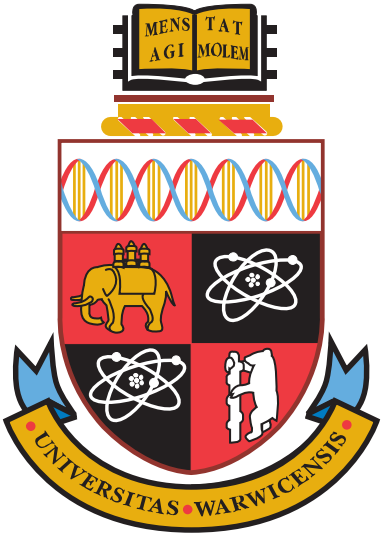

Modelling of the interaction between peptides and graphitic surfaces

by

Susana de Miranda Tomásio

Thesis

Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy

Chemistry Department

July 2010

Dedicado ao meu pai, à minha mãe e ao Gustavo . . .

To my father, my mother and to Gustavo . . .

Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my sincere gratitude to my supervisor Tiffany Walsh without whom the work presented in this thesis would have not been possible. Thank you very much for your availability, your positive attitude, your support and your enthusiasm. Special thanks to Becky Notman for the useful scientific discussions and for her valuable comments, not only during the revision of my thesis, but also all along these last three years. I also would like to thank Becky Notman for proof-reading my thesis. I would like to thank Nadia Grossiord and Ralf Brodeck for the pleasant scientific discussions and Sara Fortuna for her helpful comments. I also would like to give a special thank you to David McMahon for taking the time to proof-read my thesis and for his helpful comments. I would like to thank the past research group members Adam Skelton, Taining Liang, Dan Harding and Becky Notman for having provided an exciting and motivating environment. In addition, I would like to thank Matt Bano for being a nice office-mate during this difficult stage. This research would have not been possible without the computing facilities of the Centre for Scientific Computing and the National Grid Service. My thanks go to these facilities. I am also grateful for the financial support of the University of Warwick. Finally, I would like to thank my parents João Tomásio and Isabel Arráis and my boyfriend Gustavo Diniz for their support, encouragement and assistance both during the PhD and during the writing of this thesis.

Declarations

This thesis represents original material, and is the author's own work. No part of this thesis has been previously created or published by a person other than the author, unless otherwise acknowledged in the text. This thesis has not been submitted for a degree at an establishment other than the University of Warwick. Part of the work presented in this thesis is published in the following papers: S. M. Tomasio and T. R. Walsh, J. Phys. Chem. C. 113 8778 (2009); T. R. Walsh and S. M. Tomasio, Mol. BioSyst. 6 1707 (2010).

Do not despair that the way is long;

If you go on, you can reach it.

Do not despair that the load is heavy;

If you lift it, you can carry it.

Chinggis Khaan (1162-1227)

Abstract

The aim of this thesis is to understand the interactions of peptides with graphitic surfaces such as carbon nanotubes and graphite, in order to help establish guiding principles for the design of peptide sequences with controllable affinity to graphitic surfaces. Atomistic molecular dynamics (MD) simulations with our extended polarisable AMOEBAPRO force-field, which includes parameters for graphitic surfaces is used throughout. The peptide sequences studied were identified by phage-display experiments for their strong affinity to CNTs, and are rich in tryptophan and histidine residues [94]. The importance of the tryptophan residues on the binding affinity to CNTs is investigated by mutating each tryptophan by either tyrosine and phenylalanine. In addition, the effect of the surface curvature on the binding affinity is also explored. It is found that sequences containing tryptophan residues have more affinity to graphitic surfaces than those containing tyrosine or phenylalanine. Furthermore, it is suggested that these peptide sequences were selected for interfacial shape, since in the case of graphite, a compromise between having all the aromatic residues close to the surface and also allowing the non-aromatic residues to approach the surface is found. Following this study, the interaction of peptide sequences with CNTs is again studied, but this time with the aim to investigate the order of the residues, on the binding affinity to CNTs. The influence of the peptide sequence on the binding affinity to CNTs is studied by scrambling the sequence (HWKHPWGAWDTL). This study suggests that binding affinity is strongly dependent on the order of the content of the peptide sequences and gives some useful insights to the identification of principles that may help in the design of peptide sequences with controllable binding affinity to CNTs. For instance, it is found that strong binding may be due to the presence of isolated pairs of tryptophans, while weaker binding may be due to the presence of two tryptophan residues intercalated by another residue. The interactions of water with graphitic surfaces – CNTs, fullerenes and graphite – are also considered and it is found that the water structuring at the interface is weak and that there are no more than tree layers of structured water on the graphitic surfaces. Finally, the effect of the presence of OH defects on CNTs on the binding affinity to peptides is investigated. The results show that the binding affinity is not significantly affected by the presence of OH defects, but a general increase in the peptide mobility is noticed, giving insights for the applications of real CNTs with peptides.
The work described in this thesis helps to understand what are the key residues involved in the interaction with CNTs, why do these key residues bind better to CNTs and provide insights on the mechanisms of peptided binding to CNTs, by demonstrating the role of peptide conformation.

Abbreviations

AMBER	Assisted Model Building with Energy Refinement
AFM	Atomic Force Microscopy
AMOEBA	Atomic Multipole Optimized Energetics for Biomolecular Applications
ASP	Atomic Solvation Parameters
BCS	Bacterial Cell-Surface
BSSE	Basis Set Superposition Error
CNT	Carbon NanoTube
CD	Circular Cichroism
COM	Centre of Mass
CHARMM	Chemistry At Harvard Molecular Mechanics
DFT	Density Functional Theory
DMA	Distributed Multipole Analysis
DNA	DeoxyriboNucleic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
FM	Fluorescence Microscopy
GB	Generalised Born
GCMC	Grand Canonical Monte Carlo
GEPI	Genetically Engineered Polypeptide for Inorganics

GGA	Generalised Gradient Approximation
HA	HydroxyApatite
LJ	Lennard-Jones
NMR	Nuclear Magnetic Resonance
MD	Molecular Dynamics
MC	Monte Carlo
MM	Molecular Mechanics
MWCNT	Multi-Walled Carbon NanoTube
NVT	constant Number of atoms, Volume and Temperature
NPT	constant Number of atoms, Pressure and Temperature
PD	Phage-Display
PEL	Potential Energy Landscape
QCM	Quartz Crystal Microbalance
QM	Quantum Mechanical
RMSD	Root Mean Squared Deviation
RNA	RiboNucleic Acid
SAM	Self-Assembled Monolayer
SEM	Scanning Electron Microscopy
SPC	Simple Point Charge water model
SPC/E	Extended Simple Point Charge water model
SPR	Surface Plasmon Resonance spectroscopy
SWCNT	Single-Walled Carbon NanoTube
VDW	Van der Waals

Glossary

This Glossary includes the notation used for peptides. Throughout the thesis each peptide residue will be represented by a one-letter code. The codes are presented in Table 1.

Table 1: One letter-code, polarity and charge at pH 7.0 for the standard amino acids.

Residue	**Code**	**Polarity**	**Charge at pH=7.0**
Alanine	A	non-polar	neutral
Arginine	R	polar	positive
Asparagine	N	polar	neutral
Aspartic acid	D	polar	negative
Cysteine	C	polar	neutral
Glutamic acid	E	polar	negative
Glutamine	Q	polar	neutral
Glycine	G	non-polar	neutral
Histidine	H	polar	neutral
Isoleucine	I	non-polar	neutral
Leucine	L	non-polar	neutral
Lysine	K	polar	positive
Methionine	M	non-polar	neutral
Phenylalanine	F	non-polar	neutral
Proline	P	non-polar	neutral
Serine	S	polar	neutral
Threonine	T	polar	neutral
Tryptophan	W	non-polar	neutral
Tyrosine	Y	non-polar	neutral
Valine	V	non-polar	neutral

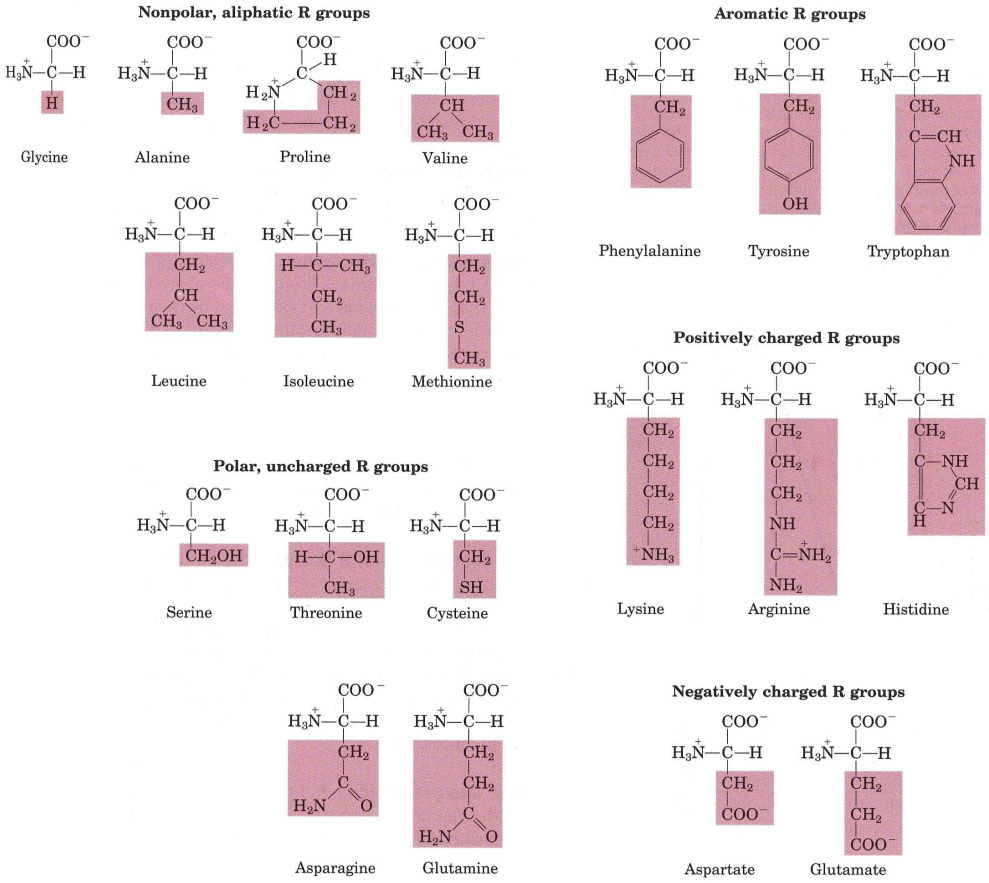

Figure 1: Schematic of the twenty amino acids. The structural formulas show the state of ionisation that predominates at pH 7.0. (From [3].)

Chapter 1

Introduction

1.1 General overview of the problem

In Nature, interactions of biomolecules such as peptides with inorganic materials exert a remarkable level of control over the growth and nucleation of these materials [1]. Hard tissues such as bones, dental tissues (dentine and enamel), spicules, spines, shells and bacterial nanoparticles are all examples of biomaterials (Figure 1.1) [1]. The unique structure of these biomaterials endow them with distinctive functions such as magnetic, optical and mechanical. This ability to direct the assembly of nanoscale structures has motivated the laboratory-based search for methods that mimic this behaviour found in biological systems; the field of *biomimetics*. Biological tissues are synthesised under genetic control, in aqueous environments and under mild physiological conditions using biomacromolecules (usually proteins but also carbohydrates and lipids), and as a consequence the selection processes in biological systems give rise to specific molecular recognition [2]. On the other hand, engineering (*i.e.* artificial) materials are synthesised using a combination of approaches (such as melting or solidification processes) in

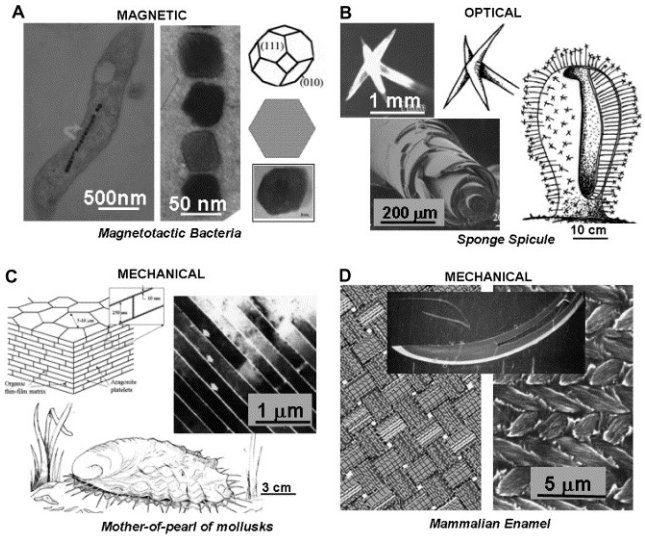

Figure 1.1: A few examples of functional biological materials systems: (A) Magnetotactic bacteria, *e.g. Aquaspirillum magnetotacticum*. (B) Spicule (*Rosella racovitzea*. (C) Mollusk *Haliotis rufescens* (D) Mammalian enamel comprises $\sim$100% hydroxyapatite(the right is an SEM image of the fractured surface and left is a schematic of the woven structure). Taken from Ref. [24].

which the final product is directed by the kinetics and thermodynamics of the system. The fundamental aspects of how peptides bind to inorganic materials as will be further explained, constitutes a challenge. Understanding the mechanisms that govern peptide binding and recognition to inorganic surfaces is crucial to the design of novel peptides and to tailor the peptide binding and promises to revolutionise the fields of materials science, bionanotechnology and medicine.

1.2 Peptides and proteins

Proteins are the most abundant biological molecules occurring in all living cells [3] and have an extensive diversity of functions such as structural support, enzymatic catalysis,

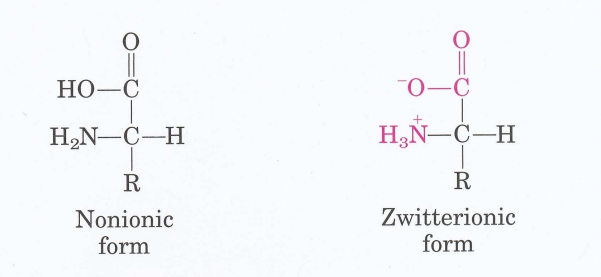

Figure 1.2: Schematic of a general structure of an amino acid (non-ionic and zwitterionic forms). The zwitterionic form predominates at neutral pH. (From [3]).

binding, transport and storage, immune protection *etc*. Proteins are the molecular pathways through which genetic information is expressed. Amino acids are the building blocks of peptides and proteins. They have a carboxyl group and an amino group bonded to the same carbon atom (the α carbon) (Figure 1.2). There are 20 naturally occurring amino acids which differ from each other in their side-chains, or R groups, which vary in structure, size, and electric charge, and which influence the solubility of the amino acids in water. Peptides and proteins are dehydrated polymers of amino acids residues joined by an amide linkage called the peptide bond. Such a linkage is formed by removing the elements of water when two or more amino acids combine. What remains of each amino acid is called an amino acid residue [3].

1.3 Biomimetics

The ability of biomolecules such as enzymes, nucleic acids and anti-bodies to recognise target molecules has enabled the emerging of life [4]. Nature has developed objects and materials with high performance functions that go from the macro-scale to the nanoscale. These functions have inspired scientists to engineer nanomaterials, nanodevices and pro-

cesses that mimic nature. A molecular understanding of these natural systems is required in order to mimic and create artificial nano- and micro-structures whose complexity and architecture is similar to those found in biology. Using biological inspiration to design or adapt materials and devices from nature is referred to as "biomimetics" [1]. The emerging field of molecular biomimetics is highly interdisciplinary since it draws on the knowledge base of biological functions, structures and principles of various objects found in nature by biologists, chemists, physicists and material scientists, and the design and fabrication of various materials and devices by engineers, material scientists, chemists and others. Molecular biomimetics combines biological molecular tools with synthetic products for the development of new hybrid materials at the molecular scale, and for this reason, it can be seen as a bio-synergetic engineering [5]. Molecular biomimetics uses proteins and peptides in the control and synthesis of nanoscales devices and self-assembly of functional nano-structures, by using their recognition properties. Peptides and proteins hold a large information content and due to their specific recognition and interactions with biomolecules, proteins are the "machinery" for innumerous functions, such as spacer, builder assembler or growth promoter. Proteins provide three distinctive advantages in the development of new materials and systems: molecular recognition, self-assembly and the genetic manipulation of their composition and structure [6].

A genetically engineered polypeptide for inorganics (GEPI) is defined as a sequence of amino acids that specifically and selectively binds to a target inorganic surface [1]. A peptide that binds to a specific molecule is often called aptamer. During the last two decades, combinatorial-peptide selection methods such as bacterial cell-surface display and phage-display have been used to select for a variety of GEPIs. Hundreds of peptide sequences specific to noble metals (Au [7; 1; 8; 9], Ag [10; 11], Pt [12] and Pd [9]), oxides and semiconductors (Cu_2O [13], GaAs [14], CdS, ZnSe [15], TiO_2 [16; 17], and

ZnO [18]), minerals (mica, hydroxyapatite [19], calcite) have been identified. In molecular biomimetics the selection of peptide sequences is accomplished through a process denoted herein as combinatorial mutagenesis. A brief overview of these combinatorial biology methods is given in the next Section.

1.3.1 Selection of inorganic-binding peptides using combinatorial mutagenesis

Throughout this thesis the terms "affinity" and "specificity" will be frequently mentioned. A peptide that binds strongly to an inorganic surface is denoted having high "affinity", while a peptide that binds to a specific surface and not other surfaces is denoted as binding with "specificity".

In combinatorial biology techniques, a large random peptide library (with a diversity of typically 10^9-10^{11} sequences) with the same number of amino acids, but varying compositions, is used to screen for sequences that strongly bind to a given inorganic surface [7; 14; 20]. Bacterial cell-surface (BCS) and phage-display (PD) libraries are commonly used to select for a variety of GEPIs. Typically, these libraries are generated by inserting random nucleotides within genes encoded on phage (virus) genomes or on bacterial plasmids (step 1 in Figure 1.3). This process leads to the display of a random peptide sequence of desired length (typically 7-14 amino acids long) within a protein of the host (the coat protein of a phage or an outer membrane or flagellar protein of a cell; step 2). A different randomised peptide is, as a consequence, produced and displayed by each cell or phage (step 3). The host library is then put in contact with the target substrate (step 4). Weak or non-specific binders are removed by breaking the weak interactions with several washing cycles, while phages with strong affinity to the surface will remain bound to the surface during this process. Bound phages or cells are then

eluted from the surface (step 6). In the case of PD, a second host (bacteriophage) is required to re-infect the host for the amplification of the eluted phages, while in the case of BCS display, cells are allowed to grow (steps 7 and 8). This entire process represents a round of biopanning. In order to refine the search for strong binders, three to five cycles of biopanning are usually performed. The choice of elution buffers (solution used to wash out weak binders) should be considered carefully since they can modify or deteriorate the surface of the target substrate.

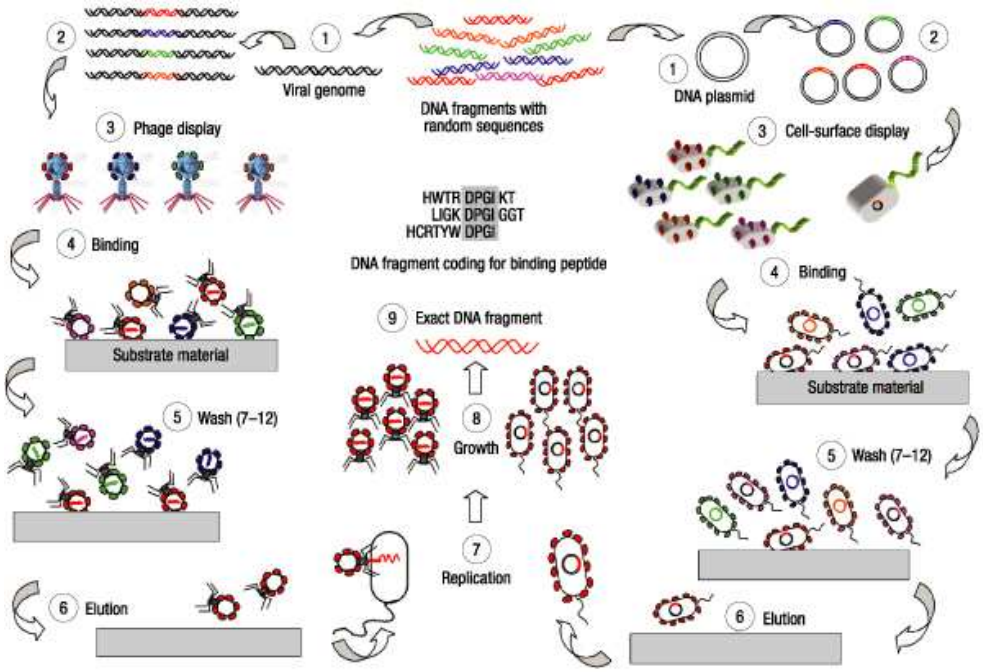

Figure 1.3: Schematic of phage-display and cell-surface display selection processes. (Taken from Sarikaya *et al.* [1])

The sequences of the strongest binding peptides can be finally deduced (step 9) from the DNA codon sequencing. In these techniques peptides can be expressed in two geometric configurations: either unconstrained (denoted herein as having a "linear"

architecture) or constrained by the formation of a disulfide bond between cystein residues (denoted herein as having a "cyclic" architecture). In PD, the peptide sequence linear form comprises either 7 or 12 amino acids and the cyclic displays 7 amino acids [12; 19]. In BCS display, peptide sequences can vary from 7 to 15 residues in the linear form and 12 residues in the constrained form [18]. Several peptide sequences can be identified for a given inorganic surface with various degrees of binding strengths. These peptide binders constitute the "first generation" peptides and serve as a starting point for subsequent improvement and/or modification. In order to refine the results several molecular-tailoring strategies can be used to further "tune" affinity and specificity of these sequences.

1.3.2 Molecular characterisation of GEPIs

The combinatorial techniques described in Section 1.3.1 only provide an elementary knowledge regarding the affinity of a peptide to a certain inorganic surface. In order to use peptides in practical applications, it is necessary to have information regarding the mechanisms and strength of binding, kinetics, and assembly behaviours of the GEPIs with regard to a particular solid [1; 6; 21; 22]. Also important is the quantitative evaluation of the degree of peptide specificity to a given material, as compared to non-specific interactions with other substrates of similar chemical compositions or surface structures (*e.g.*, binding to gold versus platinum [23]). This evaluation could be assisted by having a detailed understanding of the physical and chemical features of the material as well as an understanding of the molecular structures of the GEPI and its recognition mechanism on the solid. A fundamental routine of the screening protocol for affinity and selectivity searching is the use of relatively simple techniques such as fluorescence microscopy (FM) and enzyme-linked immunosorbent assay (ELISA) for quick monitoring

of peptide adsorption and binding. FM is used to image large molecules and is based on the phenomenon that certain molecules emit visible light when irradiated with light of a specific wavelength. Molecules can be either fluorescents or non-fluorescents; in the latter the molecule can be conjugated with a fluorescent dye. ELISA is used to detect the presence of an antigen-antibody complex in a sample. In ELISA there is an amount of antigen fixed to the surface which is not known. A specific antibody is applied by washing over a surface, such that the antibody binds to the antigen. Exposure of light of the appropriate wavelength, causes the fluorescence of the antigen-antibody, and the amount of antigen can be determined from the magnitude of the fluorescence.

However, these two techniques yield only qualitative data about the binding. Quantitative data can be obtained by determining thermodynamic and kinetic quantities involved in the binding process (such as change in free energy, ΔG, entropy, ΔS, enthalpy, ΔH, adsorption rate, k_{ads}, desorption rate, k_{des} *etc.*) through the use of modified surface plasmon resonance (SPR) spectroscopy (see below) [24; 25; 12; 8; 26]. Recent reports have coupled SPR studies with quartz-crystal microbalance (QCM) (see below), because in the former adsorption is detected by changes of the refractive index of the interface while in QCM adsorption is detected by net mass changes [6]. One of the limitations associated with these techniques is the exclusive application to small molecules. Another complementary approach is the use of various spectroscopy techniques such as circular dichroism (CD), nuclear magnetic resonance (NMR) spectroscopy and molecular modelling to characterise the molecular structure of a strong binder and peptide recognition mechanisms. Atomic force microscopy (AFM) can be used to obtain quantitative imaging data at the molecular level, such as surface interactions with proteins [27]. However, the resolution is limited to 5 nm, meaning that for a 12 residue peptide it is not feasible. Therefore, other approaches must be used,

for instance the triplication of the peptide sequence [27]. A brief description of these quantitative techniques is given in the next Section.

1.3.2.1 Surface plasmon resonance (SPR)

SPR is used to measure rates of adsorption (k_{ads}) and desorption (k_{des}) of molecules at the surface and thus enables the calculation of equilibrium constants (K_{eq}), since at equilibrium the rate of adsorption is equal to the rate of desorption:

$$K_{eq} = \frac{k_{ads}}{k_{des}} \tag{1.1}$$

SPR is an optical phenomenon which results from the excitation of surface plasmons by light. Surface plasmons are electron oscillations existent at the interface between metal and another material, that propagate parallel along the surface. These oscillations are highly sensitive to any change of the metal surface such as the adsorption of molecules. Surface plasmons can be excited by irradiating the sample with p-polarised light (polarisation parallel to the plane of incidence) and can be detected by measuring the changes in the index of refraction of the p-polarised light. By determining the resonance energy the concentration of adsorbed molecule can be inferred [6]. Adsorption and desorption rate constants (k_{ads} and k_{des}) are obtained by plotting the observed rate constant (k_{obs}) as a function of the molecule concentration.

1.3.2.2 Quartz crystal microbalance (QCM)

QCM is used to measure the amount of adsorbate at a surface by measuring the change in frequency of a quartz crystal resonator. It uses a quartz crystal disk coated with an electrode (typically gold) and the surface of interest. When disturbed by the adsorbate the quartz crystal resonator changes its frequency. From this change in frequency the amount of peptide adsorbed on the surface can be determined [6]. This technique

is useful for measuring the affinity of peptides on a surface. One of the limitations associated with the use of SPR and QCM is the need to deposit the material of interest in a thin film. Sometimes this is not possible and modifications to the instrumentation are required.

1.3.2.3 Atomic force microscopy (AFM)

AFM is used for imaging and manipulating atoms and structures on a variety of surfaces [28]. This information is collected by scanning the surface with a mechanical probe which consists of a cantilever (a beam supported on only one end) with a sharp tip (probe) at its free end, which is used to scan the surface. When the tip is put near the surface, forces between the tip and the sample cause a deflection of the cantilever which can be measured. AFM is useful to spatially and temporally monitor surface interactions with proteins giving an insight into the dynamics of the adsorption event.

1.3.3 Circular dichroism (CD)

CD is the difference in absorption of left- and right- circularly polarised light [29]. CD is used to infer the structure of macromolecules, such as the secondary structure of proteins, and therefore estimate the fraction of a molecule that is in the α-helix conformation, the β-turn conformation or random coil conformation [30]. Despite CD not giving information about the conformation of peptides adsorbed on surfaces, it can provide information about the conformation of the peptide in solution, which can then be used to help understand the peptide conformation on the surface.

1.3.4 Implementation of solid-binding peptides in bio- and nanotechnology

The control and assembly of biomolecules on inorganic surfaces is currently one of the main objectives of bio-nanotechnology [24; 31]. A major ambition is to have a library of well-characterised GEPIs, for use in a wide range of applications, from synthesis through to the molecular and nanoscale assembly of solid materials. Through their recognition properties, GEPIs offer a distinctive possibility for the self- and directed assembly of molecules [32]. GEPIs may be used as molecular building blocks due to their specific ability to synthesise, display and assemble. They can be used as a linker between two nanomaterials and they could also be used as a ligand in a functional protein, where the GEPI is genetically fused to the protein, creating a bi-functional molecule with the ability to control the formation of inorganic nanostructures while still retaining the functionality of the protein [33]. Another example is the usage of GEPIs as molecular "displayers" in the directed self-assembly at specific locations of the solid, through their linkage to enzymes [34]. In addition, a GEPI could also be chemically fused onto a synthetic polymer, to create a structure with several functions [35]. These systems are often called "hybrid structures", which means the material contains a biomolecule (such as a protein) and an inorganic substrate. In all cases, the role of the GEPI is to provide an additional function to the hybrid structure. GEPIs can also be used as electron (or proton) transporting molecular bridges [22], to study biomolecular and cellular functions by building a communication interface between biological materials and electronic components. Furthermore, studies on the material specificity of GEPIs suggest their usage as specific binding agents and molecular "displayers" for a variety of bionanotechnology applications ranging from biosensors to nanophotonic entities [23].

In the field of nanotechnology, the cooperation of inorganic nano-structures and

designed peptides will enable the fabrication of distinctive materials with several functionalities as well as optical and electronic properties. As an example, in a recent study a strategy was developed to produce quantum dot nano-arrays mediated by GEPIs, which gave rise to an improvement of up to 15-fold increase in surface-plasmon-enhanced fluorescence [36]. A few examples of the applications of the GEPIs are given below in order to demonstrate their potential usage.

1.3.4.1 Synthesis of nano-inorganics aided by GEPIs

Even though the mechanisms that influence the morphology of minerals upon binding by GEPIs are not fully understood, it is believed that once accomplished it will have great potential applications in tissue regeneration. An example of this is the use of hydroxyapatite (HA) binding peptides on the biomineralisation of calcium phosphate [19]. It was demonstrated that depending on the affinity of the HA binding peptides, they could be used to control and regulate mineralisation creating crystals of different sizes. These findings have a great impact since it shows that these GEPIs could be used to finely select and synthesise minerals with desired properties. Another example is the use of gold binding peptides to control the morphology and crystal of gold nanoparticles [37; 1]. It was found that gold binding peptides could accelerate the rate of crystal growth while changing the morphology of the gold nanoparticles from the usual shape (cubo-octahedral) to flat, triangular or pseudo-hexagonal particles [37]. These findings demonstrate that GEPIs can work as catalysts in the synthesis and morphogenesis of inorganic nanomaterials.

1.3.4.2 Directed and oriented self-immobilisation of nanoparticles and biomolecules

Patterning of biomolecules on selected substrates is essential for protein micro-arrays used in proteomics and clinical assays, as well as for biosensors and microchips [38; 39; 40; 41]. The fabrication of these micro-arrays is possible if the protein is immobilised on the inorganic surface. Recently, protein immobilisation became a key problem in bio-nanotechnology. Immobilisation offers a physical support to the molecule while improving the stability and activity and additionally, it helps to separate proteins from the reaction mixture which enables its continuous use [42; 43]. Traditional approaches to the immobilisation of biomolecules include the use of glass or metal substrates and usually requires surface functionalisation (*via* covalent bond formation) by self-assembled monolayers (SAMs) of bi-functional molecules. Mostly are based on non specific adsorption such as intermolecular forces and hydrophobic or polar interactions. The problem associated with these traditional functionalisations by bi-functional molecules is that they can cause random orientation of the protein and require multi-step chemical reactions. These reactions often lead to the attachment of proteins on the surface. In addition, the assembled monolayers can be unstable during the immobilisation. In order to avoid these issues, peptides with specificity to the solid of interest can be used to directly link the substrate and the protein. The advantages of using GEPIs in protein immobilisation is the high-inorganic binding affinity and specificity of the GEPIs [35; 44]. For example, Kacar *et al.* proposed an application for quartz-binding peptides as a linker for self-assembly and as an "ink" for micro-contact printing of photoactive molecules immobilised on a quartz surface through quartz-binding peptides [44]. The authors also demonstrated the dual functionality of the structures based on GEPIs such as the surface-substrate binding ability and the ability to transport target molecules with specific functions [44]. In addition, Kacar *et al.* used gold-binding peptides as a site-

specific molecular linker genetically fused into an enzyme for its oriented immobilisation while retaining the enzyme activity [34]. Moreover, Zin *et al.* [45] demonstrated that micro-patterned gold-binding peptides are good linkers in the direct assembly of gold nanoparticles . Alternatively, in a still new area, GEPIs could be used in the targeted immobilisation of functional inorganic nanoparticles on biomolecular substrates, such as designed proteins [46], viruses [47] and DNA [48].

1.3.5 Applications of solid-binding peptides in Medicine

The potential applications of GEPIs in medicine span a wide range of domains such as tissue regeneration, molecular scaffolds (platforms) for tissue restoration, biosensors, and drug delivery. A few recent examples of these applications are given below in order to illustrate the usage of GEPIs in the field of medicine.

1.3.6 Regenerative medicine

The repair or replacement of hard tissues such as enamel, dentin, cementum, or bone represents a major challenge in regenerative medicine [49]. As previously mentioned, GEPIs might be able to control nucleation, growth and morphogenesis of HA nanoparticles. In addition, a recent study demonstrated that specific titania-binding peptides induced the precipitation of titania, but not silica [50], demonstrating the possibility of inducing the formation of one oxide while lacking the ability to form a similar oxide. These findings might have important implications in the applications to implants.

An exciting and revolutionary approach to regenerate tissues or organs consists of combining both the cells of the patients and polymer hydrogels that can be injected into the body [51]. For example, mineralisation could be controlled by the incorporation of GEPIs with hydrogels to provide this additional functionality [22]. Therefore,

hydrogels are used as platforms (scaffolds) in the engineering of new tissues [52], and the incorporation of GEPIs into peptide hydrogels present promising developments in regeneration of hard and soft tissues.

1.3.6.1 Nanomaterial-based biosensors

A biosensor is an analytical device that combines a biological component with a detector and therefore provides quantitative or semi-quantitative information. For example, a very common commercial bionsensor is the blood glucose biosensor, which breaks blood glucose down with the enzyme glucose oxidase. Biosensors rely on the affinity between the probe molecule and the analyte, in particular the specifity of the probe to the analyte is a key criteria for the design of efficient biosensors [53]. The advantage of nanomaterial-based biosensors is the unique property of biological nanomaterials to recognise a target molecule. These biosensors are generally, of fast response, high sensitivity and small size when compared to other sensors and are used to detect diseases and monitor medical therapies. Using GEPIs such as the Ag-binding peptide or the quartz-binding peptide fused in a protein assembled on a specific material, might be the principle for creating extremely sensitive biosensors [21].

1.3.6.2 Drug delivery

As mentioned before, a GEPI can have a dual functionality by simultaneously having a binding affinity to an inorganic and organic surface. This property could be applied to drug delivery by binding a GEPI to an inorganic nanoparticle with desirable functionality. By immobilising a drug to a nanoparticle with the help of a linker peptide, it could be delivered to the target tissue [6]. In addition, peptide hydrogels are also being studied as potential molecular scaffolds to drug delivery [51].

1.3.7 Using molecular simulation as a complementary tool

Used in conjugation with experimental approaches, simulation techniques will help to characterise the peptide-inorganic interface and elucidate the modes of peptide-binding affinity and specificity. Use of molecular simulation is imperative in the disentangling of this complex interplay between composition, sequence and structure which govern binding affinity and specificity [32]. It was already described in Section 1.3.2 how we can obtain qualitative and quantitative data regarding the affinity of a GEPI to a given substrate, such as the thermodynamic and kinetic properties. Information regarding the peptide conformation in solution can also be obtained by experiment. Despite being possible to obtain information on the conformation of a peptide adsorbed on a surface as well as the key residues involved in the binding by experimental techniques [54], the fundamental basis of how peptides recognise an inorganic surface and how this could be manipulated to tailor binding affinity is not yet well understood. This is where molecular simulation may help in understanding the mechanisms involved in the binding of a peptide to a certain surface. Molecular simulations provide structural data such as distances of the residues to the surface, orientational data such as the orientation of the residues relative to the surface and backbone conformation as well as energetic information of the peptide at a given surface. In addition, molecular simulation also can provide thermodynamic quantities such as free energy differences, which are important to understand binding affinity.

1.4 Carbon Nanotubes

The breakthrough in the field of nanothechnology arised with the discovery of fullerenes in 1985 [55], the third allotropic form of carbon next to diamond and graphite. Even

though carbon filaments had been observed in the 50's, the official discovery of cylindrical fullerenes was only in 1991, when Iijima reported for the first time the preparation of multi-walled carbon nanotubes (MWCNTs) by arc-discharge of graphite electrodes [56]. Two years later, Iijima synthesised carbon nanotubes (CNTs) with only one layer, [57] the so called single-walled carbon nanotubes (SWCNTs). Since then CNTs have been the object of frenetic and intense research. Single-walled carbon nanotubes (SWCNTs) are graphene sheets rolled-up into cylinders with carbon atoms bonded by sp^2 hybrid orbitals, closed at both ends by semi-spherical caps. Multi-walled carbon nanotubes (MWCNTs) consist of sets of concentric graphite tubes that fit inside each other. The lenght of SWCNTs can vary from several hundred nanometers to several micrometers and the diameter can vary from 0.4 to 2 nm. In the case of MWCNTs the diameter can vary from 2 to 100 nm. The atomic structure of a carbon nanotube is described by the chiral indices (n, m), that specify the perimeter of the carbon nanotube on the graphene sheet as shown schematically in Figure 1.4. Also, the way the graphene sheet is wrapped gives rise to different indices (n, m) and different nanotubes. If $m = 0$, the nanotubes are called 'zigzag'. If $n = m$, the nanotubes are called 'armchair' and if they are different, the nanotubes are called 'chiral'. Depending on diameter length, CNTs can have conductor or semiconductor properties. If $n = m$ or $n - m$ is a multiple of 3, the nanotube is a semiconductor but the band gap is zero, and so we can use the term "metallic". Otherwise, it is a moderate semiconductor. Therefore, all armchair $(n = m)$ CNTs are metallic. This unique structure endows CNTs with an exceptional thermal stability, as well as outstanding mechanical, thermal and electrical properties that make them a good candidate for a wide variety of potential applications in the fields of nanotechnology and medicine.

