\infty} z_i q \ \exp\left(\frac{-z_i q\Psi(\vec{r})}{k_B T}\right)\lambda(\vec{r}) \qquad Eq.: 2.40$$

where $\rho^{f}(\vec{r})$ is the fixed charge distribution of the molecule, $\varepsilon(\vec{r})$ is the position dependent dielectric constant, c_i^{∞} is the concentration of ion i at an infinite distance from the molecule, z_i is its valency, q is the proton charge, k_B is the Boltzmann constant, T is the temperature and $\lambda(\vec{r})$ describes the accessibility to ions at point $\vec{r}$.

An analytical solution of the equation can be obtained only for very simple geometries and charge distributions. When the method is applied to proteins, the equation can be numerically solved using the finite difference methods on a grid. Firstly, a spherical probe with the size of the water radius (1.4 Å), passing all over the protein, will generate the molecular surface. A grid will be therefore generated and, according to the surface, a high dielectric constant value will be then assigned to each grid points exposed on the external side of the macromolecule. Conversely, a low dielectric constant value will be assigned to the remaining grid points. Similarly, the atom charge density is also defined and the results are improved by applying the mesh and smoothing algorithms. Finally, the electrostatic potential is numerically estimated for each grid point and the energy of the charged grid is calculated.

2.3 Electronic structure methods

Classical methods, neglecting the electrons, cannot describe chemical problems involving electronic aspects, such as bond formation and breaking, polarization, charge transfer etc. On the other hand, by explicitly treating both nuclei and electrons, electronic structure methods are particularly suitable to address such chemical

problems. These methodologies are divided into three main classes: *i) semi-empirical methods*, which use experimental parameters to simplify the solution of the Schrödinger equation, *ii) ab-initio methods*, which are based on the laws of quantum mechanics and have the electronic wavefunction as central variable to solve the Schrödinger equation, and *iii)* Density Functional Theory (DFT) methods, based on the electron density of the system.[35] In Chapter 3 of this thesis the hybrid exchange-correlation functional M06[36] has been employed to describe non-covalent interactions and is now briefly described.

2.3.1 The Schrödinger equation

The laws of quantum mechanics describe the particle-like and wave-like features of an entity, such as electron, by means of the Schrödinger equation:

$$\left\{\frac{-h^2}{8\pi^2 m}\nabla^2 + V\right\}\Psi(\vec{r},t) = \frac{ih}{2\pi}\frac{\partial\Psi(\vec{r},t)}{\partial t} \qquad Eq.: 2.41$$

Equation 2.40 describes the time-dependent wavefunction, Ψ, of a particle of mass m as the probability distribution of the particle. h is the Planck's constant and V is the potential field in which the particle moves. Energy and other properties can be derived by solving the Schrödinger equation, corresponding to different stationary states of the system. Due to the extensive computational costs that this approach might require, approximations to a time-independent Schrödinger equation have been introduced. Notably, if V is not a function of time, the equation can be rewritten as a spatial function and a time function:

$$\Psi(\vec{r},t) = \psi(\vec{r})\tau(t) \qquad Eq.: 2.42$$

Substituting Eq. 2.42 in 2.41, two equations will be obtained that will be function separately of position and time. Therefore the time-independent Schrödinger equation will have the following form:

$$\hat{H}\psi(\vec{r}) = E\psi(\vec{r}) \qquad Eq.: 2.43$$

where E is the energy of the particle and H is the *Hamiltonian operator*, equal to:

$$\widehat{H} = \frac{-h^2}{8\pi^2 m}\nabla^2 + V \qquad Eq.: 2.44$$

The following sections will focus on the method used in this thesis to solve Eq. 2.42, namely the hybrid exchange-correlation DFT functional, M06. Hybrid functionals are constructed by combining the exact exchange from Hartree–Fock theory (the least expensive *ab-initio* method) with exchange and correlation from DFT methods.

2.3.1.1 Density Functional Theory (DFT)

The DFT approach is based on a strategy of modeling electron correlation *via* general functionals of the electron density. At the origin of this theory, in 1964 the Hohenberg-Khon theorem was published which demonstrates the existence of a unique functional determining the ground state energy and density exactly.[37] In particular, as a result of the Born-Oppenheimer approximation, the Coulomb potential arising from the nuclei is treated as a static external potential V:

$$V = -\sum_{\alpha}\frac{Z_\alpha}{|\vec{r} - R_\alpha|} \qquad Eq.: 2.45$$

where R_α are the nuclear positions. The remainder of the electronic Hamiltonian includes the kinetic part and the electron- electron interaction:

$$F = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_i\sum_{j\neq i}\frac{1}{|\vec{r_i} - \vec{r_j}|} \qquad Eq.: 2.46$$

The ground-state $\Psi(r_i;\, R_\alpha)$ for this Hamiltonian gives rise to a ground-state electronic density $\rho_0(r_i)$:

$$\rho_0 = \int \Psi_0^* \Psi_0 \, d\vec{r} \qquad Eq.: 2.47$$

Both the ground-state wave function and density are functionals of the number of electrons N and the external potential $V(R)$. Therefore, DFT is based on the Hohenberg-Khon theorem according to which the external potential is determined by the corresponding ground-state electron density and *vice versa.* Furthermore, the ground state expectation value of any observable is a unique functional of the ground-state density. Additionally, the variational principle holds also for the electron density, meaning that the total energy is minimal for the ground-state density $\rho_0(\vec{r})$ of the system. The resulting Schrödinger equation can be now rewritten as a function of the electron density ρ:

$$E[\rho] = F[\rho] + \int drV(\vec{r})\rho(\vec{r}) \qquad Eq.: 2.48$$

$\hat{F}$ is a universal functional of ρ and the ground-state density is found by minimizing the functional $E[\rho]$ with respect to ρ. The exact solution of the functional would in principle make DFT exact. However, being not the case further studies by Khon and Sham derived the approximate functionals, currently used in the DFT methods, that partition the electronic energy into several terms:

$$E[\rho(\vec{r})] = T_s\rho(\vec{r}) + \frac{1}{2}\int\int \frac{\rho\vec{r}\rho\vec{r}'}{|\vec{r_i} - \vec{r_i'}|} d\vec{r}d\vec{r}' + E_{xc}[\rho(\vec{r})] + \int \rho(\vec{r})\, V(\vec{r})d\vec{r} \quad Eq.: 2.49$$

where T_s is the kinetic energy term of non-interacting electrons and E_{xc} is the exchange-correlation term which describes the remaining part of the electron-electron interactions. Both term are not solvable in this form, so that T_s can be calculated if the electron density $\rho(r)$ is build up from wavefunctions:

$$T_s\rho(\vec{r}) = \frac{1}{2}\sum_1^N \int \Psi_i^*(\vec{r})\, \nabla^2 \Psi_i d\vec{r} \qquad Eq.: 2.50$$

The proposed Kohn-Sham equations results:

$$[\nabla^2 + V_H + V_{xc} + V_{ext}]\Psi_i(\vec{r}) = \in_i \Psi_i(\vec{r}) \qquad Eq.: 2.51$$

where V_H is the Hartree potential, V_{xc} the exchange-correlation term, V_{ext} the external potential and $\Psi_i(\vec{r})$ are the Kohn-Sham orbitals. These equations are then solved iteratively until self consistency is attained.

The exchange correlation potential V_{xc} follows from equation:

$$V_{xc} = \frac{\delta E_{xc}[\rho(\vec{r})]}{\delta \rho(\vec{r})} \qquad Eq.: 2.52$$

The Kohn-Sham equations have the great privilege of being single-particle equations which, in principle, exactly collectively solve a many-body problem. However, the lack of an explicit analytical formulation of the exchange-correlation functional leads to requirement of approximated forms for it.

According to the Local Density Approximation (LDA), the electron density is thought to behave locally as a uniform electron gas, the exchange-correlation functionals of which becomes:

$$E_{xc}^{LDA} = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \int \rho \frac{4}{3}(\vec{r}) d\vec{r} \qquad Eq.: 2.53$$

The LDA approximation, although successful in the description of a broad variety of highly covalent systems, exhibits heavy deficiencies in describing hydrogen-bonds, which are crucial for studies on biologically relevant objects.

It followed the Generalized Gradient Approximation (GGA), according to which the exchange-correlation functional was gradient-corrected by introducing the gradient of the density in the functional form of E_{xc}.

The main problem in DFT practical implementations arise from the approximation inherent in the exchange-correlation potential leading to serious drawbacks in calculations of thermochemistry of molecules using the LDA or GGA standard functionals.[38] Relative energies of states with different spin multiplicity are often poorly described, radicals or atom transfer transition structures are predicted to be too stable, and activation barriers are severely underestimated.[39] In 1993, Becke introduced the hybrid functionals as a new class of approximation of the exchange-

correlation functional that combine exact exchange from HF theory with exchange and correlation from DFT methods.[40] One of the most popular GGA functionals is given by the combination of the Becke exchange functional and the Lee, Yang and Parr correlation functional (BLYP). Among the most popular and reliable hybrid functionals, B3LYP and the PBE functionals are worth to be mentioned. This combination allows overcoming the well-known deficiencies of the HF methods in describing chemical bonding, improving at the same time the performances of the exchange-correlation approximation. E^{XC} is therefore defined as:

$$E_{xc}^{hybrid} = c^{HF}E_x^{HF} + c^{DFT}E_{xc}^{DFT} \qquad Eq.: 2.54$$

where c are constants.

However, a well-known drawback of such commonly-used functionals is the poor description of the dispersion interactions while at short electron–electron distances, the standard functionals describe the corresponding effects rather well because of their deep relation to the corresponding electron density changes.[41] Aiming at overcoming this limitation, implementation of such exchange-correlation functionals generates dispersion-corrected DFT approaches including nonlocal vdW density functionals, highly parameterized form of standard meta-hybrid approximations, semiclassical corrections and dispersion- correcting atom-centered one-electron potentials.[41]

2.3.1.2 Hybrid meta exchange-correlation functional

The M06 functional has been designed for the description of non-covalent interactions, modeling of organo- and inorganometallic chemistry and herein chosen for modeling the non-covalent interaction between selected amino acids adsorbed onto a reduced size model CNT.[36] Within the M0XX family of functionals, M06 has been chosen for its efficiency in describing non-covalent interactions with reasonable computational costs. The M06 exchange functional is given by:

$$E_{M06}^{X} = \sum_{\sigma} \int dr\left[F_{X_\sigma}^{PBE}(\rho_\sigma, \nabla\rho_\sigma) f(w_\sigma) + \varepsilon_{X_\sigma}^{LSDA} h_X(x_\sigma, z_\sigma)\right] \qquad Eq.: 2.55$$

where

$$h_X(x_\sigma, z_\sigma) = \left(\frac{d_0}{\gamma(x_\sigma, z_\sigma)} + \frac{d_1 x_\sigma^2 + d_2 z_\sigma}{\gamma_\sigma^2(x_\sigma, z_\sigma)} + \frac{d_3 x_\sigma^4 + d_4 x_\sigma^2 z_\sigma + d_5 z_\sigma^2}{\gamma_\sigma^3(x_\sigma, z_\sigma)}\right) \qquad Eq.: 2.56$$

$F_{X_\sigma}^{PBE}(\rho_\sigma, \nabla\rho_\sigma)$ is the exchange energy density of the PBE exchange model,[42] $\varepsilon_{X_\sigma}^{LSDA}$ is the local spin density approximation for exchange:

$$\varepsilon_{X_\sigma}^{LSDA} = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \rho_\sigma^{4/3} \qquad Eq.: 2.57$$

and $f(w_\sigma)$ is the spin kinetic-energy-density enhancement factor:

$$f(w_\sigma) = \sum_{i=0}^{m} a_i w_\sigma^i \qquad Eq.: 2.58$$

Concerning the correlation functional, the opposite and parallel-spin correlation are treated differently. The opposite-spin M06 correlation energy is expressed as:

$$E_{\alpha\beta}^C = \int e_{\alpha\beta}^{UEG}\left[g_{\alpha\beta}(x_\alpha, x_b) + h_{\alpha\beta}(x_{\alpha\beta}, z_{\alpha\beta})\right] dr \qquad Eq.: 2.60$$

where $g_{\alpha\beta}(x_\alpha, x_b)$ is defined as:

$$g_{\alpha\beta}(x_\alpha, x_b) = \sum_{i=0}^{n} c_{C\alpha\beta,i}\left(\frac{\gamma_{C\alpha\beta}\left(x_\alpha^2 + x_\beta^2\right)}{1 + \gamma_{C\alpha\beta}\left(x_\alpha^2 + x_\beta^2\right)}\right)^i \qquad Eq.: 2.61$$

and $h_{\alpha\beta}(x_{\alpha\beta}, z_{\alpha\beta})$ defined as:

$$h_{\alpha\beta}(x_{\alpha\beta}, z_{\alpha\beta}) = \left(\frac{d_0}{\gamma(x_\sigma, z_\sigma)} + \frac{d_1 x_\sigma^2 + d_2 z_\sigma}{\gamma_\sigma^2(x_\sigma, z_\sigma)} + \frac{d_3 x_\sigma^4 + d_4 x_\sigma^2 z_\sigma + d_5 z_\sigma^2}{\gamma_\sigma^3(x_\sigma, z_\sigma)}\right) \qquad Eq.: 2.62$$

with $x_{\alpha\beta}^2 \equiv x_\alpha^2 + x_\beta^2$ and $z_{\alpha\beta} \equiv z_\alpha + z_\beta$.

The parallel spins are expressed as:

$$E_{\sigma\sigma}^{C} = \int e_{\sigma\sigma}^{UEG}[g_{\sigma\sigma}(x_\sigma) + h_{\sigma\sigma}(x_\sigma, z_\sigma)]D_\sigma dr \qquad Eq.: 2.63$$

where $g_{\sigma\sigma}(x_\sigma)$ is defined as:

$$g_{\sigma\sigma}(x_\sigma) = \sum_{i=0}^{n} c_{C\sigma\sigma,i}\left(\frac{\gamma_{C\sigma\sigma}x_\sigma^2}{1+\gamma_{C\sigma\sigma}x_\sigma^2}\right)^i \qquad Eq.: 2.64$$

and $h_{\sigma\sigma}(x_\sigma, z_\sigma)$ is defined as Eq. 2.62. D_σ is the self-interaction correction factor equal to:

$$D_\sigma = 1 - \frac{x_\sigma^2}{4(z_\sigma + C_F)} \qquad Eq.: 2.65$$

It was demonstrated that D_σ vanishes for any one-electron system and that $e_{\alpha\beta}^{UEG}$ and $e_{\sigma\sigma}^{UEG}$ in Eqs. 2.60 and 2.63 are the uniform electron gas (UEG) correlation energy density for the anti-parallel-spin and parallel spin cases.[36] The total M06 correlation energy of the new correlation functional is given by:

$$E^C = E_{\alpha\beta}^{C} + E_{\alpha\alpha}^{C} + E_{\beta\beta}^{C} \qquad Eq.: 2.66$$

The hybrid exchange-correlation energy of the M06 functional can be written as follows:

$$E_{hyb}^{XC} = \frac{27}{100}E_{HF}^{X} + \left(1 - \frac{27}{100}\right)E_{DFT}^{X} + E_{DFT}^{C} \qquad Eq.: 2.67$$

where $\frac{27}{100}E_{HF}^{X}$is the 27% of nonlocal Hartree–Fock (HF) exchange energy (optimized value along with the parameters in the new meta exchange and correlation functional by fitting the value to the data in the training set), E_{DFT}^{X} is the local DFT exchange energy, and E_{DFT}^{C} is the local DFT correlation energy.

2.3.1.3 Basis set approximation

A basis set is a mathematical description of the orbitals within a system. The approximation of the MO uses a linear combination of a pre-defined set of one-electron functions, known as *basis functions*. An individual MO is defined as:

$$\psi_i = \sum_{\mu=1}^{N} c_{\mu i}\chi_\mu \qquad Eq.: 2.54$$

where $c_{\mu i}$ is the molecular orbital expansion coefficient. Many of the electronic structure methods programs employ Gaussian-type (Gaussian Type Orbitals, GTO) atomic functions with the general form:

$$g(\alpha, \vec{r}) = cx^n y^m z^l e^{-\alpha r^2} \qquad Eq.: 2.55$$

where $\vec{r}$ is composed of x, y and z. α is a constant determining the size of the function. $e^{-\alpha r^2}$ is multiplied by powers (possibly 0) of x, y and z, and a constant for normalization so that:

$$\int_{all\ space} g^2 = 1 \qquad Eq.: 2.56$$

Thus, c depends on α, l, m and n.
Linear combination of *primitive gaussians* are used to form the actual basis functions. These functions are called *contracted gaussians* and have the following form:

$$\chi_\mu = \sum_{p} d_{\mu p} g_p \qquad Eq.: 2.57$$

where $d_{\mu p}$ are fixed constants within a given basis set. Substituting Eq. 2.57 in Eq. 2.54, following expansion for molecular orbitals results:

$$\psi_i = \sum_{\mu} c_{\mu i} \left(\sum_{p} d_{\mu p} g_p \right) \qquad Eq.: 2.58$$

For a closed shell system the Roothaan-Hall equations can be applied:[35,38]

$$FC = SC_{\varepsilon} \qquad Eq.: 2.59$$

where each element is a matrix. ε is a diagonal matrix of orbital energies, each of its element ε_i is the one-electron orbital energy of molecular orbital χ_i. F is the *Fock matrix* representing the average effects of the field of all electrons on each orbital. For a closed shell system its elements are:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda=1}^{N} \sum_{\sigma=1}^{N} P_{\lambda\sigma} \left[\left(\mu\nu|\lambda\sigma - \frac{1}{2}(\mu\lambda|\nu\sigma) \right) \right] \qquad Eq.: 2.59$$

where $H_{\mu\nu}^{core}$ is another matrix representing the energy of single electron in the field of the bare nuclei and P is the *density matrix* defined as:

$$P_{\lambda\sigma} = 2 \sum_{i=1}^{occupied} c_{\lambda i}^{*} c_{\sigma i} \qquad Eq.: 2.60$$

The coefficients are summed over the occupied orbitals only, and the factor of two comes from the fact that each orbital has two electrons. Finally, the S matrix from Eq.: 2.66 is the *overlap matrix*, which indicates the overlap between the orbitals. The $\mu\lambda|\nu\sigma$ in Eq. 2.67 is the two-electron repulsion integrals. In the HF approximation of the MO, each electron sees all the other electrons as an average distribution, while no instantaneous electron-electron interaction is included. The solution produces a set of orbitals, both *occupied* $(\psi_{i,j\ldots})$ and *virtual* (unoccupied, denoted as $\psi_{a,b\ldots}$). The total number of orbitals is equal to the number of basis functions used.

The minimal basis sets contain the minimum number of basis functions using a fixed-size atomic-type orbitals.[1] Those are called Slater Type Orbitals (STO), which are the best suited for electronic structure calculations. However, the calculations of STO exponential functions are expensive, thus almost universally replaced by the GTO. STO-3G basis set is the minimal basis set that approximates Slater orbitals with Gaussian functions. A common way to increase the basis set size is the *Split Valence basis set*, such as 3-21G or 6-31G, that uses a number of primitives Gaussians for core orbitals (3 or 6 in these examples) and two (or more) sizes of basis functions for valence orbitals. The *double zeta basis sets* combines two sizes of functions for each atomic orbital. Similarly, *triple split valence basis sets* uses three sized of contracted functions for each orbital-type. Often polarization and diffuse functions are added to the basis set to improve the description of electronic correlation and polarization, and systems with loosely bound electrons.[35,38]

[1] For the thoroughly description of Atomic type orbitals the reader is redirected to ref. 37.

2.4 References

[1] M. Calvaresi, F. Zerbetto, *Nanoscale* **2011**, *3*, 2873–2881.

[2] M. Calvaresi, F. Zerbetto, *ACS Nano* **2010**, *4*, 2283–2299.

[3] W. H. Noon, Y. Kong, J. Ma, *Proc. Natl. Acad. Sci. USA* **2002**, *99 Suppl 2*, 6466–6470.

[4] S.-T. Yang, H. Wang, L. Guo, Y. Gao, Y. Liu, A. Cao, *Nanotechnology* **2008**, *19*, 395101–395108.

[5] M. F. Cabral, J. D. Barrios, E. M. Kataoka, S. A. S. Machado, E. Carrilho, C. D. Garcia, A. A. Ayon, *Colloids Surf. B. Biointerfaces* **2013**, *103*, 624–629.

[6] E. P. D'yachkov, P. N. D'yachkov, R. I. Zhdanov, *Dokl. Phys. Chem.* **2011**, *437*, 65–67.

[7] Z. Kuang, S. N. Kim, W. J. Crookes-Goodson, B. L. Farmer, R. R. Naik, *ACS Nano* **2010**, *4*, 452–458.

[8] Y. Cheng, D. Li, B. Ji, X. Shi, H. Gao, *J. Mol. Graph. Model.* **2010**, *29*, 171–177.

[9] W. Forli, S. Halliday, R. Belew, A. Olson, *autodock.scripps.edu* **2010**, 1–49.

[10] R. Huey, G. Morris, *Scripps Res. Institute, USA* **2008**, 1–56.

[11] V. S. Lee, P. Nimmanpipug, O. Aruksakunwong, S. Promsri, P. Sompornpisut, S. Hannongbua, *J. Mol. Graph. Model.* **2007**, *26*, 558–570.

[12] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639–1662.

[13] F. J. Solis, R. J.-B. Wets, *Math. Oper. Res.* **1981**, *6*, 19–30.

[14] R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell, *J. Comput. Chem.* **2007**, *28*, 1145–1152.

[15] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

[16] P. J. Goodford, *J. Med. Chem.* **1985**, *28*, 849–857.

[17] P. F. W. Stouten, C. Frömmel, H. Nakamura, C. Sander, *Mol. Simul.* **1993**, *10*, 97–120.

[18] J. MacQueen, in *Proc. Fifth Berkeley Symp. Math. Stat. Probab. Vol. 1 Stat.*, The Regents Of The University Of California, **1967**.

[19] L. A. Kelley, S. P. Gardner, M. J. Sutcliffe, *Protein Eng.* **1996**, *9*, 1063–1065.

[20] M. L. Brian S. Everitt, Sabine Landau, *Cluster Analysis*, Wiley Online Library, London, **2001**.

[21] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, AMBER11, University of California, San Francisco **2010**.

[22] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, AMBER12, University of California, San Francisco**2012**.

[23] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781–1802.

[24] W. C. Swope, *J. Chem. Phys.* **1982**, *76*, 637.

[25] L. Verlet, *Phys. Rev.* **1967**, *159*, 98–103.

[26] W. F. Van Gunsteren, H. J. C. Berendsen, *Mol. Simul.* **1988**, *1*, 173–185.

[27] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[28] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712–725.

[29] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[30] O. N. de Souza, R. L. Ornstein, *Biophys. J.* **1997**, *72*, 2395–2397.

[31] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[32] H. Gohlke, C. Kiel, D. A. Case, *J. Mol. Biol.* **2003**, *330*, 891–913.

[33] F. Fogolari, A. Brigo, H. Molinari, *J. Mol. Recognit.* **n.d.**, *15*, 377–392.

[34] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.

[35] Ae. F. James B. Foresman, *Exploring Chemistry With Electronic Structure Methods*, Gaussian, Inc., **1996**.

[36] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2007**, *120*, 215–241.

[37] P. Hohenberg, *Phys. Rev.* **1964**, *136*, B864–B871.

[38] F. Jensen, *Introduction to Computational Chemistry*, Wiley & Sons, Inc, **2006**.

[39] O. V. Gritsenko, B. Ensing, P. R. T. Schipper, E. J. Baerends, *J. Phys. Chem. A* **2000**, *104*, 8558–8565.

[40] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[41] S. Grimme, *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.

[42] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

3. Supramolecular interactions of targeting proteins with graphitic surfaces

This chapter presents two examples of possible applications of functionalized carbon-based nanomaterials in combination with proteins, in bionanotechnology.

The chapter is divided in two main parts, and specifically in section 3.1 the deep understanding of interaction between the monoclonal antibody (mAb) Cetuximab (Ctx) and CNTs is tackled computationally and further accompanied by experimental evidences of the retained activity of this new biohybrid material. The experimental part of the project was carried out by *Dr. Riccardo Marega* and *Florent Pineaux* of the *University of Namur (UNamur)* and *Namur Research College (NARC)*, Namur (Belgium), *Prof. Carine Micheils* and *Lionel Flamant* of the *University of Namur (UNamur)*, *Prof. Yann Garcia* and *Dr. Anil Naik* of the *Institute of Condensed Matter and Nanoscience*, *MOST-Inorganic Chemistry Université Catholique de Louvain*, Louvain (Belgium), who produced, synthesized and biologically tested the biohybrid materials. The computational studies presented herein were developed jointly with *Dr. Alessandra Magistrato* and *Dr. Jacopo Sgrignani* of *International School for Advanced Studies (SISSA/ISAS)* and *CNR-IOM-DEMOCRITOS National Simulation Center* at *SISSA*, Trieste (Italy).

In the second part of the chapter, section 3.2, an *ex novo* β-sheet was designed to specifically and directionally be adsorbed on the graphene surface, applying an exhaustive computational protocol which directs the subsequent experimental production of the selected ideal proteinic candidate for the intended final purpose. The project was carried out in collaboration with *Dr. Alessandra Magistrato* of *International School for Advanced Studies (SISSA/ISAS)* and *CNR-IOM-DEMOCRITOS National Simulation Center* at *SISSA*, Trieste (Italy). The experimental section is an ongoing work, thanks to the joint activities performed by Dr. Riccardo Marega and Laure-Elie Carloni of the *University of Namur (UNamur)* and *Namur Research College (NARC)*, Namur (Belgium).

3.1 Cancer cells targeting via Antibody functionalized Fe-filled multiwalled CNTs

The project presented herewith was part of a collaborative Initial Training Network (ITN) project, named "Cavity-confined Luminophores for advanced Photonic Materials", which was aimed at exploring the potentialities of the exo- and endohedral functionalization of CNTs towards their application in material science and in bionanotechnology. The developments reported as follow are indeed focused on the biological application of Fe-filled multiwalled CNTs (Fe@MWCNTs) coupled with the targeting mAb Ctx.

3.1.1 Background

In a context in which cancer still represents one of the diseases with high incidence and mortality worldwide,[1] it is of crucial importance to provide new and innovative tools for the diagnosis, the effective treatment and eradication of tumors and metastasis. As highlighted in Chapter 1, the merge of GNMs and biomolecules have arisen increasing interest for applications in life science and nanomedicine with a special focus on cancer.[2–5] In particular, the application of CNT derivatives have been widely explored experimentally and proposed as alternative nanostructure for imaging,[6] tissue engineering,[7] drug delivery[8–10] and anticancer platform.[11,12] Interestingly, it has been demonstrated that externally modified CNTs are able to efficiently target cancer cells *in vivo*,[13,14] *in situ* deliver anticancer drug,[15] or to achieve non-invasive hyperthermal treatment.[16,17] In fact, the intrinsic thermal conductivity properties along with the endohedral functionalization of the inner cavity of CNTs, was reported to produce thermal cytotoxicity that finds applications for cancer treatment. Noteworthy, by enabling also the shielded confinement of molecular guests, such as drugs, radioactive species or magnetically-active nanoparticles within the CNTs framework it can be attained the magnetic drug-targeting and/or delivery, magnetic resonance imaging (MRI) or magnetic fluid hyperthermia (MFH).[18] Clearly, the loading and the nature of the encapsulated species might determine the efficiency and applicability of the resulted hybrid

material. For example, iron-filled CNTs with a relatively low % of loaded material (3 wt% of Fe), non-covalently functionalized with pluronic F127 were developed for *in vitro* MR-guided laser-induced thermotherapy with compelling perspectives for MRI applications, but insufficient for a magnetically-induced curative response.[16]

In this respect, with the aim to achieve the construction of new responsive biomaterials, we selected MWCNTs as multivalent and multitasking scaffold envisaging a twofold functionalization of them. Specifically, the new biohybrid presented herewith was designed to be cancer-selective thanks to a targeting moiety grafted on its external sidewalls, and magnetic responsive, owed to the internalization of a magnetically active Fe phase (α-Fe and Fe_3C). The idea was therefore to fill the MWCNTs with the Fe phase during their synthesis itself, thereby avoiding toxicological concerns about the exposition of the so-produced capped Fe-filled MWCNTs (Fe@MWCNTs). On the other hand, a covalent functionalization was envisaged for the external attachment of the targeting moiety to unambiguously address the targeting activity to the biohybrid. Doing so, the new magnetically active CNTs would be able to selectively recognize and bind cancer cells by means of the covalent external functionalization, whereas upon application of a magnetic field a sorting filtration would be achieved. Additionally the magnetic properties might be also exploited for MFH treatment following the selective recognition.

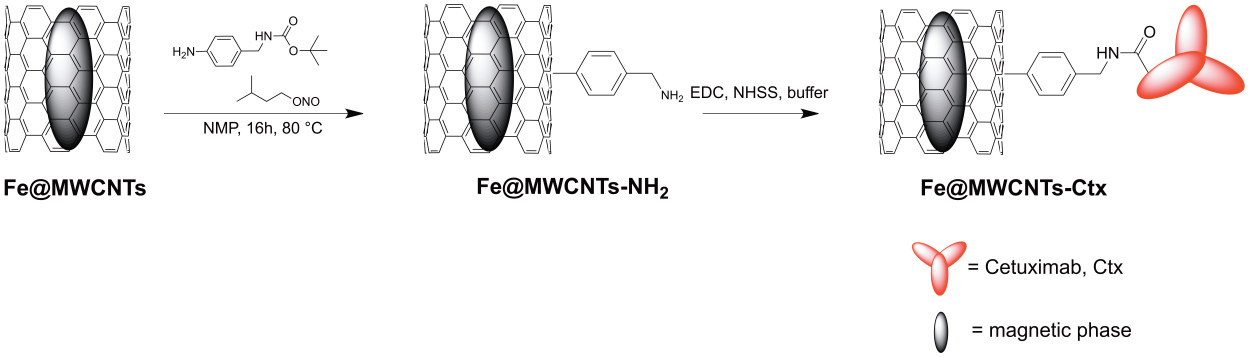
Among all the possibilities, we decided to select the monoclonal antibody (mAb) Cetuximab (Ctx)[19] as the targeting unit. Generally, antibodies are big Y-shaped glycoproteins normally produced by the immune system, which are able to selectively bind with high affinity foreign entities (antigens) present in the body, consequently activating the immune response to expel them.[20] On the other hand, mAbs are engineered Abs that, taking advantage of the high specificity and affinity of these class of proteins, are suitable targeting agents used for therapeutic purposes.[21] In 2004, Ctx was approved by the FDA (Erbitux^{TM}) for its antitumor activity towards colorectal, head and neck cancers and their metastatic forms.[22,23] Ctx's antitumor activity was ascribed to its ability to selectively bind the epidermal growth factor receptor (EGFR),[24] a plasma membrane of the receptor tyrosine kinase (RTK) family that is over-expressed, among the others, in the abovementioned cancer cells.[25] EGFR bears an extracellular binding domain, a transmembrane and two intracellular regions, which are all implicated in the ligand-induced activation. In fact, the binding

of the endogenous ligands promotes a conformational change of the protein that undergoing homo- or hetero-dimerization process, activates the intracellular tyrosine kinase cascade involved in the cell survival, proliferation and metabolism.[26] Additionally, EGFR can be also internalized and translocated at the nuclear level where can exerts a co-transcriptional activity.[26] Being implicated in such pivotal cellular proliferation processes, the over-expression of EGFR turns to be a clear cancer marker and target.[27] The EGFR path inhibition upon binding of Ctx was demonstrated to comprehend multiple approaches ranging from the competition with the endogenous ligands impeding their binding, to the steric prevention of the EGFR conformational change, necessary for its dimerization, thus activation; from the arrest of the cancer cell cycle, to the promotion of the receptor internalization and degradation; from the decrease of the expression of pro-angiogenic factors, to the affection of the apoptosis/survival equilibrium, and finally the Ab-dependent cellular cytotoxicity.[26,28] A scheme of all these mechanisms of inhibition of Ctx is provided in Figure A.1, *Appendix*. Despite the multiple inhibitory mechanism that this mAb can exert, likewise mechanisms of resistance to the chronic exposure to Ctx have been observed due to, for example, a re-activation of pro-angiogenic factors, or to a reduction of the available target proteins as a result of a dysregulation of the EGFR internalization, degradation or subcellular localization, or to an alternative hetero-dimerization of the RTK that activates in any case the kinase cascade.[26,29]

Hence, our new hybrid system would just benefit from the targeting ability of Ctx for the cancer cell recognition, while it would bypass all the resistance issues normally encountered in the mAb monotherapy by eradicating the recognized cancer cells through the magnetic-induced removal.

3.1.2 Experimental results

The production of the designed biohybrid counts three main steps as reported in Scheme 3.1: the endohedral functionalization towards the production of Fe@MWCNTs, the sidewall functionalization with a benzylamino moiety (Fe@MWCNTs-NH_2) that will enable the following step of amidation reaction with the targeting agent Ctx (Fe@MWCNTs-Ctx).



Scheme 3.1. Schematic representation of the synthetic path towards the production of Fe@MWCNTs, Fe@MWCNTs-NH_2 and Fe@MWCNTs-Ctx. The mAb Ctx is schematized as a red Y and the internalized magnetic phase as a grey oval.

The Fe@MWCNTs were successfully produced with a ferrocene-catalyzed chemical vapor deposition (CVD) process yielding a mixture of zig-zag, armchair and chiral capped nanotubes with magnetic responsivity thanks to the inner protected Fe phase (Fe@MWCNTs). As a confirmation, microscopic (transmission electron microscopy TEM, scanning electron microscopy SEM, atomic force microscopy AFM) and spectroscopic (^{57}Fe Mossbauer, Raman, X-ray photoelectron spectroscopy XPS, X-ray diffraction XRD and energy dispersive spectroscopy EDS) characterizations proved the successful endohedral functionalization of a considerable magnetic phase (10.6 ± 3.68 wt% of Fe). Afterwards, the covalent attachment of the targeting moiety, *i.e.* the mAb Ctx, was attained by first functionalizing the Fe@MWCNTs with the benzylamino moiety (Fe@MWCNTs-NH_2), with the exposed free amino groups, allowing then the external amidation reaction with the free carboxyl groups of Ctx. Extensive characterization (TEM, SEM, AFM, XPS, EDS) confirmed the effective covalent grafting of Ctx on the CNT sidewalls (Fe@MWCNTs-Ctx), whose biochemical activity was then tested performing ELISA and confocal microscopy experiments, evaluating the cancer cells sorting through magnetic filtration. The former experiment was carried out screening the binding selectivity of Fe@MWCNTs-Ctx to three different cell lines, namely model cancer cells that overexpress the target receptor (EGFR+, A321 in red in Figure 3.1), and two cells lines that do not overexpress it (EGFR-, EAhy926 and CHO green and blue respectively). As shown in the histogram in Figure 3.1, a remarkable concentration-dependent selectivity towards the A321 cells was detected, while a very low specific

binding towards the EGFR- cells was observed, suggesting a conserved Ctx activity and specificity also within the biohybrid material.

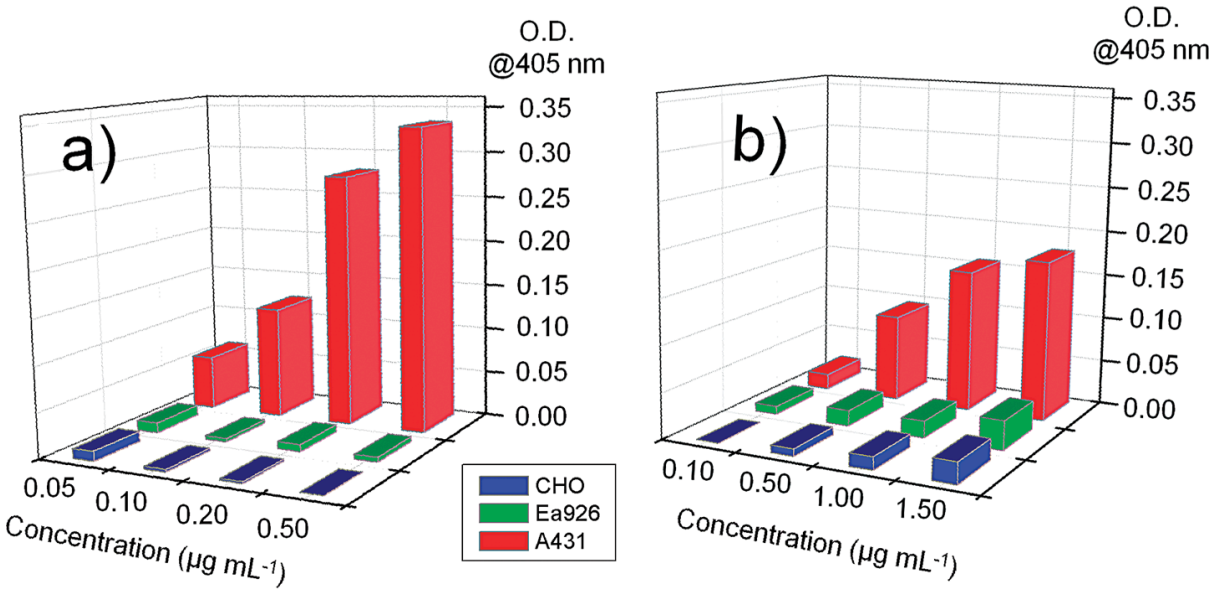


Figure 3.1. ELISA results on A321 (red), Ea926 (green) and CHO (blue) cells of the free mAb Ctx (a) and of Fe@MWCNTs-Ctx as a function of the Ctx concentration.

Concerning the confocal microscopic experiments, a stepwise process was designed to assess the magnetic sorting ability as schematized in Figure 3.2a. Firstly, the population of the three different cell lines (EGFR+ A321, EGFR- EAhy926 and CHO red, green and blue respectively in Figure 3.2a), was mixed, then the Fe@MWCNTs-Ctx were added to the medium at an equivalent concentration of 100 μg mL^{-1} (step II in Figure 3.2a); after 8 min of magnetic filtration with a permanent magnet with a coercive field strength of 900 KA m^{-1} (step III in Figure 3.2a) the separation between the supernatant (step IV in Figure 3.2a) and the magnetic precipitate was achieved through a micropipette extraction (step V in Figure 3.2a). All these steps were imaged by phase-contrast confocal microscopy (Figure 3.2b), which shows the complete and selective abstraction of the model cancer cells (red cells) from the population of the two healthy cell lines (blue and green) operated by the covalent magnetic hybrid Fe@MWCNTs-Ctx. Notably, by adding the Fe@MWCNTs-Ctx to the mixed cellular population (Figure 3.2b, first column) and testing the magnetic treatment, model of healthy cells are exclusively present in the separated supernatant (Figure 3.2b, second column), while the magnetic precipitation

reveals the efficient and selective attainment of the cancer cell extraction (Figure 3.2b, third column).

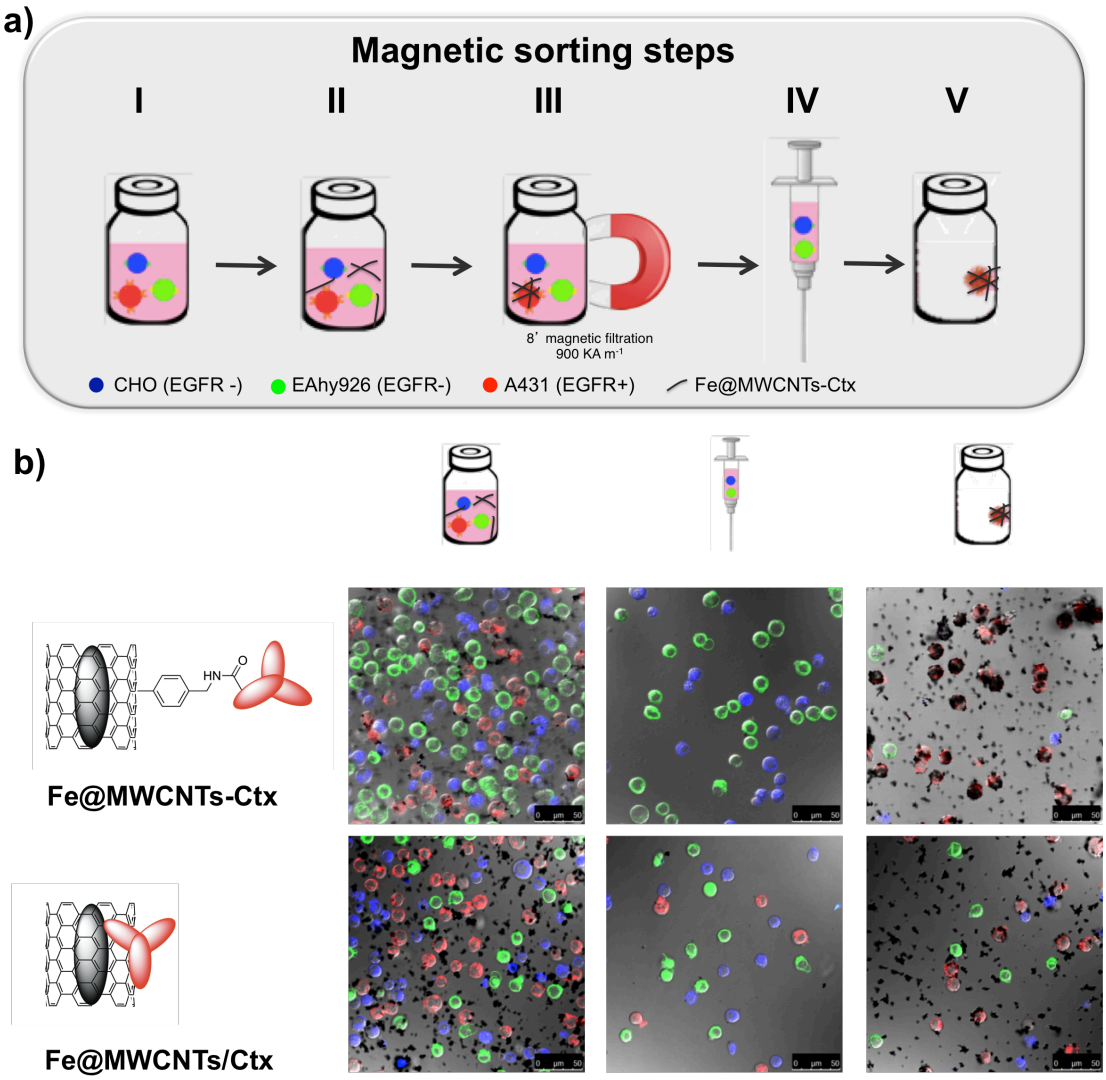


Figure 3.2. Schematic representation of the magnetic sorting steps (a): I) mixing of the three model cells, II) addition of CNTs derivatives at an equivalent concentration of 100 μg mL^{-1}, III) 8 min of magnetic filtration with a permanent magnet, IV) supernatant extraction and V) magnetic precipitate separation. (b) Confocal microscopy images of suspensions containing EGFR- (green and blue) and EGFR+ cells in the presence of Fe@MWCNTs-Ctx and Fe@MWCNTs/Ctx. From left to right the steps II, IV and V are imaged along the experiments. The efficiency of the Fe@MWCNTs-Ctx hybrid is clear in step IV (second column) where no model cancer cells are left after the magnetic extraction.

Interestingly, control experiments of the physical mixture Fe@MWCNTs/Ctx, where Ctx was non-covalently adsorbed on the CNTs sidewall, revealed a retained activity, although non-quantitative, already of the physisorbed adduct, which enables the magnetic and selective removal of cancer cells from the mixed cell population

(Figure 3.2b, bottom). Furthermore, proof of principle stimuli-induced selective cytotoxicity experiments demonstrated the selective suppression of the EGFR+ cells through magnetic fluid hyperthermia (MFH) upon application of an alternating magnetic field, ultimately supporting the efficiency of the design of these new CNT-based magnetically responsive biohybrids.

3.1.3 Computational challenge

The promising experimental evidences gave proofs of the effectiveness of the conjugating strategy resulting in a responsive biohybrid. In principle, one can assume the correct exposition of the targeting moiety for the cancer cells recognition, as well as strong non-covalent interactions that drive even the physisorbed Fe@MWCNTs/Ctx affording the extraction of the recognized cells. However, a thorough comprehension of the biohybrid formation is still missing, which might allow the general mastering and implementation of the system. In fact, although extensive experimental characterizations were carried out, they do not allow a detailed description of the interaction between CNTs and Ctx at the atomistic level, leaving open important questions about structural aspects that might address the real and intimate mechanism of action of the two counterparts facing each other. Is there a predictable way of interfacing between the CNTs and Ctx? Which are the aspects that mainly influence their supramolecular interaction? Does the nanomaterial affect anyhow the structural aspects of the protein, thereby compromising its activity? Which are the driving forces that determine the exceptional stability of the produced biohybrid such to afford the magnetic removal? Finding answers to all these questions would complete and complement the important experimental evidences of the designed bioactive Fe@MWCNTs-Ctx system. More generally, it will shed lights on the interaction between big proteins such as Abs and the tubular CNT structure, relating to the important issues of biosafety and biocompatibility in nanobiotecnology.

Consequently, extensive theoretical efforts will be required to unravel the interactions ruling the conjugation process of such biohybrids. In this chapter, we present a stepwise computational approach to assess manifold issues related to the designed CNT-Ctx conjugate ranging from the prediction of their orientation with respect to each other, the right exposition of the targeting moiety, the structural and

dynamical stability of the conjugate and the nature of the interactions involved in the adduct formation. Chapter 1 clearly evidences the influence of the shape and size of the nanomaterial in the assessment of the GNMs-protein way of interaction. By the same measure, the 3D folded state of the considered protein largely affects the approach to the nanomaterial surface. Therefore, it will follow a brief description of the mAb Ctx, the CNTs, the CNT•Ctx conjugate and the related approximations adopted for their modeling.

3.1.3.1 Approximation of the model

3.1.3.1.1 Structure and parameters of the Cetuximab model

In general, antibodies are big (~150 kDa) Y-shaped globular proteins constituted by two identical polypeptide *light chains* (L-chain) and two longer *heavy chains* (H-chain); L-chain has one extremely variable amino acidic sequence (V region) and one constant (C region), while H-chain has one V and three C regions (Figure 3.3).

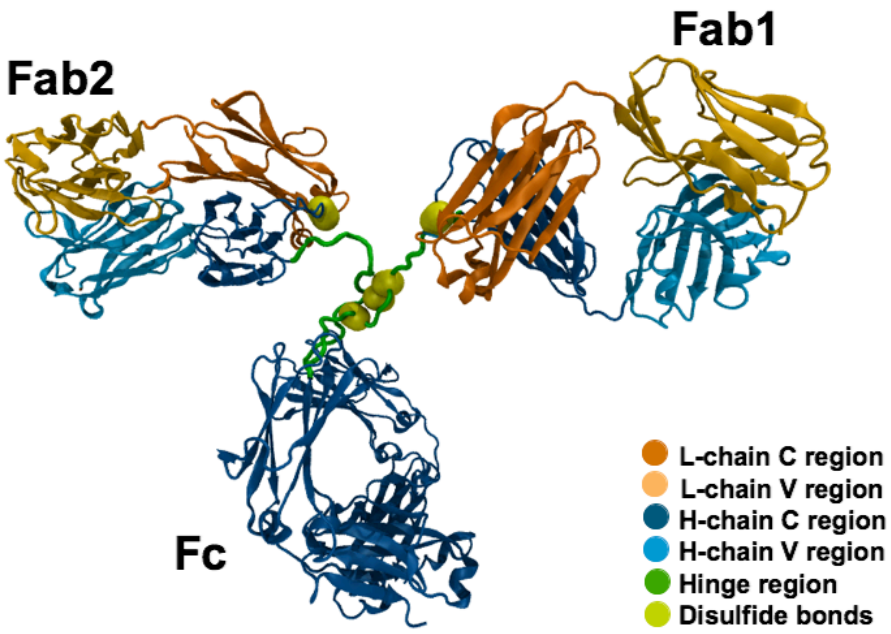


Figure 3.3. Typical Y-shaped structure of Ab with light (orange) and heavy (blue) chain where the variable regions (light colors) form the Fabs, the constant regions (dark colors) at the base of the Y form the Fc. The hinge region is highlighted in green and the disulfide bonds in yellow.

The V regions of the H and L-chains, located in the arms of the Y, form two globular fragment antigen binding (Fabs) regions, while the so-called fragment crystallizable (Fc), located at the base of the Y, is constituted by two C regions of the H-chains. Fabs and Fc are then connected by a central highly-flexible hinge region that through disulfide bonds holds together the whole structure conferring to Abs structural unity and peculiar plasticity at the same time (Figure 3.3).

The Fabs are appointed to the antigen recognition and binding, while the Fc is responsible of the activation of the immune system (effector functions) and determination of the isotype. In fact, the isotopic determinant has been found in the C region of the Fc and changes of the sequence at this level are responsible of the effector functions that lead to the differentiation of the Abs into five human isotypes, called immunoglobulin G (IgG), IgA, IgM, IgD and IgE. Among these isotypes, IgG is the major antibody in the blood and, thanks to its effector functions, is one of the most important for the role exerted in the immune response. In fact, due to the suppression effect *via* effector function located in the Fc region, together with the long half-life and its intrinsic stability, this class of IgG-derived antibodies is the most exploited for engineering therapeutic monoclonal antibodies.[30] As far as Ctx is concerned, it belongs to the IgG1 isotype.

Unfortunately, due to their intrinsic intra-protein flexibility, deriving from the multi-globular organization governed by the flexible core hinge region, Abs represent giant proteins hard to be entirely crystallized. Indeed numerous are the crystals of Ab fragments reported with high-resolution details, whereas few are the crystal structures of entire mAbs so far fully resolved as reported in the Protein Data Bank (PDB). The fully resolved crystals of Abs have been reported by Harris *et al.* (PDB code: 1IGT and 1IGY),[31–33] Saphire *et al.* (PDB code: 1HZH)[34] and Guddat *et al.* (PDB code: 1MCO),[35] which all belong to the IgG isotype.

Regarding the Ctx model system, we chose the intact murine IgG2a antibody (PDB code: 1IGT)[32] because, although murine, the isotype matches the one of Cetuximab (murine nomenclature of IgG subclasses differs from the human ones).[36] Furthermore, a 1IGT:Ctx sequence similarity of 68% for the H-chains and of 71% for the L-chains was detected and considered sufficient for using 1IGT for our simulations. Human IgG1k (PDB code 1HZH)[34] and human IgG1 (PDB code 1MCO)[35] were instead discarded due to the high asymmetry of the first solved

structure and for the hinge deletion of the second, features that could have affected the behavior of the protein in the simulations of the supramolecular interactions. Noteworthy, the peculiar Ab flexibility was one of the aims of the study presented herein.

Hence, the mAb was modeled starting from the crystal structure 1IGT of intact murine IgG2a antibody.[32] The Amber parm99SB[37] force field was employed to describe the protein and glycam06[38] force fields to describe the sugar residues. In the initial stage when it was studied separately, the Ctx system was solvated with explicit water considering a minimal distance of 13 Å between the solute and the edge of cubic box. TIP3P[39] model was used to represent the water molecules.

3.1.3.1.2 Structure and parameters of the CNT model

Concerning the CNT, whose main structural features are summarized in Chapter 1, we started considering the in-house production of Fe@MWCNTs employed experimentally. As typically observed also for industrial CNTs batches, our Fe@MWCNTs sample was characterized by a heterogeneous distribution of the tubes in diameter (1-100 nm), length (up to several μm) and chirality (armchair/zig-zag and achiral CNTs). First approximation of our CNT model is the internal magnetic phase functionalization that buried under the multiple layer of the tube was considered negligible in our study since our focus was on the interfacial interactions between the protein and the nanomaterial. Therefore we decided to model the nanomaterial as an empty tube and considered only the external wall approximating the system to an uncapped single-walled CNT with H-saturated tips. Regarding the variable diameter, aware of the effect that this aspect might have on the interfacing process with Ctx, different CNTs with increasing diameter, namely of 20.2 (CNT20), 33.8 (CNT30), 40.6 (CNT30), 60.9 (CNT60) and 81.1 (CNT80) Å, were first simulated in a preliminary docking investigation (see below). Then as adequate tradeoff between the computational time and costs, CNT20 was selected for the subsequent Molecular Dynamics (MD) simulations as a reduced model with respect to the CNTs used experimentally. Moreover, we have selected its length in a way that its surface could interact with the longest side of Ctx, that is 100 Å. However we simulated an open-ended CNT20, although aware that in reality such narrow CNT would readily

nucleate curved leading to the tube closure.[40,41] As far as the CNT chirality is concerned, taking into account the extreme heterogeneity of the experimental batch containing a mixture of chiral (zig-zag and armchair) and achiral CNTs, we decided to model armchair CNTs as one of the most commonly simulated model of this tubular nanomaterial.

Hence, in our simplified systems armchair (15,15), (25,25), (30,30), (45,45) and (60,60) CNTs were modeled empty, uncapped and single-walled, whose initial coordinates were generated with the Nanotube builder tool implemented in VMD (v. 1.8.7).[42] AM1-mulliken[43] atomic charges were computed using the antechamber module of Amber11[44] and the force field parameters were assigned to the simplified armchair (15,15) CNT atoms (100 Å in length, 20 Å in diameter for a total 2760 atoms) according to the general Amber force field (gaff).[45] The CNT was chosen to be unrestrained and saturated with H atoms, solvated with a 13 Å explicit TIP3P water box when simulated separately.

3.1.3.1.3 Structural properties of the CNT•Ctx conjugate

In order to determine the most likely approximate model of the covalent Fe@MWCNTs-Ctx conjugate, we first map the distribution of the Ctx available acidic residues prone to the covalent coupling to the benzylamino functionalized CNTs as shown in Figure 3.4. Notably, within the Ctx structure a uniform distribution of 124 amino acid residues was found (60 aspartates, 60 glutamates, 2 C-terminal cysteines and 2 C-terminal arginines), of which a restricted number (ca. 40%) is buried in the mAb's tertiary structure, shielded by other side chains. The other carboxylic acid residues, roughly 60%, are homogenously distributed on the outer antibody surface, all being potentially reactive toward amidation reactions with any of the terminal NH_2 groups exohedrally exposed on the functionalized CNT (Figure 3.4).

Thus, it is highly not trivial and not fully reliable, to construct a unique covalent model of the CNT•Ctx biohybrid. For these reasons, we have decided to exclusively investigate the non-covalent interactions of the experimental CNT•Ctx adducts, assuming that this can provide an explanation to its retained biological activity. This approximation was further supported by the experimental evidences as underlined previously, that showed that the simple physical mixture Fe@MWCNTs-

Ctx was already responsive, suggesting that the non-covalent interactions between the mAb and CNT can already lead to the formation of bioactive aggregates.

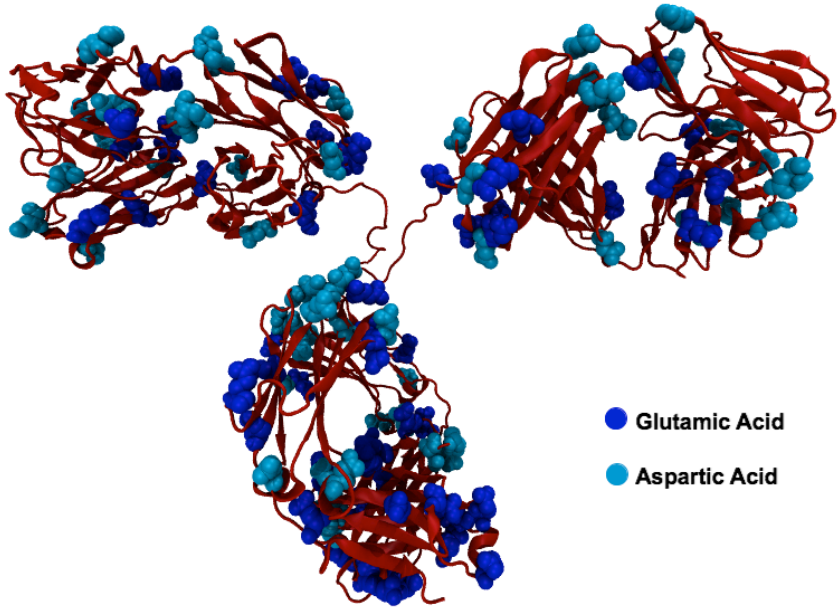


Figure 3.4. Mapping of the free carboxylic acid residues available on the Ctx surface. The aspartic residues are colored in sky-blue and the glutamic residues in blue.

3.1.4 Methods sections

3.1.4.1 Docking studies

Autodock 4.0,[46] one of the most cited programs to generally dock small organic molecules, but also used to study interactions between protein and fullerene derivatives,[47,48] was used to simulate the supramolecular interactions between Ctx and CNTs and select the most probable CNT•Ctx non-covalent adducts for the further studies. We previously selected three main conformations from the 33 ns Cetuximab trajectory, which were first prepared by AutoDockTools 1.4[46] along with all models of CNTs with different diameters. This was done by adding polar hydrogen atoms, computing Gasteiger charges, fixing 10 rotatable bonds for each ligand, i.e. the CNTs, and keeping rigid the entire Cetuximab. In analogy with the work by Yang *et al.*[48] a 'blind docking' was carried out with a large grid volume (126*126*126 with 1 Å grid spacing) covering the Ctx entire surface, in order to find the preferred binding regions

for CNT. First the CNT20 was docked on three MD extracted conformations of Ctx to explore the effect of the flexibility of the protein. Afterwards the increasing diameter CNTs (CNT20, 30, 40, 60, 80) were docked onto the same Ctx conformation to investigate the effect of the nanomaterial size. 50 docking runs were performed applying the Lamarckian Genetic Algorithm employing the default settings of Autodock with 150 number of individuals in population, 250000 number of energy evaluations and 27000 number of generations. Doing so, 50 binding models were generated and, ranked according to their relative populations, were grouped in a minimum of 1 to a maximum of 4 main clusters. In fact, it must be taken into account that the docking scoring functions have been developed for small ligands and are not fully reliable when applied to large systems as CNTs. Due to this limitation, the results of the docking are ranked on the basis of the recurrence of the docked poses rather than on the basis of their energetic values. The three most representative poses of CNT adsorbed onto the Ctx surface were selected from the systematic docking studies to be further processed and refined with MD simulations. They are named pose-fc (*p*-fc), fab (*p*-fab) and back (*p*-back) from the position that the CNT adopts with respect to the Ctx structure (see Results and discussions). However, during the MD simulation the CNT maintained the orientation of the docking extracted poses but was replaced by a smaller nanotube of 20 Å diameter to reduce the computational cost.

3.1.4.2 Force-field based MD simulations

Classical MD simulations were performed for the 20 Å diameter CNT with 100 Å of length, the Ctx and the three most representative docked poses (*p*-fc, *p*-fab and *p*-back). The separate simulations of CNT and Ctx were aimed at checking the reliability of our parameterization for the CNT and to compare the behavior of Ctx in aqueous solution or in complex with CNT. Table 3.1 summarizes the details of the simulated systems carried out in this study.

The AMBER11 package[49] was employed. In all the cases, the initial structures were relaxed by short minimization runs of 2000 steps, using the conjugate gradient energy minimization algorithm (convergence criterion $1.0E^{-4}$ kcal mol^{-1} Å).

Table 3.1. Summary of all the MD simulations performed.

System	Time (ns)	# of atoms
CNT	28	20172
Ctx	33	355780
***p*-fc**		375316
***p*-fab**	40	305566
***p*-back**		305581

Due to the large size of the systems (almost 400000 atoms) the simulations were performed in canonical ensemble. 1.15 ns of MD at constant volume (NVT) were then performed during which the systems were gradually heated to 300 K. In particular, 100,000 steps were run at 100 K using 1.5 fs time step, 100,000 steps at 200 K with 2 fs time step, 200,000 steps at 250 K, and 200,000 steps at 298 K. The MD simulations were performed at constant temperature (298 K) using Berendsen thermostat[50] using a 2 fs time step. Periodic boundary conditions were applied. Non-bonded cut-off of 8 Å was set and Particle Mesh Ewald (PME) summation[51] was used to calculate the long-range electrostatic interactions. The bonds involving hydrogen atoms were kept fixed using the SHAKE algorithm. No other restraints were applied.