However, for biological and biomedical applications, the extreme hydrophobicity

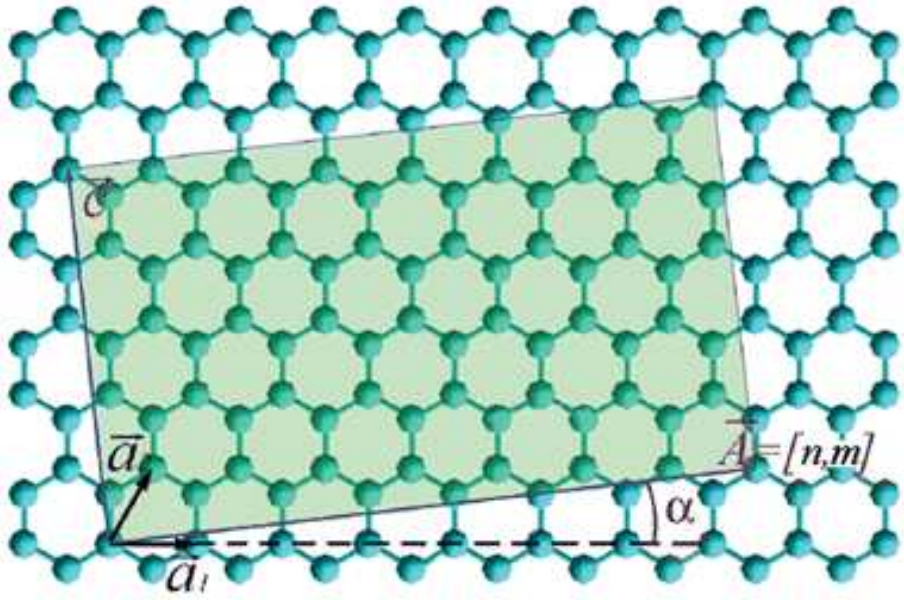

Figure 1.4: Schematic structure of a graphene with basis vector a_1 and a_2. The shadowed rectangle is the radial projection of carbon nanotube $(7, 1)$ with perimeter A and helical angle α.(Taken from reference [58].)

and lack of dispersion of CNTs constitutes a big limitation to the accomplishment of these applications [59]. Solubility of SWCNTs in water can be improved by chemically attaching various functional groups to the nanotubes. However, these modifications can perturb the intrinsic properties of SWCNTs, such as the electrical properties [60; 61]. Therefore, alternative approaches to solubilise SWCNTs have been proposed and investigated, such as the use of noncovalent [62] adsorption of surfactants [63], polymers [64; 65] and biomolecules [66; 59; 67]. In addition, there is the concern that exposure of CNTs may be toxic to humans [68]. CNT toxicity has been attributed to several factors including length, concentration, type of functionalization, duration and method of exposure [69]. Recent studies suggest that if CNTs below a given threshold (1 to 5 μm) [70] are wrapped with biomolecules such as peptides [71], they have limited toxicity.

On the other hand, longer CNTs are potentially toxic since they can accumulate in the intestine and kidneys. It is therefore important to study the interactions of peptides with CNTs, in order to tailor the binding affinity for specific applications.

1.4.1 CNT biomedical applications

Many potential applications have been proposed for CNTs, in the fields of nanotechnology, biology, and medicine. The increased interest in biomedical applications is due to the combination of properties that make CNTs unique, such as ordered structure, high aspect ratio, light weight, high mechanical strength and electrical and thermal conductivity [72]. CNTs have been explored as potential delivery vehicles for drugs [72; 73], peptides [74], proteins [75; 76], DNA [76] and RNA [77]. The cargo can be carried either on the outside of the CNT (attached to the side walls) or on the inside of the CNT. In addition, studies suggest that CNTs may be used in cancer phototherapy [78], by injecting multi-functionalised CNTs into the tumor and subsequently irradiating the tumor with a laser. Furthermore, it was shown that CNTs functionalised with biomolecules can localise the mitochondria of normal and cancerous cells and when irradiated with infrared light could induce cell apoptosis (cell death) [79]. A recent study, showed that CNTs can be used as antibacterial agents that attack the bacteria and degrade the cell integrity, and therefore causing its death [80]. CNTs have also been studied as a suitable substrate for the growth of cells for tissue regeneration [81]. In addition, CNTs can be used to build nanomotors that can enter inside cells to treat diseases [82].

1.5 Literature review of inorganic-binding peptides

It was Stanley Brown [7] who pioneered the selection of peptides that could bind to gold and chromium, out of 5 million different sequences by the phage-display technique.

Following this important work, Belcher *et al.* [14] identified peptide sequences that could bind to one inorganic surface over a range of others (*e.g.* binding specifically to GaAs but not to Si). In addition, Belcher *et al.* showed that peptides could differentiate between different crystallographic orientations of surfaces (*e.g.* binding to GaAs(001), but not to GaAs(111)). Hundreds of similar experiments have been reported since the publication of this influential work, mainly using the phage-display method to identify peptide sequences with strong affinity for a wide range of inorganic materials [83; 10; 16; 1; 84; 11; 85; 86; 87] such as metals, oxides and semiconductors.

The degree of binding affinity of a peptide onto a surface has been studied using techniques such as FM [14] and AFM [88; 89; 84; 90; 91; 27]. Hayashi *et al.* [89] used AFM to elucidate the mechanism of binding between Ti-binding peptide and the Ti surface. Furthermore, Goede *et al.* used AFM to study peptide adhesion on semiconductors (GaAs, Si, InP, GaP and Al_xGa_yAs). Recently, a number of studies have quantified the adsorption kinetics of the peptides on a surface by QCM [23; 25] and SPR [25; 12; 26]. Tamerler *et al.* [25] reported for the first time the kinetic parameters (such as the equilibrium constant, K_{eq}) and the binding energy by two different methods (QCM and SPR). Quantitative information on the peptide conformation at the surface has also been obtained by NMR experiments [54; 92] by characterising the secondary structure of the peptides.

A number of studies have also used combinatorial mutagenesis to identify sequences that bind to graphitic surfaces such as CNTs [93; 94; 88; 95; 96; 97], fullerenes [98] and nanohorns [99]. As mentioned in the previous Section, there is a significant interest in the applications of peptide binding CNTs, in the fields of nanotechnology [100] and medicine [61; 60]. Wang *et al.* [94] identified two peptide sequences (B1 `HWKHPWGAWDTL` and B3 `HWSAWWIRSNQS`) with strong affinity for CNTs by phage-display

experiments. Their results demonstrated the important role of the aromatic content in the sequences found. The histidine (H) and tryptophan (W) residues present in the sequences were found to be a key factor for the strong binding to CNTs. Since this paper was published [94], the importance of the aromatic content in the interaction between peptides and CNTs has been highlighted [88; 101], with recent emphasis on the tryptophan residue [90; 95; 102; 101]. Dieckmann *et al.* [88] showed that the interactions between peptide-coated nanotubes can be manipulated by specifying the amino acid content. The presence of aromatic residues such as phenylalanine (F), yields an interaction with the nanotube surface similar to the interaction between pyrene and CNTs [103], disfavouring peptide aggregation and promoting peptide-nanotube association. Furthermore, Dieckmann *et al.* [104] systematically varied the electron density of the aromatic residue phenylalanine by introducing an electron-donating (hydroxyl) and an electron-withdrawing (nitro) group on the benzene ring of phenylalanine. AFM measurements revealed that the ability to disperse CNTs increases with the increasing electron density of the aromatic residue [104].

The question of binding affinity of different peptide sequences to CNTs is not as straightforward as it seems. The nature of the peptide content and order of that content is crucial [32], and it is currently proposed that the interplay between sequence, structure and binding underpins the phenomena of binding affinity and specificity. Bioinformatic studies in combination with experiments [32] show that composition alone is not adequate to represent peptide-inorganic affinity and specificity. It is the sequential arrangement of the residues in a peptide together with the residue content that must be responsible for affinity and specificity behaviours. We cannot decide which residue yields greater binding to the CNT, because the binding energy is not necessarily the sum of the individual binding energies for each residue.

Although there is considerable growth in the experimental study of peptide-inorganic recognition, there are not as many modelling studies in this area. It is essential to understand the nature of molecular recognition and the degree of affinity of a peptide to a given surface, so that it can be rationalised, predicted and optimised in several applications [105]. As explained in Section 1.3.7, molecular simulation can function as a complementary tool along with the experimental characterisation methods in the realisation of this objective. Model approaches to describe protein-surface interactions have been proposed [106; 87; 107] and vary from coarse-graining [106; 108; 109; 110; 111] to atomistic descriptions [112; 113; 114; 115; 86; 112; 116; 9] and describe the non-bonded interactions either solely with Lennard-Jones (LJ) terms, or at best in combination with point-charge electrostatics. Molecular dynamics (MD) simulations of several phage-display selected peptides with different levels of affinity to the Pt(100) surface [108] were reported, showing good agreement with experimental data despite the fact that this study was done in vacuum and using only van der Waals (VDW) terms to describe the non-bonded interactions. MD simulations of selected peptides interacting with the Pt(111) surface in water was recently reported [112], and presented a relationship between sequence, structure and binding for metal binding peptides. Monte Carlo simulations of peptides in bulk water at blank charged surfaces [109] emphasised the important role of water in the conformational arrangement of the peptides at surfaces. Carravetta and Monti [113] suggested that MD simulations can capture the main characteristics of peptide-TiO_2 surface interactions in solution by comparison with *ab initio* calculations. Interactions between a histidine-tagged peptide and several surfaces Ni/Cu/Au were studied by MD simulations [115], and showed that peptides have different affinities to different metal surfaces. Heinz *et al.* [9] modelled the binding affinity of several short peptides to gold, palladium and palladium-gold bi-metal by MD sim-

ulations. Their findings suggested that strongly binding peptides tend to adopt a flat two-dimensional conformation in order to get most of the residues in close contact with the metal surface. On the other hand, weakly binding peptides tend to preserve most features of their three-dimenisonal conformation in solution, so that water molecules can be present between the surface and the residues. Skelton *et al.* [117] studied the adsorption of a hexapeptide at aqueous titania interface by MD simulations. Their results highlighted the important role of proline in the rigidity of the peptide backbone and consequently its role on the peptide binding to the surface. They proposed that a balance between peptide flexibility/rigity and the strength of the interaction between the peptide and the surface should exist in order to be capable of finding and maintain a good binding configuration. In our previous work [118] we performed MD simulations of single-walled CNTs interacting with strong and weak binding peptides as identified by Wang *et al.*. Our findings confirmed the experimental observations [94] of the differences in binding affinity for the sequences studied. In this thesis, the interactions between peptides and CNTs are studied; a review of peptides binding to CNTs along with a discussion of the past modelling studies will be given in Chapter 3.

1.6 Objective and outline of the thesis

Using peptides as synthesisers and scaffolds in the design and assembly of functional inorganic materials requires an understanding of the nature of the peptide recognition and binding to the solid. Although the number of materials-specific peptide sequences selected by combinatorial techniques keeps increasing, it is still not clear yet how peptides recognise an inorganic surface and how this could be tailored to control the behaviour of the material [1; 22]. Understanding how peptides interact with inorganic materials will enable the identification of "rules" or "guiding principles" by which interactions are

controlled and to explain and predict the behaviour. Molecular modelling of peptide-solid surfaces have the potential to provide further insights into the mechanisms that govern those interactions.

As mentioned in the previous Section, several studies have addressed the importance of the aromatic content on the peptide-CNT binding affinity [94; 88; 104; 119; 102; 101; 96]. In line with these recent studies, the aim of the work described in this thesis is to study the interaction between peptides and CNTs and other graphitic materials by molecular dynamics (MD) simulation.

The outline of this thesis is as follows:

Chapter 2 covers the main theoretical methods used in Computational Chemistry. This chapter does not have the objective of giving an exhaustive overview but rather to describe the main approaches employed in this work.

Chapter 3 is dedicated to the study of the influences of the aromatic content and interfacial shape on the binding affinity of peptides for graphitic surfaces.

Chapter 4 discusses the influence of the peptide sequence on the binding affinity to CNTs. The effects of the order of the residues in the peptide will be presented.

Chapter 5 focuses on the atomistic simulation of liquid water and graphitic surfaces: carbon nanotubes, fullerenes and graphite with an extended polarisable force-field. The study aims to demonstrate that the force-field is appropriate to use in future simulations of graphitic surfaces, biomolecules and water.

Finally, **Chapter 6** deals with the study of CNTs with defects and their impact on the binding affinity of peptides to CNTs.

The main conclusions of this thesis are highlighted in the last Chapter**Conclusions and Outlook**.

Chapter 2

Theoretical methods

In this project, in which the interaction of peptides with graphitic surfaces is studied, atomistic molecular dynamics (MD) simulations were used throughout. In this chapter, the available simulation techniques and their advantages and disadvantages are outlined.

2.1 Molecular Modelling: overview

Molecular simulation is concerned with solving problems related to representing the potential energy landscape (PEL) of chemical systems, and exploring the relevant regions of this landscape. This problem first demands a physically-reasonable inter-atomic potential to describe the chemical system. Particularly, in the case of peptide-inorganic interfaces a potential is required that simultaneously describes biomolecules and inorganic surfaces with reasonable faithfulness. Even though very few potentials have been designed to simultaneously describe both entities, some potentials have been specially tailored to describe these interactions. In terms of exploring the PEL (the conformational sampling issue), so far no approach can guarantee an exhaustive exploration of the PEL, and while a better approach to sample the PEL remains essential, there is growing

evidence suggesting that the peptide-inorganic interface typically supports many different strong binding configurations [118; 117; 120]. Therefore decent conformational sampling is critical as will be explained in Section 2.3.

2.2 Potential energy landscapes

The basis of potential energy landscapes (PEL) is the *Born-Oppenheimer approximation*, which proposes the separation of the electrons spatial distribution from the nuclear positions due to the nuclear mass being much heavier than the electron mass, and therefore the electrons are able to rapidly adjust to any changes in the nuclei positions. The consequence of this approximation is that the energy of a molecule in a given electronic state can be treated as a parametric function of the nuclear coordinates. The PEL has its origins in the fact that atoms and molecules are constantly moving. The absolute positions of the atoms in a molecule and the molecule itself are continuously moving as well as the relative positions of its bonds. Particular molecular conformations can be considered as wells on a $3N - 6(5)$-dimensional "surface' called the *potential energy surface* or PEL (sometimes called *hypersurface*) which relates the potential energy with the geometry (*i.e.* spatial arrangement of the nuclei) of the molecule [121; 122]. The PEL is incorporated in a $(3N + 1 - 6(5))$-dimensional space ($3N - 5$ and $3N - 6$ for linear non-linear molecules, respectively), where the extra dimension is the value of the potential energy. The potential energy of each molecular geometry is represented in the PEL by a single point on this hypersurface. The stable conformations of a molecule correspond to local minima in the PEL (where the first derivative of the energy is zero and and all second derivatives are positive). The points at which the potential energy is at a minimum correspond to stable configurations of the molecule. The minimum that corresponds to the most stable conformation has the lowest energy on the PEL

and is called the *global minimum*. Molecules may have many minima on the PEL and the pathway linking one minimum to another involves a transition to a point where the potential energy is higher than either of the two minima, the maximum of which is referred to as a *saddle point*. Minima and saddle points are also called stationary points since the first derivative of the potential energy with respect to all coordinates is zero. Transition states are first order saddle points. Figure 2.1 shows a schematic representation of a PEL, illustrating the more relevant points of the landscape.

As will be explained in the next Section (2.3), a PEL can be constructed with either quantum or molecular mechanics. In addition the motion of the nuclei can be explored either with classical or quantum mechanics. Sections 2.4 and 2.5 describe two techniques used to explore the most important regions of a PES.

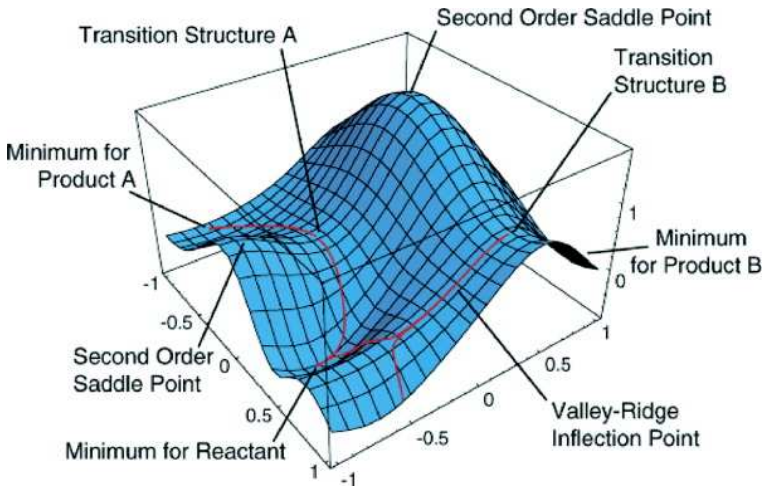

Figure 2.1: Schematic representation of a PEL, showing minima, transition states and saddle points. (From [123]).

2.3 Statistical mechanics: linkage between microscopic to macroscopic world

The aim of statistical mechanics is to provide a linkage between quantum mechanics and thermodynamics by deducing macroscopic properties of a system from the properties of molecules of the system. Macroscopic properties include entropy, internal energy, free energy, surface tension, dielectric constant, viscosity, etc., while molecular properties include molecular geometries, intra- and intermolecular forces [124; 125].

If we wish to measure a macroscopic property of an equilibrium system, we take the average of the property over the time of measurement. This is called a *time average*. Whereas if we wanted to calculate the value of that macroscopic property we would have to average over the micro-states (the quantum state) of the system. However, a calculation of a time-average is not feasible since it is not accessible by molecular simulation since the timescale is too short. Boltzmann and Gibbs developed *Statistical Mechanics* in which a single system evolving in time is replaced by a hypothetical collection of non-interacting replicas of the system. Each of these replicas is macroscopically identical, despite being in different micro-states, since each macroscopic state supports a wide variety of micro-states. This collection of identical systems is called an *ensemble*, and it is postulated that the time average of a macroscopic property is, in the long time limit, equal to the average of that property in the ensemble and is usually referred to as an *ensemble average*. This enables the replacement of the complex calculation of a time average by the calculation of an average over the ensemble.

For a given thermodynamic system of fixed composition N, volume V and temperature T (canonical ensemble), the quantum states with the same energy have equal probability of occurring. The probability p_j, of each micro-state depends on the tem-

perature T, and energy E_j:

$$p_j = ae^{-\beta E_j}, \tag{2.1}$$

where $\beta = 1/k_B T$ (being k_B the Boltzmann constant) and a a function of N, T and V. a can be obtained by normalisation: $\sum_j p_j = a \sum_j e^{-\beta E_j} = 1$ so equation 2.1 becomes:

$$p_j = \frac{e^{-\beta E_j}}{\sum_j e^{-\beta E_j}} \tag{2.2}$$

$$= \frac{e^{-\beta E_j}}{Z_{NVT}}, \tag{2.3}$$

where Z_{NVT} is called the *canonical partition function*:

$$Z_{NVT} = \sum_j e^{-\beta E_j}, \tag{2.4}$$

and is the sum over all possible quantum states of the system for a given composition and volume. The partition function constitutes a bridge between the quantum mechanical energy states of a macroscopic system and its thermodynamic properties [125]. Since the quantum energy levels of a system with a large number of molecules are highly degenerate *i.e.* different quantum states having the same energy, the probability of a system having energy E_i is given by:

$$p(E_i) = W_i e^{-\beta E_i}/Z_{NVT}, \tag{2.5}$$

where W_i is the *degeneracy*, *i.e.*, the number of states with energy E_i [124; 125].

The characteristic thermodynamic function for a canonical ensemble is the *Helmholtz free energy*, A, in which the natural variables of A are N, T and V. A can be derived in terms of Z by taking into account the fact that $A = E - TS$ (where E is the internal energy and S is the entropy of the system):

$$A = -k_B T \ln Z_{NVT}. \tag{2.6}$$

In the case of the isothermal-isobaric ensemble the corresponding thermodynamic function is dependent on the composition N, pressure P, and temperature T, and is the *Gibbs free energy*, G:

$$G = -k_B T \ln Z_{NPT}, \tag{2.7}$$

where Z_{NPT} is the isothermal-isobaric partition function. The change in free energy ΔG depends on the change in enthalpy, ΔH(H = E + PV) and on the change in entropy, ΔS according to:

$$\Delta G = \Delta H - T\Delta S. \tag{2.8}$$

The free energy difference between two macro-states tells us if it is favourable or not for the system to move to another state. In other words, it tells us about the population of both states at equilibrium.

Entropy is proportional to the number of states W, where W corresponds to the total number of energy levels that have a considerable probability of being occupied:

$$S = k_B \ln W. \tag{2.9}$$

Equation 2.9 is the *Boltzmann principle* and tell us that as the number of micro-states increases, the entropy of the system increases. The entropy is therefore a measure of how many states there are in a specific region of phase space.

If we wish to calculate entropy-related ("statistical") quantities such as ΔA, ΔG or even ΔS we face some difficulties associated with the simulation techniques. The problem is that these quantities are directly related with the partition function Z and not with its derivatives. If we could sample all the phase space, we could find the probabilities of all states and we would then be able to sum over all the possible micro-states of the system and therefore obtain Z. In practice this is not possible, since even for small systems the number of micro-states is huge. In order to reduce

the amount of phase space that has to be sampled, simulation techniques only sample the relatively low energy states, since these are the ones that have a big contribution to the partition function Z. The high energy states do not contribute much to the sum $Z = \sum_j e^{-\beta E_j}$. The consequence of this is that entropic or thermal properties (such as free energy, chemical potential, entropy) are difficult to determine accurately by simulation without resorting to special techniques. This is why several techniques have been specially developed to study these quantities as outlined in the next Section.

2.3.1 Free Energy calculations

Free energies are important if we wish to know the relative stability of several states of a system. According to the second law of thermodynamics, the entropy S of a closed system with energy E, volume V, and number of particles N, is maximum when the system is in equilibrium. From this principle, it can be deduced the corresponding equilibrium conditions for systems that can exchange heat, particles or volume with a reservoir. For a system at fixed temperture T, volume V, and number of particles N that is in contact with a heat bath, the Helmholtz free energy, $A = E - TS$, will be at a minimum in equilibrium. Similarly, for a system of fixed N particles, pressure P and T, the Gibbs free energy, $G = A + PV$, will be at a minimum. Hence, in order to compare the stability of two phases at a given temperature we just need to compare the free energies of the two phases. Unfortunately, in terms of simulation, this is not as simple as it appears since as explained in Section 2.3, it is not possible to measure the free energy (or entropy) directly in a simulation [126].

Before considering the problem of calculating free energies let us take a look at the concept of *phase space*. A system containing N atoms requires $6N$ values to describe its state (three coordinates per atom and three components of the momentum), where

each set of $3N$ positions and $3N$ momenta define a point in the $6N$-dimensional phase space. Therefore, an *ensemble* is a collection of points in the phase space. It is impossible to scan all the points of the phase space with molecular dynamics (MD) simulations or the configurational space with Monte Carlo (MC) for anything but the most trivial system (see Section 2.3). If it was possible, the trajectory would be called *ergodic* and would be independent of the initial configuration. In addition, we would be able to calculate the partition function just by adding the values of $exp(-E/k_BT)$ (where E is the potential energy, k_B is the Boltzmann constant and T is the temperature) as will be explained in Section 2.3. The fundamental assumption is that if the MD simulation is sufficiently long so that representative conformations of the system have been sampled, time averages give a good approximation to ensemble averages. Because the phase space of the systems like the ones I have been studying in this project is enormous, simulations only provide an approximation to the "true" energies, and in order to improve the sampling of the phase space several simulations using different starting configurations should be carried out.

One of the most important properties in thermodynamics is the *free energy* which can be expressed as the *Helmholtz function*, A, or the *Gibbs function*, G. The Helmholtz free energy is appropriate in systems with a constant number of particles, volume and temperature (NVT), and the Gibbs free energy is appropriate in systems with a constant number of particles, temperature and pressure (NPT). One of the limitations of molecular modelling approaches such as MC and MD simulation is that these methods only explore regions of relatively low energy and cannot necessarily access regions that require large energetic barriers to be traversed. Since flexible macromolecules typically support many minima separated by barriers [122], it is challenging to accurately calculate the free energy in these systems, since naive implementation of MD or MC might not

sufficiently sample the high energy regions. In addition, the related properties such as changes in entropy and the chemical potential,μ, are also difficult to calculate for large molecules. All energy calculations involve a difference between two states (in which the reference state is explicit). Free energy differences is a problem that can be determined with MD or MC. Several methods have been proposed to tackle the problem of calculating free energy differences, such as thermodynamic perturbation, thermodynamic integration and slow growth; a brief description of each of these techniques follows.

2.3.1.1 Thermodynamic Integration

The Gibbs free energy G, can be expressed by:

$$G = -k_B T \ln Z(N, P, T) \tag{2.10}$$

where k_B is the Boltzmann constant and Z is the partition function. The derivative of the potential energy, U, can be used to calculate the free energy difference. Thermodynamic integration involves performing a series of simulations corresponding to discrete values of λ between 0 and 1, where $U(\lambda = 0)$ corresponds to the potential energy of the reference state and $U(\lambda = 1)$ corresponds to the potential energy of the state of interest. The subsequent calculation of the area under the graph will give the free energy change.

$$G(\lambda = 1) - G(\lambda = 0) = \int_{\lambda=0}^{\lambda=1} \langle \frac{\partial U(\lambda)}{\partial \lambda} \rangle_\lambda d\lambda, \tag{2.11}$$

where λ is a parameter that characterises the path between the reference state and the state of interest, ensuring that the path is reversible, *i.e.*, the path in the opposite direction is exactly the same. The accuracy of the calculated free energy change is dependent on the number of λ values used, usually between 10 and 100 λ values are used.

2.3.1.2 Thermodynamic Perturbation

According to thermodynamic perturbation theory, the free energy difference in going from the state A to the state B is simply $G_{AB} = G_B - G_A$ and this method is used when the energy difference between the two states is much larger than k_BT, and therefore one of the states is poorly sampled. Since the free energy is a state function, it does not depend on the path, *i.e,* the way the system achieved that state, and consequently an intermediate state between A and B can be introduced in order to increase the overlap in the phase space and thus increase the sampling:

$$\begin{aligned} \Delta G &= G(B) - G(A) && (2.12)\\ &= (G(B) - G(1) + (G(1) - G(A)) && (2.13)\\ &= k_BT \ln\Big[\frac{Z(B)}{Z(1)}.\frac{Z(1)}{Z(A)}\Big], && (2.14) \end{aligned}$$

where Z is the partition function. Since the intermediate term cancels out we can choose as many intermediate states as we like in order to obtain good overlaps and reliable free energy differences:

$$\Delta G = k_BT \ln\Big[\frac{Z(B)}{Z(X)}.\frac{Z(X)}{Z(X-1)}.\frac{Z(X-1)}{Z(X-2)}\cdots\frac{Z(2)}{Z(1)}.\frac{Z(1)}{Z(A)}\Big], \quad (2.15)$$

where X is the number of intermediate states.

2.3.1.3 The Slow Growth method

The slow growth method is similar to the thermodynamic integration method, but instead of λ being constant for a given trajectory, it changes by a small constant amount at each step of the simulation. The free energy difference is thus given by:

$$\Delta G = \sum_{i=1;\lambda=0}^{i=N_{step};\lambda=1} (U_{i+1} - U_i), \quad (2.16)$$

where N_{step} is the number of steps in the simulation and U_i is the potential energy at the timestep i.

In summary, for peptide-inorganic surface systems, all such calculations as outlined above are very challenging and no one yet has been able to properly calculate genuine adsorption free energies, ΔG_{ads}. One of the main reasons is, as mentioned before, the insufficient sampling of phase space. The massive phase space of biomolecules makes a complete sampling of these landscapes impossible.

2.4 Monte Carlo simulations

In Monte Carlo (MC) simulations, a sequence of points in the configurational space is created from an initial configuration, by making random changes to the coordinates of a randomly chosen particle (molecule or atom), using a random number generator. These random changes are also known as a "random kick" [127; 128; 122]. If the energy of the new configuration is lower than the energy of the previous configuration, the new configuration is accepted and a random kick is again applied. If the energy of the new configuration is higher than the previous one, then the *Boltzmann factor* of the energy difference $exp(-\Delta U/k_BT)$ (where U is the potential energy and T is the temperature) is calculated and compared to a random number in the interval between 0 and 1. If the Boltzmann factor is greater than the random number, then the new configuration is accepted, otherwise it is rejected and the previous configuration is used in the next move. This *Metropolis* procedure can be summarised by the following relation:

$$rand(0,1) \leq exp(-\Delta U(\mathbf{r})/k_BT). \tag{2.17}$$

The configurations in the ensemble are assured to obey a Boltzmann distribution by this Metropolis procedure. States with low energy have a greater probability of being

accepted than configurations with a higher energy. However, the opportunity to accept configurations of higher energy gives MC methods the chance to explore the configurational space by avoiding being stuck in a local minimum, and by permitting moves to states of higher energy, *i.e.*, giving a better coverage of the configurational space. The desired properties are calculated for each accepted configuration and at the end of the simulation the average $\langle A \rangle$ of these properties is determined by averaging over the number of calculated values, M:

$$\langle A \rangle = \frac{1}{M} \sum_{i=1}^{M} A(\mathbf{r}). \tag{2.18}$$

As opposed to molecular dynamics (MD) simulations (2.5), MC is a non-deterministic method (stochastic method), since each configuration only depends on the previous one and on random numbers and therefore two simulations starting with the same configuration will not necessarily create the same sampling (since the generated random numbers could be different). While MD simulations measure time averages in an ensemble, MC simulations probe the canonical ensemble, yielding ensemble averages. An advantage of MC is that it is does not require the forces on the atoms to be calculated, and thus it is not always necessary to re-calculate the energy of the entire system but only the contributions corresponding to the part of the system that were changed in that step. This significantly reduces the computational time. Another advantage is that thermodynamic properties calculated from a MC simulation usually converge rapidly [127]. MC methods are better at exploring the translational and rotational space of small molecules since for large molecules such as peptides, several simultaneous perturbations are required in order to generate reasonable conformational changes. Such movements are difficult to generate by random perturbations of either the Cartesian or internal coordinates since even small movements may lead to a large energy increase. The consequence of this is a slow configurational space coverage. Another disadvantage of MC is that this method

is time-independent and therefore it is not appropriate for the study of time-dependent properties.

So while MD is more appropriate to explore the local phase space, MC is more effective for conformational changes because since it can go "uphill" in the potential energy, it might end up in a different area of the configurational space [127]. In the study of the interaction of peptides with graphitic surfaces, MD simulation was chosen over MC due to its ease of use, since the generation of random "kicks" for big molecules can be potentially difficult. Like in the case of MD simulations, MC simulation involves two stages: an equilibration period and a production stage. During the equilibration period, the thermodynamic properties and other parameters are monitored until stable values are reached. After equilibration has been reached, the production period can commence.

2.5 Molecular dynamics simulations

The basis of molecular dynamics (MD) simulations is the integration of Newton's second law, $F = ma$, to generate consecutive configurations of the system. This results in a trajectory which describes the positions and velocities of the molecules in the system as a function of time [122]. Taking into account that the forces and the derivatives of the potential energy are related in the following way:

$$\frac{\delta U_i(\mathbf{r}_i, t)}{\delta \mathbf{r}_i} = -F_i(\mathbf{r}_i, t) \tag{2.19}$$

where the vector $\mathbf{r}_i(x_i, y_i, z_i)$ describes the position of the particle i and U_i is the potential energy of the particle of mass m_i at position $\mathbf{r}_i$. By solving the differential form of Newton's second law we obtain a trajectory of the particles as a function of time:

$$F_i = m_i a_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2} \tag{2.20}$$

and therefore,

$$\frac{-dU_i}{d\mathbf{r_i}} = m_i \frac{d^2 \mathbf{r}_i}{dt^2}. \tag{2.21}$$

2.5.1 Integration algorithms

Several methods exist to integrate the equations of motion. An ideal algorithm, as like any other computer algorithm, should be efficient, require minimal memory and be easy to program. However, the most important requirements in MD simulations are that the integration algorithm should conserve energy and momentum, be time-reversible, and should permit a long time step, δt, to be used [122]. All algorithms are based on the approach of numerical integration of the equations of motion over the timestep δt. During each step the total force on each particle is assumed to be constant. For each particle and each step the total force F_i is calculated as the sum of the negative nuclear derivatives of the contributions to the interaction energy U_i:

$$F_i = -\sum_j \nabla U_i \tag{2.22}$$

where F_i is the total force on each particle i, $U_i = \sum_{j \neq i} U_j$ is the potential energy of particle i and the gradient of U_i is given by $\nabla = \frac{\partial U_i}{\partial x}\mathbf{e_x} + \frac{\partial U_i}{\partial y}\mathbf{e_y} + \frac{\partial U_i}{\partial z}\mathbf{e_z}$, where e_x, e_y, e_z are the unit vectors. Once the forces have been calculated by taking the derivatives of the the potential energy, Newton second's law is solved and the accelerations of the particles can be determined. These can be used to calculate the positions and velocities at time t, and can then be used to calculate the positions and velocities at a time $t + \delta t$.

Common to all the integration algorithms is the assumption that the positions and dynamic properties (such as velocities and accelerations) can be described as a

Taylor series expansion, meaning that for a given coordinate r it is possible to predict the coordinates at a small time δt ahead of the current time, *i.e.*, $r(t+\delta t)$:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + \frac{1}{6}\delta t^3\mathbf{b}(t) + \cdots \tag{2.23}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \delta t\mathbf{a}(t) + \frac{1}{2}\delta t^2\mathbf{b}(t) + \frac{1}{6}\delta t^3\mathbf{c}(t) + \cdots \tag{2.24}$$

$$\mathbf{a}(t+\delta t) = \mathbf{a}(t) + \delta t\mathbf{b}(t) + \frac{1}{2}\delta t^2\mathbf{c}(t) + \cdots \tag{2.25}$$

$$\mathbf{b}(t+\delta t) = \mathbf{b}(t) + \delta t\mathbf{c}(t) + \cdots \tag{2.26}$$

where **r** are the positions, **v** is the velocity (the first derivative of the positions with respect to time), **a** is the acceleration (the second derivative of the positions with respect to time), **b** is the third derivative, $\cdots$and so on.

2.5.1.1 Verlet algorithm

The Verlet algorithm [129] was used in this study and is one of the most extensively used method to integrate the equations of motion in MD simulations. In the Verlet algorithm, the positions at $t+\delta t$ [$\mathbf{r}(t+\delta t)$] are calculated by using the positions and accelerations at time t, and the positions from the former step $\mathbf{r}(t-\delta t)$. The relationship can be derived from:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + \cdots \tag{2.27}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) - \cdots \tag{2.28}$$

The sum of these expressions gives:

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2\mathbf{a}(t). \tag{2.29}$$

One of the drawbacks of this method is that the velocities cannot be calculated explicitly. Despite the velocities not being necessary for the time evolution, they are required to

the calculation of the kinetic energy *K*, which is necessary to test the conservation of the total energy ($E = K + U$) and to calculate the temperature. However, there are other ways to determine the velocities from these relations. For example, velocities can be calculated at the half-step, $t + \frac{1}{2}\delta t$:

$$\mathbf{v}(t + \frac{1}{2}\delta t) = [\mathbf{r}(t + \delta t) - \mathbf{r}(t)]/\delta t. \tag{2.30}$$

Other disadvantages are related with the loss of precision upon the calculation of the positions **r**($t + \delta t$), since we are adding a small term (δt**a**(t)) to the difference of two larger terms, 2**r**(t) and **r**($t - \delta t$). In addition, since the velocities are not calculated explicitly they are only available after the positions have been calculated in the next step. The Verlet algorithm is unable to self-start, since the new positions **r**($t + \delta t$) are obtained from **r**(t) and **r**($t - \delta t$), and at $t = 0$ only one set of positions is available, and therefore, the positions at **r**($t - \delta t$) need to be determined by other means. On the other hand, the Verlet algorithm is easy to implement, accurate and stable and has minimal storage requirements since it only stores two sets of positions ($\mathbf{r}(t)$ and $\mathbf{r}(t - \delta t)$) and the accelerations $\mathbf{a}(t)$. It is time-reversible, force and momentum conservative, and it is shown to conserve the energy properties even with long time steps [128].

2.5.2 Timesteps

The timestep in MD simulations is the interval δt in which the total force is assumed to be constant. The timestep should be small enough to capture all the relevant information about the system but not too small, otherwise the phase space would be poorly sampled. An integration timestep of 1 fs was used in our simulations which is short enough to cover the smallest possible molecular timescale (in this case the period of the X—H bond vibration).

2.5.3 Thermostats

The simulations performed in this study were carried out at room temperature since this is the temperature at which most of the relevant experiments are carried out. Thermostats are used to keep the temperature of a system constant. Before proceeding to the description of the thermostat used in these simulations it is important to understand the meaning of "constant temperature". The instantaneous temperature of a particle is related to its kinetic energy through the linear momentum $\mathbf{p}$.

$$\sum_{i=1}^{N} \frac{|\ \mathbf{p}_i^2\ |}{2m_i} = \frac{k_b T}{2}(3N - N_c), \tag{2.31}$$

where $\mathbf{p}_i = m_i.\mathbf{v}_i$ is the linear momentum of particle i with mass m_i and $3N - N_c$ is the total number of degrees of freedom, where N_c is the total number of independent internal constraints (such as fixed bond lengths and angles). This is based on the principle of *equipartition of energy* which states that each degree of freedom has an average energy of $k_B T/2$. Therefore, constant temperature does not mean the kinetic energy per particle being constant, as we would not be simulating the true constant temperature of the ensemble. In fact, the instantaneous kinetic temperature fluctuates while keeping the average ensemble temperature constant [126]. The Nosé-Hoover thermostat was used in our simulations to keep the temperature around an average room temperature. The basic idea of this thermostat is the consideration of a heat bath as an integral part of the system, by adding an additional degree of freedom s which represents that reservoir and is associated with an artificial mass Q and a conjugated momentum $\mathbf{p}_s$ [128]. This extra degree of freedom has the effect of an external system on the system of study, allowing the flow of the energy from the bath to the system and vice-versa. The equations of motion can be re-written by replacing the scaled momentum p_s with a new "real" momentum $\mathbf{p}$: $[(\mathbf{p}_s/s) \rightarrow \mathbf{p}]$. The new set of equations of motion are as

follows:

$$\frac{d\mathbf{r}}{dt} = \mathbf{p}/m \tag{2.32}$$

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} - \frac{\mathbf{p}_s}{Q}\mathbf{p} \tag{2.33}$$

where,

$$\frac{d\mathbf{s}}{dt} = \frac{s\mathbf{p}_s}{Q} \tag{2.34}$$

$$\frac{d\mathbf{p}_s}{dt} = \sum \Big[\frac{\mathbf{p}^2}{m} - k_B T\Big]. \tag{2.35}$$

The scaling parameter s can be completely ignored by writing the equations as follows [130]:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} - \xi\mathbf{p} \tag{2.36}$$

where ξ is the friction coefficient ($\xi \propto \mathbf{p}_s$) and is given by the first-order differential equation:

$$\frac{d\xi}{dt} = \frac{f}{Q}(k_B\tau - k_B T), \tag{2.37}$$

and Q is the thermal inertia parameter (thermostat degree of freedom) which controls the rate of temperature fluctuations and can vary between 10-10000 kcal/mol ps^2, f is the number of degrees of freedom, T is the desired temperature and τ is the current kinetic temperature at time t. τ is related to the kinetic energy K of the system by the following relation:

$$K = \frac{3N}{2}k_B T, \tag{2.38}$$

where N is the number of molecules in the system. The Nosé-Hoover thermostat is used to maintain a rigorous canonical ensemble.

2.5.4 Barostats

In the simulations of graphitic surfaces in water, the pressure and temperature were kept constant. In a MD simulation at constant pressure the volume is a dynamical variable that is allowed to vary in order to keep the pressure constant. The pressure was kept constant with the Nosé-Hoover barostat, which works in a similar way to the Nosé-Hoover thermostat. In an (N, P, T) ensemble both the temperature and pressure are controlled by the combination of the Nosé-Hoover thermostat and barostat. The equations of motion are as follows [128]:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} - (\chi + \xi)\mathbf{p} \tag{2.39}$$

$$\chi = \frac{dV}{dt}/3V \tag{2.40}$$

$$\frac{d\chi}{dt} = \frac{P_0 - P}{t_P^2 k_B T}V \tag{2.41}$$

$$\frac{\xi}{dt} = (\sum_i \mid \mathbf{p}_i^2 \mid m - f k_B T)/Q \tag{2.42}$$

where $\mathbf{p}$ is the momentum, χ is the volume scaling factor, t_P is the relaxation time for the pressure fluctuations and can vary from a few hundred ps to a few ns, V is the volume of the system and P_0 and P are the instantaneous and desired pressure, respectively.

2.5.5 Periodic boundary conditions

In the study of graphitic surfaces in water we modelled an infinite CNT and an infinite graphite surface which was accomplished by the use of *periodic boundary conditions* in order to avoid edge effects and to enable the simulation of a relatively small number of particles. When using periodic boundary conditions, an infinite lattice is formed by the replication of the original simulation cell throughout space in three dimensions. In this way, there is a periodic image of each molecule in the original (central) cell in each

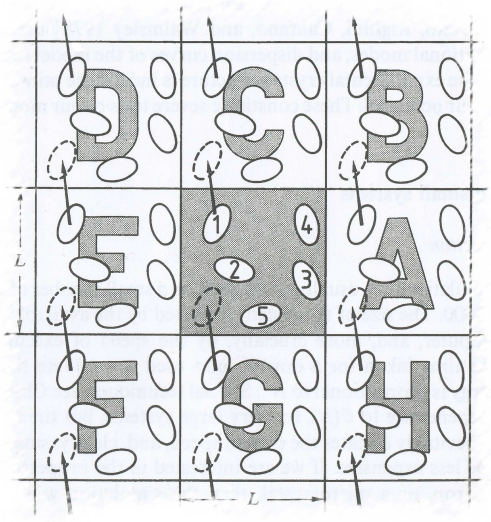

Figure 2.2: Schematic of a two-dimensional periodic system. As molecule 1 leaves the box it is replaced by its image which moves from box G into the central box (Taken from [128]).

of the neighbour cells in such a manner that as a molecule in the original cell moves, its periodic image in each of the adjacent cells moves in the same way. If a molecule leaves the central cell during the simulation, it is replaced by one of its periodic images that enters through the opposite side of the cell, as schematically illustrated in Figure 2.2. This process enables the conservation of the number of molecules in the central cell [128; 122]. Despite the cubic box being the easiest to visualise and implement, other shapes are also used depending on what is more convenient for a given simulation. However, one of the biggest limitations associated with the use of periodic boundary conditions is the inability to simulate fluctuations of a system at a wavelength greater than the length of the cell (*e.g.* of importance for simulation of a liquid close to the gas-liquid critical point).

2.5.5.1 Potential truncation and minimum image convention

The calculation of the non-bonded energies is one of the most time-consuming aspects of molecular simulation. In principle, in the calculation of the non-bonded interactions, all pairs of atoms in the system and corresponding images of the adjacent cells should be considered. But in this case we would have an infinite number of terms. Since the Lennard-Jones (LJ) potential falls off very quickly with distance, a *non-bonded cutoff* can be applied together with the *minimum image convention*. In the minimum image convention, the energy (or force) is calculated by considering each molecule at the centre of a region which has the same size and shape as the basic unit cell. Each molecule interacts with all the molecules whose centres lie within that region, that is the closest molecule or periodic images of the other $N-1$ molecules (where N is the number of molecules in the system). The application of this convention assures that a given atom does not interact with its own image or with the same molecule twice. When a cutoff is appied, the interactions between all pairs of atoms that are further from the cutoff value are set to zero. When periodic boundary conditions are applied, the cutoff distance should not be greater than half the shortest length of the box. The drawback of this is that by truncating the calculation we are no longer calculating the exact property but an approximation, so cell size effects should be explored for the properties under study.

2.5.5.1.1 Ewald Summation Handling long-range interactions arising from, *e.g.* charge-charge and dipole-dipole interactions, can be problematic due to their range usually being greater than half the box length. The problems related with the sum of the electrostatic interactions are the slow convergence and both the positive and negative terms form a divergent series if added alone. In addition, the sum is dependent on the order in which the terms are considered. The potential energy between the charges in

the central box and all the periodic boxes can be written as:

$$U_{elec} = \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 \mid \mathbf{r}_{ij} + \mathbf{n} \mid}, \tag{2.43}$$

where q_i and q_j are the partial charges on particles i and j, r_{ij} is the distance between the particles i and j and ε_0 the permittivity of free space. The sum over $\mathbf{n}$ is the sum over all cubic lattice points, $\mathbf{n} = (n_x L, n_y L, n_z L)$ with n_x, n_y, n_z being integers and L the size of the central box. The issue with truncating these interactions is that it gives rise to considerable inaccuracies. Several methods exist to deal with the problem of the long-range forces; the one that was used in this work is the *Ewald summation*. The Ewald sum is a method to efficiently sum the interaction between an ion and all its periodic images [128]. This is accomplished by surrounding each point charge with a charge distribution of equal magnitude and opposite sign which spreads out the charge, such as a Gaussian distribution. The effect of adding a Gaussian distribution is that it acts like an ionic species, screening the interaction between the neighbouring charges. The consequence of this is that the screened interactions are short-ranged and therefore the total screened potential can be calculated by adding the interactions between all molecules in the central box and all their images in the real space. Since a charge distribution of the same magnitude and same sign as the original charge is also added, it cancels the overall potential of the arrangement of screening charges. This distribution of cancelling charges is summed in reciprocal space using Fourier transforms. The final electrostatic energy is then the sum of the real space contribution and the reciprocal space contribution minus the interaction of the cancelling distribution centred at $\mathbf{r}_i$ with itself. The sum can be expressed as follows:

$$\varphi(\mathbf{r}) = \varphi_{sr}(\mathbf{r}) + \varphi_{lr}(\mathbf{r}) \tag{2.44}$$

where $\varphi_{sr}(\mathbf{r})$ is the short-range term which is summed in real space and $\varphi_{lr}(\mathbf{r})$ is the long-range term which is added in reciprocal space.

2.5.6 Initial conditions

The initial configuration of a system should be carefully selected, since it can significantly influence the progress of the simulation. It is usual to initiate the simulation with a configuration similar to the state we intend to simulate, making sure that there are no interactions of high energy, since this may trap the system in a high energy state or lead to the system exploding (if in an NPT ensemble). In addition, simulations with different initial configurations can be performed in order to improve the search for the lowest energy geometries. Another important aspect in initialising a MD simulation is the assignment of the initial velocities of the atoms. The initial atom velocities are usually determined from a random distribution such that the overall linear momentum is zero:

$$P = \sum_{i=1}^{N} m_i v_i = 0, \tag{2.45}$$

where the velocities v_i are randomly selected from a *Maxwell-Boltzmann* distribution at the temperature of interest, which describes the probability of an atom i of mass m_i of having velocity v_i in the x direction at a temperature T:

$$p(v_{ix}) = (\frac{m_i}{2\pi k_B T})^{\frac{1}{2}} exp[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}] \tag{2.46}$$

The temperature can be calculated from the velocities using the relation in 2.31. The Maxwell-Boltzmann distribution is a Gaussian distribution and is created by a random number generator between 0 and 1 [122; 128]. The generated random number is then converted into a Gaussian distribution:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{(x- < x >)^2}{2\sigma^2}], \tag{2.47}$$

where $p(x)$ is the probability of generating a value from a Gaussian distribution with mean $< x >$ and variance $\sigma^2 =< (x- < x >)^2 >$. The same equations are applied to the y and z components.

2.6 Construction and representation of the Potential Energy Landscape (PEL)

As mentioned before, the Born-Oppenheimer approximation enables the separation of the electronic from the nuclear motion. This enables solving the electronic Schrödinger equation (described in Section 2.6.1) for a given spatial arrangement of the nuclei using the nuclear positions as parameters. The PEL is defined as the set of the total potential energies (*e.g.* as obtained by solving the Schrödinger equation in 2.48) for all possible nuclear arrangements for a given electronic state. Solving the electronic Schrödinger equation for a given set of nuclear coordinates is computationally expensive for large systems. Quantum mechanics can be used to construct and explore a PEL, however, it is currently impossible to construct a PEL for large molecules. In addition, the motion of the nuclei on the PEL can be described with either classical or quantum dynamics. However, quantum dynamics on a PEL is beyond the scope of this thesis.

Useful information can, however, be obtained by limiting the calculations to an interesting part of the PEL, as for example the low-energy nuclear geometries. The PEL, can be calculated with quantum mechanics "on-the-fly" (*ab initio* dynamics), however quantum mechanical calculations for large molecules are prohibitively expensive. An alternative, is the use of a parameterised potential which is a function of internal nuclear coordinates only and is, in general, based on experimental data or quantum mechanical data. Such a potential is called an *empirical force-field*; the term "force-field" has also

been used by spectroscopists for the set of force constants [131]. Force-field methods are also known as *molecular mechanics* (MM) methods, and a description will be given in 2.6.2. Methods in which the objective is solving the electronic Schrödinger equation are usually referred to as *electronic structure calculations*. A very brief review of these main techniques is presented in the next Section (2.6.1). Another widely used quantum mechanical method is *density functional theory* (DFT) in which the effect of electron correlation is modelled as a function of the electron density, and the electron density is (in principle) the major physical quantity and not the wavefunction. An abbreviated outline of DFT is given in Section 2.6.1.2.

2.6.1 Electronic structure theory methods

Since electrons are very light particles they cannot be correctly described by classical mechanics. Therefore, in order to describe the electron distribution in detail we need to use quantum mechanics.

The time-independent form of the Schrödinger equation for a single particle (*e.g.* an electron) of mass m, moving through space with a position vector $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ in an external field $V(r)$ is given by:

$$\{-\frac{\hbar^2}{2m}\nabla^2 + V(r)\}\Psi(\mathbf{r}) = E\Psi(\mathbf{r}), \tag{2.48}$$

where Ψ is the *wavefunction*. It is the square of the modulus of Ψ that has a physical meaning: $\|\Psi\|^2 = \Psi.\Psi^*$ gives the probability distribution of the particle, where Ψ^* is the complex conjugate of the wavefunction.
E is the total energy of the particle and $\nabla^2 = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})$ is the *Laplacian operator*. The left-hand side of equation 2.48 can be abbreviated to $\mathbf{H}\Psi$, where $\mathbf{H}$ is

known as the *Hamiltonian*; this operator corresponds to the total energy of the system:

$$\mathbf{H} = -\frac{\hbar^2}{2m}\nabla^2 + V. \tag{2.49}$$

The Schrödinger equation for one particle is thereby symbolically written as:

$$\mathbf{H}\Psi = E\Psi. \tag{2.50}$$

The Schrödinger equation is an eigenvalue problem. It is not possible to solve exactly the Schrödinger equation for molecular systems with more than one electron and thus solutions for larger systems are only approximations to the real solution. The approximations involved concern the Hamiltonian and/or its operand which can be the wavefunction (*ab-initio* and semi-empirical methods) or the electron density distribution (DFT and Kohn-Sham-DFT- methods). The Schrödinger equation in equation 2.50 yields an energy and a corresponding wavefunction for every possible electronic state (with a given spin multiplicity). *ab initio* means " from first principles". In contrast to semi-empirical models, *ab initio* models are derived without any recourse to experimental data (apart from natural constants). On the other hand, semi-empirical models are in part informed by experimental data (such as molecular energies and geometries). Despite these methods being computationally cheaper than *ab initio* methods, they are limited to systems for which those parameters are known.

2.6.1.1 Hartree-Fock (HF) theory

HF is the simplest type of *ab initio* electronic structure theory. In the HF model electrons interact through coulombic and exchange and the correlation energy is exactly zero. Each electron can be described by a spin-orbital, a product of a spin-function and a spatial function. The total wavefunction is given by the antisymmetrized product of

all spin-orbitals, for example one or a fixed linear combination of several Slater determinants [127]. This is because electrons are indistinguishable particles and therefore the overall wavefunction must be antisymmetric, *i.e.*, change sign upon exchanging any pair of electrons.

Because HF only considers the average electron-electron interactions, the major limitations of HF is the neglect of electron correlation. HF is a *variational* method since its equations are derived from the *variational principle*, which states that the energy calculated with an approximation to the exact wavefunction will always be greater than or equal to the exact energy. Therefore, the lower the energy, the closer the approximate wavefunction will resemble the true wavefunction. The HF wavefunction is obtained by imposing the condition that the energy should be stationary, *i.e.*, the first derivative of the energy, δE, with respect to the spin orbitals is zero, under the constraint of the orthonormality of the spin-orbitals. In addition, HF theory is a *self-consistent* approach, which means that in the calculation the spin-orbitals are improved step-by-step until the point where the electronic solutions that correspond to the total energy are unchanged for all the electrons and the total energy cannot get any lower, *i.e.*, self-consistency is achieved. The solutions of the HF equations, the HF orbitals, constitute a starting point from where improvements or further assumptions can be added, leading to *post-Hartree-Fock* methods and *semi-empirical* methods, respectively [127]. Post-Hartree-Fock methods have been developed to specifically treat electron correlation [127]; Configuration-interaction (CI), Coupled-cluster (CC), Møller-Plesset (MP) perturbation theory are the most important methods that incorporate correlation effects. The inclusion of further assumptions (*i.e.* neglect and/or parameterisation of integrals in the HF equations) leads to the computationally cheaper (but less accurate) semi-empirical methods (like AM1, PM3, etc). However, an overview of these approaches is

beyond the scope of this thesis.

2.6.1.2 Density Functional Theory (DFT)

DFT is based on the fact that for a fixed arrangement of nuclei and a fixed number of electrons the ground state electronic energy is entirely determined by the electron density [122; 127]. It was first shown by Hohenberg and Kohn that the exact electron density could describe the ground state as well as other properties of a system [132] making the total electronic energy a functional (a functional maps a function onto a number) of the electronic density distribution. It can be written as follows:

$$E[\sigma(\mathbf{r})] = \int V_{ext}(\mathbf{r})\sigma(\mathbf{r})d\mathbf{r} + F[\sigma(\mathbf{r})] \tag{2.51}$$

where the first term describes the classical interaction between the electrons with the external potential $V_{ext}(\mathbf{r})$ (the Coulomb interaction), $\sigma(\mathbf{r})$ is the electron density and $F[\sigma(\mathbf{r})]$ represents the sum of the kinetic energy of the electrons and inter-electronic interactions. In DFT there is a one-to-one correspondence between the electron density and the electronic energy, however the universal functional which connects these two properties is unknown. Therefore the objective of DFT is to find this functional that connects the electron density with the energy.

DFT has a big limitation, which is the fact that it does not reliably capture dispersion at all separations, and exchange is also an approximation (unlike in HF). In addition, with this approach it is difficult to consistently improve the results.

2.6.2 Molecular mechanics (MM)

The main aim of molecular mechanics (MM) is to use the information derived from other related molecules to predict the geometry of a new molecule. The basic idea

behind MM is that empirical functions are based on the picture of atoms in molecules as connected by balls and springs. In this sense MM is a method used to construct the PEL and enable the calculation of optimised geometries of molecules. The force-field potential is a sum of several contributions (intramolecular and intermolecular) and is given by:

$$U_{total} = U_{elec} + U_{LJ} + U_{str} + U_{bend} + U_{torsion} + U_{imp} \tag{2.52}$$

where U_{total} is the total energy, U_{elec} is the non-bonded electrostatic contribution, U_{LJ} is the non-bonded Lennard-Jones contribution, U_{str} is the bonding contribution, U_{bend} the angle-bending contribution, $U_{torsion}$ the dihedral contribution and U_{imp} the improper torsion contribution to the energy. A brief overview of each contribution as well as how it is modelled is presented below.

2.6.2.1 Intermolecular interactions

The existence of intermolecular interactions is the reason why liquids and solids exist, otherwise all matter would be in the gas phase. In MM the description of non-bonded interactions is usually divided in two groups: the van der Waals (vdw) interactions and the electrostatic interactions. The VDW interactions include dispersion and repulsion interactions. The electrostatic interactions include the interaction between charges and higher-order multipoles (*e.g.* dipole-dipole) in general, as well as related phenomena such as induction interactions. In addition, the intermolecular interactions can be divided into two main types: *long-range*, where the energy of interaction can usually be expressed as an inverse power of distance r between the atoms, and can be of three kinds: *electrostatic*, *induction* and *dispersion*; and *short-range*, where the interaction energy decreases exponentially with distance [133]. A brief description of each of these interactions will be given below.

2.6.2.1.1 Dispersion The force arising from the long-range dispersion interaction is also known as the *London Force* since London was the first to explain this dispersive force by quantum mechanics. This interaction arises from the fact that electrons in an atom or molecule are continuously moving which creates temporary dipole moments in the charge density. These instantaneous dipoles can induce the formation of another temporary dipole in a neighbouring atom or molecule. This gives rise to an attractive interaction between the two particles that is always present at all separations. The dispersion interaction is always attractive and therefore always lowers the interaction energy. The larger the molecule or atom the stronger the dispersion interaction since molecular polarisability increases with the number of electrons. The dispersion energy between molecules is approximately given by:

$$U_{disp} = -\frac{3\alpha^4\hbar\omega}{4(4\pi\epsilon_0)^2 r^6}, \tag{2.53}$$

where α is the polarisability, ω is the angular frequency and is related with the force constant k by $\omega = \sqrt{k/m}$, $\hbar = h/2\pi$, where h is the Planck constant, r is the distance between the atoms and ϵ_0 is the permittivity of free space. Equation 2.53 shows that the long-range dispersion energy is proportional to r^{-6}.

2.6.2.1.2 Repulsion Forces arising from repulsive interactions (also called *exchange forces* or *overlap forces*) arise when two molecular species approach together such that their electron clouds overlap, causing an increase in the energy. According to Pauli's exclusion principle, electrons in a system cannot have exactly the same set of quantum numbers. The origin of this interaction is due to electrons of the same spin in close proximity and therefore it is an exchange effect. This short-range interaction effectively prohibits electrons of the same spin occupying the same internuclear region. At short internuclear distances the interaction energy varies with $1/r$, however at large separations

the energy decays exponentially with $exp(-2r/a_0)$, where a_0 is the Bohr radius.

2.6.2.1.3 Induction When a polar molecule approaches an apolar molecule it induces a dipole moment in the apolar molecule. This phenomenon is also known as *polarisation* and will be outlined in Section 2.6.2.4. Induction is non-additive and is always attractive, varying with r^{-3}, r^{-4} or r^{-6}. This interaction is often ignored in many force-fields due to its small contribution to the potential energy. However, studies have shown the importance of induction in the description of the molecular conformational dependence when using multipoles to describe the electrostatic interactions [134].

2.6.2.2 Modelling Van der Waals (VDW) interactions

The VDW energy (E_{vdw}) represents the repulsion and attraction between non-bonded atoms and embodies approximations to both the dispersion and repulsion contributions to the intermolecular interaction. E_{vdw} is zero at an infinite interatomic distance and is large and positive at short distances due to the short-ranged repulsion term. At intermediate distances there is a minimum energy; the well-depth gives the equilibrium VDW interaction between the two species. The most common function that describes these behaviours is the *Lennard-Jones 12-6* potential which has the following form:

$$U_{LJ} = 4\epsilon\Big[\Big(\frac{\sigma}{r_{ij}}\Big)^{12} - (\frac{\sigma}{r_{ij}})^{6}\Big] \tag{2.54}$$

where ϵ is the well depth of the function, σ is the separation at which the energy is zero and r_{ij} is the separation between the particles i and j.

However, in the case of the AMOEBAPRO force-field a buffered $14-7$ functional form is adopted [135]:

$$U_{vdw}(ij) = \epsilon_{ij}\Big(\frac{1+\delta}{\rho_{ij}+\delta}\Big)^{(n-m)}\Big(\frac{1+\gamma}{\rho_{ij}^{m}+\gamma} - 2\Big), \tag{2.55}$$

where ϵ_{ij} is the potential well depth, $\rho_{ij} = R_{ij}/R^0_{ij}$ with R_{ij} as the $i-j$ separation (in angstroms) and R^0_{ij} the minimum energy distance. δ and γ are *buffering constants* and following Halgren [135], the fixed values $n = 14$, $m = 7$, $\sigma = 0.07$, and $\gamma = 0.12$ were used. These constants were determined from fits to rare-gas data and are applied to all interactions between the atoms i and j. For heterogeneous atoms pair, the combination rules are as follows:

$$\epsilon_{ij} = \frac{4\epsilon_{ij}\epsilon_{jj}}{(\epsilon_{ii}^{1/2} + \epsilon_{jj}^{1/2})^2}, \tag{2.56}$$

$$R^0_{ij} = \frac{(R^0_{ii})^3 + (R^0_{jj})^3}{(R^0_{ii})^2 + (R^0_{jj})^2}. \tag{2.57}$$

The terms involving the buffering constants, δ and γ, produce the repulsive part of the potential (formal exponent $n = 14$), while the δ-buffered term and the constant term, -2 describe the dispersion interactions (formal exponent $n - m = 7$). The buffered terms keep the potential finite as R_{ij} approaches zero, and thus avoids the strong divergence that is found in the unbuffered LJ potential [135]. The use of $n - m > 6$ allows a better description of the dispersion interaction in equation 2.53, while the positive value of δ serves to damp the dispersion term at small R_{ij}. In addition, this buffered potential gives a softer repulsion region than the LJ potential, and reaches the repulsion region at shorter ranges. It was found that this potential gave better descriptions of liquid properties and noble gases than the LJ potential, providing a better fit to gas phase *ab initio* results [135]. Figure 2.3 shows a schematic of the LJ, Buckingham exp-6 and buffered 14-7 potentials.

2.6.2.3 Electrostatic Interactions

The long-range electrostatic interactions are the first to arise when two molecules start approaching each other. They can be either attractive or repulsive and are strictly pair-

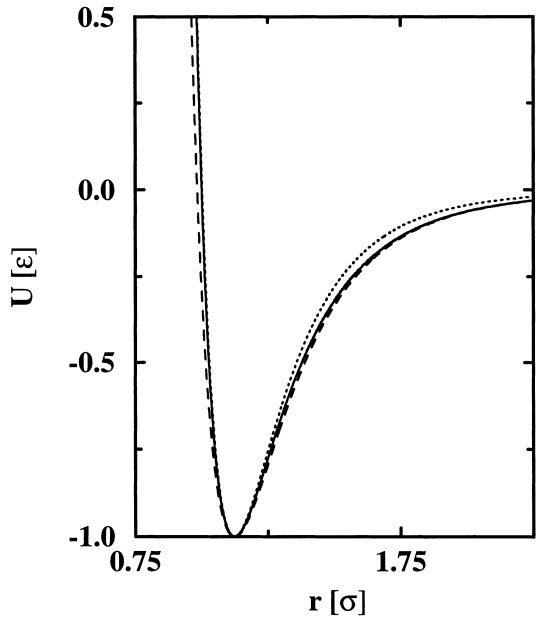

Figure 2.3: Comparison of the Lennard-Jones 12-6 (solid line), the Buckingham exp-6 (dashed line), and the buffered 14-7 (dotted line) potentials. From [136].

wise additive. This means that for instance, for three molecules A, B and C, the total energy is the sum over the pair contributions $U_{AB} + U_{BC} + U_{AC}$, where U_{AB} is the potential energy between the molecules A and B and is evaluated assuming molecule C is not present, and so on.

2.6.2.3.1 Example: The dipole-dipole interaction The dipole-dipole interaction arises from molecules with permanent dipole moments (polar molecules). The difference in electronegativity of the atoms in a molecule moves the electron distribution near the atom with higher electronegativity becoming more negatively charged and thus making the other atom more positively charged. This very strong interaction (around 2.0 kcal/mol) arises when one charge in a molecule interacts with the other charge of opposite sign of another molecule. The dipole-dipole interaction can be attractive or repulsive depending on whether the molecules are in a favourable or unfavourable orientation and is proportional to r^{-3} at long-range. The dipole-dipole interaction energy has the form [133]:

$$U_{\mu\mu} = -\frac{\mu^A \mu^B}{4\pi\epsilon_0 r^3}(2\cos\theta_A \cos\theta_B - \sin\theta_A \sin\theta_B \cos\varphi), \tag{2.58}$$

where μ^A and μ^B are dipoles of molecules A and B, respectively, and θ_A, θ_B and $\varphi = \varphi_A - \varphi_B$ are the polar angles of molecules A and B.

2.6.2.3.2 Modelling electrostatic interactions The uneven charge distribution in a molecule arises from the difference in electronegativity between atoms in such a way that the more electronegative atom attracts more electrons than the less electronegative atoms. There are several ways of describing this charge distribution, the simplest being the one that distributes partial charges over the atoms in the molecules. Coulomb's law is then used to calculate the classical interaction between pairs of point charges, as

follows:

$$V = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \tag{2.59}$$

where N_A and N_B are the number of atoms in the two molecules A and B, respectively. The major limitation of using partial atomic charges is the neglect of geometry dependence. Alternatively, there is an approach to describe electrostatic interactions called the *distributed multipole approximation* which is able to represent the anisotropic behaviour of a molecular charge distribution. This approach focuses on the inclusion of higher moments of the electronic multipole distribution: the charge, dipole, quadrupole, and higher multipoles through the molecule, and are generally represented by q (charge), μ (dipole), Θ (quadrupole) and Φ (octopole).

2.6.2.3.3 Point-charge model Point-charge electrostatic models are a simpler way of describing the electrostatic interactions. In order to accurately describe the electrostatic properties of a molecule, it might be necessary to assign partial charges at sites other than at the nuclei. For example, the nitrogen molecule has a dipole moment and a total net charge of zero and therefore by using the atom partial charge model a zero charge on each nucleus would be assigned. However, since the nitrogen molecule has a quadrupole moment the "simplest" way of describing this is by adding a $-q$ charge at each nucleus and a "+2q" at the centre of mass. Alternatively, the electrostatic interaction may be described by assigning bond dipole moments in the molecule and the electrostatic energy is thereby given by the sum of dipole-dipole interaction energies [122].

2.6.2.3.4 The Distributed multipole approximation In the distributed multipole expansion the molecule is divided in specific regions which are described by each multipole moment. This procedure enables a more accurate representation of the electrostatic

interaction than the point charge model and enables a representation of the anisotropic behaviour of the molecular charge distribution. In addition, the geometry dependence can be corrected by modelling polarisation as explained in the next Section (2.6.2.4). An exact method to expand the electrostatic potential in terms of multipole moments distributed at several regions in a molecule was developed by A. Stone and is called the *Distributed Multipole Analysis* (DMA) [137; 138]. In the DMA method the multipoles are calculated from *ab initio* calculations in which the wavefunction is defined by Gaussian basis functions. The permanent atomic multipoles are put at each atomic centre and include the monopole (charge), dipole and quadrupole moments, written as a vector M:

$$M = [q_i, \mu_{ix}, \mu_{iy}, \mu_{iz}, Q_{ixx}, Q_{ixy}, Q_{ixz} \cdots, Q_{izz}]^T, \tag{2.60}$$

where q_i is the point charge located at the centre of the atom i, μ_{ix} is the x dipole component of atom i, Q_{ixx} is the xx quadrupole component of atom i, Q_{ixy} is the xy quadrupole component of atom i, and so on. One of the problems is that dipole and quadrupole moments are dependent on the choice of origin [133]. The atomic multipoles of a molecule obtained by DMA calculations can indeed be viewed as the sum of the permanent electrostatic components with the contribution from the inter- and intra-molecular polarisation [134]:

$$\mathbf{M}_i = \mathbf{M}_i^{perm} + \mathbf{M}_i^{ind}, \tag{2.61}$$

where $\mathbf{M}_i$ is the resulting multipole on each atom and $\mathbf{M}_i^{perm}$ and $\mathbf{M}_i^{ind}$ are the permanent and induced moments, respectively. As will be explained in the next Section (2.6.2.4), this contribution ($\mathbf{M}_i^{ind}$) must be removed from the multipoles in order to avoid double counting upon the application of the polarisable force-field.

2.6.2.4 Polarisation

In the previous Sections (2.6.2.3.4) we focused on the permanent electrostatic multipoles of a molecule. However, changes in the charge distribution of a molecule or atom due to the application of an external field also gives rise to electrostatic interactions. This phenomenon is called *polarisation* and refers to the displacement of charge in a particle as a result of the application of an electrostatic field [139]. At the microscopic level, when an electric field E is applied to a molecule with a permanent dipole moment, the field causes a re-distribution of the charges, polarising the molecule and originating an *induced dipole moment*, μ_{ind}. The induced dipole moment is proportional to the electric field, $\mathbf{E}$:

$$\mu_{ind} = \mu_{perm} + \alpha\mathbf{E}, \tag{2.62}$$

where μ_{perm} is the permanent dipole moment and α is the *dipole polarisability* of the molecule. In the case of isolated atoms, the polarisability is *isotropic* which means that it only depends on the direction of the electric field (because an atom is represented as a sphere it looks the same in any orientation). However, in the case of molecules it gets more complicated since the orientation of the induced dipole does not always have the same direction as the electric field. This is why the polarisability of a molecule is *anisotropic*. The polarisability of a molecule can therefore be modelled by adding isotropic atomic polarisabilities. Several methods to include molecular polarisation have been proposed. In the case of the AMOEBAPRO force-field, Thole's interactive dipole polarisability model is used, which is based on a damping interaction method [140]. While some polarisability methods do not model induction between atoms within the same molecule, in Thole's model, atoms are polarised by induced dipoles of atoms belonging to the same molecule as well as atoms outside the molecule. In this way the

field E of equation 2.62 is the sum of the fields created by the permanent multipoles and induced dipoles:

$$\mu_{i,\alpha}^{ind} = \alpha_i(\sum_{\{j\}} T_{\alpha}^{ij} M_j + \sum_{\{j'\}} T_{\alpha\beta}^{ij'} \mu_{j',\beta}^{ind}). \tag{2.63}$$

Where $\mu_{i,\alpha}^{ind}$ is the induced dipole moment at each atomic site i, α_i is the atomic polarisability at atomic site i, $T_{\alpha}^{ij} = [T_{\alpha}, T_{\alpha 1}, T_{\alpha 2}, T_{\alpha 3}, \cdots]$ is the interaction matrix between site i and j (the subscript α stands for $\{x, y, z\}$), M_j is the permanent atomic multipoles given by equation 2.61, and $\mu_{j',\beta}^{ind}$ is the induced dipole at atomic site j' (the subscript β stands for $\{x, y, z\}$). The set $\{j\}$ comprises all atomic sites outside the molecule containing i, while the set $\{j'\}$ includes all atomic sites except i. The first term in equation 2.63 is what is called *direct induction* and is the induction caused by permanent multipoles, while the second term is the *mutual induction* and corresponds to the induction due to induced dipoles at other sites.

Amongst the proposed schemes, the one proposed by Thole provide several advantages. The main advantage is that it avoids the *polarisation catastrophe* at very short-ranges due to the existence of polarisation between atoms from the same molecule. This polarisation catastrophe is avoided by including a scheme that restrains one of the atomic multipole moments in each pair of interactions, to a finite value as the separation between the atoms approaches zero [140]. The charge distribution used in the AMOEBAPRO force-field to prevent this catastrophe is an exponential form and is usually referred to as a smearing function:

$$\sigma(u) = \frac{3a}{4\pi} e^{-au^3}, \tag{2.64}$$

where $u = r_{ij}/(\alpha_i \alpha j)^{1/6}$ is the effective distance as a function linear separation r_{ij} and atomic polarisabilities at sites $i(\alpha_i)$ and $j(\alpha_j)$ and a is a dimensionless width parameter related to the charge distribution and that controls the damping strength. In addition,

Thole's model is capable of exhibiting anisotropic behaviour to an external field only by using isotropic atomic polarisabilities and the atomic polarisabilities are transferable to many molecular polarisabilities.

The resulting atomic multipoles $\mathbf{M}_i$, obtained by *ab initio* calculations incorporate intra-molecular polarisation and must be subtracted from the $\mathbf{M}_i$ in order to obtain the true permanent atomic multipoles:

$$M_{i,\alpha}^{ind} = \alpha_i(\sum_{\{j\}} T_\alpha^{ij}(M_j - M_j^{ind}) + \sum_{\{j'\}} T_\alpha^{ij'} M_{\{j'\}}^{ind}). \tag{2.65}$$

The second term in equation 2.65 is the same as in equation 2.63 but the induced dipole was replaced by generalised induced moments. This equation can be solved iteratively for all atomic sites at the same time.

2.6.2.5 Intramolecular interactions

The intramolecular interaction energy includes the contributions from bond stretching, angle bending and torsional terms. A brief description of how these contributions are modelled is given below.

2.6.2.5.1 Bond stretching The total molecular potential energy stored in a bond where the atoms are separated by a distance l, can be descrived with an harmonic potential given by:

$$U_{str} = \frac{1}{2} K_B (l - l_0)^2, \tag{2.66}$$

where K_B is the force constant, l is the bond distance and l_0 is the equilibrium bond distance. This function cannot describe bond dissociation. A typical value of K_B is around $300 - 700$ kcal mol^{-1} Å^{-2}. In addition, the forces between atoms are very strong being typically in the range of $35 - 225$ kcal mol^{-1}.

In the case of the AMOEBAPRO force-field an anharmonicity term is included by the use of a deviation from the ideal bond length (l_0):

$$U_{str} = K_B(l - l_0)^2[1 - 2.55(l - l_0) + (7/12)2.55(l - l_0)^2]. \quad (2.67)$$

l_0 and K_B are parameters that are obtained from the parametrisation of the force-field. An overview about *parametrisation* is given in the last Section of this Chapter (2.8).

2.6.2.5.2 Angle Bending Hooke's law (the harmonic potential) is used to describe the angle bending potential (U_{bend}):

$$U_{bend} = \frac{1}{2}K_{bend}(\theta - \theta_0)^2 \quad (2.68)$$

where K_{bend} is the force constant, θ is the angle and θ_0 is the equilibrium angle. The distortion of an angle typically requires less energy that stretching or compressing a bond and therefore the force constants K_{bend}, are smaller than K_B. K_{bend} values are typically between $0.009-0.0120$ kcal mol^{-1}deg^{-1} and again, K_{bend} and θ_0 are parameters derived from the force-field parametrisation (Section 2.8).

In the AMOEBAPRO force-field higher order terms are included in order to improve the accuracy [141]:

$$\begin{aligned} U_{bend} &= K_{bend}(\theta - \theta_0)^2[1 - 0.014(\theta - \theta_0) + 5.6 \times 10^{-5}(\theta - \theta_0)^2 \quad (2.69) \\ &\quad -7.0 \times 10^{-7}(\theta - \theta_0)^3 + 2.2 \times 10^{-8}(\theta - \theta_0)^4]. \end{aligned}$$

2.6.2.5.3 Torsional motions In the case of bond-stretching and angle-bending a considerable amount of energy is required in order to cause distortions from the equilibrium values. Whereas the rotation about chemical bonds (torsion) is one of the main terms responsible for variations in the geometry as well as the energy. Another important cause for these variations is the non-bonded interactions, discussed earlier. Let us

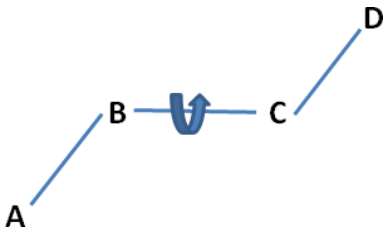

Figure 2.4: Schematic of a dihedral angle between four bonded atoms A, B, C, and D.

consider the dihedral angle ABCD between four bonded atoms A, B, C and D (Figure 2.4). It is usual to distinguish two types of dihedral angles, the *proper* dihedrals where a full rotation about the B—C bond occurs, and *improper (out-of-plane)* dihedral angle where the rotation around the same bond is limited [139; 122]. The rotation around the C—C bond in ethane is an example of a proper torsion. The common form of the dihedral angle potential, $U_{torsion}$ is:

$$U_{torsion} = \frac{U_n}{2}[1 + cos(n\phi - \phi_0)], \tag{2.70}$$

where U_N is the barrier height, n is the multiplicity (integer) and gives the periodicity of the potential in the interval [0,2π] and ϕ_0 is the equilibrium dihedral angle. In Figure 2.5 a schematic of the variation of the torsional energy, $U_{torsion}$, as a function of the torsion angle ϕ, is shown.

In the case of the AMOEBAPRO force-field a Fourier expansion is used to describe the torsional energies:

$$U_{torsion} = \sum_n K_{n\phi}[1 + \cos(n\phi \pm \delta)], \tag{2.71}$$

where $K_{n\phi}$ is the dihedral force constant, n is the multiplicity, ϕ is the dihedral angle and δ is the phase factor which determines where the torsion angle passes through its minimum value.

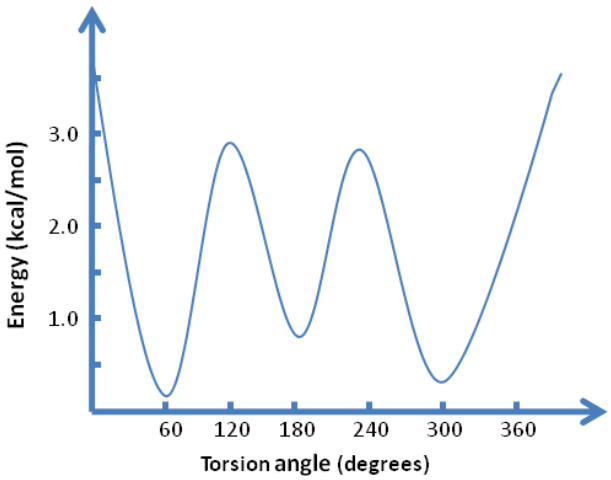

Figure 2.5: Schematic of the variation of the torsional energy with the torsion angle. (Reproduced from [122].)

2.6.2.6 Improper dihedral angle and out-of-plane bending motions

The geometry or chirality of atoms can be corrected by using improper dihedral angles. For example, modelling cyclobutanone with a force-field that only treats bond-stretching and bond-bending terms will incorrectly yield a geometry in which the oxygen is out of the plane. To obtain the correct geometry an additional term must be included in the force-field. The easiest way of doing this is by using an *out-of-plane* bending term. One of the techniques to include the out-of-plane bending term is to consider an "improper" torsion angle where the four atoms are not bonded in the sequence 1—2—3—4. The torsional potential to this method is:

$$U_{oo} = K_{oo}(1 - cos2\omega) \tag{2.72}$$

where K_{oo} is the out-of-plane bending constant and ω is the out-of-plane angle.

In the case of the AMOEBAPRO force-field a Wilson-Decius-Cross function is

used to restrain the out-of-plane bending [141]:

$$U_{oo} = K_{\chi}\chi^2, \tag{2.73}$$

where χ is the out-of-plane angle.

2.6.2.7 Cross Terms

One of the consequences related with the decreasing of the angle bending is the stretching of the bond between adjacent atoms in order to reduce the interaction between the atoms. This is an example of several important cross terms which are believed to be necessary in some force-fields that aim to predict the vibrational spectra. In the AMOEBAPRO force-field the coupling between stretching and bending is handled by the inclusion of a cross term which has the following form:

$$U_{b\theta} = K_{b\theta}[(b - b_0) + (b' - b'_0)](\theta - \theta_0), \tag{2.74}$$

where b and b' are the bond lengths and θ and θ_0 are the bond angles.