3.1.4.3 Calculated properties

The Root-Mean Square Deviation (RMSD) and the Root-Mean Square Fluctuation (RMSF) on the Cα, and on the C, N backbone atoms of Ctx and on C atoms of CNT were calculated over the MD trajectories. The gyration radius (Rg) was also calculated. All properties were calculated for the three docking poses of the CNT•Ctx adduct and for the interaction of CNT with each Ctx domain separately (namely, CNT-Fab1, CNT-Fab2 and CNT-Fc). The *ptraj* module of AMBER11[49] was used for all the above mentioned properties. Secondary structure was analyzed using the DSSP method.[52] The difference in the Solvent Accessible Surface Area (SASA) due to the formation of the adduct ($\Delta SASA = SASA_{(CNT \bullet Ctx)} - SASA_{(Ctx)} - SASA_{(CNT)}$) was calculated along all the trajectories for each pose with the *Cpptraj* module of AMBER11.[49]

The electrostatic surface potentials for the mAb Ctx was calculated by solving the Poisson-Boltzmann equation with the APBS program.[53] The results were visualized using a VMD (v 1.8.7) interface.[42]

The interaction energies, namely the electrostatic (EEL) and the van der Waals energies (vdW) between CNT and Ctx were calculated over the last 25 ns of the MD trajectory using a cut-off for non-bonded interaction of 8 Å employing the NAMD Energy plugin (v 1.4) as implemented in VMD (v 1.9.1) according to Eq. 2.16.[42] For the three different poses, the energetic contribution of each single amino acid present at the interface between the CNT and Ctx was also calculated with the same method. The amino acids with a mean vdW interaction energy (<vdW>) value < 2 kcal mol^{-1} were discarded from the analysis. However, since this allows detecting, still a large number of residues (≥ 20) for each pose we introduced a second threshold (T) as the mean <vdW> value (<<vdW>>±<<*SD*>>), (T=<<vdW>>-<<*SD*>>) to further reduce the number of residues selected.

3.1.4.4 DFT calculations

The residues displaying the largest vdW interaction energy with CNT were used to perform single point DFT calculations on reduced model systems comprising the residue side chain and a small section of CNT (Figure 3.5). Each residue was found to interact with a minimum of two aromatic rings to a maximum of four. Therefore, we used two different models systems of the CNT fragment with a naphtalene-like and pyrenyl-like structure (two and four aromatic rings, respectively). The non-canonical numbering of the atoms starts from the closest carbon atom of the benzyl rings that the charged moiety of the residues faces in the interaction (Figure 3.5a). In particular, Lys841, Lys555 and Glu512 were selected for *p*-fab, Arg522 and His521 for *p*-fc. The conformation of each residue with respect to the CNT was chosen from the snapshot of the trajectory with the lowest calculated vdW interaction energy between the residue and the CNT (example in Figure 3.5b).

The binding enthalpy (*ΔH)* of each residue and the fraction of CNT, was calculated at the DFT level of theory using the M06 hybrid exchange correlation functional,[54] as it allows to account for dispersion interactions,[55] with the 6-31G* basis set. The correction for the basis set superposition error (BSSE) was applied using the Boys-Bernardy counterpoise technique.[56,57] Single point energy

calculations were also performed on the fraction of CNT in the presence and not of the selected residues. All calculations were performed using the Gaussian09 package.[58] Moreover the Bader charge analysis was carried out using the program written by Arnaldsson *et al*.[59] The charges of the CNT fragment were calculated in the presence and in the absence of the selected residue.

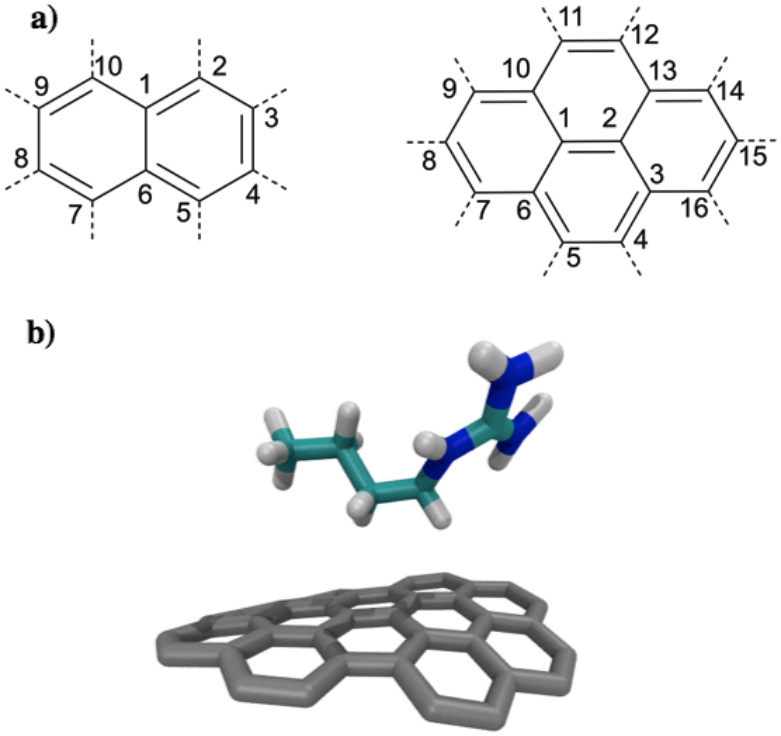


Figure 3.5. a) Scheme and the non-canonical numbering of the CNT benzyl rings interacting with the aa side chains. b) An example of the reduce model system comprising a small section of CNT and the aa side chain.

3.1.5 Results and discussions

With the aim of elucidating the supramolecular interactions stabilizing the bioconjugate formed by CNT and Ctx, we have initially performed force-field based MD simulations of the CNT and the Ctx alones to evaluate the accuracy of the parameterization. In a second avenue, docking calculations and classical MD simulations on the most representative docked poses of the CNT•Ctx bioconjugate were performed. Concerning the isolated CNT, the simulation corroborates the reliability of our parameterization displaying a constant average value in both the RMSD (0.20 ± 0.004 Å) and RMSF (324.96 ± 4.73 Å) plots after 28 ns of MD

(Figure A.2, *Appendix*), suggesting that the structure of the tubular framework reached the equilibrium.

As a model of Cetuximab mAb, we selected the crystal structure of murine IgG2a (PDB code: 1IGT)[32] because of its correspondence to the andibody isotype (murine IgG2a matches the human IgG1)[36] and for its sequence similarity of ~70 %. The intrinsic flexibility of the mAb was reflected by the 33 ns MD trajectory of the isolated Ab, showing that the RMSD of the protein backbone does not converge on a stable value.[1] In fact, being the Ctx structure constituted by three functional regions (two Fabs and one Fc connected by the highly-flexible hinge region), intense structural fluctuations were observed during the MD simulations, suggesting high mobility for the entire structure (Figure 3.6). On the contrary, when calculating the RMSD on separate Fab1, Fab2, and Fc regions, a stable trend was observed already after 5 ns (Figure 3.6b), thus suggesting a fair stability of the isolated domains over the time, consistent with the literature reports.[60,61]

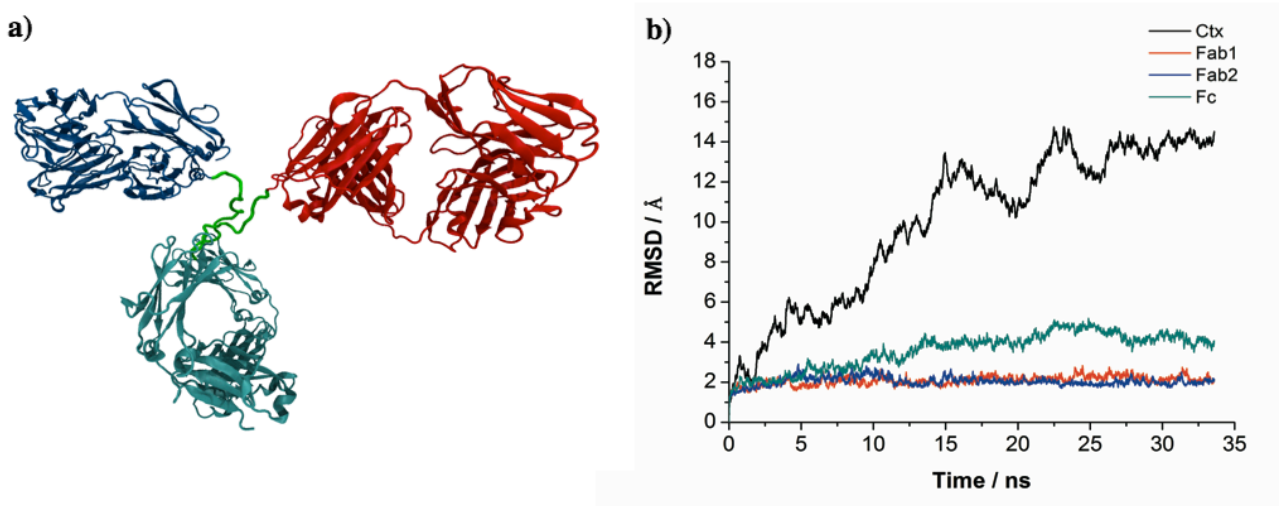


Figure 3.6. a) The Ctx structure with the Fab1, Fab2 and Fc domains colored in red, blue and cyan, respectively. b) RMSD plot for the whole Ctx and its domains calculated over the 33 ns of MD simulation. The RMSD of the entire protein is depicted as a black line, whereas those of the Fab1, Fab2 and Fc domains are colored in red, blue and cyan, respectively.

This behavior is also reflected by the Rg plot, where, conversely to the whole protein (Rg = 46.97 ± 1.77 Å), a stable trend is observed if the Ctx fragments are analyzed separately (Rg_{Fab1} = 24.41 ± 0.13 Å, Rg_{Fab2} = 24.49 ± 0.13 Å, Rg_{Fc} = 24.96 ±

[1] Due to the large size of the systems we were not able to run longer simulations on standard computational resources.

0.19 Å). Notably, the secondary structure (SS) of each domain is preserved during the entire simulation time (Table 3.2). For these reasons, all subsequent analyses are reported for each Ctx domain separately. This confirms that the sampling done during the 33 ns MD simulation is sufficient to confidently carry on our computational studies

Table 3.2. SS content (average number of residues in α-helix and β-sheet along the simulated time) is reported for all the simulated systems

SS form	Ctx	*p*-fc	*p*-fab	*p*-back
β-sheet	635 ± 6	640 ± 8	639 ± 7	639 ± 10
α-helix	38 ± 6	42 ± 5	42 ± 5	39 ± 5

As previously highlighted, from the experimental evidences a strong non-covalent affinity between the two macromolecular counterparts, was assumed. Thus, among the different envisaged reaction pathways, it is realistic to think that the mAb macromolecule firstly adsorbs onto the CNT surface and then undergoes covalent grafting through amidation reaction between the Ctx's carboxylic acids and the neighboring NH2 groups present on the CNT surface. Therefore, in order to understand how the Ctx and the CNT macromolecules can interface with each other, as this could significantly influence the protein structural and conformational properties and hence, that of the two Fabs, docking calculations were carried out.

At first, the electrostatic surface potential map of the Ctx nicely displays a rather uniform surface polarity (Figure 3.7), thus clearly indicating the absence of a particular hydrophobic pocket that might drive, at a first glance, a preferential interaction with the CNT framework.

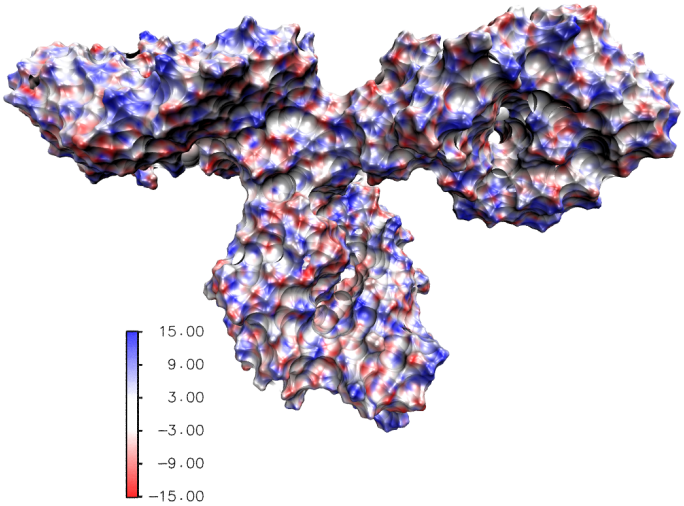


Figure 3.7. The Electrostatic surface potential is mapped on the Ctx structure displaying in red the hydrophobic and in blue the hydrophilic regions.

Furthermore considering the previous data, it is clear that the antibody structure cannot be efficiently represented by a single conformation. A cluster analysis was then performed on the 33 ns trajectory, from which the three most representative conformations could be obtained (red, blue and green Ctx conformations in Figure 3.8a). Therefore, to predict the interaction mode between the CNT exosurface and the Ctx, we performed docking simulations of a 3D model of CNT with a 20 Å in diameter and 100 Å in length, labeled as CNT20 onto the three most representative conformations of Ctx as calculated above. As clearly observed in Figure 3.8b, independently on the initial conformation of the Ctx, the docking simulations generated the same distribution of the binding positions.

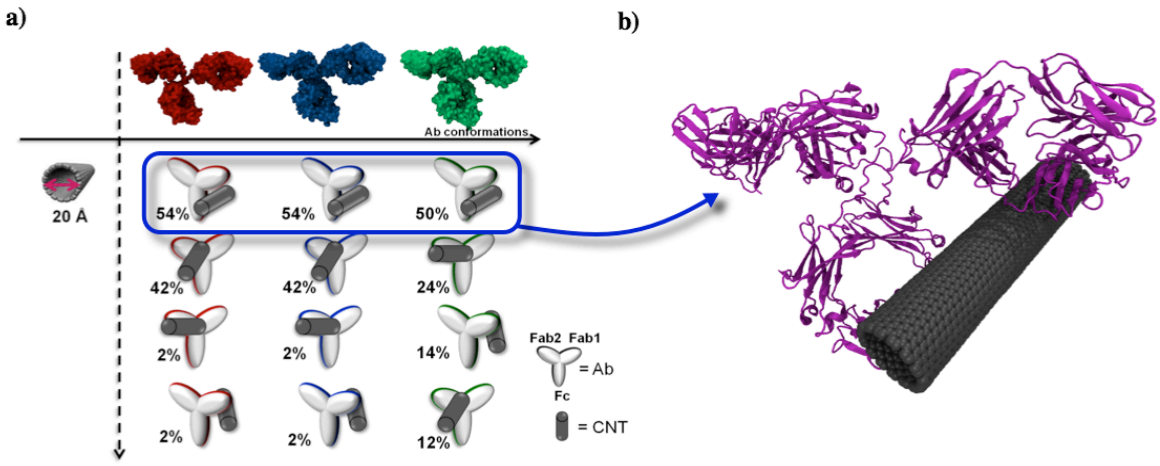


Figure 3.8. a) The interaction modes, predicted by docking calculations between CNT20 and the three Ctx conformations, are sketched as cartoons and colored according to the conformations red, blue and green. The 50 adsorption modes as returned from each docking simulation runs are ranked, from the top to the bottom, in 4 main clusters, from the highest to the lowest populated (%). The most populated cluster for each Ctx conformation results the same (blue frame) with the CNT nested in the cavity between Fab1 and Fc whose graphical representation is reported in c) coloring the CNT in grey and the Ctx in magenta.

Specifically, the most populated docked adsorption mode (≥50% occupancy), recurring in all the simulations for CNT20, develops the CNT framework nested in the cavity between the Fab1 and Fc regions (i.e., Fab-Fc cavity), in which hydrophobic interactions are maximized (Figure 3.8b). This adsorption mode obviously depends on the CNT's diameter, suggesting that its dimension plays a critical role in determining, among the most likely positions, the final structure. This means that such prediction is diameter-dependent and not significantly representative of the supramolecular interactions between all class of CNTs and the Ctx.

Consequently, a systematic docking analysis with CNTs of increasing diameters (*d*CNTs = 30, 40, 60, and 80 Å, labeled as CNT30, CNT40, CNT60, and CNT80, respectively) was carried out on the Ctx equilibrated geometry (Figure 3.9a). From this systematic analysis, it clearly emerges that, increasing the CNTs' diameter, the predicted interaction modes for the small-diameter CNTs converge to a unique cluster in which the CNTs move away from the Fab-Fc cavity. In fact, CNTs featuring *d*CNTs ≥ 80 Å interact with the Ctx only through the Fc region (pose-fc, *p*-fc), leaving unaltered the Fabs regions (Figure 3.9b).

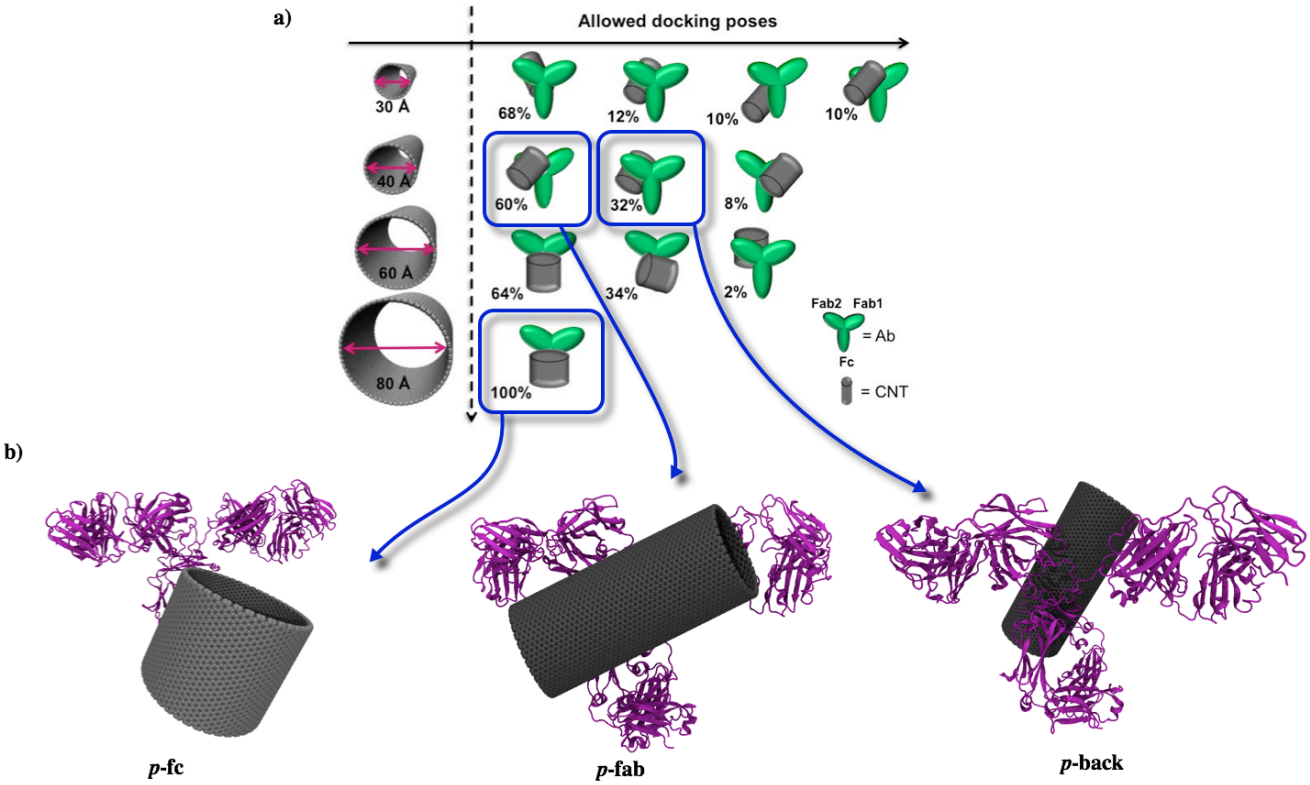


Figure 3.9. a) Systematic docking studies for different CNTs of increasing diameter values (30, 40, 60 and 80 Å labeled as CNT30, CNT40, CNT60 and CNT80, respectively). The predicted adsorption poses, sketched as cartoons in each row, are grouped into clusters from the highest (left) to the lowest (right) populated one. b) The three different docking poses selected to perform the MD simulations are, from the left to the right, *p*-fc (100%, as obtained from CNT80, blue frames), *p*-fab (60%, as obtained from CNT40) and *p*-back (32%, as obtained from CNT40) the latter pose being recurrent also for CNT30 and CNT60. The CNT is colored in grey and Ctx in magenta.

Notably, in any of the discussed cases, the presence of the CNT structure does not seem to sterically affect the conformational or chemical properties of the Fabs, thus further supporting the experimental biochemical findings for which the Ctx's binding properties are largely preserved in the hybrids. However, since an evident diameter-dependent orientation behavior has been also observed, we decided to not exclude other probable orientations that could possibly provide further structural insights into the CNT•Ctx supramolecular association. In fact, we have decided to consider two more poses, namely those derived from the interaction of Ctx with smaller CNTs (Figure 3.9b). Specifically, the pose in which the tube interacts with the

frontal Fab2 domain of Ctx (pose-fab, *p*-fab) and that with the CNT interacting on the Ab's backside with Fab2 (pose-back, *p*-back, Figure 3.9b).

CNT20 model has been used for the MD simulations. This approximation allows us to reduce the computational costs and run all-atoms simulations of systems of about 400,000 atoms on our computational resources. For each pose, we performed 40 ns of MD simulations in explicit water. Due to their large size the simulated systems are not fully equilibrated at this time. However, we are confident that this does not qualitatively affect our findings. At a visual inspection, the *p*-fc is stable during the entire simulation time. In fact, for the *p*-fc, the Fc region is largely stabilized as observed by the RMSD plot (Figure 3.10a), whereas in the free Ctx trajectory the Fc domain fluctuates more than the two Fabs (Figure 3.6b). This suggests that the presence of the graphitic surface stabilizes the Fc domain. On the contrary, the two Fab domains, which are relatively far from the interaction site with the graphitic surface, display higher and fluctuating RMSD values in the simulated timeframe (40 ns).[2] Notably, at 5 ns an abrupt increase of the RMSD is observed for all three domains, suggesting significant conformational rearrangements of the Fab domains. An additional peak is also observed at ~26 ns, where the two Fabs sequentially approach the Fc region. The structural rearrangements, occurred during the trajectory of the *p*-fc, are clearly visible if we compare the initial and the last snapshots of the simulation. Namely, the two Fabs get closer to the Fc region (Figure 3.10a) and tangentially approach the CNTs' framework without adhering to it.

[2] Due to the large size of the systems we were not able to run longer simulations on standard computational resources.

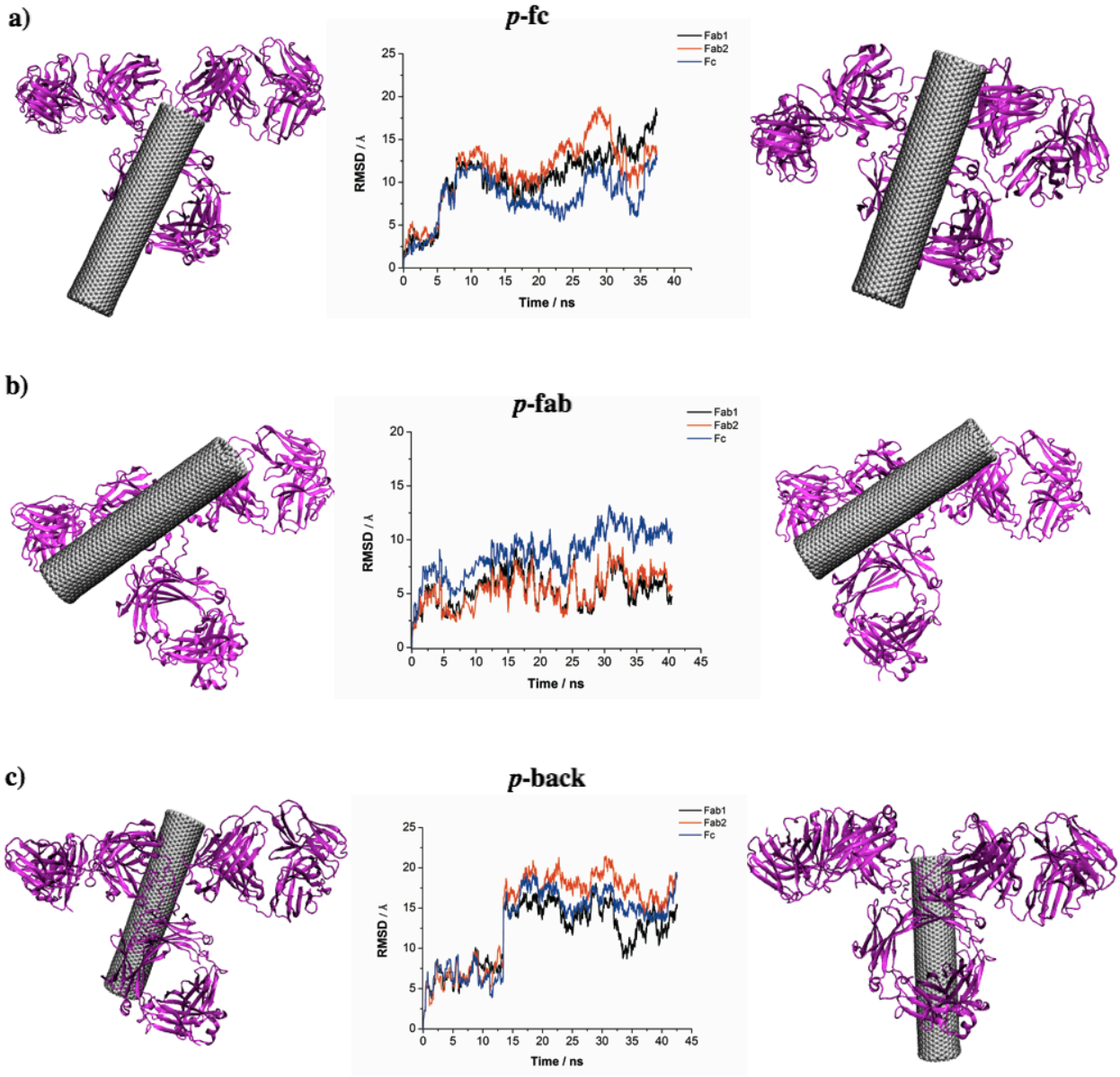


Figure 3.10. The starting structure (left panel) of the MD simulation, the RMSD resulting from the 40 ns MD trajectories (middle panel), and the last snapshot (right panel) of the simulation for *p*-fc (a), *p*-fab (b) and *p*-back (c) poses. In each plot, the black line represents the Fab1 region, the red the Fab2 and the blue the Fc one.

The lack of an intimate contact leaves the Fabs free to move, and thus still undergoing significant fluctuations, as shown by the calculated RMSF (Figures 3.11a-3.12a). In principle, the fact that the two Fabs are highly mobile and not hindered by the presence of the graphitic surface suggests that both domains are structurally and sterically unaffected by the presence of the graphitic tube. Thus, most probably, the two Fabs conserve their natural binding properties towards EGFR as observed in our experiments with EGFR-overexpressing cancer cells.

Notably, despite the considerable changes in the relative positions of the carbon framework with respect to the Ctx, the secondary structure of Ctx is not affected, conversely gaining few residues more in β-sheet and α-helix as compared to Ctx alone (Table 3.2). To localize the effect exerted by the tubular framework on the Ab flexibility, we have calculated the difference in RMSF (ΔRMSF = $RMSF_{(CNT.Ctx)}$-$RMSF_{(Ctx)}$) for the antibody residues in each docking pose along the MD trajectories (Figure 3.12). The ΔRMSF displays the set of amino acidic residues that are stabilized or destabilized by the interaction with the CNT. As a general trend, one can observe that the fluctuations of the amino acidic residues interfacing the graphitic wall are significantly reduced upon formation of the CNT•Ctx complex (Figures 3.12 and A.3a, *Appendix*). In particular, for *p*-fc pose, the decrease of the fluctuations is limited to residues belonging to Fab1 and Fc domains (residues 100-200 and 400-500, respectively). An opposite situation has been observed for the trajectory of pose *p*-fab. Here, the Fab regions are stabilized by the direct interaction with the CNT external wall, as shown by the RMSD and ΔRMSF plots (Figures 3.10b-3.11b and 3.12b). In fact, the RMSD displays a quasi-homogeneous trend for the three domains, with a low average value for those regions intimately contacting the graphitic surface. The amino acidic residues involved in the interaction in *p*-fab belong to Fab1 (residues from 100-210 and 300-410) and Fab2 domains (residues from 750-850) and, as previously observed for pose *p*-fc, they display a reduced fluctuation in the CNT•Ctx complex (Figures 3.12b-A.3b, *Appendix*). As observed above, also for this pose the interaction mode of the CNT framework does not affect Ctx's secondary structure, which essentially remains unaltered during the entire MD simulation (Table 3.2).

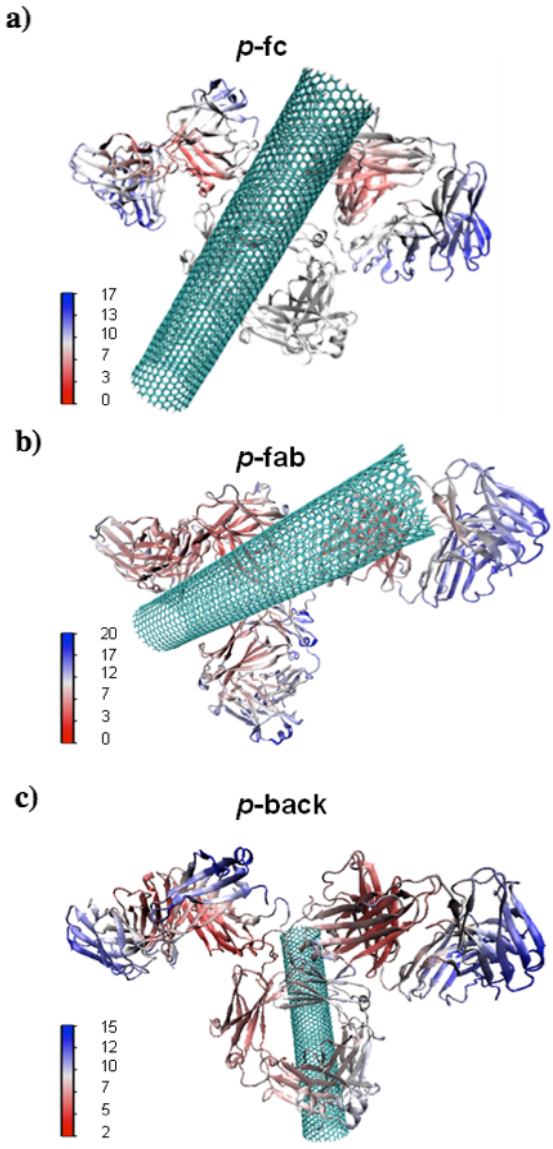


Figure 3.11. RMSF of the backbone atoms of Ctx, calculated over the 40 ns MD trajectory for p-fc (a), p-fab (b) and p-back (c) poses. Red (blue) colors describe protein regions that are more (less) rigid.

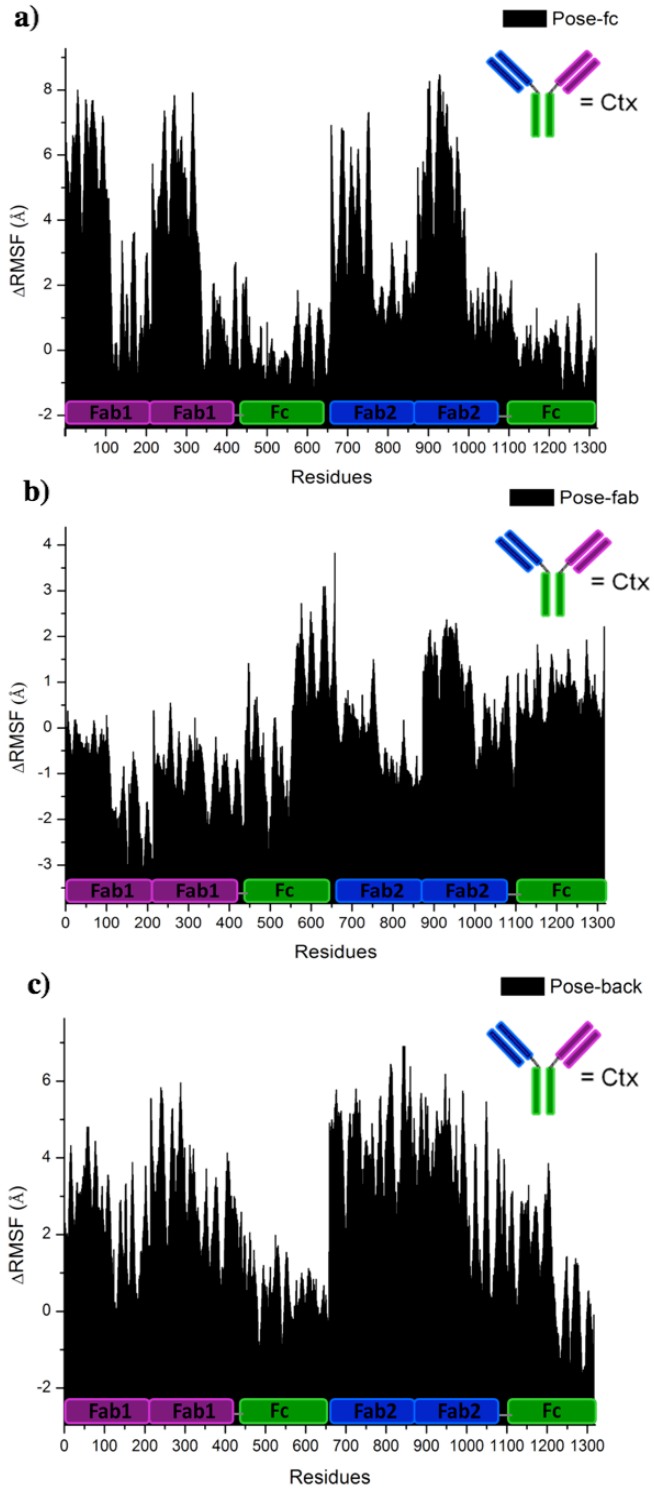


Figure 3.12. Calculated ΔRMSF for each Ctx residue (axe x) contacting the CNT surface along the trajectories for the *p*-fc (a), *p*-fab (b) and *p*-back (c) poses. The bottom bars (magenta for Fab1, blue for Fab2 and green for Fc) indicate the different domains to which the residues belong. Negative (positive) values refer to residues that are more (less) rigid.

Surprisingly, for the *p*-back mode a remarkable rearrangement of the CNT•Ctx complex is observed during the simulation (Figure 3.10c). Specifically, a translocation of the carbon framework from an initial central position on the backside of the Ctx body to a pose in which the graphitic wall uniquely interacts with the Fc back region occurred. Clearly, the chosen adsorption mode was not a stable binding pose, consistently with the fact that in the docking simulations a CNT of a larger diameter was used. Nevertheless, the tube gets closer to the Fc region unaffecting the Fab domains and leaving once again the EGFR recognition moieties free. The RMSD of the overall *p*-back is characterized by a remarkable flexibility. In particular, at about 14 ns of the trajectory, a sizable increase of the RMSD occurs due to the pose rearrangement evolving towards the Fc domain (Figure 3.10c). After this, the Fc region is stabilized by the interaction with the graphitic surface as confirmed by the low RMSF (Figures 3.11c-3.12c). In addition, ΔRMSF analysis for this pose suggests that the Fab domains significantly fluctuate in the CNT•Ctx complex, in contrast to the Fc residues, which feature a reduced mobility (Figure 3.12c). No loss of the secondary structure was observed for this pose (Table 3.2).

Of note, despite the fact that the CNT framework could easily interact through its terminal edges with the antibody, the rearrangements observed for the three poses do not involve the H-saturated CNT tips,[62] but it instead tends to interact with Ctx through its external sidewalls. As also shown by the docking studies describing the diameter-dependent behavior, this aspect also validates the general concept for which the CNT•Ctx supramolecular interaction is dominated by the protein adsorption on the most available portion of the CNT surface area. On the other hand, we are aware that this could be caused by the approximation of using a SWCNT at the place of a MWCNT.

To elucidate the dynamics of the absorption process, we also analyzed the variation of the Solvent Accessible Surface Area (ΔSASA) during the simulations. The largest decrease in the SASA is observed for *p*-fab, suggesting that a larger number of residues interacting with the graphitic wall is present in this pose than in the other two.[63] Clearly, the more the CNT framework and Ctx interact, the less the SASA is available and the largest number of residues are involved in the adsorption. A significant decrease in the SASA is observed for *p*-fc, whereas it revealed to be virtually unaffected for *p*-back (Figure 3.13).

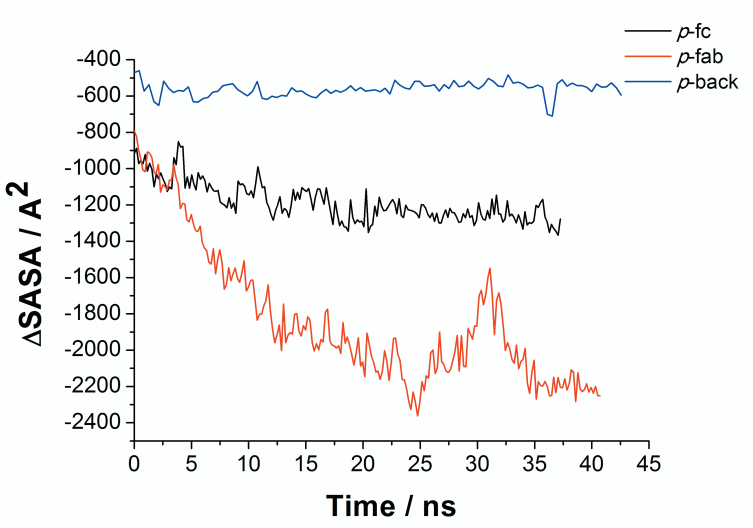


Figure 3.13. Calculated ΔSASA *vs* simulation time for *p*-fc (black line), *p*-fab (red line) and *p*-back (blue line) poses.

In order to determine at the qualitative level the energetically-favored binding pose, we calculated the non-bonded interaction energies, namely the electrostatic (EEL) and van der Waals (vdW), between Ctx and the CNT surface over the last 25 ns of the MD trajectories. These energies are highly qualitative due to the approximate nature of the force-field. However a comparison of the interaction energies of the different docking poses can provide qualitative information on the relative stabilities of the CNT•Ctx complexes.

Table 3.3. Summary of the average non-bonded interaction energies, electrostatic (EEL) and van der Waals (vdW) calculated for the three poses. Standard deviations are also reported.

Pose	<EEL>[a]	<vdW>[a]
***p*-fc**	-5 ± 2	-131 ± 6
***p*-fab**	-8 ± 4	-267 ± 21
***p*-back**	-1 ± 1	-58 ± 5

[a] (kcal mol^{-1}).

As expected, the EEL contribution is relatively small if compared to the vdW non-bonded energy for all poses (Figure 3.14a). However, both *p*-fc and *p*-fab poses show non-negligible EEL energies with respect to *p*-back, whose value is about 0 (Table 3.3). As it clearly appears from the calculated non-bonded interaction energies reported in Table 3.3, the hydrophobic nature of CNTs predominantly drives the formation of the complex through vdW interactions (Table 3.3 and 3.14b).[64]

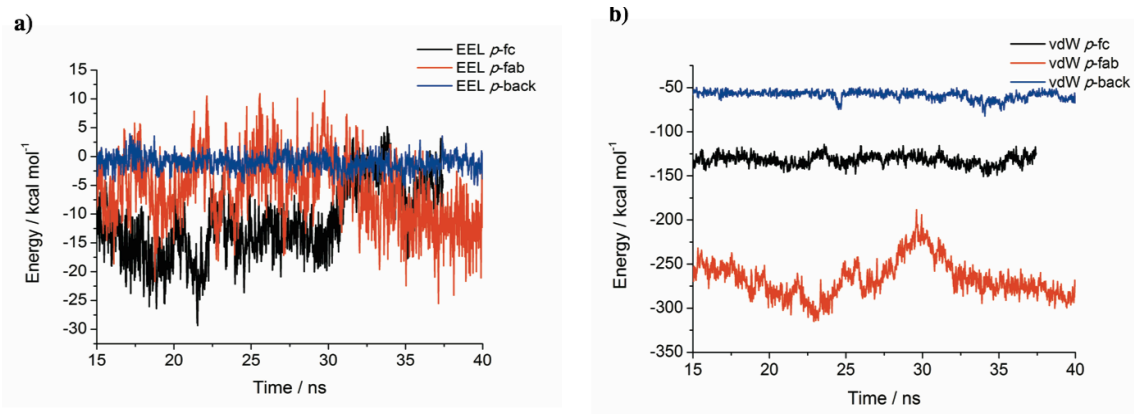


Figure 3.14. Electrostatic (EEL) and van der Waals (vdW) interaction energies (in (a) and (b), respectively) calculated over the last 25 ns of the trajectories of the three binding poses under study. The energies are depicted in black, red and blue for *p*-fc, *p*-fab and *p*-back, respectively.

Large differences in the vdW energy have been observed for the different binding poses. In particular, pose *p*-fc features an average vdW energy value of -131 ± 6 kcal mol^{-1}, while *p*-fab is characterized by the largest vdW interaction energy of -267 ± 21 kcal mol^{-1}. Notably, for *p*-fab a significant change of both vdW and EEL energy is observed at 30 ns (Figure 3.14), consistently with a rearrangement of the complex, which allows maximizing the interactions between Ctx and CNTs. Instead, *p*-back pose displays a reduced average value for the vdW energy of -58 ± 5 kcal mol^{-1}. In strong agreement with the docking prediction, the reduced vdW energy value also suggests that the *p*-back interacting mode is certainly the less stable pose. One can thus conclude that both *p*-fab and *p*-fc poses are representative interaction modes for the supramolecular CNT•Ctx complex and that the non-bonded interaction energy is stronger within the *p*-fab pose. This could in principle account for diameter-dependent simulations for which the binding pose depends on the chosen tube's diameter. In fact, while for CNT20 the tubular framework can fit in a superficial nook of the antibody, CNTs with larger diameters (namely CNT60-80) are hampered, as they cannot interlock with the Ctx in the same fashion. For this reason, the *p*-fab pose can be reasonably considered as the most representative interaction mode for small-diameter CNTs (≤40Å). On the other hand, the trend displayed by *p*-fc pose results a plausible and stable way of interacting that, as predicted by the docking studies, could occur when the antibody interacts with bigger CNTs. Thus, as it has been intuitively reveled by the ELISA binding experiments (see section 3.1.2), the non-covalent structure for the CNT•Ctx complexes depends on the size and dimension of CNTs.

Hence, since both the predicted poses are reliable, we can reasonably conclude that, under the current experimental conditions in which a heterogeneous mixture of CNTs is used, the Ctx proteins facing small diameter CNTs probably interact in *p*-fab fashion, whereas with bigger diameter CNTs in a *p*-fc pose. To further rationalize at the molecular level the physical origin of the supramolecular interactions between the graphitic surface and the Ctx domains, the energetic contribution of each amino acid at the CNT•Ctx interface has been calculated. Clearly, as observed by MD simulations, the electrostatic contribution was negligible (Table 3.4), while significant is the role of the vdW interactions (Figure 3.15 and Table 3.4). In *p*-fab, all three domains are involved in the adsorption and we found fourteen residues mainly contributing to the formation of this pose.

Table 3.4. Average electrostatic (<EEL>) and van der Waals (<vdW>) interaction energies (kcal mol^{-1}) of selected residues and CNT, corrected by their standard deviations are reported for the three binding poses.

	***p*-fab**			***p*-fc**			***p*-back**		
Ctx Domain	Residue	<EEL>	<vdW>	Residue	<EEL>	<vdW>	Residue	<EEL>	<vdW>
Fab1	Gly166	-0.3±0.5	4.8±0.8	Arg202	-2.7±2.2	10.9±4.0	-	-	-
	Ser167	0.01±0.3	6.1±0.9	-	-	-	-	-	-
	Lys213	-0.3±4.1	8.3±4.3	-	-	-	-	-	-
	Thr216	-0.4±0.7	6.8±2.5	-	-	-	-	-	-
Fab2	Thr728	-0.6±0.7	7.8±1.2	-	-	-	-	-	-
	Val730	0.3±0.6	5.7±0.9	-	-	-	-	-	-
	Pro731	0.02±0.1	7.7±0.9	-	-	-	-	-	-
	Lys841	-0.2±0.5	9.9±2.0	-	-	-	-	-	-
	Ser1025	-0.2±0.4	5.6±0.5	-	-	-	-	-	-
	Glu753	0.1±0.4	4.1±2.4	-	-	-	-	-	-
	Thr1075	0.03±0.2	-4.4±1	-	-	-	-	-	-
Fc	Val511	-0.5±0.4	-6.4	Ser483	0.3±0.7	-6.9±4	Asp1159	-0.4±0.7	-7.6±1
	Glu512	0.4±1.1	-7.7	Leu484	0.3±0.4	-7.6±0.9	Thr1178	-0.3±0.6	-8.1±0.9
	Lys555	-0.4±0.7	-10.6	His521	-0.4±0.4	-0.8±1.5	His1179	-0.1±0.6	-8.7±0.9
	-	-	-	Arg522	-0.6±0.6	12.0±1.8	Arg1180	0.02±0.6	14.9±0.8
	-	-	-	Glu523	0.1±0.6	-7.8±0.8	-	-	-
	-	-	-	Asp524	-0.3±0.5	-6.4±0.9	-	-	-
	-	-	-	Ser485	-0.03±0.5	-5.9±0.6	-	-	-
	-	-	-	Gln519	-0.1±0.3	-7.3±1.2	-	-	-
	-	-	-	Thr527	-0.3±0.4	-9.9±2.5	-	-	-

[a] (kcal mol^{-1}).

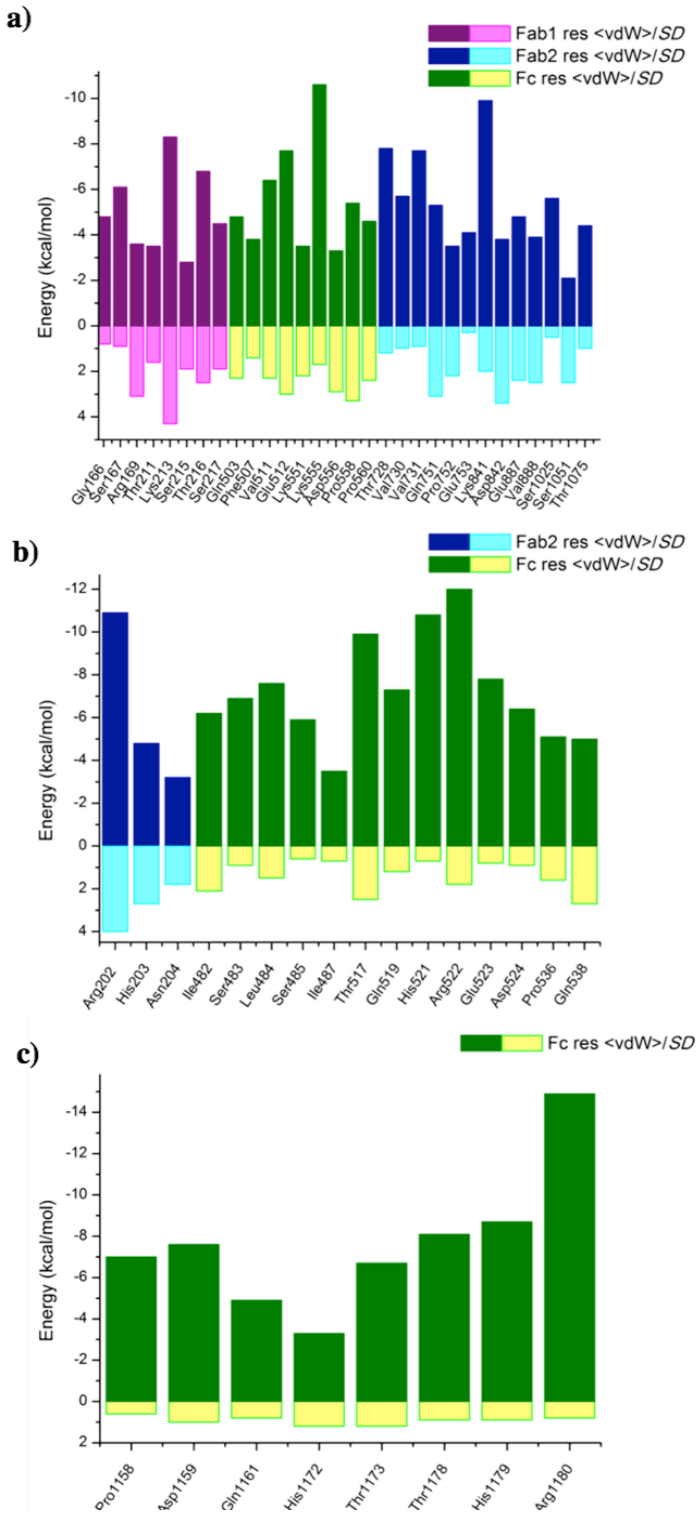


Figure 3.15. Average vdW energy (<vdW>) and standard deviation (SD) calculated for each residue at the CNT•Ctx interface, for the pose *p*-fab (a), *p*-fc (b) and *p*-back (c). In each plot, the x axis reports the residues of Ctx and the bars in different colors refer to the globular domains of Ctx to which the residues belong (Fab1-violet, Fab2-blue, Fc-green).

Surprisingly, the most relevant contribution to the vdW interaction energy is given by positively charged amino acidic residues, such as Lys213, Lys555 and Lys841, belonging to Fab1, Fab2 and Fc regions, respectively (Figure 3.15a).[63,65,66] A non-negligible contribution from negatively charged residues such as Glu753 and Glu512, polar residues such as Ser167 and Thr728 and hydrophobic residues such as Val511 and 730 (Table 3.4) has been also observed. Conversely, the residues stabilizing the *p*-fc pose are mainly located on the Fc region (Figures 3.12a-3.15b-A.3a, *Appendix* and Table 3.4). In addition, in this case the largest contribution to the total interaction energy is given by the positively charged residues (Arg202 and Arg522). Selected polar residues such as His521, Thr527, Ser483 and Ser485 also contribute significantly along with a few negatively charged amino acids (Glu523 and Asp524). Finally, the *p*-back pose displays a very small number of amino acids giving a relevant contribution to the vdW interaction energy, thus supporting our hypothesis that this pose is the less favored (Figure 3.15c). Also in this case, the higher contribution derives from the charged (Arg1180 and Asp1159) and the polar (His1179 and Thr1178) fragments.

The importance of positively charged residues in the stabilization of hybrid biomaterials formed by carbon nanostructures such as nanotubes or fullerenes has been already reported in the literature[63,65–68] (see Chapter 1), and their effect is greater than usually thought.[69,70] In fact, it has been recently reported that polar and charged residues can interact with hydrophobic surfaces, maintaining their H-bonding abilities with the solvent, acting, therefore, like surfactants.[63] Consistently with these findings, a visual inspection of the adsorption site shows that the Lys, Arg and Glu residues interact with the graphitic surface through the hydrophobic chains (Figures 3.16), apparently ruling out any electrostatic or polarization effects.

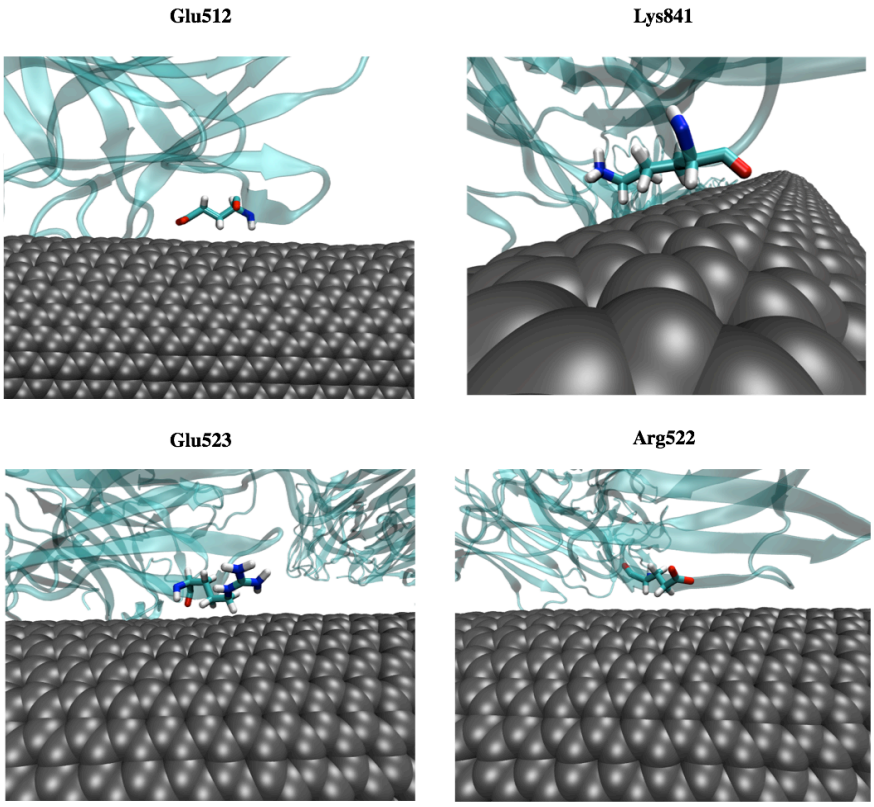


Figure 3.16. Orientations of the residues giving the largest contribution to the vdW interaction energy with the CNT: the Glu512 and Lys841 within *p*-fab (top), the Glu523 and Arg522interacting with CNT within the *p*-fc (bottom).

In order to probe the existence of polarization effects, the dipole moments (μ), the Bader charges and ΔH of binding between the graphitic surface and the selected residues have been estimated by means of first-principles DFT calculations *in-vacuo* on models of reduced size (Table 3.5). For the *p*-fab pose, the selected residues are Lys555, Lys841 and Glu512, while for *p*-fc His521, Arg522 and Glu523 were chosen. Pose *p*-back was excluded from this analysis because it is the less-likely binding mode. The orientation of the residues was chosen from the snapshot of the trajectory for which each amino acid has the largest vdW interaction energy value as calculated previously (Table 3.4). From the energetic point of view, the contribution of the selected residues was confirmed by DFT calculations, displaying the same trend as that observed for the classical MD simulations. The residues with the largest binding energy are those bearing a long alkyl chain, namely the Lys555 and Lys841 for p-fab and Arg522 for p-fc. Furthermore, monitoring the molecular dipole moment (μ) and the Bader charges located on the graphitic surfaces in the presence or in the absence

of each residue (Table 3.5), only a faintly polarization has been detected on the carbon exosurface by the charged or polar residues (Tables A.1 and A.2). No correlation could be found between the dipole moments and the ΔH values (Table 3.5), thus suggesting that the prevalent interactions between the amino acidic residues and the graphitic surface are essentially governed by the hydrophobic alkyl chains.[63,71–74]

Table 3.5. Dipole moment (μ) and ΔH of binding between the amino acids residues and the CNT. From left to right we report μ calculated for the CNT fragment alone (**CNT**), for the CNT fragment in the presence of the residue (**CNT**-RES), for the residue alone (**RES**), for the residue in the presence of the CNT (CNT-**RES**) and for the complex (**CNT-RES**).

Pose	Residue	μ[a]					ΔH[b]
		CNT	**CNT**-RES	**RES**	CNT-**RES**	**CNT-RES**	
p-fab	Lys841	1.20	1.51	19.91	19.95	14.51	-21.47
	Lys555	0.76	0.99	20.27	20.30	14.77	-19.02
	Glu512	0.91	1.09	18.69	18.49	12.40	-13.73
p-fc	Arg522	0.74	0.88	24.53	24.56	21.06	-10.93
	His521	0.83	1.07	3.16	3.17	2.30	-9.11
	Glu523	1.24	1.37	25.49	25.61	18.07	-6.20

[a] (D). [b] (kcal mol^{-1}).

3.1.6 Generalization of the model

An overview on the differences and similarities within the vast world of mAbs might draw interesting inferences on the structural prediction of the proposed hydrophobic-driven interaction model. As we have seen in section 3.1.3.1.1, IgG is the major antibody in the blood and exerts a fundamental role in the immune response. Additionally, it is worth recalling that this class of IgG-derived antibodies is the most exploited for engineering therapeutic monoclonal antibodies.[30] In fact, nearly 30 IgG-like mAbs have been approved by the FDA in the US and/or Europe and several others are in clinical developments (Table 3.6).[21] For these reasons, we have limited our structural screening on the human IgG Ab isotype.

Table 3.6. FDA approved therapeutic monoclonal antibodies.

Name	Licensed indication(s) and year of first approval-region	Origin and isotype
Capromab pendetide	Diagnostic imaging (1996-US)	Murine IgG1
Muronomab-CD3	Transplant rejection (1992-US	Murine IgG2a
Tositumumab iodine 131	NHL[a] (2003-US)	Murine IgG2
Ibritumomab tiuxetan	NHL (2002-US, 2004-EU)	Murine IgG1
Basiliximab	Prophylaxis for transplant rejection (1998-US)	Chimeric IgG1
Brentuximab vedotin	ALCL[b] and Hodgkin lymphoma (2011-US)	Chimeric IgG1
Catumaxomab	Malignant ascites (2009-EU)	Chimeric IgG2a/b
Cetuximab	Colorectal, head and neck cancer (2004-US, EU)	Chimeric IgG1
Infliximab	RA[c], ankylosing spondylitis, Crohn's disease, ulcerative colitis (1998-US, 1999-EU)	Chimeric IgG1
Rituximab	B cell non-Hodgkin lymphoma (1997-US, 1998-EU)	Chimeric IgG1
Alemtuzumab	B-CLL[d] (2001-US)	Humanized IgG1
Bevacizumab	Colorectal, lung, breast cancer (2004-US, 2005-EU)	Humanized IgG1
Daclizumab	Prophylaxis for transplant rejection (1997-US, 1999-EU (withdrawn 2009))	Humanized IgG1
Eculizumab	Paroxysmal nocturnal haemoglobinuria, atypical haemolytic-uremic syndrome (2007-US)	Humanized IgG2/4
Efalizumab	Psoriasis (2003-US (withdrawn 2009))	Humanized IgG1
Gentuzumab Ozogamcin	Leukemia (2000-US)	Humanized IgG4
Natalizumab	Multiple sclerosis (2004-US, 2006-EU)	Humanized IgG4
Omalizumab	Severe asthma (2003-US, 2005-EU)	Humanized IgG1
Palivizumab	Prevention of RSV infection in neonates (1998-US)	Humanized IgG1
Ranibizumab	Macular degeneration (2006-US, 2007-EU)	Humanized IgG1
Tocilizumab	Castleman's syndrome, RA (2010-US, 2009-EU)	Humanized IgG1
Trastuzumab	HER-2 positive breast cancer (1998-US)	Humanized IgG1
Adalimumab	RA, Crohn's disease, plaque psoriasis (2002-US, 2003-EU)	Human IgG1
Belimumab	Systemic lupus erythematosus	Human IgG1
Canakinumab	Muckle-Wells syndrome (US, EU-2009)	Human IgG1
Denosumab	Osteoporosis, bone metastasis (2009-US, EU	Human IgG1
Golimumab	RA, psoriatic arthritis, ankylosing spondylitis (2009-US, EU	Human IgG1
Ipilimumab	Advanced melanoma (2011-US, EU)	Human IgG1
Ofatumumab	CLL (2009-US, 2010-EU)	Human IgG1
Panitumumab	Colorectal cancer (2007-US, EU)	Human IgG1
Ustekinumab	Plaque psoriasis (2009-US, EU)	Human IgG1
Pertuzumab	HER-2 positive breast cancer (2012-US	Humanized IgG1

[a] NHL, non-Hodgkin lymphoma;
[b] ALCL, anaplastic large cell lymphoma;
[c] RA, rheumatoid arthritis;
[d] B-CLL, B-cell chronic lymphocytic leukemia.

First, it is interesting to note that, despite their heterogeneity and difference in the targeting properties in all reported structures (see section 3.2.3.1.1) the entire protein exhibits the same multiglobular Y-shape architecture (Figures 3.17a-c).

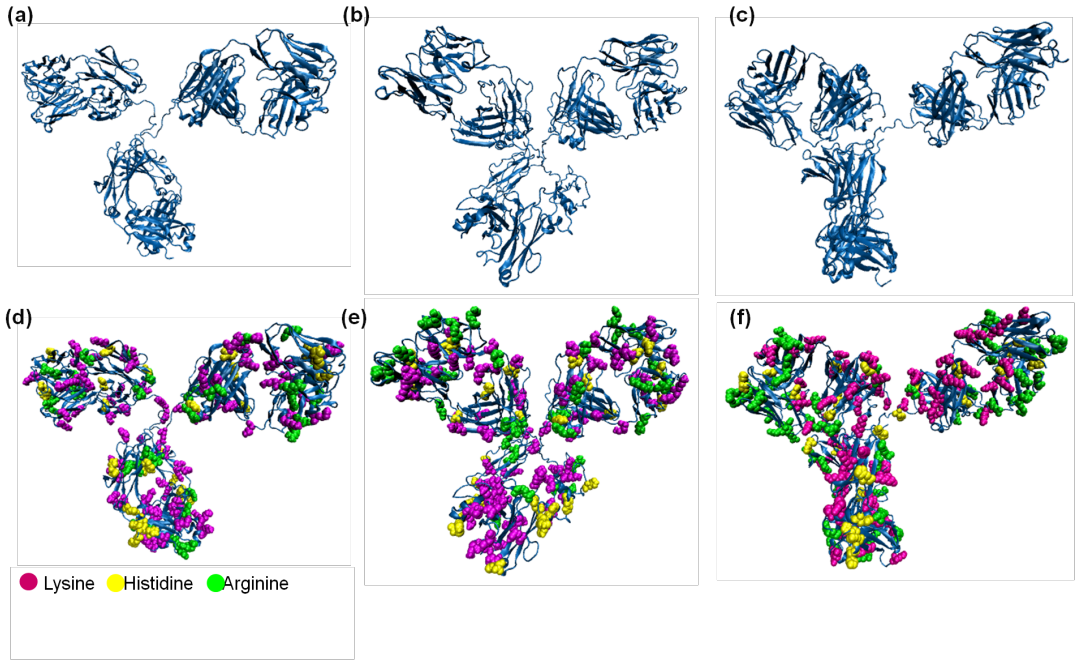


Figure 3.17. Y-shaped structure of the crystal structure 1IGT (a), 1IGY (b) and 1HZH (c) as a common feature. Influent residues distribution on the Ab surface of the crystal structure 1IGT (d), 1IGY (e) and 1HZH (f). The lysines (magenta), the histidines (yellow) and the arginines (green) are represented in vdW equally distributed over the Abs surfaces.