2.6.2.8 Amoeba vs standard force-fields

A reliable force-field is imperative when simulating biological molecules. The force-field used in this work is an extension of the AMOEBAPRO force-field developed by Ponder and co-workers [142; 143; 144] and has already been used in our preliminary studies of peptide-CNT recognition [118]. There are three main features that distinguishes AMOEBAPRO from the other force fields: the electrostatic interactions are described with the distributed multipole approximation as described in Section 2.6.2.3.4 and intra- and intermolecular polarisation is included to model induction effects and the conformational dependence of the multipoles on each atom. The AMOEBAPRO force-field

is designed to model peptides and proteins, and includes atomistic and continuum solvation models. The force-field was extended to include parameters for the non-bonded interactions between peptides and graphititic surfaces (carbon nanotubes [118], graphite [145] and fullerenes) at the same level of approximation (distributed multipoles and polarisation effects) [118]. Most of the *standard* force-fields describe the electrostatics with a simple point-charge distribution. If we had chosen to use such a force-field, the non-bonded interactions would be reduced to the VDW terms only, since an ideal carbon nanotube (CNT) carries little to no charge on each carbon atom. The use of only VDW terms is a very poor approximation to the real intermolecular interactions. A polarisable force-field allows the inclusion of multipolar electrostatic interactions for flexible molecules, which is very important for describing the π-stacking interactions between aromatic peptide side-groups and the nanotube. Therefore, we believe that the expense of a polarisable force-field is justifiable, since the aromatic residues are the dominant residues involved in the interaction with the graphitic surfaces. Therefore the distributed multipoles and the inclusion of polarisation help in capturing conformational changes due to ring orientation of the peptide relative to the CNT surface. In addition, in our previous paper [118] we showed that the use of only VDW interactions does not compare well with first-principle calculations. To give the reader an idea of how expensive the AMOEBAPRO force-field is with respect to other "off-the-shelf" force-fields, a MD simulation comprising a box of water with 522 water molecules was run for 0.5 ns with the AMOEBAPRO force-field and with TIP3P water. In both cases, the minimum image convention was applied (no spherical cutoff applied) and in the case of the AMOEBAPRO force-field the induction threshold was set to 10^{-3} D (the default is 10^{-6}). On the same computer, with the TIP3P water, it took 30 hours, while with the AMOEBAPRO it took 83 hours. This gives an estimate that the AMOEBAPRO

force-field is around 2.7 times slower than a standard force-field. This extra expense is due to the presence of multipoles and the polarisation. To understand where does this extra computational expense come from let us consider a water dimer and let us count the number of terms in the electrostatic interaction by using the point charge model and by using multipoles. With point charges there is one interaction (charge-charge) for each pair of atoms and since there are nine possible atom-pairs for a water dimer, the electrostatic interaction has nine contributions, and can be calculated by summing over the interaction of each pair of point charges using equation 2.59. In the case of multipoles there are six interactions (charge-charge, charge-dipole, charge-quadrupole, dipole-dipole, dipole-quadrupole and quadrupole-quadrupole) for each pair of atoms, and therefore there are 36 terms contributing to the electrostatic interaction. However, the description of the electrostatics with distributed multipoles is not just a mere sum of all these terms. There is an additional expense which is related to the fact that the dipole-dipole, dipole-quadrupole and quadrupole-quadrupole interactions are dependent on the choice of origin and therefore during an optimisation or dynamic simulation it is necessary to keep track of the orientation and the position of each atom. Each molecule is defined by a "global" coordinate frame which is defined by a set of Cartesian axes fixed in space and a "local" coordinate frame which is bound to the molecule, moving as the molecule moves. The atomic multipoles remain constant with respect to the local frame as the molecule moves. The relation between the global and local frames is described by a rotation and a translation: a rotation into the desired orientation and a translation to the desired positions [133]. In order to calculate the multipole interaction each local frame must be moved to the global frame by applying a rotation operation for each multipole site in the molecule. The application of rotation operations (via matrices) is a time-consuming operation leading to the extra computational time. The extra

computational expense also arises from the calculation of the induction energy and is related with the fact that the induction energy is calculated via an iterative procedure until convergence is achieved. It took 48 hours to run the same box of water for 0.5 ns with the AMOEBAPRO force-field but with the polarisation switched off.

2.7 Implicit solvation model

Implicit solvent models are used to represent the solvent without including explicitly the solvent molecules (at the atomistic level). The advantage of this is a decrease in the computational time, while still accounting for solvent effects, such as changes in the properties or in the behaviour of the molecules. The disadvantage is that the solvent ordering contribution to the binding free energy is neglected. In the study of the graphitic surfaces with peptides the Atomic Solvation Parameters (ASP) model [146] was used, which works by including an extra term in the force-field that is dependent on the derivative of the free energy of solvation ΔG_s:

$$\Delta G_s = \sum_{atoms,i} \Delta\sigma_i A_i, \tag{2.75}$$

where σ_i is the atomic solvation parameter for atom i, and A_i is the solvent accessible area for each atom or group. The solvent interaction potential describes protein-water interactions based only on the atomic solvation parameters ($\Delta\sigma_i$) and their solvent accessible areas A_i. The atomic solvation parameters are derived from vacuum-to-water free energies of transfer of amino acid side-chain analogs, which express the hydrophobicity of each atom i and is based on the accessible surface area of atoms in the protein. The solvent accessible area A_i, of a protein atom is defined as the area covered by a spherical solvent (probe), as the probe rolls over the surface of the protein.

The force $\mathbf{F}_i$ on atom i can be expressed as:

$$\mathbf{F}_i = -\frac{\delta(\Delta G_s)}{\delta \mathbf{r}_i} \tag{2.76}$$

$$= \sum_{atoms j \neq i} \left(\Delta\sigma_i \frac{\delta A_i}{\delta \mathbf{r}_j} - \Delta\sigma_j \frac{\delta A_j}{\delta \mathbf{r}_i} \right), \tag{2.77}$$

where $\mathbf{F}_i$ is the force on atom i with position $\mathbf{r}_i$. Equation 2.76 describes solvation forces in terms of derivatives of the accessible areas with respect to the atomic positions.

2.8 Parametrisation and Transferability

In addition to the form of the intermolecular potential, a force-field also comprises a set of parameters (such as constants, K_B, K_{bend}, σ_{ij}, equilibrium bond lengths etc.) for each type of atom or atom pair. For example, a force-field includes a distinct set of parameters for an oxygen atom in different environments *e.g.* in a carbonyl group and in a hydroxyl group. It is important to notice that the same functional form can be used in two force-fields, however with completely different parameters. A force-field typically contains a large number of parameters and therefore its parametrisation and validation can be a very difficult and time-consuming task. A force-field should be parametrised according to the properties we are aiming to determine and study. For example, if we aim to study the structural properties of molecules, the geometries and energies of key molecules should be included in the parametrisation [122]. Quantum mechanics calculations can be used to provide the data for the parametrisations of force-fields, since most of the time experimental data is hard to obtain or simply does not exist. Once one has identified the data set for the parametrisation, several approaches can be applied to obtain the parameters. The parameters can be tested by "trial and error", in which an iterative procedure is applied to fine-tune the parameters until consistency is reached. Alternatively, the set of parameters can be obtained by applying least-squares fitting to

determine the optimal fit to the data provided. Parameter sets and functional forms are specified to perform together and therefore the parameters from one force-field should not be used with a potential from another force-field without re-parametrisation.

One of the most important features of a force-field is the *transferability*. The same set of parameters should be appropriate to model a number of related molecules, without the need to define new parameters for each molecule. This ability to transfer parameters from one molecule to another is crucial, otherwise parametrisation would be impossible and the force-field would not have the ability to predict the properties of other molecules. For example, the same set of parameters for a given force-field should be suitable to model all n-alkanes [122]. This feature is very important if we want to use a force-field to make predictions. To allow transferability, the same set of parameters should be used for the maximum number of molecules as possible. For example, usually the same set of VDW parameters can be used for all atoms of the same element (sp^3, sp^2, sp, etc.).

Chapter 3

Modelling the binding affinity of peptides for graphitic surfaces by mutation of the aromatic content

3.1 Abstract

In this Chapter we study the interactions between peptide sequences and graphitic surfaces — carbon nanotubes (CNTs) and graphene — using molecular dynamics (MD) simulations with our polarisable force-field. It was observed that peptide sequences with strong affinity to CNTs selected by phage-display [94] are rich in tryptophan (W) and histidine (H) residues. The aim of this study is, therefore, to investigate the importance of the tryptophan residue in the peptide sequences B1(HWKHPWGAWDTL) and B3(HWSAWWIRSNQS) [94] by mutating each tryptophan by either tyrosine or phenylalanine. Another aspect of this project was to explore the effect of curvature on the binding

affinity. Our findings demonstrate that sequences containing tryptophan residues have more affinity for the graphitic surfaces than those containing tyrosine (Y) or phenylalanine (F). Our results also suggest that these peptide sequences were also selected for interfacial shape. For the graphene surface our results suggest that there is a compromise between having all the aromatic residues close to the surface and allowing the non-aromatic residues to also approach the surface.

3.2 Introduction

Due to their fascinating structure, properties and potential applications, CNTs have attracted much attention in recent years [100]. Unfortunately, the extreme hydrophobicity of CNTs presents a major obstacle to their usage in aqueous solutions; a problem for biological or medicinal applications. In addition, due to the difficulty of purifying and separating CNTs, it has been complicated to assemble CNTs into practical structures in aqueous solution. As a result, dispersing CNTs in aqueous solution and increasing their solubility in water, in order to control dispersions, has been the subject of extensive research. Solubility of CNTs in water can be improved by covalently attaching functional groups to the nanotube walls [147; 64]. However, these modifications of the CNT sidewall can locally perturb the electronic (sp^2) structure of the tube, and therefore can compromise the electronic and optical properties that make CNTs unique. Therefore, non-covalent functionalisation of CNTs is preferable for many applications. Non-covalent functionalisation might not only increase the solubility of CNTs but also preserve its electronic and mechanical properties. Several approaches for solubilising CNTs using non-covalent adsorption of surfactants [148; 149], polymers [150; 65], and biomolecules [151; 152; 153; 154; 155; 71; 156] have been proposed and tested. In the cases where an interface between CNTs and biological systems is required (for example, biosensors),

biomolecules such as DNA, RNA, and peptides are appropriate [151; 157]. Peptides have particularly shown advantages over non-covalent functionalisation by detergents because they are not necessarily toxic to mammalian cells. Furthermore, peptides are well-known for their high specificity for other biomolecules and materials [22; 21] and for their ability to self-assemble into a wide range of complex, functional structures such as molecular linkers and erectors.

A number of peptide sequences with specific affinity for carbon nanotubes [93; 94; 95; 97], fullerenes [98] and nanohorns[99] have been reported. Common to all of these sequences is the presence of aromatic residues. The importance of the aromatic content on the peptide-CNT binding affinity has been highlighted in several works [94; 95; 88; 97]. In addition, a number of studies have also focused on the importance of tryptophan in the peptide-CNT binding affinity [90; 104; 119; 102; 101; 96; 118; 120]. Experimental measurements of pyrene binding to CNTs [103] indicated that this interaction is quite strong due to the highly aromatic nature of the π- stacking and therefore, it is not surprising that aromatic residues also have an affinity to graphitic surfaces. AFM measurements and optical absorption spectra suggested that the ability to disperse single-walled carbon nanotubes (SWCNTs) may be increased by increasing the electron density of the aromatic residue [104]. Furthermore, Li *et al.* [90] demonstrated that proteins containing poly-tryptophan showed a stronger adhesion on CNTs than those containing poly-lysine, by direct AFM measurements. Salzmann *et al.* [102] demonstrated the importance of the balance between hydrophilic (charged) and aromatic residues in the dispersion of CNTs. Furthermore, measurements of the dispersion effect of copolymers (a polymer formed by two or more different monomers) comprising mixed sequences of lysine/tryptophan and lysine/tyrosine demonstrated a better solubilization for the tryptophan-containing peptides. Additionally, Xie *et al.* [101] studied

the interactions of three aromatic residues (tryptophan, phenylalanine, tyrosine) SWCNTs, by designing a range of surfactant peptides. Their findings suggested that the sequence containing tryptophan supported the highest affinity for CNTs (both bundle and SWCNTs) while the tyrosine was found to be more selective for individual SWCNTs and phenylalanine the aromatic residue with the lowest affinity for CNTs. Honek *et al* [119] investigated the nature of the π—π interactions between the tryptophan residue and CNTs by single-point mutation and the incorporation of non-natural tryptophan analogs to alter the properties of the peptide chain. These binding affinities to CNTs were experimentally evaluated and further investigated by computational modelling. Their results suggested that the π—π interaction between the tryptophan side chain and the CNT depends on the orbital interactions between the highest occupied molecular orbital (HOMO) of the tryptophan indole ring and the lowest unoccupied molecular orbital (LUMO) of the CNT. This implicates the sidewall of the CNT behaving as an electron acceptor and the indole ring as an electron donor. However, previous studies on the interactions of CNTs with anthracene derivatives reported by Zhang *et al.* [158] suggested that the nanotube acted as the electron donor and the anthracene an electron acceptor. Therefore, these studies indicate that CNTs can either act as electron donors or acceptors in these specific interactions.

A number of modelling studies have attempted to understand the nature of the peptide-CNT interaction [118; 120; 114; 159; 160; 161; 162; 163; 164]. Several studies have used a variety of computational methods, such as MD simulations [114; 159; 160; 118; 120] and DFT calculations [165; 163] in an attempt to gain insight into the atomic scale mechanisms underlying the strong affinity of peptide sequences identified by Wang *et al.* [94] to CNTs. Cheng *et al.* [114] reported binding free energies for CNTs and those peptides sequences [94] based on a combined simulation method of

MD with a continuum solvation model. The continuum water medium solvent was used to calculate energetic contributions based on the thermodynamic theory. The authors took into account both contributions of the solute and solvent, however, the entropic change of the solute has been neglected. These authors did not calculate real binding free energies, since they assumed that all the phase space had been covered and merely substituted the calculated binding interaction energies by binding free energies. As explained in Section 2.3.1 of Chapter 2 the phase space of biomolecules, such as peptides, is enormous and therefore it is challenging its complete exploration. Only if we were able to sample the entire phase space we would we be able to make the assumption made by these authors. More recently, Kyani *et al.* [159] proposed a model to calculate binding free energies for the same peptide sequences using a linear interaction energy approach in conjugation with MD simulations. This method is based on a linear fit of computed interaction energies between the peptide and the CNT. Again, these authors have not calculated real binding free energies but interaction energies for the peptide-CNT system in the free and bound state. Both authors [114; 159] claimed good qualitatively agreement with experimental observations and concluded that the strong affinity of the aromatic residues for the CNT are driven by the van der Waals (VDW) interaction. However, these authors used single-point charges to describe the electrostatic interactions, and because carbon atoms on the CNT were modelled as uncharged particles, the calculation of the non-bonded interactions was reduced to the evaluation of the VDW terms only. Fan *et al.* [163] used a a DFT tight-binding method with an empirical VDW force correction to study the same peptides adsorbed on CNTs with different chiralites and different diameters. The authors performed single-point calculations with periodic boundary conditions in two dimensions under the local density approximation (LDA). Their results showed a competition between π-π and hydrogen-π stacking in the bind-

ing to CNTs and demonstrated that peptides can be spontaneously attracted to the sidewall of the CNTs. They did not observe any linear relationship between the binding energy and the tube diameter, but despite the chirality dependence of the binding energies not being obvious, a general binding energy increase was noticed upon increasing the tube diameter. Moreover, their calculations predicted that the CNT non-covalent modification by peptides might increase the electron transfer capabilities of SWCNTs. However, in this study the authors only considered one peptide configuration. Recently, Gianese *et al.* [160] reported a study, which performed a comparative investigation of the adsorption of a peptide sequence selected by Wang *et al.* [94] on two different carbon surfaces - a graphene sheet and a CNT - by calculating free energies of folding and binding based on a combination of solvation energy, formation of hydrogen bonds and the amount of hydrophobic contact surface between the peptide and the surface. In both cases tryptophan was found to be the most strongly bound residue and different arrangements of the histidine and tryptophan residues enabled a multitude of different bound configurations. Although in agreement with the experimental observations [94], that tryptophans play a dominant role in the binding affinity to CNTs, the electrostatic interactions were modelled by single-point charges and water was described by a simple point charge (SPC) model. Chiu *et al.* [166] demonstrated with atomistic MD simulations that for a designed α-helical peptide structure, the π-stacking interaction between the aromatic groups of the peptide and the CNT surface is dominant. Recently, they extended their work to explore the peptide-peptide interactions and peptide-CNT interactions, by performing MD simulations of a six and five peptide bundle around a SWCNT [161]. Their results strongly suggest that peptides wrapped around CNTs are good dispersal agents. Roman *et al.* [165] performed a study with DFT, using the generalised gradient approximation (GGA) for the exchange and correlation energy, in partnership

with periodic boundary conditions, on the adsorption of single amino acids glycine, cysteine, histidine and phenylalanine on CNTs and graphite. While histidine and cysteine did not show considerable substrate dependence, glycine was shown to bind stronger on the CNT than on the graphite whilst phenylalanine showed the opposite behaviour, which seemed to be due to the weakening of π-stacking on CNT curvature. However, the authors did not describe which functional they used. Recently, Wang and Ai [164] used MPWB1K and MP2 methods to study the adsorption of tripeptides (GXG where X is any residue and G is glycine) onto CNTs. Their results indicated that the hydrophobic residues yield a strong adsorption on CNTs and that the $\pi \cdots \pi$ interaction is likely to be the most important interaction. In addition, amongst the three aromatic tripeptides, GWG exhibited the strongest adsorption in both gas and aqueous phase, where the solvent effects were taken into account with an implicit model. Recent works [167; 168] have used atomistic MD simulations to study reversible cyclic peptides [169] and their binding to CNTs. In our previous work [118], we performed MD simulations of SWCNTs interacting with 'strong-binder' and 'weak-binder' aptamers, as identified by Wang *et al*, with our validated extension of the AMOEBAPRO force-field [134; 170], which treats both the molecule and CNT as polarisable. Our results confirmed the experimental observations [94] of the relative binding affinity for the systems studied. In this study we also calculated adsorption energies of several residue analogs, including aromatic residue analogs, with the extended force-field and compared them with data from electronic structure theory. A similar study was recently published by Rajesh *et al.* [162], which used quantum mechanical methods (DFT using the plane-wave pseudopotential approach within GGA and MP2) to calculate the interaction energy of aromatic residue analogs (histidine, phenylalanine, tyrosine and tryptophan) with a planar graphene sheet and CNTs and found that tryptophan is the aromatic residue

Table 3.1: Original and mutated peptide sequences and a weak-binder control peptide sequence NB1 used in this study.

Original sequences	
B1	HWKHPWGAWDTL
B3	HWSAWWIRSNQS
Mutated sequences	
B1Y	HYKHPYGAYDTL
B1F	HFKHPFGAFDTL
B3Y	HYSAYYIRSNQS
B3F	HFSAFFIRSNQS
Control sequence	
NB1	LPPSNASVADYS

with most affinity for CNTs. With the exception of our previous work, all simulation studies used traditional force-fields that are based on a distributed charge model for the electrostatics and do not include a description of polarisation effects.

Taking into account the importance of the aromatic content, the aim of this project is to explore whether mutations of the experimentally determined 'strong binder' sequences, HWKHPWGAWDTL (denoted B1) and HWSAWWIRSNQS (denoted B3), can recover similar binding affinity to the CNTs. To this end MD simulations were performed of B1 and B3, and the corresponding mutants B1Y, B1F, B3Y and B3F (where in each case W was replaced by either Y or F), adsorbed on a carbon nanotube with the aim of determining if there was any change to the binding affinity of the mutants. For example, the B1Y mutant corresponds with the sequence HYKHPYGAYDTL. Table 3.1 shows the peptide sequences studied. In addition, experimental studies have indicated that peptide selection can distinguish between sequences that bind to CNTs and sequences that bind to graphite [94; 96]. To this end, we also ran similar simulations of these peptides adsorbed on a graphene sheet. Because the chemical composition

of graphene and CNTs is the same, we considered these two systems to be the ideal test case for investigating the influence of the interfacial shape on the binding affinity of peptides. Distances and angles between the aromatic residues and the graphitic surfaces were calculated and compared with the original sequences. Our molecular simulations provide a detailed examination of peptide-CNT/peptide-graphene interactions for a system consisting of an (8,0) zigzag segment/graphene sheet and one of each of the six aptamers (each with 12 residues).

3.3 Methods

We performed MD simulations of systems comprising one surface (CNT or graphene sheet) and one each of the peptide sequences B1, B3, B1Y, B1F, B3Y, B3F and the control weak-binder NB1(LPPSNASVADYS) [94], using our extended AMOEBAPRO force-field [134; 170]. Details of the force-field extension for the description of CNTs [118] and graphene [145], has previously been given and good agreement was obtained when compared with results from electronic structure theory. Our previous results using the CNT extension to the force-field also yielded excellent agreement with experimental observations [94]. As explained in Section 2.6.2.8 the AMOEBAPRO force-field has three main features that makes it distinct from a standard force field: the electrostatic interactions are described with the multipole approximation [137; 138] instead of point-charges and intra- and intermolecular polarisation is included to model induction effects and conformational dependence of the multipoles on each atom. Since the aromatic residues are the main ones involved in the interaction between graphitic surfaces and peptides, we believe that the expense of a polarisable force-field is justifiable because we are modelling graphitic surfaces which carry little to no charge, and therefore the use of a standard force-field would implie the use of only VDW. Using distributed multipoles in

this work enabled an appropriate description of the π—stacking interactions between the aromatic groups and the surface. Furthermore, previous experimental work suggested that there is a weak charge transfer interaction between aromatic groups and CNTs [104]. This phenomenon can be approximated with this polarisable force-field.

In the case of the peptide-CNT systems, a hydrogen-terminated zigzag (8,0) nanotube with 976 carbon atoms was used. This CNT was at least four times the contour length of the peptide. In the case of the peptide-graphene interface a sheet comprising 2772 carbon atoms was modelled. All simulations were carried out in the canonical ensemble at room temperature. The Verlet [129] algorithm was used to solve Newton's equations of motion with an integration timestep of 1.0 fs and a cutoff of 8 Å applied to all non-bonded interactions. The systems were equilibrated for 1 ns, followed by an additional 1 ns production run. In order to identify many different binding configurations, six initial geometries were used for each aptamer. To achieve this, several tactics were used, for instance, peptide configurations were constructed and placed by hand near the surface. In this case, geometries varied from extended backbone configurations aligned with the long axis of the nanotube, to helical geometries that could wrap around the CNT by changing the backbone torsion angles. Simulations of the peptide without the surface, in effective continuum solvent conditions, were also run to create a variety of geometries which were then placed close to the CNT surface; in addition, the peptide-nanotube geometries that yielded strong binding were used to perform a number of "on the fly" mutations in both the forward (W to Y/F) and reverse (Y/F to W) directions.

All simulations were performed using the TINKER [171] package. To account for solvation effects, we used a continuum effective solvent (as described in Section 2.7 of Chapter 2), since the inclusion of water at the atomistic level would increase the

simulation time by several orders of magnitude. The implicit solvent employed was the ASP model [172; 146], used in conjunction with a modified background dielectric. This approach was found to give physically reasonable behaviour when used previously [118]. Furthermore, a comparative study of the structural properties of a strong-binding B1-nanotube configuration in both implicit and explicit solvent, has been done by Walsh [120]. These data clearly showed that the peptide configuration has not changed significantly by the presence of explicit water, and that the trends of the properties, such as the distances of the aromatic residues to the CNT surface and the relative orientation of the aromatic rings with respect to the CNT normal had not changed substantially. Due to these satisfactory results, in this case, this is a reasonable approximation. This approximation seems to work due to the lack of solvent structuring at the nanotube water-interface [173], as opposed to the strong water structuring at hydrophilic surfaces such as the rutile TiO_2 surface [174]. Furthermore, the conjecture of the water structuring around the peptide and the surface not being so pronounced as it would be for a hydrophilic surface is supported by evidence of the dynamics of water at a hydrophobic surface being faster than that in bulk [175]. In addition, due to the high hydrophobic content of these peptides sequences, it is expected that the structural changes if the explicit solvation was accounted for would be minimum. Therefore, while in full understanding that the use of implicit solvation is not an ideal approach, we conclude in this case it is acceptable.

To analyse our simulation data we used both structural and energetic metrics. To quantify binding affinity, we calculated the normalised interaction energy, E_N, (the interaction energy between the peptide and surface, divided by the number of atoms in the peptide); since each peptide contains a different number of atoms, the normalisation enable us to make a fairer comparison between the systems. The interaction energy,

E_{int}, was calculated within TINKER and is defined by the difference between the energy of the peptide adsorbed at the CNT and the energy of the molecular fragments at infinite separation, *e.g.*

$$E_{int} = E_{peptide+CNT} - (E_{peptide} + E_{CNT}). \tag{3.1}$$

Two other ways of normalising the binding affinities were tested with the intention of showing that the trend in the binding affinity did not change upon the use of several normalisation schemes. For these alternatives, instead of assigning a weight of 1 to each atom in the peptide as we did in the denominator in our calculation of E_N, two other weighting schemes were tried: in one of the methods a weight of 1 was assigned to any peptide atom within 6 Å of the surface and a weight of zero for all peptide atoms further from the surface; in the other scheme a weight of 1 was assigned to any atom closer than 4 Å from the surface, and an exponentially decaying weight for all atoms more distant (weight $=$ exp[-a(x-4)]), where the decay constant a was set to 0.5. E_N was then calculated by dividing the interaction energy by the sum of the weight. These data are presented in Section 3.4.2. Distribution profiles of the ring centre-of-mass to the CNT suface distances for the aromatic residues were performed by calculating the distance from the centre-of-mass of each aromatic ring to the surface. For the ring—surface distances the data is presented as histograms averaged over the production run trajectories. In addition, the ring-tilt angles with respect to the surface normal were determined by calculating the angle between the aromatic ring normal vector and the CNT normal vector. A schematic of a ring-tilt angle is illustrated in Figure 3.1. Cumulative moving averages (CA$_i$) of the ring-surface distance and of the interaction energy for each peptide adsorbed on the CNT were calculated by averaging the sequence of i values x_1, $\cdots$, x_i up to the current time:

$$CA_i = \frac{x_1 + \cdots + x_i}{i}. \tag{3.2}$$

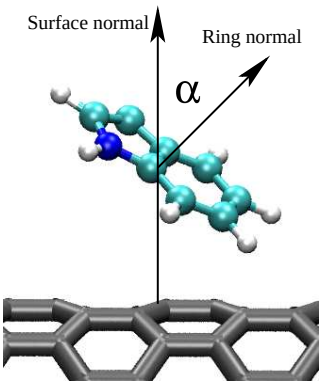

Figure 3.1: Schematic of the ring-tilt angle α. The ring-tilt angle is the angle between the aromatic ring and the nanotube and was calculated by measuring the angle between the aromatic ring normal and the nanotube normal.

3.4 Results

3.4.1 Equilibration checks

In order to determine if equilibration had been reached, structural and energetic criteria were used during the equilibration runs, such as measuring the fluctuations in the average distance between the CNT surface and the aromatic ring centre-of-mass, and the fluctuations in the average interaction energy between the peptide and the surface. A typical plot of such properties during the equilibration period is shown in Figure 3.2 for the B1Y-nanotube system. These data indicate that the properties are changing very little with time and therefore after 1 ns equilibration has been reached. Furthermore, the same properties in the post-equilibration (production) period were calculated. In Table 3.2 are the fluctuations (calculated as the root mean square deviation) of the cumulative moving averages of the ring-surface distance for each peptide adsorbed on the CNT in the 1 ns post-equilibration period. The small fluctuations of the moving averages indicate that this property is stable and as a result the ring-nanotube distances

and the interaction energy are not changing dramatically with time. Therefore, these data show that equilibration have been achieved.

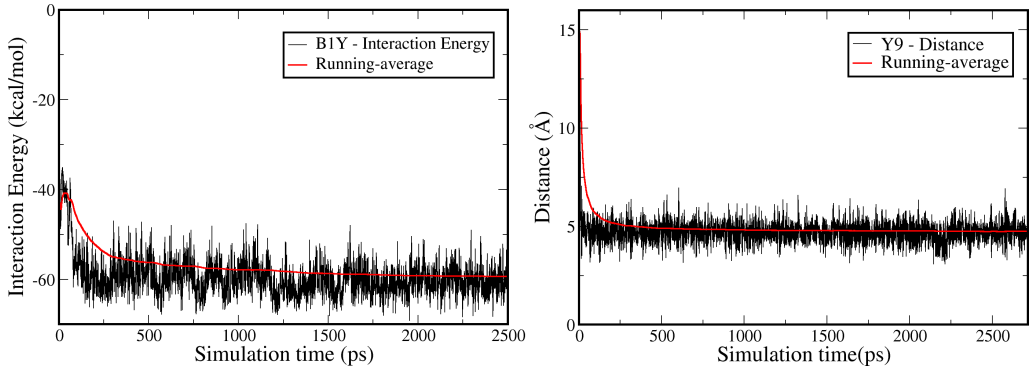

Figure 3.2: Two properties of the B1Y-nanotube system, as a function of time during the equilibration period of the simulation; interaction energy (left) and ring-surface distance for (Y9) right. The red curve shows the cumulative moving average of the properties with time.

3.4.2 Peptide—Nanotube Simulations

For each aptamer, the normalised interaction energy, E_N, was averaged over the top four distinct binding trajectories that resulted in the best binding affinities (*i.e.* the lowest interaction energy). Profiles of E_N from the "best" trajectory (*i.e.*, the trajectory of strongest affinity) of B1, B3, B3Y and NB1 interacting with the CNT are shown in Figure 3.3. The best trajectories of B1Y, B1F and B3F are similar to B3Y, so are omitted from Figure 3.3 for clarity. Normalised interaction energies (averaged over the four best trajectories) for all cases are presented in Table 3.3, along with the standard error of the mean (SEM), as defined by the standard deviation in the sample means. The breakdown in E_N for each of the four trajectories for each aptamer is presented in Table 3.5. In addition, as mentioned in the Methods, E_N was calculated using two

Table 3.2: Root mean square deviation (RMSD) of the CA_i of the ring-surface distances (Å) and interaction energy (kcal/mol) (for each peptide adsorbed on the CNT) in the 1 ns post-equilibration (production) period, showing stability of the CA_i of these properties. See Figure 3.2 for a typical plot of these CA_i during the equilibration period. [120]

B3W	**RMSD**
W2	0.005
W5	0.003
W6	0.009
Inter. Energy	0.0028
B1W	**RMSD**
W2	0.011
W6	0.004
W9	0.002
Inter. Energy	0.146
B1Y	**RMSD**
Y2	0.009
Y6	0.024
Y9	0.012
Inter. Energy	0.134
B1F	**RMSD**
F2	0.006
F6	0.008
F9	0.047
Inter. Energy	0.103
B3Y	**RMSD**
Y2	0.082
Y5	0.014
Y6	0.020
Inter. Energy	0.111
B3F	**RMSD**
F2	0.054
F5	0.019
F6	0.010
Inter. Energy	0.915

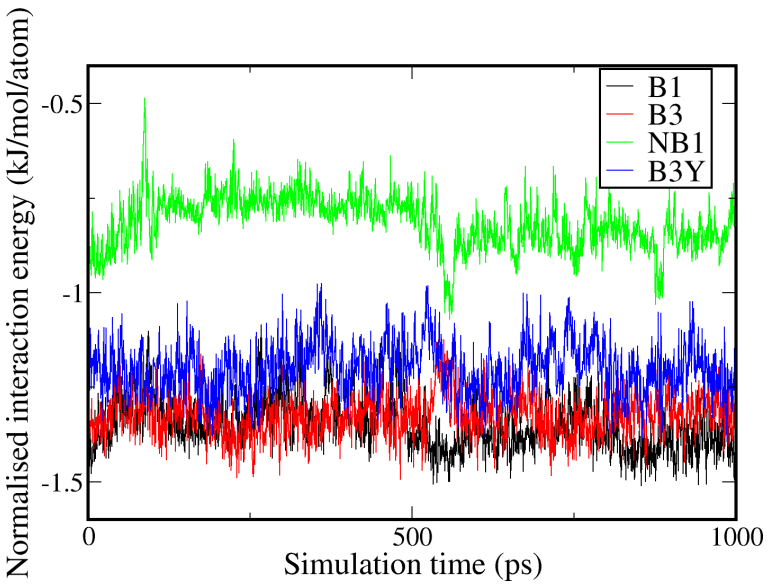

Figure 3.3: Normalised interaction energies as a function of time for original sequences B1 and B3, a mutated sequence B3Y, and the weak-binder control NB1 interaction with the nanotube, over the last 1ns of production run.

other different methods. Table 3.4 demonstrates that the trend in the binding affinity does not change upon using different normalisation schemes.

We first draw attention to the differences between the strong-binder peptide sequences B1 and B3, the mutant sequences B1Y, B1F, B3Y, B3F and the weak-binder control NB1 interacting with the CNT. Table 3.3 demonstrates that the strong-binder peptides B1 and B3 exhibit a slightly greater binding affinity than the mutant peptides. Despite the gap between the normalised energies being small, if we take into account that each peptide sequence has approximately 200 atoms, it is still a considerable difference in the binding affinity. Furthermore, the differences in E_N between the original and mutant sequences are greater than the SEM associated with each case. The differences among the binding affinities of the tyrosine and phenylalanine mutants is small as well. In terms of the "best" trajectory for each aptamer (Table 3.5), the tyrosine mutants

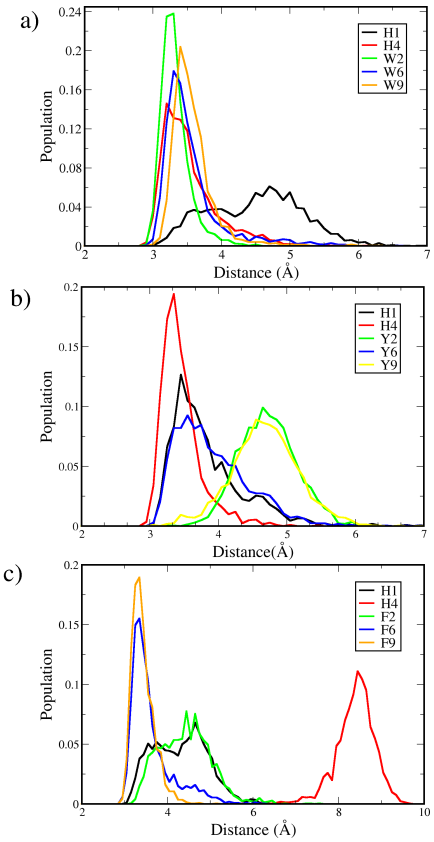

Figure 3.4: Distribution of the distance from the nanotube surface to the ring centre-of-mass for all the aromatic residues H, W, Y and F for the peptide sequences (a) B1, (b) B1Y, and (c) B1F.

Table 3.3: Average normalised interaction energies, E_N (kJ mol^{-1}atom^{-1}), and the standard error of the mean (SEM) of E_N, taken from the four distinct trajectories of lowest energy for each aptamer, for the peptide—nanotube interface.

Aptamer	E_N	SEM
B1	-1.32	0.04
B3	-1.29	0.03
B1Y	-1.15	0.06
B3F	-1.14	0.03
B3Y	-1.12	0.07
B1F	-1.07	0.08
NB1	-0.87	0.06

show slightly greater binding than the phenylalanine mutants. However, these effects are very subtle and the ordering between the tyrosine and phenylalanine mutants might change, *e.g.* if these simulations were carried out in explicit water.

It is proposed that the origin of these energetic differences lie in the differences in binding geometry. In Figure 3.4 the distribution profile of the ring centre-of-mass to CNT surface distances for H, W, Y and F for B1, B1Y and B1F is shown. These data clearly show that all tryptophans [Figure 3.4(a)] spend more time closer to the nanotube surface compared with the corresponding tyrosine and phenylalanine groups [Figure 3.4 (b) and (c)]. This trend was found for all the configurations studied and also for the other peptide sequences B3, B3Y and B3F (Figure 3.5). The best (lowest energy) trajectories of B1 and B3 yielded binding geometries where all the aromatic residues (tryptophans and histidines) were bound to the CNT, whereas in the case of the mutant peptides, only two or three aromatic residues adopt binding configurations reasonably close to the surface. In general, the spatial proximity of the histidine in the first position was found to be more diffuse (as evidenced by broad peaks in Figure 3.4 and 3.5, even for B1 and B3), as it is at the extremity of the peptide chain.

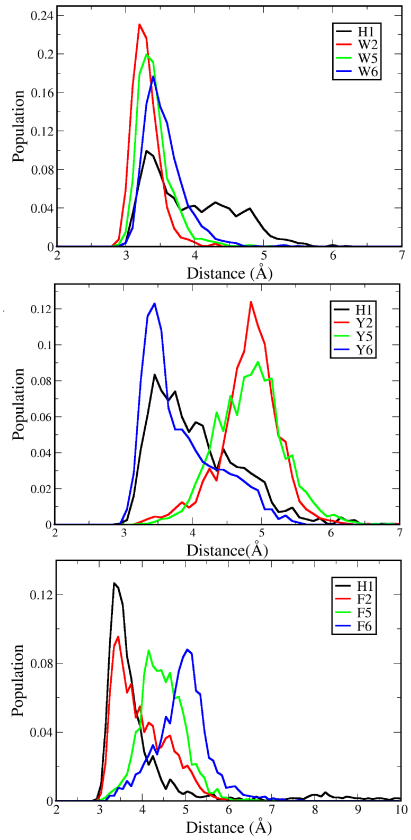

Figure 3.5: Distribution of the distance from the nanotube surface to the ring centre-of-mass for all the aromatic residues H, W, Y and F for the peptide sequences (a) B3, (b) B3Y, and (c) B3F.

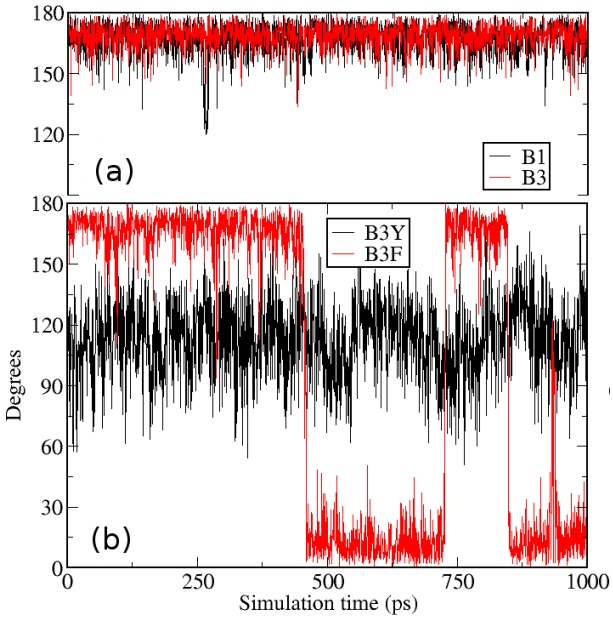

Figure 3.6: Example profiles of the ring-tilt angle (relative to the surface normal) at the peptide nanotube-interface as a function of time for (a) W2 in both original sequences B1 and B3 and (b) Y2 and F2 in B3Y and B3F, respectively.