Furthermore, when the amino acidic sequence of the Ab 1IGT model, used for our simulations, is compared to the other available intact Ab crystal structures, a considerable sequence similarity is observed (Table 3.7) for different isotype.

Table 3.7. Sequence similarity between the Ab crystal structures and 1IGT. The PDB codes marked with asterisk (*) refer to multi-methods derived crystal structures.

PDB code	Ab isotype	Sequence similarity[a] to 1IGT	
		L-chains	H-chains
1IGY	Murine IgG1	70	50
1HZH	Human IgG1k	70	50
1MCO	Human IgG1	< 30	50
1R70*	Human IgA	< 30	40
1ZVO*	Human IgD	< 30	< 30

[a] %

In fact, a very high (70%) sequence similarity is detected between 1IGT and different IgG subclasses, as shown by the very similar composition and distribution of the amino acid residues playing a role in the formation and stabilization of the CNT•Ctx conjugate; namely, polar (His), negatively (Glu and Asp) and positively (Arg and Lys) charged residues. All the interacting residues within the CNT•Ctx complex are conserved in all Abs structures regardless the subclass or the isotype (Table 3.8) and display the same equal distribution within the Abs surface as observed for 1IGT (Figure 3.17d-f). Furthermore, when the amino acidic sequence of the Ab 1IGT model, used for our simulations, is compared to the other available intact Ab crystal structures, a considerable sequence similarity is observed for different isotype (IgA and IgD, Table 3.7 and Figure 3.18). Besides, as Ctx, virtually all monoclonal mAbs approved by the FDA and currently in use, belong to IgG1 isotype subclass. Hence, taking into account the observed structural similarities, the 1IGT model used herein is structurally representative for all the mAbs listed in Table 3.6 and, possibly, it is also representative in its way of interfacing with the curved graphitic surface of CNTs.

Table 3.8. Number of amino acids of interest within the whole Ab crystal structures and 1IGT. The PDB codes marked with asterisk (*) refer to multi-methods derived crystal structures.

PDB code	Ab isotype	Influent residues				
		Glu	Asp	His	Arg	Lys
1IGT	Murine IgG2a	60	60	28	38	82
1IGY	Murine IgG1	62	60	30	40	86
1HZH	Human IgG1k	66	56	29	54	88
1MCO	Human IgG1	56	40	22	26	86
1R70*	Human IgA	66	54	24	51	63
1ZVO*	Human IgD	76	50	26	67	63

Therefore, we can likely postulate that all Abs belonging to the five isotypes could in principle interact with CNTs in the same manner as that proposed for CNT•Ctx complexes, namely through the hydrophobic tail of charged and polar amino acids in such a way that the globular Y-shape is retained.

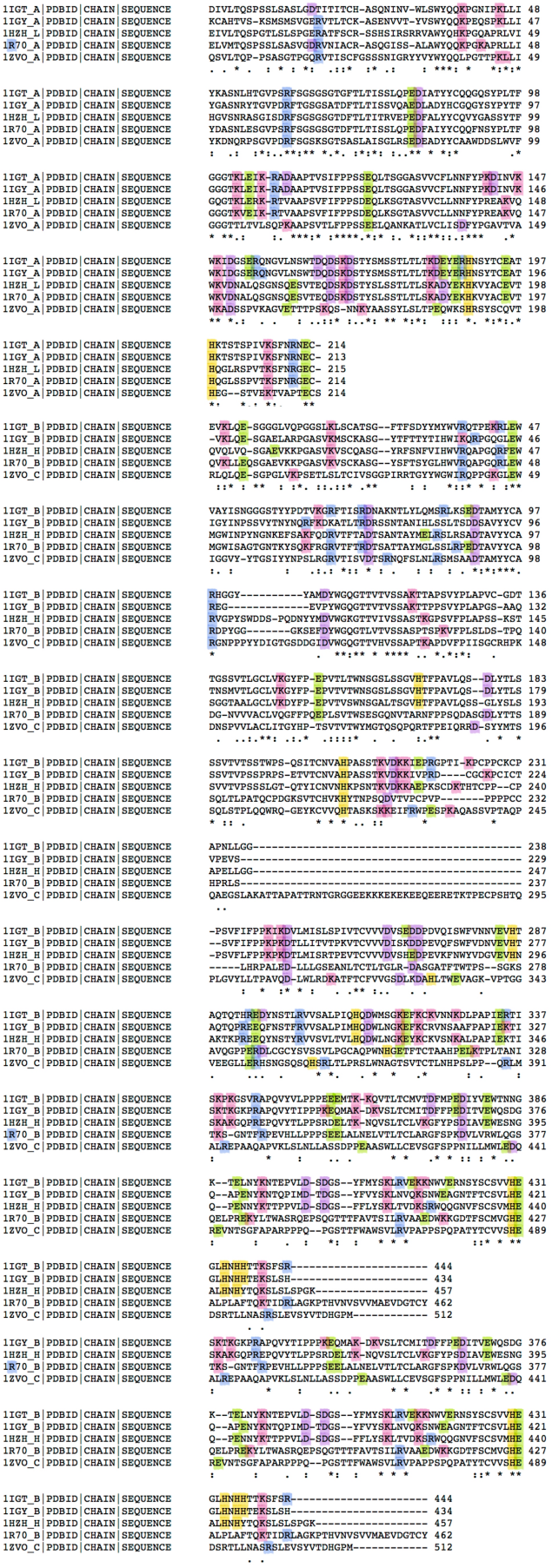

1IGT_A|PDBID|CHAIN|SEQUENCE DIVLTQSPSSLSASLGDTITITCH-ASQNINV-WLSWYQQKPGNIPKLLI 48
1IGY_A|PDBID|CHAIN|SEQUENCE KCAHTVS-KSMSMSVGERVTLTCK-ASENVVT-YVSWYQQKPEQSPKLLI 47
1HZH_L|PDBID|CHAIN|SEQUENCE EIVLTQSPGTLSLSPGERATFSCR-SSHSIRSRRVAWYQHKPGQAPRLVI 49
1R70_A|PDBID|CHAIN|SEQUENCE ELVMTQSPSSLSASVGDRVNIACR-ASQGISS-ALAWYQQKPGKAPRLLI 48
1ZVO_A|PDBID|CHAIN|SEQUENCE QSVLTQP-PSASGTPGQRVTISCFGSSSNIGRYYVYWYQQLPGTTPKLLI 49
1IGT_A|PDBID|CHAIN|SEQUENCE YKASNLHTGVPSRFSGSGSGTGFTLTISSLQPEDIATYYCQQGQSYPLTF 98
1IGY_A|PDBID|CHAIN|SEQUENCE YGASNRYTGVPDRFTGSGSATDFTLTISSVQAEDLADYHCGQGYSYPYTF 97
1HZH_L|PDBID|CHAIN|SEQUENCE HGVSNRASGISDRFSGSGSGTDFTLTITRVEPEDFALYYCQVYGASSYTF 99
1R70_A|PDBID|CHAIN|SEQUENCE YDASNLESGVPSRFSGSGSGTDFTLTISSLQPEDFAIYYCQQFNSYPLTF 98
1ZVO_A|PDBID|CHAIN|SEQUENCE YKDNQRPSGVPDRFSGSKSGTSASLAISGLRSEDEADYYCAAWDDSLWVF 99
1IGT_A|PDBID|CHAIN|SEQUENCE GGGTKLEIK-RADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVK 147
1IGY_A|PDBID|CHAIN|SEQUENCE GGGTKLEIK-RADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVK 146
1HZH_L|PDBID|CHAIN|SEQUENCE GQGTKLEIK-RTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQ 148
1R70_A|PDBID|CHAIN|SEQUENCE GGGTKVEIK-RTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQ 147
1ZVO_A|PDBID|CHAIN|SEQUENCE GGGTTLTVLSQPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVA 149
1IGT_A|PDBID|CHAIN|SEQUENCE WKIDGSERQNGVLNSWTDQDSKDSTYSMSSTLTLTKDEYERHNSYTCEAT 197
1IGY_A|PDBID|CHAIN|SEQUENCE WKIDGSERQNGVLNSWTDQDSKDSTYSMSSTLTLTKDEYERHNSYTCEAT 196
1HZH_L|PDBID|CHAIN|SEQUENCE WKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVT 198
1R70_A|PDBID|CHAIN|SEQUENCE WKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVT 197
1ZVO_A|PDBID|CHAIN|SEQUENCE WKADSSPVKAGVETTTPSKQS-NNKYAASSYLSLTPEQWKSHRSYSCQVT 198
1IGT_A|PDBID|CHAIN|SEQUENCE HKTSTSPIVKSFNRNEC- 214
1IGY_A|PDBID|CHAIN|SEQUENCE HKTSTSPIVKSFNRNEC- 213
1HZH_L|PDBID|CHAIN|SEQUENCE HQGLRSPVTKSFNRGEC- 215
1R70_A|PDBID|CHAIN|SEQUENCE HQGLSSPVTKSFNRGEC- 214
1ZVO_A|PDBID|CHAIN|SEQUENCE HEG--STVEKTVAPTECS 214
1IGT_B|PDBID|CHAIN|SEQUENCE EVKLQE-SGGGLVQPGGSLKLSCATSG--FTFSDYYMYWVRQTPEKRLEW 47
1IGY_B|PDBID|CHAIN|SEQUENCE -VKLQE-SGAELARPGASVKMSCKASG--YTFTTYTIHWIKQRPGQGLEW 46
1HZH_H|PDBID|CHAIN|SEQUENCE QVQLVQ-SGAEVKKPGASVKVSCQASG--YRFSNFVIHWVRQAPGQRFEW 47
1R70_B|PDBID|CHAIN|SEQUENCE QVKLLEQSGAEVKKPGASVKVSCKASG--YSFTSYGLHWVRQAPGQRLEW 48
1ZVO_C|PDBID|CHAIN|SEQUENCE RLQLQE-SGPGLVKPSETLSLTCIVSGGPIRRTGYYWGWIRQPPGKGLEW 49
1IGT_B|PDBID|CHAIN|SEQUENCE VAYISNGGGSTYYPDTVKGRFTISRDNAKNTLYLQMSRLKSEDTAMYYCA 97
1IGY_B|PDBID|CHAIN|SEQUENCE IGYINPSSVYTNYNQRFKDKATLTRDRSSNTANIHLSSLTSDDSAVYYCV 96
1HZH_H|PDBID|CHAIN|SEQUENCE MGWINPYNGNKEFSAKFQDRVTFTADTSANTAYMELRSLRSADTAVYYCA 97
1R70_B|PDBID|CHAIN|SEQUENCE MGWISAGTGNTKYSQKFRGRVTFTRDTSATTAYMGLSSLRPEDTAVYYCA 98
1ZVO_C|PDBID|CHAIN|SEQUENCE IGGVY-YTGSIYYNPSLRGRVTISVDTSRNQFSLNLRSMSAADTAMYYCA 98
1IGT_B|PDBID|CHAIN|SEQUENCE RHGGY----------YAMDYWGQGTTVTVSSAKTTAPSVYPLAPVC-GDT 136
1IGY_B|PDBID|CHAIN|SEQUENCE REG-------------EVPYWGQGTTVTVSSAKTTPPSVYPLAPGS-AAQ 132
1HZH_H|PDBID|CHAIN|SEQUENCE RVGPYSWDDS-PQDNYYMDVWGKGTTVIVSSASTKGPSVFPLAPSS-KST 145
1R70_B|PDBID|CHAIN|SEQUENCE RDPYGG-------GKSEFDYWGQGTLVTVSSASPTSPKVFPLSLDS-TPQ 140
1ZVO_C|PDBID|CHAIN|SEQUENCE RGNPPPYYDIGTGSDDGIDVWGQGTTVHVSSAPTKAPDVFPIISGCRHPK 148
1IGT_B|PDBID|CHAIN|SEQUENCE TGSSVTLGCLVKGYFP-EPVTLTWNSGSLSSGVHTFPAVLQS--DLYTLS 183
1IGY_B|PDBID|CHAIN|SEQUENCE TNSMVTLGCLVKGYFP-EPVTVTWNSGSLSSGVHTFPAVLQS--DLYTLS 179
1HZH_H|PDBID|CHAIN|SEQUENCE SGGTAALGCLVKDYFP-EPVTVSWNSGALTSGVHTFPAVLQSS-GLYSLS 193
1R70_B|PDBID|CHAIN|SEQUENCE DG-NVVVACLVQGFFPQEPLSVTWSESGQNVTARNFPPSQDASGDLYTTS 189
1ZVO_C|PDBID|CHAIN|SEQUENCE DNSPVVLACLITGYHP-TSVTVTWYMGTQSQPQRTFPEIQRRD-SYYMTS 196
1IGT_B|PDBID|CHAIN|SEQUENCE SSVTVTSSTWPS-QSITCNVAHPASSTKVDKKIEPRGPTI-KPCPPCKCP 231
1IGY_B|PDBID|CHAIN|SEQUENCE SSVTVPSSPRPS-ETVTCNVAHPASSTKVDKKIVPRD----CGCKPCICT 224
1HZH_H|PDBID|CHAIN|SEQUENCE SVVTVPSSSLGT-QTYICNVNHKPSNTKVDKKAEPKSCDKTHTCPP--CP 240
1R70_B|PDBID|CHAIN|SEQUENCE SQLTLPATQCPDGKSVTCHVKHYTNPSQDVTVPCPVP-------PPPPCC 232
1ZVO_C|PDBID|CHAIN|SEQUENCE SQLSTPLQQWRQ-GEYKCVVQHTASKSKKEIFRWPESPKAQASSVPTAQP 245
1IGT_B|PDBID|CHAIN|SEQUENCE APNLLGG-- 238
1IGY_B|PDBID|CHAIN|SEQUENCE VPEVS--- 229
1HZH_H|PDBID|CHAIN|SEQUENCE APELLGG-- 247
1R70_B|PDBID|CHAIN|SEQUENCE HPRLS--- 237
1ZVO_C|PDBID|CHAIN|SEQUENCE QAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEQEERETKTPECPSHTQ 295
1IGT_B|PDBID|CHAIN|SEQUENCE -PSVFIFPPKIKDVLMISLSPIVTCVVVDVSEDDPDVQISWFVNNVEVHT 287
1IGY_B|PDBID|CHAIN|SEQUENCE --SVFIFPPKPKDTLLITVTPKVTCVVVDISKDDPEVQFSWFVDNVEVHT 277
1HZH_H|PDBID|CHAIN|SEQUENCE -PSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHN 296
1R70_B|PDBID|CHAIN|SEQUENCE ------LHRPALED-LLLGSEANLTCTLTGLR-DASGATFTWTPS--SGKS 278
1ZVO_C|PDBID|CHAIN|SEQUENCE PLGVYLLTPAVQD-LWLRDKATFTCFVVGSDLKDAHLTWEVAGK-VPTGG 343
1IGT_B|PDBID|CHAIN|SEQUENCE AQTQTHREDYNSTLRVVSALPIQHQDWMSGKEFKCKVNNKDLPAPIERTI 337
1IGY_B|PDBID|CHAIN|SEQUENCE AQTQPREEQFNSTFRVVSALPIMHQDWLNGKEFKCRVNSAAFPAPIEKTI 327
1HZH_H|PDBID|CHAIN|SEQUENCE AKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTI 346
1R70_B|PDBID|CHAIN|SEQUENCE AVQGPPERDLCGCYSVSSVLPGCAQPWNHGETFTCTAAHPELKTPLTANI 328
1ZVO_C|PDBID|CHAIN|SEQUENCE VEEGLLERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPP--QRLM 391
1IGT_B|PDBID|CHAIN|SEQUENCE SKPKGSVRAPQVYVLPPPEEEMTK-KQVTLTCMVTDFMPEDIYVEWTNNG 386
1IGY_B|PDBID|CHAIN|SEQUENCE SKTKGKPRAPQVYTIPPPKEQMAK-DKVSLTCMITDFFPEDITVEWQSDG 376
1HZH_H|PDBID|CHAIN|SEQUENCE SKAKGQPREPQVYTLPPSRDELTK-NQVSLTCLVKGFYPSDIAVEWESNG 395
1R70_B|PDBID|CHAIN|SEQUENCE TKS-GNTFRPEVHLLPPPSEELALNELVTLTCLARGFSPKDVLVRWLQGS 377
1ZVO_C|PDBID|CHAIN|SEQUENCE ALREPAAQAPVKLSLNLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQ 441
1IGT_B|PDBID|CHAIN|SEQUENCE K--TELNYKNTEPVLD-SDGS--YFMYSKLRVEKKNWVERNSYSCSVVHE 431
1IGY_B|PDBID|CHAIN|SEQUENCE Q--APENYKNTQPIMD-TDGS--YFVYSKLNVQKSNWEAGNTFTCSVLHE 421
1HZH_H|PDBID|CHAIN|SEQUENCE Q--PENNYKTTPPVLD-SDGS--YFLYSKLTVDKSRWQQGNVFSCSVMHE 440
1R70_B|PDBID|CHAIN|SEQUENCE QELPREKYLTWASRQEPSQGTTTFAVTSILRVAAEDWKKGDTFSCMVGHE 427
1ZVO_C|PDBID|CHAIN|SEQUENCE REVNTSGFAPARPPPQ--PGSTTFWAWSVLRVPAPPSPQPATYTCVVSHE 489
1IGT_B|PDBID|CHAIN|SEQUENCE GLHNHHTTKSFSR---------------------- 444
1IGY_B|PDBID|CHAIN|SEQUENCE GLHNHHTEKSLSH---------------------- 434
1HZH_H|PDBID|CHAIN|SEQUENCE ALHNHYTQKSLSLSPGK------------------ 457
1R70_B|PDBID|CHAIN|SEQUENCE ALPLAFTQKTIDRLAGKPTHVNVSVVMAEVDGTCY 462
1ZVO_C|PDBID|CHAIN|SEQUENCE DSRTLLNASRSLEVSYVTDHGPM------------ 512
1IGY_B|PDBID|CHAIN|SEQUENCE SKTKGKPRAPQVYTIPPPKEQMAK-DKVSLTCMITDFFPEDITVEWQSDG 376
1HZH_H|PDBID|CHAIN|SEQUENCE SKAKGQPREPQVYTLPPSRDELTK-NQVSLTCLVKGFYPSDIAVEWESNG 395
1R70_B|PDBID|CHAIN|SEQUENCE TKS-GNTFRPEVHLLPPPSEELALNELVTLTCLARGFSPKDVLVRWLQGS 377
1ZVO_C|PDBID|CHAIN|SEQUENCE ALREPAAQAPVKLSLNLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQ 441
1IGT_B|PDBID|CHAIN|SEQUENCE K--TELNYKNTEPVLD-SDGS--YFMYSKLRVEKKNWVERNSYSCSVVHE 431
1IGY_B|PDBID|CHAIN|SEQUENCE Q--APENYKNTQPIMD-TDGS--YFVYSKLNVQKSNWEAGNTFTCSVLHE 421
1HZH_H|PDBID|CHAIN|SEQUENCE Q--PENNYKTTPPVLD-SDGS--YFLYSKLTVDKSRWQQGNVFSCSVMHE 440
1R70_B|PDBID|CHAIN|SEQUENCE QELPREKYLTWASRQEPSQGTTTFAVTSILRVAAEDWKKGDTFSCMVGHE 427
1ZVO_C|PDBID|CHAIN|SEQUENCE REVNTSGFAPARPPPQ--PGSTTFWAWSVLRVPAPPSPQPATYTCVVSHE 489
1IGT_B|PDBID|CHAIN|SEQUENCE GLHNHHTTKSFSR---------------------- 444
1IGY_B|PDBID|CHAIN|SEQUENCE GLHNHHTEKSLSH---------------------- 434
1HZH_H|PDBID|CHAIN|SEQUENCE ALHNHYTQKSLSLSPGK------------------ 457
1R70_B|PDBID|CHAIN|SEQUENCE ALPLAFTQKTIDRLAGKPTHVNVSVVMAEVDGTCY 462
1ZVO_C|PDBID|CHAIN|SEQUENCE DSRTLLNASRSLEVSYVTDHGPM------------ 512

Figure 3.18. Sequences alignment of the selected Ab crystal structures (1IGY, 1HZH, 1MCO, 1R70 and 1ZVO) to 1IGT. The position of amino acids of interest preserved within the six structures is highlighted in yellow (His, H), blue (Arg, R), pink (Lys, K), violet (Asp, D) and green (Glu, E).

3.1.7 Concluding remarks

We report for the first time all-atom explicit solvent Molecular Dynamics (MD) simulations investigating the structural and dynamical properties of a non-covalent bioconjugate in which the monoclonal Cetuximab antibody (Ctx) is adsorbed on a CNT surface. Upon selection of the three most representative adsorption modes as obtained by docking studies, force-field MD and DFT simulations unambiguously showed that hydrophobic interactions mainly govern the adsorption of the protein on the graphitic surface. Two main adsorption poses have been predicted: a pose-fab (*p*-fab) and pose-fc (*p*-fc), the former being favored with small-diameter tubes ($\leq$ 40 Å). In all the predicted poses, the secondary structure of Ctx is largely unaffected by the presence of the graphitic surface and, consistently with previous literature studies, our simulations reveal that positively charged amino acidic residues, such as Lys and Arg, predominantly contribute to the stabilization of the CNT•Ctx complex acting like surfactants. The predicted structural models are consistent with the experimental data, for which the immobilization of the antibody on CNTs does not disrupt the structural and recognition properties of the Ctx, consequently supporting the reliability of the used bioconjugation strategy for engineering stable and responsive hybrid nanomaterials for therapeutic applications. Moreover, a remarkable structural similarity of Ctx with antibodies of different isotypes suggests that in principle the CNT framework can interact in the same manner with all Abs currently used in clinical applications.

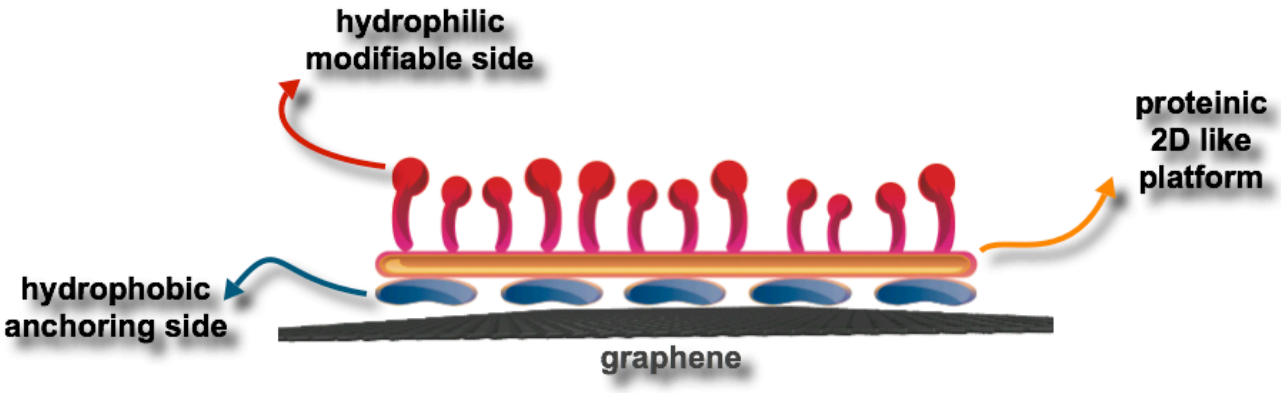
3.2 Design of Janus-type β-sheet for the hierarchical functionalization of graphene

3.2.1 Aim of the project

The project presented in section 3.1 was generally a prime example of the use of CNT-Ab biohybrid as an effective functionalization strategy for the CNT exploitation in nanomedicine. Through a combination of experimental and computational studies we probed the effectiveness of such functionalization approach to yield stable and biological active conjugates as the grafted protein conserved its structural and functional properties. Meanwhile, since its discovery in 2004[75] increased interest has been noted, for graphene as an emerging nanomaterial suitable for biomedical and biosensing purposes.[76–78] Graphene is particularly appealing because of its high available surface area that, thanks to its peculiar 2-dimensional (2D) geometry, bears two modifiable sides, thereby increasing the load of possible functionalization at a lower concentration, compared to CNTs.[79] Furthermore, along with its extremely high conductivity, impressive stress resistance and flexibility, graphene-based materials offer improved functionalities and previously unavailable options that found promising applications in the biomedical field.[76,80] However, graphene still maintains the typical drawbacks of GNM, such as the poor biocompatibility, solubility, and toxicity. Despite the efforts done for overcoming these limitations they still represent open issues, thus requiring further surface engineering investigations.[80]

At this respect, encouraged by the above compelling results, we found inspiration for investigating further the devised functionalization of GNMs towards biological applications by exploring the graphene's frontiers. So, taking advantage of the high available 2D surface, we envisaged the possibility of transferring also to graphene the Ab-targeting approach applied to CNTs, as a promising way of implementing our strategy. This would improve the control of the tailored biohybrid properties handling with different graphitic nanomaterial. However, passing from a 1D to a 2D GNM, its functionalization strategy would require a personalized and specific approach. For instance, the endohedral functionalization cannot be certainly applied to graphene, but we might want to exploit the high absorbance and

conductivity of the planar GNM for thermal and photothermal treatment.[81,82] This would thus imply a non-covalent functionalization in order to preserve its electronic and optical integrity. In fact, the application of the covalent Ab-coupling through amidation reaction, as used in the previous case, would destroy the aromatic conjugated pattern, while the direct non-covalent adsorption of a targeting unit, such as the mAb Ctx, would not assure the correct exposition and the tunability of the system. Hence, we envisage a non-covalent functionalization of graphene with a functional interface able to follow the extended 2D graphitic structure allowing for a high-density surface coverage. Then, aiming at controlling and mastering the functionalization of the flat GNM, a directional adsorption of such interface on the graphene surface will attain the univocal and selective adhesion to the surface with one side, meanwhile exposing the targeting function from the solvent exposed side. This will be achieved by designing a "Janus-like" structure that, due to its intrinsic bi-functional arrangement, will assemble in a directional manner exposing towards the nanomaterial a hydrophobic graphene anchoring side, and to the medium hydrophilic targeting functions, as shown in Scheme 3.2. Additionally, deciding for a bi-functional platform of proteinic nature, the functional interface between graphene and the targeting portion would also enhance the biocompatibility of the designed graphene-based targeting tool.



Scheme 3.2. Schematic representation of the graphene interfacing targeting tool: it will be made from a 2D like proteinic platform (orange) that will expose the hydrophobic anchoring side (blue) and hydrophilic groups (red) prone to the functionalization with a targeting unit.

Designed for our Ab-targeting strategy, the bi-functional interfacing platform presented herein results not only a specific graphene-functionalizing tool due to its 2D

shape equipped with a graphene-anchoring side, but also an extremely adaptable and versatile system that can be modified with any possible targeting agent.

Hence with a reverse approach as compared to the previous project, we present first a computational study to develop an *ex novo* designed new tool to non-covalently, directionally and bi-functionally modify the flat graphene surface preserving both its electronic and optical properties as well as the 2D structural arrangement. In a second avenue, the proposed system will be then corroborated by the experimental results.

3.2.2 Design of the interfacing targeting tool

As described above, we envisaged an interfacing targeting tool to non-covalently functionalize graphene with specific characteristics so to obtain a 2D-like, proteinic and bi-functional array. In particular, the two functionalities will consist of graphene anchoring moieties from one side of the 2D platform, and of modifiable targeting-oriented groups from the other. Hereafter, we describe separately each component used to address the required properties of the tailored system.

3.2.2.1 Proteinic, 2D-like, Janus-type platform: the β-sheet structure

Starting with the interfacing platform (Scheme 3.2), we need a 2D like and bi-functional structure in order to follow the peculiar flat geometry of graphene. Additionally, aimed at increasing the nanomaterial biocompatibility we though of a proteinic array as ideal biomimetic moiety. Hence, analyzing the possible proteinic structures, as highlighted in Chapter 1, we might consider the four possible levels of arrangement that polypeptides can adopt: primary, secondary, tertiary and quaternary structures. At the first level a polypeptide can have a primary structure, which consists of the simple aminoacidic sequence characterized by an amino-terminal (N-term) and a carboxy-terminal (C-term) group as show in Figure 3.19a. Notably, the rotation around the bond of the backbone atoms of each aminoacid (aa), namely the C_α-N and

the C_α-C bonds, defines two important torsion angles, called φ and ψ, crucial for the following folding process (Figure 3.19a).

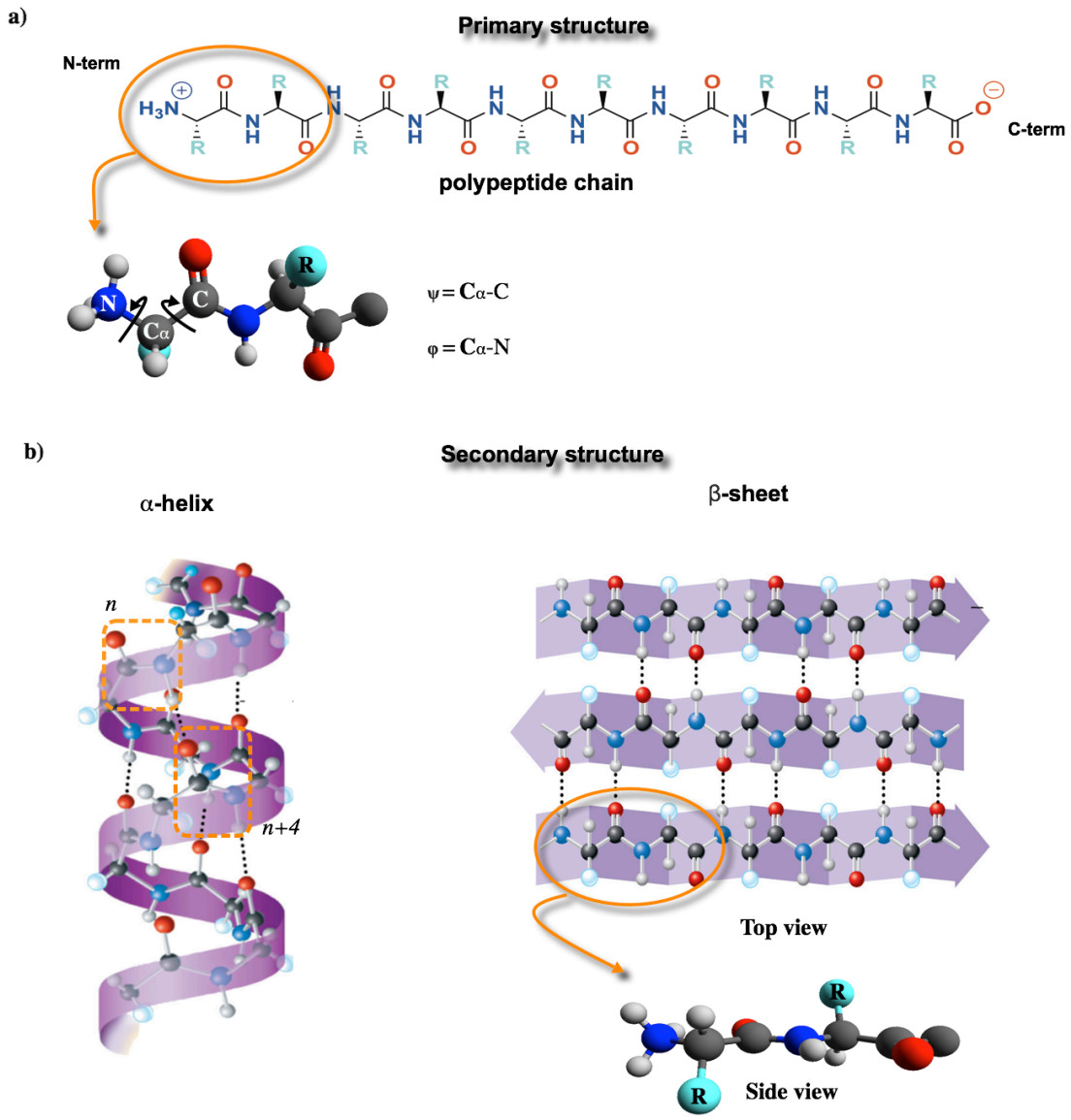


Figure 3.19. a) The primary structure of a protein consisting of the aminoacidic sequence, whose atoms are colored in red (O), blue (N), dark grey (C) and turquoise (R). The latter is the common way to generalize the different side chains of each aa. A 3D representation of a portion of the polypeptide sequence is reported highlighting the torsion angles φ and ψ around the backbone bonds Cα-N and the Cα-C. b) Two possible secondary structure arrangements, namely the α-helix and β-sheet, where the H-bonds are highlighted with dotted lines. The *n+4* residues involved in H-bonds within the α-helix are framed in orange dotted squares. A 3D representation of the side view of a portion of the β-sheet arrangement highlights the opposite displacement of the R groups as compared to the backbone plane.

In fact, adopted values of φ and ψ of -57° and -47°, respectively, correspond to a polypeptide folded in α-helix, while values of -139° and 135° lead to the folding in planar geometry know as the β-sheet secondary structure (Figure 3.19b).[83–85] The secondary structure is therefore defined by the degree of rotation of the backbone torsion angles, whose plane will define either the helical or the sheet arrangement. Of note, the displacement of the side chains of each residue, due to the intrinsic chirality of the aa, will be located externally to the α-helix framework (turquoise spheres in Figure 3.19b), and alternated at the two sides of the planar arrangement within the β-sheet structure (inset in Figure 3.19b). In particular based on this concept, by alternating hydrophilic/hydrophobic aa, β-sheet peptides with a tailored amphiphilic character have been engineered as reported elsewhere.[83,84,86] Importantly, these secondary structure arrangements are stabilized by non-covalent interactions occurring between both side chains (vdW interactions) and the backbone atoms (Hydrogen-bonds, H-bonds). Noteworthy, among these the H-bonds result particularly important and peculiar in both the secondary structure as a regular and repetitive H-bonding pattern is indeed induced by each peculiar arrangement. Specifically, within the α-helix, each backbone carbonyl group of residues in position *n* forms H-bond with the hydrogen of the NH group of residue *n+4* (Figure3.19b), whereas within the planar β-sheet the backbone carbonyl group of a residue in a polypeptide chain interacts trough H-bond with the NH group of an adjacent chain (Figure3.19b).[83] However, in the β-sheet case we can further distinguish two main H-bonding patterns, namely the parallel and antiparallel (Figure 3.20). The former is characterized by the parallel alignment of each polypeptide chain, called here strand, where all the "N-term to C-term" orientation maintains the same direction. This induces a distortion on the planarity of the inter-strands backbone H-bonds that occur between the residue *n* of one strand and the residue *n+1* of another strand (Figure 3.20a).

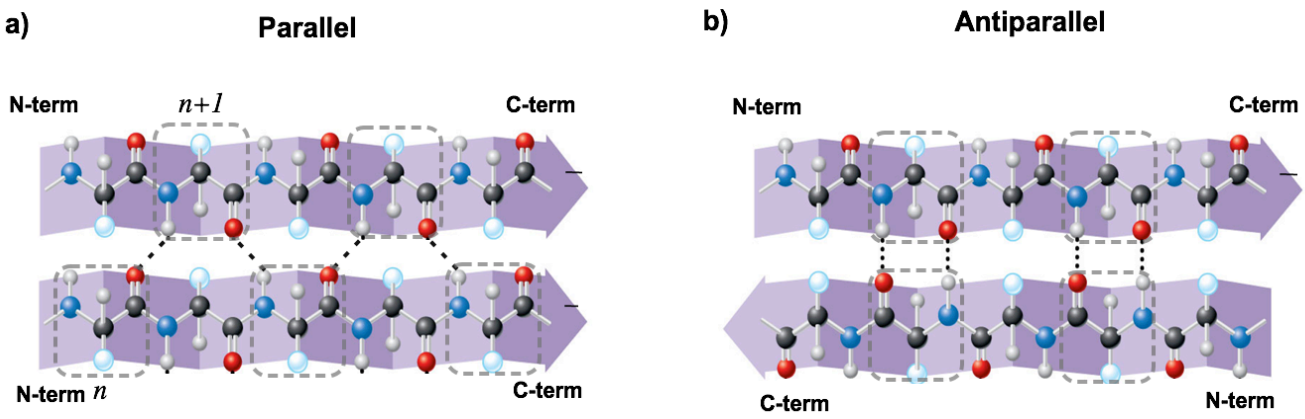


Figure 3.20. a) The two possible H-bonding patterns that a β-sheet can adopt according to the orientation of each strand: the same (parallel, a) or the opposite (antiparallel, b). In the first the aa *n* forms H-bond with the *n+1* aa of the adjacent strand (grey dotted squares), while in the antiparallel arrangements the interacting residues are one in front of each other.

The antiparallel one is instead characterized by the opposite alignments of the strands where the N-term faces the C-term group of the following strand, generating a more stable pattern due to the perfect planarity of the inter-strands backbone H-bonds (Figure 3.20b). Hence, it goes without saying that the planarity, the versatility and the bi-functionality that characterize the β-sheet structure perfectly fulfill the requirements to construct our 2D interface for the graphene functionalization.

An accurate design of the β-sheet will be then carried out to obtain the desired directional interface. The next step is therefore the engineering of the β-strand sequence that will then assemble in an extended β-sheet array. As already pointed out, the well-recognized approach of alternating hydrophobic/hydrophilic aa in a polypeptide chain can achieve the production of amphiphilic β-sheet,[83,84,86] while Brown *et al.* proved the surface-assisted assembly of a newly designed amphiphilic β-sheet on graphite.[87] Based on these evidences, we choose an alternating β-srand sequence of hydrophilic/hydrophobic aa that will automatically differentiate the polarity of the two sides of the resulting β-sheet. Giving priority to the preservation of the correct 2D arrangement, among the available hydrophobic and hydrophilic aa (see section 1.3, Figure 1.3) we select alanine and threonine (three and one letter abbreviations Ala, A and Thr,T) as the smallest and less sterically hindered residues, which might not perturb the β-sheet assembly. A structural representation of the preliminary portion of designed β-strand is reported in Figure 3.21. Before defining

the length of the $(TA)_n$ strand, we first analyzed the other components of the bi-functional proteinic interface.

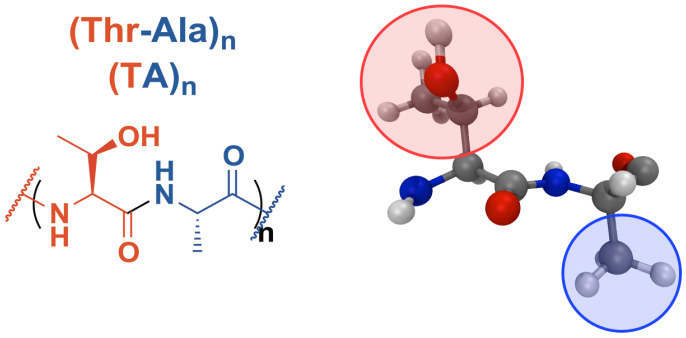


Figure 3.21. The chemical structure and the 3D representation of the preliminary $(TA)_n$ sequence of the designed β-strand are reported colored in red the hydrophilic portions and in blue the hydrophobic ones. The three and one letter abbreviations are also reported. The 3D sketch on the right highlights the alternated confinement of the hydrophilic side chain (red circle) and the hydrophobic side chain (blue circle).

3.2.2.2 Graphene anchoring point

The $(TA)_n$ strand assembling in β-sheet is expected to expose as hydrophobic side the Ala side chains that will direct the adsorption onto graphene maximizing their surface contact thanks to the 2D geometry fitting, while the Thr as hydrophilic side will be solvent-exposed enhancing the biohybrid solubility.

Moreover, inspired by the works by Dichtel and co-workers,[88–90] who studied and optimized the ability of aromatic tripodal compounds to adsorb on graphene, we decided to insert a pyrenyl lateral modification within the β-strand sequence that can act as a graphitic anchoring moiety, enhancing in this way the β-sheet adsorption performances trough π-π stacking. Aiming at this, we envisaged the insertion of the pyrenyl group through the chemical modification of a glutamate (Glu, E) aa with the commercially available pyren-1-ylmethanamine, whose structural representation as a preliminary portion of designed β-strand is reported in Figure 3.22. The modified Glu residue (Glu-pyr) will be fundamental for the pyrene insertion because, thanks to its aminoacidic structure, will preserve the correct backbone arrangement of the β-sheet maintaining its proteinic nature, but at the same time will allow the exposition of the

jutting anchoring portion towards the hydrophobic side. Glu-pyr will be inserted in the β-strand sequence as a hydrophobic residue and placed at the two extremities, N-term and C-term, of each β-strand sequence, thereby promoting the lateral anchoring of the β-sheet structure onto the graphene surface.

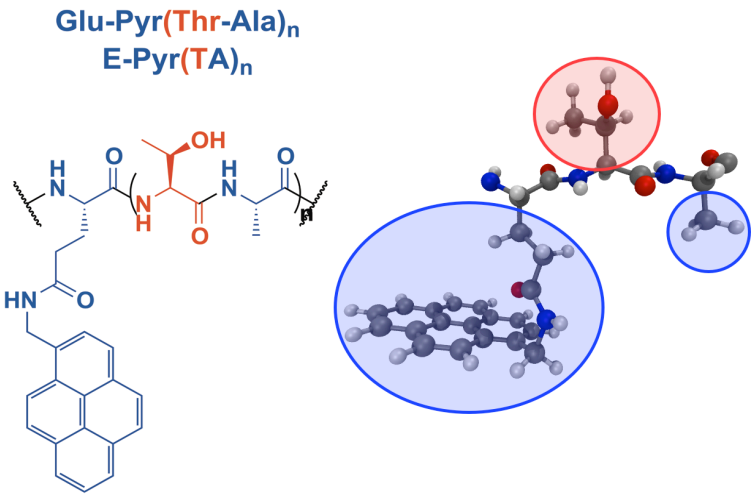


Figure 3.22. The chemical structure and the 3D representation of the preliminary E-Pyr(TA)$_n$ sequence of the designed β-strand where the hydrophilic portions are colored in red and the hydrophobic ones in blue. The three and one letter abbreviations are also reported. The 3D sketch on the right highlights the alternated confinement of the hydrophilic side chain (red circle) and the hydrophobic side chain (blue circle), which now includes the pyrenyl modified Glu.

3.2.2.3 Targeting oriented functional groups

As far as it concerns the hydrophilic side, this will be responsible of the functionalization of the entire biohybrid tool with a targeting unit. The idea is to add a residue that might be easily and efficiently coupled with the targeting protein, for example with the mAb Ctx, but that can be also versatile and applicable to other possible targeting unit. Following the successful amidation reaction strategy proposed in section 3.1, we therefore identified in the lysine aa (Lys, K) the ideal candidate for this purpose. In fact, Lys is classified as a positively charged aa (Figure 1.3) due to the nature of its pentylamino chain. The insertion of such residue will therefore conserve the polarity of the hydrophobic side, exposing the amino groups prone to the amidation reaction with the carboxylic groups of any possible targeting protein

(Figure 3.23). Furthermore, the aa will once again preserve the proteinic character of the designed interfacing platform without affecting the β-sheet backbone disposition.

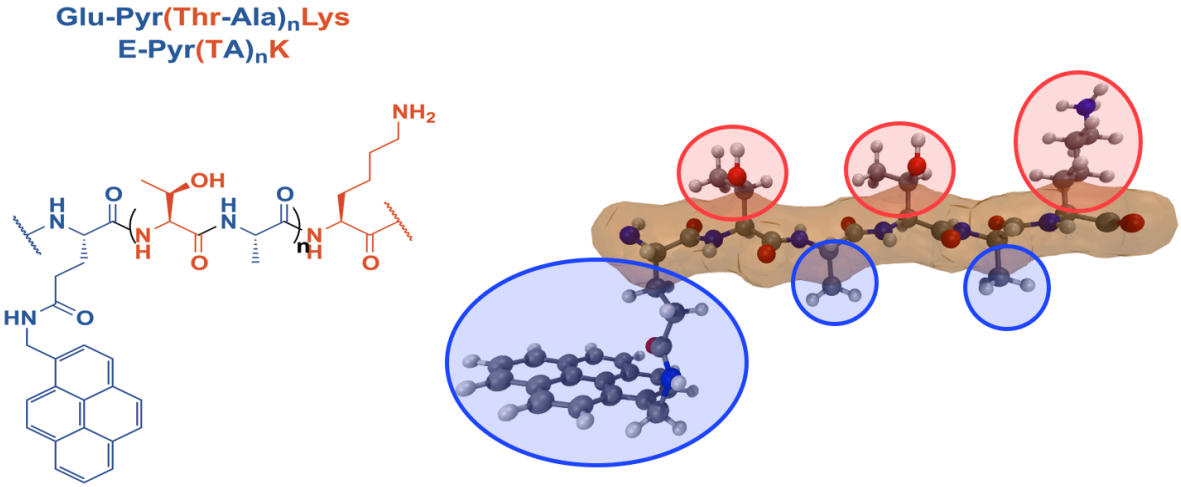


Figure 3.23. The chemical structure and the 3D representation of the preliminary E-Pyr(TA)$_n$K sequence of the designed β-strand where the hydrophilic portions are colored in red and the hydrophobic ones in blue. The three and one letter abbreviations are also reported. The 3D sketch on the right highlights the alternated confinement of the hydrophilic side chain (red circle) and the hydrophobic side chain (blue circle), included the pyrenyl modified Glu. The planarity of the β-strand backbone is highlighted in light orange.

3.2.2.4 Assembling process

Having identified all the functional building blocks of the β-strand unit, it is now of crucial importance to define the modality according to which the assembling process will be achieved. It is worth recalling that with the herein designed β-sheet system we want to achieve the extended high-density coverage of the graphitic surface in order to maximize the ratio between the available graphene surface and its functionalization rate. Additionally, the accurate design will have to assure the univocal, directional and hierarchical assembly of the β-strands in an extended β-sheet in the easiest and automated way. In other words, our system should be able to autonomously form a unique polymeric network forming stable covalent bonds between the strands able at the same time through a self-sorting, self-recognition process to yield the most stable product. This strategy bears the name of dynamic covalent chemistry (DCC). DCC consists of reversible reactions that under

thermodynamic control are able to select among all the possible species the one with the relatively higher stability in a self-repairing approach.[91] Exemplar case of DCC illustration is shown by the natural process of disulfide exchange during protein folding, which directs the biomolecule structure towards its active conformation.[92]

Hence, inspired by the natural efficiency and by the emerging interest in applying DCC to engineer functional systems,[93] we plan the insertion of cysteine (Cys, C) aa at the two β-strand extremities in order to promote the assembling process among strands through disulfide bonds. In fact, this sulfur-containing aa, renewed to bridge through disulfide bonds the proteinic structure, will allow the linkage between the strands conserving at the same time the proteinic nature of the backbone of the β-sheet interface. However, in order to obtain the univocal and uniform assembly, fundamental prerequisite lies in the symmetry of the designed β-strand. Consequently, the desired symmetric strand will bear the targeting-oriented Lys placed centrally, while the two sides will grow symmetrically with *n* repetitions of the TA sequence so to obtain the hydrophilic/hydrophobic alternation fundamental for the β-sheet formation. Afterwards, the modified Glu-pyr anchoring groups will be located laterally and finally the Cys aa at the two extremities, as terminal assembling points. The resulted β-strand is shown in Figure 3.24. By this way, the correct hydrophobic/hydrophilic alternation will be satisfied, resulting in a symmetric β-strand able to undergo univocal disulfide directed assembly.

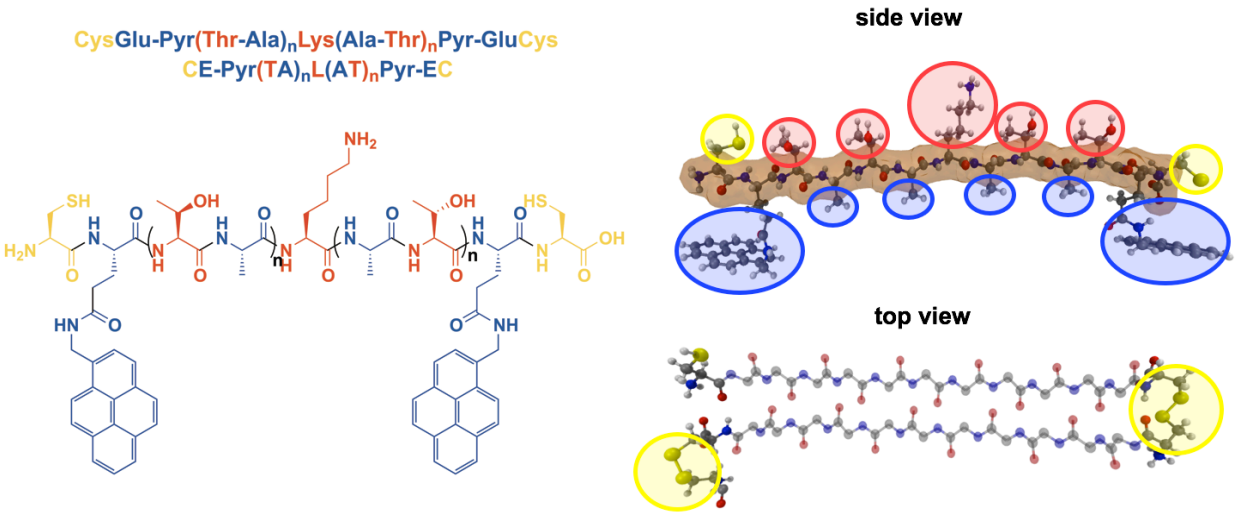


Figure 3.24. The chemical structure of the final β-strand sequence is reported highlighting in yellow the lateral Cys, in blue the hydrophobic aa (the Glu-pyr and the Ala) and in red the hydrophilic ones (the Thr and the Lys). The three and one letter abbreviations are also reported. The side view of the 3D sketch on the right highlights the alternated confinement of the hydrophilic side chains (red circle), the hydrophobic side chains (blue circle) and the terminal assembling point (yellow circles), i.e. the Cys. The top view of the 3D sketch shows the disulfide bonds between two β-strands, of which for clarity only the backbone atoms are displayed.

3.2.3 Computational challenge

The accurate design of the β-strand sequence as described in the previous sections should in principle assure the formation of the desired β-sheet interface. However, important structural aspects have still to be assessed for the finalization of the β-strand device. In fact, the sequence length, the type of H-bonding pattern and the ideal number of pyrenyl anchoring functionalities represent crucial aspects that might affect the stability and functionality of the final proteinic interface. At this respect, theoretical calculations and especially classical MD simulations would be applied to model and predict the structural stability of different β-sheets as a function of their modifications. We will thereby propose the ideal β-strand candidate, suitable for the functionalization of the graphitic interface in the shortest and most efficient way.

First of all, taking into account that the stability of our designed β-sheet partly relies on the alternation of the hydrophilic/hydrophobic nature of the residues, it will be of paramount importance to define the β-strand sequence length in particular of the $(TA)_n$ portion. In fact it has been reported in the literature that the β-sheet stability increases with the strand length both in parallel[94] and antiparallel[95,96] arrangements. So, taking fixed the number and position of the lateral Cys, the modified Glu-pyr and the central Lys, we modulated the number of residues in the alternating hydrophilic/hydrophobic pair $(TA)_n$ and we first hypothesized two possible building blocks (bb). Specifically, bb1 bearing the shortest (TA) repetition (n = 1) counts 9 aa per strand, while bb2 with n = 2 counts 13 aa (Figure 3.25).

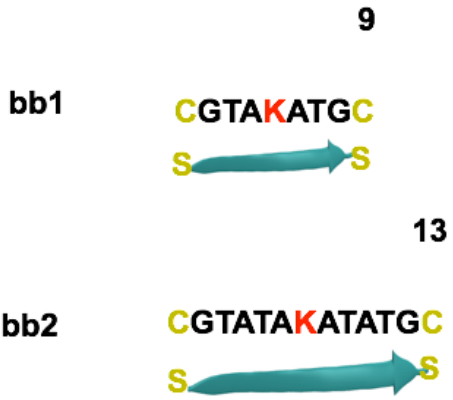


Figure 3.25. By modulating the repetition of the $(TA)_n$ pair, two possible aminoacidic sequences have been identified: building block 1 (bb1, n=1) and 2 (bb2 n=2) counting 9 and 13 aa, respectively. The 3D representation of β-strand is also reported as a cyan arrow, whose direction indicates the chain orientation from the N-term to the C-term group.

Secondly, the backbone intermolecular H-bonding pattern represents the constitutive structural parameter that mainly determines the stability of the final assembled β-sheet.[97,98] Thanks to the symmetric β-strand sequence, the same, univocal and repetitive functional group displacement will be expected in the β-sheet formation, so to obtain the desired bi-functional interface. However, if one considers all the possible orientations that the simplest case, *i.e.* two β-strands facing each other, can adopt (parallel and antiparallel orientations, same or opposite hydrophilic/hydrophobic confinement) a situation as depicted in Figure 3.26 will be encountered. Taking the Lys residue as reference of the hydrophilic side, when the two β-strands approach each other pointing the Lys towards the same side, will be able to form correctly the renowned parallel or antiparallel orientation as clearly

demonstrated by the shown H-bonds patterns in Figure 3.26a, top and bottom. On the other hand, when the Lys residues are pointed towards the opposite sides, the β-strands are placed in a specular fashion meaning that each NH/CO group of one strand faces the NH/CO group of other one. This hypothetical arrangement induces the occurrence of NH-NH and CO-CO interstrands repulsive interactions either in parallel and antiparallel (Figure 3.26b top and bottom). By this way, any H-bonds will be formed but the repulsive interactions implying energetically prohibitive β-strands aggregation as reported in the literature.[99,100]

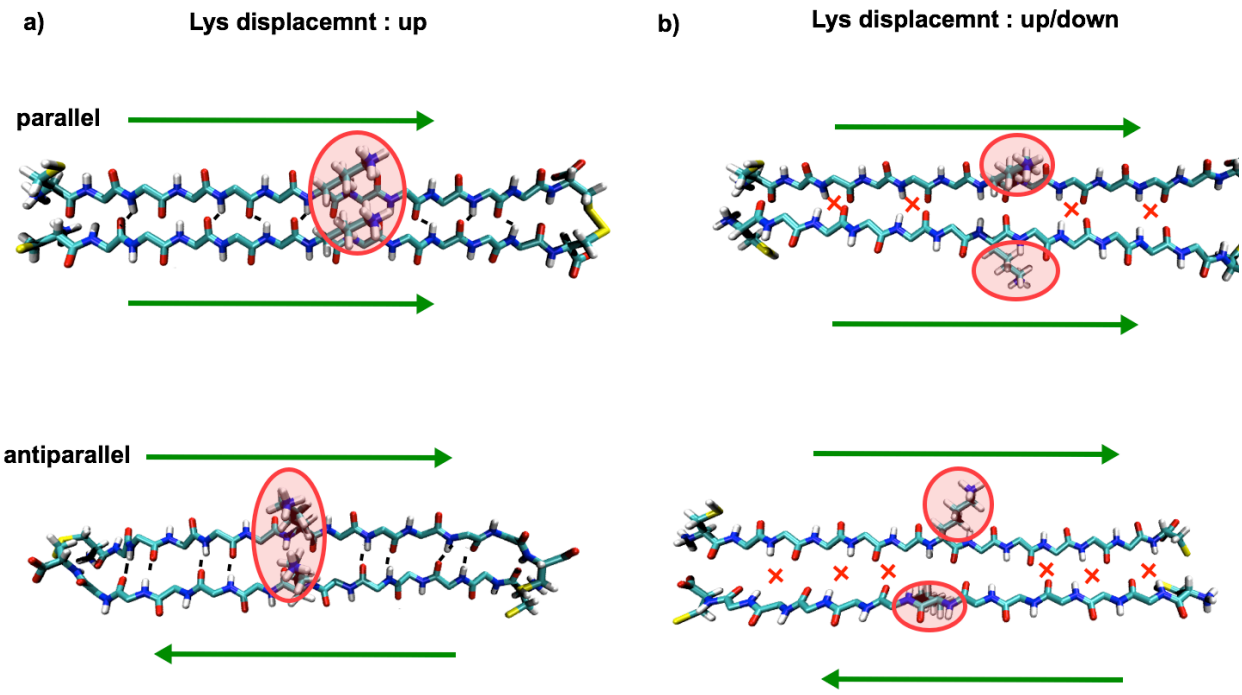


Figure 3.26. β-hairpin model of two strands approaching with a parallel (top) or antiparallel (bottom) arrangement either pointing the hydrophilic Lys aa (framed in red circles) towards the same side (a) or to opposite sides (b). Attractive interactions such as H-bonds are labeled with a black line between the atoms involved, repulsive interactions are instead labeled with red crosses.

Consequently, by excluding the arrangements shown in Figure 3.26b, both parallel and antiparallel patterns in Figure 3.26a guarantee the correct bi-functional orientation. Hence, it still remains to verify which of the two H-bonding patterns attain the most structurally stable β-sheet. To this aim, MD simulations will be performed evaluating the effect of the parallel/antiparallel on the structural integrity of the designed proteinic β-sheet and consequently on its functional attainment.

Furthermore, the presence of the pyrenyl functionalities, fundamental for the graphene anchoring, represents the only non-proteinic moiety within the designed β-sheet, thereby resulting a possible weak point for the stability of the whole system. Consequently, the presence as well as the number of the pyrenyl functionalities will be investigated considering two different systems with two or one pyrenyl functionalities per strand as depicted in Figure 3.27. To be thorough, the β-strand without any pyrenyl insertion will be studied as a reference system.

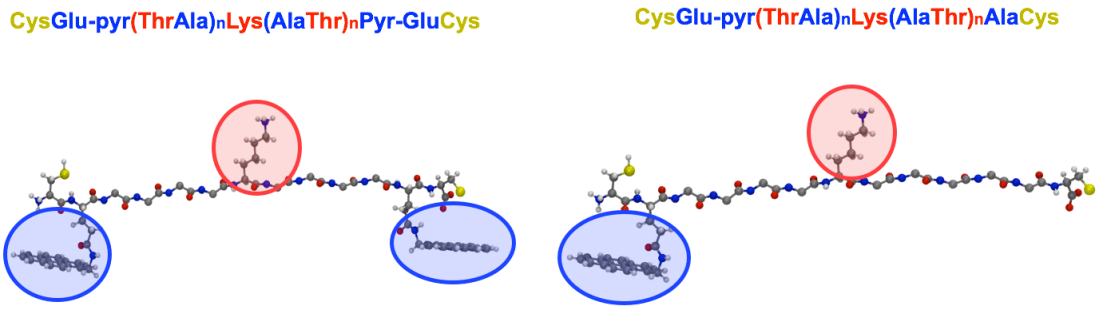


Figure 3.27. The 3D representation of two possible β-strands with two (left) and one (right) pyrenyl functionality inserted within the sequence. For clarity only the backbone atoms, the Lys aa as hydrophilic residue (red circles) and the Glu-pyr (blue circles) residues are show.

Finally, the H-bonding patterns, the strand length and the number of pyr groups will be accurately investigated as a measure of the structural stability of the designed β-sheet. Once identified the structural features that might provide the most stable β-sheet candidate, the latter will be also simulated adsorbed on the graphene surface, ultimately assessing the effect of the graphitic surface on the structural stability and the functionality of the whole biohybrid system. In summary, a complete and thorough computational protocol will be therefore carried out to accurately tackle the design of a graphene functionalization interface, whose efficiency will ultimately tested experimentally.

3.2.4 Methods sections

3.2.4.1 Model systems

Proteinic β-sheet. In order to simulate the proteinic β-sheet, the designed single strand polypeptide sequence was engineered with the peculiar alternation of hydrophobic/hydrophilic aa (ThrAla), the insertion of a central targeting point (Lys), of lateral anchoring groups (Glu-pyr) and polymerization attacks (Cys). The resulting sequence CE-pyr $(TA)_nK(AT)_n$E-pyrC was therefore characterized by an amphiphilic nature and a symmetric structure to ensure the univocal assembling. The strand was built with two different lengths, namely with n = 1 and 2 for a total of 9 and 13 aminoacids (aa), labeled as building block 1 and 2 respectively (bb1 and bb2, Figure 3.25). Taking into account the cooperative inter-strand interactions occurring in the β-sheet folding process,[101,102] and the expected extensive multistrand β-sheet formation by means of disulphide polymerization,[103] we decided to simulate six-stranded β-sheets as the minimum β-sheet extension to adequately represent the expected experimental conditions with reasonable computational efforts. 11 systems have been chosen as representative and significant to thoroughly sieve the β-sheet elements.