Table 3.4: Average normalised interaction energies E_N (kJ mol^{-1}), for the lowest-energy trajectory of B1 and B3, the mutant sequences and the weak-binder NB1, interacting with a CNT. E_N was calculated using three different methods: dividing the interaction energy by **(a)** the number of peptide atoms; **(b)** the number of atoms within 6 Å of the surface, and **(c)** using a decay weighting function to any atom further than 4 Å(see Section 3.3).

Aptamer	E_N (kJ mol^{-1})		
	(a)	**(b)**	**(c)**
B1	−1.35	−1.54	−1.72
B3	−1.32	−1.60	−1.79
B1Y	−1.20	−1.45	−1.64
B3Y	−1.20	−1.41	−1.60
B3F	−1.17	−1.42	−1.61
B1F	−1.17	−1.46	−1.63
NB1	−0.81	−1.37	−1.46

An orientational analysis of the ring-tilt angle with respect to the surface normal was then undertaken to elucidate the origin of these differences in the ring-surface distances. An example of the profiles of the orientation with time of tryptophan in position 2 (W2) for both B1 and B3, contrasted with the same profiles for B3Y and B3F is shown in Figure 3.6(a). For the original sequences B1 and B3, W2 remains steadily in a predominantly flat orientation (such that the rings lie tangential to the surface plane), while the mutants exhibit less orientational stability. Figure 3.6(b) helps to understand the reasons for this instability; the example trajectory given for F2 in B3F yields instantaneous changes from 180 to 0° indicating ring-flipping events between flat orientations. On the other hand, the example trajectory of Y2 in B3Y, shows a profile where the ring fluctuates around an almost perpendicular orientation. These examples clearly illustrate the differences in the ring orientational behaviour of the original and mutated sequences and its subsequent effects on the binding affinity and do not necessarily suggest that Y

Table 3.5: Breakdown of the normalised interaction energy, E_N, (kJ mol^{-1} atom^{-1}) for each of the four distinct trajectories of lowest energy for each aptamer, for the peptide—nanotube interface.

Aptamer	E_N
B1-1	−1.36
B1-2	−1.34
B1-3	−1.29
B1-4	−1.28
B3-1	−1.33
B3-2	−1.30
B3-3	−1.27
B3-4	−1.27
B1Y-1	−1.20
B1Y-2	−1.18
B1Y-3	−1.18
B1Y-4	−1.06
B3Y-1	−1.20
B3Y-2	−1.13
B3Y-3	−1.12
B3Y-4	−1.03
B1F-1	−1.17
B1F-2	−1.10
B1F-3	−1.03
B1F-4	−0.98
B3F-1	−1.17
B3F-2	−1.16
B3F-3	−1.16
B3F-4	−1.10
NB1-1	−0.95
NB1-2	−0.87
NB1-3	−0.86
NB1-4	−0.81

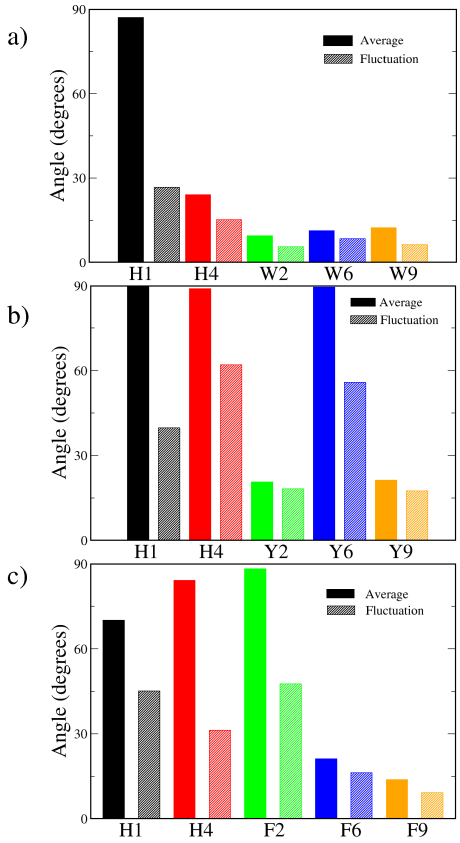

Figure 3.7: Average and fluctuation of the ring-tilt angle (relative to surface normal) at the peptide-nanotube interface for all aromatic residues in (a) the original sequence B1 and the mutated sequences (b1) B1Y and (c) B1F. A value of approximately 0° indicates an orientation parallel with the surface.

mutations always oscillates and F mutations always yield ring-flips. Instead, these examples illustrate two different behaviours that contribute to the decrease of the binding affinity of the mutated peptides. For instance, a ring that flips, such as in the case of F2 in B3F, often puts its ring centre-of-mass further from the surface compared to a ring that does not flip, giving rise to a lower affinity. Similarly, a ring being oriented perpendicularly to the surface (such as Y2 in B3Y is) also unable to get the ring as close to the surface as a ring that lies flat on the surface.

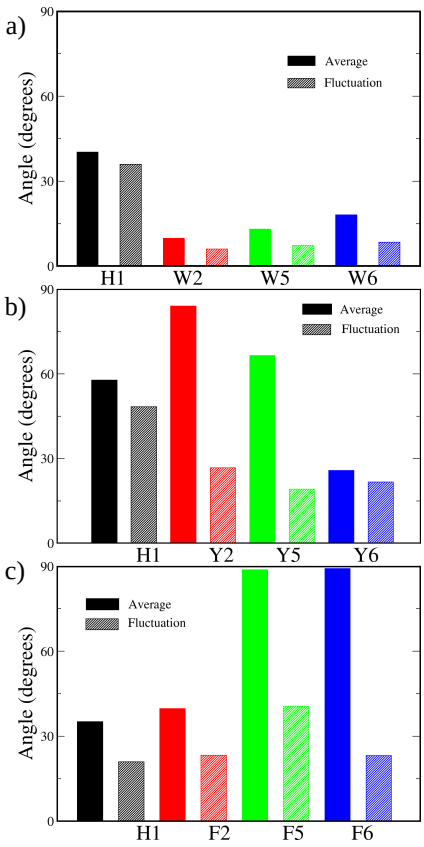

Figure 3.8: Average and fluctuation of the ring-tilt angle (relative to the surface normal) at the peptide-nanotube interface for all aromatic residues in (a) the original sequence B3 and the mutated sequences (b) B3Y and (c) B3F. A value of approximately 0° indicates an orientation parallel with the surface.

It was found that all tryptophans in B1 and B3 and the non-terminal histidine in B1 (H4), always adopted a flat orientation at the surface. This observation is supported by the average ring-tilt angle and the fluctuation in this angle (calculated as the root-mean-square deviation of the tilt angle), presented as a bar chart in Figures 3.7 and 3.8 for the best trajectories in each case. An average angle of around 0° indicates the ring is oriented perfectly flat on the surface, whereas a tilt of 90° indicates a perpendicular orientation of the ring. In addition, large fluctuations indicate that the side chain of the aromatic residue is flexible and therefore that the aromatic ring is not so strongly bound, such that the ring can adopt several orientations relative to the surface. Figure 3.7(a) indicates that tryptophans in B1 spend most of the time flat on the surface. The corresponding fluctuation is very small for all rings except H1, suggesting that once tryptophans bind to the CNT surface they remain stable with no significant changes in orientation. On the other hand, in each of the mutated sequences, B1Y and B1F [Figure 3.7 (b) and (c)], at least one of Y or F exhibits a large average tilt with a greater corresponding fluctuation. These data provide evidence that in general, the orientation of the rings in the tyrosine and phenylalanine side groups is more free than is seen for the indole ring in tryptophan when interacting with the surface. The same phenomenon was evidenced for the other configurations studied and also for the other peptide sequences B3, B3Y and B3F (Figure 3.8), reinforcing our assertion that Y and F are more mobile and bind less tightly at the interface compared to W. In the case of H4 in B1, a small average tilt and a small fluctuation indicates that this residue is also flat and stable on the nanotube surface. However, for the mutated sequences B1Y and B1F, the corresponding H4 average tilt is around 90°, with a greater fluctuation. In each case, one of the nearby aromatic residues is also orientationally unstable (Y6 in Figure 3.7 b) and F2 in Figure 3.7c)), which suggests that structural effects might affect

the binding of these "key" residues (Y6 and F2). Additionally, in agreement with the distance analysis, a general trend was also found for H1; in all configurations studied, H1 yielded a large fluctuation in the tilt angle.

How our findings changed upon small modifications of the force-field were also considered. Changing multipole moments on atomic sites without recalculating the distributed multipoles on the entire molecular fragment is not appropriate [133]. But examining how binding changed when we changed the polarisation of the carbon atoms of the nanotube is valuable. Table 3.6 reveals that the trends in E_N remain unchanged upon employing a CNT carbon polarisability either lower or higher (2.0 and 5.0 Å^3 respectively) than our original value (3.5 Å^3).

Table 3.6: Average normalised interaction energies, E_N (kJ mol^{-1}atom^{-1}), of the "best" trajectories for each aptamer, calculated as a function of atomic polarisability (Å^3) of the CNT carbon atoms, for the peptide-nanotube interface. All results reported in this work were obtained using an atomic polarisability of 3.5 Å^3.

Polarisability	**3.5**	**2.0**	**5.0**
System		E_N	
B1W	−1.35	−1.34	−1.35
B1Y	−1.20	−1.19	−1.20
B1F	−1.17	−1.16	−1.17
B3Y	−1.20	−1.19	−1.21
B3F	−1.17	−1.16	−1.17
NB1	−0.81	−0.81	−0.82

These findings show that mutations of the tryptophan content by either tyrosine or phenylalanine cause a decrease in the binding affinity to CNTs, however none of the mutated sequences reaches the lower affinity of the weak-binder control NB1. In order to investigate the influence of the interfacial shape, similar simulations were performed with the same peptides adsorbed on a graphene sheet.

3.4.3 Peptide—Graphene Simulations

Now let us focus our attention on the peptide sequences interacting with the graphene sheet. Normalised interaction energies (averaged over the four best trajectories) for each aptamer, and the corresponding SEM for the peptide-graphene interface are presented in Table 3.7. The breakdown in E_N for each of the four trajectories for each aptamer are given on Table 3.8. As found for the peptide-nanotube interface, these data show that B1 and B3 have greater binding affinity to the graphene sheet compared with the mutants; However, in this case, the spread of interaction energies (as shown in Table 3.8), especially for the original sequences, is wider than for the peptide-nanotube case. Therefore, there is not such a clear distinction in the normalised interaction energies between the original sequences and the mutant sequences at the peptide-graphene interface. Again, the mutations for either phenylalanine or tyrosine do not bring the binding affinity into the lower range supported by the NB1 weak binder control peptide. However, a slight diminishing of the binding affinity was noted for the mutant sequences. Again, the differences in E_N amongst the mutants is very small. In the case of the peptide-nanotube interface, the best trajectory for each aptamer (Table 3.5) shows that the B1Y mutant binds slightly better than the B1F mutant, with the same occurring for the B3 mutants. However, in the case of the peptide-graphene interface, the B3F mutant binds slightly better than B3Y, with the opposite being true for the B1 mutants. Still, inclusion of explicit solvation might change this ordering.

The ring to surface distance data are similar to that of the CNT, since again the original sequences maximise the ring-surface contact, compared with the mutants. In Figures 3.9 and 3.10 is the distribution profile of the distance of the ring centre-of-mass of the aromatic residues to the graphene sheet for the best trajectories of B1, B1Y and B1F and B3, B3Y, and B3F, respectively. In contrast to the CNT case, all rings in

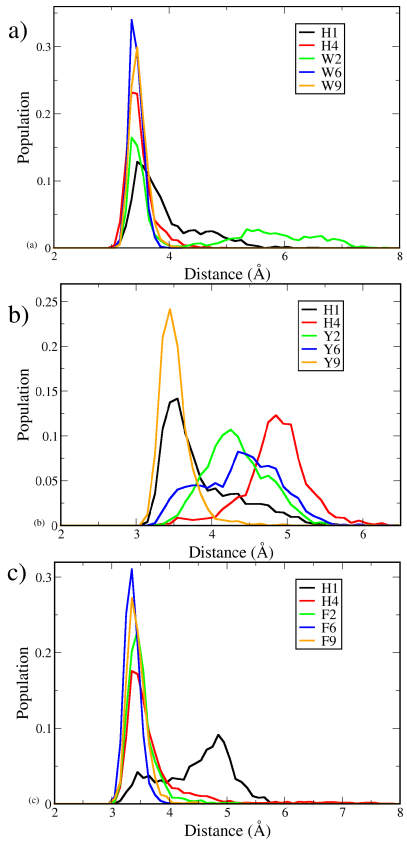

Figure 3.9: Distribution of the distance from the graphene surface to the ring centre-of-mass for all the aromatic residues H, W, Y and F for the peptide sequences (a) B1, (b) B1Y, and (c) B1F.

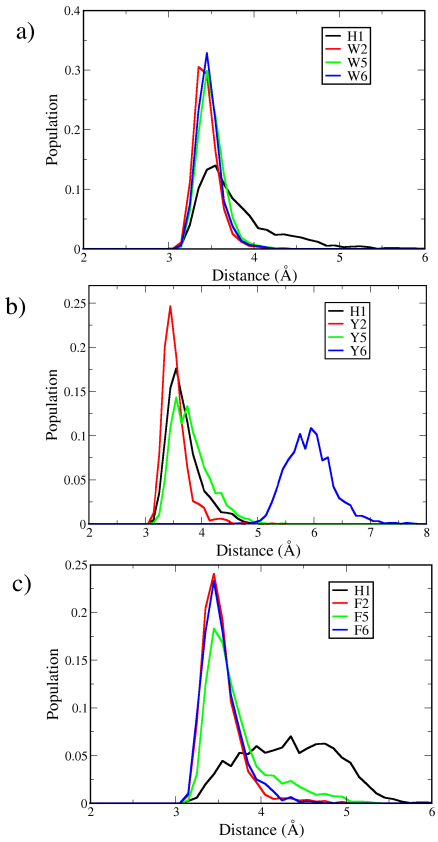

Figure 3.10: Distribution of the distance from the graphene surface to the ring centre-of-mass for all the aromatic residues H, W, Y and F for the peptide sequences (a) B3, (b) B3Y, and (c) B3F.

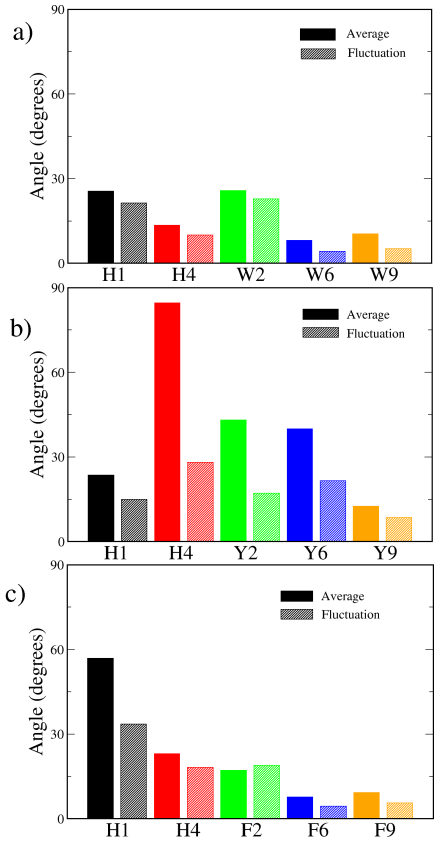

Figure 3.11: Average and fluctuation of the ring-tilt angle (relative to the surface normal) at the peptide-graphene interface for all aromatic residues in (a) the original sequence B1 and the mutated sequences (b) B1Y and (c) B1F. A value of approximately 0° indicates an orientation parallel with the surface.

Table 3.7: Normalised interaction energies, E_N (kJ mol^{-1}atom^{-1}), and the standard error of the mean (SEM) of E_N, averaged over the four distinct trajectories of lowest energy for each aptamer, for the peptide–graphene interface.

Aptamer	E_N	SEM
B1	−2.01	0.18
B3	−1.93	0.22
B3F	−1.88	0.07
B1Y	−1.87	0.08
B3Y	−1.84	0.07
B1F	−1.83	0.11
NB1	−1.61	0.03

the B1 sequence, including H1, are very close to the surface, with distribution peaks that are considerably less broad than noted for the nanotube interface. B1F also has more rings in contact with the surface, with only H1 not binding to the surface. B1Y shows unusual behaviour too; in this case the number of rings in very close contact with the surface is lower than in the case of the CNT (decreased from three rings close to the surface to two). Instead, this system supports two further rings with moderate surface contact *e.g.*, a separation of 4.3 Å to Y2 and 4.5 Å to Y6. Once more, the corresponding orientational analysis highlights the increased structural stability of the original sequences at the graphene surface. Figures 3.11 and 3.12 show the average ring-tilt and corresponding fluctuation in each tilt angle for B1, B1Y, and B1F, and B3, B3Y and B3F, respectively. These data are in good agreement with the distance profiles, which show that all the rings of B1 have very flat orientations, while for B1F, only the H1 ring exhibits a noticeable deviation from a flat orientation on average. For B1Y [Figure 3.11(b)], these data show the end-point regions of the chain are bound to the surface *via* the flat orientations of H1 and Y9, while the orientations of the central rings in H4, Y2 and Y6 are considerably more floppy, being in agreement with the fact

Table 3.8: Breakdown of the normalised interaction energy, E_N, (kJ mol^{-1} atom^{-1}) for each of the four distinct trajectories of lowest energy for each aptamer, for the peptide—graphene interface.

Aptamer	E_N
B1-1	-2.15
B1-2	-2.11
B1-3	-1.99
B1-4	-1.75
B3-1	-2.16
B3-2	-2.08
B3-3	-1.75
B3-4	-1.71
B1Y-1	-1.99
B1Y-2	-1.83
B1Y-3	-1.83
B1Y-4	-1.83
B3Y-1	-1.89
B3Y-2	-1.89
B3Y-3	-1.84
B3Y-4	-1.74
B3F-1	-1.96
B3F-2	-1.88
B3F-3	-1.87
B3F-4	-1.80
B1F-1	-1.92
B1F-2	-1.88
B1F-3	-1.82
B1F-4	-1.68
NB1-1	-1.66
NB1-2	-1.61
NB1-3	-1.60
NB1-4	-1.58

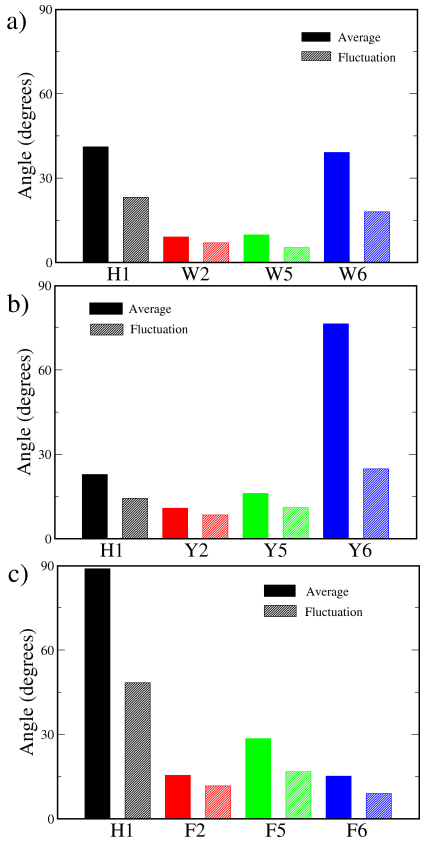

Figure 3.12: Average and fluctuation of the ring-tilt angle (relative to the surface normal) at the peptide-graphene interface for all aromatic residues in (a) the original sequence B3 and the mutated sequences (b) B3Y and (c) B3F. A value of approximately 0° indicates an orientation parallel with the surface.

that these residues do not get close to the surface [Figure 3.9 (b)].

3.4.4 Discussion

As explained in our previous work [118], the objective of this extended force-field is not to reproduce absolute binding energies, but rather to capture trends in binding. Despite the overestimate of the binding energies, our extended force-field yielded correct trends [118] compared with first-principles calculations [176]. Therefore, it is possible

that these calculated binding energies are an overestimate of the observed binding affinities. However, it is maintained that these binding affinities exhibit realistic trends upon mutation. According to the experiments of Wang *et al.* [94], *all* sequences, including NB1, showed some degree of binding to CNTs, as also found with the simulations using the AMOEBAPRO force-field. We believe that the expense in employing this force-field is justifiable for this study, since we have evidence [118] that the structural and orientational features of the aromatic rings play a key role in the binding affinity. The multipole-based description of the electrostatic interactions and inclusion of polarisation has been shown to be important in capturing changes in the binding energy related to the ring orientation relative to the surface [118]. Our previous work demonstrates that non-bonded interactions based only on VDW contributions do not necessarily compare well with first-principles calculations, (with more pronounced effects for the indole and imidazolium rings), whereas our force-field gives a satisfactory performance [118]. Therefore, it is argued that the use of this force-field is justified. However, for other peptide-CNT systems, where the aromatic rings are not a key feature, it might be sufficient to use the standard force-field.

By comparing the results obtained for aptamers interacting with a CNT or a graphene sheet it is concluded that B1 and B3 are strong binders for both surfaces, with the tyrosine and phenylalanine mutations yielding slightly less binding affinity. The results also suggest that tryptophan has more affinity for the surfaces than tyrosine and phenylalanine. These findings are in good agreement with the ranking in binding affinities reported by experimental studies on similar systems [101; 102], emphasising the importance of tryptophan in binding to carbon nanotubes [90; 119]. In our previous work we concluded that the aromatic content of the sequences is the key factor for the strong binding to CNTs. While tyrosine and phenylalanine are also aromatic, these residues

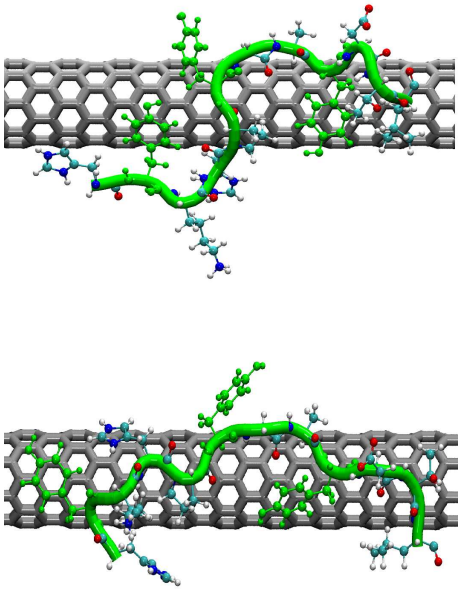

Figure 3.13: Snapshots taken from two different strong-binding trajectories (with different starting configurations) of the B1Y peptide interacting with the CNT. The backbone and the tyrosine residues are highlighted.

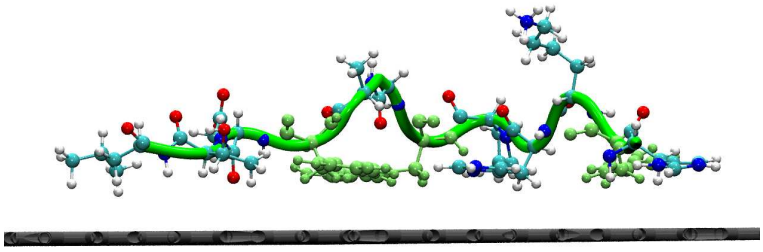

Figure 3.14: Snapshot from the "second-best" trajectory for B1Y on the graphene surface, illustrating the degree of backbone buckling at the interface.

are smaller than tryptophan. The greater stability of tryptophan at the interface can be attributed to the long aspect ratio of the indole group, which enables configurations where *all* the atoms can interact with the surface. While tyrosine and phenylalanine, despite binding with comparable strength (relative to tryptophan) on these graphitic surfaces on an atom-for-atom basis, they lack the same degree of stabilisation conferred by the indole group. Therefore, it is proposed it is this structural stability of the indole group on the CNT surface that gives rise to the energetic stability of the tryptophan-containing peptide sequences.

As in our previous work, the trajectories of B1, B3, and their mutants show that strong binders support a multitude of different spatial and arrangements of the aromatic groups on the nanotube/graphene surface, giving further evidence to support our earlier hypothesis [118] that peptide affinity can be attributed, in part, to the notion that "strong-binder" aptamers support a variety of strong binding configurations [118]. For illustration an example is shown in Figure 3.13, in which two different strong-binding configurations taken from different trajectories of the B1Y peptide are interacting with the CNT, emphasising the different possible strong-binding conformations of the peptide backbone.

It was noted in the Results Section (3.4.3), that the best trajectory (strongest binding) of the B1Y mutant on the graphene surface actually supported fewer rings in close contact with the graphene sheet compared with the nanotube simulations. By examining the normalised interaction energies of B1Y and B1F (Table 3.8) it is noted that the two are indeed very close in binding affinity, despite the fact that B1F supports a greater number of rings close to the surface. Actually, B1Y configurations with all five aromatic residues bound on the sheet were found; however, in these cases, the binding affinity was *weaker*. This apparent contradiction can be rationalised by examining the

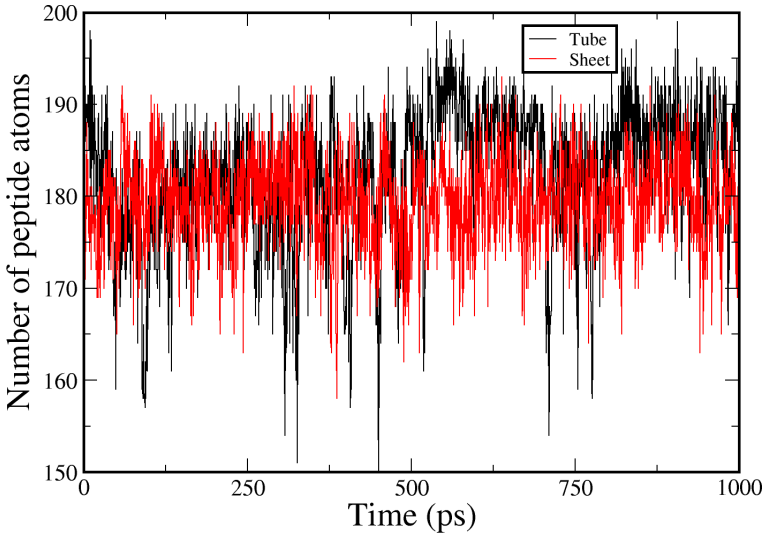

Figure 3.15: Number of peptide atoms within 6 Å of the surface as a function of time for the B1 peptide, for both the peptide-nanotube and peptide-graphene interfaces.

remaining (non-aromatic) residues in the chain. In the case of the weaker-bound B1Y configuration that supported more ring-surface interaction (denoted as the second best trajectory), some of the non-aromatic residues in the peptide were pointing away from the surface and therefore not contributing to the interaction with the surface, diminishing the overall binding affinity. Figure 3.14 clearly shows a buckling of the peptide backbone in the regions of the chain between the ring positions. This backbone buckling is proposed to come from the fact that the presence of the aromatic rings close to the flat surface does not allow the non-aromatic residues to also approach the surface. In the case of the "best" trajectory, some of the rings (Y6, Y9) are maintained at a medium distance away from the surface, and the peptide backbone is not buckled, which allows the remaining residues to also approach the surface, giving rise to a competitive binding affinity compared with B1F (see Table 3.8). This buckling behaviour was not noted in any of the peptide—nanotube simulations. Based on these data it is proposed that

these CNT binder sequences, have not only been selected to bind strongly to a graphitic surface, but also to bind best at *curved* graphitic surfaces. As an example, the number of peptide atoms within 6 Å of the surface in the case of the "best" B1 simulations on both the CNT and graphene were calculated. Figure 3.15 clearly shows that more peptide atoms are closer to the surface at the peptide-nanotube interface. It is likely that the strongest graphite-binding peptides will contain residues in specific sequences that allow closer contact at the surface for most atoms of the peptide chain. Furthermore, the modes of binding at the graphene interface appear to be different from those seen for the CNT interface.

3.5 Conclusions

The binding affinity of several peptide sequences adsorbed at two graphitic surfaces, graphite and a carbon nanotube, was studied using MD simulations. Two tryptophan-rich peptide sequences with strong affinity to CNTs were mutated by changing the tryptophan residues by either tyrosine or phenylalanine. In agreement with recent experimental observations, it was found that none of the mutants could outperform the original sequences in terms of binding affinity. In addition, none of the peptide mutants reached the lower binding affinity supported by the weak-binder peptide. The relatively greater structural stability of the indole ring in tryptophan appears to be the origin of these differences. It was also found that the original sequences showed the strongest binding to the graphitic surfaces regardless of the interfacial shape. However, these data suggest that aromatic groups may be more abundant in peptides selected for binding to curved graphitic surfaces rather than flat surfaces.

Chapter 4

Modelling the effect of the peptide sequence on the binding affinity for carbon nanotubes

4.1 Abstract

In this Chapter we revisit the study of the interactions of peptide sequences with single walled carbon nanotubes (SWCNTs) by molecular dynamics (MD) simulations. Recent studies on the adsorption of peptides onto inorganic surfaces have pointed out the importance of the conformational structure of the peptides on the binding affinity. The purpose of this study is to better understand how the order of the residues of the peptide sequences may affect interactions with a surface; in this case a carbon nanotube. To achieve this, an experimentally-identified [94] peptide with strong affinity for carbon nanotubes (CNTs) was used, namely B1, with sequence (`HWKHPWGAWDTL`). Fifteen peptide sequences with the same residue content as the original sequence, but with these

residues presented in a different order were created. Interestingly, one peptide sequence with a stronger binding affinity than the original sequence (B1) was found. While none of the scrambled sequences reach the lower binding affinity shown by a weak-binder control peptide (as discussed in Chapter 3), some of the sequences studied here support a binding affinity comparable to those for mutant sequences (where the tryptophan content was replaced by either tyrosine or phenylalanine). These findings show that the binding affinity to CNTs is dependent on the order of the residue content; in other words, that the peptide sequence matters. These observations draw some useful insights into the possible mechanisms by which peptides bind to inorganic surfaces. In addition, these findings also form a preliminary basis for identifying sequence motifs that may lead to designed sequences with predictable binding affinities for CNTs. For instance, it was found that weak binding may be due to the presence of two tryptophan (W) residues intercalated by another residue (r) (WrW), while strong binding may be due to the presence of two tryptophans together (WW).

4.2 Introduction

In Nature, the hierarchical structure and highly functional biological hard tissues are accomplished in part *via* molecular recognition between solids and biological molecules such as peptides [177; 14]. In the same way, the fabrication and utility of advanced biomaterials, engineered tissues and therapeutic devices are dependent on the affinity and selectivity of biomolecules at inorganic surfaces [1]. Consequently, combinatorial selection methods such as phage-display and cell-surface display have become the major techniques for the selection of a wide variety of peptides with strong affinity to a given inorganic surface [7; 14; 10; 1; 18; 178; 179]. In addition, there has been a growing interest in the use of inorganic-binding peptides for the synthesis and assembly

of materials with controlled properties for applications in nanotechnology and medicine [94; 6; 22]. While the number of inorganic-binding peptides reported in the literature keeps increasing, there is still limited knowledge on the mechanisms underpinning the basis of the peptide-inorganic surface recognition. As pointed out in Section 1.3.1 of Chapter 1, a peptide that binds strongly to an inorganic surface is referred to as having high "affinity", while a peptide that binds to a specific surface and not to other surfaces is denoted as binding with "specificity". Due to the poor understanding of the complex interaction between peptides and inorganic materials it has been difficult to identify *a priori* a sequence that has a binding specificity to a desired target material [180; 87]. Understanding the nature of the molecular recognition and the degree of affinity of a selected peptide to a given surface would enable tailoring the peptide binding affinity and specificity towards a given target surface [21]. Several studies have aimed to understand peptide-inorganic recognition, mainly by focusing on the amino acid composition of the peptides [16; 94; 178; 181; 85; 181; 17]. However, recent studies have pointed out the importance of not only the composition but also the order of the content (which is the sequence), as being important for the binding affinity and specificity [84; 85]. For instance, Goede *et al.* [84] observed a siginificantly drop in binding affinity to silica by randomising a peptide sequence. In addition, this randomised sequence also gave rise to a loss of binding specificity (binding to both silica and GaAs, where the original sequence did not bind well to GaAs). Simlarly, Belcher *et al.* [85] demonstrated the importance of not just the composition but also the sequence in the modulation of binding strength and observed differences in binding affinity to CdS, ZnS and Au by rearranging the sequence (and maintaining the composition). In addition, Belcher *et al.* defined a criterion for tuning peptide binding affinity and specificity to CdS, ZnS and Au, based on the idea that certain residues are responsible for the direct binding to the target surface,

while other residues locally modulate the environment in which the binding residues are incorporated [85].

Another relevant work is the study of Choe *et al.* [13] which probed the influence of the sequence and conformation on Cu_2O binding affinity. The effect of peptide conformation was studied by using linear and cyclic versions of the peptide. Their results suggested that in some cases, conformation may be more important than composition itself, in determining peptide-inorganic affinity. Oren *et al.* [108] pointed out the nature of the binding affinity as being an interplay between sequence, conformation and binding, highlighting that the peptide composition is necessary but not sufficient for identifying differences in affinity. Hayashi *et al.* [91] systematically modified the peptide sequence in order to clarify the role of each residue in the interaction with TiO_2 and demonstrated that the strength of binding decreased upon shuffling the sequence. Their results indicated that not only polar and charged residues play an important role in the strength of the binding, but also the order of the residues was found to be relevant [91]. In addition, the results of Hayashi *et al.* demonstrated that the peptide structure determines the arrangement of charged residues and therefore affects significantly the binding strength [91]. This Ti-binding peptide was studied in more detail by atomistic molecular dynamics (MD) simulations [117] and the results suggested that not just the electrostatic interactions were relevant for the binding but intra-peptide interactions may also be important, again underscoring the importance of being able to control the conformation of the peptide at the interface. A recent study of Oren *et al.* [182] gave further evidence that the local molecular environment in which the key residues are integrated plays a major role in the peptide-inorganic binding. All of these studies lead to the conjecture that the sequence plays an important role in affecting the binding affinity of peptides adsorbed on inorganic surfaces. The work described herein aims to

investigate the effect of the peptide sequence on the binding affinity to CNTs.

The extraordinary physical and mechanical properties of CNTs, with promising applications in the fields of nanotechnology and medicine, have attracted the attention of many researchers. Nevertheless, these potential applications face the problem that CNTs do not disperse well in aqueous solution. As mentioned in Chapter 3, the ability of biomolecules, such as peptides, to recognise different inorganic materials gives rise to new expectations for the selective dispersion and controlled manipulation of CNTs. The importance of the aromatic groups on the binding affinity to CNTs has been reported previously [120; 97; 88], and tryptophan has been underlined as being particularly important for the peptide-nanotube interaction [120; 101; 90; 119; 102](see Chapter 3). A complete literature review on peptides interacting with CNTs has been given in Chapter 3.

Table 4.1: Original sequence B1 and scrambled peptide sequences used in this study.

Aptamer	Sequence
B1(original)	HWKHPWGAWDTL
B1-1	WHPGLHTWKADW
B1-2	HKWWWHPGADTL
B1-3	KGHAWDPHWLWT
B1-4	LWKWGDAHHWTP
B1-5	HWKLDHWWTAGP
B1-6	HWWAPDKTGWHL
B1-7	PDHAWWKWTHLG
B1-8	HHWAGKPWDTLW
B1-9	WWLHHPGTKADW
B1-10	HWWHPADLTGWK
B1-11	WTPWKGWHADLH
B1-12	LWPWHDAGTWKH
B1-13	GHTAWDWHWPLK
B1-14	AWLKPHHWWDGT
B1-15	AWDHPHLKWGWT

Common to the use of combinatorial selection methods, such as phage-display

and cell-surface display is the lack of prior knowledge about the interaction between peptides and the target inorganic material [6; 183; 1]. It would be useful to understand the influence of the residue sequence and conformational structure on the recognition process at the inorganic interface. Understanding the relation between binding and structure would enable the tuning of the binding properties to a specific surface by choosing the peptide sequence. Based on the importance of the aromatic content to the strong affinity for CNTs, and on the recent studies which suggest that the sequence also plays a role in determining binding affinity, the aim of this work is to study the influence of the order of the amino acids on the experimentally determined strong binder sequence, HWKHPWGAWDTL (denoted B1) selected by phage-display experiments. Herein the importance of the order of the content of a peptide sequence on the binding affinity to CNTs is demonstrated, by scrambling B1. Fifteen peptide sequences with the same residue content as B1, but in a different order, interacting with a CNT were studied by atomistic MD simulations. In Table 4.1 the scrambled sequences and the original sequence B1 are presented. It is anticipated that, in terms of binding affinity, the sequences that have the aromatic residues together (W and H), or even the three tryptophans together may have a weaker binding affinity compared to the sequences that have the aromatic content spread in the sequence. This is because the presence of aromatic residues close together may give rise to steric effects between the aromatic rings which prevent the non-aromatic residues from getting close to the CNT surface. On the other hand, having the aromatic content spread throughout the sequence may avoid this effect. However, it is kept in mind that other factors may also affect the affinity.

4.3 Methods

MD simulations of systems consisting of one CNT and one each of the fifteen scrambled peptide sequences were performed with our extended AMOEBAPRO [142; 170] force-field. Most of the scrambled peptide sequences were created by mixing the amino acid content of the strong binder B1 sequence (HWKHPWGAWDTL) [94], using a random number generator, however three sequences were created by hand. In total, fifteen sequences with the same amino acid content of B1 were generated.

A hydrogenated zigzag (8,0) CNT with 976 atoms was modelled; this was at least four times the length of the extended peptide. All the simulations were carried out in the canonical ensemble at room temperature. The simulation details were the same as described in the Methods of Chapter 3. Depending on the equilibration period, each simulation was run for 2-10 ns where the last 1 ns was used to analysed the trajectories. To attempt to identify many different binding configurations, between 5 and 7 different starting configurations were considered for each peptide, yielding a total of around 100 distinct peptide-CNT simulations. Several methods were used to create initial configurations for these simulations. First, peptide configurations were constructed by hand and placed near the CNT surface. The geometries ranged from simple extended backbone configurations roughly aligned along the nanotube axis, through to configurations where the backbone torsion angles were set to create a pseudo-helical peptide geometries that wrapped the nanotube. Second, simulations of the peptide without the nanotube were performed in implicit solvent and configurations from these simulations were used and placed close to the CNT. Finally, the dihedral angles from the B1-CNT geometries that yielded strong binding were taken and used to subsequently construct configurations for the scrambled sequences.

Energetic and structural metrics were used to analyse the results. Binding affinity

was measured, again in terms of interaction energy between the peptide and the CNT. To analyse the structure at the interface the distance from the centre-of-mass of each aromatic ring to the surface, and the orientation of each aromatic ring (with respect to the surface normal) were calculated. In addition, the distance from the centre-of-mass of the side chain of each residue adjacent to an aromatic residue to the CNT surface was calculated. The distance between residues in the chain was also calculated by averaging the centre-of-mass separation of the side-chains associated with these residues. Furthermore, the number of peptide atoms and backbone atoms within 6.0 Å to the CNT surface along with the respective fluctuation were calculated for each sequence. The average end-to-end peptide distance was calculated by determining the distance between each carbon of each of the two terminal methyl groups (blocking groups) and averaging this distance over the production run. The three closest peptide atoms to the CNT were also indentified as a function of time for the three best binders and the three worst binders. This was achieved by calculating the distance between all the peptide atoms to the CNT surface and considering only the atoms that were within the cutoff distance (8 Å). The frequency at which each peptide atom was the closest to the surface was then determined by counting the number of times each atom was the closest and dividing by the number of frames of the production run. It is emphasised that due to the large number of peptide sequences, in some cases, data is shown only for certain sequences in order to highlight any relevant change in the properties.

4.4 Results

As mentioned in the Methods, three of the fifteen sequences were made by hand while the others were created with a random number generator. The sequences created by hand were B1-1, B1-2 and B1-3. B1-1 was created with the intuit to study the effect of

Table 4.2: Interaction Energies, E_{int} (kcal mol^{-1}), and Normalised Interaction Energies, E_N (kJ mol^{-1}atom^{-1}), and the Standard Error of the Mean (SEM) of E_{int}, averaged over the four and two distinct trajectories of lowest energy for each aptamer.

	4 Best trajectories			**2 Best trajectories**		
Aptamer	E_{int}	SEM	E_N	E_{int}	SEM	E_N
B1 (original)[120]	68.79	1.98	−1.32	−70.42	0.98	−1.35
B1-6	**−69.95**	2.67	**−1.34**	**−72.05**	1.80	**−1.38**
B1-5	−67.18	6.05	−1.29	−70.87	1.34	−1.36
B1-7	−66.54	3.53	−1.28	−68.78	3.98	−1.32
B1-9	−65.68	1.98	−1.26	−66.77	0.27	−1.28
B1-11	−65.66	5.11	−1.26	−69.97	0.57	−1.34
B1-1	−65.46	6.14	−1.26	−70.75	1.20	−1.36
B1-10	−64.65	2.69	−1.24	−66.47	1.17	−1.28
B1-8	−61.96	5.51	−1.19	−66.06	2.50	−1.27
B1-2	−61.41	5.70	−1.18	−65.21	1.71	−1.25
B1-3	−60.94	6.47	−1.17	−64.23	1.39	−1.24
B1-15	−60.60	1.00	−1.16	−61.45	0.01	−1.18
B1-4	−60.29	8.27	−1.16	−62.53	6.42	−1.20
B1-13	−59.79	5.62	−1.15	−64.35	1.43	−1.24
B1-12	−59.69	2.43	−1.14	−61.09	0.78	−1.17
B1-14	**−59.19**	1.02	**−1.14**	**−59.93**	0.04	**−1.15**
NB1	−36.50	2.40	−0.87	−38.03	2.50	−0.909

having tryptophan residues at the chain ends, while B1-2 was created to study the effect of having all tryptophans and one histidine together. Prior to the simulations, it was expected that B1-2 would bind weakly to the CNT due to steric effects of having three tryptophans together (`WWW`), while B1-1 was expected to bind more strongly (but less strongly than the original sequence) since it has one tryptophan at each end and another tryptophan in the middle of the chain, and therefore could "anchor" the CNT. Similarly, it was anticipated that B1-14 would bind weakly to the CNT due to the presence of two histidines and two tryptophans together.

For each aptamer, the interaction energies, E_{int}, averaged over both the top two and the top four distinct binding trajectories (that yielded the best binding affinities,

i.e. the lowest interaction energy) are reported. The normalised interaction energies, E_N (interaction energy between peptide and surface, divided by the number of peptide atoms) are also reported in order to make a comparison with the energies reported in our previous work [120]. These data are presented in Table 4.2 along with the standard error of the mean (SEM), as defined by the standard deviation in the sample means. While none of the scrambled peptide sequences diminishes the binding affinity to a level comparable with NB1 (the weak-binder control peptide), some peptide sequences exhibit a reduction in the binding affinity comparable to the mutation of the tryptophan content by either tyrosine or phenylalanine [120]. Furthermore, peptide sequences with different affinities to the nanotube, including affinities higher than the original sequence B1, were found. This suggests that the order of the residues, that is, the sequence, is important to the binding affinity. For each sequence, the best trajectory was analysed and the ranking obtained from the average over the four best trajectories was considered.

It is proposed that the origin of differences in the binding affinity is related to differences in the conformational structure of the peptides. In an attempt to understand the origin of these differences the scrambled peptide sequences are divided in two groups; one group includes the sequences with the highest affinity to the CNT, (denoted as *strong binders* herein) and the other group includes the sequences which affinity to the CNT is comparable to the affinity of the tyrosine/phenylalanine mutants (denoted *medium binders*) as found in our previous study [120]. Table 4.3 shows the peptide sequences for each group. The strong binders group includes the original sequence B1 and all the sequences with similar affinity. The line between strong and medium binders was drawn at -65 kcal/mol, since the typical interaction energy of the tyrosine/phenylalanine mutants are for energies greater than this value. In order to search for particular domains responsible for stronger and/or weaker binding, we sought for common residue motifs

Table 4.3: Scrambled peptide sequences divided in two groups; strong and medium binders. Sequences are ordered from highest to lowest affinity to the CNT.

Aptamer	Sequence
Strong binders	(Comparable to B1)
B1-6	HWWAPDKTGWHL
B1(original)	HWKHPWGAWDTL
B1-5	HWKLDHWWTAGP
B1-7	PDHAWWKWTHLG
B1-9	WWLHHPGTKADW
B1-11	WTPWKGWHADLH
B1-1	WHPGLHTWKADW
Medium binders	(Comparable to Y/F mutations)
B1-10	HWWHPADLTGWK
B1-8	HHWAGKPWDTLW
B1-2	HKWWWHPGADTL
B1-3	KGHAWDPHWLWT
B1-15	AWDHPHLKWGWT
B1-4	LWKWGDAHHWTP
B1-13	GHTAWDWHWPLK
B1-12	LWPWHDAGTWKH
B1-14	AWLKPHHWWDGT

present in each group (either strong or medium). In Table 4.4 is presented the frequency of each possible motif for both the strong and medium binder sequences. These data show that strong binder sequences have predominantly WK, WW and LH as diad motifs, whereas medium binders tend to feature WD, HH, PH, AW and LK. Both classes of binders heavly feature HW, and moderately featured WH, HP and WT and thus cannot be attributed to strong or weak binding affinity. In terms of triad motifs, with the exception of B1-8, the medium binder sequences feature either two tryptophans separated by another residue (W-r-W, where *r* is any residue)(B1-3, B1-4, B1-12, B1-13 and B1-15) or four aromatic residues together, (denoted as quartet, ar-ar-ar-ar, where *ar* is an aromatic residue)(B1-2, B1-10 and B1-14) or an aspartic acid and histidine residue separated by

one (D-r-H) or two residues (D-r-r-H)(B1-3, B1-4, B1-10, B1-13, B1-14 and B1-15). The strong binder sequences tend to have tryptophan residues assembled in pairs (B1-5, B1-6, B1-7, B1-9); the medium binder sequences also have two tryptophans together, however they are incorporated in a greater motif such as WWW, HWW or WWH and therefore do not count. On the other hand, the medium binder sequences strongly feature two tryptophan residues intercalated by any other residue (the W-r-W motif), which usually pushes the intercalated residue further away from the tube surface – usually beyond 8 Å (B1-3, B1-4, B1-12, B1-13 in the medium binders). This is because the contact between tryptophan residues and the CNT surface is always maximised.

Figures 4.1 and 4.2 show the distribution profiles of the centre-of-mass to CNT surface distances for the aromatic residues H and W and for the residues that are adjacent, for the strong and medium binder peptides, respectively. Figure 4.2 shows the residues (r) in between two tryptophans (WrW) are found far away (beyond 8 Å) from the nanotube surface, particularly L10 in B1-3 (8.36$\pm$0.81Å), P3 in B1-12 (7.30$\pm$0.33 Å), and D6 and H8 in B1-13 (9.74$\pm$0.41 Å and 10.05$\pm$0.61 Å, respectively). The centre-of-mass average distance of these residues (L, P, D and H) to the CNT surface when not in WrW is 4.53$\pm$0.36, 4.81$\pm$0.90, 7.71$\pm$1.19 and 4.80$\pm$1.64 Å, which is significantly smaller than when in the motif WrW. Interestingly, the B1-7 sequence in the strong binders also have two tryptophans separated by one residue (WKW). Figure 4.1 shows the two tryptophan residues in B1-7 (W6 and W8) in close contact with the CNT surface, while lysine (K7) is far from the CNT surface. The difference compared with the medium binders, is that in this case, K7 is still within the cutoff distance (5.48$\pm$0.73 Å). A similar phenomenon is seen for a tryptophan and a histidine residue separated by another residue (W-r-H or H-r-W, where *r* stands for any residue, Table 4.4), such as A4 and T9 in B1-7 in the strong binders (Figure 4.1) and D3 in B1-15 in

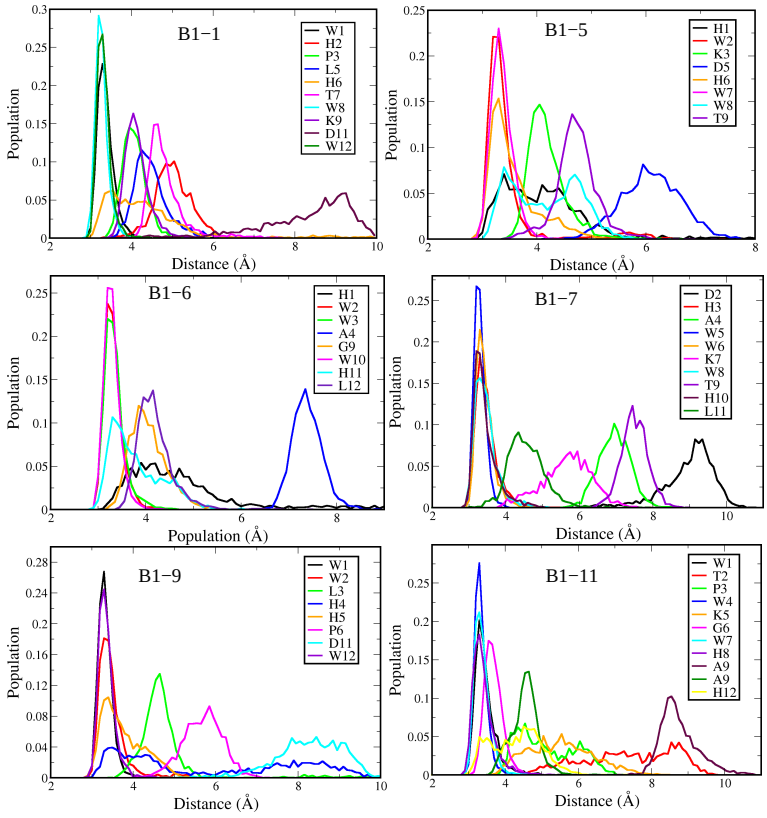

Figure 4.1: Distribution profiles of the distance from the CNT surface to the ring centre-of-mass for the aromatic residues, H and W, and for the nearby residues for the strong binders.

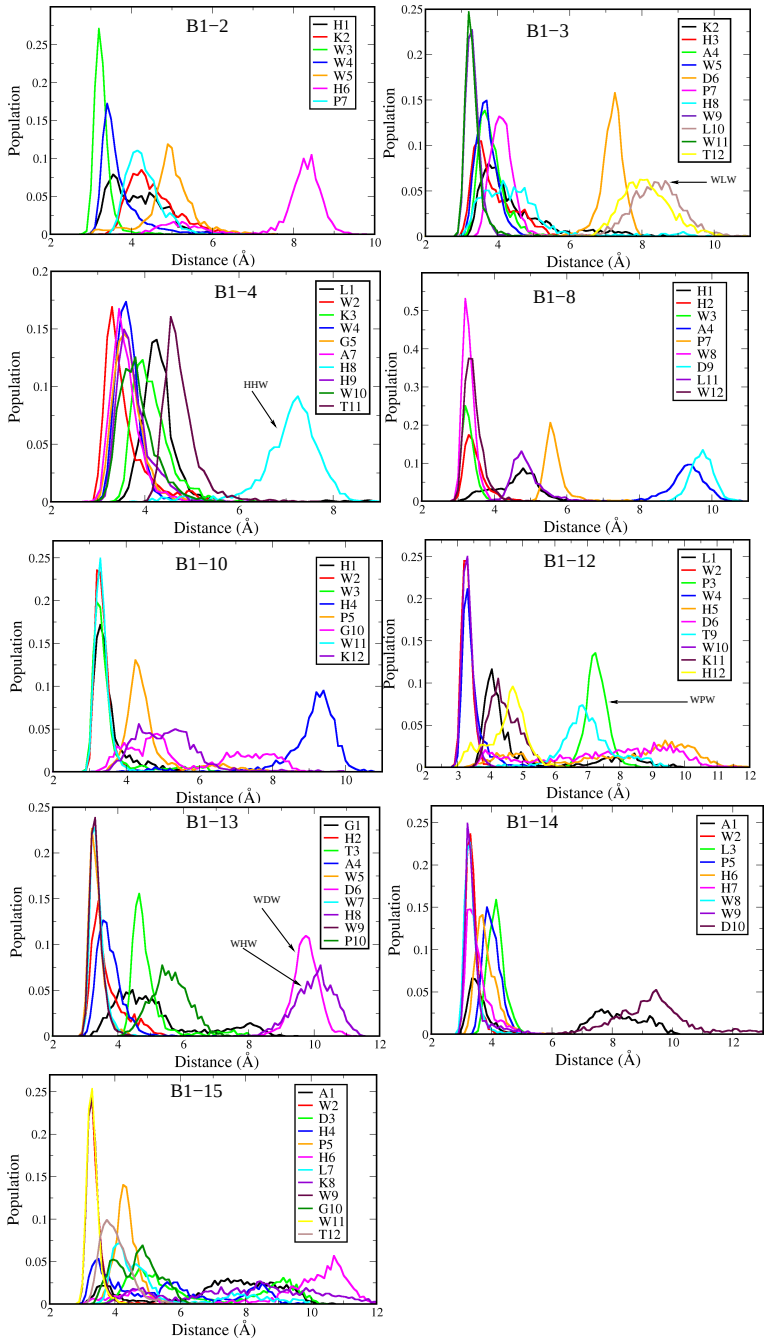

Figure 4.2: Distribution profiles of the distance from the CNT surface to the ring centre-of-mass for the aromatic residues, H and W, and for the nearby residues for the medium binders.

Table 4.4: Motifs featured in the peptide sequences along with their frequency in the strong and medium binders, respectively. *ar* stands for any aromatic residue (H or W) and *r* stands for any residue, including aromatic.

2aa Motif	Strong binders	Medium binders
HW	4/7	6/9
WH	3/7	3/9
HH	1/7	**3/9**
WW	**4/7**	3/9
HP	3/7	3/9
PH	—	**3/9**
HK	—	1/9
KH	1/7	1/9
HL	2/7	1/9
LH	**3/7**	—
HT	—	1/9
TH	1/7	—
WK	**5/7**	2/9
KW	—	3/9
WA	1/7	1/9
AW	1/7	**4/9**
WP	—	1/9
PW	2/7	2/9
WD	1/7	**5/9**
DW	2/7	1/9
WT	3/7	3/9
TW	1/7	1/9
DK	1/7	—
KL	1/7	—
KL	1/7	—
LK	—	**3/9**
KG	1/7	1/9
GK	—	1/9
AD	2/7	1/9
DA	—	2/9
AP	1/7	—
PA	—	1/9
GD	—	1/9
DG	—	1/9
GP	1/7	—
PG	2/7	1/9
TA	1/7	1/9
TL	1/7	2/9
LT	—	1/9

3aa Motif	Strong binders	Medium binders
HWW	2/7	2/9
WWH	—	2/9
HHW	—	3/9
WWW	—	1/9
WKH	1/7	—
HKW	—	1/9
HWK	2/7	—
WrW	1/7	**7/9**
ar-ar-ar	1/7	**6/9**
ar-r-ar	8/7	19/9
arKar	2/7	3/9
WKW	1/7	1/9
WKH	1/7	1/9
D-r-H	1/7	2/9
H-r-D	1/7	—

4aa Motif	Strong binders	Medium binders
ar-ar-ar-ar	—	**3/9**
HHWW	—	1/9
HWWH	—	1/9
WWWH	—	1/9
D-r-r-H or **H-r-r-D**	—	**6/9**
D-r-r-H	—	2/9
H-r-r-D	—	4/9

the medium binders (Figure 4.2), however in this case the residue in between histidine and tryptophan is positioned such that it is not as far away as in the case of WrW. An exceptional case of a histidine and tryptophan separated by another residue (H-r-W) is the motif HHW, found in the sequences B1-4 and B1-8 of the medium binders, where the intermediate residue (r) is an histidine (Table 4.4). In both sequences, it is not the residue in the middle of histidine and tryptophan that suffers the penalty of being further away from the surface, but rather the histidine in the extremity. For instance, in Figure 4.2 it is possible to see H8 in B1-4 and H1 in B1-8 far from the CNT surface. However, this might be an exception because it is at the chain end. Another exceptional case of a histidine and tryptophan separated by another residue is the motif WWH where in this case, the intermediate residue is a tryptophan. This motif is featured in the sequences B1-2 and B-10 and in both instances is incorporated in a quartet (ar-ar-ar-ar, where *ar* stands for an aromatic residue, Table 4.4). Again, as shown in Figure 4.2, it is not the intermediate residue (W in this case) that is far from the CNT surface, but the extremal histidine (H6 in B1-2 and H4 in B1-10). In the case of the motif HWW, a similar behaviour is found for W8 in B1-5 and H1 in B1-6 (Figure 4.1), *i.e.*, the residue in the extremity of the motif is further away from the CNT surface. However, in the case of B1-10 and B1-14 in the medium binders, there is no evidence of such behaviour. However, in these cases the motif HWW exists within a quartet (HWWH in B1-10 and HHWW in B1-14); it is suggested that other factors might be contributing to the lower binding affinity and it will be discussed later. Similarly, the motif WWW featured in B1-2 also puts one of the tryptophans in the extremity (W5) further from the surface (Figure 4.2). But again, this motif is incorporated into a quartet and perhaps attributing this behaviour to the presence of three aromatic residues together, *per se*, might not be appropriate. A proline residue in between two histidines in B1-15 is also found to be

considerably distant from the CNT surface (Figure 4.2). In general, it seems that two aromatic residues separated by another residue (ar-r-ar) will place one of the residues further from the surface. However, B1-4 features two tryptophans spaced by a lysine residue (WKW), and none of the residues is very far from the CNT surface; however, the binding affinity is low. This apparent inconsistency can be understood by examining the remaining residues of the B1-4 peptide chain. One suggestion to explain this is a possible weak interaction between H8 and W10, and H9 and T11, which seems to cause these residues to be positioned further away, giving rise to a tryptophan ring-tilt angle higher than that previously seen in Chapter 3 (Figure 3.7 in Chapter 3). Table 4.5 summarises the average and fluctuation of the ring-tilt angle for the tryptophan residues in the top three strongest binders and the four worst medium binders, and the average ring-CNT distance and fluctuation for the best binder B1-6, and the medium binders B1-2, B1-10 and B1-14. The ring-tilt average of W10 in B1-4 is 50.12°, which is significantly higher than in the other cases. These observations suggest that the presence of motifs comprising two aromatic residues spaced by another residue, (ar-r-ar), is a characteristic of weaker binder peptides and may help in the distinction of weak from strong binder peptides.

Intra-peptide interactions in the chain similar to the interactions observed in B1-4 (*i.e.* R1×R2, where R1/R2 is histidine) were also found in other sequences. Table 4.6 summarises the proposed relevant intra-peptide interactions. This evidence is only based on distances and needs to be supported with energies. It is important to note that this behaviour was only found for the medium binder peptides. Aspartic acid (D) is always, on average, relatively far from the surface and in the case of B1-3, the histidine (H8) is far from the surface in order to interact with D6. Figure 4.3 shows a snapshot of D6 in close contact with H8 in the B1-3 sequence. In addition, W5 is slightly further away from the

Table 4.5: Average and fluctuation of the distance (in Å) from the CNT surface to the ring centre-of-mass for the tryptophan residues in the strongest binder B1-6 and the medium binders B1-2, B1-10 and B1-14, and average and fluctuation of the ring-tilt angle (relative to surface normal) (in degrees) for tryptophan residues in the top three strongest binders and the four worst medium binders. Sequences are ordered from the highest to the lowest affinity.

Distances									
Aptamer	Residue	Average	RMSD	Residue	Average	RMSD	Residue	Average	RMSD
B1-6	**W2**	3.36	0.18	**W3**	3.39	0.22	**W10**	3.34	0.17
B1-2	**W3**	3.36	0.26	**W4**	4.51	1.06	**W5**	5.16	1.05
B1-10	**W2**	3.40	0.21	**W3**	3.57	0.66	**W11**	3.39	0.20
B1-14	**W2**	3.40	0.24	**W8**	3.38	0.27	**W9**	3.46	0.38
Ring-tilt angles									
Aptamer	Residue	Average	RMSD	Residue	Average	RMSD	Residue	Average	RMSD
B1-6	**W2**	8.70	5.12	**W3**	10.68	6.96	**W10**	8.09	4.90
B1-orig.	**W2**	9.54	5.58	**W6**	11.28	8.40	**W9**	12.28	6.30
B1-5	**W2**	38.59	10.40	**W7**	11.20	5.89	**W8**	30.53	18.06
B1-4	**W2**	20.02	19.76	**W4**	24.03	6.34	**W10**	50.12	10.97
B1-13	**W5**	9.89	6.67	**W7**	13.25	7.92	**W9**	10.14	6.39
B1-12	**W2**	8.65	4.91	**W4**	21.56	11.39	**W10**	8.36	5.18
B1-14	**W2**	73.10	30.88	**W8**	38.21	18.60	**W9**	71.31	28.07

Table 4.6: Average distance (Å) between the centre-of-mass of the residues and respective fluctuation in the sequences B1-3, B1-4 and B1-13.

Aptamer	**Residues**	Average	RMSD
B1-3	**D6 H8**	6.77	0.85
B1-4	**H8 W10**	7.22	1.25
	H9 T11	6.82	0.71
B1-13	**D6 H8**	7.17	0.61

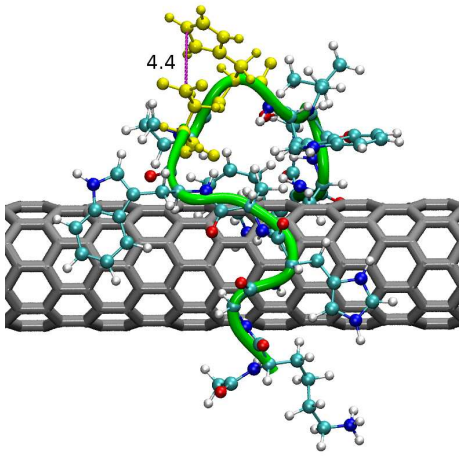

Figure 4.3: Snapshot taken from the best trajectory of the B1-3 peptide interacting with a CNT showing the residues D6 and H8 (in yellow) in close contact and the distance (4.4 Å) between two carbons in each of the residues. The backbone is highlighted.

surface, as evidenced by the broader peak in the distance distribution profiles in Figure 4.2, as a result of this interaction. Again, there might be a weak interaction between the residues D6 and H8 in the B1-13 sequence; interestingly, these two residues are in between two tryptophans. These observations suggest that aromatic residues separated by any other single residue might either give rise to intra-peptide residue interactions causing the residues to be distant from the CNT, or might cause the adjacent residues to be directed away from the CNT surface and therefore be further from the surface. Both situations give rise to a lower binding affinity due to the decreased number of atoms interacting with the nanotube surface and should be taken into account if we wish to design a peptide with controllable affinity to CNTs.

In the case of the peptide sequences containing a quartet (B1-2, B1-10 and B1-14), at least one tryptophan is always further from the nanotube on average, as evidenced by the broader peaks of the distribution profiles (Figure 4.2). However, in the

Table 4.7: Average and fluctuation of the distance (Å) between the aspartic acid (D) and lysine (K) side chain centre-of-mass to the CNT surface and the two adjacent residues (motif). Sequences are ordered from the strongest to the weakest binder.

Strong binders						
	Aspartic Acid			**Lysine**		
Aptamer	Average	RMSD	Motif	Average	RMSD	Motif
B1-6	5.66	0.82	**PDK**	4.33	0.40	**DKT**
B1 (original)	6.35	0.83	**WDT**	4.54	0.89	**WKH**
B1-5	6.09	0.55	**LDH**	4.19	0.36	**WKL**
B1-7	**8.96**	0.77	**PDH**	**5.48**	0.73	**WKW**
B1-9	8.18	0.98	**ADW**	4.28	0.56	**TKA**
B1-11	6.91	2.03	**ADL**	5.39	0.90	**WKG**
B1-1	7.87	1.20	**ADW**	3.94	0.36	**WKA**
Medium binders						
B1-10	8.80	1.14	**ADL**	5.11	0.90	**WK**
B1-8	**9.73**	0.31	**WDT**	4.01	0.27	**GKP**
B1-2	8.41	1.44	**ADT**	4.51	0.58	**HKW**
B1-3	7.17	0.32	**WDP**	4.73	1.17	**KG**
B1-15	6.59	2.00	**WDH**	7.87	2.09	**LKW**
B1-4	8.06	0.80	**GDA**	4.14	0.39	**WKW**
B1-13	**9.74**	0.41	**WDW**	4.59	0.54	**LK**
B1-12	7.51	1.98	**HDA**	4.45	0.48	**WKH**
B1-14	**9.34**	1.24	**WDG**	**8.50**	1.02	**LKP**

case of B1-14 there is no such clear distinction in the distances of these residues from the surface. But an examination of the ring-tilt of these tryptophan residues (Table 4.5) reveals that even though W8 and W9 are, on average, close to the CNT, they are oriented almost perpendicular to the surface, as opposed to the flat orientation that tryptophan typically adopts. In addition, the fluctuation of this average angle is also greater than usually found for tryptophan. Even though B1-14 is the worst binder, the aromatic residues are all, on average, considerably close to the CNT, which means that there must be another factor playing a major role in dictating weak binding affinity. From Figure 4.2 it is noticed that alanine (A1) and aspartic acid (D10) are further from the surface (beyond 8 Å). But alanine is a small residue and therefore the loss for not having this residue interacting with the CNT is not huge, and despite the aspartic acid being slightly bigger, it is noted that it is rarely found close to the CNT. An inspection of the other residues in the chain helps to understand the weak binding in B1-14 phenomenon. In Table 4.7 is the average and fluctuation of the distance between the centre-of-mass of the side-chains of aspartic acid (D) and lysine (K) respectively to the CNT surface, along with the motif in which these residues are incorporated, for the strong and medium binders. This table shows that not only D10 in B1-14 is very far from the CNT, as previously concluded, but also K4. The lysine side-chain is long (with four methylene spacers) and the consequence of not having this residue interacting with the CNT is a lowering in the binding affinity. In the case of B1-2, it is clear from the distribution profiles (Figure 4.2) that W4 and W5 are more distant from the CNT than usual (4.51 Å and 5.16 Å, respectively, instead of around 3.4 in B1-6, Table 4.5). In addition, from Table 4.7 it is clear that D10 is also very far from the CNT (8.41 Å). In B1-10, all tryptophan residues are reasonably close to the CNT, but there is a histidine (H4) (Figure 4.2) and the aspartic acid (D7) very far from the surface (Table 4.7).

Table 4.8: Average number of peptide atoms (peptide) and average number of backbone atoms (backbone) within 6.0 Å of the surface and respective root mean squared deviation (RMSD) and the average peptide end-to-end distance and RMSD for the strong and medium binders. Sequences are ordered from the highest to the lowest affinity.

Strong binders						
Sequence	Peptide	RMSD	Backbone	RMSD	End-end	RMSD
B1-6	**186.78**	**7.30**	**59.58**	**3.00**	30.97	2.84
B1 (original)[120]	**182.19**	**7.47**	**61.27**	**2.95**	27.32	1.39
B1-5	**189.12**	**7.73**	**61.03**	**2.95**	28.62	2.63
B1-7	174.73	5.52	59.58	3.00	24.16	5.31
B1-9	176.78	7.56	55.28	3.25	26.69	4.04
B1-11	176.11	7.82	54.62	4.47	17.95	4.78
B1-1	176.51	13.06	52.63	5.21	25.78	3.18
Medium binders						
Sequence	Peptide	RMSD	Backbone	RMSD	End-end	RMSD
B1-10	164.74	13.38	49.41	10.07	26.82	5.86
B1-8	171.04	5.08	42.77	3.10	30.12	1.80
B1-2	181.95	7.11	61.70	3.65	26.18	4.29
B1-3	163.31	8.37	52.19	4.27	25.09	3.50
B1-15	152.04	13.23	47.42	6.93	29.59	4.23
B1-4	179.22	6.09	60.55	2.29	25.27	2.00
B1-13	**166.98**	**7.82**	**46.84**	**3.96**	16.79	3.52
B1-12	**159.74**	**10.09**	**52.83**	**6.39**	14.44	2.42
B1-14	**156.50**	**15.95**	**45.46**	**8.04**	26.44	4.40

To further rationalise the observed differences in the binding affinity, the average number of peptide atoms and peptide backbone atoms within 6 Å to the CNT surface were calculated. Table 4.8 presentes the average and fluctuation of the number of peptide and peptide backbone atoms, respectively, for the strong and medium binders. These data show that the strong binder sequences have considerably more atoms (both peptide and peptide backbone) within 6 Å than the medium binder sequences. This difference is accentuated for the three strongest binders and the three worst binders (in bold in Table 4.8). A greater fluctuation is noticed in the case of the medium binders. This result helps to reinforce the previous conjecture that sequences with stronger affinity facilitate surface contact for more of the peptide atoms. Strong binders do not necessarily have all the aromatic residues flat on the surface; instead, it is the compromise between maintaining strong ring-surface interactions and also allowing the other residues to approach the surface that govern the peptide-CNT binding affinity, just like it was found for the peptide-graphite interface in Chapter 3. This idea will be more carefully discussed in the next Section. Furthermore, in Table 4.8 the average and fluctuation of the peptide end-to-end distances for both the strong and medium binders is shown. With the exception of the worst binder B1-14, the second and the third worst binders (in bold) support relatively smaller peptide end-to-end distances than in the case of the three best binders (in bold). It is understandable that an extended peptide may allow more atoms to interact with the CNT. However, it is suggested that this is not necessarily a requirement for strong binding affinity, since the CNT is curved and certain sequences may be able to wrap the CNT (as previously found), therefore supporting smaller peptide end-to-end distances.

In addition, the three closest peptide atoms (within the 8.0 Å cutoff) to the CNT for the three strongest binders (B1-6, B1-original and B1-5) and the three worst binders

(B1-13, B1-12 and B1-14) were determined. Table 4.9 summarises the three closest peptide atoms, the residue to which the atom belongs to, the type (backbone (B) or side-chain (S)) and the frequency. It is evident from Table 4.9 that in the case of the strongest binders all the closest atoms belong to the peptide backbone, while in the case of the B1-12 and B1-14 (weakest binders) one of the atoms belongs to the peptide side-chain. This suggests that strong binders have more peptide backbone atoms close to the surface, and therefore more atoms overall close to the CNT, supporting the previous results in Table 4.8, that strong binders have more peptide atoms (including backbone) interacting with the CNT, while in the case of the worst binders the peptide backbone cannot approach the surface so well. In addition, in the case of the three best binders, many of the closest atoms belong to aromatic residues (W and H), while in the case of the three worst binders the closest residues belong to other residues such as proline (P5) in the case of B1-14, glycine (G8) in the case of B1-12 and alanine (A4) in the case of B1-13.

4.5 Discussion

The data presented in this Chapter suggest that the peptide sequence is crucial in determining the strength of the binding affinity to CNTs. This is because the order of the residues in the chain, that is the sequence, determines the peptide conformation which in turn can play a large role in facilitating binding to the surface. Hence, it is proposed that not only the aromatic content (W and H) is essential for the strong affinity to CNTs, but the local environment in which each of the aromatic residues is incorporated may also play an important role. Furthermore, these findings form a basis for further studies with the aim to identify dependencies and establish principles that may help in the design of peptide sequences with controllable binding affinity to CNTs.

Specifically, the weaker affinity of the medium binder sequences identified in this work, is the result of an interplay between several motifs. The next step will be to systematically disentangle these dependencies between the motifs.

The number of different permutations of 12 residues, where there are three indistinguishable tryptophan residues, and two indistinguishable histidine residues is given by:

$$\frac{12!}{3!2!} = 11! = 39,916,800. \tag{4.1}$$

Even though the set of sequences studied in this work constitute only a small sample, the data give a glimpse of how the order of the content may change the binding affinity of peptides with respect to CNTs. The probability for each residue to be at a certain position in the chain is not the same for all the residues, since there are three tryptophans and two histidines. For all the residues other than tryptophan and histidine, the probability of being in a certain position in the chain is 1/12, however in the case of tryptophan and histidine is 3/12 and 2/12, respectively. Even though, it is well known that the aromatic residues bind well to CNTs, we should bear in mind that, in this case, tryptophan and histidine are more likely to be present in the featured motifs (by virtue of their relative abundance). However, the probability of finding the motifs in the sequences can be calculated. For instance, the probability of finding just one HW, P(1HW) in the sequences is given by the probability of finding at least one HW, minus the probability of finding two HWs in the sequence (since in this case, it is possible to have two HWs, but not more in the sequence):

$$\begin{aligned} \mathrm{P(1HW)} &= (\frac{11!}{2!} - 9!)/11! && (4.2) \\ &= 0.49 && (4.3) \end{aligned}$$

The P(1HW) is the same as the probability of finding just one WH, P(1WH). This

probability tells us that there is almost a 50% chance of having one HW in any sequence with this particular content (both strong and medium binders); our range of sequences (Table 4.3) tallies with this. Similarly, the probability of finding just two tryptophans together (WW) is:

$$\begin{aligned} \mathrm{P(WW)} &= (\frac{11!}{2} - \mathrm{P(WWW)})/11! && (4.4) \\ &= (\frac{11!}{2} - \frac{10!}{2})/11! && (4.5) \\ &= 0.45 && (4.6) \end{aligned}$$

where P(WWW) is the probability of finding three adjacent tryptophans. Again, this value tells us that the probability of finding just one WW in the sequences is about 50%. However, this motif is underrepresented in the medium binders, while appearing in roughly of the strong binders (see Table 4.3). This uneven distribution of the presence of the WW motif suggests that WW may be a key feature in the strong binder sequences.

The data presented in this Chapter support the idea of previous studies that binding affinity is the result of a complex interplay between structure, sequence and binding [32]. In addition, in agreement with previous studies [184], these findings show that the content primarily determines the relative strength of peptide binding to the surface. However, the binding affinity is the product of not just the content but many other factors, such as sequence and structure [13], and the environment in which the residues directly bind to the surface also plays a major role [85; 182]. These data show that by only mixing the peptide content of B1, the binding affinity drops to a level comparable to the mutations of the tryptophan content by either tyrosine and phenylalanine. Based on the differences in binding affinity observed in this work, it would be interesting if these peptide sequences could be experimentally tested.

In agreement with our previous studies [118; 120], it was found that all sequences

were able to bind at the CNT surface in a wide variety of configurations. An example is illustrated in Figure 4.4, showing two different strong binding configurations taken from two different trajectories of the B1-6 peptide interacting with the CNT, emphasising the different strong binding conformations that the peptide backbone can adopt.

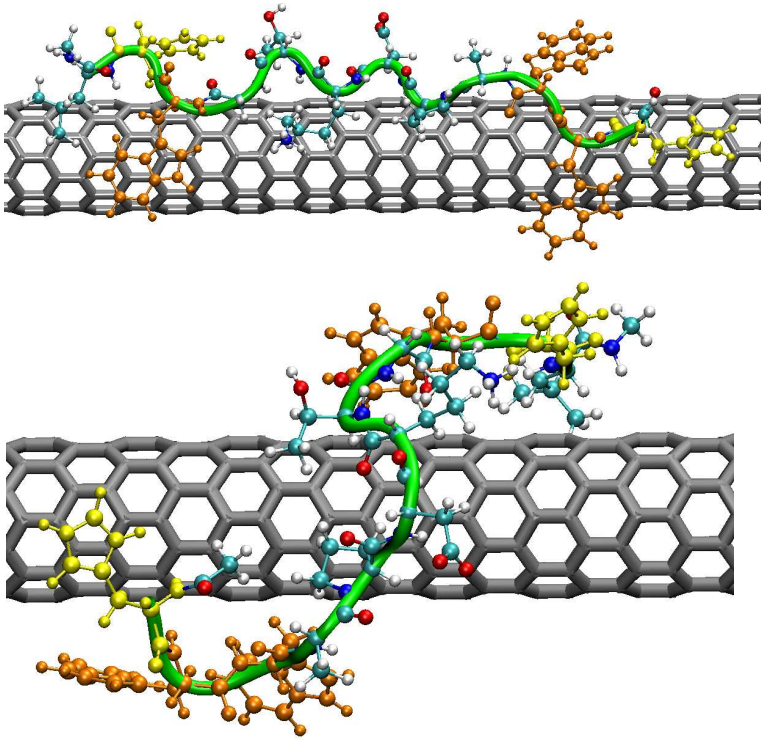

Figure 4.4: Two snapshots taken from two different strong binding trajectories (with different starting configurations) of the B1-6 peptide interacting with the CNT. The backbone is highlighted in green, the tryptophans in orange and the histidines in yellow.

As previously mentioned, these data suggest that there is a compromise between maintaining strong ring-CNT interactions and also allowing the other residues to approach the surface, as evidenced for instance by the small tryptophan-CNT distances in the case of B1-14 peptide (the worst binder) and the slightly greater distance of W8 in B1-5 (Table 4.2). This is similar to what was found in Chapter 3 for the case of the

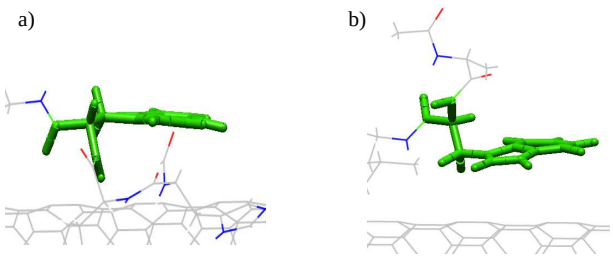

Figure 4.5: Schematic of the two different tryptophan configurations adopted in the peptide-CNT simulations. a) all side-chain atoms parallel to the surface and b) the tail of the side-chain which is connected to the backbone perpendicular to the surface.

peptide-graphite interface. However, in Chapter 3 we adjusted the interfacial shape to the sequence, while here, we adjusted the sequences while keeping the shape of the CNT fixed. Inspection of the trajectories revealed that the tryptophan side-chains can adopt two different configurations at the surface; one in which all the tryptophan side-chain (the indole ring and the methylene spacer) bind to the CNT, (referred to *parallel* herein), and another one in which only the indole is close to the CNT and the methylene spacer that connects the indole ring to the backbone is perpendicular to the CNT (referred to *perpendicular*). Figure 4.5 illustrates these configurations. It was found that for the best binders B1-6 and B1-5, two tryptophans adopted the parallel configuration and one adopted the perpendicular orientation during the simulations, while in the case of B1 one adopted the parallel orientation and one the perpendicular orientation, while the other one adopted a mixture of both. However, in the case of the three worst binders it was found that B1-14 (worst binder) supported the three tryptophans in a perpendicular orientation, while B1-12 and B1-13 supported two tryptophans in perpendicular and one in parallel. These findings may also help to rationalise why the worst binders have less atoms interacting with the surface.