For sake of clarity the nomenclature used herein to label the systems understudy is explained as follow:

- **6** is the number of strand per β-sheets;
- **bb** stands for the building block 1 or 2 that differ in the sequence length (9 or 13 aa respectively);
- **para** is the parallel H-bonding pattern; when not specified the antiparallel pattern is used;
- **pyr** indicates the insertion of the pyrenyl functionalities on each lateral Glu, whereas **1pyr** indicates the functionalization of only one Glu per strand;
- **ala** indicates that the lateral Glu have been substituted with Ala aa;
- **g** indicates that the β-sheets have been simulated adsorbed parallel to graphene. The pyrenyl moieties of the Pyr-modified β-sheets were arbitrary simulated parallel to the graphene surface;
- **cnt** indicates that the β-sheets have been simulated onto CNT;

- **II or ⊥** denotes the parallel or perpendicular orientation of the β-sheets adsorbed on CNT respectively.

The parm99SB[37] force field was employed to describe the β-sheet proteins. The pyrenyl moiety and the modified Glu to which is attached, were also parameterized according to the AMBER procedure. TIP3PBOX[39] model was used to simulate explicitly water molecules. Table 3.8 summarizes the details of the simulated systems carried out in this study.

Table 3.8. Summary of all the simulations performed.

System	Time (ns)	# of atoms
6bb1	15	16046
6bb2	50	22919
6bb2-para	15	17477
6bb2-ala	50	30546
6bb1-pyr	15	20096
6bb2-pyr	50	28040
6bb2-1pyr	50	28364
g-6bb2	50	24654
g-6bb2-ala	50	30546
g-6bb2-pyr	50	37224
g-6bb2-1pyr	50	33399

Graphene. A small graphene unit was parameterized according to the general AMBER procedure: a single point energy calculation was performed at B3LYP/6-31G level of theory and the resp charges were derived using the antechamber module of Amber12[104] and the force field parameters were therefore assigned to the graphene building block according to the general Amber force field (gaff).[45] Therefore an extended surface of graphene was built sticking together multiple units, up to the chosen dimension of 92*84 Å, four times bigger than the proteinic β-sheet.

3.2.4.2 Force field based MD simulation

Classical MD simulations were performed on all the systems described in the above section. All the systems were solvated with explicit water considering a minimum distance of 10 Å between the solute and the edges of the cubic box. The Lys residues were simulated positively charged, therefore the whole β-sheet was neutralized by adding 6 Cl^- ions. When the graphene surface was simulated, it was located at the bottom of the box. The initial structures were firstly relaxed by short minimization runs of 10000 steps, using the conjugate gradient energy minimization algorithm (convergence criterion $1.0E^{-4}$ kcal mol^{-1} Å). The MD simulations were performed at constant pressure (1 atm) and temperature (298 K) using Langevin thermostat with a 2 fs time step. Periodic boundary conditions were applied. Non-bonded cut-off of 8 Å was set and Particle Mesh Ewald (PME)[51] summation was used to calculate the long-range electrostatic interactions. The bonds involving hydrogen atoms were kept fixed using the SHAKE algorithm. No other restraints were applied. The calculations were performed with the AMBER12 program.[104]

3.2.4.3 Calculated properties

For all the simulated systems the Root-Mean Square Deviation (RMSD) on the Cα, and on the C, N backbone atoms of the β-sheet structure, Radius of Gyration (Rg), the number of contacts between the side chains of each system and the distance between the center of mass of the pyr functionalities when present were calculated along the entire trajectories. The *ptraj* module of AMBER12 was used for all the above properties. The number of donor moieties involved in H-bonds was calculated along the 50 ns simulations with the *Cpptraj* module of AMBER12.[104] The H-bonding content (Hbc) was calculated considering the ratio between the number of donors actually involved in H-bonds and the total number of available H-bonding donors; the <% of occupancy> (<%occ>) was calculated by averaging the % of occupancy of the donors involved in H-bonds along the simulated time; the % of donors was also plotted as a function of the % of occupancy to assess in details the singular donor contribution to the total Hbc. The secondary structure (SS) was also

analyzed by using the DSSP method[52] and the number of residues found in the β-sheet was plotted as a function of the simulated time. The total number of backbone residues available for H-bonds is 42 within **6bb1** (54 tot aa subtracted of the 12 lateral Cys involved in turns secondary structure elements), whereas 66 for **6bb2** (78 tot aa subtracted of the 12 lateral Cys). Hence, the SS content, namely the ratio between the total number of residues actually involved in the β-sheet organization and the total number of backbone residues available for H-bonds gives a further idea of the retention of SS structural information. As explained in section 3.2.2.1, the secondary structure arrangement of a polypeptide depends on the backbone rotation around the torsion angles φ and ψ. The easiest way to visualize these backbone dihedral angles, as a measure of the secondary structure of a protein, is the ramachandran plot.[105] The φ and ψ dihedral angles represent the axes of the ramachandran plot and range from -180° to 180°. The ideal φ and ψ values that the different secondary structure elements may adopt, define the favored regions (sky-blue), while variation of those angles up to certain values defines the allowed regions (pale blue). In particular, φ and ψ values of about -57° and -47°, respectively, correspond to the α-helix structure and are located in the bottom left square within the sky blue region as shown in Figure 3.28a. On the other hand, values of 135° and -139° refer to the β-sheet secondary structure, which is located in the top left square within the sky blue region (Figure 3.28b).[83]

Due to the circularity of the dihedral angles (0° is the same as 360°), the bottom left sky blue area is a continuation of the bigger top left area, and the top right a continuation of the top left. The ramachandran plots were calculated by using the RAMPAGE online server.[106] The interaction energies, namely the electrostatic (EEL) and the van der Waals (vdW) energies of the β-sheet as well as between the graphene surface and the β-sheets, were calculated along the simulations with a cut-off, of non-bonded interaction energies of 8 Å employing the NAMD energy plugin (v 1.4) as implemented in VMD (v 1.9.1) according to the Eq. 2.16.[42] To estimate the binding free energy among the graphene surface and the β-sheets, a molecular mechanics/Generalized–Born Surface Area (MM-GBSA)[107] analysis of the trajectories was also performed.

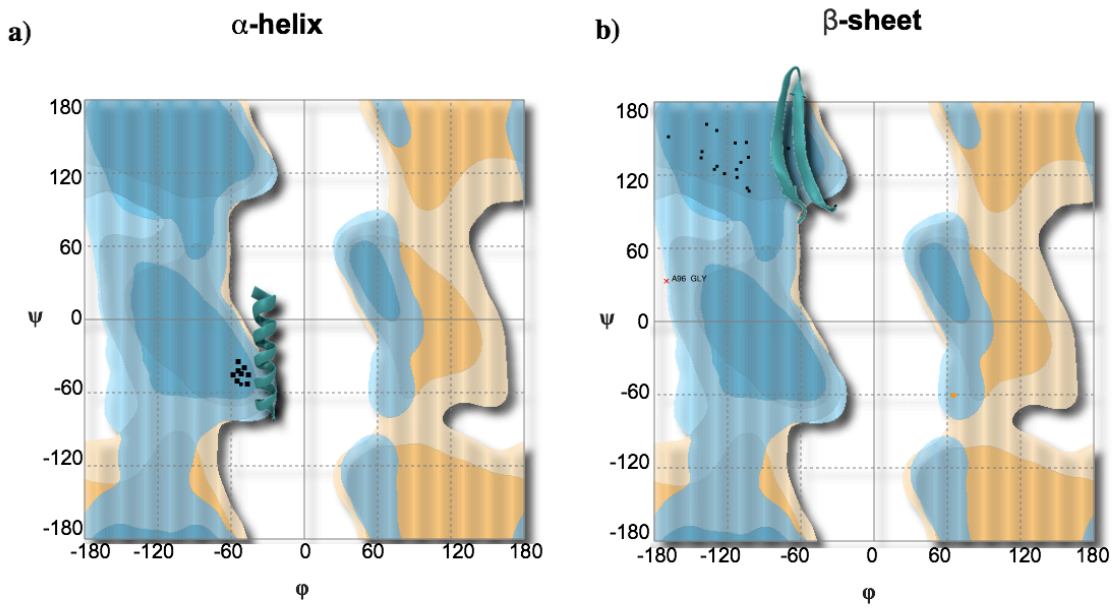


Figure 3.28. Examples of two secondary structure elements visualized by the ramachandran plot: the α-helix (a) and β-sheet (b), for which the backbone dihedral angles of each residues are marked with black dots. The φ and ψ axes of the ramachandran plot ranges from -180° to 180°, defining the general favored and allowed regions in sky blue and pale blue, respectively. In orange and pale orange are highlighted also the favored and allowed regions of glycine that has less restricted rotation due to its small side chain.

3.2.5 Results and discussions

A *de-novo* bi-functional proteinic β-sheet was designed to selectively and directionally functionalize graphene surface. To do so, the β-sheet was engineered to expose one hydrophobic face directed to the adsorption onto graphene, while the other fully hydrophilic face is solvent exposed and available for final targeting purposes. The adsorption tendency of the hydrophobic face is enhanced by the insertion of pyrenyl moieties at the strand extremities. However, several aspects that can undermine the β-sheet stability have to be considered such as the β-strand length as well as the presence and the number of the pyrenyl functionalities. In fact, the proper modulation of these aspects can ultimately lead to the identification of the ideal targeting graphene-binding interface. The β-sheet was selected and validated by a thorough computational protocol performing classical MD simulations of different systems as reported in Table 3.8.

This section is therefore organized in five main parts: *"β-strand length and H-bonding pattern"*, which will address the choice of the length of the single strand and verify the stability of the parallel or antiparallel H-bonding patterns; *"presence and number of pyrenyl functionalities"*, which will give an account of the ideal number and location of the pyrenyl anchoring points; *"adsorption on graphene",* will evaluate the stability of the selected β-sheet once adsorbed onto graphene; *"directional adsorption on graphene",* will evaluate the effectiveness of the selected β-sheet to be directionally adsorbed onto graphene. Finally, in *"2D vs 1D: comparison with CNTs",* the graphene adsorption ability of the selected β-sheet will be compared with the absorption on CNT.

3.2.5.1 β-strand length and H-bonding pattern

The identification of the ideal β-strand length was tackled by simulating for 15 ns timescale two different building blocks with the sequence $CE(TA)_nK(AT)_nEC$, namely **bb1** and **bb2** (n = 1 and 2, respectively) arranged in a 6-stranded β-sheet (**6bb1** and **6bb2**) first in antiparallel fashion. In order to evaluate the intrinsic stability of the so-designed β-sheet, the pyr anchoring groups were not considered for the moment.

By comparing the starting conformations and the last snapshot after the short simulated time, it emerges the evident tendency of **6bb1** to undergo misfolding of the β-sheet structure, while the conformation in the **6bb2** is retained after the considered timescale (Figure 3.29).

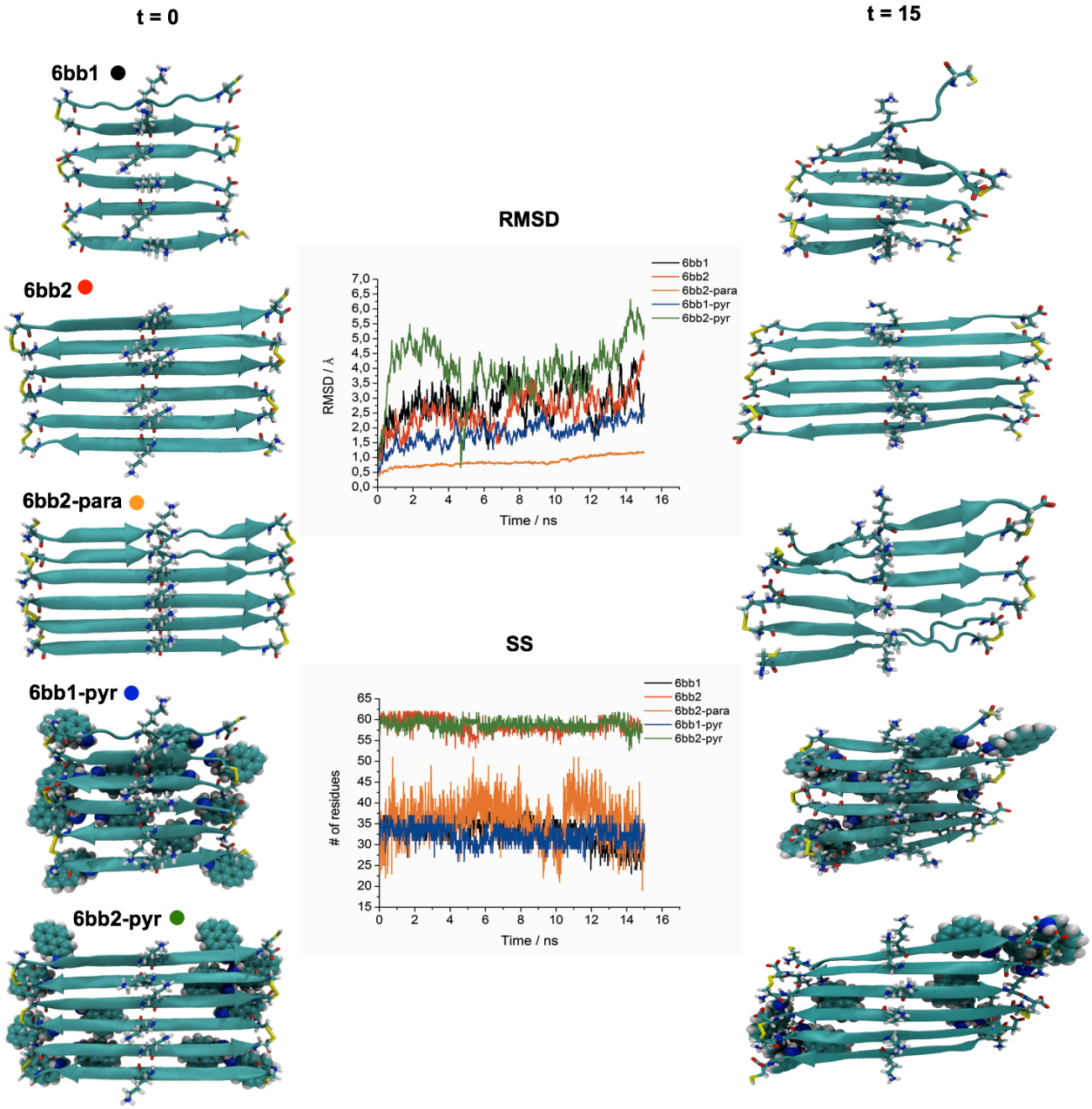


Figure 3.29. The starting structure (left panel), the RMSD (top) and the SS (bottom, middle panel), and the last snapshot (right panel) of the 15 ns MD simulation are reported for **6bb1** (black), **6bb2** (red), **6bb2-para** (orange)**, 6bb1-pyr** (blue) and **6bb2-pyr** (green). The 3D representations of the β-strand are drawn as cartoon arrows whose direction evidences the antiparallel or parallel secondary structure; for clarity all the side chains are not shown but the hydrophilic face is marked by the Lys residues drawn in licorice, while the assembling points among the strands are evidenced by lateral disulphide bonds between Cys drawn in licorice; when present, the pyrenyl functionalization is drawn in vdW.

The calculated RMSD as reported in Figure 3.29 (black and red lines) reveals that the systems are fairly stable as compared to the starting structure. As a confirmation of the visually observed misfolding, calculating the secondary structure (SS) content along the time, its actual reduction is indeed detected in the **6bb1** system (33 out of 42, 33/42 ± 3 residues in β-sheet), while a stable SS content is observed for **6bb2** (59/66 ± 2 in β-sheet, Figure 3.29 bottom, black and red lines).

Moreover, by plotting the backbone dihedral angles, namely φ and ψ, in the ramachandran plot, significant differences were highlighted in the β-sheet region. The φ and ψ dihedral angles axes of the ramachandran plot, ranging between -180° and 180°, define different regions within the plot that correspond to secondary structure elements. In particular, the perfect antiparallel organization of the β-sheet is characterized by backbone dihedral angles φ and ψ of about -139° and 135°,[83] respectively, therefore located in the top left square of the ramachandran plot. Dihedral angles close to these values occupy the so-called "favored" regions (black dots), whereas values deviated from the reference values occupy "allowed" (orange dots) up to "outliner" (red dots) regions. As shown in Figure 3.30, the **6bb1** system unveils the 100% of residues in the favored region (sky blue region marked with black dots), albeit their displacement is spread over the β-sheet area. This means that the backbone dihedral angles of **6bb1** adopt deviated values as compared to the expected values (-139° and 135°), thereby leading to the distortion of its β-sheet structure as visually observed (Figure 3.29).

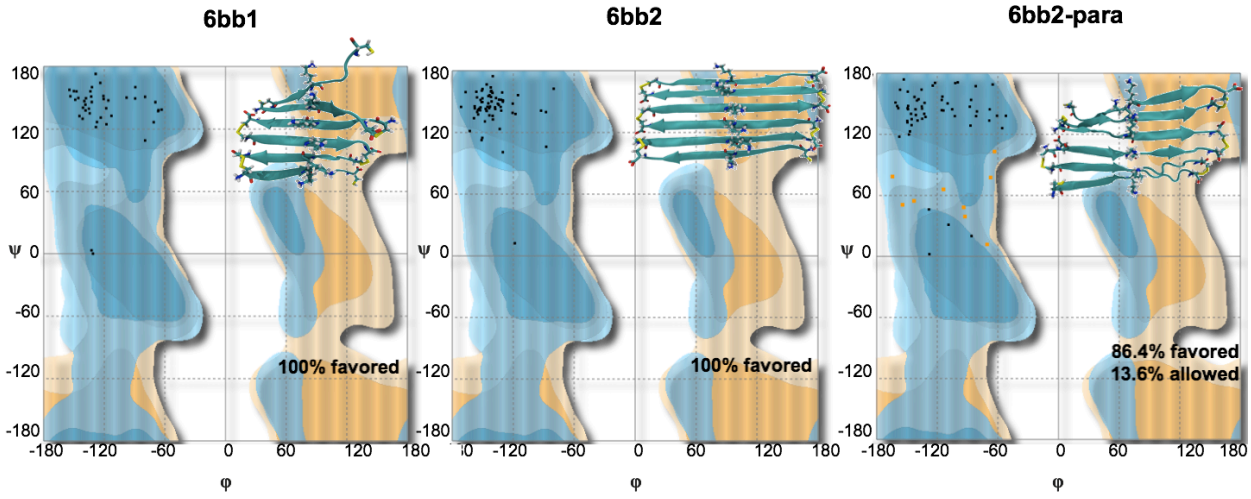


Figure 3.30. The ramachandran plots are reported with the % of residues in favored (black dots) and allowed region (orange dots) for **6bb1**, **6bb2** and **6bb2-para** from the left to the right.

Conversely, even though the same 100% of favored occupied region is observed for **6bb2**, the backbone dihedral angles are now matching the β-sheet values so that their better disposition within the top left region represent the sign of the proper secondary structure conformation of the longer β-strand **bb2** (Figure 3.30). Additionally, in order to qualitatively quantify the stability of the simulated systems, the non-bonded interaction energies, namely the van der Waals (vdW) and the electrostatic (EEL), were calculated and the average values are reported in Table 3.9. It is worth recalling that the stability of a β-sheet derives from both the side chains and H-bonds interactions, the latter fundamental for the compactness of the system. In fact, the EEL energy, accounting for H-bonding interactions, is found to contribute the most as compared to the vdW (Table 3.9) most likely accounting for the side chains contacts. The energetic analysis clearly shows a major stability of **6bb2** (<EEL> = -925 kcal mol^{-1}) as compared to **6bb1** (<EEL> = -442 kcal mol^{-1}), that we can reasonably ascribe to the better stabilization of its H-bonding pattern and the β-sheet organization as suggested by the above analysis of the secondary structure (Figures 3.29-30). Although based on short simulation time, the structural evidences and the energetic analysis suffice to exclude **6bb1** as plausible β-sheet arrangement, while the most stable **6bb2** pattern confirms the length-dependent effect on the β-sheet SS, as also reported in the literature.[95]

Table 3.9. Summary of the average electrostatic (EEL) and van der Waals (vdW) energies and standard deviations calculated for the systems understudy along the 15 ns trajectories.

System	<EEL>[a]	<vdW>[a]
6bb1	-442 ± 91	-117 ± 11
6bb2	-925 ± 74	-155 ± 12
6bb2-para	-43 ± 143	-151 ± 14
6bb1-pyr	-2182 ± 64	-0.71 ± 15
6bb2-pyr	-2828 ± 17	-35 ± 17

[a] (kcal mol^{-1}).

Having selected **6bb2** as the most stable system for the β-sheet formation, the H-bonding patterns were investigated in order to exclude any other possible arrangements. Aiming at this, the **6bb2** was also simulated with the parallel orientation of the six strands (**6bb2-para**). As shown in Figure 3.29, after the 15 ns MD trajectory, despite the stability revealed by the RMSD (orange line), the last

snapshot of the β-sheet structure is visibly deformed comparing to the starting conformation. In line with this, the SS profile is evidently affected (orange line in Figure 3.29), prompting a poor organization of the β-sheet in such H-bonding pattern. Notably, **6bb2-para** with its 36/66 ± 5 residues in β-sheet resembles more the SS content of the shorter **6bb1** system (33/42) than its antiparallel version **6bb2** (59/66 Figure 3.29, orange, black and red lines respectively). As a confirmation of this behavior, the ramachandran plot was also calculated. Now the perfect parallel organization of the β-sheet is characterized by backbone dihedral angles φ and ψ of about -119° and 113°,[83] respectively, therefore located in the second square of the top left area of the ramachandran plot. Interestingly, only the 86.4% of the backbone dihedral angles of **6bb2-para** occupies a favored region (sky blue region marked with black dots, Figure 3.30) over which they are disorderly distributed.

Remarkably, the poor structural organization and the reduced H-bonds formation of **6bb2-para** are also reflected in the energetic profile: if the <vdW> is almost unvaried as compared to **6bb2** (-155 and -151 kcal mol^{-1} respectively), the <EEL> energy is drastically changed being 20 times higher than those of **6bb2** (-43 and -925 kcal mol^{-1}, respectively Table 3.9). All these findings suggest the poor reliability of this arrangement, thus discarding it from the plausible H-bonding patterns that the designed β-sheet may adopt.

Additionally, in order to further corroborate the selection of **6bb2** over **6bb1**, explorative 15 ns MD simulations were also run on the pyrenyl modified systems **6bb1-pyr** and **6bb2-pyr**. At the visual inspection, we have previously observed that the **6bb1** system clearly loses its structural compactness (Figure 3.29). Conversely, the pyrenyl-modified β-sheet **6bb1-pyr** seems to retain its conformation as compared to the starting point, while damped RMSD fluctuations are observed (blue line in Figures 3.29). The explanation of this findings can be that the shorter strands length of the **6bb1-pyr** array, induce the pyr to be in strict vicinity among them, thus depicting a crowded situation as shown in Figure 3.31a. On the other hand, the longer β-strand of **6bb2-pyr** significantly separate the pyrenyl functionalities located at the N and C-term edges (Figure 3.31a). So, aiming at quantifying such separation, the distance between all the pyr functionalities located at the N-term edge and those at the C-term was calculated as a function of the simulated time (Figure 3.31b). Indeed the crowded vicinity of the pyr groups in **6bb1-pyr** corresponds to a distance in the range

of 20 Å, which is significantly smaller with respect to the pyr distance in the **6bb2-pyr** system (35 Å, Figure 3.31b). Consequently this vicinity can explain the observed rigidity of the **6bb1-pyr** structure. The SS profile is rather stable along the time and unvaried compared to the non-modified **6bb1** (32 and 33/42 ± 2 residues in β-sheet for **6bb1-pyr** and **6bb1**, blue and black lines respectively in Figure 3.29).

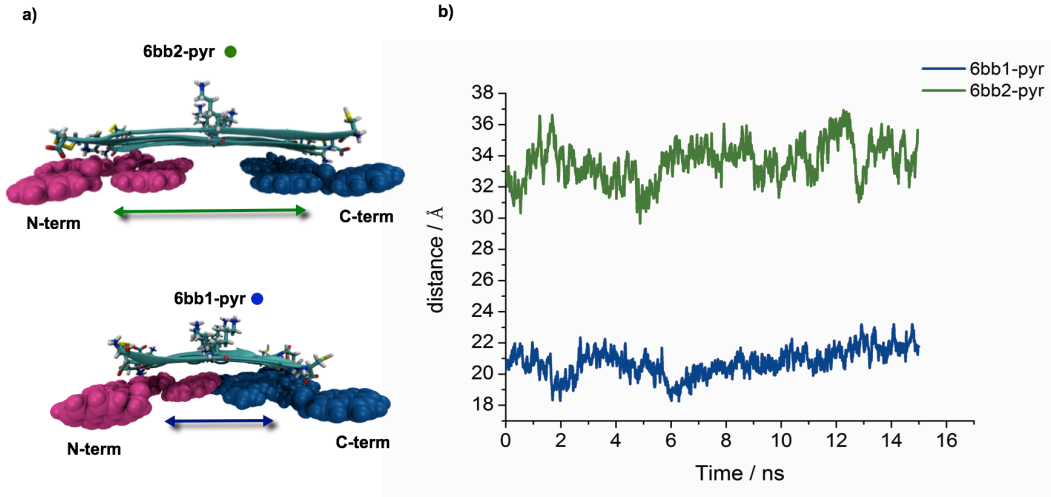


Figure 3.31. (a) Side view of the starting structure of **6bb2-pyr** (top) and **6bb2-pyr** (bottom) where the pyr functionalities at the N-term edge are colored in magenta, the ones at the C-term in blue. The distances between all the pyr at the two edges are evidenced by an arrow and are plotted as a function of time in green for **6bb2-pyr** and blue for **6bb1-pyr** (b).

On the other hand, thanks to the major distance among the pyr, the **6bb2-pyr** system results more mobile as the RMSD demonstrates (green line in Figure 3.29). Nevertheless no affection of the SS content was detected as displayed by the constant profile along the simulated time (green line in Figure 3.29). Furthermore, by visualizing the backbone dihedral angles φ and ψ a distortion of the β-sheet was observed for **6bb1-pyr** with almost the 5% of backbone angles in allowed regions and a more spread displacement of the residues in the β-sheet region (Figure 3.32). This is probably due to the crowded vicinity of the pyrenyl functionalities inserted in the shorter strands that induces the distortion of the backbone angles. In fact, the increased distance between the pyrenyl functionalization avoids the backbone distortion in **6bb2-pyr**, thus obtaining a ramachandran plot with only 1.5% of residues in allowed regions and a better displacement of the residues in the expected

top left β-sheet region (Figure 3.32), resulting just slightly changed compared to the **6bb2** system (100%, Figure 3.30).

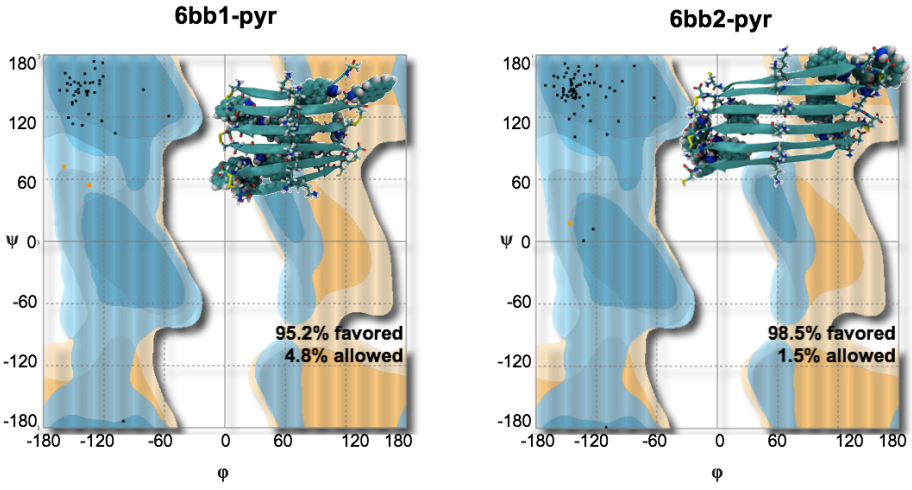


Figure 3.32. The ramachandran plots are reported with the % of residues in favored (black dots) and allowed region (orange dots) for **6bb1-pyr** (left) and **6bb2-pyr** (right).

When it comes to the evaluation of the energetic profiles, the interactions between the pyr seems to be unfavorable both in **6bb1-pyr** and in **6bb2-pyr** as they drastically affect the <vdW> energy, dropped to negligible values (-0.71 and -0.35 kcal mol^{-1}, correspondingly Table 3.9). On the other hand, the sharp difference in the <EEL> energy found between the shorter **6bb1-pyr** and the longer **6bb2-pyr** (-2182 and -2828 kcal mol^{-1}, respectively, Table 3.9), confirms, along with the structural analysis the major stability of the **6bb2-pyr** system. Hence, availing of all these evidences, from this preliminary analysis we were able to endorse the antiparallel β-sheet formed by longer strands as the most stable option for our purposes, thus abandoning the **6bb1** system and its pyr modified β-sheet as valuable choice. In the next sections further studies on the **bb2** system will therefore follow.

3.2.5.2 Presence and number of pyrenyl functionalities

In the previous section, the stability of **6bb2** over **6bb1** was assessed and preliminary studies on the pyr-modified **6bb2-pyr** were carried out. Starting from this partial investigation, we now want to evaluate the effect of the presence of the anchoring groups, namely the pyrenyl moieties (pyr), along the time by elongating the **6bb2-pyr** simulation up to 50 ns. As previously observed this system is characterized by two pyr-coupled Glu per strand (Glu-pyr), one at each N- and C-term extremity. The six-stranded β-sheet will therefore count 12 functional groups in total. In parallel, we envisaged another modified system, **6bb2-1pyr**, aimed at addressing the effect of the number of the pyr groups within the designed β-sheet. Hence, **6bb2-1pyr** was designed with one Glu-pyr per strand located at the C-term extremity of each strand, whereas the second pyr was replaced by an Ala aa so to maintain the side hydrophobicity. Doing so, **6bb2-1pyr** system will result in an alternating functionalized β-sheet with an increased distance between the pyr. Hence, the comparison of the behavior of the **6bb2-pyr** and **6bb2-1pyr** will clarify the effect of the presence and number of the pyrenyl anchoring groups on the stability of the β-sheet. Meanwhile, 50 ns simulation of **6bb2** were also run as a reference system. However, as described in section 3.2.4 **6bb2**, exposes free Glu towards the hydrophobic side, thus probably disturbing the full hydrophobicity of that face. Hence, 50 ns MD simulation of **6bb2-ala** were also run as a second reference system, where each free Glu is replaced by an Ala residue (the resulting single strand sequence is C**A**$(TA)_2$K$(AT)_2$**A**C), thereby preserving the hydrophobic character of the side.

Starting with the pyr-modified systems (**6bb2-1pyr** and **6bb2-pyr**), by comparing their starting and last snapshots a remarkable structural rearrangement is observed, which leads to the change of the conformation from a planar, flat to a pronounced β-sheet curvature in both the cases (Figure 3.33). This behavior significantly differs from the reference systems (**6bb2** and **6bb2-ala**) that maintain their planarity even after 50 ns of simulation (Figure 3.33). In fact, the peculiar conformational changes of the modified systems are notable in the calculated RMSD (red and blue lines in Figure 3.33, top plot), compared to the most stable trend of

6bb2 and **6bb2-ala** (black and green lines respectively in Figure 3.33, top plot), fluctuate more along the simulated time.

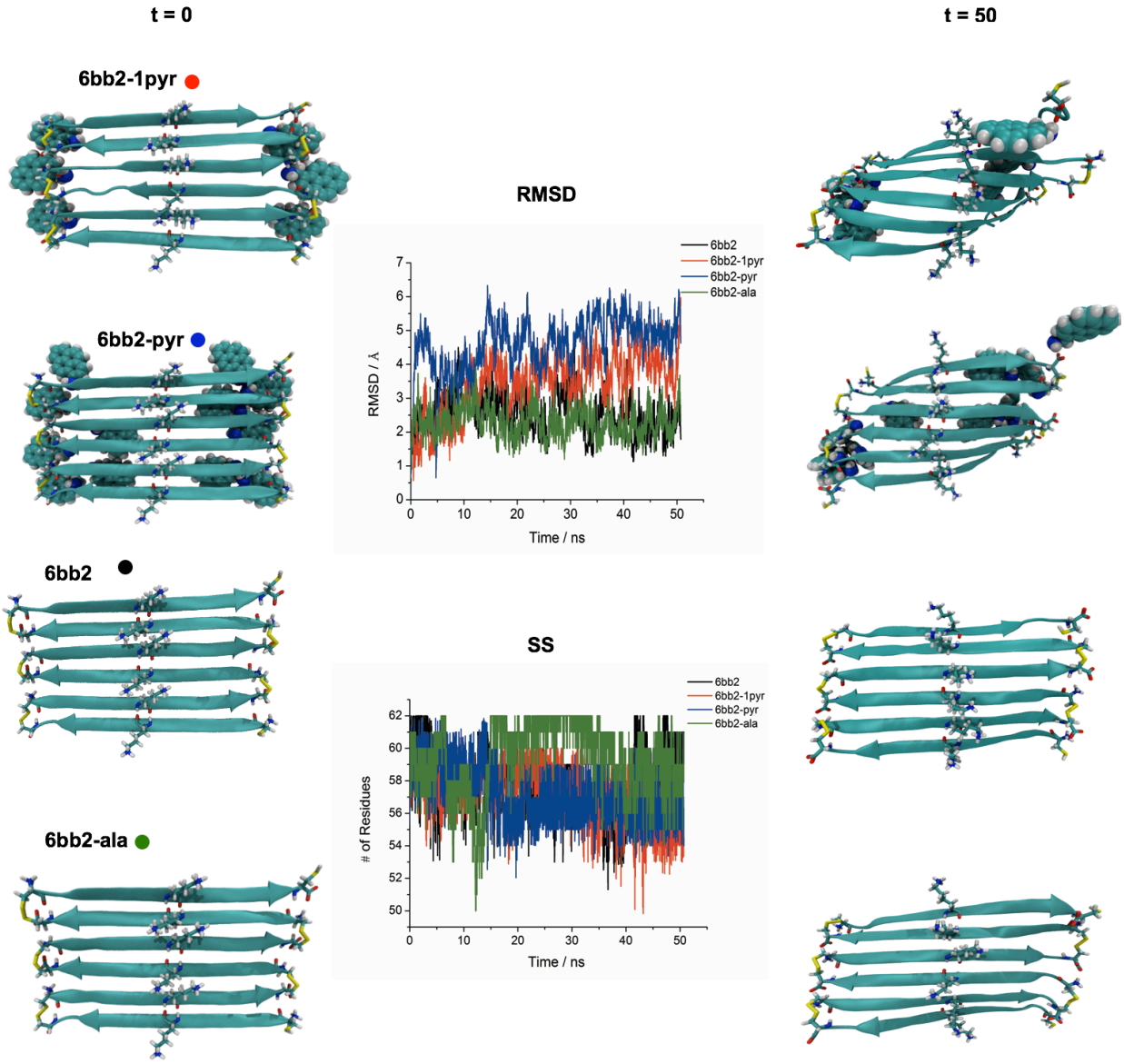


Figure 3.33. The starting structure (left panel), the RMSD (top) and the SS (bottom, middle panel), and the last snapshot (right panel) of the 50 ns MD simulation are reported for **6bb2-1pyr** (red), **6bb2-pyr** (blue), **6bb2** (black) and **6bb2-ala** (green) from the top to the bottom. The 3D representations of the β-strand are drawn as cartoon arrows whose direction evidences the antiparallel secondary structure; for clarity all the side chains are not shown but the hydrophilic face is marked by the Lys residues drawn in licorice, while the assembling points among the strands are evidenced by the lateral disulphide bonds between Cys drawn in licorice; when present, the pyrenyl functionalization is drawn in vdW.

Hence, the structural differences between modified and reference systems must necessarily be caused by the presence of the pyr functionalities and their

intermolecular interactions. In order to quantify those pyr interactions, the distance between the center of mass of each pyr and its closest neighbor was calculated along the 50 ns trajectories. In particular, the pyr located externally, such as pyr 1 in Figure 3.34a, can interact only with one neighbor, namely pyr 2. On the other hand, the pyr located in the middle (pyr 2 in Figure 3.34a) can interact with its two neighbors (pyr 1 and 3). Therefore the total number of interacting pyr pair is to N-n/2, where N is the total number of pyr and n is the number of pyr located externally interacting only with one neighbor (/2).

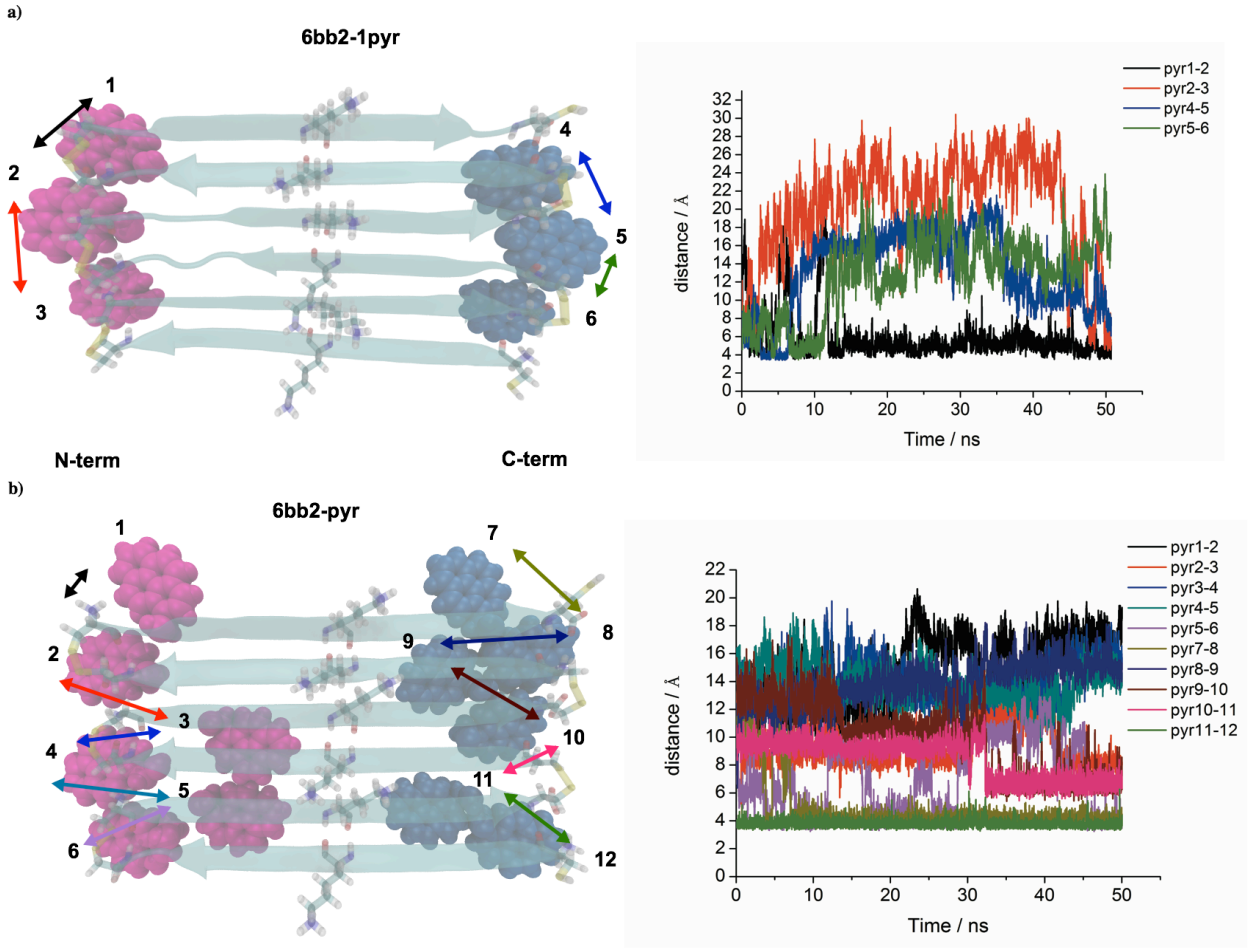


Figure 3.34. Top view of the starting structure of **6bb2-pyr** (a) and **6bb2-pyr** (b) where the pyr functionalities at the N-term edge are colored in magenta and those at the C-term in blue. The distances between the pyr, numbered in increasing order from the N-term to C-term edge, are marked by arrows and plotted as a function of time on the right panel.

Hence, **6bb2-1pyr** with 6 pyr and 4 external pyr can form 4 pairs (Figure 3.34a), while **6bb2-pyr** with 12 pyr and 4 external groups can form 10 possible pyr

pairs (Figure 3.34b). Interestingly some pyr couples are found to rearrange in order to undergo π-stacking (distance of about 4 Å, Figure 3.34). Specifically, in **6bb2-1pyr**, being the pyr alternated one per strand, it is noted that they are more separated in distance (from a minimum of 4 to a maximum of 26 Å) and only the couple pyr 1-2 is found to stably stack along the trajectory (black line in Figure 3.34a). On the other hand, within the **6bb2-pyr** the 12 pyr enable an increased number of possible contacts and rearrangement towards favorable interactions. In particular, couples 5-6, 7-8 and 11-12 undergo stable π-stacking (violet, olive and green lines in Figure 3.34b), whereas couples 9-10, 10-11 and 2-3 (red, burgundy and pink lines in Figure 3.34b) achieve a closer distance only during the last ns of the simulation. As emerged from these findings and the visual inspection, the pyr moieties force the β-sheet backbone to curve in order to stack between them, therefore being responsible of the RMSD fluctuations and of the affection of the backbone planarity of the 2D platform. Although **6bb2-pyr** and **6bb2-1pyr** undergo considerable structural rearrangement no SS loss is observed (Figure 3.33). In fact, the SS content is identical in the two modified cases (57/66 ± 2 residues in β-sheet, red and blue lines in Figure 3.33), and discloses similar profile with respect to the reference systems (59/66 ± 2 residues in β-sheet, black and green lines in Figure 3.33). This suggests a good intrinsic stability of the designed sequence assembled in β-sheet, further confirmed by the analysis of the backbone dihedral angles. In fact, **6bb2-pyr** and **6bb2-1pyr** show the correct distribution of the backbone dihedral angles in the ramachandran plot, thus an additional proof of the proper β-sheet settlement (100 and 97% of occupied favored region, respectively Figure 3.35a). In line with the above analysis also the reference cases unveil exemplar ramachandran plots (100%, Figure 3.35b).

Up to now, despite the visible conformational changes occurring in the pyr-modified systems, a general structural stability of the systems understudy can be stated.

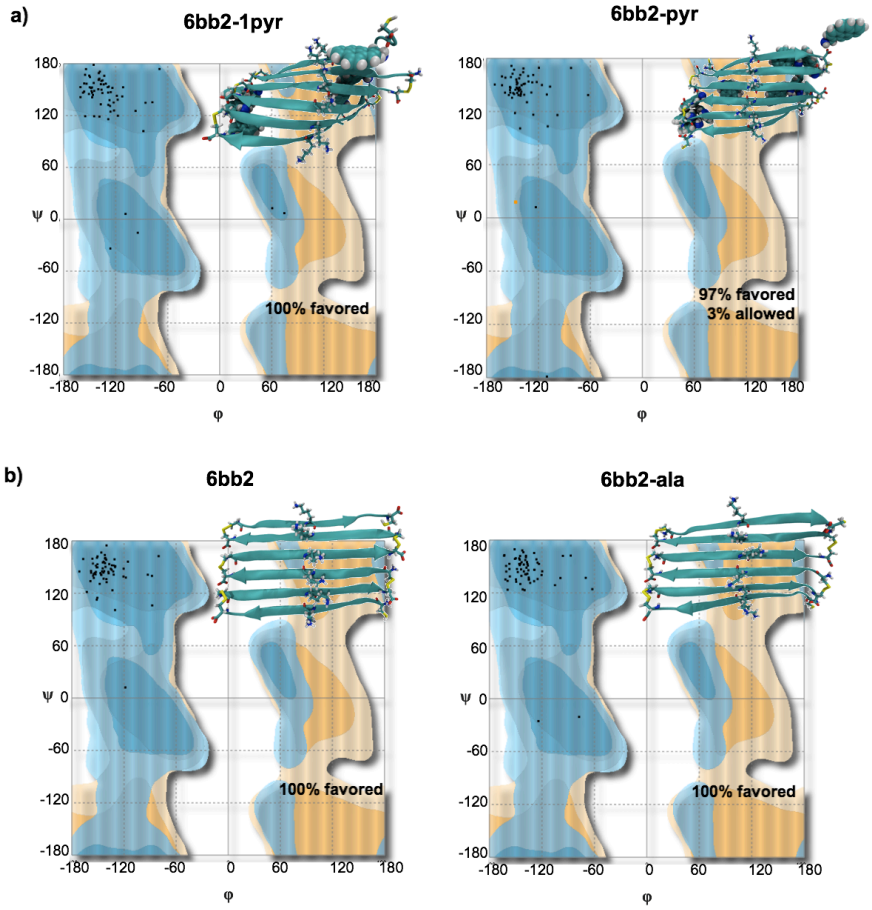


Figure 3.35. The ramachandran plots are reported with the % of residues in favored (black dots) and allowed region (orange dots) for **6bb2-1pyr** (a, left), **6bb2-pyr** (a, right), **6bb2** (b, left) and **6bb2-ala** (b, left).

To further evaluate the stability and compactness of the designed proteinic β-sheet the radius of gyration (Rg) was calculated as a function of time (Figure 3.36a). Between the modified systems, **6bb2-1pyr** presents the lowest average Rg value (15.4 Å Figure 3.34b), which refers to a more compact conformation as compared to **6bb2-pyr** (15.8 Å). Such difference, even though small, was then explained by calculating the distance between the two groups of pyr at the N- and C-term edges as previously calculated (Figure 3.36c). In fact, the distance between the pyr at the two edges is mainly constant along the 50 ns of simulation in **6bb2-pyr** (blue line in Figure 3.36c), while it visibly decreases along the **6bb2-1pyr** trajectory (red line in Figure 3.36c), most likely explaining its relatively major curvature.

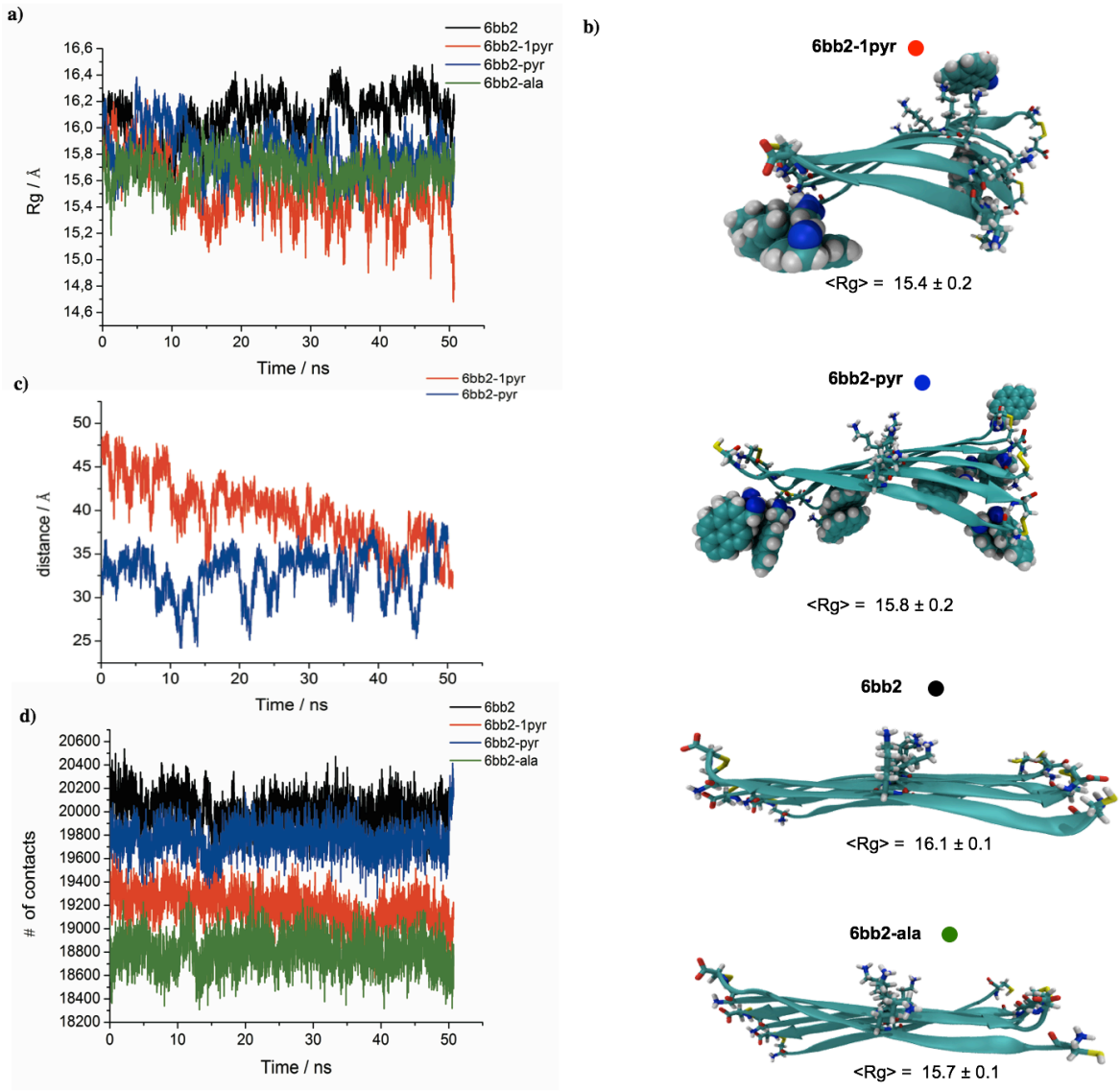


Figure 3.36. The gyration radius (Rg) is plotted as function of time (a). The last snapshot of **6bb2-1pyr, 6bb2-pyr, 6bb2** and **6bb2-ala** is drawn as cartoon in side view to evidence the degree of curvature of the structure; the pyrenyl functionalization are drawn in vdW when present (b). The distance between the pyr moieties at the N-term and the C-term edges is plotted as a function of time for **6bb2-1pyr** and **6bb2-pyr** (c). The number of contacts among all the side chains is plotted as a function of the simulated time (d). In each plot, **6bb2-1pyr** is colored in red, **6bb2-pyr** in blue, **6bb2** in black and **6bb2-ala** in green.

In fact, the pyr moieties, being in constant proximity in **6bb2-pyr** enable a dense network of side chains contacts (average number of 19756), whereas getting closer only at the last ns of the **6bb2-1pyr** trajectory, induce the formation of less side chains contacts (19142, blue and red lines, Figure 3.36d). Concerning the reference systems, **6bb2-ala** results more compact than **6bb2** (15.7 and 16.1 Å respectively,

Figure 3.34b). Despite small, this difference in compactness is now explained by the peculiar secondary non-covalent interactions occurring within **6bb2** system. In fact, **6bb2** is simulated with the free lateral Glu that are able to form H-bonds with the N-term groups present at their proximity (Table A.3 and Figure A.4, *Appendix*). The H-bonds formation provokes the increase distance among the strands and consequently the rise of the **6bb2** Rg. Conversely, the **6bb2-ala** is simulated with lateral Ala substituting the free Glu, thus any Coulombic interactions can take place. Moreover the low % of occupancy of the aforementioned H-bonds explains the small gap in Rg between the **6bb2** and **6bb2-ala** (Table A.3). As a consequence, the H-bonding interactions between the Glu side chains significantly contribute to the major number of side chains contacts of **6bb2** (20006) with respect to **6bb2-ala** (15372, black and green lines in Figure 3.36c).

As a matter of fact, the backbone H-bonds play a fundamental role in the formation and stabilization of the secondary structure of a protein-like system. For this reason the β-sheet backbone H-bonds were further analyzed from a dynamical point of view. For this purpose the Hydrogen-bonding content (Hbc) was calculated as the % of all H-bonding donors within each β-sheet actually involved in H-bonds during the trajectories over the total number of the available donors ($N_{H\text{-}bonds}$ / $tot_{H\text{-}bonds}$ *100, values in Table 3.10). Additionally, the % of donor as a function of the occupancy is also plotted to picture the distribution of the individual Hbc of the donors during the trajectories (Figure 3.37).

Table 3.10. The Hbc and the % of occupancy of the total number of donors actually involved in H-bonds are reported for the four cases in the absence (left) and in the presence of the graphene surface (right). The standard deviation is also reported.

System	Hbc %	<%occ>	System	Hbc %	<%occ>
6bb2	81	58 ± 12	g-6bb2	73	41 ± 26
6bb2-ala	84	59 ± 14	g-6bb2-ala	80	65 ± 12
6bb1-pyr	80	56 ± 14	g-6bb1-pyr	82	57 ± 20
6bb2-pyr	80	59 ± 15	g-6bb2-pyr	80	53 ± 21

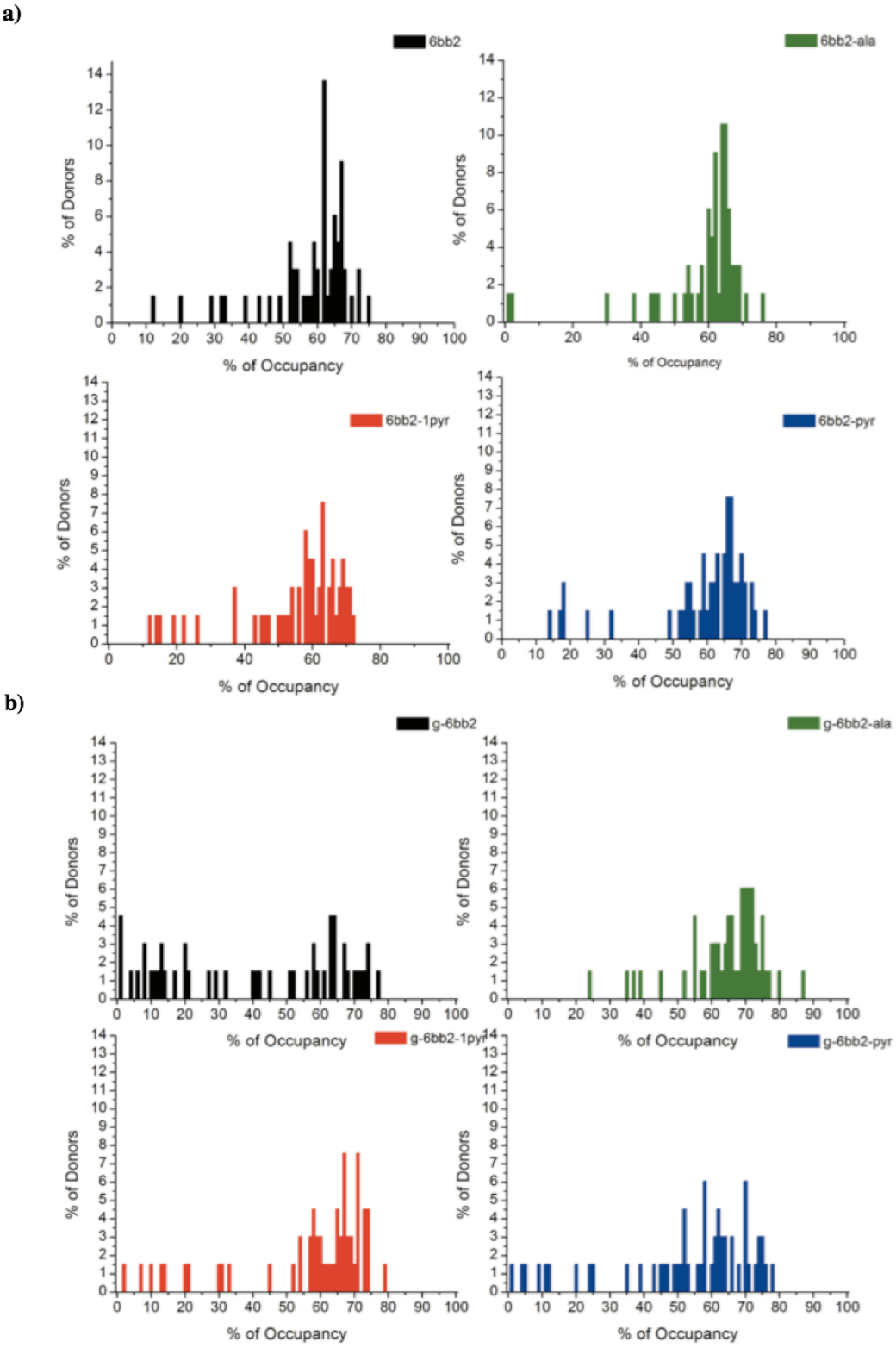


Figure 3.37. The % of donors as a function of the % of occupancy is plotted for **6bb2** and **g-6bb2** (black, a and b respectively), **6bb2-ala** and **g-6bb2-ala** (green, a and b respectively), **6bb2-1pyr** and **g-6bb2-1pyr** (red, a and b respectively), **6bb2-pyr** and **g-6bb2-pyr** (blue, a and b respectively).

As reported in Table 3.10, the four systems present high Hbc (≥ 80%) with an average occupancy of around 58%, meaning that the 80% of donors involved in H-bonds are actually occupied for the 58% of the simulation. Afterwards aiming at visualizing the dynamics of the H-bonds along the trajectories, the donors are plotted as a function of their H-bonds occupancy as shown in Figure 3.37a. From this analysis emerges an equal distribution over the four cases with the majority of donors presenting an occupancy shifted towards high % over time (from 50 to 80 % of occupancy, Figure 3.37a) regardless of the presence or the absence of the pyrenyl functionalities. Therefore, by quantifying the structural integrity in terms of SS, through ramachandran plot and H-bonding analysis the overall stability of the designed β-sheet was confirmed and all the structural differences analyzed did not evidenced substantial alteration of the arrangements both in the presence and in the absence of the pyr anchoring groups.

Finally, the non-bonded interaction energies, EEL and vdW, were calculated for the four systems (Figure A.5, *Appendix*) and the average values over the last 30 ns of the simulations are reported in Table 3.11. As observed in the previous section, the vdW contribution is ascribable to the side chains contacts and in this case it results constant and consistent with the structural analysis carried out so far (Figure A.5a *Appendix* and Table 3.11). For instance, between the pyr-modified systems, **6bb2-1pyr** has a higher <vdW> energy (-74 kcal mol^{-1}) as compared to **6bb2-pyr** (-44 kcal mol^{-1}) ascribable to the more stable π-stack occurring among the pyr functionalities (Figure 3.31), while displays a lower <EEL> contribution compared to **6bb2-pyr** (-2207 and -2794 kcal mol^{-1}, respectively) in line with the profile pictured by the % of occ of their backbone H-bonds (56 and 59 %, respectively, Table 3.10).

Table 3.11. Summary of the average electrostatic (EEL) and van der Waals (vdW) energies and standard deviations calculated for the systems understudy over the last 30 ns of the trajectories.

System	<EEL>[a]	<vdW>[a]
6bb2-1pyr	-2207 ± 49	-74 ± 14
6bb2-pyr	-2794 ± 44	-44 ± 17
6bb2	-867 ± 78	-158 ± 11
6bb2-ala	-1077 ± 63	-130 ± 11

[a] (kcal mol^{-1}).

Similarly, the major number of contacts observed for **6bb2** corresponds to the highest vdW energy (<vdW> = -158 ± 11 kcal mol^{-1}), which decreases in the **6bb2-ala** case accordingly to the reduction of the side chains interactions (-130 ± 11 kcal mol^{-1}). Conversely, the EEL contribution, accounting for the H-bonds, nicely correlates with the Hbc and % of occupancy. In fact, among the reference systems, **6bb2** is less stable compared to the **6bb2-ala** (<EEL> = -867 and -1077 kcal mol^{-1} Table 3.11), in agreement with the lower Hbc just analyzed (81 and 84 % respectively). From the whole analysis it emerges a general stability along the time of the pyr-modified systems (**6bb2-1pyr** and **6bb2-pyr**) that, compared to the reference systems (**6bb2-ala** and **6bb2**), were demonstrated to conserve their structural integrity and compactness. Importantly, neither affections of the β-sheet organization nor secondary structure losses are detected when the pyr functionalities are inserted in the β-sheet structure regardless of their number and positions. Lastly, the β-sheet structural stability was also assessed with or (**6bb2-ala**) without (**6bb2**) a full hydrophobic graphene interacting side, implying that the intrinsic stability of the system relies on the backbone alternation of the hydrophobic/hydrophilic aa and it is not affected by the lateral insertion of a hydrophilic aa (Glu) or a modified version of it (Glu-pyr). Hence, in order to select the ideal bi-functional proteinic graphene interfacing platform, the simulations of the behavior of these systems once absorbed onto the graphene surface are required.

3.2.5.3 Adsorption on graphene

Consequently, 50 ns MD simulation of the systems understudy in the previous section were carried out modeling the adsorption onto graphene (**g-6bb2-1pyr**, **g-6bb2-pyr**, **g-6bb2** and **g-6bb2-ala,**). By analyzing the trajectories and the structural aspects some differences start emerging (Figure 3.38). At first glance, comparing the starting and the last snapshot of the trajectories of the pyr-modified systems no remarkable structural changes seems to occur (Figure 3.38, first and last panel). In fact, the presence of the pyrenyl moieties, by anchoring the graphene surface, clearly dumps the fluctuations of the β-sheets backbone yielding a stable and constant RMSD profile, whilst slightly more stable for **g-6bb2-1pyr** than for **g-6bb2-pyr** (red and blue lines in Figure 3.38). Despite small, this difference might be ascribed to the higher number of the pyr groups in **g-6bb2-pyr** that induces more structural changes, in order to find a stable arrangement. In fact, by calculating the distance between all the pyr at the N-term and those at the C-term edge of the two modified systems, it is obeserved that the pyr distance in **g-6bb2-1pyr** increases along the simulation (red line in Figure 3.38) implying a well separation among the anchoring groups occupied in the interaction with graphene. On the other hand in **g-6bb2-pyr**, the distance between the pyr decreases along the simulated time (blue line in Figure 3.38), so by getting closer the anchoring groups induce the distortion of the entire β-sheet backbone as observed in the RMSD. In line with this, an unperturbed SS profile is observed for **g-6bb2-1pyr** (59/66 ± 1 residues in β-sheet), while a reduction of its content is detected for **g-6bb2-pyr** (53/66 ± 2 residues in β-sheet, red and blue lines in Figure 3.38). As a confirmation, the calculated ramachandran plot results exemplar for **g-6bb2-1pyr** with 100% of backbone dihedral angles found in the favored regions, while a reduced % of residues are found in the favored region for **g-6bb2-pyr** (96%) spread over the β-sheets region (Figure 3.39a)

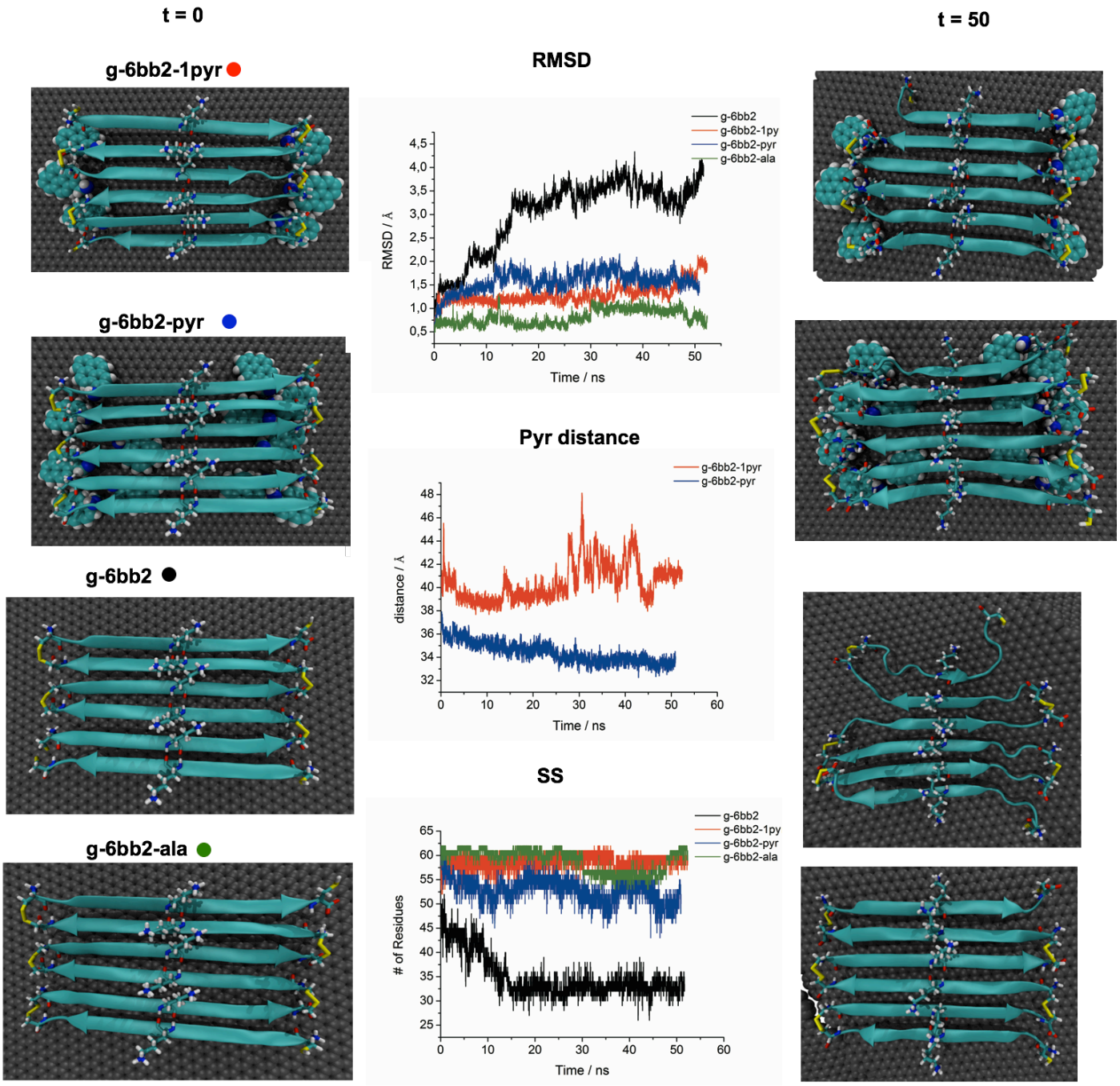


Figure 3.38. Top view of the starting structure (left panel), the RMSD (top) and the SS (bottom, middle panel), and the last snapshot (right panel) of the 50 ns MD simulations are reported for **g-6bb2-1pyr** (red), **g-6bb2-pyr** (blue), **g-6bb2** (black) and **g-6bb2-ala** (green) from the top to the bottom. The 3D representations of the β-strand are drawn as cartoon arrows whose direction evidences the antiparallel secondary structure; for clarity all the side chains are not shown but the hydrophilic face is marked by the Lys residues drawn in licorice, while the assembling points among the strands are evidenced by the lateral disulphide bonds between Cys drawn in licorice; when present, the pyrenyl functionalization is drawn in vdW.

.

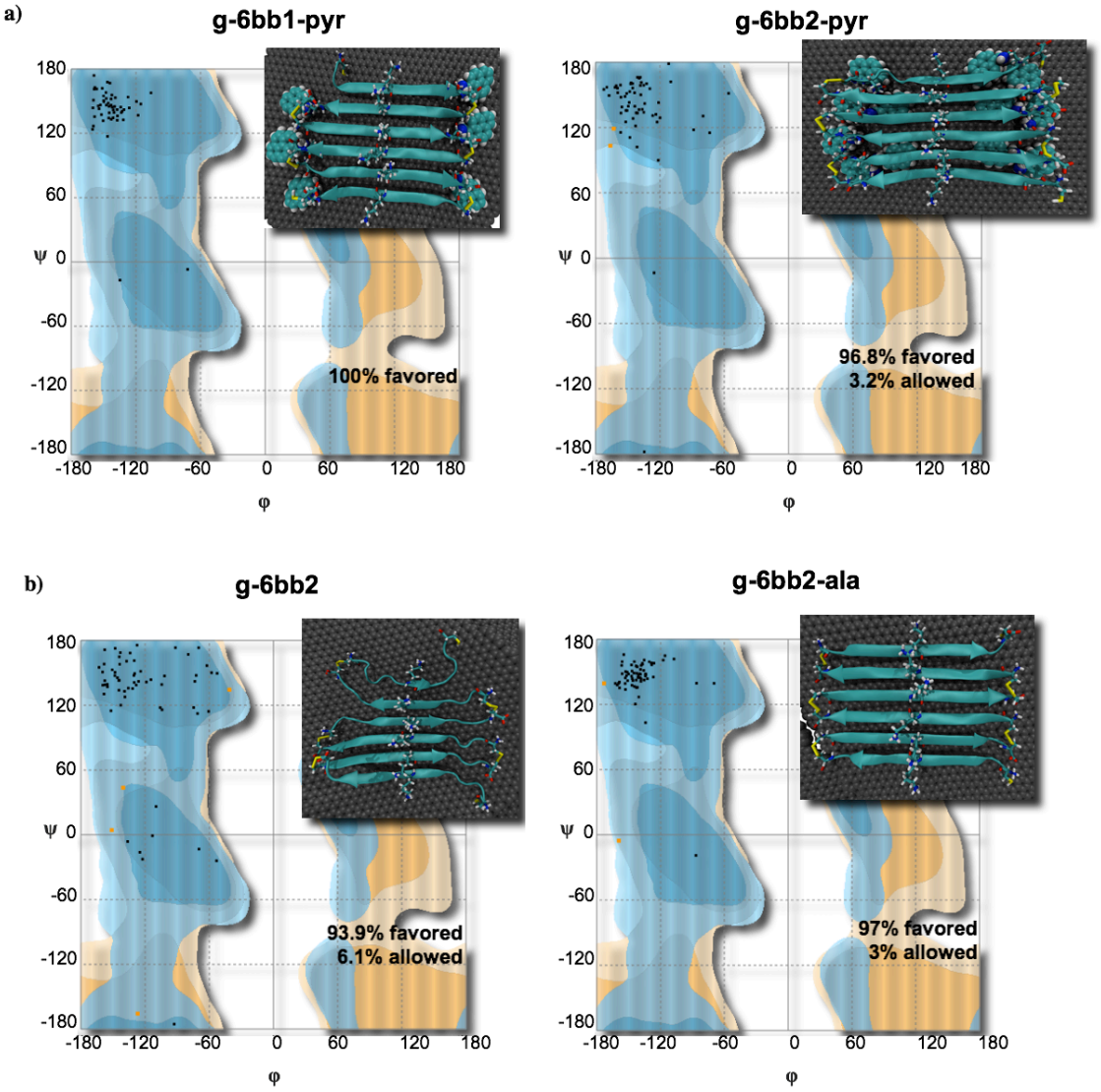


Figure 3.39. The ramachandran plots are reported with the % of residues in favored (black dots) and allowed region (orange dots) for **g-6bb2-1pyr** (a, left), **g-6bb2-pyr** (a, right), **g-6bb2** (b, left) and **g-6bb2-ala** (b, right).

As far as the reference systems (**g-6bb2** and **g-6bb2-ala**) are concerned, after 50 ns MD simulation, the conformation of **g-6bb2** appears visibly affected upon adsorption on graphene, while **g-6bb2-ala** does not undergo structural changes. From a dynamical point of view the RMSD highlights further this discrepancy showing the **g-6bb2** profile (black line in Figure 3.38) largely fluctuating as compared to the starting structure, whereas the **g-6bb2-ala** behavior (green line) results the most stable and constant among the systems understudy (Figure 3.38). Notably, in agreement with the interaction mode observed in the CNT-Ctx study (section 3.1.5), the lateral Glu residues are found to interact with the graphene surface exclusively through their

alkylic side chains, while forcing the charged moiety towards the amino groups of the strand as already observed in **6bb2** simulated alone in solution (Table A.4 and Figure A.6, *Appendix*). Of note, due to the presence of graphene the carboxylic groups are most constantly pointing towards the N-term groups, thereby yielding H-bonding interactions with a significantly higher % of occupancy (from 3 to 30% Table A.3 and A.4, *Appendix*), compared to **6bb2** (from 2 to 8% Tables A.3). As shown in Figure A.7, those collateral interactions induce structural rearrangements responsible of the fluctuation of the RMSD (black line in Figure 3.38), instead absent when the Glu are replaced by Ala in **g-6bb2-ala** (green line in Figure 3.36). As a consequence, also the SS content of **g-6bb2** is drastically affected over time (35/66 ± 4 residues in β-sheet, black line Figure 3.38). Conversely, the SS profile of **g-6bb2-ala** displays a stable trend along the simulated time (59/66 ± 2 residues in β-sheet, green line in Figure 3.38), in line with the visual inspection (Figures 3.38). Moreover, the evaluation of the ramachandran plots (Figure 3.39b) confirms the poor stability of the **g-6bb2** system with its reduced % of residues in the favored region (93%) and a spread location allover the β-sheet area. In contrast with the **g-6bb2-ala** system where a perfectly concentrated reduced % of residues is observed in the β-sheet area (97% in favored region).

Hence, from all these findings we can clearly say that the presence of the pyr does not disturb the structural integrity of the β-sheet. At the same time it is worth noting that the presence of a hydrophilic residues, such as Glu, drastically affect the β-sheet conformation when adsorbed onto the graphene surface, thereby highlighting the importance of the fully hydrophobicity of the interacting side.

To further confirm these findings, the Rg was also calculated. Regarding the modified systems, although their <Rg> values are similar (18.3 and 18.2 Å), **g-6bb2-1pyr** presents a more stable and constant Rg profile compared to **g-6bb2-pyr** (red and blue lines, figure 3.40a).

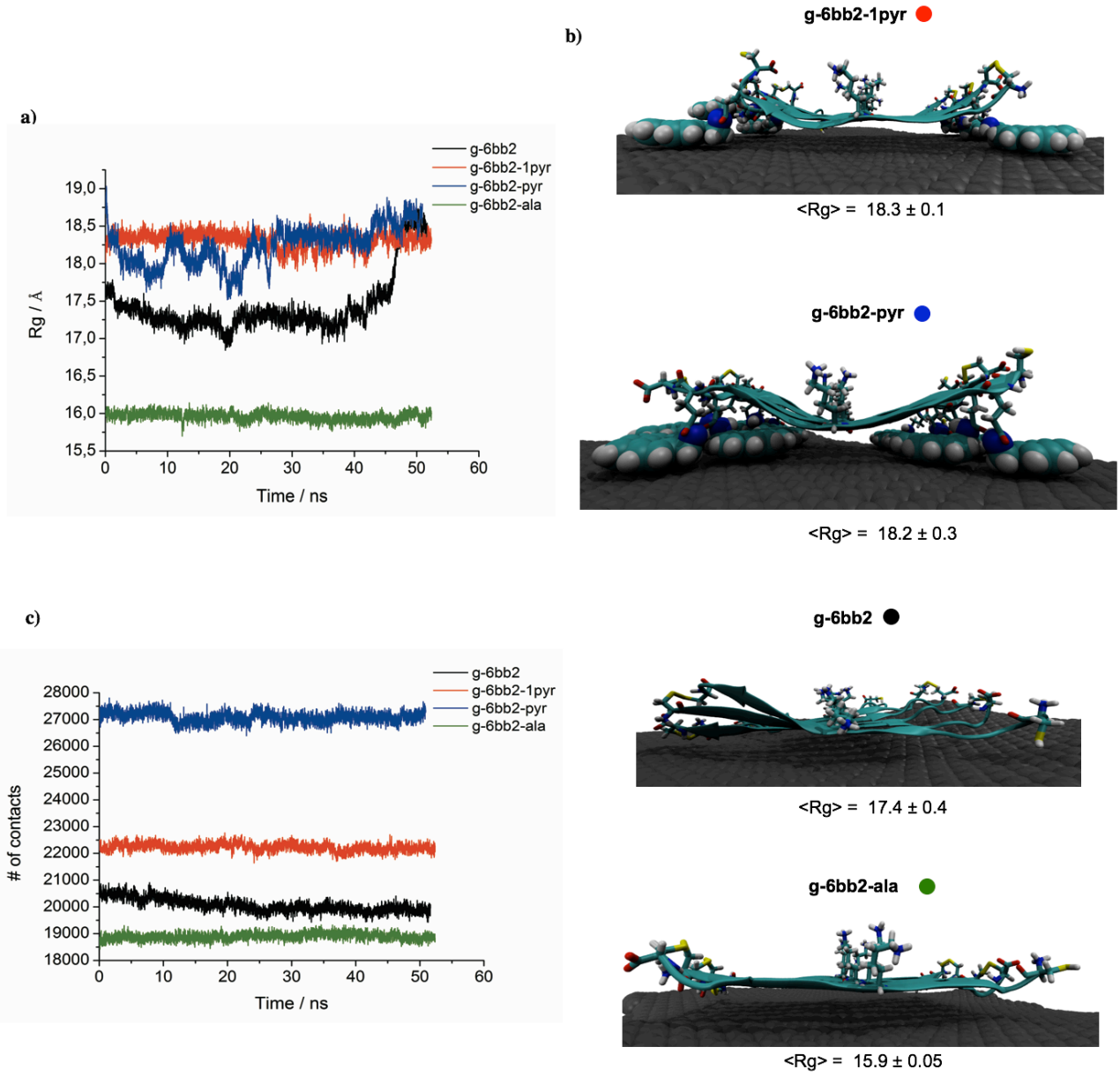


Figure 3.40. The gyration radius (Rg) is plotted as a function of time (a). The last snapshot of **g-6bb2-1pyr** (red), **g-6bb2-pyr** (blue), **g-6bb2** (black) and **g-6bb2-ala** (green) is drawn as cartoon in side view to evidence the distortion of the structure when occurred; the pyrenes functionalization are drawn in vdW when present (b). The number of contacts is plotted as a function of the simulated time for the four systems (c).

In fact, by monitoring the distance between the pyr functionalities (Figure 3.38), a more crowed situation in **g-6bb2-pyr** was described, most likely responsible of the structural changes and the fluctuating profiles in the RMSD and the Rg. This is also deciphered in the highest number of contacts among the side chains in **g-6bb2-**

pyr than in **g-6bb2-1pyr** (blue and red lines in Figure 3.40c). Meanwhile, the Rg analysis of the reference systems reveals that **g-6bb2** loses its compactness and planarity (<Rg> = 17.4 Å) along the simulated time, most probably due to the lateral distortion induced by the Glu interaction mode. The perfect planarity and protein compactness is instead conserved in **g-6bb2-ala** as the <Rg> = 15.9 Å clearly demonstrates. Finally, thanks to the contribution of the encountered Glu/N-term H-bonds the number of contact of **6bb2** is higher than **6bb2-ala** also when adsorbed on graphene.

Afterwards, by analyzing the interaction energies (EEL and vdW) between each β-sheet and the graphene surface, the adsorption results reasonably driven by the vdW energy (Table 3.12 and Figure A.7, *Appendix*). Interestingly, the pyr-modified systems display different profiles: in fact, bearing the double of pyrenyl functionalities **g-6bb2-pyr** is more strongly interacting with graphene (<vdW> = -454 ± 13 kcal mol^{-1}) than **g-6bb2-pyr** (<vdW> = -340 ± 9 kcal mol^{-1}). However, it is worth noting that by doubling the number of pyrenes the energetic contribution does not double correspondingly. Conversely, the observed crowding, provoked by increasing the number of pyr functionalities, causes structural rearrangements that affect the β-sheet stability.

Table 3.12. ΔG of binding and non-bonded interaction (vdW and EEL) energies, averaged over the last 30 ns of the simulations are reported for the **g-6bb2**, **g-6bb2-ala**, **g-6bb2-1pyr** and **g-6bb2-pyr** trajectories.

System	<vdW>[a]	<EEL>[a]	ΔG[a]
g-6bb2-1pyr	-340 ± 9	-2.1 ± 1.7	-257 ± 9
g-6bb2-pyr	-454 ± 13	-2.3 ± 2.2	-381 ± 9
g-6bb2	-304 ± 22	-9.5 ± 4.1	-2.7 ± 25
g-6bb2-ala	-249 ± 7	-0.57 ± 1.6	-

[a] kcal mol^{-1}

Regarding the reference systems, despite their structural differences evidenced so far the energetic profiles are similar (<vdW> = -304 and -249 kcal mol^{-1} for **g-6bb2** and **g-6bb2-ala** respectively, Table 3.12 and Figure A.7). Nevertheless, due to the free lateral Glu a major contribution of the EEL is observed for **g-6bb2** than for **g-6bb2-ala** (<EEL> = -9.5 and -0.57 kcal mol^{-1} respectively, Table 3.12 and Figure A.7). Additionally, for a more reliable estimation of the interaction energy a

molecular mechanics/Generalized-Born solvent area (MM/GBSA) analysis was performed on the trajectory calculating the binding free energy (ΔG) as reported in Table 3.12. The latter energetic analysis reveals that the adsorption of **g-6bb2** is evidently not favored (ΔG = -2.7 ± 25 kcal mol^{-1}), while the role of the pyr moieties, revealing to be fundamental for the stabilization of the systems and the interaction with graphene (Table 3.12).

Finally, the structural analysis of the simulated complexes was completed by investigating the Hbc, the % of occupancy and its distribution (Table 3.10 and Figure 3.37b), to precisely account for the H-bonding pattern stability. Compared to **6bb2-1pyr**, **g-6bb2-1pyr** presents an increased Hbc (from 80 to 82%) and % of occupancy (from 56 to 57%), while **g-6bb2-pyr**, compared to **6bb2-pyr**, displays a stationary Hbc value (80%) but a reduced % of occupancy (from 59 to 53%). Additionally, the distribution of the donor group actually involved in H-bonds along the time appears unaltered for **g-6bb2-1pyr** (high % occupancy from 60 to 80% Figure 3.37b) but visibly affected for **g-6bb2-pyr** (% occupancy shifted from 40 to 70% Figure 3.37b). This is in good agreement with the altered SS profiles of **g-6bb2-pyr** (blue line in Figure 3.38) and the ramachandran plot analysis (Figure 3.39) suggesting that the increase of the number of the pyr functionalities does not necessarily ameliorates the adsorption performance of the β-sheet, but in our designed system affects its structural stability and consequently its efficiency in adhering to the surface. Concerning the reference systems, all the values result drastically reduced for **g-6bb2** as compared to **6bb2** (Hbc from 81 to 73% and <occ> from 58 to 41% Table 3.10 and Figure 3.37b) further confirming the loss of the SS and the poor β-sheet organization of this derivative when adsorbed onto the surface. On the contrary, even though **g-6bb2-ala** has a reduced Hbc as compared to the simulation of the β-sheet alone (from 84 to 80%), the H-bonding donors are now actually occupied for the 65% of the simulation, instead of the 59%.

In conclusion, the presence of the pyr anchoring groups is shown to favor a stable interaction of the designed β-sheet with the graphitic surface. Nevertheless, the presence of the two pyr per strand (**g-6bb2-pyr**), instead of yielding enhanced performance of adsorption onto graphene leads to a crowded situation that partly affect the structural stability of the system understudy. The β-sheet presenting one pyr per strand (**g-6bb2-1pyr**) is instead demonstrated to preserve the structural features,

fundamental for the stability and functionality of the designed interface. For these reasons, **g-6bb2-1pyr** has been chosen as the ideal proteinic β-sheet for the directional and controlled functionalization of graphene. Finally, the comparison of the pyr-modified systems with the references **g-6bb2** and **g-6bb2-ala** further proved that the presence of the anchoring groups does not induce significant affection of the β-sheet stability. It additionally probed the importance of the full hydrophobicity of the interacting side, that when modified can remarkably affect the structural stability of the designed proteinic interface.

3.2.5.4 Janus-type adsorption

In this section the directionality of the designed interfacing platform will be assessed. In order to prove the effective and preferential interaction of the proteinic β-sheet through the hydrophobic side of its bi-functional structure, the chosen **g-6bb2-1pyr** system and **g-6bb2-ala** adsorbed through the hydrophobic side as the best reference system were compared to the reverse adsorption of the latter through the hydrophilic side (**g-6bb2-ala-r**).

After 50 ns MD simulation, comparing the starting and last snapshot of the trajectory the reverse adsorbed **g-6bb2-ala-r** does not show any particular structural rearrangement (Figure 3.41). In fact, the calculated RMSD displays a constant trend along the simulated time, even more than **g-6bb2-ala** (lilac and green lines in Figure 3.41). Moreover, any losses of the SS are detected as shown in the plot in Figure 3.41. Hence, despite the reverse adsorption, **g-6bb2-ala-r** seems to behave similarly to the two systems interacting with the hydrophobic side. In line with these findings, by analyzing the backbone dihedral angles as plotted in the ramachandran plot the exact values referring to the proper β-sheet formation (top left area of the plot) are detected for the reverse adsorbed system, similarly as the other two cases (Figure 3.42).

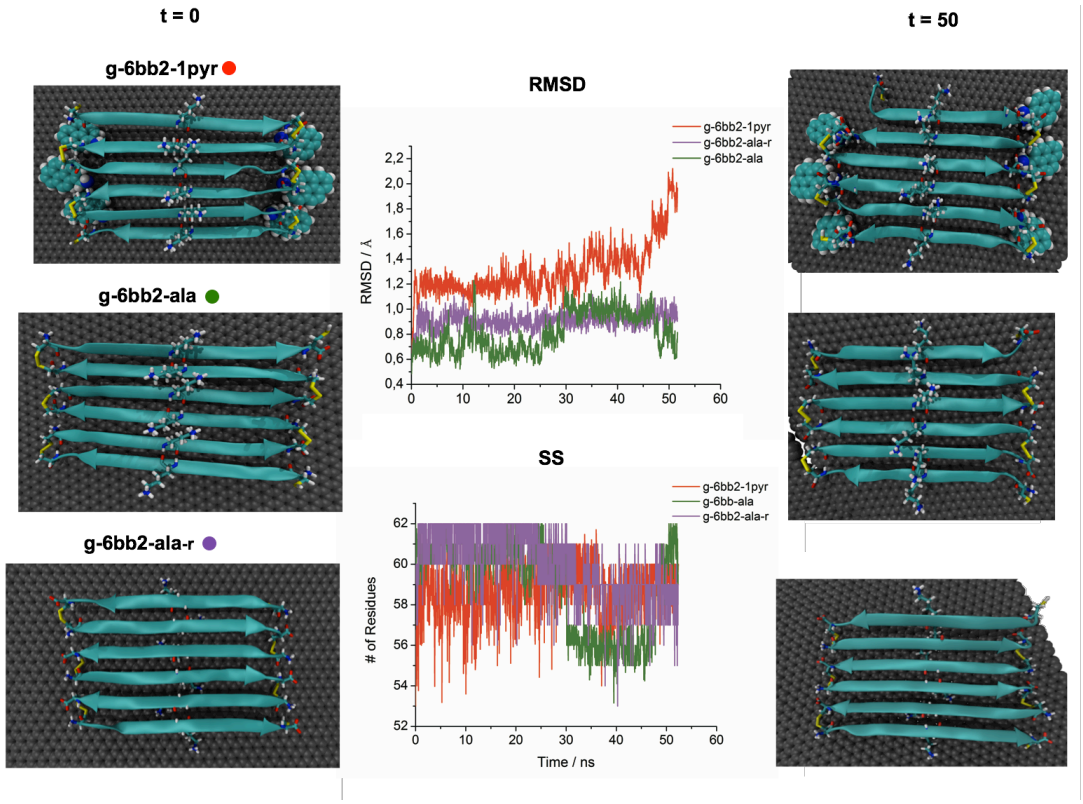


Figure 3.41. Top view of the starting structure (left panel), the RMSD (top) and the SS (bottom) (middle panel), and the last snapshot (right panel) of the 50 ns MD simulations are reported for **g-6bb2-1pyr** (red), **g-6bb2-ala** (green) and **g-6bb2-ala-r** (lilac) from the top to the bottom. The 3D representations of the β-strand are drawn as cartoon arrows whose direction evidences the antiparallel secondary structure; for clarity all the side chains are not shown but the hydrophilic face is marked by the Lys residues drawn in licorice, while the assembling points among the strands are evidenced by the lateral disulphide bonds between Cys drawn in licorice; when present, the pyrenyl functionalization is drawn in vdW.

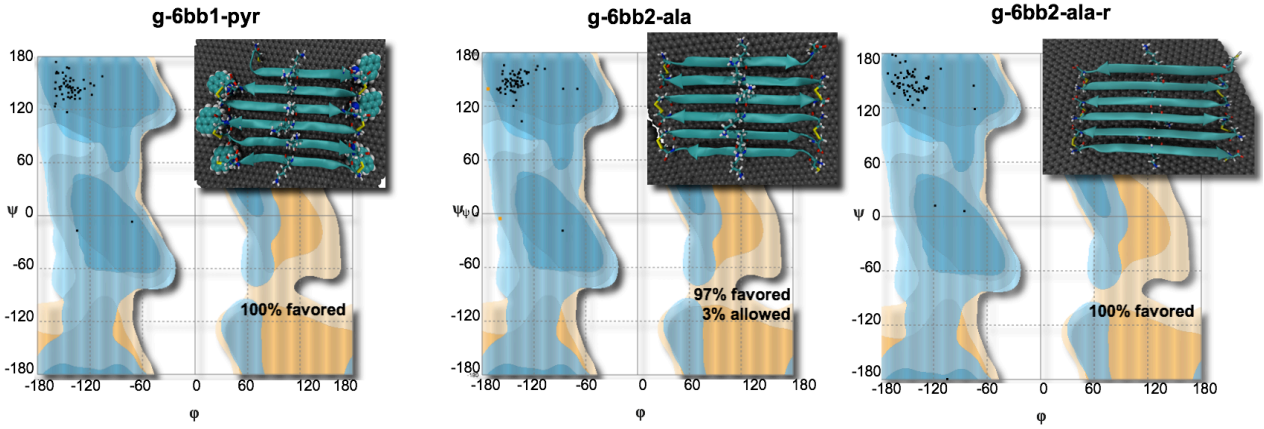


Figure 3.42. The ramachandran plots are reported with the % of residues in favored (black dots) and allowed region (orange dots) for **g-6bb2-1pyr** (left), **g-6bb2-ala** (middle) and **g-6bb2-ala-r** (right).

Proceeding with the estimation of the compactness of the designed β-sheet now adsorbed through the hydrophilic side, the Rg was also calculated along the trajectory revealing the same profile as **g-6bb2-ala** (lilac and green lines respectively in Figure 3.44a) and the identical <Rg> value of 16 Å (Figure 3.43b). However, by looking at the side view of the last snapshot of the **g-6bb2-ala-r** simulation it emerges a significant curvature of its hydrophilic side that tend to detach with the long edge of the strands from the graphene surface (Figure 3.43b).

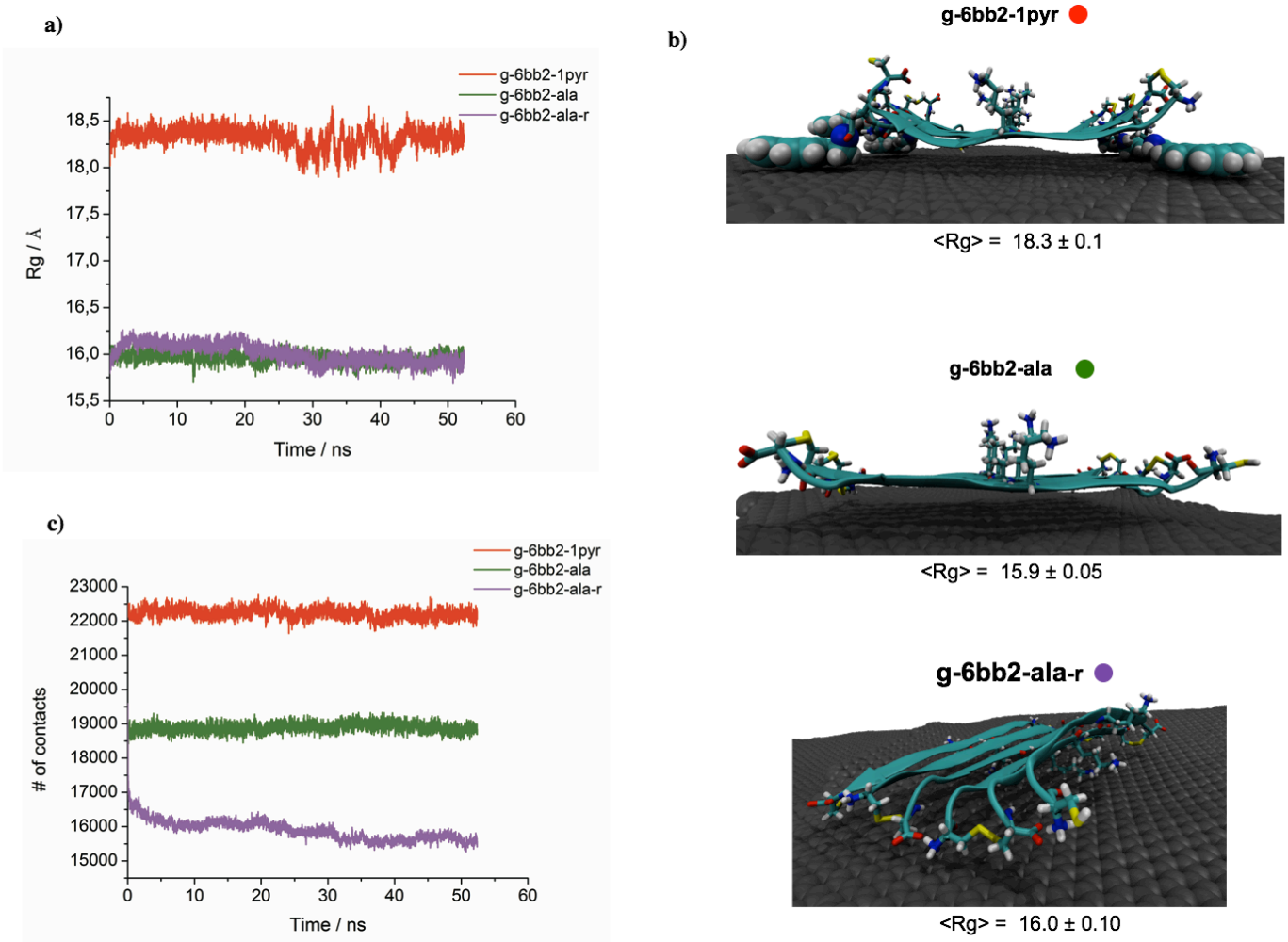


Figure 3.43. The gyration radius (Rg) is plotted as a function of time (a). The last snapshot of **g-6bb2-1pyr** (red), **g-6bb2-ala** (green) and **g-6bb2-ala-r** (lilac) is drawn as cartoon in side view to evidence the distortion of the structure when occurred; the pyrenes functionalization is drawn in vdW when present (b). The number of contacts is plotted as a function of the time for the three systems (c)

As it is clearly visible from Figure 3.43b, the Lys residues interact through the alkyl chains with graphene, pointing at the same time their amino group towards the

solvent. Something not observed in the case of the Thr residues, which bear shorter alkyl chains. This surfactant-like behavior, which resembles the Glu adsorption mode observed in the **6bb2** simulations, is also consistent with the way of interaction found in the CNT-Ctx conjugate as reported in section 3.1.5. As a consequence of this distortion, the total number of contacts that the side chains can entertain in **g-6bb2-ala-r** are drastically reduced (average value of 15814) compared to **g-6bb2-ala** (18897) and **g-6bb2-1pyr** (18452, Figure 3.43c). Moreover, to further investigate the stability of **g-6bb2-ala-r** the H-bonding analysis was carried out measuring the Hbc, the % of occupancy (Table 3.13) and the distribution of occupancy along the trajectory (Figure 3.44). Interestingly, **6bb2-ala** presents higher <%occ> when adsorbed through the hydrophobic side (65% **g-6bb2-ala**) than when adsorbed through the hydrophilic one (61% **g-6bb2-ala-r**), while the slightly lower <%occ> of **g-6bb2-1pyr** (57%) is compensated by higher or equal Hbc with respect to the other two systems (82 vs 80 and 82, Table 3.13).

Table 3.13. The Hbc and the % of occupancy of the total number of donors actually involved in H-bonds are reported for for the **g-6bb2-1pyr**, **g-6bb2-ala**, and **g-6bb2-ala-r** trajectories.

System	Hbc %	<%occ>
g-6bb2-1pyr	82	57 ± 20
g-6bb2-ala	80	65 ± 12
g-6bb2-ala-r	82	61 ± 11

Then, when it comes into the evaluation of the distribution of the occupancy of the donor actually involved in the H-bonds it is worth noting that the predominant high % of occupancy of the majority of donors both in **g-6bb2-1pyr** and **g-6bb2-ala** (in the range between the 60 to 80%, Figure 3.44) is shifted towards lower % in **g-6bb2-ala-r** (from 40 to 70%, Figure 3.44), actually suggesting a labile secondary structure of β-sheet when adsorbed through its hydrophobic face. Finally the non-bonded interaction energies, namely the EEL and vdW, were calculated along the 50 ns MD simulation and the values averaged over the last 30 ns of the simulations are reported in Table 3.14. As observed in the previous section, the adsorption is driven by the vdW energy, thereby representing the parameter of reference.

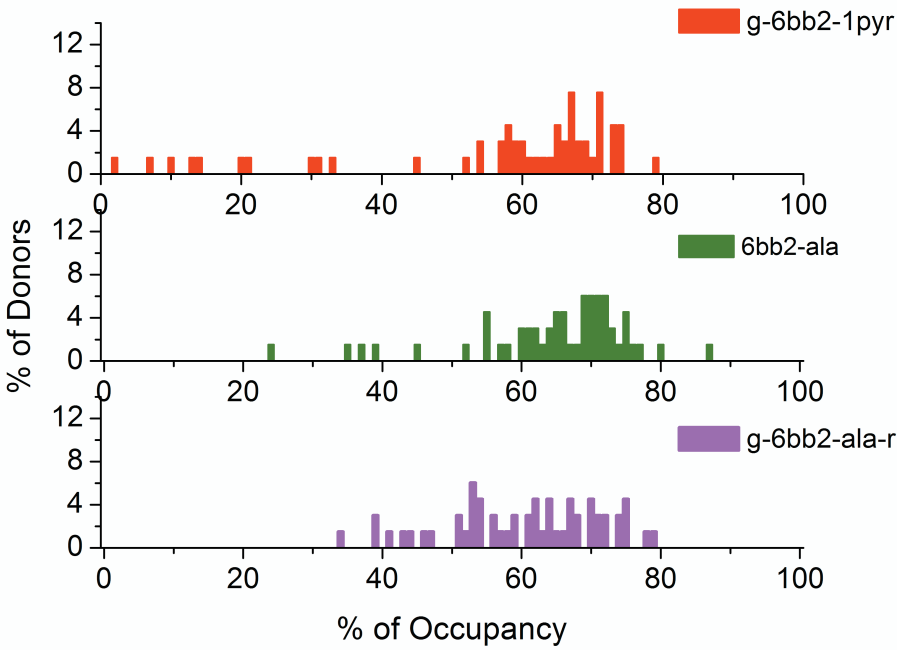


Figure 3.44. The % of donors as a function of the % of occupancy is plotted for **g-6bb2-1pyr** (red), **g-6bb2-ala** (green) and **g-6bb2-pyr** (lilac).

Notably, despite **g-6bb2-ala-r** presenting the lowest <vdW> value among the three systems (Table 3.14) it is slightly lower than **g-6bb2-ala** (-243 ± 24 and -249 ± 7 kcal mol^{-1} respectively), whereas the presence of the pyr groups significantly enhances the adsorption performance (-340 ± 9 kcal mol^{-1}). This suggests that the interaction of the two sides of the **6bb2-ala** system is completely nonspecific since both the adsorbed sides behave similarly when the β-sheet structural features are considered. On the other hand, when the pyr functionalities are inserted in the hydrophobic side, they increase the propensity of this designed β-sheet to stably interact with the flat graphitic surface, ultimately confirming the importance of the anchoring groups to achieve the directional functionalization of graphene.

Table 3.14. Non-bonded interaction (vdW and EEL) energies averaged over the last 30 ns of the simulations are reported for the **g-6bb2-1pyr**, **g-6bb2-ala**, and **g-6bb2-ala-r** trajectories.

System	<vdW>[a]	<EEL>[a]
g-6bb2-1pyr	-340 ± 9	-2.1 ± 1.7
g-6bb2-ala	-249 ± 7	-0.57 ± 1.6
g-6bb2-ala-r	-243 ± 24	-4.9 ± 2.9

[a] kcal mol^{-1}

3.2.5.5 2D vs 1D: comparison with CNT

In the previous sections we have presented a newly designed 2D interface as a directional and bi-functional approach for the graphene functionalization. Nevertheless, to fully explore and evaluate the potentialities of this functionalization approach we simulated the chosen **g-6bb2-1pyr** β-sheet also on the 1D structure of CNT performing 50 ns MD simulations. We therefore modeled the behavior of **6bb2-1pyr** on the 20 Å CNT, as parameterized in section 3.2. Two main orientations between the CNT and the pyr-modified β-sheet were envisaged placing **6bb2-1pyr** parallel (**cnt-6bb2-II**) or perpendicular (**cnt-6bb2-⊥**) to the long axe of the tubular framework (Figure 3.45).

Interestingly, regardless of the starting β-sheet orientations (**II** or ⊥) the **6bb2-1pyr** tends to wrap around the CNT yielding in both **cnt-6bb2-II** and **cnt-6bb2-⊥** the same final adsorption mode (Figure 3.45). According to this finding, at the first ns of the trajectories (at 2-3 ns) a sharp peak in the RMSD is observed (grey and violet lines in Figure 3.45) corresponding to the structural rearrangement of the β-sheet around the CNT. After the systems reach the equilibrium during the last simulated time (Figure 3.45). Although pronounced, this structural rearrangement just slightly alters the β-sheet SS in **cnt-6bb2-II** (55 ± 1 residue in β-sheet) and in **cnt-6bb2-⊥** (57 ± 2) as compared to **g-6bb2-1pyr** (59 ± 1, Figure 3.45). These findings are also confirmed by the ramchandran plot analysis that displays the 97% of residues in allowed regions for **cnt-6bb2-II** and the 98.5% for **cnt-6bb2-⊥** slightly diminished with respect to the 100 % of residues in allowed regions presented by **g-6bb2-1pyr** (Figure 3.46). The tendency of **6bb2-1pyr** to wrap around the tubular framework was also observed in the calculated Rg that indeed decreases along the 50 ns for both the β-sheet orientation adsorbed on the CNT (Figure 3.47a) resulting 1 Å smaller than the graphene adsorbed system **g-6bb2-1pyr** (Figure 3.47b). Hence this proteinic β-sheet results dynamic and versatile enough to adapt its conformation to a tubular framework, maintaining at the same time a constant number of contacts among the side chains, which might help the stabilization of the systems (Figure 3.47c).[100]

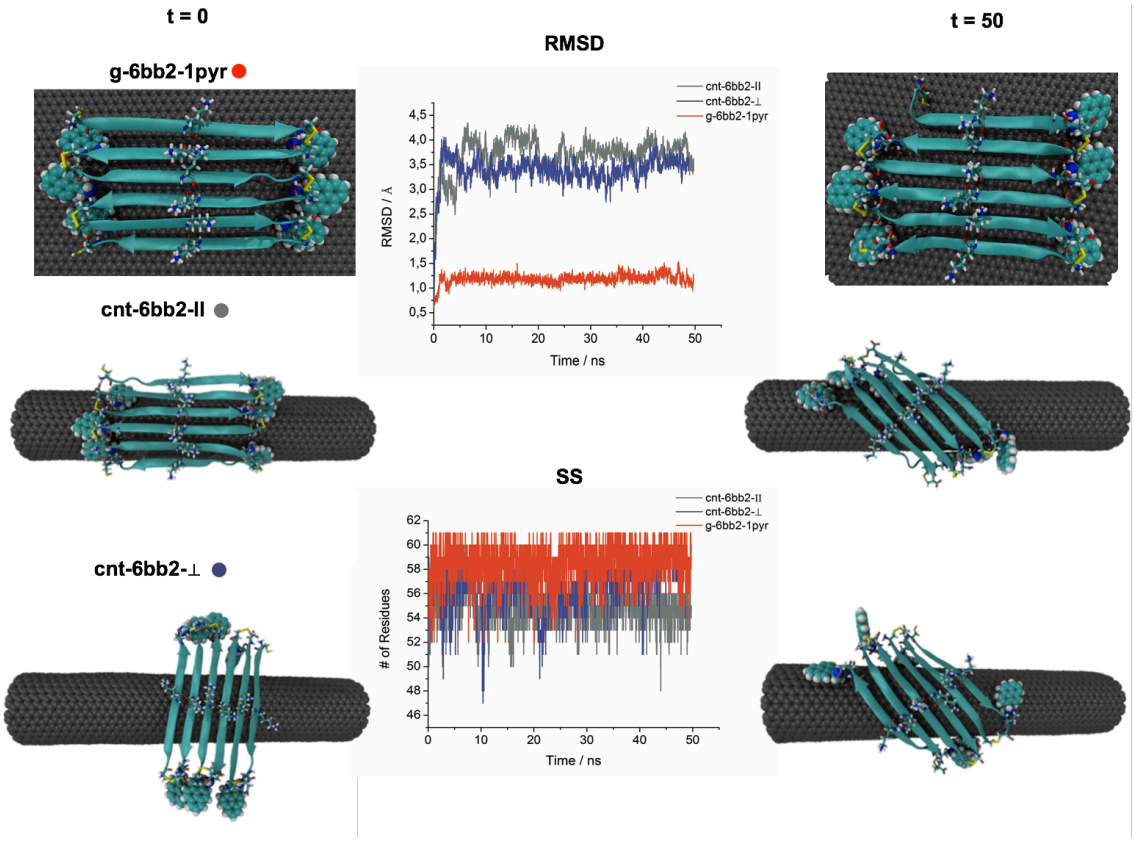


Figure 3.45. Top view of the starting structure (left panel), the RMSD (top) and the SS (bottom) (middle panel), and the last snapshot (right panel) of the 50 ns MD simulations are reported for **g-6bb2-1pyr** (red), **cnt-6bb2-II** (grey) and **cnt-6bb2-⊥** (violet) from the top to the bottom. The 3D representations of the β-strand are drawn as cartoon arrows whose direction evidences the antiparallel secondary structure; for clarity all the side chains are not shown but the hydrophilic face is marked by the Lys residues drawn in licorice, while the assembling points among the strands are evidenced by the lateral disulphide bonds between Cys drawn in licorice; when present, the pyrenyl functionalization is drawn in vdW

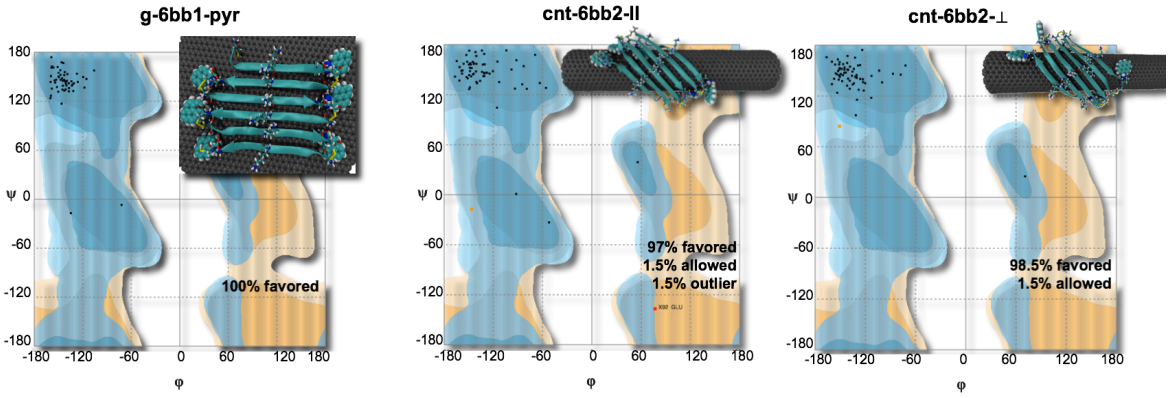


Figure 3.46. The ramachandran plots are reported with the % of residues in favored (black dots), allowed region (orange dots) and outlier (red dots) for **g-6bb2-1pyr** (left), **cnt-6bb2-II** (middle) and **g-6bb2-⊥** (right).

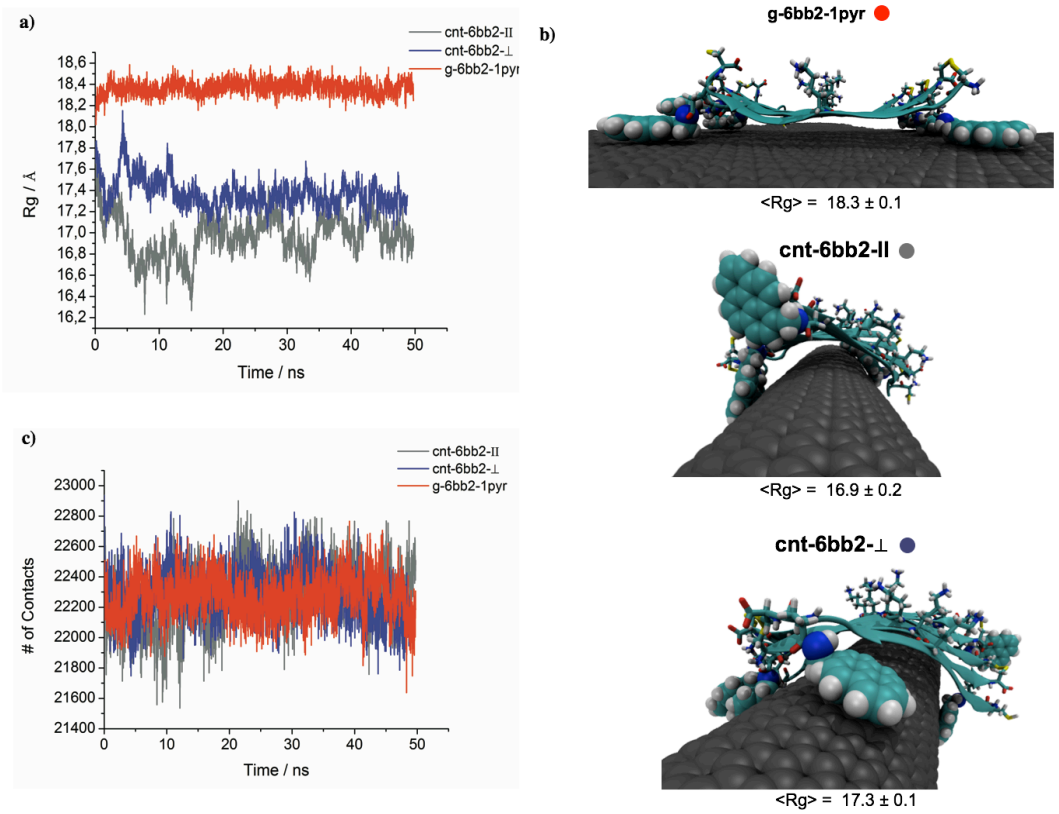


Figure 3.47. The gyration radius (Rg) is plotted as a function of time (a). The last snapshot of **g-6bb2-1pyr** (red), **cnt-6bb2-II** (grey) and **g-6bb2-⊥** (violet) is drawn as cartoon in side view to evidence the distortion of the structure when occurred; the pyrenes functionalization is drawn in vdW when present (b). The number of contacts is plotted as a function of the time for the three systems (c)

Evaluating further the stability of the **6bb2-1pyr** adsorbed on CNT, the Hbc, the % of occupancy (Table 3.15) and the distribution of the occupied H-bonding donors along the trajectories were also calculated (Figure 3.48). From this analysis it emerges that both the orientations adsorbed on CNT, resembling the **g-6bb2-1pyr** system, present the 82% of donors actually involved in H-bonds, which are occupied for the 54% and 59% of the simulated time for **cnt-6bb2-II** and **cnt-6bb2-⊥,** respectively (Table 3.15).

Table 3.15. The Hbc and the % of occupancy of the total number of donors actually involved in H-bonds are reported for **g-6bb2-1pyr**, **cnt-6bb2-II** and **cnt-6bb2-⊥.**

System	Hbc %	<%occ>
g-6bb2-1pyr	82	57 ± 20
cnt-6bb2-II	82	54 ± 23
cnt-6bb2-⊥	83	59 ± 20

Additionally, in all the three cases the major % of donors occupied in H-bonds present an equal distribution with mostly high % of occupancy along the time (from 60 to 80%, Figure 3.48).

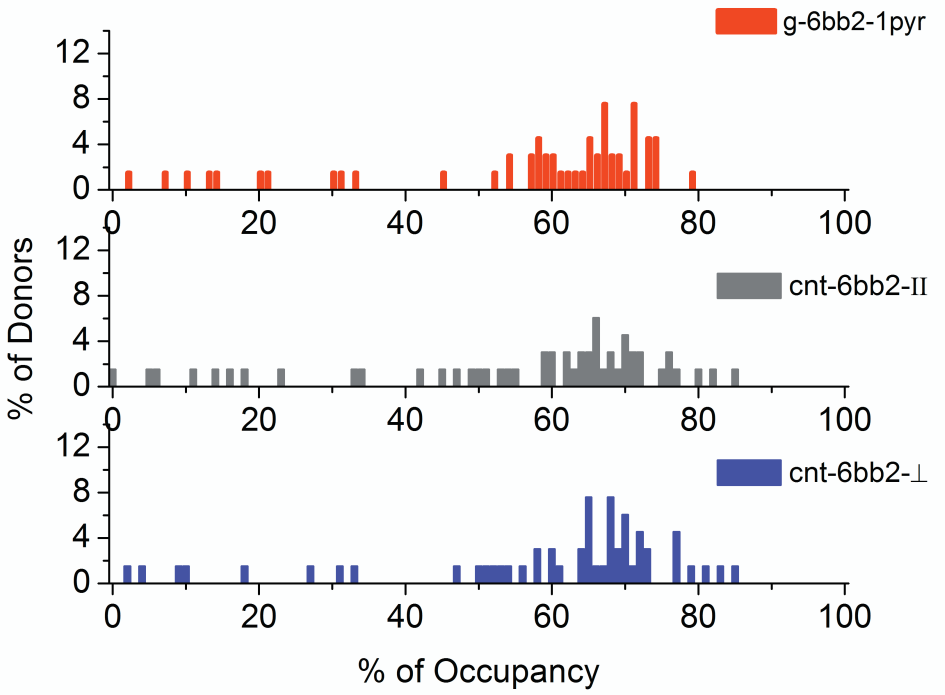


Figure 3.48. The % of donors as a function of the % of occupancy is plotted for **g-6bb2-1pyr**, **cnt-6bb2-II** (grey) and **g-6bb2-⊥** (violet).

Hence from a structural point of view, the CNT-adsorbed β-sheet visibly undergoes conformational changes in order to maximize the contact surface with the tubular framework. Despite the visible structural rearrangements the secondary structure integrity is preserved as observed in Figure 3.45, 3.48 and Table 3.15. Only small distortions of the backbone dihedral angles are found in the ramachandran plot (Figure 3.46) that, ascribable to the induced curvature of the β-sheet to adapt to the CNT structure, can be neglected. However, analyzing the non-bonded interaction energies EEL and vdW, calculated between the β-sheet and the CNT, **cnt-6bb2-II** and **cnt-6bb2-⊥** are less strongly adsorbed on CNT than on graphene (<vdW> = -235, -227 and -340 kcal mol^{-1} respectively, Table 3.16), even less than the reference systems without the pyr functionalization (<vdW> = -249 and -242 kcal mol^{-1} for **g-6bb2-ala** and **g-6bb2-ala-r**, respectively). Consequently, even though the structural integrity of the interfacing platform is preserved upon adsorption on CNT, the

conformation adopted yields the lowest interaction energy found in this study. Importantly, the energetic analysis performed in this section carries the penalty of the different parameterization of CNT (AM1-bcc charges) and graphene (RESP charges) that might affect the energetic evaluation. Therefore we consider these results highly qualitative and we did not perform any further energetic analysis. Additionally, it must be pointed out that as explorative investigation we sample the behavior of the anchoring β-sheet only on a 20 Å CNT, prompting a possible diameter-dependent effect in the adsorbing performance of this interfacing platform. Nevertheless, the observed structural stability also on CNT represents a compelling and promising starting point for further theoretical and experimental studies aimed at exploring the versatility of this newly designed targeting platform.

Table 3.16. Non-bonded interaction (vdW and EEL) energies averaged over the last 30 ns of the simulations are reported for the **g-6bb2-1pyr**, , **cnt-6bb2-II** and **g-6bb2-⊥** trajectories.

System	<vdW>[a]	<EEL>[a]
g-6bb2-1pyr	-340 ± 9	-2.1 ± 1.7
cnt-6bb2-II	-235 ± 23	-0.9 ± 1.4
g-6bb2-⊥	-227 ± 27	-0.3 ± 1.1

[a] kcal mol^{-1}

3.2.6 Concluding remarks and perspectives

In summary, we have been able to design a new proteinic β-sheet able to directionally functionalize graphene in the view of biological applications. In particular we provide computational evidences that a 13-mer polypeptide can fold stably in an antiparallel β-sheet exposing a hydrophobic side, prone to the anchoring to graphene, and a hydrophilic face devoted to functionalization purposes. Additionally, pyrenyl functionalities have been inserted in the hydrophobic side and their presence was demonstrated to be fundamental to improve the β-sheet adsorption on graphene. The ideal number of pyr functionalities was found to be one per strand, selecting hence the **6bb2-1pyr** system as the best designed β-sheet for the purpose we have presented herein. In collaboration with Dr. Riccardo Marega and Laure-Elie Carloni of our group, we are now validating experimentally the theoretical design by producing the β-sheet protein and testing it onto graphene.

3.3 References

[1] "Globocan 2012," can be found under http://globocan.iarc.fr/, **2012**.

[2] K. K. Jain, *Technol. Cancer Res. Treat.* **2005**, *4*, 407–416.

[3] K. K. Jain, *Drug Discov. Today* **2005**, *10*, 1435–1442.

[4] V. Sanna, N. Pala, M. Sechi, *Int. J. Nanomed.* **2014**, *9*, 467–483.

[5] Y. Rupali Prashant, C. Anuradha Nirangan, M. Chandrakant Shripal, M. Shrinivas Krishna, *Int. J. Pharm. Sci. Rev. Res.* **2013**, *20*, 210–217.

[6] K. Kostarelos, A. Bianco, M. Prato, *Nat. Nanotechnol.* **2009**, *4*, 627–633.

[7] S. L. Edwards, J. A. Werkmeister, J. A. M. Ramshaw, *Expert Rev. Med. Dev.* **2009**, *6*, 499–505.

[8] G. Pastorin, W. Wu, S. Wieckowski, J.-P. Briand, K. Kostarelos, M. Prato, A. Bianco, *Chem. Commun.* **2006**, 1182–1184.

[9] W. Wu, S. Wieckowski, G. Pastorin, M. Benincasa, C. Klumpp, J.-P. Briand, R. Gennaro, M. Prato, A. Bianco, *Angew. Chem. Int. Ed.* **2005**, *44*, 6358–6362.

[10] C. Ménard-Moyon, E. Venturelli, C. Fabbro, C. Samorì, T. Da Ros, K. Kostarelos, M. Prato, A. Bianco, *Expert Opini. Drug Discov.* **2010**, *5*, 691–707.

[11] C. Samorì, H. Ali-Boucetta, R. Sainz, C. Guo, F. M. Toma, C. Fabbro, T. Da Ros, M. Prato, K. Kostarelos, A. Bianco, *Chem. Commun.* **2010**, *46*, 1494–1496.

[12] C. Gaillard, G. Cellot, S. Li, F. M. Toma, H. Dumortier, G. Spalluto, B. Cacciari, M. Prato, L. Ballerini, A. Bianco, *Adv. Mater.* **2009**, *21*, 2903–2908.

[13] Z. Liu, W. Cai, L. He, N. Nakayama, K. Chen, X. Sun, X. Chen, H. Dai, *Nat. Nanotechnol.* **2007**, *2*, 47–52.

[14] A. A. Bhirde, V. Patel, J. Gavard, G. Zhang, A. A. Sousa, A. Masedunskas, R. D. Leapman, R. Weigert, J. S. Gutkind, J. F. Rusling, *ACS Nano* **2009**, *3*, 307–316.

[15] A. Bianco, K. Kostarelos, M. Prato, *Curr. Opin. Chem. Biol.* **2005**, *9*, 674–679.

[16] X. Ding, R. Singh, A. Burke, H. Hatcher, J. Olson, R. A. Kraft, M. Schmid, D. Carroll, J. D. Bourland, S. Akman, F. M. Torti, S. V. Torti, *Nanomedicine* **2011**, *6*, 1341–1352.

[17] C. J. Gannon, P. Cherukuri, B. I. Yakobson, L. Cognet, J. S. Kanzius, C. Kittrell, R. B. Weisman, M. Pasquali, H. K. Schmidt, R. E. Smalley, S. A. Curley, *Cancer* **2007**, *110*, 2654–2665.

[18] R. Marega, D. Bonifazi, *New J. Chem.* **2014**, *38*, 22–27.

[19] L. B. Saltz, N. J. Meropol, P. J. Loehrer, M. N. Needle, J. Kopit, R. J. Mayer, *J. Clin. Oncol.* **2004**, *22*, 1201–1208.

[20] K. D. Elgert, *Immunology: Understanding The Immune System*, Wiley & Sons, Inc, United State of America, USA, **1998**.