Table 4.9: The three closest (on average) peptide atoms (within the 8.0 Å cutoff) during the production run, for the three best strong binder sequences and the three worst medium binders. *B* stands for peptide backbone atom and *S* stands for side-chain atom. % is the frequency at which the atom is the closest.

Top 3 strongest binders												
Pept.	Atom	Res.	Type	%	Atom	Res.	Type	%	Atom	Res.	Type	%
B1-6	H	**W10**	*B*	18.2	H	**W2**	*B*	15.2	H	**K7**	*B*	8.0
B1-origi.	H	**W9**	*B*	21.0	H	**W9**	*B*	9.4	H	**W2**	*B*	9.1
B1-5	H	**H6**	*B*	22.7	H	**W2**	*B*	7.8	H	**W7**	*B*	7.4
Top 3 weakest binders												
B1-13	H	**A4**	*B*	23.8	H	**H2**	*B*	11.6	H	**K12**	*B*	10.4
B1-12	H	**H12**	*S*	19.2	H	**W10**	*B*	16.1	H	**G8**	*B*	7.2
B1-14	H	**P5**	*B*	25.0	H	**H7**	*B*	19.6	H	**W9**	*S*	8.6

4.6 Summary and Outlook

The binding affinity between a CNT and a range of peptide sequences with the same content as the B1 peptide (but in a different order) were investigated by molecular dynamics simulations. Fifteen peptide sequences were created by mixing the amino acid content of the strong binder sequence B1 (`HWKHPWGAWDTL`), using a random number generator and by hand. Peptide sequences with different affinities to the CNT, including affinities higher than the original sequence B1 were found. While none of the scrambled peptide sequences diminished the binding affinity to a level comparable to the control peptide NB1, some peptide sequences reduced the binding affinity to a level comparable to the mutations of the tryptophan content by either tyrosine or phenylalanine (Chapter 3).

These findings suggest that the order of the amino acids, *i.e.* the sequence, plays an important role on the binding affinity to CNTs. In order to further the understanding of why certain peptide sequences have more binding affinity to the CNT than others,

it would be interesting to study these peptides in water and determining the flexibility of each of the sequences by calculating the peptide dihedral angles and determine how much the angles change during the production run. A knowledge of the peptide flexibility would give a further insight into the understanding of the peptide-CNT binding affinity. In addition, it would be interesting to isolate the main motifs identified in the strong and medium binders in order to establish the dependencies between the several motifs. In other words, it would be interesting to design some strong and weak binders and test them. The work presented in this Chapter represents a first step towards understanding the design rules for peptide-inorganic interfaces for both the chemistry and the shape of the interface. For instance, these findings suggest that strong binding may be due to the presence of two tryptophans together (WW), while weak binding may be due to two tryptophan residues intercalated by another residue (r) (WrW).

Chapter 5

Simulation of liquid water on graphitic surfaces using a polarisable force-field

5.1 Abstract

The interactions of water with graphitic surfaces and biomolecules are of considerable importance in the fields of bio- and nanotechnology [59]. Here, molecular dynamics (MD) simulations of the interface between water and graphitic surfaces are reported. An existing polarisable force-field that uses distributed multipoles up to and including quadrupoles for the electrostatics is extended to include a description of the non-bonded interactions between water and graphitic surfaces: carbon nanotubes, fullerenes and graphite. The spatial and orientational distributions of water around the graphitic surfaces are shown to reproduce the hydration behaviour very well, when compared with previous work.

5.2 Introduction

The structure of water at an interface is different from the bulk phases and may play a key role in biological systems [185; 186]. This deviation of the water properties at the liquid-solid interface gives rise to multiple interesting behaviours. A deep understanding of such behaviours would give insights into the potential applications of these materials.

The discovery of CNTs and fullerenes has led to a new era of nanoscience and nanotechology [187; 188]. Due to their intriguing physical and chemical properties [189; 190] these systems have been the subject of many experimental and theoretical studies. The outstanding potential applications in the fields of biology, medicine [61] and materials science [191] require an understanding of the behaviour of nanotubes in aqueous solution. Many theoretical studies have addressed the interface between water and graphitic surfaces (graphite [192; 175; 193; 194; 195; 196; 197; 198; 199; 200; 201], nanotubes [202; 173; 203; 204] and fullerenes [205; 206; 207; 208]). In addition, several authours have studied the adsorption of water clusters on bare [209] and functionalised graphitic surfaces [210; 211; 212; 213; 214; 215; 216]. A review of these studies of the interaction of water clusters with functionalised graphitic surfaces will be given in Chapter 6.

The use of quantum chemistry calculations to study the interface between water and graphitic surfaces is to date not practical, since for large systems these calculations can be prohibitively expensive. On the other hand, MD simulations have given further insight, and a wide range of intermolecular potentials for water and water-carbon have been proposed to describe these interactions [217; 194; 198]. However, an issue common to all of these potentials is their transferability between different systems. For example, the molecular water-carbon potential used in Refs [173; 218] and [219], is based on interaction parameters from the Bojan-Steele potential [220] who derived these param-

eters from the adsorption of molecular oxygen on graphite, suggesting these authors assumed that the oxygen atoms in water are similar to those in molecular oxygen. The most popular model to describe water-carbon interactions is the simple point-charge (SPC) (on water only) and Lennard-Jones (LJ) potential [217]. Similar models with an additional quadrupole-quadrupole interaction between carbon atoms and water have been proposed [173; 200; 194; 198]. Several studies have been done to optimise the parameters for this interaction potential [194; 204] compared with experiment. A more sophisticated model potential has been reported that includes point-charges (on water), LJ interaction and tube polarisation [203].

Graphite is a hydrophobic surface, and CNTs and fullerenes also exhibit low solubility in water [221; 222; 223; 224; 225]. Grunze and Pertsin studied the behaviour of the graphite interface in water [199; 226; 200]. They used grand canonical Monte Carlo (GCMC) simulations to test three water-graphite model potentials to describe the behaviour of water confined between two parallel graphite sheets [199] by reproducing and comparing their results with the MP2 calculations of Feller and Jordan [227], later revised by Karapetian and Jordan [228]. In this work of Karapetian and Jordan [228] it was concluded that the binding energy of a water molecule on a single graphite layer estimated in a previous work [227], was approximately twice as large in magnitude as the one obtained from their model potential calculations. Reasons for this discrepancy were pointed out as being the use of a very large correction for basis set superposition error (BSSE) [229]. Pertsin and Grunze [199] concluded that the thermodynamics and structure of water at the graphite interface was extremely sensitive to the range and orientation dependence of the model potential. These conclusions [199] were then demonstrated to be valid for weaker water-graphite interactions by exploring the phase and shear behaviour of water confined in graphitic slit pores of different widths with two

orientation-dependent model potentials [200]. Gordillo *et al.* reported a MD simulation of water confined between two graphite sheets with separations ranging from 7 to 15 Å [193]. They found two orientational distributions when water molecules are close to the walls and the graphite sheets separated by 15 Å: one with a dangling bond pointing towards the surface and other with both OH bonds parallel to it. However, their studies have exclusively made use of a simple point charge (SPC) potential to describe the water-water interactions and a Lennard-Jones potential to describe the water-graphite interactions using the usual Lorentz-Berthelot mixing rules. The dependence of the contact angle of water droplets on graphite with the water-carbon potential was studied by Werder *et al.* [194] and were compared with experimental values. Unfortunately, the reported experimental contact angles span a range of values from 30° to 86°, and therefore it is not sensible to compare the contact angles obtained from MD simulations with the experimental data. Several interaction potentials from the literature were used to model graphite-water systems, where all were based on a pairwise additive LJ potential between the oxygen atoms of the water and the carbon atoms. These data suggested a linear relationship between the water contact angle and the water monomer binding energy on graphite. The authors adjusted the interaction parameters for the LJ potential acting between the oxygen and carbon atoms such that the simulations recovered a contact angle of 86°. The rigid SPC/E model (which only includes a carbon-oxygen LJ potential) was used because it has the lowest bulk energy and the contact angles reported thus represented an upper bound to those that would be obtained with other water models with weaker water-water interactions. This work was extended by Jaffe and Werder [204] and new parameters for the water-carbon potential were presented in order to reproduce the experimental macroscopic contact angle of 84-86°. This study included the variation of C-O LJ well depth with the contact angle and the effects of

adding pairwise water hydrogen-carbon LJ terms to the potential and also the effects of changing the long-range cut-off used in the truncation of the interactions. They also presented a continuum model of the graphite in which the interaction energy depends only on the surface-oxygen normal, reducing the computational time needed to calculate the water surface interaction. Very recently, several works have reported use of electronic structure theory to get better estimates of the strength of interaction between a single water molecule and a graphene sheet [201; 197; 195]. Previous studies have shown that the interaction energy between water and graphite is well described by using a single graphite sheet as opposed to a many-layered model of graphite [227; 228; 199; 200]. Huff *et al.* [201] calculated the potential energy surface of a single water molecule over a coronene ($C_{24}H_{12}$) molecule using DFT and MP2 calculations. These authors found that both methods predicted a minimum energy structure for water with the hydrogens facing the ring. In agreement with these findings is the work of Rubes *et al.* [197] who used the density-functional/coupled-cluster (DFT/CC) level of theory to calculate the interaction of water with graphene, also obtaining a water minimum structure with both hydrogen atoms of water pointing toward the surface. Jenness and Jordan [195] used density fitting-density functional theory-symmetry-adapted perturbation theory (DF-DFT-SAPT) to calculate the individual contributions of the interaction between a water molecule with various acenes and fitted the parameters in six water-graphite potentials, including the AMOEBA potential. However, the authors did not mention which atomic polarisabilities were used.

Various computational studies have also been performed to investigate the properties of fullerenes with water. Bedrov *et al.* studied the solvation of a single C_{60} fullerene molecule as well as pairs of C_{60} fullerenes [205] by using a LJ potential between carbon atoms and water and the TIP4P potential [230] to model water. These

authors extended their study to SWCNTs and graphene sheets [206] and found a correlation between water density and the extent of hydrogen bonding at the interface, with the degree of the surface curvature. A reduction in the water density and extent of hydrogen bonding was noticed upon increasing the radius of curvature of the nanotube. Choudhury [208] proposed a coarse-grained model to study the solvation behaviour of C_{60} in water, and compared this model with atomistic MD simulations. Their atomistic data show that the water near the fullerene surface is densely packed relative to the bulk and a disruption of the hydrogen bond network occurs in the regions near the fullerene. OH bond orientation distributions show that water molecules show preferential orientation with one OH bond pointing away from the surface while the other OH bond is slightly inclined toward the surface. Distributions of the orientation of the plane of the water molecules show that near the fullerene walls, the waters are mostly oriented tangential to the fullerene surface [208].

In the case of the nanotubes, Walther *et al.* [173] studied the structural properties of water surrounding a (16,0) carbon nanotube by performing fully atomistic MD simulations in water, with the objective of testing the influence of solvation on the vibrational modes of the nanotube. In this study the CNT was described as fully flexible, and the water by the flexible SPC model of Teleman *et al.* [231] with partial charges. The nonbonded interactions between the water molecules comprised an O-O LJ term and the Coulomb potential. The carbon-water interaction was described with a C-O LJ interaction, plus an interaction between quadrupoles on the carbon atoms and the partial charges on the water atoms. Their results indicate that the presence of bulk water has a negligible influence on the breathing frequency mode of the CNT. The dominant orientation of the water molecules at the carbon-water interface was found to have the HOH plane nearly tangential to the interface as for the graphite-water system

[209], with a characteristic layering observed from the water radial density profiles. A different model was proposed by Zhao and Johnson [198] who derived a potential that incorporates polarisability and atomic quadrupoles on graphite and atomic quadrupoles on water. These authors found that for weakly polar fluids the polarisation contributions were negligible compared with the LJ interactions. However, for strongly polar fluids, such as water, the contribution from the polar terms to the total potential energy was found to be significant. The first MD simulation using a polarisable nanotube immersed in liquid water was done by Moulin *et al.* [203], where they used a LJ term between the carbon and oxygen and an electrostatic term between charges on water and the induced dipoles on the polarisable nanotube, with the TIP4P potential [230] for water. Despite the fact that CNTs have a very large polarisability, they concluded that the influence of the nanotube polarisation was negligible on the arrangement of the water molecules around the nanotube and on their adsorption energy. However, it is not surprising since water is also polarisable and this contribution was neglected. Our work makes use for the first time of a polarisable CNT and polarisable water by using an extended version of the AMOEBAPRO force-field which has been modified to treat the interactions between water and graphitic surfaces.

The aim of this study is to demonstrate that the polarisable force-field is suitable for use in further simulations that include graphitic surfaces, biomolecules and water. To this end, we have run MD simulations of these surfaces in water, using our novel potential, and have analysed the interfacial structure, orientational order, water residence times in several regions of the interface and the average number of hydrogen bonds. A clear understanding of the role that water plays in these systems is vital for understanding future studies of mechanisms of biomolecules interacting with the aqueous interface of graphitic surfaces. In addition, MD simulations of a CNT with one OH

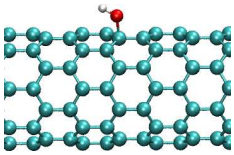

Figure 5.1: CNT with an OH defect attached to exterior wall.

defect attached to the exterior wall of the CNT were performed in order to investigate the structural properties of the water molecules surrounding the OH defect. This study aims to complement the studies of the effect of the presence of defects on the binding affinity to peptides presented in the next Chapter (Chapter 6). Furthermore, this is the first simulation study of liquid water and a defective CNT. A CNT with one defect is shown in Figure 5.1.

5.3 Methods

MD simulations of the graphitic surfaces in water were carried out using the Tinker package [171]. In the case of the CNT, an (8,0) nanotube of 96 carbon atoms was held rigid along the y-axis in a periodic cell of initial dimensions 26.65× 12.79× 26.65 Å with 279 water molecules. In the case of the fullerene, the C—C bonds were kept rigid, and the fullerene was placed in the centre of a box of initial dimensions 28.00× 28.00× 28.00 Å with 676 water molecules. For the graphite, a cell of initial dimensions 12.782×14.76×47.0 Å was used, incorporating three graphene layers, each with 72 carbon atoms, and 210 water molecules.

The force-field developed in this work is an extension of the AMOEBAPRO force-field [142; 143; 144]. The AMOEBAPRO force-field includes an atomistic model to describe the water molecules. Our extension includes parameters for the non-bonded

interaction between graphitic surfaces and water at the same level of approximation. The atom-centred distributed multipoles up to and including the quadrupole were calculated by T. R. Walsh with the GDMA program [232] with input wavefunctions calculated using GAUSSIAN03 [233] at the HF/6-31G* level of theory. Atomic polarisabilities for the graphite were taken from Zhao and Johnson [198].

Newton's equations of motion were solved using the Verlet algorithm with an integration step of 1.0 fs. In the case of the fullerene a cutoff of 8 Å applied to all non-bonded interactions while in the case of the CNT and and graphite, the minimum image convention was applied (no spherical cutoff applied). The Ewald summation was used to handle the long-range electrostatic interactions. Coordinates were saved every 0.5 ps (500 steps). The first 1 ns of each simulation was used as an equilibration period and removed before subsequent analysis. The simulations were carried out in the isothermal-isobaric (NPT) ensemble with periodic boundary conditions applied in all directions. The temperature of the system was coupled to 298 K by means of the Nosé -Hoover thermostat. All simulations were carried out at a target pressure of 1 atm. Anisotropic constant P simulations were used in the case of the CNT and graphite systems. In these cases, the barostat was only applied in the dimensions not spanned by C atoms (*i.e.* the barostat was only applied in the x and z direction in the case of the CNT and only in the z direction in the case of graphite). Isotropic constant P was applied to the fullerene system. To calculate water residence times in the first layer, each simulation was extended by 300 ps and the coordinates were saved more frequently (every 10 fs).

To analyse the results structural metrics were used. The oxygen and hydrogen density profiles as a function of the distance from the surface were obtained by dividing the simulation cell into slices (cylindrical, spherical and rectangular box, respectively for

the CNT, fullerene and graphite), and by counting the number of oxygen and hydrogen atoms (respectively) in each slice and averaging over time, and by normalising with respect to the volume of the slice. In the Results Section it is described how the density profile is used to divide the box into layers corresponding to the arrangement of water relative to the surface. The orientational probability distributution of water within a given layer, was determined by calculating the cos(θ) for all water molecules within that layer, where θ is the angle between either the dipole vector, O—H bond vector or plane normal vector of water and the CNT/fullerene/graphite normal vector, and by counting the frequency of the cos(θ) in that layer.

5.4 Results

The aim of this work is to study the structural properties of water surrounding three graphitic surfaces (CNT, fullerene and graphite) and whether or not the curvature of the surfaces can influence these properties. The behaviour of water molecules around the surfaces was studied by considering both the spatial and orientational distributions of water. Oxygen and hydrogen density profiles around the CNT, fullerene and graphite, which show the density of water as a function of the distance to each surface, are shown in Figure 5.2, while the exact values appear on Table 5.1. In all cases the distance is measured from the surface to the water oxygen atoms. The density profiles show that the water near the CNT/fullerene/graphite surfaces is densely packed relative to the bulk. The almost coincident first peaks of oxygen and hydrogen distributions indicate that the plane of the water molecule is almost tangential to the surfaces. However, hydrogen is slightly closer to the surfaces than oxygen, which is probably due to a smaller van der Waals radius of hydrogen.

To gain insight into the alignment of water molecules on the surfaces and

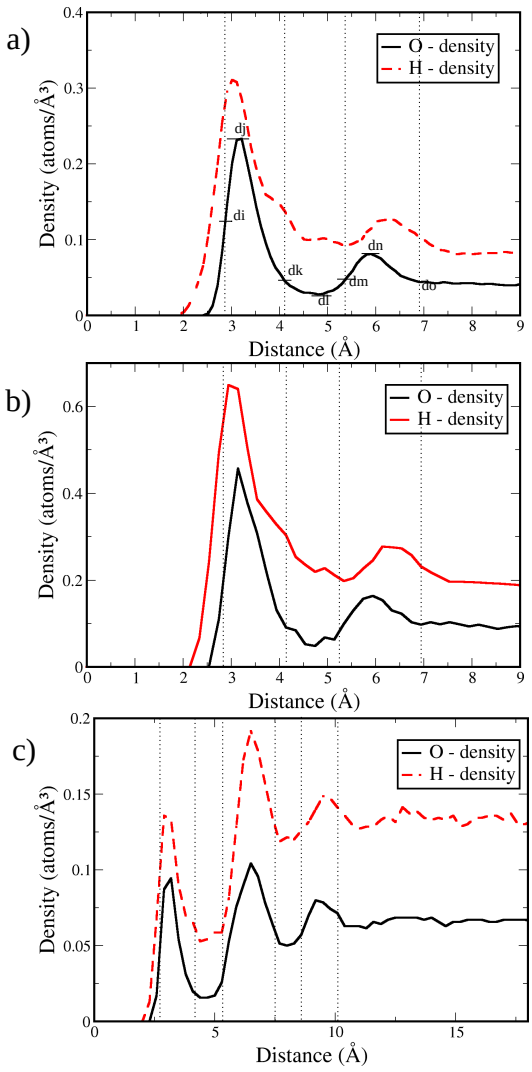

Figure 5.2: Atom density of oxygen and hydrogen around (a) carbon nanotube, (b) fullerene and (c) graphite as measured from each surface. Dotted lines represent layer boundaries.

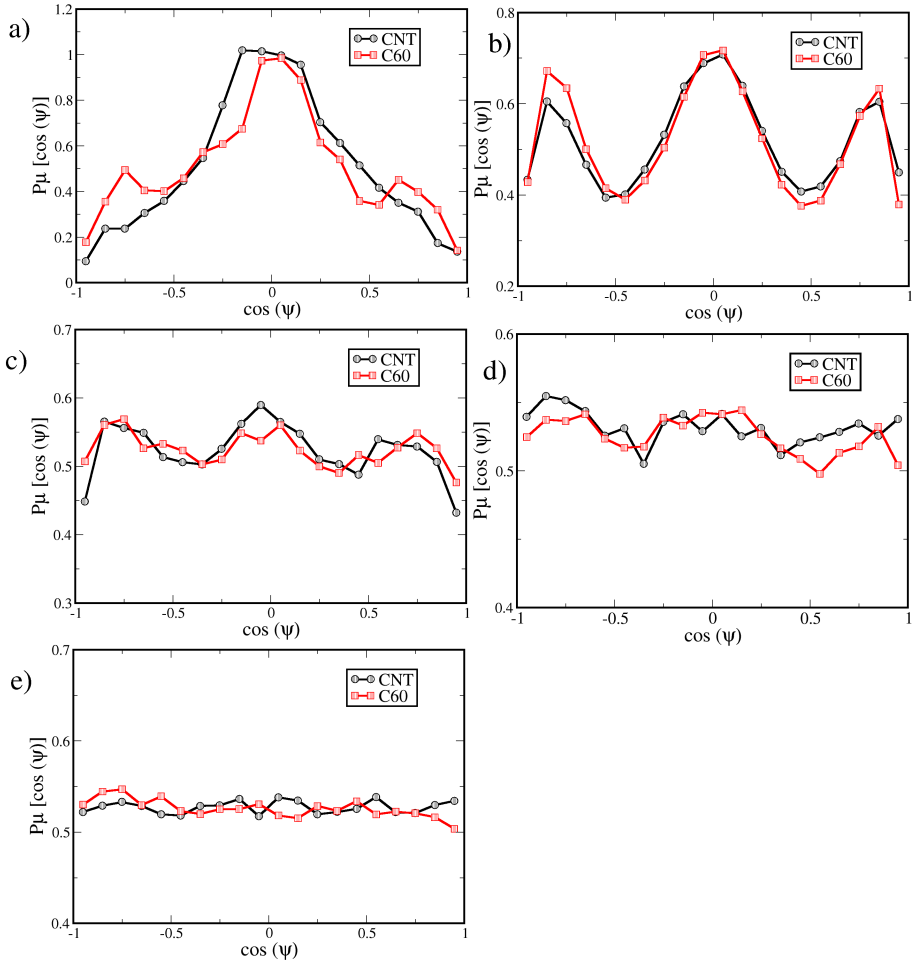

Figure 5.3: Dipole moment distribution $P_\mu(cos\psi)$ vs $cos(\psi)$, where ψ is the angle between the dipole moment vector and the CNT/fullerene normal vector. Plots a) to e) corresponds to layers 1 to 5.

Table 5.1: Summary of the peak positions (r) and heights [$\rho(r)$] of the oxygen and hydrogen density profiles for the three surfaces. Peak positions (r) are in Å.

CNT						
	1st Peak		1st Trough		2nd Peak	
	r	*ρ(r)*	*r*	*ρ(r)*	*r*	*ρ(r)*
O	3.16	0.23	4.76	0.028	5.76	0.084
H	2.96	0.31	5.36	0.092	6.26	0.130
Fullerene						
O	3.14	0.22	4.74	0.023	5.94	0.079
H	2.94	0.31	5.34	0.0953	6.14	0.133
Graphite						
	1st Peak		1st Trough		2nd Peak	
	r	*d(r)*	*r*	*d(r)*	*r*	*d(r)*
O	3.2	0.098	4.4	0.0154	6.5	0.0947
H	2.9	0.126	4.4	0.0530	6.5	0.180
Graphite						
	2nd Trough		3rd Peak			
	d(r)	*d(r)*	*r*	*d(r)*		
O	7.7	0.0536	9.5	0.078		
H	8.0	0.12	9.5	0.144		

compare with other simulation data, three orientational probability distributions of the water molecules — a distribution of the dipole moment $P_{\mu}[cos(\psi)]$, OH bond vector $P_{OH}[cos(\theta)]$, and the plane-normal vector $P_{p}[cos(\phi)]$ of the water molecule were calculated. The angles ψ, θ, and ϕ are the angles made, respectively, by the dipole moment vector, the OH bond vector, and the plane-normal vector with the CNT/fullerene/graphite normal vector. To investigate the orientation of the water molecules near the surface as well as away from it the space around the CNT/fullerene/graphite was divided into cylindrical/spherical/rectangular box layers respectively, as summarised in Tables 5.2 and 5.3. The division into layers has been considered previously [234; 235]. In the calculation of the different orientational distribution functions, we considered only the water molecules that are within a specified shell. It will be referred in the text to

Table 5.2: Summary of cylindrical/spherical shell boundaries considered for the calculation of the orientational distributions around the CNT and fullerene, respectively.

	CNT		**Fullerene**		
Layer No.	r_1(Å)	r_2(Å)	r_1(Å)	r_2(Å)	Remarks
1	0.00	2.86	0.00	2.83	Repulsive zone around the solute
2	2.86	4.10	2.83	4.14	Around the first peak in the O density profile
3	4.10	5.36	4.14	5.24	Between the first and the second peak in the O density profile
4	5.36	6.90	5.24	6.94	Around the second peak in the O density profile
5	6.90	10.0	6.94	9.00	Beyond the second peak

layers 1 and 2 as the interfacial region and layers 3 and 4 as the intermediate region, for all cases. The systems were partitioned into layers, by dividing the region of the oxygen density profile of the CNT interface into layers, based on the previous work of Choudhury [208]. In order to be consistent with the CNT layer division, a rule was applied when choosing the layers for the fullerene and graphite interfaces. For each CNT layer, a ratio (d_i/d_j, d_k/d_j etc.) between the optima of the CNT O density profile in that layer and the density at the respective layer boundary was calculated (See Figure 5.2). The calculated ratios were then applied to the fullerene and graphite oxygen density profiles to determine the boundaries of the layers. As an example, for the CNT in layer 1 the ratio between the density of the first boundary and the density of the first peak (d_i/d_j) was calculated, and then used to determine the first boundary (d_i) in the fullerene and graphite systems; to determine the position in the second boundary the ratio d_k/d_j was calculated for the CNT system and then this value was used to determine the second boundary (d_k) in the fullerene and graphite systems.

The results for the $P_\mu[cos(\psi])$ of the water in the five different layers for the

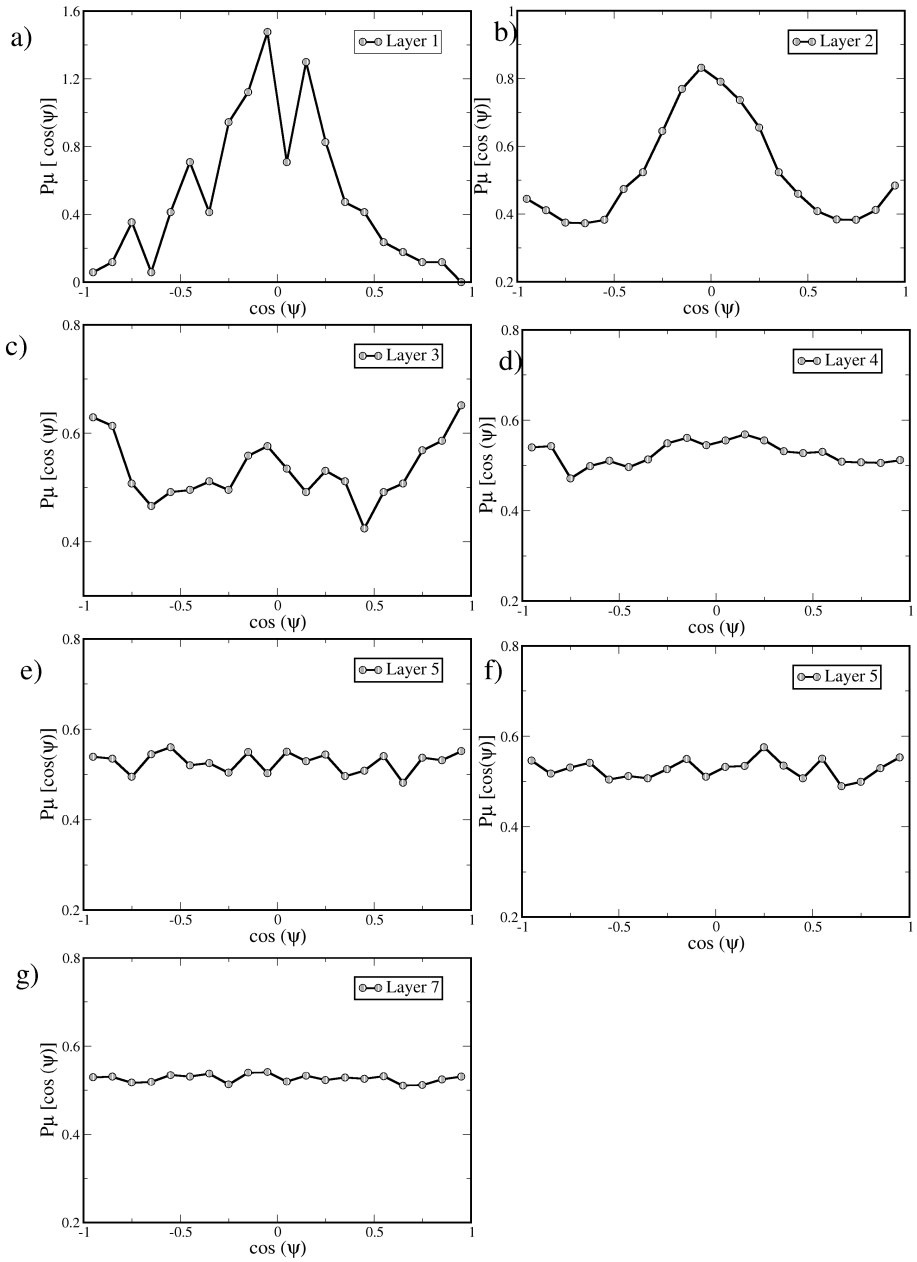

Figure 5.4: Dipole moment distribution $P_\mu[(cos(\psi)]$ vs $cos(\psi)$, where ψ is the angle between the dipole moment vector and the graphite surface normal vector. Plots a) to g) corresponds to layers 1 to 7.

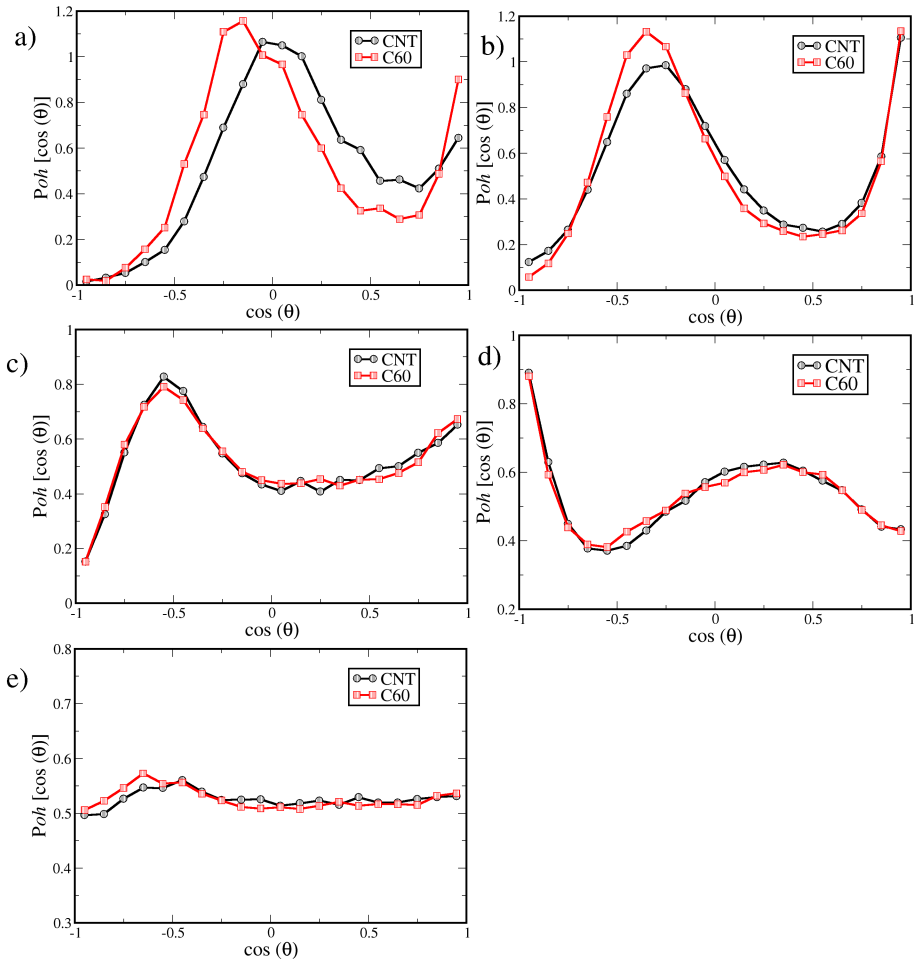

Figure 5.5: OH bond distribution $P_{OH}(cos\theta)$ vs $cos(\theta)$, where θ is the angle between the OH bond vector and the CNT/fullerene normal vector. Plots a) to e) corresponds to layers 1 to 5.

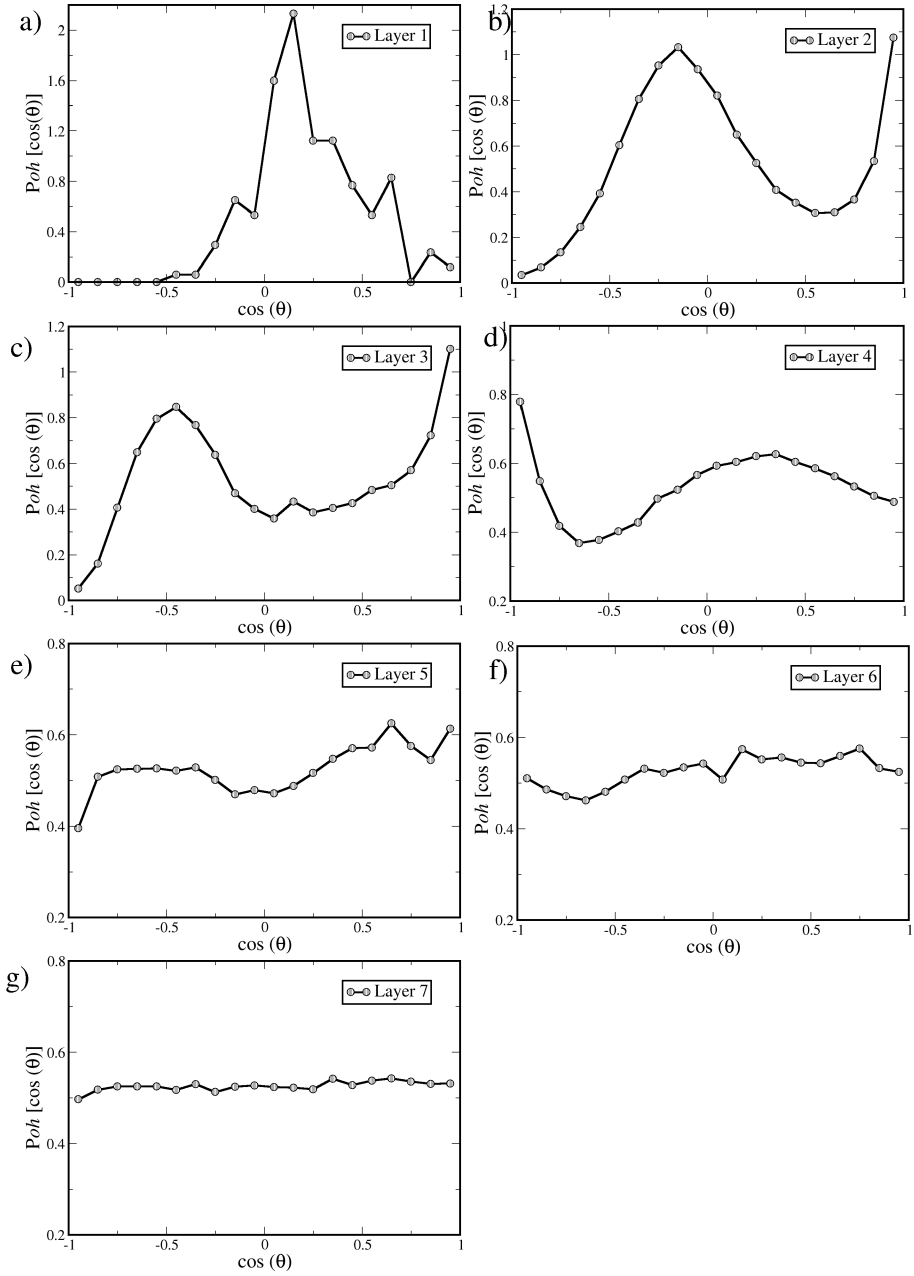

Figure 5.6: OH bond distribution $P_{OH}[cos(\theta)]$ vs $cos(\theta)$, where θ is the angle between the OH bond vector and the graphite surface normal vector. Plots a) to g) corresponds to layers 1 to 7.

Table 5.3: Summary of layer boundaries considered for the calculation of the orientational distributions for the graphite sheet.

Layer No.	r_1(Å)	r_2(Å)	Remarks
1	0.00	2.72	Repulsive zone around the solute
2	2.72	4.18	Around the first peak in the O density profile
3	4.18	5.32	Between the first and the second peak in the O density profile
4	5.32	7.50	Around the second peak in the O density profile
5	7.50	8.57	Between the second and the third peak in the O density profile
6	8.57	10.10	Around the third peak
7	10.10	20.00	Beyond the third peak

CNT and fullerene are shown in Figure 5.3, while the results for the dipole distribution of graphite around the seven layers are shown in Figure 5.4. Near the hydrophobic surface (layer 1) the three surfaces present a peak in the distribution around $\psi = 90^\circ$ suggesting that dipole moments show preference to be tangential to the solute surface. These results are in agreement with recent studies of the orientation of water molecules on CNTs [173], fullerenes [208] and graphite [196]. Similar to layer 1, the dipole orientation of water molecules around the first peak (layer 2) has, for all surfaces, a broad distribution of dipoles at $\phi = 90^\circ$. However, in the case of the nanotube and the fullerene a large fraction of water molecules also orient their dipoles around $\phi = 0^\circ$ and $\phi = 180^\circ$. This is not the case for the graphite surface. The distribution in layer 3 for the CNT and fullerene, (Figure 5.3) is similar to the distribution in layer 2 but is flatter. In the case of graphite, layer 3 presents a small bump at around $\phi = 90^\circ$ as in the previous layers and two peaks at 0° and $180\ ^\circ$, suggesting that some of the water molecules orient their dipole moments perpendicular to the graphite sheet and other water molecules orient their dipoles tangential to the surface. The distribution around

the second peak (layer 4) for the CNT and fullerene is approaching bulklike behaviour while in layer 5 it is more-or-less homogeneous, indicating a bulk phase. In the graphite system, the dipole orientation in layer 4 exhibits a small broad bump between 80° and 100°. This suggests that in this layer dipoles are slightly oscillating from the tangential orientation to the surface. In layers 5 and 6 there is still a weak structuring, but in layer 7 it reaches the bulk phase. These $P_{\mu}[cos(\psi)]$ distributions are comparable to the dipole distributions of water near planar graphene plates [234].