[21] N. A. P. S. Buss, S. J. Henderson, M. McFarlane, J. M. Shenton, L. de Haan, *Curr. Opin. Pharmacol.* **2012**, *12*, 615–622.

[22] "2004 - FDA Approves Erbitux for Colorectal Cancer," can be found under http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2004/ucm108244.htm, **2004**.

[23] D. Cunningham, Y. Humblet, S. Siena, D. Khayat, H. Bleiberg, A. Santoro, D. Bets, M. Mueser, A. Harstrick, C. Verslype, I. Chau, E. Van Cutsem, *N. Engl. J. Med.* **2004**, *351*, 337–345.

[24] B. R. Voldborg, L. Damstrup, M. Spang-Thomsen, H. S. Poulsen, *Ann. Onc.* **1997**, *8*, 1197–1206.

[25] J. B. Johnston, S. Navaratnam, M. W. Pitz, J. M. Maniate, E. Wiechec, H. Baust, J. Gingerich, G. P. Skliris, L. C. Murphy, M. Los, *Curr. Med. Chem.* **2006**, *13*, 3483–3492.

[26] T. M. Brand, M. Iida, D. L. Wheeler, *Cancer Biol. Ther.* **2011**, *11*, 777–792.

[27] R. I. Nicholson, J. M. Gee, M. E. Harper, *Eur. J. Cancer* **2001**, *37 Suppl 4*, 9–15.

[28] S. Li, K. R. Schmitz, P. D. Jeffrey, J. J. W. Wiltzius, P. Kussie, K. M. Ferguson, *Cancer Cell* **2005**, *7*, 301–311 .

[29] D. L. Wheeler, S. Huang, T. J. Kruser, M. M. Nechrebecki, E. A. Armstrong, S. Benavente, V. Gondi, K.-T. Hsu, P. M. Harari, *Oncogene* **2008**, *27*, 3944–3956.

[30] P. J. Carter, *Exp. Cell Res.* **2011**, *317*, 1261–1269.

[31] L. J. Harris, S. B. Larson, K. W. Hasel, J. Day, A. Greenwood, A. McPherson, *Nature* **1992**, *360*, 369–372.

[32] L. J. Harris, S. B. Larson, K. W. Hasel, A. McPherson, *Biochemistry* **1997**, *36*, 1581–1597.

[33] L. J. Harris, E. Skaletsky, A. McPherson, *J. Mol. Bio.* **1998**, *275*, 861–872.

[34] E. O. Saphire, P. W. Parren, R. Pantophlet, M. B. Zwick, G. M. Morris, P. M. Rudd, R. A. Dwek, R. L. Stanfield, D. R. Burton, I. A. Wilson, *Science (80-.).* **2001**, *293*, 1155–1159.

[35] L. W. Guddat, J. N. Herron, A. B. Edmundson, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4271–4275.

[36] E. Mix, R. Goertsches, U. K. Zett, *J. Neurol.* **2006**, *253 Suppl* , 9–17.

[37] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712–725.

[38] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.

[39] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[40] S. Iijima, *Nature* **1991**, *354*, 56–58.

[41] A. Maiti, C. Brabec, C. Roland, J. Bernholc, *Phys. Rev. B* **1995**, *52*, 14850–14858.

[42] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.

[43] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2000**, *21*, 132–146.

[44] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

[45] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[46] O. A. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, *J.Comput. Chem.* **1998**, 1639–1662.

[47] V. S. Lee, P. Nimmanpipug, O. Aruksakunwong, S. Promsri, P. Sompornpisut, S. Hannongbua, *J. Mol. Graph. Model.* **2007**, *26*, 558–570.

[48] S.-T. Yang, H. Wang, L. Guo, Y. Gao, Y. Liu, A. Cao, *Nanotechnology* **2008**, *19*, 395101–39518.

[49] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J.

Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, AMBER11, University of California, San Francisco **2010**.

[50] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[51] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[52] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577–2637.

[53] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.

[54] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2007**, *120*, 215–241.

[55] J. Sgrignani, D. Franco, A. Magistrato, *Molecules* **2011**, *16*, 442–465.

[56] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553–566.

[57] S. Simon, M. Duran, J. J. Dannenberg, *J. Chem. Phys.* **1996**, *105*, 11024–11031.

[58] D. J. F. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had, **2009**.

[59] W. Tang, E. Sanville, G. Henkelman, *J. Phys.-Condens. Matter* **2009**, *21*, 084204–084211.

[60] J. P. Brandt, T. W. Patapoff, S. R. Aragon, *Biophys. J.* **2010**, *99*, 905–913.

[61] X. Wang, S. Kumar, S. K. Singh, *Pharm. Res.* **2011**, *28*, 3128–3144.

[62] T. Panczyk, W. Rudzinski, A. Jagusiak, *Colloids Surf. A* **2012**, *409*, 149–158.

[63] M. Calvaresi, S. Hoefinger, F. Zerbetto, *Chem. - A Eur. J.* **2012**, *18*, 4308–4313.

[64] G. Hummer, J. C. Rasaiah, J. P. Noworyta, *Nature* **2001**, *414*, 188–190.

[65] A. de Leon, A. F. Jalbout, V. A. Basiuk, *Chem. Phys. Lett.* **2008**, *457*, 185–190.

[66] A. de Leon, A. F. Jalbout, V. A. Basiuk, *Int. J. Quantum. Chem.* **2009**, *110*, 953–959.

[67] M. Calvaresi, F. Zerbetto, *ACS Nano* **2010**, *4*, 2283–2299.

[68] M. Calvaresi, F. Zerbetto, *Nanoscale* **2011**, *3*, 2873–2881.

[69] D. Nepal, K. E. Geckeler, *Small* **2007**, *3*, 1259–1265.

[70] D. Nepal, K. E. Geckeler, *Small* **2006**, *2*, 406–412.

[71] G. Zuo, W. Gu, H. Fang, R. Zhou, *J. Phys. Chem. C* **2011**, *115*, 12322–12328.

[72] G. Zuo, Q. Huang, G. Wei, R. Zhou, H. Fang, *ACS Nano* **2010**, *4*, 7508–7514.

[73] G. Zuo, X. Zhou, Q. Huang, H. Fang, R. Zhou, *J. Phys. Chem. C* **2011**, *115*, 23323–23328.

[74] G. Zuo, S.-G. Kang, P. Xiu, Y. Zhao, R. Zhou, *Small* **2012**, *9*, 1546–1556.

[75] K. S. Novoselov, A. K. Geim, S. V Morozov, D. Jiang, Y. Zhang, S. V Dubonos, I. V Grigorieva, A. A. Firsov, *Science (80-.).* **2004**, *306*, 666–669.

[76] D. Bitounis, H. Ali-Boucetta, B. H. Hong, D.-H. Min, K. Kostarelos, *Adv. Mater.* **2013**, *25*, 2258–2268.

[77] V. C. Sanchez, A. Jachak, R. H. Hurt, A. B. Kane, *Chem. Res. Toxicol.* **2012**, *25*, 15–34 .

[78] H. Y. Mao, S. Laurent, W. Chen, O. Akhavan, M. Imani, A. A. Ashkarran, M. Mahmoudi, *Chem. Rev.* **2013**, *113*, 3407–3424.

[79] Z. M. Markovic, L. M. Harhaji-Trajkovic, B. M. Todorovic-Markovic, D. P. Kepić, K. M. Arsikin, S. P. Jovanović, A. C. Pantovic, M. D. Dramićanin, V. S. Trajkovic, *Biomaterials* **2011**, *32*, 1121–1129.

[80] S. Shi, F. Chen, E. B. Ehlerding, W. Cai, *Bioconjug. Chem.* **2014**, DOI 10.1021/bc500332c.

[81] K. Yang, J. Wan, S. Zhang, B. Tian, Y. Zhang, Z. Liu, *Biomaterials* **2012**, *33*, 2206–2214.

[82] K. Yang, S. Zhang, G. Zhang, X. Sun, S.-T. Lee, Z. Liu, *Nano Lett.* **2010**, *10*, 3318–3323.

[83] H.-D. J. Norbert Sewald, *Peptides: Chemistry and Biology*, Wiley-VCH Verlag GmbH & Co. KGaA, **2009**.

[84] C. L. Nesloney, J. W. Kelly, *Bioorg. Med. Chem.* **1996**, *4*, 739–766.

[85] L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 729–740.

[86] K. Yaakobi, Y. Liebes-Peer, A. Kushmaro, H. Rapaport, *Langmuir* **2013**, *29*, 6840–6848.

[87] C. L. Brown, I. A. Aksay, D. A. Saville, M. H. Hecht, *J. Am. Chem. Soc.* **2002**, *124*, 6846–6848.

[88] J. Rodríguez-López, N. L. Ritzert, J. A. Mann, C. Tan, W. R. Dichtel, H. D. Abruña, *J. Am. Chem. Soc.* **2012**, *134*, 6224–6236 .

[89] J. A. Mann, W. R. Dichtel, *ACS Nano* **2013**, *7*, 7193–7199 .

[90] J. A. Mann, T. Alava, H. G. Craighead, W. R. Dichtel, *Angew. Chem. Int. Ed.* **2013**, *52*, 3177–3180.

[91] S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders, J. F. Stoddart, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 898–952.

[92] E.-K. Bang, M. Lista, G. Sforazzini, N. Sakai, S. Matile, *Chem. Sci.* **2012**, *3*, 1752–1763.

[93] A. Wilson, G. Gasparini, S. Matile, *Chem. Soc. Rev.* **2014**, *43*, 1948–1962.

[94] F. Freire, A. M. Almeida, J. D. Fisk, J. D. Steinkruger, S. H. Gellman, *Angew. Chem. Int. Ed.* **2011**, *50*, 8735–8738.

[95] H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Giriat, T. Muir, S. H. Gellman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12015–12020.

[96] G. Bratzel, M. J. Buehler, *J. Mech. Behav. Biomed. Mater.* **2012**, *7*, 30–40.

[97] A. W. Fitzpatrick, T. P. J. Knowles, C. A. Waudby, M. Vendruscolo, C. M. Dobson, *PLoS Comput. Biol.* **2011**, *7*, e1002169.

[98] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333–66.

[99] Z. Su, C. L. Dias, *J. Phys. Chem. B* **2014**, DOI 10.1021/jp504798s.

[100] G. D. Rose, P. J. Fleming, J. R. Banavar, A. Maritan, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16623–16633.

[101] F. A. Syud, H. E. Stanger, H. S. Mortell, J. F. Espinosa, J. D. Fisk, C. G. Fry, S. H. Gellman, *J. Mol. Biol.* **2003**, *326*, 553–568.

[102] S. R. Griffiths-Jones, M. S. Searle, *J. Am. Chem. Soc.* **2000**, *122*, 8350–8356.

[103] T.-B. Yu, J. Z. Bai, Z. Guan, *Angew. Chem. Int. Ed.* **2009**, *48*, 1097–1101.

[104] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, AMBER12, University of California, San Francisco **2012**.

[105] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* **1963**, *7*, 95–99.

[106] S. C. Lovell, I. W. Davis, W. B. Arendall, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, D. C. Richardson, *Proteins* **2003**, *50*, 437–450.

[107] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, *Acc. Chem. Res.* **2000**, *33*, 889–897.

4 Structural validation of peptide-based motogenic Au surfaces by means of MD simulations

The following chapter will concentrate on the computational studies carried out to give structural insights into the design of peptide-based self-assembled monolayer (SAM) on gold (Au) surface used for cellular migration purposes.

The chapter is divided in four main parts: section 4.1 and 4.2 includes a general introduction concerning the aim of the project, section 4.3 discusses the experimental evidences achieved in our group in designing motogenic Au surfaces through SAM of thiolated IGDQK peptide; section 4.4 describes the computational goals aimed at comprehend the structural differences observed experimentally; section 4.5 focuses on the proposed computational methods used to unravel the dynamical properties of the peptide-based SAM, and finally section 4.6 goes through the results and discussions.

The experimental part of the project was carried out by *Dr. Valentina Corvaglia* and *Dr. Riccardo Marega* of the *University of Namur (UNamur)* at the *Namur Research College (NARC)*, Namur (Belgium), in collaboration with *Prof. Carine Micheils* of the *University of Namur (UNamur)* who synthesized, produced and biologically tested the IGDQK-SH peptide SAM on Au surfaces. We acknowledge *Prof. Alessandro De Vita*, *Dr. Massimo Riello* and *Giovanni Doni* for the fruitful discussions on the computational developments during my two-week stage at *King's College*, London (UK).

4.1 Introduction

In the wide world of bionanotechnology, the combination of biomolecules and surfaces goes beyond the use of GNMs, as discussed in detail in Chapter 1. In fact, the use of inorganic surfaces, including metals (such as gold Au, silver Ag, aluminum Al) and many metal oxide species has been extensively explored for material science, nanomedicine, biology, sensing and electronics.[1–6] Among these surfaces, the interaction of proteins with Au proves to be particularly fascinating, due to the

biocompatibility and stability of this metal, as it is inert to corrosion and oxidation. Au nanoparticles are a potential candidate for biomedical applications,[7] such as in cancer treatment,[8] imaging[9] and drug delivery,[10] while Au surfaces coupled with peptides and proteins have applications in material science such as in optical[11] and biosensing devices.[12,13] In this context, it is of paramount importance to understand protein-surface interactions (adsorption, chemisorption, orientation) with the aim of engineering, in a controlled and tunable manner, new hybrid materials for nanotechnological and biomedical purposes.[14–18]

Focusing on Au, such an understanding has been probed by both experimental and computational tools, for which efforts have provided significant insights into the study of the non-covalent adsorption of peptides and protein onto the metal substrate. Regarding structural aspects of the surface, between all the possible crystal lattices that Au might adopt, the close-packed (111) facet is the most thermodynamically stable. In this respect, Heinz *et al.* carried out a computational study that compared the adsorptions of different short peptides derived by phage–display technique on both (111) and (100) even Au surfaces.[19] Simulations demonstrate that all peptides currently being studied preferably adsorb more tightly on the (111) facet (adsorption energy of -63 kcal mol^{-1}) compared to the (100) Au substrate (0 kcal mol^{-1}). Additionally, the residues involved in binding were found to be in direct contact with the specific site of the surface, as their structural features can act as an approximate prolongation of the fcc lattice. In fact, within the highly dense Au (111) surface, the hexagonal lattice with projected interatomic distance of ~1.6 Å and angles of 120° (as shown in Figure 4.1a) can be distinguished.[19,20] Consequently, chemical groups with similar bond lengths and angles, such as aromatic residues and sp^2-hybridized structures (the CO_2^- of Asp), fit well on top of the Au (111) arrangement.

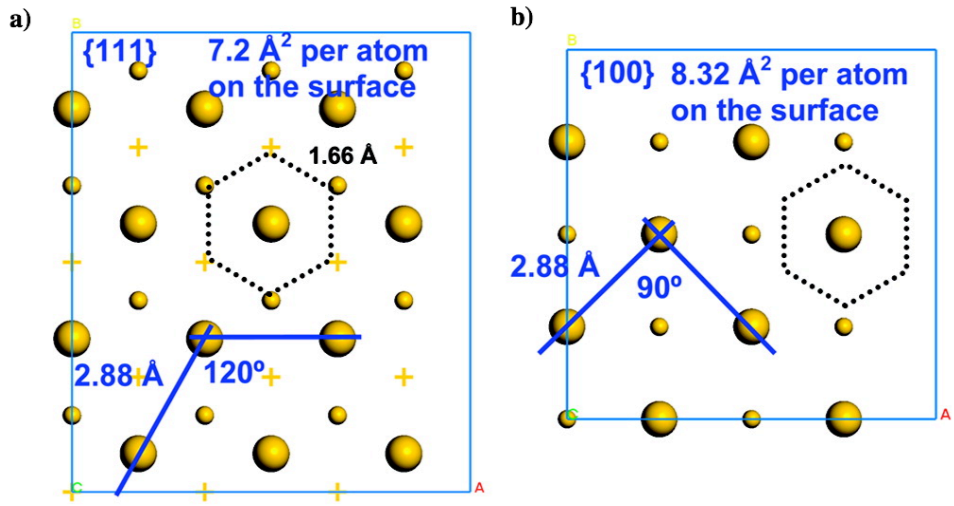


Figure 4.1. Schematic representation of the geometry of Au (111) (a) and (100) (b) crystal lattices, showing the first atomic layer with yellow spheres and the underlying layers with smaller spheres and crosses. Dotted black lines represent favorable orientations that aromatic or sp^2-configured groups may adopt on the Au (111) surface. Reprinted with permission from Ref. 19.

On the other hand, this complementary behavior cannot occur in the Au (100) facet, which is characterized by lower atomic density, higher interatomic distance (2.88 Å) and smaller angles (90°, Figure 4.1b). This arrangement creates "empty" fcc lattice sites that are more easily filled with solvent molecules rather than with the peptide sequence. In line with these findings, other theoretical[20,21] and experimental[19,22] works were reported to confirm that peptide adsorption occurs preferentially on the Au (111) rather than (100) facet.

Regarding the interacting biomolecules, computational methods were largely used to unravel Au-protein interactions, including all structural levels of complexity, from amino acids (aa) to polypeptides and globular proteins. These provided an in-depth comprehension of their intrinsic propensity and modality of spontaneous interaction. Hoefling *et al.*[23] systematically studied the adsorption of the 20 natural aa on Au (111) by performing classical molecular dynamics (MD) simulations, modeled with a polarizable force-field. Consistent with the binding mechanism proposed by Heinz *et al.*,[19] they found that aromatic aa were the strongest binders (absolute value of interacting free energy ranging from 44 to 34 kj mol^{-1}), followed by sulfur-containing aa (39 to 38 kj mol^{-1}) and positively-charged aa (36 to 30 kj mol^{-1}) for which the geometry fitted the Au crystal lattice better. In increasing order, also polar aa (29 to 23 kj mol^{-1}) and aliphatic ≈ negatively charged aa (26 to 17 kj mol^{-1})

bound with a certain affinity to the Au substrate. In parallel, the authors complemented the study with a combination of DFT and MD simulations, and reported a general tendency of the interacting aa to lay on the surface, thereby maximizing surface contact area through backbone and side chains interactions.[24] Additionally, other theoretical calculations confirmed the same trend in affinity exerted by the adsorbed aa on Au (111), as reported by Hoefling *et al.* (aromatic > S-containing > positively charged > polar > aliphatic ≈ negatively charged aa).[25,26] Prediction of the adsorption process of aa and polypeptides on Au has triggered further interest by researchers; by means of classical MD simulations, the three main steps of *diffusion*, *association* and *binding* (Figure 4.2a)[20,27–29] have been described.

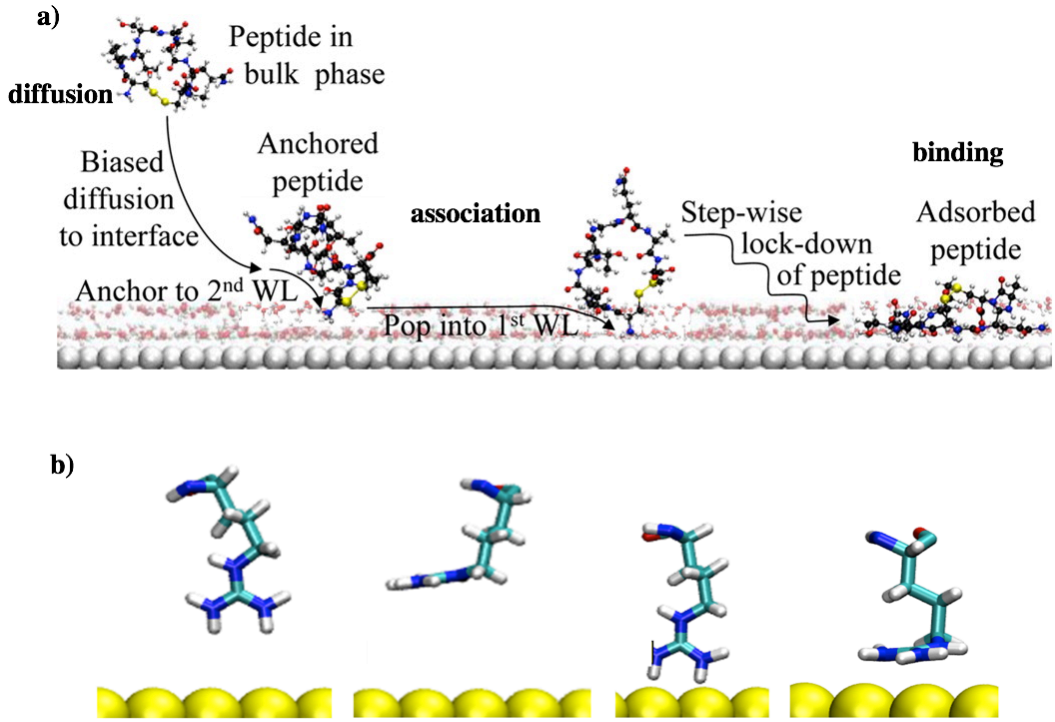


Figure 4.2. Generic peptide adsorption mechanism consisting of three main steps: *diffusion*, *association* and *binding* (a). Adapted with permission from Ref. 29. Snapshots of the simulated adsorption process of the Arg residue on Au (111), which adopts different conformations, ultimately acting as an anchoring group by breaking the first water layer (water not shown for clarity) (b). Adapted with permission from Ref. 28.

Specifically, in the *diffusion* stage, an aa or polypeptide is freely diffusing in the solvent as it does not feel the presence of the surface (distance greater than 10 Å).[27] On approaching the surface (distance smaller than 7 Å), the biomolecule penetrates

the first hydration shell thanks to moieties that enable the *association* step, such as hydrophilic (serine, Ser)[20,29] or positively-charged (arginine, Arg)[28] side chains (Figure 4.2b). Finally, molecules reach the typical distance of physisorpion (lower than 4 Å) in the *binding* step, adopting a flat conformation on the metal surface (Figure 4.2a).

The above-mentioned studies provide a profiling of the so-called Au-binding aa as useful clues for the design of specific Au-binding peptides and proteins. In parallel, numerous combinatorial and phage-display techniques identified selective Au-binding peptides for which the compositions were confirmed in the aforementioned findings.[4,30,31] The polypeptides are deemed to be Au-selective binders if they have aromatic and S-containing aa (Tyr, Met, Phe) as strong binding residues, and charged or polar aa (Ser, Arg) as anchoring points for the *association* step (Figure 4.3a). However, when going from a single aa to a sequence of them, it goes without saying that the affinity of a polypeptide is not directly the sum of the affinity of each individual residue.[26] In fact, the propensity of a polypeptide to bind to Au relies not only on the presence of strong binding residues, but also on the ability of the sequence to expose them to the surface. In other words, the sequence-encoded affinity of a polypeptide can be effective if such a chain is intrinsically flexible, in order to enable conformational changes to make the residues available for the contact with the surface. In this respect, extensive theoretical investigations of such phage display-identified peptide probed into the importance of the presence of residues with small side chains (*i.e.* Gly or Ala) that confer flexibility to the backbone (Figure 4.3b).[20] MD simulations proved that backbone conformational freedom was of utmost importance to guarantee the possibility of multiple plausible conformations suitable for the adsorption on the surface;[20,32–35] the affinity was then verified by binding experiments.[32,33]

Another aspect that enhances the affinity of a peptide to the Au substrate is its tendency to maximize backbone contact.[23] In line with this, proteins with a high β-sheet secondary structure propensity are ideally-suited for interactions with such a surface due to the tendency of directing the entire backbone to the material interface.[28,36] As observed during polypeptide adsorption, also during proteinic β-sheet adsorption on Au, positively charged[28] and polar aa[36] play a determining role

in the *association* step (Figure 4.2a). They can indeed cooperate with the first hydration shell, thereby contributing to protein-surface recognition.

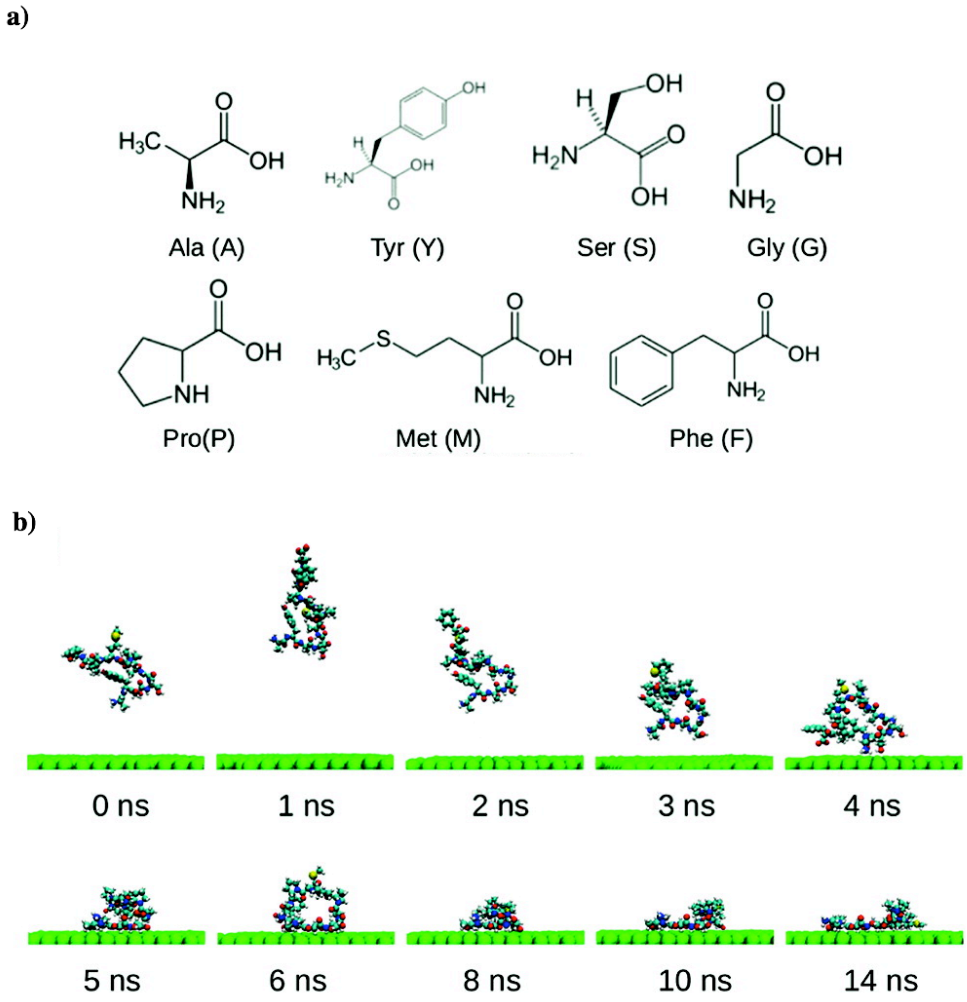


Figure 4.3. Structures of all aa present in a typical Au-binding peptide (a). Snapshots of the simulated stepwise adsorption process of an Au-binding peptide identified by a phage-display technique (b). The intrinsic backbone flexibility enables adoption of the preferred conformation for strong binding (water molecules are not shown for clarity). Adapted with permission from Ref. 20.

It is important to note that electrostatic interactions were repetitively found to play a key role in driving protein adsorption on metal surfaces, selecting the preferred conformation. The establishment of interactions of charged aa (such as Arg) with metals guides the orientation of the protein, for which immobilization is then stabilized by the protein's flexibility at the surface interface.[37] One must also consider the polarizability of metal surfaces, for which accurate modeling required the implementation of known force-field to describe both the inorganic components and their interactions with biomolecules.[38–40] During protein adsorption, a polarization

effect can be observed, and charge transfer might be possible.[19] However, it has been reported that both effects will have limited influence, especially in the presence of a solvent (*i.e.* water) with a similar or higher polarity than the considered biomolecule adsorbed on even surfaces.[19,40]

A separate discussion must be reserved for sulfur-containing aa, particularly cysteine (Cys) which displays a peculiar behavior on Au surface. It is worthy to note that this aa, thanks to the strong affinity of S to metals in general, is able to form a covalent bond with Au, as reported by experimental[41,42] and theoretical studies,[43,44] thereby making it exploitable as a prospective protein anchoring point on inorganic supports.[45] S-Au interactions between proteins and Au are reminiscent of the more general affinity of thiol molecules for inorganic surfaces, which have attracted increasing interest due to their ability, when placed as the "anchoring group" of organic compounds, to drive the adsorption of such molecules in an organized SAM.[46,47] Notably, SAM is an efficient, versatile and simple way of achieving high-coverage functionalization of Au, yielding a spontaneous, crystalline-like, functional patterned surface. Generally, SAM molecules are characterized by the presence of an "anchoring group" that is devoted to the adhesion to the surface, and a spacer link - usually an alkane chain that aids the packing of molecules and separation of the anchoring group from the terminal group, which ultimately defines the functionality of the entire surface (Figure 4.4).

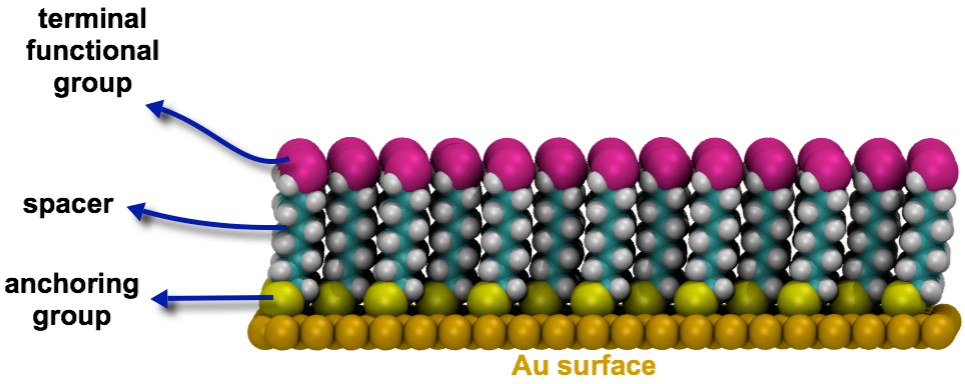


Figure 4.4. Schematic view of the classic alkanethiol SAM showing the constituting parts anchoring group (yellow), spacer (cyan) and terminal functional groups (magenta).

The merging of such a highly-ordered surface with a myriad of possible functional groups provides an incredibly appealing, biocompatible and biomimetic tool that finds wide applications in chemical and biological fields, stimulating the scientific community to develop new tailored patterned surfaces that are able to interact with biological matters.[48]

4.2 Aim of the project

In Chapter 3, we focused on the design of new functional hybrid materials based on GNMs as targeting agents towards cancer cells. In the present chapter, we want to apply the biohybrid construction approach to investigate the cellular migration, another crucial aspect related to the pathological condition of cancer cells. In particular, understanding fundamental individual[49,50] and collective[51] cellular migration mechanisms could, in principle, help current therapeutic developments as the stimulation of malignant cells to migrate allows them to detach from a primary tumor undergoing metastatic growth in distal organs.[52] Hence, by studying this important pathological process, our goal is to build a biological model surface that would allow us to control the migration of cancer cells in a directional and regulated manner. Consequently, the induction of cellular migration by external stimuli requires a motogenic functional group. In this regard, it is well known that fibronectin plays an important role as a crucial protein involved in adhesion, growth and cell migration events.[53] Interestingly, a highly-conserved sequence, the Ile-Gly-Asp (**IGD**) has been recently identified in the I_7 and I_9 modules of the gelatin-binding domain[54,55] and in the migration stimulating factor[56] of fibronectin. Extensive studies by Schor *et al.* showed that low concentrations of **IGD**-containing peptides stimulate high motogenicity in human dermal fibroblasts (HDFs).[57,58] Other studies revealed that cancer cells can proteolitically modify fibronectin into polypeptides exposing **IGD** residues, the latter stimulating motility.[59]

Hence, in order to engineer a biomimetic surface, the SAM approach was considered to be particularly suitable for the desired purpose, while the motogenic properties would be achieved by the sequence-encoded functionality of an **IGD**-based peptide. The idea is therefore to form SAMs of an **IGD**-based peptide on Au by producing a chemical gradient concentration of the motogenic agent thereby inducing directional and controlled cellular migration along the surface. To do so, we designed

a modified decanthiol molecule where the thiol and alkane chains represent the anchoring group and the spacer respectively; the terminal group is rendered by the **IGD**-containing peptide bearing the motogenic sequence L(+)-isoleucine, glycine, L(-)-aspartic acid, L(-)-glutamine (**IGDQ**) and is further exposed by an additional molecular linker L(-)-lysine (K) (**IGDQK-SH**, Figure 4.5). To prove its motogenic activity, a concentration gradient of **IGDQK-SH** on Au surface was first envisaged in order to attain control of cell migration. At this stage, the patterned surface would allow both recognition of the cancer cells through the motogenic sequence and their migration following the increase in concentration of exposed motogenic agent. In a second stage, the tailored modification with bidirectional gradient will be proposed to accomplish spatial control of the directed movement of cells.

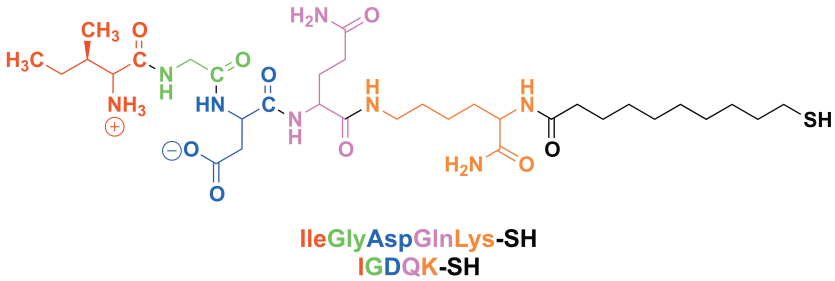


Figure 4.5. Chemical structure of **IGD**-containing thiolated pentapeptide used in this study as chemotactic agent (**IDGQK-SH**).

The project presented herewith consisted of an experimental part in which modified SAMs were successfully produced and tested biologically. In light of the experimental evidence, a computational part was developed to unravel the structural and dynamic aspects of the designed motogenic surface. Hence, in the next section, the experimental achievements are presented first, followed by the complementary *in silico* studies.

4.3 Experimental results

It is known that the formation of well-ordered and stable organized SAMs on Au surface is crucial for the suitability of these surfaces towards the final goal, which in our case is cell recognition and migration. A first attempt at building an **IGD**-based motogenic SAM, the **IGDQK-SH** pentapeptide (Figure 4.5) was successfully

synthesized and a gradient SAM (**∂IGDQK-S·Au**) was produced. In order to characterize its topography and surface organization, water contact angle (WCA) measurements and AFM analyses were performed. The former described the wettability change along the Au substrate through the WCA variation (ΔWCA); the latter showed the SAM organization and gradual distribution. Unexpectedly, the surface produced showed a change in wettability going from a hydrophobic to a more hydrophilic surface (WCA from *ca.* 40° to 35° ΔWCA = -4.9°, Table 4.1) and poor surface organization, since the height profile of the AFM analysis shows large peptide aggregates in the range of 60-100 nm (Figure 4.6)

Table 4.1. Static WCA values (°) as a function of the position along the Au surfaces, averaged over three measurements of the sample at the same distance.

SAMPLE	Position	WCA (°)	Δ (°)[a]
∂IGDQK-S·Au	1 mm	40.1 ± 2.3	-4.9
	5 mm	36.5 ± 1.3	
	9 mm	35.2 ± 2.0	
∂IGDQK-S·Au/ MT(PEG)$_4$-S·Au	1 mm	37.9 ± 0.3	7.1
	5 mm	40.6 ± 1.1	
	9 mm	45.0 ± 1.4	
∂IGDQK-S·Au/ C$_8$H$_{17}$-S·Au	1 mm	48.7 ± 1.0	22.5
	5 mm	59.3 ± 0.7	
	9 mm	71.2 ± 0.7	

[a] Difference in WCA (°) between the initial point (1 mm) and the final point (9mm) of the Au surfaces.

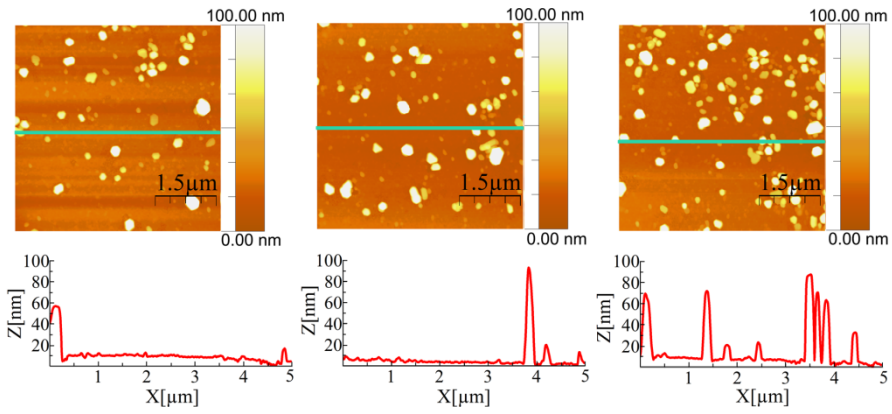


Figure 4.6. AFM images (top) with their relative cross section profiles (bottom) along the **∂IGDQK-S·Au**.

The observed hydrophobicity, together with the large formation of aggregates is a first sign of poor organization of such a surface. However, **∂IGDQK-S-Au** was then tested in the presence of a model cancer cell (MDA-MB-231) and of model healthy cells (human dermal fibroblast HDF, Figure 4.7) to measure its biological response. Indeed, by depositing the two cell lines onto the **∂IGDQK-S-Au** SAM and imaging their behavior over time using an optical microscope, the replication phenomenon of the seeded cells could be recognized, showing a visibly increased number of both cancer and healthy cells in the deposition area (Figure 4.7).

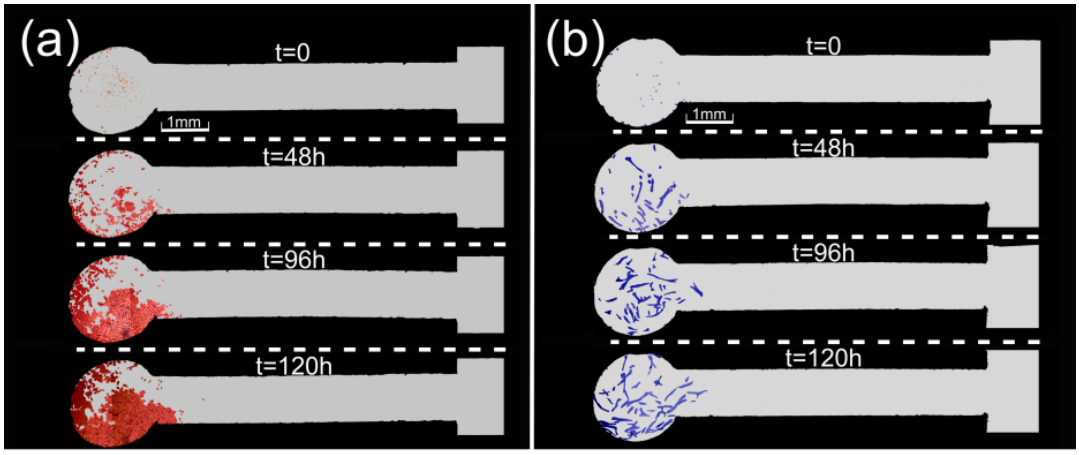


Figure 4.7. Optical microscopy images of MDA-MB-231 cells (a), HDFs (b) onto **∂IGDQK-S-Au**. All migratory experiments were repeated at least 5 times.

Due to these unexpected results, *mixed* SAMs of **IGDQK-SH** with increasing gradient density were envisaged using molecular filler, namely *n*-alkanethiol derivatives, to backfill the peptide gradient. It is reported in the literature that the introduction of alkenethiols into thiolated peptides aids in the formation of ordered and organized SAMs.[60–62] Hence, taking inspiration from the known ability of pegylated thiols to act as a non-fouling protein surface,[63–67] the methyl-PEG_4-thiol (**MT(PEG)$_4$** Figure 4.8a) was first selected to backfill the **IGDQK-SH**-modified surface to help SAM organization and consequently the exposition of the recognition motif.

However, even in **∂IGDQK-S-Au/MT(PEG)$_4$-S-Au**, the surface wettability was hardly changed along the gradient, with ΔWCA = 7° from low to high peptide concentration (Table 4.1).

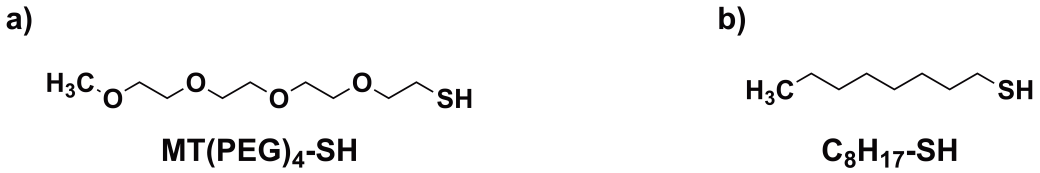


Figure 4.8. Chemical structure of methyl-PEG$_4$-thiol (**MT(PEG)$_4$-SH**) a) and 1-octanthiol (**C$_8$-CH$_3$-SH**) b) molecular filler chosen for the *mixed* SAMs.

Furthermore, in this surface, large peptide aggregates were observed in the AFM analysis (height profile: 50-90 nm, Figure 4.9) suggesting that the presence of the pegylated filler did not suffice to significantly improve SAM organization.

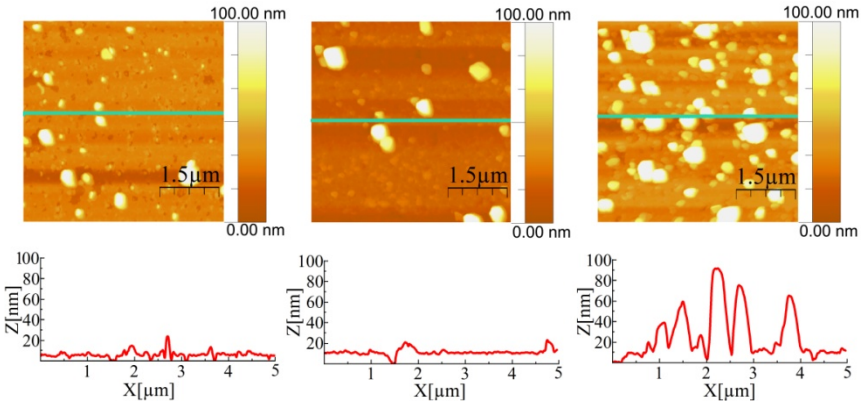


Figure 4.9. AFM images (top) with their relative cross section profiles (bottom) along **∂IGDQK-S-Au/MT(PEG)$_4$-S-Au**.

To confirm this, when the motogenic activity was tested in the presence of model cancer cells MDA-MB-231, cell replication behavior was mostly observed to reach confluence inside the deposition spot, while a very minor fraction of individual cells migrated towards the increased concentration of **IGDQK-S-Au** as shown in Figure 4.10.

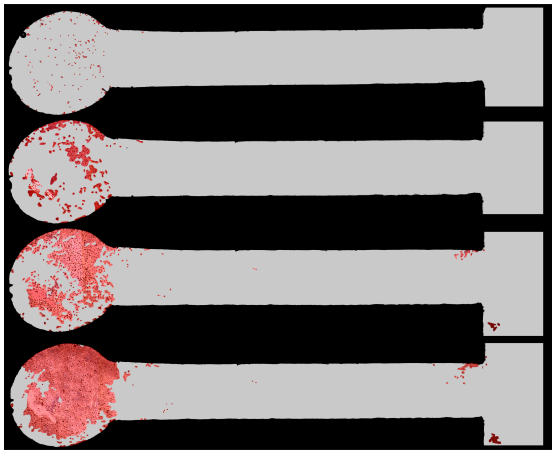


Figure 4.10. Optical microscopy images of MDA-MB-231 cells on **∂IGDQK-S-Au/MT(PEG)$_4$-S-Au**. All migratory experiments were repeated at least 5 times.

Hence, as a second attempt to enhance SAM organization and expose the motogenic sequence, a more hydrophobic alkenethiol filler was envisaged, namely 1-octanethiol **C$_8$H$_3$-SH** (Figure 4.8b). Interestingly, when the new *mixed* SAM **∂IGDQK-S-Au/C$_8$H$_{17}$-S-Au** was produced and characterized, it emerged as having the most significant ΔWCA (22.5°, Table 4.1), meaning that along the gradient, better exposition of the peptide is achieved which gave a more hydrophobic surface. Additionally, the AFM analysis confirmed the gradient organization of the patterned surface (height profiles: 10-20 mm in Figure 4.11), which significantly differed from the previous analyzed surfaces.

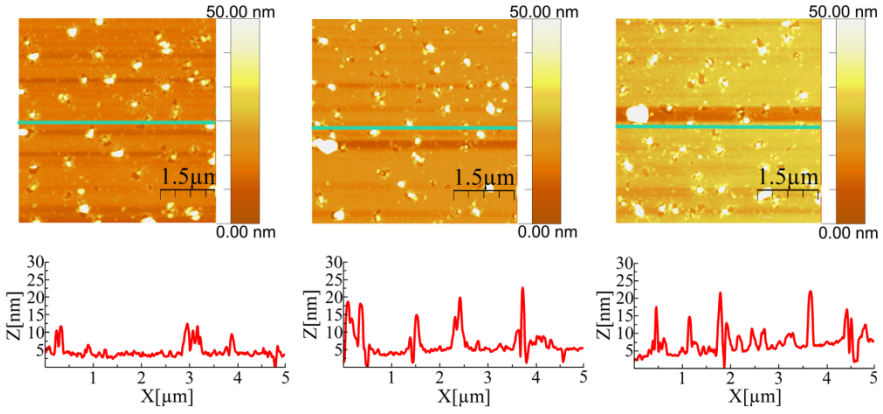


Figure 4.11. AFM images (top) with their relative cross section profiles (bottom) along the **∂IGDQK-S-Au/C$_8$H$_{17}$-S-Au**.

Most importantly, effective and selective cellular migration of the model cancer cells was achieved within this patterned surface (Figure 4.12). It should be noted that after

just 24 h, the cells had fully adhered and replicated, progressively occupying the deposition area. At 48 h, migration towards increasing surface density of **IGDQK-S-Au** had started, whereas after 96 h, some cells had remarkably covered distances of *ca.* 8-9 mm, following the *versus* imposed by the gradient. The trend was further enhanced after 120 h. Conversely, HDFs showed complete adhesion but in a different manner to MDA-MB-231 cells, as only a few fibroblasts reached about one third of the gradient path and none had travelled the same distance as the cancer cells after 120 h (Figure 4.12b).

In summary, from these preliminary experimental outcomes, three different **IGDQK-S-Au** SAMs were envisaged and tested, allowing us to observe remarkable differences in topography, organization and biological response of the obtained patterned surfaces. Surprisingly, the combination of **IGDQK-SH** with **C_8H_{17}-SH** resulted in the best condition to get an ordered and biologically active SAM.

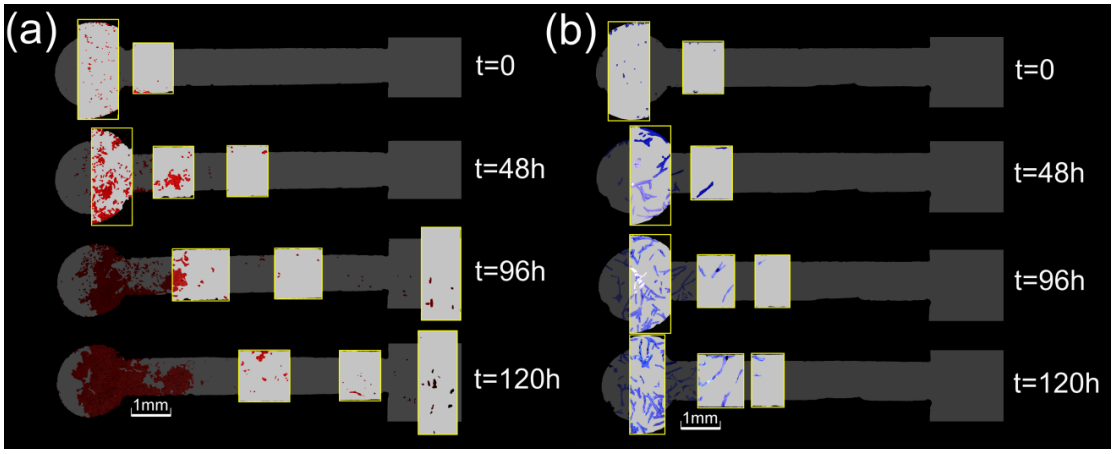


Figure 4.12. Optical microscopy images of MDA-MB-231 cells (a), HDFs (b) on **∂IGDQK-S-Au/C_8H_{17}-S-Au**. All migratory experiments were repeated at least 5 times.

4.4 Computational challenge

From the above discussion, it clearly emerged that the effective motogenic activity of **IGDQK-SH** depends on correct SAM organization, for which functionality is strictly linked to **IGDQ** exposition. Unexpectedly, the designed peptide organizes itself very differently depending on the concentration along the gradient and the presence or not of different kinds of molecular fillers. These conditions proved to be so important that they could fully compromise the biological

response of the resulting patterned surface. Thanks to extensive characterization of both the structural (AFM and WCA analysis) and functional (biological tests) aspects of the SAMs formed, it was unveiled that considerable hydrophilicity was associated with poor **IGDQ** exposition. Although it is plausible, the correlation between the actual exposition of the motogenic sequence and the nature of the surface still remain an assumption. In fact, we assume that in order to be an effective motogenic substrate, the SAM surface should properly expose the Ile residue of the **IGDQ** sequence as it was demonstrated to play a pivotal role in sequence bioactivity.[58] Due to the nature of this aa, proper Ile orientation would therefore give a highly hydrophobic surface. These conditions are indeed verified in the ***∂*IGDQK-S˙Au/C_8H_{17}-S˙Au** that displays an organized, hydrophobic surface on which the cancer cells' selective motogenic phenomenon is attained.

The role and the way in which octanthiol accomplishes the ideal confinement of the desired motogenic sequence with respect to the SAM of the pentapeptide alone (***∂*IGDQK-S˙Au**) or compared to the one obtained in the presence of the pegylated filler (***∂*IGDQK-S˙Au/MT(PEG)$_4$-S˙Au**) remains unexplained. Therefore further studies are required to explain structural aspects and the interactions governing the correct formation of **IGDQ** exposed SAMs. Computational techniques and, in particular, classical MD simulations result in a powerful and feasible approach to describe phenomena occurring at the peptide-surface interface at the atomistic level. Hence, in order to complement the experimental findings and postulate a reliable rationale to this behavior, a stepwise computational protocol was envisaged. The need to identify model systems that reliably approximate experimental conditions is of primary importance for an in depth understanding of the empirical outcomes correlating the structural aspects of the designed systems to their biological activity.

The designed pentapeptide **IGDQK-SH** bears a protonated Isoleucine (NH_3^+) and a deprotonated Aspartate (COO^-) within its sequence, which might interact between each other. Its backbone exhibits considerable flexibility, due to the presence of the Gly aa, which might induce exposition of the polar aa (Gln and Lys, Figure 4.5). Therefore, an initial investigation of the peptide conformational space is required, when it is alone in solution. Next, a series of intermolecular interactions must be accurately investigated to unravel the behavior of the peptide on Au surfaces as the SAMs produced in the presence or in the absence of different fillers were observed to behave differently. Firstly, Au-peptide interactions will be tackled since a

particular affinity between the two might play a defining role in subsequent SAM formation. Afterwards, attention will be focused on interpeptide interactions, as they can drastically affect peptide sequence exposition. Additionally, dealing with a gradient peptide monolayer, different peptide concentrations must be evaluated. Finally, the relationship between the peptide and different molecular fillers will be investigated, as divergent structural and biological trends have been described experimentally for the three tested systems.

In summary, studying the conformations of the designed **IGDQK-SH** peptide as a function of its concentration, environment (in solution or surface adsorption) and of the different natures of filler (**MT(PEG)$_4$** or **C$_8$H$_{17}$-SH**), the structural and dynamic aspects driving surface assembly and ultimately the desired functionality will be addressed. This would allow us to provide an exhaustive validation of experimental findings contributing to atomistic and molecular information in order to visualize the prepared peptide-based motogenic Au surfaces.

4.5 Methods section

4.5.1 Model approximation

Classical MD simulations of model systems were run in explicit methanol (MeOH) in time of 10 ns to evaluate the peptide conformations during the experimental conditions employed for the formation of the SAMs. In order to compare such results with the conditions in which **IGDQK-SH** SAMs are exposed to the cells, 10 ns MD simulations in explicit water were also performed on the last snapshot of the MD in MeOH. Hence, inspired by the produced surfaces and the experimental conditions, ten different systems were identified for systematically studying the **IGDQK-SH** behavior in both the solvents as summarized in Table 4.2. For the sake of clarity the nomenclature used to label the chosen systems is reported as follow:

- **IGDQK-SH/MeOH**: a single molecule of peptide in MeOH represented the starting model system that can give an idea on the behavior of the free peptide

molecule in solution. The system, solvated in a 10 Å buffer of MeOH, counts 3627 atoms.

- **1IGDQK-S-Au**: a single molecule of peptide simulated chemisorbed through the thiol moiety on a slab of Au was used to investigate the way of adsorption when it first approaches the Au surface. The solvated system counts 15272 atoms in MeOH, 8357 in water.
- **1IGDQK-S-Au/C_8H_{17}-S-Au and 1IGDQK-S-Au/MT(PEG)$_4$-S-Au** : a single molecule of peptide chemisorbed onto a slab of Au was simulated surrounded by a SAMs of either **C_8H_{17}-SH** or **MT(PEG)$_4$-SH**. The solvated systems counts 13172 and 12515 atoms in MeOH, 13565 and 16358 in water in the presence of **C_8H_{17}-SH** and **MT(PEG)$_4$-SH** SAMs, respectively.
- **2IGDQK-S-Au**: two molecules of peptide simulated chemisorbed through the thiol moiety on a slab of Au were used to study the way of adsorption of increasing peptide concentration considering the interpeptide interactions. The solvated system counts 13708 atoms in MeOH, 11809 in water.
- **2IGDQK-S-Au/C_8H_{17}-S-Au and 2IGDQK-S-Au/MT(PEG)$_4$-S-Au**: two molecules of peptide chemisorbed onto a slab of Au were simulated surrounded by a SAMs of either **C_8H_{17}-SH** or **MT(PEG)$_4$**. The second thiolated peptide is placed close to the first, spaced out by a molecule of **C_8H_{17}-SH** or **MT(PEG)$_4$-SH**. The solvated systems counts 13208 and 14806 atoms in MeOH, 15020 and 17044 atoms in water in the presence of **C_8H_{17}-SH** and **MT(PEG)$_4$-SH** SAMs, respectively.
- **∂IGDQK-S-Au**: an increasing number of peptide molecules simulated chemisorbed through the thiol moiety onto a slab of Au was used to model the gradient peptide SAM tested experimentally. The solvated system counts 15830 and 19454 atoms in MeOH and water respectively.
- **∂IGDQK-S-Au/C_8H_{17}-S-Au and ∂IGDQK-S-Au/MT(PEG)$_4$-S-Au**: an increasing number of peptide molecules chemisorbed onto a slab of Au was simulated having each peptide line spaced out by a line of either **C_8H_{17}-SH** or **MT(PEG)$_4$-SH** molecules. These systems were designed with the aim of simulating the **IGDQK-SH** chemical gradient along the Au surfaces. Indeed in the **C_8H_{17}-SH** or **MT(PEG)$_4$-SH** SAMs, an increasing number of filler molecules was replaced by peptides molecules. This situation allowed generating a single peptide placement on one side, followed by an increased

concentration, which reached the full Au surface coverage on the opposite side. These systems counts 16762 and 15132 atoms in MeOH, 25097 and 25395 in water the presence of **C_8H_{17}-SH** and **MT(PEG)$_4$-SH** SAMs, respectively.

The time scale is limited (20 ns in total for each system), but it was deemed sufficient to simulate the initial stages of adsorption, as observed elsewhere.[37,68,69]

Table 4.2. MD simulations performed onto the selected model systems.

System	Time (ns)	Number of Atoms	
		MeOH	Water
IGDQK-SH/MeOH	10	3627	
1IGDQK-S·Au	10	15272	8357
1IGDQK-S·Au/C_8H_{17}-S·Au	10	13172	13565
1IGDQK-S·Au/MT(PEG)$_4$-S·Au	10	12515	16358
2IGDQK-S·Au	10	13708	11809
2IGDQK-S·Au/C_8H_{17}-S·Au	10	13208	15020
2IGDQK-S·Au/MT(PEG)$_4$-S·Au	10	14806	17044
∂IGDQK-S·Au	10	15830	19454
∂IGDQK-S·Au/C_8H_{17}-S·Au	10	16762	25097
∂IGDQK-S·Au/MT(PEG)$_4$-S·Au	10	15132	25395

We decided to simulate the first layer where the alkanethiols are adsorbed to, approximating the Au (111) substrate to one-slab surface counting 480 atoms. Although aware that the multilayer underneath are the ones involved in charge transfer processes, their presence was neglected as we want to focus mainly on the organization of the SAM and the secondary non-covalent interpeptide interactions. Concerning the SAM structure, we must take into account that it depends on the morphology of the metal surface. Au (111) is characterized by an atomic arrangement of six-fold hexagonal symmetry which corresponds to the most efficient way of packing atoms within a single layer ("closed packed"). Thanks to the strong interaction between thiol molecules and Au (111) substrate, the sulfur chemisorbs on specific site of the surface, forming densely packed monolayer with a chain tilt of about 30° respect to the surface normal. This arrangement allows the maximum packing driven by vdW interactions between the alkyl chains and a structure with a

(√3x√3)R30° overlayer (R=rotated) lattice,[70] as schematized in Figure A.8, *Appendix*. The one-slab Au surface designed for our simulations can fit 130 thiol molecules.

4.5.2 Parameterization

IGDQK-SH starting structure was parameterized with protonated Isoleucine (NH_3^+), deprotonated Aspartate (COO^-) and one neutral sulfur atom, resulting in an uncharged molecule. Charges were computed using the restrained electrostatic potential (RESP) approach,[71] applying the antechamber module available in AMBER12,[72] and the force field (*ff*) parameters were assigned to the peptide according to the AMBER parm99SB *ff*.[73] Only for the MD simulation of **IGDQK-SH** in MeOH, the terminal thiol group was explicitly parameterized as a thiolated species SH. Similarly, **C_8H_{17}-SH** and **MT$(PEG)_4$-SH** were parameterized using the general Amber *ff* (gaff).[74] The Au-S bond between self-assembled molecules and Au atoms was placed at the distance of 2.682 (Å) and treated with a Lennard-Jones potential of 9.033 kcal mol^{-1}, as reported in the Dreiding *ff*.[75] The gold-thiol interaction described with a non-bonded potential provides a more realistic representation of the organothiols, demonstrated experimentally to be able to hop from one site to another allover the Au surface.[76,77] The Au-Au bond distance was set at 2.884 (Å) and Au atoms were restrained by setting the force imposed by organic molecules to zero. Freezing the Au atoms is justified by reported evidences that demonstrate their relaxation has a negligible effect on the conformation of the SAMs.[78] The Au-Au interactions were neglected by excluding them from the neighbor list.

4.5.3 Force field based MD simulations

10 ns classical MD simulations were performed for all the systems first in MeOH. The last snapshot of each simulation was then solvated in water and additional 10 ns of MD simulations were carried out. Table 4.2 summarizes all the simulated systems. The AMBER 12 package[72] was employed, and in all the cases the

initial structures were relaxed by short minimization runs of 2000 steps, using the conjugate gradient energy minimization algorithm (convergence criterion $1.0E^{-4}$ kcal mol^{-1} Å). The MD simulations were performed at constant volume (NVT) and temperature (298 K) using Langevin thermostat with a 2 fs time step. Periodic boundary conditions were applied. Non-bonded cut-off of 8 Å was set and Particle Mesh Ewald (PME) summation[79] was used to calculate the long-range electrostatic interactions. The bonds involving hydrogen atoms were kept fixed using the SHAKE algorithm, and the Au atoms were hold fixed along all the simulations. MEOHBOX[80] model was used to simulate methanol molecules, while TIP3P[81] model was used to represent the water molecules.

4.5.4 Calculated properties

Cluster analysis was performed on the 10 ns MD simulation of **IGDQK-SH/MeOH** using the *ptraj* module of AMBER 12.[72]

For all the systems the Root-Mean Square Deviation (RMSD) on the C_α, C, and N atoms of the **IGDQK-SH** backbone, was calculated over the MD trajectories. The Radius of Gyration (Rg) of the peptide was also calculated along the simulations together with the intermolecular H-bonds occurring within the SAMs with a minimum occupancy of the 4% along the simulations. In the gradient simulated system, the % of occupancy of the H-bonds between the same pair residues was averaged among them. Besides, the radial distribution functions (RDFs) were calculated between the oxygen atom (O) of water and the carbon atom (C) of the terminal methyl of the filler molecules in the *mixed* SAMs (*i.e.* **C_8H_{17}-S·Au** and **MT$(PEG)_4$-S·Au**). For the simulation without filler the RDF was also calculated considering the distance between the water O and the exposed gold atoms of the slab. This way of calculating the RDFs gives an idea of the degree of solvation of the surface that it is reduced when the water-surface interactions are replaced by peptide-surface contacts. Meanwhile, in order to have a precise information about the different level of solvation of the peptide as a function of its increasing concentration, the RDFs were calculated between the water O atom and the C of the terminal Ile, the terminal methyl group of **C_8H_{17}-S·Au** and **MT$(PEG)_4$-S·Au** for each residue separately, and between the water O and the gold atoms to measure also the surface-water

interactions. The *ptraj* module of AMBER12[72] was used for all the above properties. The Solvent Accessible Surface Area (SASA) of the **IGDQK-SH** molecules was calculated along all the trajectories with the *Cpptraj* module of AMBER12.[72]

The Electrostatic surface potentials for **IGDQK-SH** was calculated by solving the Poisson-Boltzmann equation with the APBS program.[82] The results were visualized using a VMD (v 1.9.1) interface.[83] The non-bonded interaction energies, namely the electrostatic (EEL) and the van der Waals (vdW), among the chemisorbed molecules (**IGDQK-SH**, **C_8H_{17}-S·Au** and **MT(PEG)$_4$-S·Au**) were calculated over the simulated time using a cutoff of 8 Å employing the NAMD Energy plugin (v. 1.4) as implemented in VMD (v 1.9.1) according to Eq. 2.16.[83]

4.6 Results and discussion

This section is divided into four main parts: *"conformational analysis of IGDQK-SH"* where the structural aspect of the solvated peptide will be addressed; *"SAMs of IGDQK-S·Au"* where **1**, **2** or a gradient *∂* of **IGDQK-SH** will be studied adsorbed onto the Au surface; *"SAMs of IGDQK-S·Au*/MT(PEG)$_4$-S·*Au"* where **1**, **2** or a gradient *∂* of **IGDQK-SH** surrounded by the PEG filler will be studied adsorbed on Au surface; *"SAMs of IGDQK-S·Au*/*C_8H_{17}-S·Au"* where **1**, **2** or a gradient *∂* of **IGDQK-SH** surrounded by the octanthiol filler will be studied adsorbed onto the Au surface.

4.6.1 Conformational analysis of IGDQK-SH

In order to evaluate the intrinsic conformational freedom of **IGDQK-SH** 10 ns MD simulations of the peptide solvated in a box of MeOH were run. To quantify the structural changes over the trajectories the RMSD of the C_α, N and C atoms of the peptide backbone, its Rg and the peptide's intramolecular H-bonds were calculated. While the former gives an indication of the different peptide conformations, the second is a measure of the peptides compactness and the latter gives an idea of the interaction network involved within the SAM organization.

By analyzing the **IGDQK-SH/MeOH** system, first the RMSD and Rg plots were calculated as represented in Figure 4.13. As a result of the degrees of freedom associated to a free **IGDQK-SH** molecule in MeOH, the RMSD plot presents large fluctuations, which are related to significant peptide conformational changes in solution (Figure 4.13a). Moreover performing a cluster analysis along the **IGDQK-SH/MeOH** trajectory, such variations resulted in the alternation of two main populated conformations. In fact, it results that **IGDQK-SH** can interchanges between a closed and extended conformation with occupancy of the 35.7 % and 64.3 %, respectively.

The oscillation between these two conformations is also confirmed by the calculated Rg as a function of time (Figure 4.13b).

Specifically, this parameter represents a measure of the mass-weighted spatial distribution of the atoms in each peptide conformation, therefore it is higher (lower) when the molecule is extended (closed) giving an estimation on the deviation of the peptide structure from its center of mass. The obtained plot shows that the Rg of **IGDQK-SH** initially decreases when the peptide adopts the more compact conformation (closed, Rg_{min} = 5.0 Å), whereas it increases when the molecule switches to the extended structure (Rg_{max} = 8.7 Å), showing Rg values oscillating between the two conformations as a function of time (Figure 4.13b). None of the H-bonds detected along this simulation had an occupancy ≥ 4%. This analysis confirms the significant backbone flexibility of the designed peptide **IGDQK-SH** that routinely interconverts between two conformations. Two Rg values were detected for the closed and extended conformation that will serve as a reference for the conformational evaluation of the peptide within the SAMs. Finally, this alternating conformational change might be the plausible picture of the peptide in solution before approaching the Au surface during the SAM formation.

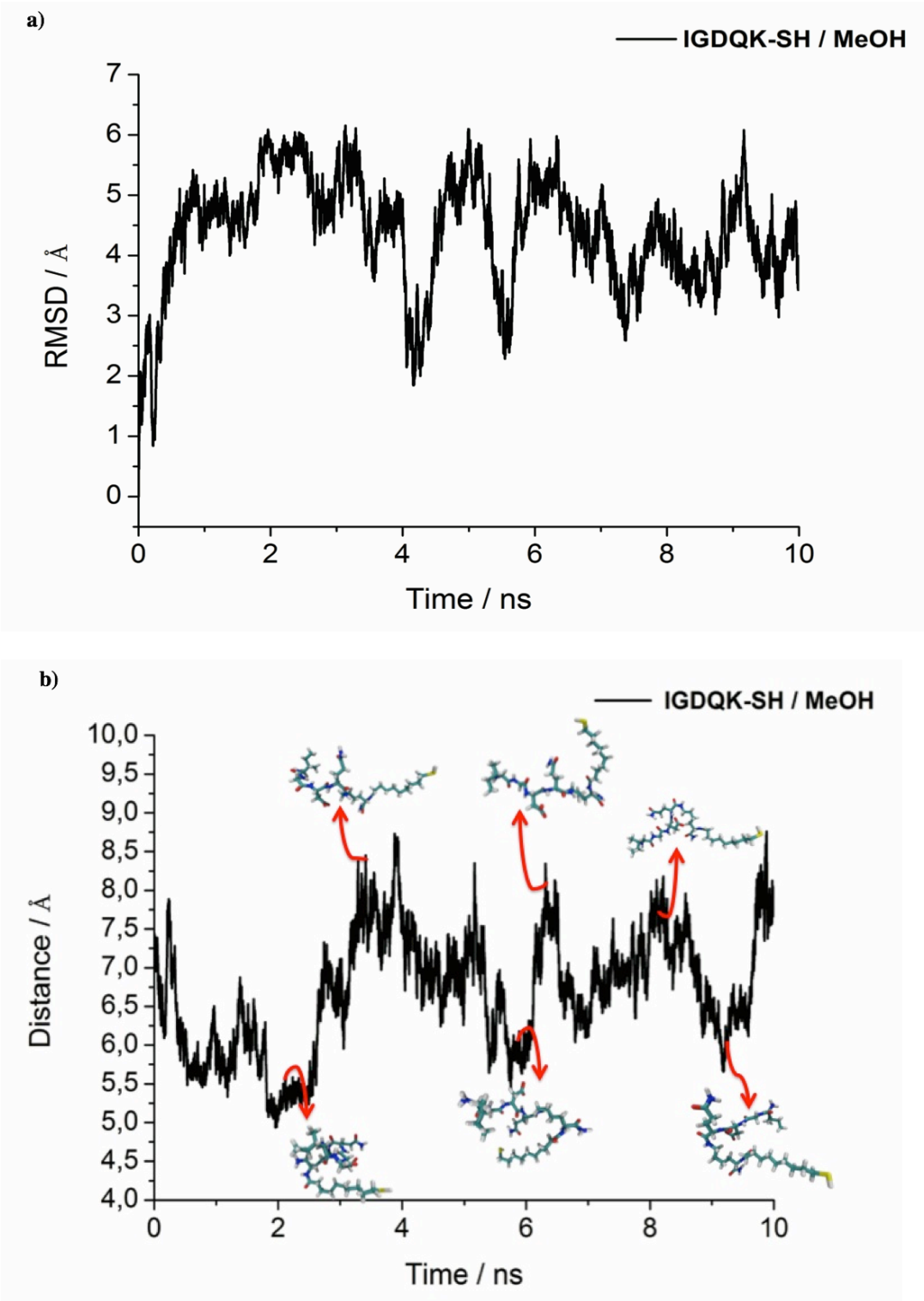


Figure 4.13. RMSD (a) and Rg (b) plots calculated along the 10 ns trajectory of **IGDQK-SH/MeOH**. The alternation of the closed (bottom of the Rg plot, b) and extended (top of the Rg plot, b) conformations adopted by **IGDQK-SH** during the simulation is highlighted with the licorice drawn molecules.

4.6.2 SAMs of IGDQK-S·Au

Taking in mind the conformational alternation of the peptide in solution, its subsequent behavior when approaching the surface was studied by simulating **1**, **2** and the gradient *∂* of **IGDQK-S·Au** immobilized onto the surface were simulated first in MeOH, then in water.

As previously described, the RMSD fluctuations and the Rg values in addition to the SASA were calculated over the 10 ns trajectories for the aforementioned systems (Figure A.10, *Appendix*). The SASA gives an idea of the degree of exposed surface within a molecule along its conformational changes as a function of the simulated time. For the sake of clarity we will refer to the average <RMSD>, <Rg> and <SASA> values, for which comparison gathers significant differences as reported in Table 4.3.

Specifically, the <RMSD> of the peptide backbone calculated for the **1IGDQK-S·Au** system in MeOH revealed large fluctuations (7.3 ± 3.5 Å), which can be compared to those observed for **IGDQK-SH/MeOH** (Table 4.3), demonstrating that structural changes persisted even when the peptide is chemisorbed onto Au surface. Interestingly, increasing the concentration of the peptides adsorbed onto the surface (**2IGDQK-S·Au** and ***∂*IGDQK-S·Au**) the <RMSD> is slightly higher in values (8.4 and 8.2 Å, respectively), meaning that large conformational changes as compared to the starting structure are still occurring, but with a lower oscillation (standard deviation, SD of 1.7 and 1.0 Å, respectively).

Reasonably, the presence of two peptides mutually influences the adsorption and conformations of the two molecules involved in intermolecular interactions. In fact, H-bonds between the side chains of the peptide sequence were detected with considerable % of average occupancy (among the others Asp-Ile 18% of occupancy, Table 4.4). When the simulations were then run in water the overall trend was generally not altered (Table 4.3). In fact, one or two peptides, freer to move, still undergo consistent fluctuation also in water (8.9 ± 2.2 and 9.4 ± 2.5, respectively), especially if one considers that the starting point was the MeOH equilibrated last snapshot. While surrounded by a gradient of the same peptide, the **IGDQK-SH** peptides present dumped fluctuations (5.7 ± 0.8 Å). In line with this the H-bonds

previously encountered persist also during the water simulated gradient system (Asp-Ile = 16%, Table 4.4).

Table 4.3 Average values of the RMSD, Rg and SASA calculated along the 10 ns MD simulations in MeOH and water for **1**, **2** or a gradient *∂* of **IGDQK-S·Au, IGDQK-S·Au/MT(PEG)$_4$-S·Au** and **IGDQK-S·Au/ C$_8$H$_{17}$-S·Au**. The standard deviations are also reported.

	<RMSD>[a]		**<Rg>[a]**		**<SASA>[b]**	
	MeOH	Water	MeOH	Water	MeOH	Water
IGDQK-SH	4.4 ± 0.9	-	6.7 ± 0.8	-	-	-
IGDQK-S·Au						
1	7.3 ± 3.5	8.9 ± 2.2	7.8 ± 0.9	5.7 ± 0.6	1035 ± 98	751 ± 75
2	8.4 ± 1.7	9.4 ± 2.5	12 ± 1.8	7.6 ± 0.5	1920 ± 128	1169 ± 100
∂	8.2 ± 1.0	5.7 ± 0.8	21.4 ± 0.7	20.2 ± 0.2	16466 ± 748	10585 ± 904
IGDQK-S·Au/ MT(PEG)$_4$-S·Au						
1	3.9 ± 0.3	4.8 ± 0.9	8.2 ± 0.7	7.2 ± 0.3	653 ± 67	546 ± 44
2	3.7 ± 0.3	3.9 ± 0.5	8.6 ± 0.4	8.5 ± 0.2	1085 ± 84	936 ± 60
∂	5.6 ± 0.8	5.6 ± 0.8	20.5 ± 0.6	20.2 ± 0.2	11056 ± 642	7770 ± 361
IGDQK-S·Au/ C$_8$H$_{17}$-S·Au						
1	3.5 ± 0.2	6.3 ± 1.7	8.3 ± 0.9	6 ± 0.3	799 ± 73	590 ± 50
2	3.9 ± 0.3	4.8 ± 1.1	8.7 ± 0.4	7.7 ± 0.2	1311 ± 92	1029 ± 70
∂	6.9 ± 1.0	5.4 ± 1.0	22.5 ± 0.7	22.5 ± 0.2	15412 ± 1187	11539 ± 641

[a] Å; [b] A^2

Consistently with these findings, when the non-bonded interaction energies, namely the EEL and the vdW were calculated for ***∂*IGDQK-S·Au** as a qualitative measure of the forces coming into play, the total energy of the system (<Tot> = -5429 ± 124 kcal mol^{-1} in MeOH and -4818 ± 180 kcal mol^{-1} in water, Table 4.5) was clearly driven by the EEL contribution (<EEL> = -4784 ± 101 and -3808 ± 88 kcal mol^{-1} respectively) as it accounts for the H-bonds interactions. On the contrary the vdW energy should refer to the interactions between the alkyl chains as a measure of the SAM organization. Indeed a lower contribution of the vdW is observed along the trajectories (<vdW> = -644 ± 55 and -975 ± 49 kcal mol^{-1} for ***∂*IGDQK-S·Au** in MeOH and water respectively), confirming the poor organization of this system. At confirmation of this energetic analysis, the electrostatics was already reported elsewhere to rule the adsorption on of proteins on inorganic surfaces.[37,68,69]

Table 4.4. The average occupancy (%) of H-bonds detected during the 10 ns trajectories of the simulated systems both in MeOH and water is reported. The residues involved in the intermolecular H-bond interaction are reported displaying the symbol name in black when is donor, in red when is acceptor.

	H-bonds		
System	Residue	<Occupancy> (%)	
		MeOH	Water
1IGDQK-S·Au	-	-	-
2IGDQK-S·Au	-	-	-
∂IGDQK-S·Au	ASP-ILE	18.3	16.3
	ASP-GLN	11.4	9.4
	GLY-LYS	9.0	-
	GLY-GLN	7.2	4.5
	GLN-GLN	6	-
	GLY-ILE	5.8	-
	LYS-LYS	5.6	5.2
	GLN-LYS	4.9	7.3
	ASP-LYS	4.4	9.5
1IGDQK-S·Au/MT(PEG)$_4$-S·Au	LYS-PEG	4	4
2IGDQK-S·Au/MT(PEG)$_4$-S·Au	ASP-ILE	23	11.8
	PEG-LYS	11.3	14.2
	ASP-LYS	-	10.6
∂IGDQK-S·Au/MT(PEG)$_4$-S·Au	ASP-ILE	16.9	17.9
	PEG-LYS	-	9.9
	PEG-ILE	9.8	-
	ASP-GLN	9	6.7
	GLN-LYS	7.2	11.3
	ASP-ASP	6.2	6.9
	GLN-GLN	5.8	9.72
	LYS-LYS	5.8	6.5
	GLY-LYS	4.6	17.2
	LYS-GLN	4.3	5.3
	ASP-GLY	4.2	42.7
1IGDQK-S·Au/C$_8$H$_{17}$-S·Au	-	-	-
2IGDQK-S·Au/C$_8$H$_{17}$-S·Au	GLN-GLN	-	14.7
	ASP-ILE	6	4.4
	ASP-GLY	-	5.8
∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au	ASP-ILE	22.7	15.9
	ASP-ASP	7.1	28.6
	GLN-ILE	18.7	-
	ASP-GLN	6.4	10.8
	LYS-GLY	10.1	-
	GLY-GLN	8.8	6.5
	LYS-GLN	8.6	7.9
	GLN-LYS	-	7.8
	GLN-ASP	-	7.8
	LYS-LYS	-	5.5
	ASP-LYS	-	4.7

Concerning the conformational aspects, the peptide tendency to reorganize onto the surface in the three systems understudy (**1-2-∂IGDQK-S·Au**) was then tackled calculating the Rg (Figure A.9a, *Appendix* and Table 4.3). The <Rg> of the single **1IGDQK-S·Au** displayed a stable trend around a mean value (7.8 Å) that, compared to the two identified conformations (Rg = 5.0 Å and 8.7 Å for closed and extended, section 4.6.2), implies a situation in between. Afterwards, the <Rg> of **1IGDQK-S·Au** simulated in water oscillates around 5.7 Å, for which value certainly refers to the closed state. In fact, the fluctuation of this value corresponds to the transition of the peptide structure from an upright to a more compact orientation flat on the gold surface, in MeOH as well as in water (Figure 4.11a). This is further corroborated by the reported tendency of aa[23,24] as well as of longer polypeptides[19,33–36] to favor a direct contact with the metal surface, preferring conformations that allow for major backbone contact[23,24] excluding any layers of water between the surface and the peptides residues with a subsequent loss in entropy.[19]

Table 4.5. Average total (<Tot>) electrostatic (<EEL>) and van der Waals (<vdW>) interaction energies (kcal mol^{-1}) corrected by their standard deviations are reported for **∂IGDQK-S·Au**, **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** and **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** simulated in MeOH and water.

System	Energy[a]	Solvent	
		MeOH	Water
∂IGDQK-S·Au	<Tot>	-5429 ± 124	-4818 ± 180
	<EEL>	-4784 ± 101	-3808 ± 208
	<vdW>	-644 ± 55	-975 ± 49
∂IGDQK-S·Au/ MT(PEG)$_4$-S·Au	<Tot>	1333 ± 156	2066 ± 233
	<EEL>	3244 ± 122	4132 ± 222
	<vdW>	-1911 ± 73	-2066 ± 35
∂IGDQK-S·Au/ C$_8$H$_{17}$-S·Au	<Tot>	-6011 ± 162	-5480 ± 159
	<EEL>	-4662 ± 109	-3861 ± 171
	<vdW>	-1349 ± 84	-1619 ± 41

[a] kcal mol^{-1}

Similarly, **2-∂IGDQK-S·Au** systems seem to maintain a certain extended conformation in MeOH (<Rg> = 12 ± 1.8 and 21.4 ± 0.7 Å, Table 4.3, Figures A.9a and 4.14b-c), while they drastically collapse onto the surface when simulated in water (<Rg> = 7.6 ± 0.5 and 20.2 ± 0.2 Å, Table 4.3, Figures A.9a and 4.14b-c). In particular, the simulated gradient displays a disordered organization in MeOH, that

becomes a perfect semi-spherical aggregation in water. This is in excellent agreement with the AFM measurements (Figure 4.6) that clearly displayed big aggregates of the ***∂*IGDQK-SH** peptide deposited on Au (111), failing in achieving a structured and organized SAM.

Afterwards, the SASA profile was calculated for the three simulations describing the exposed peptide surface as a function of its conformational change along the time. It gathers that the **IGDQK-SH** peptide has the tendency to adopt an extended conformation (<Rg> = 7.8 and 12 Å, Table 4.3) in which the exposed surface increases first in MeOH, whereas switching to the more closed conformations (<Rg> = 5.7 and 7.6 Å, Table 4.3) the <SASA> remarkably decreases along the simulated time in water (Table 4.3 and Figure A.9c). In particular, a drastic reduction of SASA is observed for ***∂*IGDQK-S·Au** that passes from 16466 to 10585 A^2 from MeOH to water, confirming the disordered aggregation with a reduced surface exposition. This might be at the base of the poor readability of the motogenic sequence in ***∂*IGDQK-S·Au** that indeed does not attain the migration of any kind of cell lines (Figure 4.7).

As a further confirmation of the experimental observations, also the electrostatic surface potential both in MeOH and water, displayed in Figure 4.14a-c, shows the tangential adsorption of the immobilized Au-peptide pointing the hydrophobic residues (blue coloration) Isoleucine (Ile) and Glycine (Gly) towards the surface, while exposing the hydrophilic aa (red coloration) Glutamine (Gln) or Aspartate (Asp) to the solvent bulk (MeOH and water respectively).

Additionally, it is worth noting that the Ile is here the N-term residue, thereby bearing a charged NH_3^+ group. This might contribute to the binding as it has been reported that positively charged aa strongly interact with Au surfaces.[23,25,26,28,29,84]

In summary, we can imagine that the stepwise formation (MeOH conditions) of **IGDQK-S·Au** starts first with the stable peptide chemisorption on Au through the S-Au bond, then followed by the spontaneous adsorption of all its structure through the backbone.

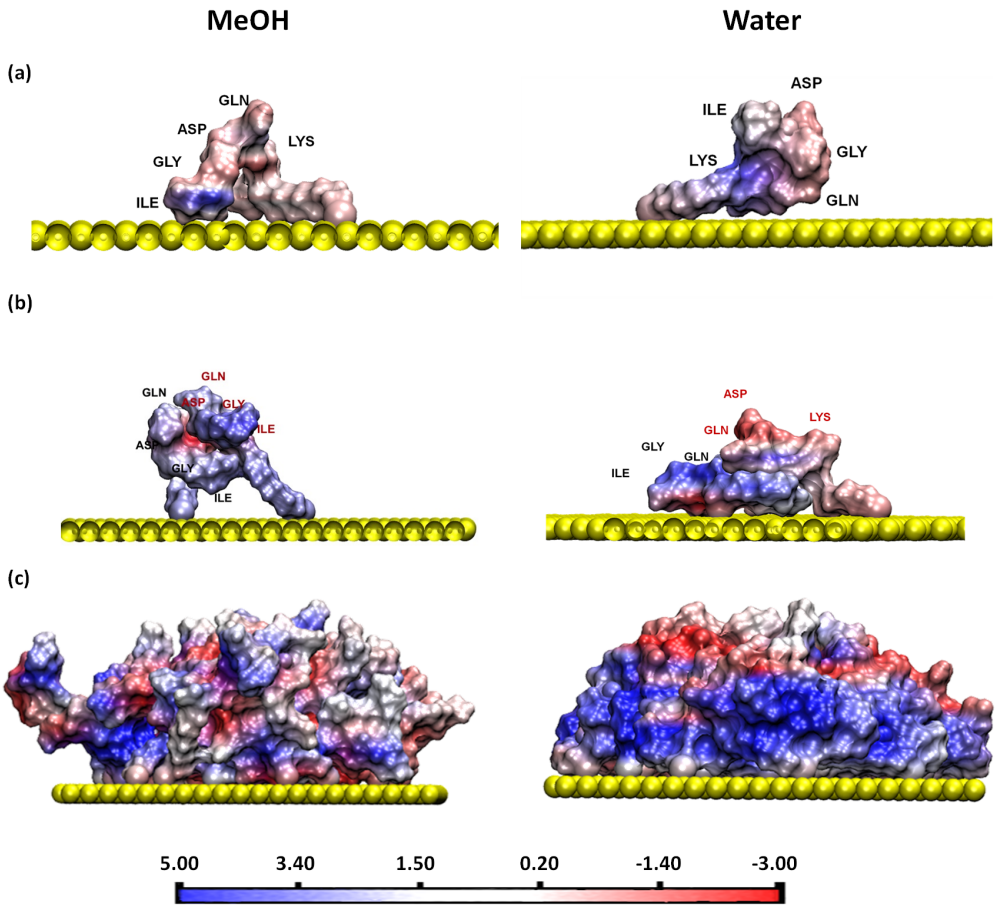


Figure 4.14. Front views of electrostatic surface potential plotted onto the peptide in the last snapshot of the **1IGDQK-S·Au** (a), **2IGDQK-S·Au** (b), and **∂IGDQK-S·Au** (c) MD simulations in MeOH (left column) and water (right column). The scale bar at the bottom is also reported. The aa residues within the peptide structures (a and b) are labeled with their symbol name, namely Isoleucine (ILE), Glycine (GLY), Aspartic acid (ASP), Glutamine (GLN) and Lysine (LYS).

Increasing the peptide concentration the self-assembling should be guaranteed by the vdW interactions among the alkyl chains. Conversely, in **2** and **∂IGDQK-S·Au** systems H-bonds occurring among the side-chains prevail over the vdW interactions affecting the correct assembling. Afterwards, the simulations in water mainly show the adoption of closed conformations stabilized by the interpeptide H-bonds. This is a first elucidation of the poor organization of full **IGDQK-SH** SAM that instead forms aggregates as observed by the AFM analysis.

4.6.3 SAM of IGDQK-S·Au / MT(PEG)$_4$-S·Au

To enhance the SAM organization and hopefully the peptide exposition, a *mixed* SAM of **IGDQK-S·Au** backfilled with **MT(PEG)$_4$-SH** molecular filler (Figure 4.8a) was envisaged and tested experimentally. Hence to reproduce the experimental tested systems, **1**, **2** and **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** first in MeOH and then in water were also investigated *in silico*. As reported in the literature,[63–67] the PEG filler is expected not only to improve the **IGDQK-S·Au** SAM formation, thus in our case the peptide exposition, but also to play as a non-fouling protein surface thereby impeding the stacking of the cells and promoting their migration exclusively thanks to the detection of the peptide gradient.

When the behavior of the peptide surrounded by the SAM of **MT(PEG)$_4$-S·Au** is simulated in the experimental condition of SAM production (MeOH), **IGDQK-S·Au** at any concentration (**1**, **2** and **∂**) displays always drastically dumped fluctuations in the <RMSD> (3.9, 3.7 and 5.6 Å respectively, Table 4.3 and Figure A.10b), which find explanations in the structural organization of the SAM. Notably, the decanthiol moieties of **IGDQK-SH** peptides are now inserted within the filler monolayer and surrounded by the PEG molecules; consequently their fluctuations are significantly reduced along the trajectories. Additionally, secondary non-covalent intermolecular interactions might also be responsible of the dumped fluctuations. In fact, in **1IGDQK-S·Au/MT(PEG)$_4$-S·Au** it starts appearing a peculiar H-bond interaction between the NH of the Lys residue and the terminal oxygen of the underlying PEG molecules. Such H-bonds act as anchoring points for the **IGDQK-SH** onto the **MT(PEG)$_4$-S·Au** SAMs, whilst with a relatively low average % of occupancy (4%, Table 4.4). This value rises to 11.3% of occupancy when the simulated peptide molecules become **2**. At the same time, in **2IGDQK-S·Au/MT(PEG)$_4$-S·Au** also remarkable interpeptide H-bonds are detected occurring, among the others, between the side chains of Asp and Ile (23%, Table 4.4). These interactions induce the peptide to assume a closed state thus contributing to the poor exposition of the motogenic sequence as the AFM-described topography demonstrates. When a gradient concentration of peptide is simulated (**∂IGDQK-S·Au/MT(PEG)$_4$-S·Au**), the H-bonds among the side chains (Asp-Ile 17% of average occupancy) prevail over the anchoring to the PEG molecules (PEG-Lys 9.8% of average occupancy, Table 4.4),

suggesting a concomitant role of both the interpeptide and peptide-PEG H-bonds in the poor exposition of the motogenic sequence. The anchoring through H-bonds between the polypeptide chain and the PEG SAM as observed herein is consistent with another study reported in the literature.[85] These findings well match with the <Rg> values of the *mixed* SAMs in MeOH that, despite range between values accounting for extended conformation (~8 Å for **1-2IGDQK-S·Au/MT(PEG)$_4$-S·Au** and 20.5 Å for **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au**, Table 4.3 and Figure A.10a), correspond to a closed arrangement spread over the PEG SAM (Figure 4.15).

Passing from MeOH to the biological test conditions (water), the <RMSD> still maintains similar constant values for the three systems **1**, **2** and **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** (4.8, 3.9 and 5.6 Å respectively, Table 4.3 and Figure A.10b), meaning that a better organization of the SAM is achieved reducing the peptide degree of conformational freedom. This is also confirmed by the increased % of average occupancy of the peptide-peptide and peptide-PEG H-bonds as reported in Table 4.4 (now % of occupancy ranges from 4 to 42%). As previously observed in MeOH, also the simulation of the gradient **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** in water displays interpeptide interactions with higher % of average occupancy with respect to the PEG-Ile H-bonds (42% and 10% respectively), albeit the latters result crucial for the anchoring of the peptides to the SAM. Hence, it gathers that at low concentration of the peptide (**1-2IGDQK-S·Au/MT(PEG)$_4$-S·Au**), the poor sequence exposition is mainly caused by the interaction of the pentapeptide with the PEG fillers. Meanwhile, when the peptide concentration increases (**∂IGDQK-S·Au/MT(PEG)$_4$-S·Au**) the anchoring to the underlying SAM is combined to a dense network of interpeptide interactions that significantly contribute to the inefficient exposition and readability of the **IGDQK-S·Au** sequence. The network of H-bonds can reasonably explain the slight reduction of <Rg> from MeOH to water observed in the three systems **1-2-∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** where the <Rg> values pass from 8.2 to 7.2, 8.6 to 8.5, 20.5 to 20.2 Å, respectively (Table 4.3 and Figure A.10a, *Appendix*).

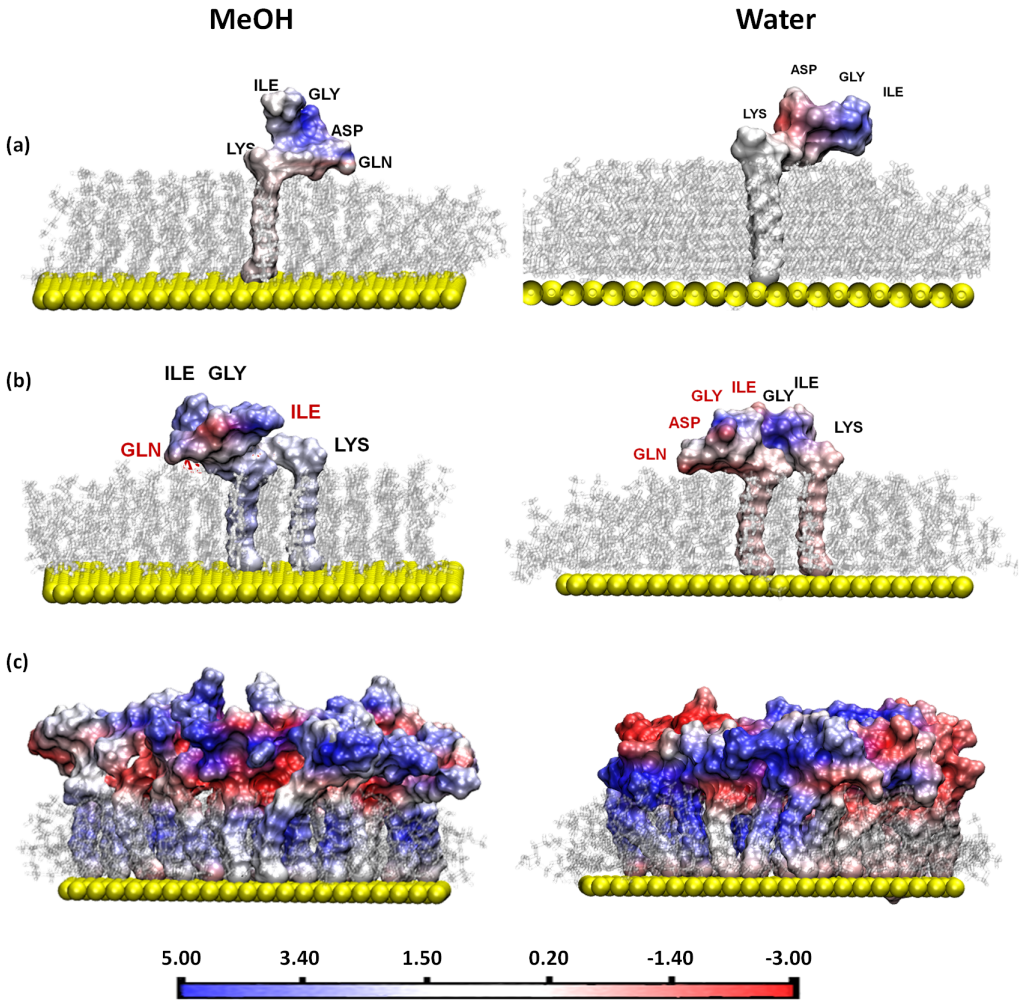


Figure 4.15. Front view of electrostatic surface potential plotted on the last snapshot of the **1** (a), **2** (b) and ***∂*IGDQK-S·Au/MT(PEG)$_4$-S·Au** (c) MD simulations in MeOH (left column) and water (right column). The scale bar at the bottom is also reported. The aa residues within the peptide structures (a and b) are labeled with their symbol name, namely Isoleucine (Ile), Glycine (Gly), Aspartic acid (Asp), Glutamine (Gln) and Lysine (Lys).

As observed for the systems simulated in MeOH, also in water the <Rg> values refer to extended conformations (Rg_{max} = 8.7 Å for **IGDQK-SH/MeOH**), but actually correspond to closed arrangement of **IGDQK-S·Au** molecules stably anchored to the PEG surface (Figure 4.15a-b). This can be more easily appreciated by the visual inspection of **1** and **2IGDQK-S·Au/MT(PEG)$_4$-S·Au** systems, where the specific orientation of the residues within the chemotactic sequence can be also localized through the electrostatic surface potential plotted on their structures (Figure 4.15a-b). In these cases the solvent exposition of the hydrophobic portion (blue regions) is

slightly enhanced, while a heterogeneous hydrophobic/hydrophilic distribution is observed in **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** (Figure 4.15c).

As reported previously, the energetic analysis was carried out only on the gradient **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** system in both the solvents. At this respect, the total non-bonded interaction energy of this system both in MeOH and water results unfavorable (<Tot> = 1333 ± 156 and 2066 ± 233 kcal mol^{-1} respectively, as reported in Table 4.5). Consistently with what observed for **∂IGDQK-S·Au**, the total energy is determined by the EEL contribution, whose positive values (<EEL> = 3244 ± 122 and 4132 ± 222 kcal mol^{-1} for **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** in MeOH and water respectively, Table 4.5) are probably due to the repulsive O-O interactions among the PEG chains. However the simulated gradients with PEG show an improved degree of SAM organization that is reflected in the increased vdW contribution as compared to the full **∂IGDQK-S·Au** systems (<vdW> = -644 and -1911 in MeOH, -975 and -2066 kcal mol^{-1} in water for **∂IGDQK-S·Au** and **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** respectively, Table 4.5), most likely accounting for the interchains interactions.

From a structural point of view, calculated SASA aids in visualizing the actual reduction of the exposed peptide surface as a further measure of its availability. The SASA variation of the simulated systems passing from MeOH to water reveals a remarkable reduction of the exposed surface for all the three systems (Table 4.3 and Figure A.10c, *Appendix*), caused by the adopted conformation tangential to the underlying PEG SAM (Figure 4.15). In particular, the **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** system varies from 11056 to 7770 A^2 of exposed surface, both the values significantly lower than those of the full **∂IGDQK-S·Au** system (16466 to 10585 A^2 from MeOH to water, Table 4.3). However it is worth recalling that the SASA is calculated on the entire **IGDQK-SH** molecule that is thus intrinsically more exposed in **∂IGDQK-S·Au** missing all the surrounding backfilled SAM. Hence the highest <SASA> values detected in **∂IGDQK-S·Au** are in part explained by the intrinsic major exposition of the peptide in that system, in part are due to the ineffective backfilling strategy (**IGDQK-S·Au/MT(PEG)$_4$-S·Au**) in enhancing the peptide sequence disclosure.

In short, **IGDQK-S·Au/MT(PEG)$_4$-S·Au** at different concentration (**1-2-∂ IGDQK-S·Au/MT(PEG)$_4$-S·Au**) results anchored to the underlying PEG SAM due to the formation of H-bonds between Lys and PEG residues. Additionally, interpeptide interactions come into play at higher concentration, which induce a

rearrangement process of the peptide that collapses on itself, thus impeding a proper exposition of the sequence as demonstrated by the calculated Rg and SASA (Table 4.4). The hydrophobic exposition is indeed barely increased only at low concentration of peptides (**1-2IGDQK-S·Au/MT(PEG)$_4$-S·Au**), whereas the gradient simulation yields a heterogeneous polarity of the resulting surface. The overall analysis is in good agreement with the experimental evidences that disclose a hydrophilic surface, poorly organized and with the sequence exposition insufficient to accomplish the cell recognition and migration processes (Figures 4.9-10).

4.6.4 SAMs of IGDQK-S·Au / C_8H_{17}-S·Au

As last attempt in the view of achieving a well-organized and effective motogenic patterned surface, *mixed* SAMs of **IGDQK-S·Au** backfilled with the hydrophobic **C_8H_{17}-S·Au** molecular filler were produced and tested experimentally (Figures 4.11-12). The systematic computational investigation was therefore completed by performing MD simulations on the **1-2** and **∂IGDQK-S·Au/C_8H_{17}-S·Au** systems in MeOH and water as assembling and biological test conditions.

The systems with one and two peptide molecules surrounded by a SAM of **C_8H_{17}-S·Au** in MeOH display <RMSD> values in the same order as the PEG systems (3.5 and 3.9 Å, respectively, Table 4.3 and Figure A.11b, *Appendix*), meaning that regardless the type of filler, its presence suffices to significantly dump the fluctuation of the peptides. However it is worth noting that the simulated gradient **∂IGDQK-S·Au/C_8H_{17}-S·Au** shows <RMSD> value of 6.9 ± 1.0 Å, slightly higher than the **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** system (5.6 ± 0.8 Å, Table 4.3), implying a major degree of freedom of the peptides in the octanthiol environment. This behavior is ascribed to the absence of the anchoring H-bonds between the peptide and the underlying filler SAM. Nevertheless, in **2IGDQK-S·Au/C_8H_{17}-S·Au** the interpeptide interaction between Asp-Lys comes into occur with an average % of occupancy equal to 6% that increases up to 23% in **∂IGDQK-S·Au/C_8H_{17}-S·Au**, along with many other H-bonding pairs (Table 4.4).

When the three systems are then switched to water, the <RMSD> values are generally higher as a confirmation of the conformational freedom of the peptides within the **C_8H_{17}-S·Au** SAM. Notably, the **IGDQK-S·Au** peptides mixed with the

C_8H_{17}-S·Au molecular filler maintain the mobility of the exposed motogenic sequence emerging from the SAM, but at the same time the ability of entertaining interpeptide interactions. In **2IGDQK-S·Au/C_8H_{17}-S·Au** Gln-Gln, Asp-Gly and Asp-Ile H-bonding pairs are encountered with 14.7, 5.8 and 4.4 % of average occupancy; whereas in **∂IGDQK-S·Au/C_8H_{17}-S·Au** the network of H-bonds is expanded as a function of the concentration of the peptide displaying among the most occupied H-bonding pairs the Asp-Asp (30%) and the Asp-Ile pairs (16%, Table 4.4).

From the energetic point of view, the evaluation of the non-bonded interaction energies reveals that among all the gradient simulated systems **∂IGDQK-S·Au/C_8H_{17}-S·Au** presents the highest <Tot> energy both in MeOH (-5429 ± 124, 1333 ± 156 and -6011 ± 162 kcal mol^{-1} for **∂IGDQK-S·Au**, **∂IGDQK-S·Au/MT$(PEG)_4$-S·Au** and **∂IGDQK-S·Au/C_8H_{17}-S·Au** respectively) and in water (-4818 ± 180, 2066 ± 233 and -5480 ± 159 kcal mol^{-1}, Table 4.5) thanks to favorable EEL and vdW contributions accounting respectively for H-bonds and the side and alkyl chains interactions (Table 4.5). Although qualitative, those values indicate a remarkable major stability of the *mixed* octanthiol SAM over the others due to the ideal combination of favorable electrostatic and vdW interactions.

In order to analyze the structural aspects of the pentapeptide immersed in the **C_8H_{17}-S·Au** SAM, the Rg was also calculated along the trajectories both in MeOH and water (Table 4.3 and Figure A.11a, *Appendix*). The **1IGDQK-S·Au/C_8H_{17}-S·Au** system shows similar <Rg> values as the systems with PEG (8.2 and 8.3 Å respectively), that although refers to an extended conformation (8.7 Å as calculated for the peptide in solution), correspond to a closed and poorly exposed arrangement. On the contrary to the abovementioned case the geometry orientations are now actually extended towards the solvent rather than stacked on the SAM filler, as Figure 4.16a clearly shows.

Moving to water, despite the <Rg> value of **1IGDQK-S·Au/C_8H_{17}-S·Au** decreases to 6 Å, the majority of the sequence is still stretched towards the solvent clearly exposing the hydrophobic Ile residue as the electrostatic surface potential reveals (Figure 4.16a). Increasing the number of peptide in **2IGDQK-S·Au/C_8H_{17}-S·Au** simulated in MeOH, the <Rg> value of 8.7 effectively corresponds to the extended conformations and the ideal Ile exposition is maintained (Figure 4.16b).

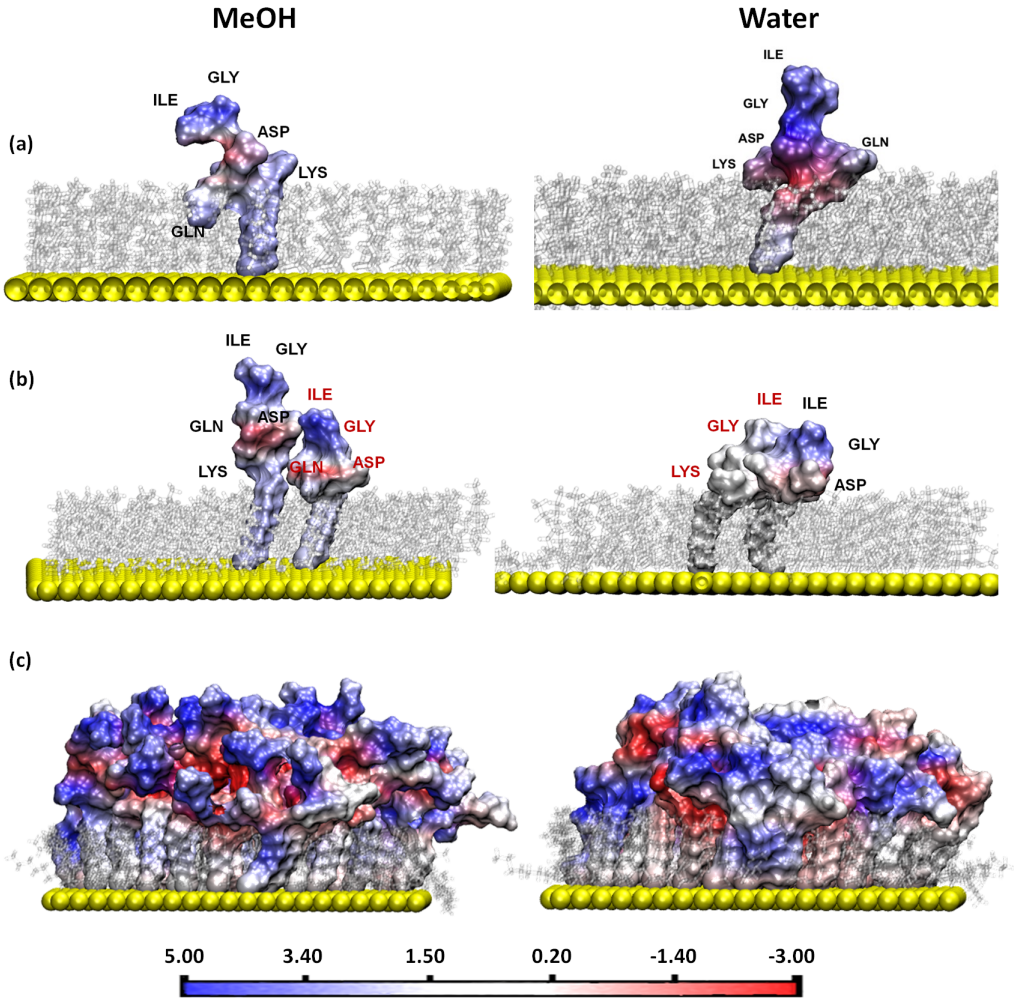


Figure 4.16. Front view of electrostatic surface potential plotted on the last snapshot of the **1** (a), **2** (b) and ***∂*IGDQK-S·Au/C_8H_{17}-S·Au** (c) MD simulations in MeOH (left column) and water (right column). The aa residues within the peptide structure are labeled with their symbol name, namely Isoleucine (ILE), Glycine (GLY), Aspartic acid (ASP), Glutamine (GLN) and Lysine (LYS). The scale bar at the bottom is also reported.

However when the system is simulated in water, the two molecules tend to lie on the SAM surface (Figure 4.16b) due to the intermolecular interactions occurring especially within the Gln-Gln pair (14% of average occupancy in Table 4.4). Additionally, it has been reported that Lys and Gln, although polar aa, bearing a long alkyl side chains, can establish hydrophobic interactions with alkanethiol SAM,[86] prompting this might also occur with our **C_8H_{17}-S·Au** SAM. Nevertheless, at low peptide concentration a better exposition and readability of the motogenic sequence is

now achieved as the hydrophobic Ile are visibly pointing towards the solvent (Figure 4.16a-b). When it comes into the evaluation of the peptide exposition in the gradient **∂IGDQK-S·Au/C_8H_{17}-S·Au**, it is worth noting that the <Rg> value of the system simulated in MeOH is significantly higher than that of both the *mixed* PEG SAM and the full peptide SAM (22.5, 20.5 and 21.4 Å respectively), and it remains unvaried also in the simulation in water (22.5, 20.2 and 20.2 Å respectively, Table 4.3).

The major sequence exposition of the **∂IGDQK-S·Au/C_8H_{17}-S·Au** system as calculated by the SASA nicely matches with the experimentally observed biological activity of such patterned surface with respect to **∂IGDQK-S·Au/MT$(PEG)_4$-S·Au** (Figure 4.12 and 4.10 respectively).

Finally, the SASA analysis unambiguously confirms the effectiveness of the octanthiol filler in improving the sequence availability. Basically, passing from the simulations in MeOH to those in water it is observed the reduction of the exposed peptide surface, measured as the <SASA> (Table 4.3 and Figure A.11c, *Appendix*), at any peptide concentration (**1-2-∂**), which is induced by the rearrangement of the peptides trying to entertain intermolecular interactions. Noteworthy, the *mixed* octanthiols SAMs display significantly higher exposed surface than the *mixed* PEG SAMs both in MeOH (15412 *vs* 11056 A^2 respectively, Table 4.3) and in water (11539 and 7770 A^2). Furthermore, taking into account that in the **∂IGDQK-S·Au** system the exposed surface is intrinsically higher due to the absence of the molecular fillers, the <SASA> values are surprisingly similar when compared to the **∂IGDQK-S·Au/C_8H_{17}-S·Au** in MeOH (16466 ± 748 and 15412 ± 1187 A^2 respectively) and in water (10585 ± 904 and 11539 ± 641 A^2 respectively). This suggests that the octanthiol filler effectively ameliorates the peptide exposition as compared to the other attempts.

Summarizing, when the **IGDQK-S·Au** peptide is immersed in a **C_8H_{17}-S·Au** SAM at low concentration (**1-2IGDQK-S·Au/C_8H_{17}-S·Au**) it perfectly exposes its sequence, pointing especially the hydrophobic Ile towards the solvent bulk (both MeOH and water, Figure 4.16a-b). When the peptide concentration increases (**∂IGDQK-S·Au/C_8H_{17}-S·Au**) it was proved a general major peptide surface exposition by calculating the Rg and SASA. Therefore it emerges that the octanthiol filler effectively aids in enhancing the peptide availability also among the simulated systems confirming the empirical data collected from the AFM, WCA measurements and biological tests.

4.6.5 Sequence readability in the gradient SAMs

In the above sections, it has been demonstrated that the gradient of **∂IGDQK-S·Au** fails in attaining an organized and ordered SAM due to the affinity of the motogenic sequence for the Au surface, which stacks on it rather than provides its structure for the cell recognition. On the other hand, secondary interactions are shown to come into paly in **∂IGDQK-S·Au/ MT(PEG)$_4$-S·Au** that, being responsible of the anchoring of the peptide on the underlying surface, impede an efficient peptide availability. Finally, the SAM **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** was proved to accommodate the ideal condition for the motogenic sequence exhibition that assures the functionality of the patterned surface, as observed experimentally.

However, despite exhaustive and detailed, the analysis carried out so far on the gradient simulated systems does not provide any precise information about the part of the peptide sequence actually visible and prone to exert the motogenic function. Although the Rg and SASA represent a clear proof of the major sequence stretching and disclosure in **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** (Table 4.3), the correct readability of the sequence remains to be addressed. It is worth recalling that within the **IGDQ** sequence the Ile acts as pivotal residue for the correct biological recognition and activity,[58] thereof considered the target aa for the readability of our peptide sequence.

In this framework, a further analysis was performed to clarify the evident discrepancy in the behavior of the peptide in the three simulated gradient, namely **∂IGDQK-S·Au, ∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** and **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au**, and verify whether the observed major exposition correspond to the actual and correct readability of the peptidic sequence. Aiming at this, the radial distribution functions (RDFs) were calculated only for the gradient systems simulated in water, since it is pivotal in the biological test conditions to demonstrate the correct sequence recognition. At first, the RDFs were calculated between the water O and the contact surface (the exposed gold atoms in **∂IGDQK-S·Au** and the C atom of the terminal methyl of the **C$_8$H$_{17}$-S·Au** or of the **MT(PEG)$_4$-S·Au** in **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** and **MT(PEG)$_4$-S·Au**, respectively) thereby giving an idea of the degree of solvation of the gradient patterned surfaces. In fact, the higher the RDFs, the biggest are the

water-surface interactions, while the most the RDFs decrease, the most water-surface interactions are replaced by peptide-surface ones.[87]

Notably the RDF at 3.25 (calculated distance with a major water density, Figure A.12, *Appendix*) for **∂IGDQK-S·Au**, at 3.75 Å for **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** and **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** are 0.9, 49.3 and 25.3 in height, respectively. This means that the **∂IGDQK-S·Au** surface (Au atoms) is completely hampered by the interaction with the peptide, while the *mixed* SAMs are exposed to the solvent, meaning the peptide is more erected. Among the *mixed* SAMs, **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** unveils the higher water density (49.3 in height) with respect to **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** (25.3 in height) due to the more hydrophilic nature of the PEG portion.

The above analysis addresses the relationship between the peptides and the surface and the consequent degree of surface exposition to the solvent considering the water-surface and peptide-surface interactions. Furthermore, from this analysis could be defined the distance with the higher water density in the three systems. Now, in order to address the water-peptide interactions as a measure of the sequence exposition to the solvent, the RDFs are now calculated for each residue separately at the defined distances (3.25 for **∂IGDQK-S·Au** and 3.75 Å for **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** and **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au**) between the water O and respectively the C of the terminal Ile of the **IGDQK-S·Au** molecules, the C atom of the terminal methyl group of **C$_8$H$_{17}$-S·Au** and **MT(PEG)$_4$-S·Au**, and between the water O and the Au atoms when no fillers are present. Doing so it has been possible to measure the patterned surfaces-water interactions and picture a precise and punctual water distribution map over the surfaces (Figure 4.17). A RDF map is therefore drawn coloring the Ile, Au and the two fillers according to the relative water density around them, thus the degree of exposition will correspond to low (blue coloration) medium (red) and high (green) water density.

The RDF map of **∂IGDQK-S·Au** shows a lower water density in correspondence of the gold atoms that, completely covered by the peptides, are indeed not water-exposed (Figure 4.17a top, light blue coloration).

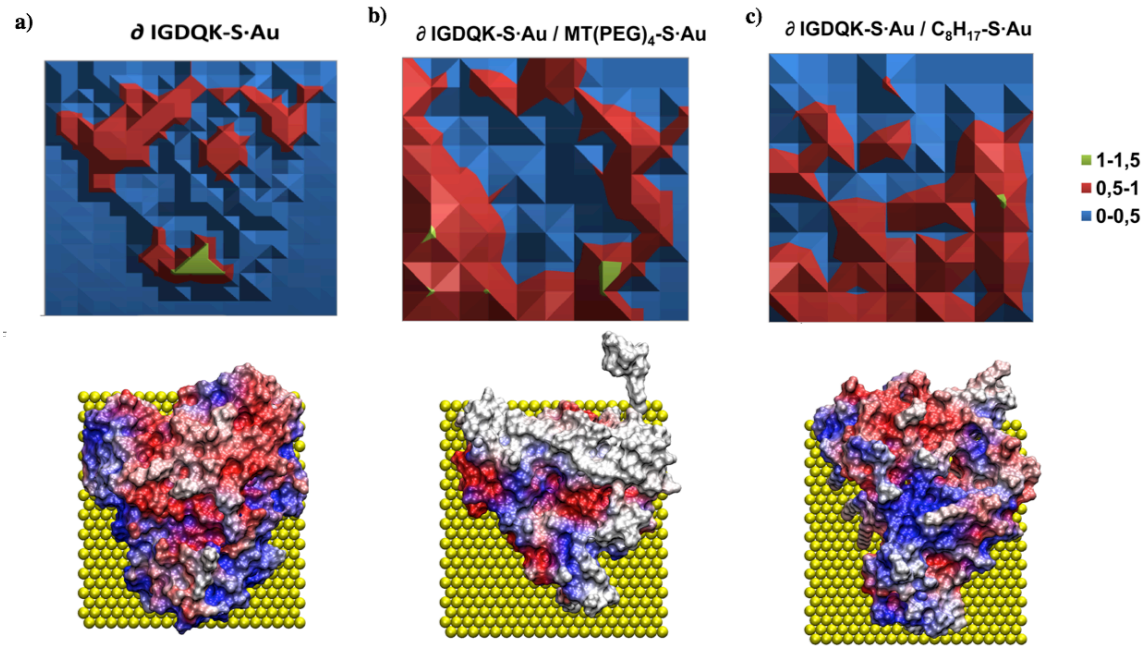


Figure 4.17. Top: RDFs calculated over the 10 ns trajectories at 3.75 Å of distance for **∂IGDQK-S·Au** (a) **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** (b) and **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** (c). From the lowest to the highest water density, the residues within the RDF maps are colored in blue, red and green respectively. Bottom: top view of the electrostatic surface potential plotted on the last snapshot of the **∂IGDQK-S·Au** (a) **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** (b) and **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** (c) MD simulations in water.

On the other hand, the presence of the peptide defines a distinct solvent density cone, which is slightly more exposed where the Ile are oriented towards the solvent (higher exposition red-green coloration), albeit the predominance of the blue coloration still evidences a poor exposition of the Ile residues. Knowing that the Ile exposition is fundamental for the motogenic activity of these patterned surfaces, these findings nicely correlate to the absence of biological response when **∂IGDQK-S·Au** SAM was tested in the presence of cancer cells and HDF (Figure 4.7). Additionally, looking at the top view of the electrostatic surface potential plotted on the last snapshot of the **∂IGDQK-S·Au** simulation in water, it appears evident the hydrophilic nature of the resulting surface (Figure 4.17a bottom, red coloration). It was proved experimentally that **∂IGDQK-S·Au** produces a homogeneous hydrophilic surface as measured by the WCA along the gradient (ΔWCA = -4.9°, Table 4.1), now further confirmed by the RDF map.

Concerning the **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** system a red-green coloration is observed in correspondence mainly of the lateral PEG molecules (Figure 4.17b

top), while the central area referring to the peptides gradient is mainly blue-colored suggesting a poor solvent exposition of the Ile at the increasing of the peptide concentration. This slightly increased Ile exposition (red coloration at low concentration, Figure 4.17b) reasonably explains the fractional migration of the cancer cells when deposited on such SAM (Figure 4.10). These findings in addition to the top view of the electrostatic surface potential of the *mixed* PEG SAM, showing a neutral coloration (Figure 4.17b bottom), confirm the poor hydrophobicity of the resulting surface described by experimental evidences. Indeed, RDF map and electrostatic surface potential are in excellent agreement with the WCA measurements that reveal an imperceptible variation in polarity along the **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** gradient (ΔWCA = 7.1°, Table 4.1).

On the other hand, the RDF map of **∂IGDQK-S·Au/C$_8$H$_{17}$-S·Au** displays a diffused red coloration, a clear sign of the water exposition of both octanethiol and Ile residues progressively exposed as a function of the peptides concentration (Figure 4.17c top). The MD simulation shows that this patterned surface remarkably exposes the Ile residues therefore yielding a hydrophobic surface as observed in the blue coloration of the electrostatic surface potential (Figure 4.17c bottom). These results are in excellent agreement with the experimental WCA values that unveil a significant variation in polarity along the gradient (ΔWCA = 22.5°, Table 4.1). Hence, the remarkable motogenic activity of the **IGDQK-S·Au** backfilled with **C$_8$H$_{17}$-S·Au** is explained by the correct orientation of the motogenic sequence, starting with Ile, as demonstrated by this *in silico* validation.

In conclusion, the RFD map was herein used as valuable and suitable analysis to precisely elucidate at the atomistic level the polarity of the produced SAMs correlating with the experimental WCA measurements. By punctually quantifying the Ile readability along the different gradients, though the RDF maps we were able to identify a structural-activity relationship ultimately clarifying the divergent biological responses of the produced SAMs as shown in section 4.3. Therefore the hydrophilicity of **∂IGDQK-S·Au** can be ascribed to the insufficient Ile orientation towards the water (Figure 4.17a), which reasonably causes no motogenic response of cancer cells deposited on such surface. The map of the **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** system unveils a negligible increase of the Ile water orientation, which from one side confirms the WCA measurements, and from the other, provides a surface that

hardly suffices to attain a considerable cell migration. In the same measure, it was nicely demonstrated that the improved performance of the **∂IGDQK-S·Au/C_8H_{17}-S·Au** in the organization and exposition of the motogenic sequence is essentially due to the major and correct orientation of the Ile residues. Hence the hydrophobic nature of the **∂IGDQK-S·Au/C_8H_{17}-S·Au**, as observed experimentally and computationally, accounts for the exposition of the Ile aa thereby providing an effective motogenic surface.

4.7 Concluding remarks

We investigate the self-assembling behaviour of Ile-Gly-Asp (IGD)-containing peptide on Au as new motogenic surface studying the migratory response of cancer cells and HDF. **IGDQK-SH** pentapeptide was used to produce full (**∂IGDQK-S·Au**) and *mixed* SAMs, the latter prepared by backfilling the peptide SAM with two different molecular fillers (**MT(PEG)$_4$-S·Au** or **C_8H_{17}-S·Au**), obtaining the **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** and **∂IGDQK-S·Au/C_8H_{17}-S·Au** systems. Both structural (AFM and WCA analysis) and functional (biological tests) characterizations reveal that the poor SAM organization can drastically affect the motogenic response as very different behaviors were observed for the three produced SAMs. In particular, only when interfacing with **∂IGDQK-S·Au/C_8H_{17}-S·Au** cancer cells exhibit selective migratory response.

We therefore performed classical MD simulations of the three produced systems both in the assembling (MeOH) and in the biological test conditions (water). By means of a thorough trajectory analysis we have been able to demonstrate that the efficient motogenic activity of **∂IGDQK-S·Au/C_8H_{17}-S·Au** is correlated to the proper organization and correct sequence orientation pointing the Ile residues towards the solvent. The simulated SAM in the presence of the octanthiol filler shows a well-organized hydrophobic surface that is in excellent agreement with the empirical data (AFM and WCA), while the correct structural arrangement, in particular of the Ile residues, clarifies the motogenic activity. On the contrary, the MD simulations demonstrated that the poor organization of the full **IGDQK-S·Au** SAM is caused by the tendency of the motogenic sequence, at low peptide concentration (**1IGDQK-S·Au**), to lye down on the Au surface, whereas at the increase of the peptide

concentration (**2-∂IGDQK-S·Au**) inter-peptides interactions impede the correct sequence exposition. These findings explain the absence of the biological activity of such system with atomistic level of details. Finally, the inefficient motogenic response of **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** finds explanation in the secondary interactions occurring between the peptides molecules and the underlying PEG fillers. Peptide-PEG H-bonds are therefore responsible of the anchoring of the peptide on the underlying surface, impeding the efficient peptide availability.

By this way we have been able to complement and validate the experimental outcomes explaining with atomistic detailed information the structural-activity relationship of the prepared peptide-based motogenic Au surface.

4.8 References

[1] I. Willner, B. Willner, *Nano Lett.* **2010**, *10*, 3805–3815.

[2] A. J. Makarucha, N. Todorova, I. Yarovsky, *Eur. Biophys. J.* **2011**, *40*, 103–115 .

[3] M. Mahmoudi, I. Lynch, M. R. Ejtehadi, M. P. Monopoli, F. B. Bombelli, S. Laurent, *Chem. Rev.* **2011**, *111*, 5610–5637.

[4] M. Sarikaya, C. Tamerler, A. K.-Y. Jen, K. Schulten, F. Baneyx, *Nat. Mater.* **2003**, *2*, 577–585.

[5] A. J. Haes, W. P. Hall, L. Chang, W. L. Klein, R. P. Van Duyne, *Nano Lett.* **2004**, *4*, 1029–1034.

[6] K. E. Sapsford, W. R. Algar, L. Berti, K. B. Gemmill, B. J. Casey, E. Oh, M. H. Stewart, I. L. Medintz, *Chem. Rev.* **2013**, *113*, 1904–2074.

[7] Y.-C. Yeh, B. Creran, V. M. Rotello, *Nanoscale* **2012**, *4*, 1871–1880.

[8] F. X. Gu, R. Karnik, A. Z. Wang, F. Alexis, E. Levy-Nissenbaum, S. Hong, R. S. Langer, O. C. Farokhzad, *Nano Today* **2007**, *2*, 14–21.

[9] Y. Jin, C. Jia, S.-W. Huang, M. O'Donnell, X. Gao, *Nat. Commun.* **2010**, *1*, 41.

[10] S. C. McBain, H. H. P. Yiu, J. Dobson, *Int. J. Nanomed.* **2008**, *3*, 169–180.

[11] P. Zijlstra, P. M. R. Paulo, M. Orrit, *Nat. Nanotechnol.* **2012**, *7*, 379–382.

[12] A. J. Haes, R. P. Van Duyne, *Anal. Bioanal. Chem.* **2004**, *379*, 920–930.

[13] T. Endo, S. Yamamura, N. Nagatani, Y. Morita, Y. Takamura, E. Tamiya, *Sci. Technol. Adv. Mater.* **2005**, *6*, 491–500.

[14] S.-T. Yang, Y. Liu, Y.-W. Wang, A. Cao, *Small* **2013**, *9*, 1635–1653.

[15] M. Rabe, D. Verdes, S. Seeger, *Adv. Colloid Interf. Sci.* **2011**, *162*, 87–106.

[16] J. J. Gray, *Curr. Opin. Struct. Biol.* **2004**, *14*, 110–115.

[17] K. Nakanishi, T. Sakiyama, K. Imamura, *J. Biosci. Bioeng.* **2001**, *91*, 233–244.

[18] R. Di Felice, S. Corni, *J. Phys. Chem. Lett.* **2011**, *2*, 1510–1519.

[19] H. Heinz, B. L. Farmer, R. B. Pandey, J. M. Slocik, S. S. Patnaik, R. Pachter, R. R. Naik, *J. Am. Chem. Soc.* **2009**, *131*, 9704–9714.

[20] J. Yu, M. L. Becker, G. A. Carri, *Langmuir* **2012**, *28*, 1408–1417.

[21] J. Yu, M. L. Becker, G. A. Carri, *Small* **2010**, *6*, 2242–2245.

[22] R. L. Willett, K. W. Baldwin, K. W. West, L. N. Pfeiffer, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7817–22.

[23] M. Hoefling, F. Iori, S. Corni, K.-E. Gottschalk, *Langmuir* **2010**, *26*, 8347–8351.

[24] M. Hoefling, F. Iori, S. Corni, K.-E. Gottschalk, *ChemPhysChem* **2010**, *11*, 1763–1767.

[25] G. Nawrocki, M. Cieplak, *J. Phys. Chem. C* **2014**, *118*, 12929–12943.

[26] Z. E. Hughes, L. B. Wright, T. R. Walsh, *Langmuir* **2013**, *29*, 13217–13229.

[27] L. Bellucci, S. Corni, *J. Phys. Chem. C* **2014**, *118*, 11357–11364.

[28] M. Hoefling, S. Monti, S. Corni, K. E. Gottschalk, *PLoS One* **2011**, *6*, e20925.

[29] M. J. Penna, M. Mijajlovic, M. J. Biggs, *J. Am. Chem. Soc.* **2014**, *136*, 5323–5331.

[30] C.-L. Chen, N. L. Rosi, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 1924–1942.

[31] J. M. Slocik, R. R. Naik, *Adv. Mater.* **2006**, *18*, 1988–1992.

[32] S. Corni, M. Hnilova, C. Tamerler, M. Sarikaya, *J. Phys. Chem. C* **2013**, *117*, 16990–17003.

[33] Z. Tang, J. P. Palafox-Hernandez, W.-C. Law, Z. E. Hughes, M. T. Swihart, P. N. Prasad, M. R. Knecht, T. R. Walsh, *ACS Nano* **2013**, *7*, 9632–9646.

[34] U. O. S. Seker, B. Wilson, J. L. Kulp, J. S. Evans, C. Tamerler, M. Sarikaya, *Biomacromolecules* **2014**, *15*, 2369–2377.

[35] C. R. So, J. L. Kulp, E. E. Oren, H. Zareie, C. Tamerler, J. S. Evans, M. Sarikaya, *ACS Nano* **2009**, *3*, 1525–1531.

[36] A. Calzolari, G. Cicero, C. Cavazzoni, R. Di Felice, A. Catellani, S. Corni, *J. Am. Chem. Soc.* **2010**, *132*, 4790–4795.

[37] K. Kubiak-Ossowska, P. A. Mulheran, *Langmuir* **2010**, *26*, 7690–4.

[38] J. Bardeen, *Phys. Rev.* **1940**, *58*, 727–736.

[39] M. W. Finnis, *Surf. Sci.* **1991**, *241*, 61–72.

[40] R. Guidelli, W. Schmickler, *Electrochim. Acta* **2000**, *45*, 2317–2338.

[41] K. Uvdal, P. Bodö, B. Liedberg, *J. Colloid Interf. Sci.* **1992**, *149*, 162–173.

[42] Q.-M. Xu, L.-J. Wan, C. Wang, C.-L. Bai, Z.-Y. Wang, T. Nozawa, *Langmuir* **2001**, *17*, 6203–6206.

[43] R. Di Felice, A. Selloni, E. Molinari, *J. Phys. Chem. B* **2003**, *107*, 1151–1156.

[44] R. Di Felice, A. Selloni, *J. Chem. Phys.* **2004**, *120*, 4906–4914.

[45] Q. Chi, J. Zhang, J. U. Nielsen, E. P. Friis, I. Chorkendorff, G. W. Canters, J. E. T. Andersen, J. Ulstrup, *J. Am. Chem. Soc.* **2000**, *122*, 4047–4055.

[46] M. D. Porter, T. B. Bright, D. L. Allara, C. E. D. Chidsey, *J. Am. Chem. Soc.* **1987**, *109*, 3559–3568.

[47] H. A. Biebuyck, C. D. Bain, G. M. Whitesides, *Langmuir* **1994**, *10*, 1825–1831.

[48] J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo, G. M. Whitesides, *Chem. Rev.* **2005**, *105*, 1103–1169.

[49] X. Jiang, D. A. Bruzewicz, A. P. Wong, M. Piel, G. M. Whitesides, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 975–978.

[50] G. Mahmud, C. J. Campbell, K. J. M. Bishop, Y. A. Komarova, O. Chaga, S. Soh, S. Huda, K. Kandere-Grzybowska, B. A. Grzybowski, *Nat. Phys.* **2009**, *5*, 606–612.

[51] P. Friedl, D. Gilmour, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 445–457.

[52] P. Friedl, J. Locker, E. Sahai, J. E. Segall, *Nat. Cell Biol.* **2012**, *14*, 777–783.

[53] S. E. Carsons, *Fibronectin in Health and Disease*, CRC Press, Florida, **1989**.

[54] M. C. Erat, B. Sladek, I. D. Campbell, I. Vakonakis, *J. Biol. Chem.* **2013**, *288*, 17441–17450.

[55] M. C. Erat, U. Schwarz-Linek, A. R. Pickford, R. W. Farndale, I. D. Campbell, I. Vakonakis, *J. Biol. Chem.* **2010**, *285*, 33764–33770.

[56] I. R. Ellis, S. J. Jones, D. Staunton, I. Vakonakis, D. G. Norman, J. R. Potts, C. M. Milner, N. A. G. Meenan, S. Raibaud, G. Ohea, A. M. Schor, S. L. Schor, *Exp. Cell Res.* **2010**, *316*, 2465–2476.

[57] S. L. Schor, I. Ellis, J. Banyard, A. M. Schor, *J. Cell Sci.* **1999**, *112*, 3879–3888.

[58] C. J. Millard, I. R. Ellis, A. R. Pickford, A. M. Schor, S. L. Schor, I. D. Campbell, *J. Biol. Chem.* **2007**, *282*, 35530–35535.

[59] X. Houard, S. Germain, M. Gervais, A. Michaud, F. van den Brûle, J.-M. Foidart, A. Noël, C. Monnot, P. Corvol, *Int. J. Cancer* **2005**, *116*, 378–84.

[60] C. R. So, C. Tamerler, M. Sarikaya, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 5174–5177.

[61] R. Levicky, T. M. Herne, M. J. Tarlov, S. K. Satija, *J. Am. Chem. Soc.* **1998**, *120*, 9787–9792.

[62] K. Leufgen, M. Mutter, H. Vogel, W. Szymczak, *J. Am. Chem. Soc.* **2003**, *125*, 8911–8915.

[63] E. Ostuni, R. G. Chapman, R. E. Holmlin, S. Takayama, G. M. Whitesides, *Langmuir* **2001**, *17*, 5605–5620.

[64] K. Prime, G. Whitesides, *J. Am. Chem. Soc.* **1993**, *115*, 10714 – 10721.

[65] P. Harder, M. Grunze, R. Dahint, G. M. Whitesides, P. E. Laibinis, *J. Phys. Chem. B* **1998**, *102*, 426–436.

[66] R. A. Latour, *Biointerphases* **2008**, *3*, FC2–FC12.

[67] N. Schwierz, D. Horinek, S. Liese, T. Pirzer, B. N. Balzer, T. Hugel, R. R. Netz, *J. Am. Chem. Soc.* **2012**, *134*, 19628–19638.

[68] D. Pellenc, R. A. Bennett, R. J. Green, M. Sperrin, P. A. Mulheran, *Langmuir* **2008**, *24*, 9648–55.

[69] P. Mulheran, K. Kubiak, *Mol. Simul.* **2009**, *35*, 561–566.

[70] L. Strong, G. M. Whitesides, *Langmuir* **1988**, *4*, 546–558.

[71] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269–10280.

[72] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, AMBER12, University of California, San Francisco **2012**.

[73] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712–725.

[74] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[75] S. L. Mayo, B. D. Olafson, W. A. Goddard, *J. Phys. Chem.* **1990**, *94*, 8897–8909.

[76] J. B. Schlenoff, M. Li, H. Ly, *J. Am. Chem. Soc.* **1995**, *117*, 12528–12536.

[77] D. M. Collard, M. A. Fox, *Langmuir* **1991**, *7*, 1192–1197.

[78] Y.-H. Kim, S. S. Jang, W. A. Goddard, *J. Chem. Phys.* **2005**, *122*, 244703.

[79] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[80] J. W. Caldwell, P. A. Kollman, *J. Phys. Chem.* **1995**, *99*, 6208–6219.

[81] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[82] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.

[83] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.

[84] L. B. Wright, N. A. Merrill, M. R. Knecht, T. R. Walsh, *ACS Appl. Mater. Interfaces* **2014**, *6*, 10524–33.

[85] V. P. Raut, M. A. Agashe, S. J. Stuart, R. A. Latour, *Langmuir* **2005**, *21*, 1629–1639.

[86] W. Soliman, S. Bhattacharjee, K. Kaur, *J. Phys. Chem. B* **2010**, *114*, 11292–11302.

[87] J. Zhao, Q. Wang, G. Liang, J. Zheng, *Langmuir* **2011**, *27*, 14876–14887.

Conclusions and perspectives

Modified graphitic nanomaterial (GNMs, such as fullerene, carbon nanotubes (CNTs) and graphene) and inorganic surfaces, have recently emerged as new multifunctional platforms for application in nanomedicine and biology. The idea of merging nanotechnology and biology represents nowadays a fascinating avenue to confer a new functional essence to GNMs through chemical modification with biomolecules, *i.e.* peptides, proteins or enzymes. Additionally, this approach allows overcoming one of the main barrier to exploit GNMs potentialities, that is their insolubility and the difficulties in separating them. Effectively, the biomolecular modification of GNMs increases their (poor) solubility and biocompatibility, leading to a resulting biohybrid compound which, thanks to the new functional property, is exploitable for several applications ranging from imaging and bio-sensing, nanomedicine and cancer therapy, gene therapy, drug delivery and tissue engineering. Consequently, the deep knowledge of the free and spontaneous tendency of GNMs to interact with proteins, interpreted as non-covalent interaction driven by their chemical nature, results of pivotal importance for the achievement of a controlled manipulation towards tailored applications. Several experimental and theoretical techniques, as I accurately reported in the first Chapter of this thesis, have been applied to study the non-covalent nanomaterial/proteins interactions ultimately leading to the control and engineering of new materials for biomedical applications with reduced toxicological implications. However, taking into consideration the extremely structural and chemical variety of protein and nanomaterials, responsible of their different ways of interaction, as well as the limitation of computational techniques in efficiently describe such interactions, the full comprehension and prediction of the forces ruling the bio-hybrid formation is still an open issue.

By applying extensive force-field based Molecular Dynamics (MD) and quantum mechanical calculations I herein presented three examples though which I contributed to this research field, probing protein/nanomaterials interfaces.

Specifically, the reader is firstly accompanied through the studying of the structural and dynamical properties of a non-covalent bioconjugate in which the monoclonal Cetuximab antibody (Ab Ctx) is adsorbed on a CNT surface. For the first

time all-atom explicit solvent MD simulations was performed on an entire CNT-Ctx biohybrid of very large size (400000 atoms). Despite the remarkable size and nature of the system we have been able to carry out explorative docking calculations observing a diameter dependent behavior of increasing CNT diameters adsorbed on the Ctx structure. Unfortunately, being the docking program developed for the prediction of small ligand interacting with protein active site, no fully reliable energetic analysis could be attained, whereas the three most representative adsorption modes were selected. The latter were refined by force-field MD and DFT simulations that unambiguously showed that hydrophobic interactions mainly govern the adsorption of the Ctx protein on the graphitic surface. Moreover, the conjugation with the nanomaterial does not affect the secondary structure of Ctx. The predicted structural models are consistent with the experimental data, thus rationalizing the intact structural stability and recognition properties of the Ctx immobilized on CNT. MD simulations clearly support the reliability of the used bioconjugation strategy for engineering stable and responsive hybrid nanomaterials for therapeutic applications. Importantly, our specific structural model can be in principle generalized thanks to a remarkable structural similarity of Ctx with antibodies of different isotypes, prompting that in principle the CNT framework can interact in the same manner with most of antibodies currently used in clinical applications. Although the studies of the Ctx/CNT interface were limited by the large size of the systems (longer simulations and further investigation of the diameter dependent effect might confirm and complete the qualitative results achieved), the biohybrid was accurately tackled giving important insight at atomistic level of details of the forces ruling such interaction.

With a reverse approach the reader is then introduced to the design of a *de novo* Janus-type proteinic β-sheet able to directionally functionalize graphene in the view of biological applications. In particular, I provide computational evidences that a 13-mer polypeptide can fold stably in an antiparallel β-sheet exposing a hydrophobic side, prone to the anchoring to graphene, and a hydrophilic face devoted to functionalization purposes. Additionally, pyrenyl functionalities have been inserted in the hydrophobic side and extensive classical MD simulations demonstrated that their presence was fundamental to improve the β-sheet adsorption on the flat graphitic surface. In fact, the absence of the anchoring moiety yielded a purely aspecific adsorption profile, ultimately valuing our pyrenyl modified β-sheet engineering for the controlled functionalization of graphene. Additionally, the ideal number of pyr per

β-strand was also predicted electing the **6bb2-1pyr** system as the best designed β-sheet for the purpose we have presented herein. With the present Janus-type proteinic β-sheet I achieved the ideation of a new face-selective graphene functionalization approach that following the peculiar 2D nanomaterial geometry would allow its high-coverage non-covalent modification in a controllable and organized manner. The theoretical design is here guiding the production of the β-sheet protein that, in collaboration with Dr. Riccardo Marega and Laure-Elie Carloni of our group, is now under experimental validation. However, despite the β-sheet structural stability was extensively and accurately investigated, the simulated systems resulted highly mobile and still evolving. Thus in parallel to the experimental developments, further theoretical confirmation of the structure stability might be required.

Finally, in Chapter 4 we investigate the self-assembly of Ile-Gly-Asp (IGD)-containing peptide on Au as new motogenic surface studying the migratory response of cancer cells and human dermal fibroblast. IGDQK-SH pentapeptide was used to produce one full and two *mixed* self assembled monolayers (SAMs), the latter prepared by backfilling the peptide SAM with two molecular fillers with different hydrophobicity. Both structural (AFM and WCA analysis) and functional (biological tests) characterizations reveal that different morphologies can drastically affect the motogenic response as very different behaviors are observed for the three produced SAMs. Therefore, classical MD simulations have been performed to model the three produced systems both in the assembling (MeOH) and in the biological test conditions (water). Remarkably, by means of a thorough trajectory analysis we have been able to demonstrate that the efficient motogenic activity is correlated to the proper organization and correct sequence orientation. The systems in which pronounced peptide affinity to the Au surface and interpeptide interactions were observed correspond to poorly organized materials where the sequence exposition was drastically affected and the surface response was compromised. Although all the simulated motogenic surfaces simulated for a relatively short timescale might be consider preliminary study, significant differences have emerged comparing the three SAMs, which, being in excellent agreement with the experimental results, clarified further the motogenic activity with important structural insights. By this way, we have been able to elucidate and complement the experimental outcomes explaining with atomistic detailed information the structural-activity relationship of the prepared peptide-based motogenic Au surface.

Interestingly, in a transversal and interdisciplinary manner, the three projects are developed in collaboration with the experimental team of our group, which contributed with precious experimental evidences. Therefore, the theoretical studies presented herein tackled crucial aspects of the interaction between different biomolecules and nanomaterials working as a powerful complement to either validate, rationalize, predict and guide the experimental findings.

Appendix

Chapter 3

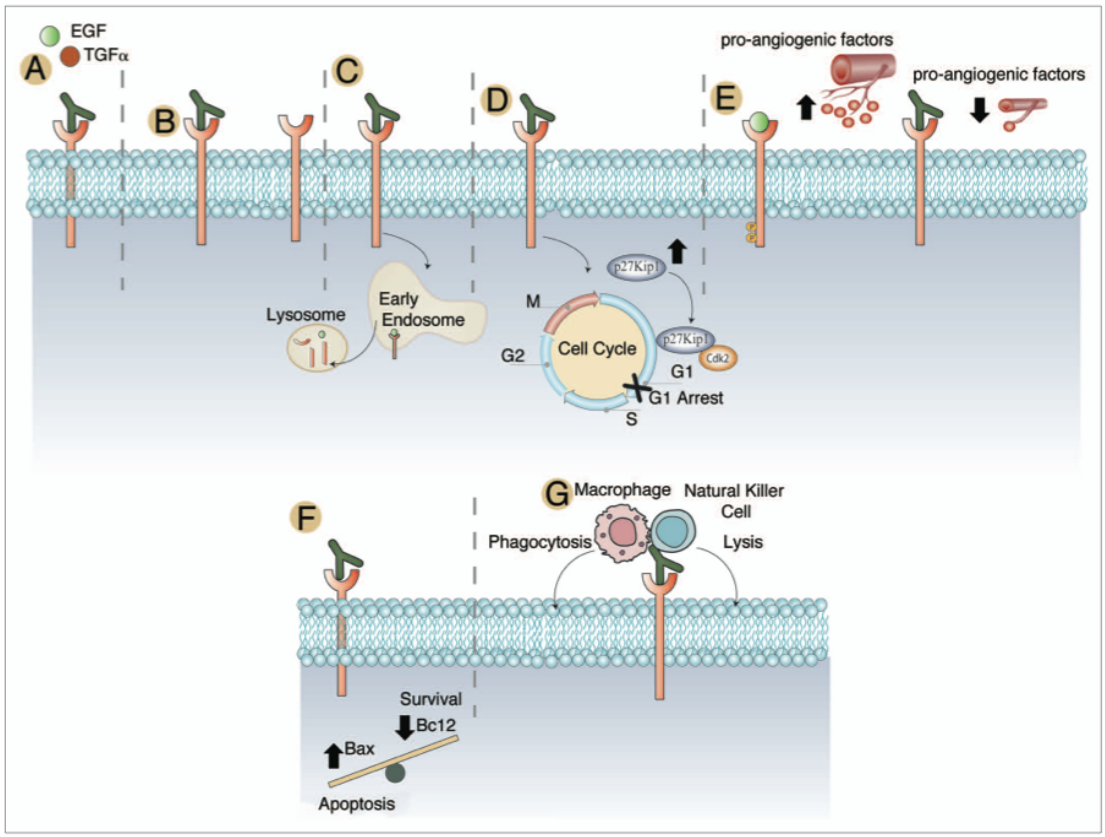


Figure A.1. Mechanism of EGFR inhibition by Ctx where Ctx is shown as a green Y, EGFR is colored in orange: A) endogenous ligand competition; B) sterically inhibition of the dimerization process; C) promotion of receptor internalization and degradation; D) induction of cell cycle arrest in G_1 phase; E) decrease of the expression of pro-angiogenic factors; F) influence of the cell apoptosis and cell survival; G) Ab-dependent cellular cytoxicity. Reprinted with permission from Ref. 26, Copyright © 2011, Landes Bioscience.

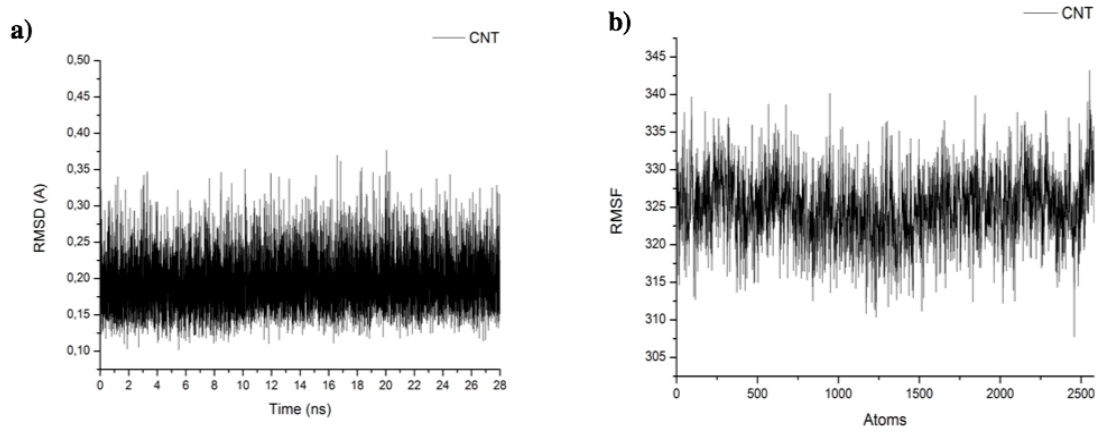
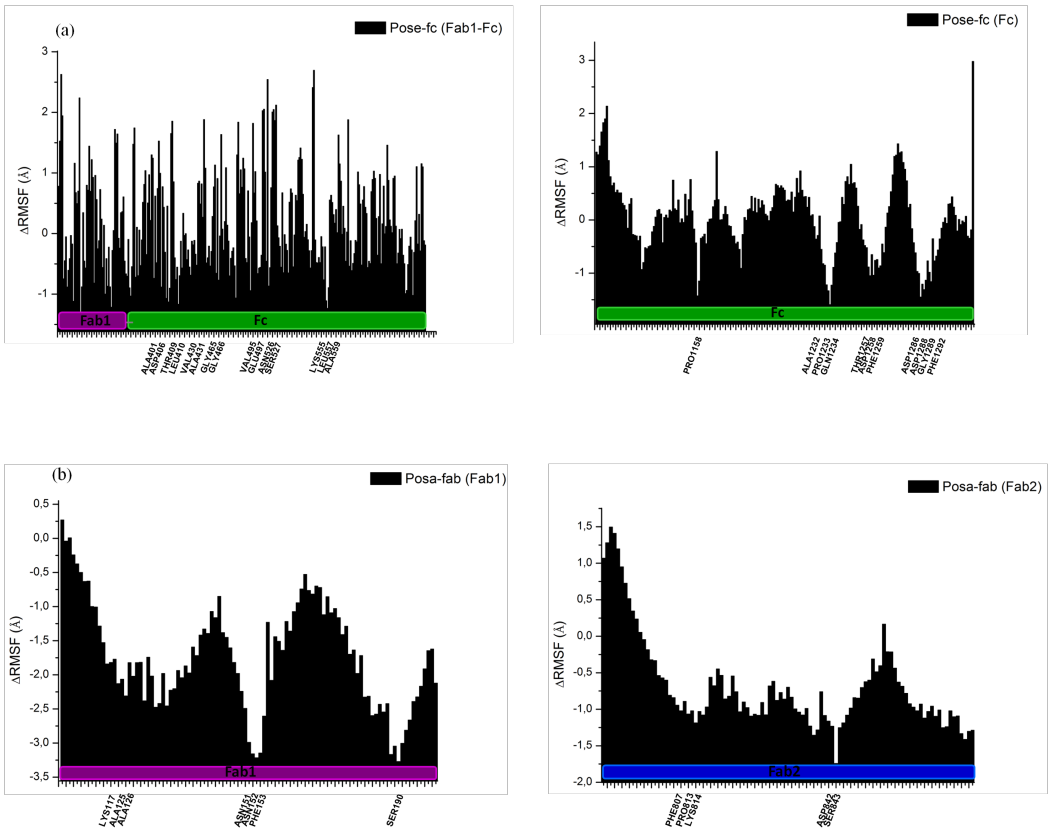


Figure A.2. The RMSD (a) as a function of time calculated along the 28 ns MD simulation of the CNT in explicit water. The RMSF (b) plotted as a function of the atoms number of the CNT along the 28 ns MD simulation.



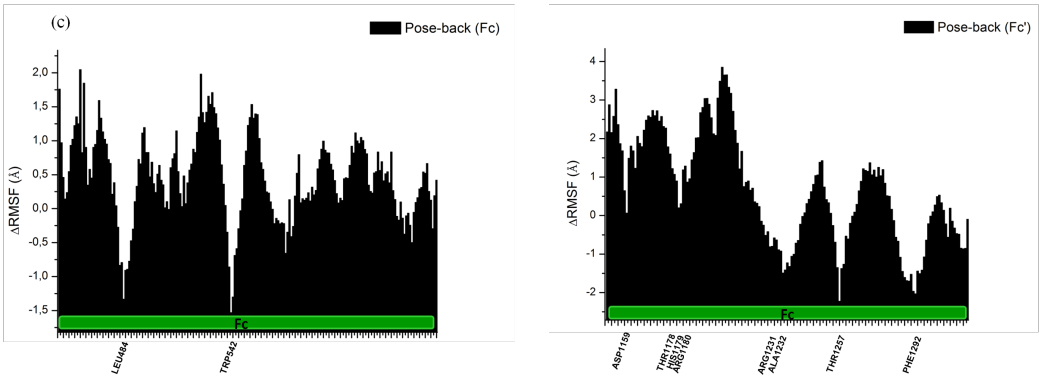


Figure A.3. Close view of the ΔRMSF for selected regions of the antibody, which are more affected by CNT binding for pose-fc, -fab and back in (a), (b), (c) respectively.

Table A.1. Summary of the Bader charges calculated for CNT benzyl rings interacting with residues Lys841, Lys555 and Glu512 in pose-fab. The charges of the CNT alone (CNT) and in presence of each residue are reported.

	Bader Charges (e)					
Carbon atom	**CNT**	**Lys841**	**CNT**	**Lys555**	**CNT**	**Glu512**
1	0.03	-0.17	0.19	-0.001	-0.10	-0.22
2	0.12	0.14	-0.13	0.27	0.04	0.14
3	-0.13	-0.12	-0.10	0.08	-0.01	0.05
4	-0.06	-0.05	-0.17	-0.02	0.12	-0.23
5	0.01	-0.28	-0.07	-0.09	0.09	0.05
6	-0.07	0.24	0.06	0.10	0.06	-0.09
7	0.14	-0.08	-0.10	-0.28	-0.17	0.02
8	-0.08	0.04	-0.22	-0.03	0.19	-0.11
9	0.02	-0.06	-0.15	-0.03	-0.04	0.09
10	0.13	-0.09	0.20	0.002	0.06	0.41
11					-0.20	-0.03
12					0.16	0.18
13					0.08	-0.22
14					-0.18	0.03
15					0.03	0.08
16					-0.17	-0.20

Table A.2. Summary of the Bader charges calculated for CNT benzyl rings interacting with residues His521, Arg522 and Glu523 in pose-fc interaction mode. The charges of the CNT alone (CNT) and in presence of each residue are reported.

	Bader Charges (e)					
Carbon atom	**CNT**	**His521**	**CNT**	**Arg522**	**CNT**	**Glu523**
1	0.11	0.21	0.01	1.35	-0.18	0.13
2	-0.22	-0.08	0.21	0.06	0.08	0.13
3	-0.02	-0.15	0.10	1.33	-0.64	0.04
4	0.13	0.15	-0.03	-0.18	-0.18	-0.46
5	-0.18	-0.13	0.04	1.23	-0.06	0.19
6	-0.17	-0.19	-0.17	-1.43	-0.04	0.05
7	0.18	1.40	-0.23	-1.55	-0.18	-0.10
8	0.04	-1.35	0.23	2.54	0.19	-0.19
9	0.0001	0.21	0.15	-2.66	-0.13	-0.06
10	0.13	-0.2	0.14	-1.11	0.03	-0.25

Table A.3. The occupancy (%) of H-bonds between the lateral Glu and the N-term groups detected during the 50 ns trajectories of **6bb2** is reported. The residues involved in the intermolecular H-bond interaction are reported displaying the symbol name and the number of the relative number within the proteinic structure.

H-bonds		
Donor	**Acceptor**	**%**
Cys40	Glu64	7.4
Cys66	Glu67	4.7
Cys14	Glu38	2.3

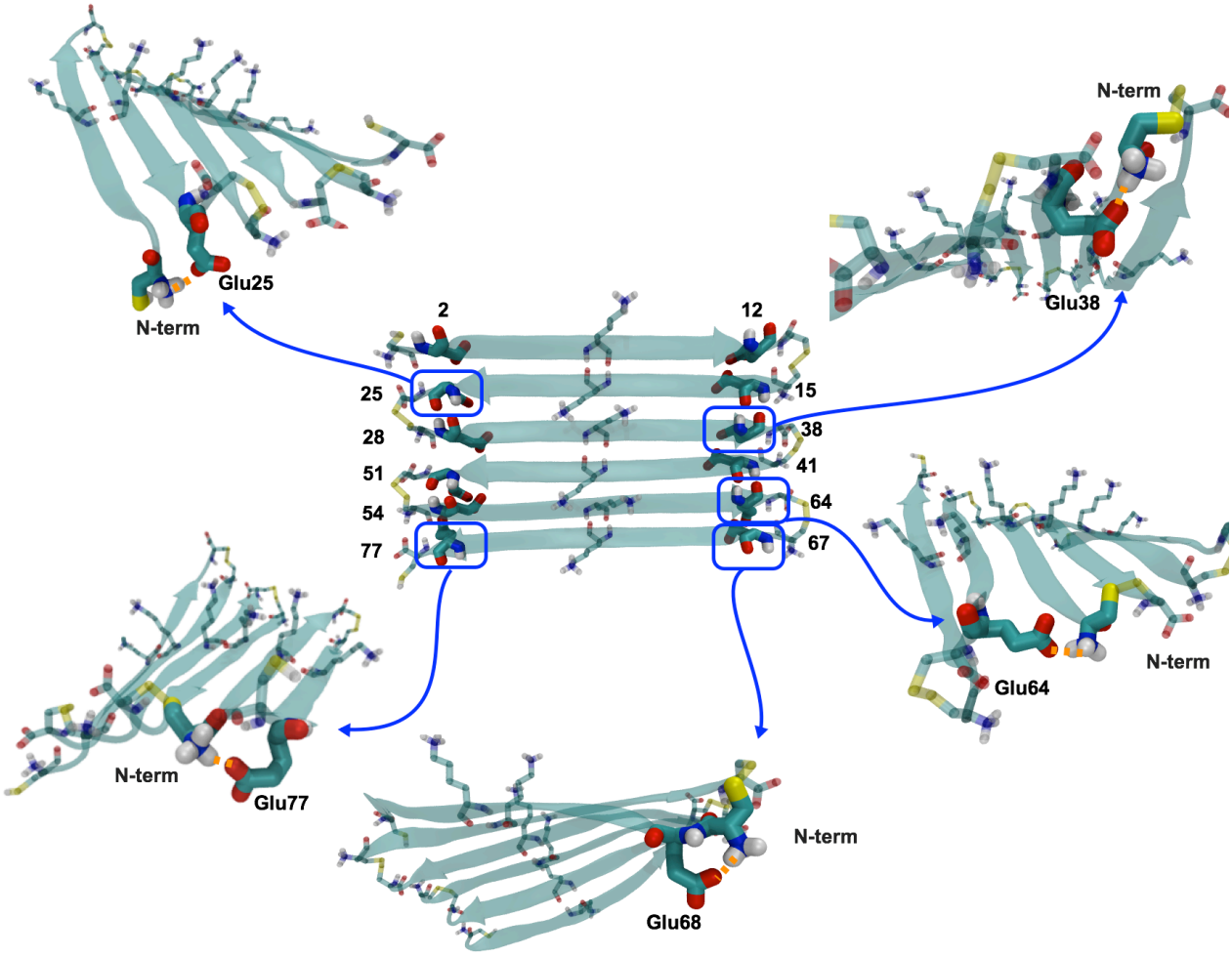


Figure A.4. In the middle it is reported the starting structure of **6bb2** where the lateral Glu are drawn in licorice and labeled with the reference number. The Glu forming H-bonds with the N-term groups are framed and for each of them is reported a snapshot extracted from the 50 ns trajectory showing the Coulombic interactions between free Glu residues and NH_3^+ groups. For clarity all the rest of the proteinic β-sheet is shown in transparent mode.

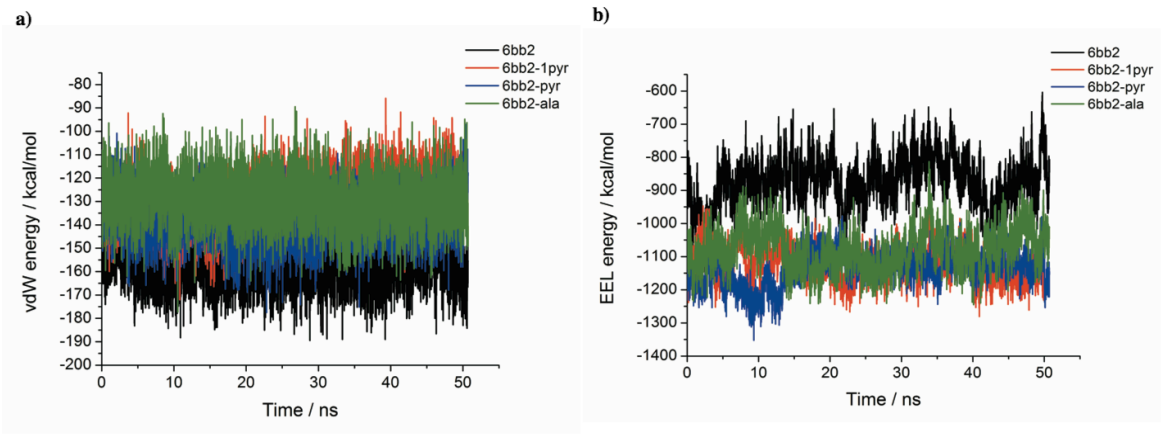


Figure A.5. The vdW (a) and the EEL (b) non-bonded interaction energies are calculated along the **6bb2** (black), **6bb2-ala** (green), **6bb2-1pyr** (red) and **6bb2-pyr** (blue) 50 ns trajectories.

Table A.4. The occupancy (%) of H-bonds between the lateral Glu and the N-term groups detected during the 50 ns trajectories of **g-6bb2** is reported. The residues involved in the intermolecular H-bond interaction are reported displaying the symbol name and the number of the relative number within the proteinic structure.

H-bonds		
Donor	**Acceptor**	**%**
Cys14	Glu38	30
Cys27	Glu51	17
Cys40	Glu41	9
Cys66	Glu64	5.5
Cys1	Glu2	4
Cys53	Glu54	3

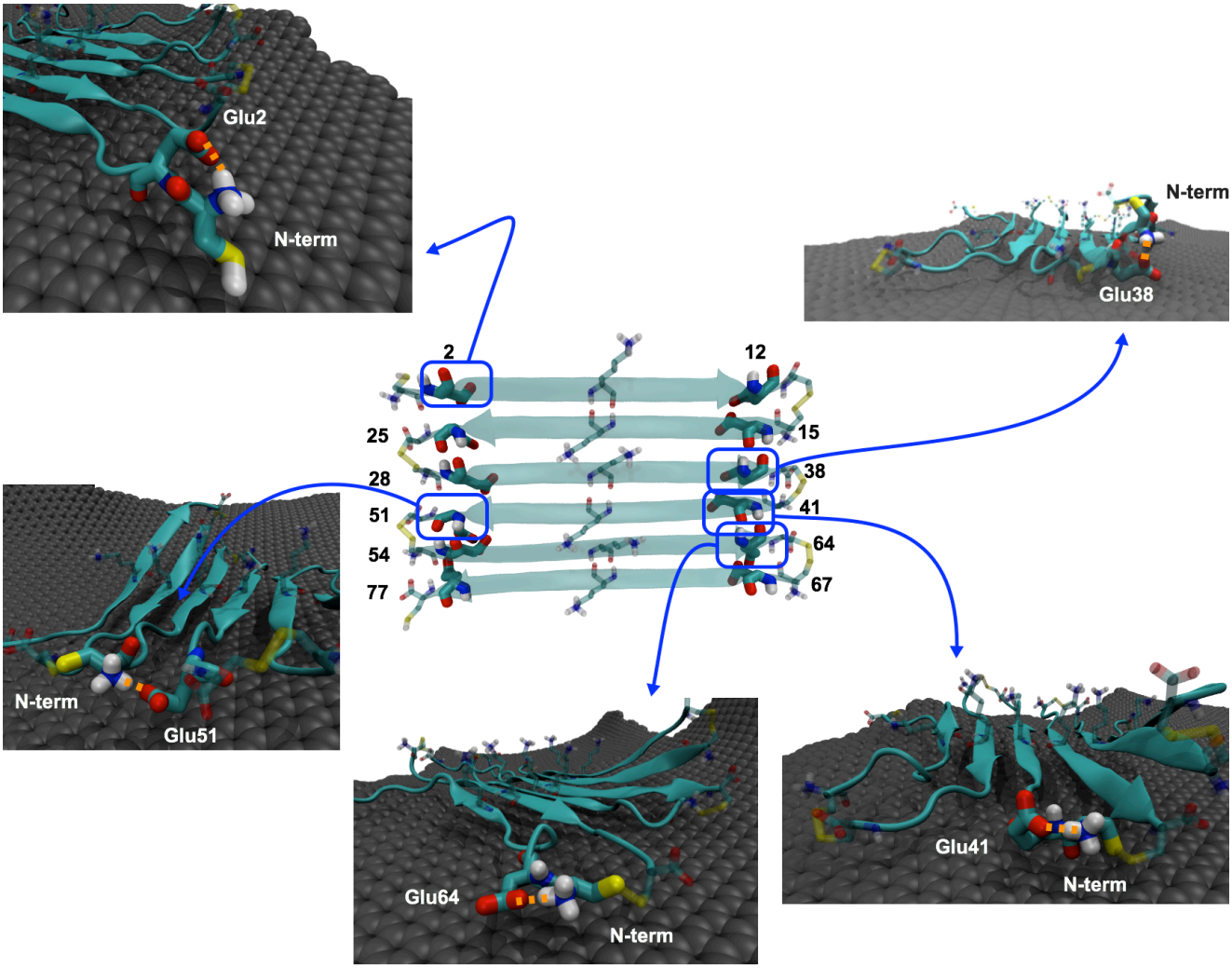


Figure A.6. In the middle it is reported the starting structure of **6bb2** where the lateral Glu are drawn in licorice and labeled with the reference number. The Glu forming H-bonds with the N-term groups are framed and for each of them is reported a snapshot extracted from the **g-6bb2** 50 ns trajectory showing the Coulombic interactions between free Glu residues and NH_3^+ groups. For clarity all the rest of the proteinic β-sheet is shown in transparent mode.

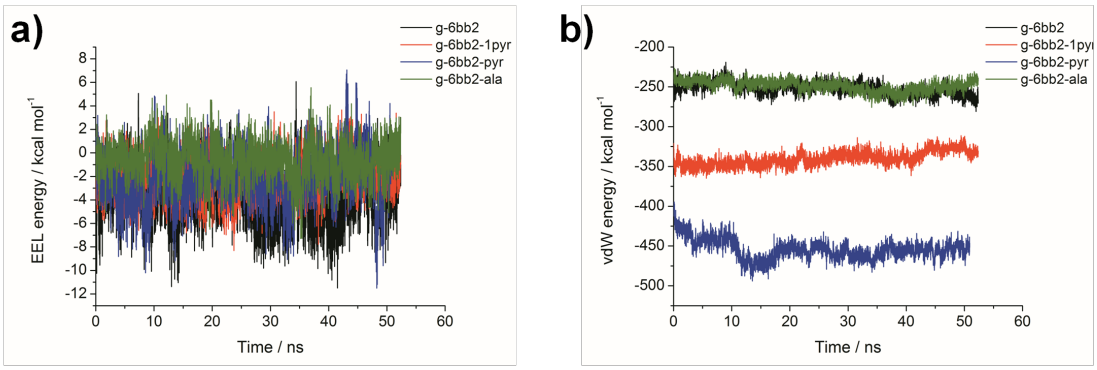


Figure A.7. The vdW (a) and the EEL (b) non-bonded interaction energies between the β-sheets and the graphene surface are calculated along the **g-6bb2** (black), **g-6bb2-ala** (green), **g-6bb2-1pyr** (red) and **g-6bb2-pyr** (blue) 50 ns trajectories.

Chapter 4

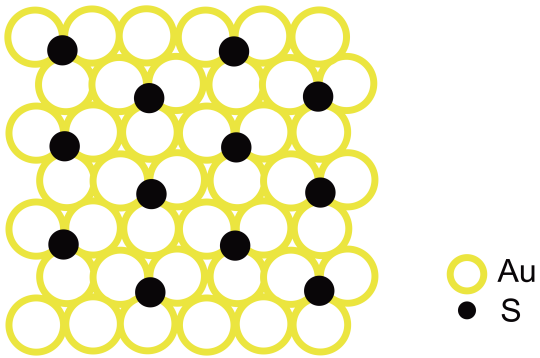


Figure A.8. Schematic diagram sketching the (√3x√3)R30° arrangement where the sulfur atoms (black dots) are positioned in the 3-fold hollows of the gold lattice (empty yellow circles). This represents the maximum coverage that the thiolates might achieve.

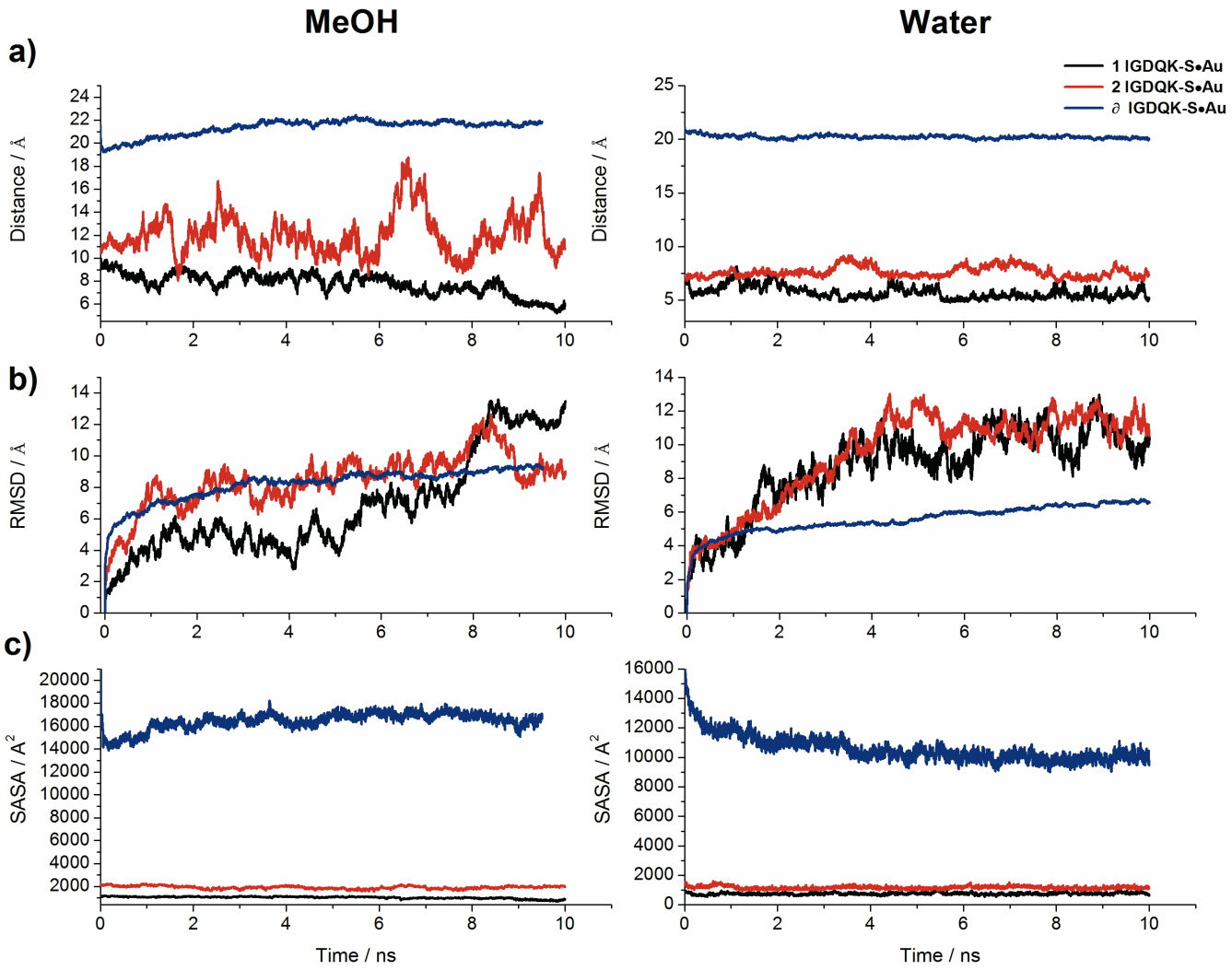


Figure A.9. Rg (a), RMSD (b) and SASA (c) plots calculated along the 10 ns trajectory of **IGDQK-S·Au** (black profile), **2IGDQK-S·Au** (red profile), and ***∂*IGDQK-S·Au** (blue profile) in MeOH (left column) and water (right column).

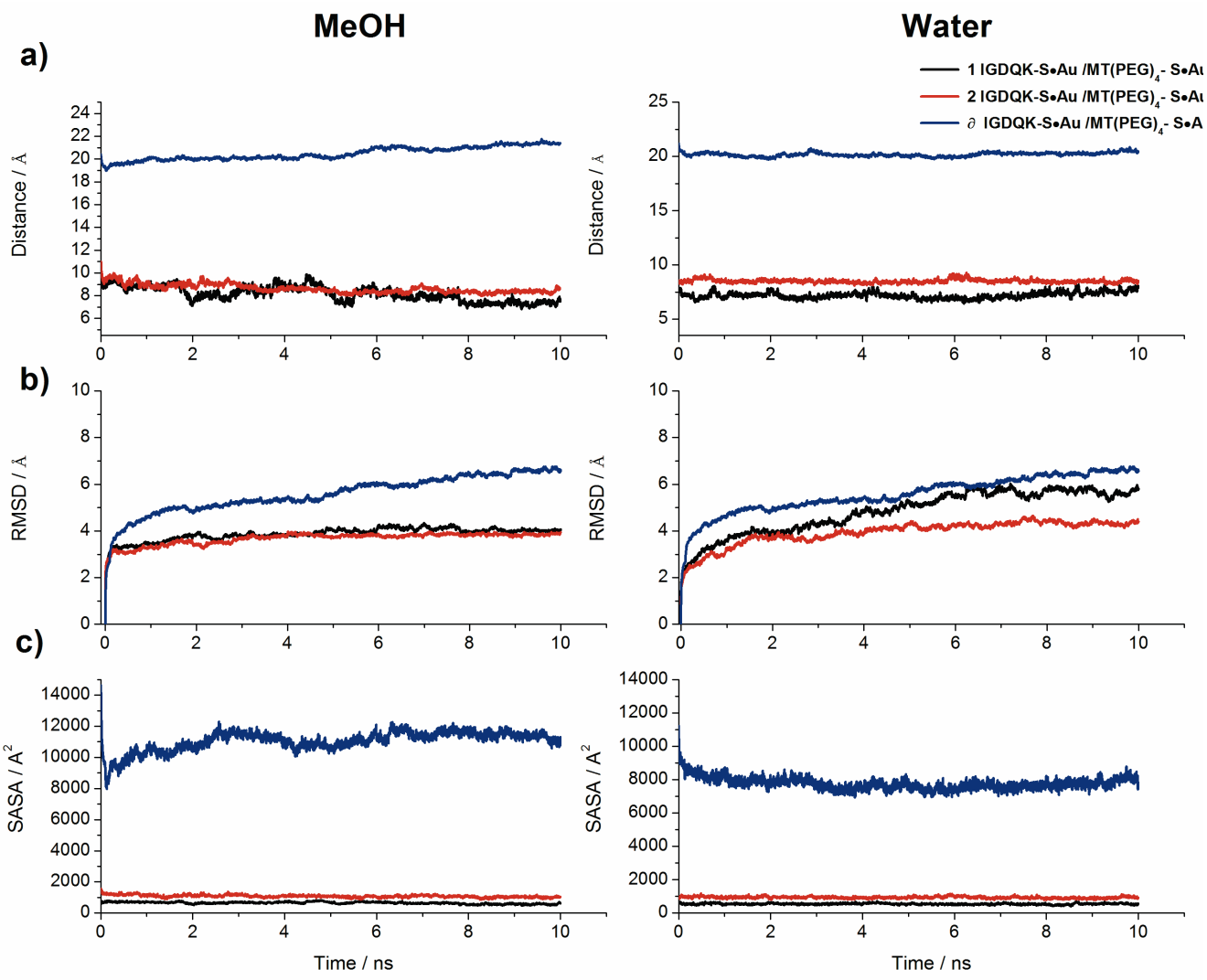


Figure A.10. Rg (a), RMSD (b) and SASA (c) plots calculated along the 10 ns trajectory of **IGDQK-S·Au/MT(PEG)$_4$-S·Au** (black profile), **2IGDQK-S·Au/MT(PEG)$_4$-S·Au** (red profile), and **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** (blue profile) in MeOH (left column) and water (right column).

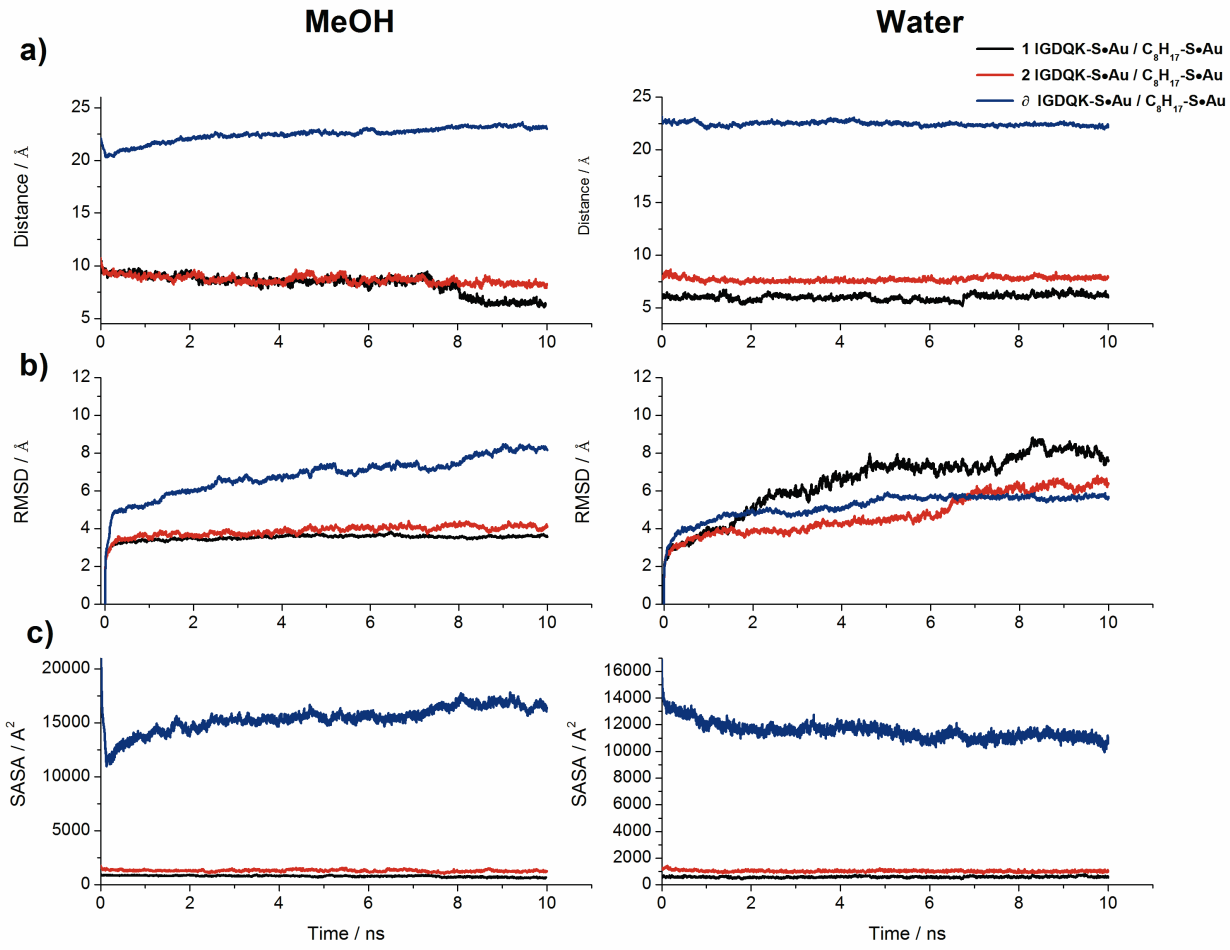


Figure A.11. Rg (a), RMSD (b) and SASA (c) plots calculated along the 10 ns trajectory of **IGDQK-S·Au/C_8H_{17}-S·Au** (black profile), **2IGDQK-S·Au/C_8H_{17}-S·Au** (red profile), and **∂IGDQK-S·Au/C_8H_{17}-S·Au** (blue profile) in MeOH (left column) and water (right column).

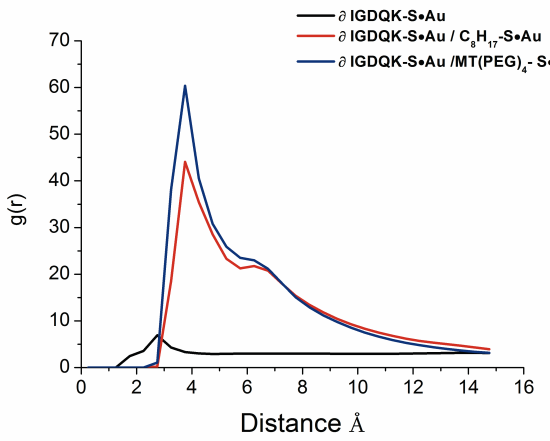


Figure A.12. RDFs plot calculated along the 10 ns trajectory of **∂IGDQK-S·Au** (black profile), **∂IGDQK-S·Au/MT(PEG)$_4$-S·Au** (blue profile), and **∂IGDQK-S·Au/C_8H_{17}-S·Au** (red profile) simulated in water.

Curriculum Vitae

Education and experiences

- **University of Namur and Namur Research College (NARC) (Belgium) 03/2010-12/2014**

Ph.D. student in computational chemistry, Department of Chemistry, Davide Bonifazi's group

- **King's College, London (UK) 12/2012**

Visiting Ph.D. student in the group of Prof. Dr. Alessandro De Vita

- **International School of Advanced Studies (SISSA/ISAS), Trieste (Italy) 2010/2012**

Visiting Ph.D. student in the group of Dr. Alessandra Magistrato

- **Universitá degli Studi di Parma (Italy) 2004/2009**

Bachelor and Master degrees in Medicinal Chemistry and Technology presenting the thesis "Building and validation of new model of the H3 receptor" under the supervision of Prof. Marco Mor.

Research Fellowship and Grant

- **iCARE - International Cancer Research Fellowship co-funded Marie Curie Action, FP7**

2015/2017

Post-Doc fellow at Fondazione Centro San Raffaele, Milano under the supervision of Dr. Giovanna Musco (Start 03/2015)

- **Marie Curie ITN, FP7 2010/2014**

Ph.D. fellow within the FINELUMEN, "Cavity-confined Luminophores for Advanced Photonic Materials: A training action for young researchers".

- **SCAI - SuperComputing Applications and Innovation - CINECA 2011-2014**

Agreement CINECA-UniTs

- **CASPUR HPC Grant 2011 on GPU cluster 2011**

Awards

International Travel Awards for travel fellowships to the 2014 Biophysical Society Annual Meeting, San Francisco, California.

Academic collaboration

The Marie Curie project frames the Ph.D. in a European context in strict collaboration with a wide network of academic (to name a few: Dr. Nicola Armaroli, Italian National Research Council in Bologna, IT, photochemistry) and industrial partners (NANOCYL, Belgium (Dr. J. Amadou); PHILIPS Aachen, Germany (Dr. P. Backmann)). Other collaborations have been established with theoretical groups (Dr. Alessandra Magistrato, International School of Advanced Studies, SISSA/ISAS, Trieste, IT; Prof. Dr. Alessandro De Vita, King's College, London, UK) and experimental group (Prof. Dr. Stelios Couris, Institute of Chemical Engineering Sciences, Patras, GR).

Scientific interests

Computational chemistry, Molecular Dynamics simulations, trajectory analysis, electronic structure methods
Molecular modeling of the interaction between carbon nanomaterial and proteins for biologic applications
Investigation of weak non-covalent interactions (vdW, H-bonding interactions) involved in the binding, self-organization and self-assembly processes.

Attended courses

- **Gaussian Workshop** **23-27/07/2012**

Ohio Supercomputer Center (OSC) in Columbus, Ohio

- **Amber Workshop** **03-06/05/2011**

Ohio Supercomputer Center (OSC) in Columbus, Ohio

Computer skills and competences

Unix/Linux, Windows and Mac OS operative systems, bash shell scripting, awk
Knowledge of different computational chemistry software:
Modeller, Suite Schrodinger (MacroModel, Maestro, Glide), Pymol, Amber and AmberTools, Gromacs, NAMD, Autodock, Gaussian, MOPAC, VMD

Technical skills and competences

Analytical chemistry (XPS, TGA, SEM, TEM, Ft-IR, Uv-Vis

Languages

Italian (native), English (fluent), Spanish (fluent), French (basic).

List of publication

C. Fabbro, S. Armani, L-E. Carloni, F. De Leo, J. Wouters, D. Bonifazi,

"2,5-Diamide-subtituted five-membered heterocycles: challenging molecular synthones."

Eur. J. Org. Chem. **2014**, doi: 10.1002/ejoc.201402654

N. Liaros, S. Couris, L. Maggini, F. De Leo, F. Cattaruzza, C. Aurisicchio, D. Bonifazi,

"NLO Response of Photoswitchable Azobenzene-Based Materials."

ChemPhysChem **2013**, *14*, 2961–72.

F. De Leo, J. Sgrignani, D. Bonifazi, A. Magistrato,

"Structural and Dynamic Properties of Monoclonal Antibodies Immobilized on CNTs: A Computational Study."

Chem. Eur. J. **2013**, *19*, 12281–12293 COVER PAGE.

R. Marega, F. De Leo, F. Pineux, J. Sgrignani, A. Magistrato, A. D. Naik, Y. Garcia, L. Flamant, C. Michiels, D. Bonifazi,
"Functionalized Fe-Filled Multiwalled Carbon Nanotubes as Multifunctional Scaffolds for Magnetization of Cancer Cells."
Adv. Funct. Mater. **2013**, *23*, 3173–3184.

L. Maggini, F. De Leo, R. Marega, H.-M. Tóháti, K. Kamarás, D. Bonifazi,
"Carbon nanotube-based metal-ion catchers as supramolecular depolluting materials." *ChemSusChem* **2011**, *4*, 1464–9.

Paper submitted

F. De Leo and D. Bonifazi,
"Non-covalent Interactions between Proteins and Graphitic Nanomaterials: from Spontaneous Attraction to Tailored Mastering."

A. Kremer, C. Aurisicchio, F. De Leo, B. Ventura, J. Wouter, N. Armaroli, A. Barbieri and D. Bonifazi,
"Walking down in the chalcogenic group of the periodic table: from fluorescent to phosphorescent organic emitters."

Paper in preparation

V. Corvaglia, R. Marega, F. De Leo, C. Michiels, S. Silvestrini, D. Bonifazi,
"Structural insight for design motogenic surfaces."

L. Đorđević, T. Marangoni, I. Papagiannouli, P. Aloukos, S. Couris, F. De Leo, F. Diederich and D. Bonifazi,
"Boosting Third-order NLO responses of fullerene-porphyrin [n]pseudorotaxanes"