In Figure 5.5 we show the $P_{OH}[cos(\theta)]$ distributions in the same five layers for the CNT and fullerene and Figure 5.6 shows the $P_{OH}[cos(\theta)]$ distributions for the graphite. The distribution in layer 1 for the CNT and fullerene [Figures 5.5(a) and 5.5(b) respectively], show two large peaks at around $\theta = 0°$ and 90° - 100°. The greatest peak (at 90° - 100°) suggests that the majority of the water molecules has one of the OH bonds tangential to the surface. Due to the HOH angle constraint, it can be inferred that the other OH bond should also be tangential to the surface. The smaller peak at $\theta = 0°$ indicates there are a small number of molecules which have one of the OH bonds pointing away from the surface. In the case of the graphite system, there is no such peak at $\theta = 0°$, which suggest that practically all the water molecules have one of the OH bonds flat on the surface. This difference in the OH orientation in layer 1 between the graphite and the CNT and fullerene might be attributed to differences in the surface curvature. However, the $P_{OH}[cos(\theta)]$ in layer 2 is similar for all the surfaces. For the three surfaces there are two large peaks at $\theta = 0°$ and 90° - 100°. Again, these data suggest that some water molecules have one of the OH bonds tangential to the surface and other water molecules have one of the OH bonds pointing away from the surface. The other OH bond is once again restricted to the HOH angle constraint. The distribution in layer 3 as shown in Figure 5.5(c) and 5.6(c), which corresponds to the

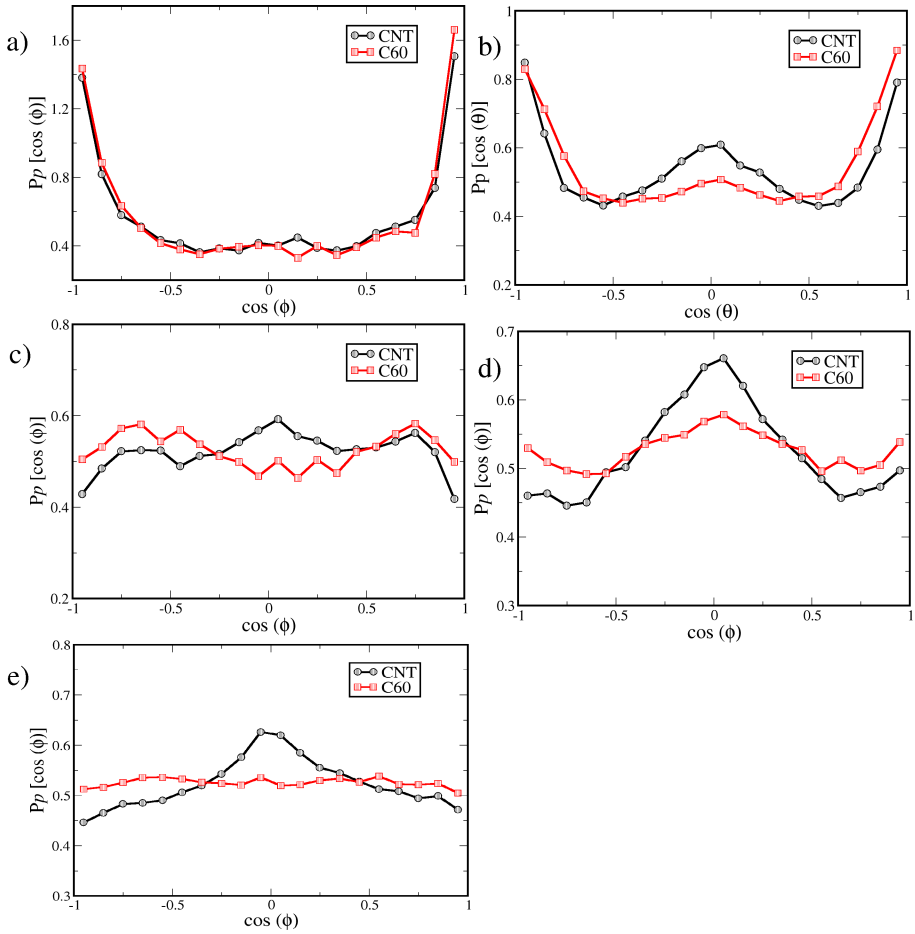

Figure 5.7: Distribution of the orientation of the plane of the water molecules $[P_P(cos\phi)]$ vs $cos\phi$, where ϕ is the angle between the normal vector to the plane of the water molecule and the vector drawn from the centre of the CNT/fullerene to the oxygen atom of the water. Plots a) to e) corresponds to layers 1 to 5.

region between the first and the second peak of the oxygen density profile has, for all surfaces, a peak at around θ = 110° - 120° and another one at 0°. The distribution around the second peak (layer 4) for all the surfaces, has a pronounced peak at around 180° and a shallower one at around 60°-70°, while in the case of the CNT and fullerene, in layer 5 water molecules are as random as in the bulk, and in the case of graphite, water reaches the bulk phase in layer 7 [Figure 5.6(g)]. Clearly, the OH orientation persists further (to layer 4) than the dipole orientation (layer 3 at best). These findings are in good agreement with the $P_{OH}[cos(\theta)]$ distributions for fullerenes [208], CNTs [173] and for graphite [196]. A small bump in layer 5 is observed in the case of the CNT and fullerene which should be further investigated.

Finally in Figures 5.7 and 5.8 we show the $P_p[cos(\phi)]$ for the same layers for the three surfaces. For all three surfaces, there are two large peaks at 0° and 180° in layers 1 and 2, indicating that water molecules are mostly oriented with their planes tangential to the surface [Figures 5.7(a) and (b) and 5.8(a) and (b)]. In the third layer, water molecules at the CNT interface are oriented with their plane normals slightly perpendicular to the surface, while in the case of the fullerene, water molecules are oriented with their plane normals slightly tilted at around θ=35°-40°. Layers 4 and 5 have uniform distributions with values around 0.5 that are characteristic of the bulk liquid. However, in the case of the CNT, in layer 5 there is a small peak at around θ=90°. A couple of tests were done in order to investigate possible origins for this small bump. First, the boundaries of layer 5 were modified by decreasing and increasing the layer limits to understand if it was a problem related with boundary delimitation. Second, the $P_p[cos(\phi)]$ was calculated in the last 0.5 ns of the trajectory to confirm that it was not an equilibration issue. Failing that, one possible cause for water not behaving as bulk in layer 5 might be the size of the box being too small. In the case of

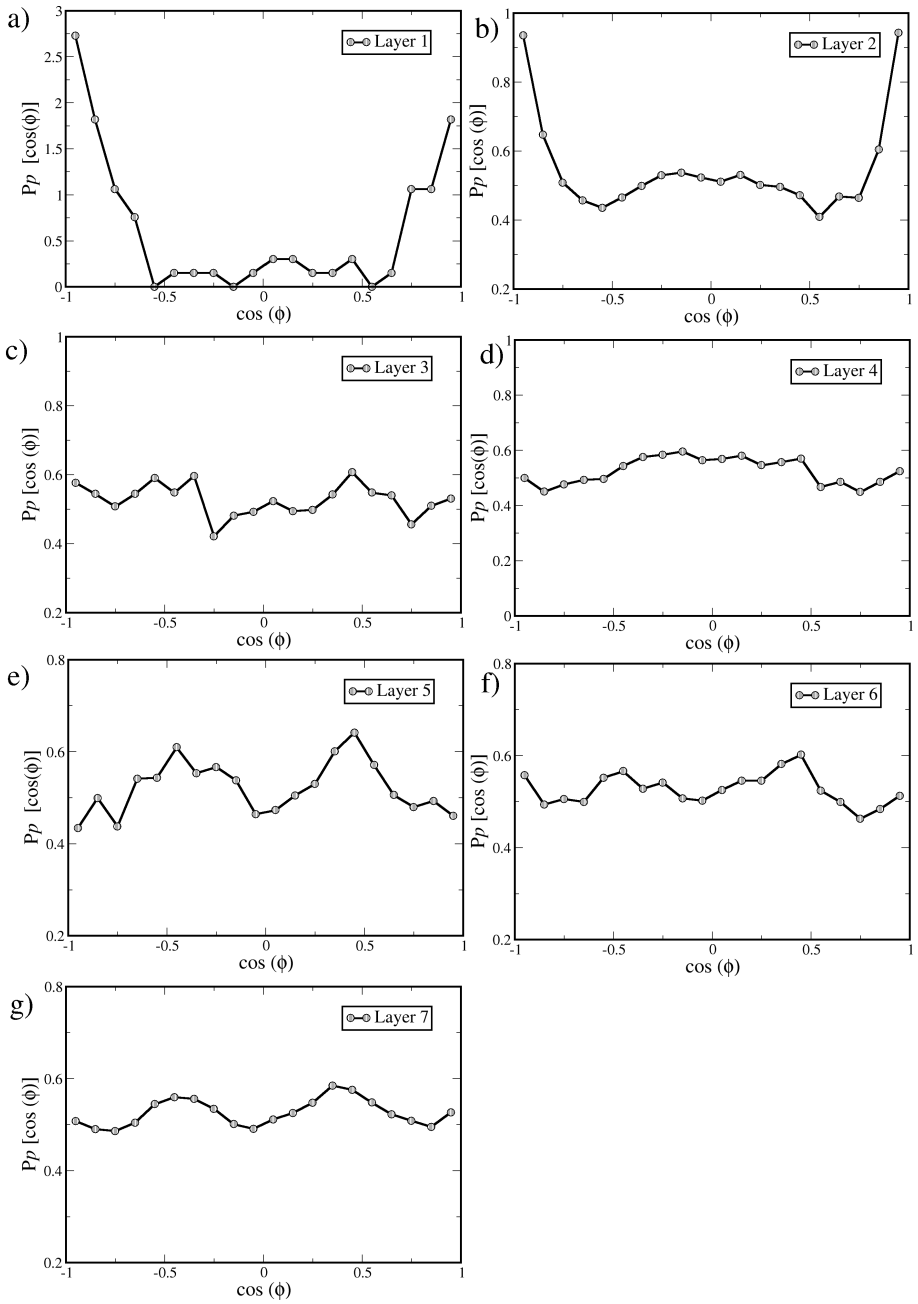

Figure 5.8: Distribution of the orientation of the plane of the water molecules $[P_P[cos(\phi)]$ vs $cos(\phi)$, where ψ is the angle between the normal vector and the graphite surface normal vector. Plots a) to g) corresponds to layer 1 to 7.

the graphite, planes of water molecules behave as random as in bulk after layer 5.

These data suggest that for all surfaces, in layer 1 the majority of the water molecules lie flat on the surface with the dipole moment and the plane of the water molecules tangential to the surface. The OH bonds are also tangential to the surface. However, in the case of the CNT and the fullerene, a small fraction of water molecules also orient one of the OH bonds away from the surface. In fact, this is in agreement with the $P_p[cos(\phi)]$, where a non-zero probability of finding a water molecule with its plane oriented perpendicular to the surface is presented [Figure 5.7(a)]. On the other hand, in the case of the graphite, the probability of finding a water molecule with its plane perpendicular to the surface is very small, which correlates well with the non existing peak at 0° in the $P_{OH}[cos(\theta)]$ distribution [Figure 5.6]. To illustrate this, let us consider the maximum of the $P_p[cos(\phi)]$ distribution and $P_p[cos(0^\circ)]$ in the three cases. If we divide the maximum of the $P_p[cos(\phi)]$ by the $P_p[cos(0^\circ)]$ in each case, for layer 1, we will get approximately 2.8, 4.7 and 11 for the CNT, fullerene and graphite, respectively. These data show, that in the case of the CNT and fullerene the $P_p[cos(0^\circ)]$ is more significant than in the case of the graphite.

5.4.1 Hydrogen bonding

It is well known that water has an extended network of hydrogen bonds that are thought to be responsible for most of its properties. It is likely that the presence of a large hydrophobic solute such as a graphitic surface will cause a disruption of the H-bond network. To further investigate the influence of the graphitic surfaces on the molecular water structure, we calculated the average number of hydrogen bonds per water molecule as a function of distance from the surface. To define hydrogen bonds in our systems, we used the geometric criteria [236] whereby two water molecules are considered to be

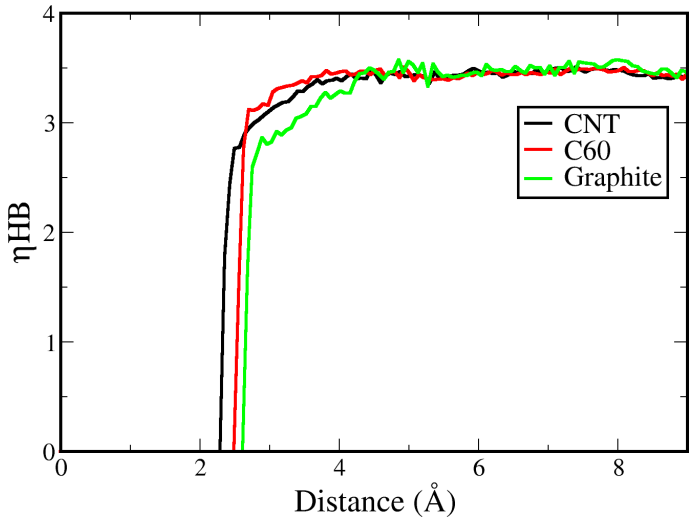

Figure 5.9: Average number of hydrogen bonds (η_{HB}) per water molecule as a function of the distance r from each surface. For clarity the fullerene and graphite profiles are off-set the origin by 0.1 Å.

hydrogen bonded only if the inter-oxygen distance is less than 3.55 Å and simultaneously the hydrogen-oxygen (hydrogen bonded) distance is less than 2.5 Å and the H—O$\cdots$O angle is less than 30°.

The hydrogen bond profile for the the three surfaces is shown in Figure 5.9 together with the corresponding averages in Table 5.4. In the case of the CNT, we find that the average number of hydrogen bonds decreases from a bulk value of about 3.4 per molecule to a value of about 2.85 at the CNT-water interface. Similarly, for the fullerene interface the average number of hydrogen bonds decreases from a bulk value of approximately 3.4- 3.5 to an interfacial value of about 2.9. For the graphite system, an equivalent behaviour is found, going from an average number of hydrogen bonds of 3.4 for the bulk phase to a value of around 2.7 at the interfacial region.

This behaviour at the interface seems to be due to the plane of the water molecules being oriented tangential to the surface. It is possible that the structured water

will prevent the formation of more than three hydrogen bonds due to all molecules being in the same orientation. This is similar to what has been reported for other hydrophobic surfaces [226; 175; 208; 235]. On the other hand, water in the layers above the first layer, is in two different orientations and therefore is free to form as many hydrogen bonds as it can, which is why the number of hydrogen bonds increases after the first peak in the density profile. This conclusion is supported by the orientation distribution analysis where we found that in layer 1 water molecules orient themselves with the dipole moment, plane and OH bonds tangential to the plane of the surface. Whereas in layer 2 some of the water molecules are oriented with their planes tangential to the surface and other water molecules are oriented with the water plane slightly oriented perpendicularly to the surface.

5.4.2 Residence times of water molecules within layers

We calculated residence times τ_{res} which are the average times spent by water molecules in each layer. We considered "continuous" residence times, *i.e.*, they are computed until a water molecule leaves the layer for the first time and thus does not allow for multiple re-entrance of water molecules. The values for the residence times τ_{res}, of water molecules in each region for the three surfaces are presented in Table 5.4. In order to calculate this property five time origins were considered. For each origin we looked for water molecules in a specific layer and monitored the time (in picoseconds) that each water molecule resided in that layer for that interval. This process was repeated for five time origins. An interval of 200 picoseconds between each origin was considered. In general, a water molecule stays longer in the interfacial and bulklike region compared to the intermediate region. This is due to the attraction that the surface exerts on the water molecules, which forces the waters to be in the region of the first peak of the oxygen density profile.

Table 5.4: Average and normalised residence times of water in each layer for all three surfaces. Layers refers back to Tables 5.2 and 5.3.

Layer No.	τ_{res}(ps)	τ_{res}^{N}(ps)
CNT		
1	0.046	0.016
2	4.16	3.35
3	0.91	0.72
4	2.85	1.85
5	4.74	1.53
Fullerene		
1	0.038	0.013
2	3.52	2.69
3	0.85	0.77
4	2.51	1.48
5	2.83	1.37
Graphite		
1	0.03	0.011
2	9.38	6.42
3	0.69	0.60
4	6.04	2.77
5	1.22	1.14
6	2.10	1.37
7	26.58	2.68

The greater residence times of the water molecules in the bulk region can be explained by the greater number of hydrogen bonds which trap the water molecules. Assuming that molecules move evenly, the residence times should be normalised with the width of the considered region in order to obtain a parameter that is independent on the size of the water layer. The normalised residence times $< \tau_{res}^{N} >$ of water in each region for all surfaces, are given in Table 5.4 along with the non-normalised values. For all three surfaces, the normalised residence times are greater for layer 2 which is the region that corresponds to the highest oxygen density of the density profile and are lower for layer 3 which corresponds to the lowest density region of the oxygen density profile. In addition, the region around the second peak of the oxygen density profile (layer 4) also displays higher residence times. In the case of the graphite interface, we also notice a slightly lower residence time in layer 5, when compared to layer 6. This is because layer 5 corresponds to a region of low oxygen density as it is the area between the second and the third peak. The higher residence time values in the region of the first peak of the oxygen density profile indicates that water spends more time closer to the solute surface than in any other region.

5.5 CNT–OH defect simulations

As previously mentioned, MD simulations of a CNT with an OH defect attached to the side walls were performed in explicit solvent with the aim to investigate the structural properties of water surrounding the OH defect. This study is essential in order to be able to perform further simulations to study the interactions of defective CNTs and peptides. Figure 5.10 presents the oxygen and hydrogen density profiles around the OH defect as a function of the distance to the surface. In order to calculate the water density profiles, a 2×1×2 super-cell was created such that the C—O bond lay along the z-axis,

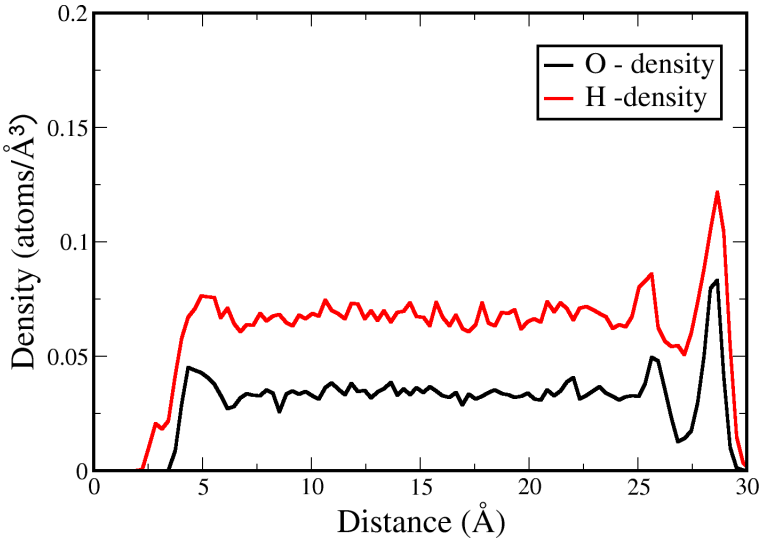

Figure 5.10: Atom density of oxygen and hydrogen around an OH defect attached to the CNT as measured from the surface.

as illustrated in Figure 5.11. Then, a cylinder of radius 1.5 Å was considered around the OH defect (dotted lines in Figure 5.11) and was divided into slices in the z direction and the number of hydrogen and oxygen atoms were counted for each slice and averaged with respect to the volume of the slice. The oxygen and hydrogen density profile show a broad peak at around 4.35 Å, and show that there is a loss of structuring compared to that observed at the ideal CNT. Between 10 and 20 Å water behaves as bulk and as expected, after this distance, water starts to behave as previously seen for the ideal CNT. The right hand side of the plot is thus, symmetric to the water density profile plot at the ideal CNT (Figure 5.2).

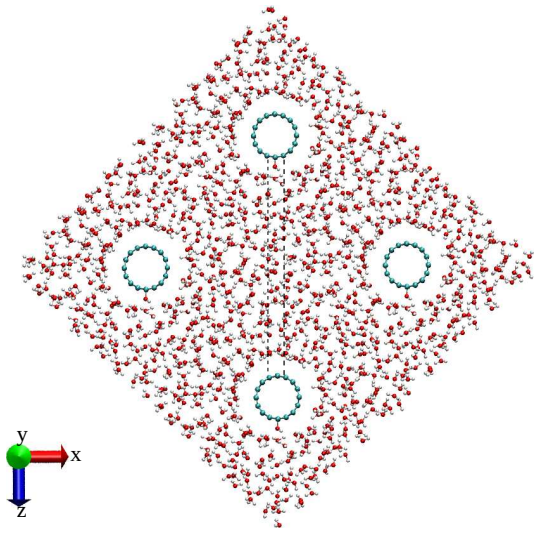

Figure 5.11: Snapshot illustrating a 2×1×2 supercell of the CNT wiht one OH defect in water in the xz plane. The C—O bond lies along the z-axis. Dotted lines correspond to the cylinder around the OH defect that goes till the CNT of the next cell.

5.6 Discussion

The results of our simulations enable us to draw some conclusions about the structure and orientation of water molecules at the liquid - graphitic surface interface. In the first layer, water molecules orient themselves with the dipole moment nearly tangential to the plane of the surface and with both OH bonds lying flat on the solute. This correlates with the lower average number of hydrogen bonds per water molecule at the interface compared to the bulk. In the case of the CNT and fullerene, the plane of the water molecules is slightly inclined toward the surface, while in the case of graphite the plane of the water molecules is mostly oriented tangential to the surface. This also correlates with the almost coinciding first peak distributions of the oxygen and hydrogen density profiles. In addition, this is in good agreement with previous MD simulations, which have also provided evidence that water molecules prefer to orient the dipole moment

nearly tangential to the plane of the CNT [173], and the fullerene [208]. The differences observed between the water plane orientation for the graphite and the CNT and fullerene could be explained by differences in the curvature of each surface. Therefore, we expect that the hydrogen bonds at the interface layer will be in a planar geometry for graphite and slightly out-of-plane for the CNT and fullerene. However, it should be pointed out that since different boundaries were used the simulations are not really comparable.

We observe that water molecules in the interfacial zone (layers 1 and 2) have greater residence times than water molecules in the intermediate zone (layers 3 and 4), and in the bulk phase residence times are slightly lower than in the interfacial region. The increased residence times at the interface correlates with the region of highest density of oxygen and this is in good agreement with previous MD simulations that also calculated the residence times of water in a graphite channel [175]. In that study the authors obtained a normalised residence time of 5.46 ps and 3.59 for the interfacial and bulk water, respectively. These results are in reasonable agreement with our calculated normalised residence times for the graphite interface (6.42 ps and 2.68 ps, respectively).

In agreement with the water density, water shows anisotropic orientational ordering for all three surfaces. Our results suggest that there are no more than two layers of structured water on the surface of CNT and fullerene and three layers of structured water at the graphite interface. In all cases, the first and second layers exhibit the most structuring. The extent of this water structuring is rather weak and does not persist further than 6 Å away from the CNT and fullerene and 10 Å away for the graphite. This is in contrast to other well-studied hydrophilic surfaces; for example, the water on the surface of titania is highly structured and the ordering propagates far from the surface [174]. This obviously has consequences when considering binding biomolecules to surfaces. Strongly structured water is more likely to extend further into the bulk and

therefore exert a greater effect on the local environment of bound molecules; whereas weakly structured water is probably easier to displace for direct interactions with the surface.

5.7 Summary and Outlook

MD simulations to investigate the interactions of water with graphitic surfaces — carbon nanotubes, fullerenes and graphite have been carried out. The existing polarisable AMOEBAPRO force-field [142; 143; 144] which includes distributed multipoles up to and including quadrupoles for the electrostatics was extended to include a description of the non-bonded interactions between water and the graphic surfaces at the same level of theory. Water density profiles, water orientational structure, average number of hydrogen bonds and water residence times have been calculated. These simulations suggest that the water ordering is fairly weak and that there are no more than three layers of structured water on the graphitic surfaces. The spatial and orientational distributions of water around the graphitic surfaces are shown to reproduce the hydration behaviour very well, when compared with previous work. Furthermore, these results show that this extended force-field can be used to model aqueous graphitic surfaces interfaces and it is therefore viable for use in future studies of biomolecules with graphitic surfaces in water. In addition, MD simulations of a CNT with one OH defect in water have been performed. Water density profiles show that water structuring at the CNT+OH interface is very weak, indicating that the OH defect has a disruptive effect on the structured interfacial water.

Chapter 6

Modelling the effect of surface defects on the binding affinity between peptides and carbon nanotubes

6.1 Abstract

In this Chapterthe interactions of peptides with a carbon-nanotube (CNT) that has defects are investigated by molecular dynamics (MD) simulations. The presence of defects on CNTs is thought to be very common and arises from the synthesis and purification process; it is thought that the presence of defects affects some of the properties of CNTs. The aim of this study is to investigate if the presence of defects affects the binding affinity between peptides and CNTs. In line with the concentration and type of defects used, the findings described herein demonstrate that the presence of OH defects

attached to the outside walls of CNTs only affects the peptide locally and the binding affinity is not significantly affected by the presence of defects on the nanotube exterior.

6.2 Introduction

Most of the modelling studies related to molecular adsorption on CNTs treat them as "ideal", *i.e.*, perfect crystalline structures without defects. However, experiments have shown that even high-quality CNTs may contain at least one defect per each 4 μm [237]. Defects on CNTs can be introduced during either the synthesis or purification processes and can be of several types: topological defects, which correspond to the presence of rings in the network other than hexagons (*e.g.* the Stone-Wales defect); structural defects such as the absence of a carbon atom (vacancy); and functionalisation or chemical defects, such as the covalent attachment of atoms or groups to the carbon lattice of the nanotube [238]. These types of defects are illustrated in Figure 6.2.

While several works are dedicated to the development of techniques that can identify and count point defects in CNTs [239; 237], others are focused on the effects of the presence of defects on carbon surfaces (*e.g.* [212; 213; 216; 215; 240; 241]). The presence of defects can substantially alter the properties of CNTs, especially the mechanical [242; 243; 244; 240] and electronic properties [238; 245; 241] and may switch on or off certain desirable properties. For instance, the conductance of CNTs can be tuned by controlling the density of defects, and induced defects may be used to modify the CNT properties [246]. Several studies focused on the effect of the Stone-Wales defects in the mechanical properties [242; 243; 244; 240], such as the fracture strength [244], indicate that even a few number of defects in the carbon lattice can give rise to a degradation of the CNT mechanical properties [242]. A number of studies have also investigated the effect of vacancy defects on CNTs on its mechanical [244; 247] and

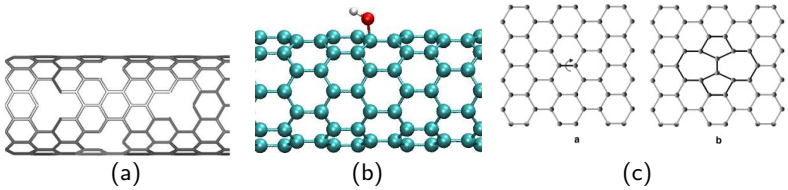

Figure 6.1: (a) Schematic of a CNT with random vacancies (From [248]), (b) CNT with a covalent functionalisation and (c) schematic of the formation of a Stone-Wales defect on a CNT. (From [252]).

electrical properties [245; 248]. In addition, a wide number of studies also investigated the effects of covalent functionalisation of CNTs and multi-walled carbon nanotubes (MWCNTs) with polymers [64] and biomolecules [66; 59; 67]. This procedure has been used not only as an effective way of solubilising and dispersing CNTs in water but also as a tool for the fabrication of nanocomposite materials [249; 250]. A review of covalent functionalisation with organic moieties, along with the different techniques used to characterise the functionalised materials can be found in Reference [251].

As mentioned in Chapter 5, the interaction of water with pure, ideal graphitic surfaces has been object of extensive experimental [253; 254] and theoretical studies, not only by modelling bulk water adsorbed on the graphitic surface ([173; 192; 199; 202; 255], for example) but also by modelling the interaction of water clusters with these surfaces [209]. From both types of studies it has been demonstrated that the adsorption interaction of water at these bare carbon surfaces is rather weak [192; 255; 202; 199]. Ulberg and Gubbins [209] demonstrated the importance of hydrogen bonding on the mechanism of adsorption of water onto carbon slit pores, by calculating adsorption isotherms. Following this work, Maddox *et al.* [256] included a description of carboxyl groups on the surface of the pore, in which they concluded that the presence of such hydrophilic sites significantly increased the adsorption of water although it was dependent on the arrangement of the active groups on the surface, and whether or not they

were isolated. Subsequent studies by Gubbins *et al.* [210; 211; 212] presented a simple model for the adsorption of water in activated carbons, based on a square-well potential to describe surface sites. Grand canonical Monte Carlo (GCMC) simulations were performed on pure water [210] and methane/water [211] vapour mixtures [210; 211] adsorbed onto porous activated carbons. In agreement with the early experimental study of Dubinin [257], Gubbins *et al.* [211] concluded that the mechanism of water adsorption in activated carbons is based on the initial adsorption of water molecules at the activated carbon followed by the clustering of water molecules and pore filling at higher pressures, with the formation of hydrogen bonds being the driving force for water adsorption. Furthermore, Gubbins *et al.* demonstrated the importance of the site density and site distribution of defects in the water adsorption by varying the configuration of sites (for a fixed concentration) [210].

In addition, there are many experimental [257] and theoretical [213; 258; 212] studies on adsorbed small molecules (including water) on activated carbon surfaces such as carbon black [216; 259], soot particles [214; 260; 261; 262; 263] and graphite [264; 265; 215]. These studies have been undertaken for a wide variety of reasons. For instance, the adsorption properties of soot particles have been considered since these particles are emitted into the troposphere by aircraft and are thought to be responsible for the nucleation of cloud ice particles [266]. Soot is typically modelled as graphitic surfaces with several functional groups attached to it, such as hydroxyls, carbonyl and carboxyl groups. The presence of such groups increases the adsorption of large quantities of water, unlike pure graphite. A series of papers reported by Picaud *et al.* have reported the adsorption of water on partially oxidised soot surfaces using both quantum mechanical (QM) calculations [214; 264] and MD simulations [260; 265; 215]. The QM calculations aimed to investigate the local structure of the activated sites as well

as the optimal arrangement of the water molecules (adsorbed as water clusters) around it. In addition, the calculations also supplied values of partial point charges for the functional groups of hydroxyl [264] and carboxyl [214] defective sites. The obtained geometries were used to perform MD simulations to characterise the adsorption of water molecules (not liquid water) on activated graphite surfaces [260; 265; 215]. These authors were the first to carry out MD simulations of water clusters adsorbed on defective graphitic surfaces, and were also the first to perform MD simulations to characterise the competition between the affinity of water—OH and water—COOH by simulating both types of sites simultaneously on the surface [215]. These studies suggest that the presence of chemical defects such as COOH groups act as a trapping site for the water molecules, which then attract other water molecules leading to the formation of large aggregates [214; 260]. However, in the case of hydroxyl groups, the OH—water interaction is weaker than the water—water interaction which leads to the clustering of water molecules at the site. GCMC simulations were also performed by Picaud *et al.* to study the adsorption isotherms of water on soot particles modelled by concentric graphite layers with randomly distributed OH and COOH groups [262]. In agreement with their previous MD simulations [260; 265] these studies highlighted that the OH sites have a weaker effect on the adsorption of water molecules compared to the COOH sites.

The adsorption of other solvents such as methanol and ethanol, on graphitised carbon black has also been considered [259]. It was found that both the presence and the configuration of the carbonyl groups attached to the surface made the surface more attractive with respect to adsorption. However, once a water monolayer was formed, the adsorption behaviour between both surfaces (with and without carbonyl groups) was found to be similar. More recently, Birkett *et al.* [216] used an idealised method to

introduce heterogeneity on a carbon surface by considering several functional groups as active sites on the surface. Functional groups were primarily modelled as fixed water molecules parallel to the carbon surface, within this model, these authors considered the three functional groups that are usually found in carbon blacks; carbonyl, hydroxyl and carboxyl groups [267]. The authors justified the use of fixed water molecules as a model to represent heterogeneity at the surface, taking into account that if these fixed water molecules could not promote water adsorption, it would be unlikely that any other functional group would promote adsorption of water molecules on the surface. Birkett *et al.* [216] concluded from these studies that the fixed water molecules functioned as clustering sites and therefore increased the adsorption of water molecules compared to the bare surface. While the adsorption of water molecules due to both carbonyl and hydroxyl goups was very weak, the adsorption of water molecules due to two carboxyl groups was found to be comparable to that of two fixed water molecules (set at a distance from the surface comparable to a carboxyl site). The higher affinity for water molecules of the carboxyl groups (compared with carbonyl and hydroxyl) was found to be related to the distance the group extends from the surface (which is similar to the distance of the fixed water molecules).

Common to all these studies is the modelling of the carbon surfaces as graphite, the inclusion of functional groups (such as hydroxyl, carbonyl and carboxyl groups), and, the modelling of water as clusters not as liquid water. Furthermore, despite the considerable number of studies addressing the influence of defects on the properties of CNTs, no one yet has reported how the presence of defects on CNTs might affect the interaction of peptides at such interfaces. Therefore binding affinity changes associated with the presence of defects has not been reported until the present study. An understanding of the effects of the presence of defects in CNTs on the binding affinity is crucial, in the

development of new materials as well as for biomedical applications. There is a wide range of new proposed therapies and applications based on CNTs, including cell destruction [268; 78], antibacterial activity [80], sensitive biosensors [59; 60], drug delivery [72], electroanalytical devices [269], artificial muscles [270] and laser heating cancer therapy [268; 271]. Despite these potential therapeutical applications, there is the concern that exposing humans to CNTs may cause significant health problems. In addition, issues related with environmental impact must also be addressed [59]. However, studies suggest that if CNTs below a given length (1 to 5 μm) are wrapped with biomolecules [71] they have limited toxicity [70], while CNTs greater than a given treshold cannot be excreted and may build up in the kidneys. Since defects are almost certainly going to be present on the CNTs, it is important to understand its effects on the interaction with peptides in order to be able to control binding affinity.

6.3 Methods

As mentioned before, the aim of this study is to examine if the presence of defects and therefore the departure from the CNT ideality affects the binding affinity between CNTs and peptides. To achieve this, several OH groups were attached to the surface of the CNT. The inclusion of the parameters for the OH defect was based on the assumption that the OH group and the carbon attached to the OH group are similar to the tyrosine residue in AMOEBAPRO. The carbon atom at which the defect is attached was also treated as similar to tyrosine. This enabled the use of the tyrosine parameters in the description of the OH defect attached to the CNT. In contrast to the carbon atom attached to the OH group in tyrosine, the carbon atom attached to the OH group in the CNT has valence 4 instead of 3. For this reason a new atom type and atom class was created in which the same bonded and VDW parameters as tyrosine were used but the

valence was 4. The dipoles and quadrupoles for the carbon attached to the OH group and for the OH group were set to zero and the carbon atom charge was adjusted in order to obtain a neutral charge overall for the COH group. No polarisation was included in the COH group. Similarly, the parameters for the bonded terms were the same as the parameters for tyrosine. For instance the C—O, the O—H stretch and the C—O—H angle were the same as for tyrosine and are summarised in Table 6.1.

Table 6.1: Potential parameters for the COH defect used in the present MD simulations. These parameters were taken from the tyrosine residue in the AMOEBAPRO force-field.

VDW parameters			
Atom	σ /Å	ϵ /kcal mol^{-1}	q/e
C	3.800	0.0890	+0.22900
O	3.405	0.1100	-0.56325
H	2.655	0.0135	+0.33425
Interacting pair	distance (Å)		
C—C	1.4214		
C—O	1.3550		
O—H	0.9470		
C—O—H angle	109.00°		

As previously mentioned, it is known from experiments that high-quality SWCNTs can contain, on average, about one defect per 4 μm. This is equivalent of one defect per 10^{12} atoms [237]. In addition, defects can be bunched together or spread over the tube, or more likely a combination of both. Because of this, we ran simulations of the peptide sequences with the CNT with OH defects arranged in two different distributions (one with the defects bunched in one site, and another with the defects spread over the CNT). Herein, I will use *spread defects* to refer to the four OH defects spread over the CNT and *bunched defects* for the 4 OH defects bunched together in one site. The spread defects were all selected from a region of width 60 Å in the centre of the

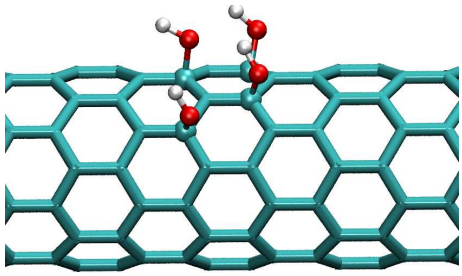

Figure 6.2: CNT with 4 OH bunched defects.

nanotube, in order to avoid having defects near the CNT ends and therefore avoiding the edge effects arising from the presence of the terminal hydrogen atoms. The first defect was randomly selected from that region and the second defect was randomly selected under the condition of it being at a distance greater than 20 Å from the first OH defect. The third OH defect was randomly selected under the condition of being at a distance greater than 15 Å from the first and second OH defects, while the fourth defect was randomly selected to be at a distance greater than 10 Å from the previous three defects. The bunched defects were chosen from two pairs of carbon atoms from two adjacent hexagons at the centre of the CNT, where each pair is intercalated by an sp2 carbon. This is because we do not have the torsional parameters for two defective sites together. Figures 6.2 and 6.3 show an illustrative example of the bunched and spread defects, respectively.

MD simulations of systems comprised of a defective CNT and one each of the peptide sequences B1 (HWKHPWGAWDTL), B3 (HWSAWWIRWNQS), B1Y (HYKHPYGAYDTL) and B3Y (HYSAYYIRSNQS) were performed with our extended AMOEBAPRO force-field which was modified to treat the interactions between peptides and the OH defect. A hydrogen-terminated zigzag (8,0) CNT with 976 atoms was used. All simulations were performed with the TINKER [171] package. Since the inclusion

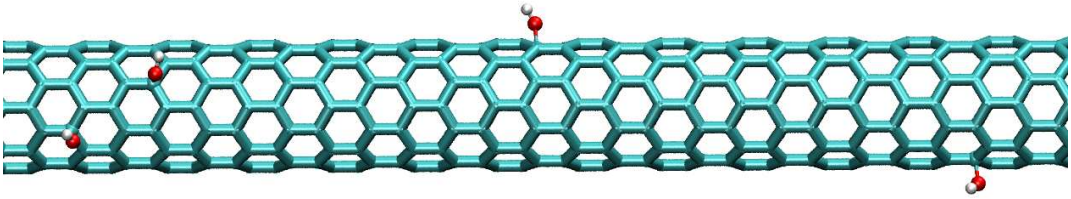

Figure 6.3: CNT with 4 OH defects spread over the tube.

of atomistic water would increase the simulation time, a continuum effective solvent model (ASP [172; 146]) was used to describe the water. As previously mentioned, this model has previously [118; 120] been shown to give good results. All simulations were carried out in the canonical ensemble at room temperature. The Verlet [129] algorithm was used to solve the equations of motion with an integration timestep of 1.0 fs and a cutoff of 8 Å was applied to all non-bonded interactions. The systems were equilibrated for 1 ns, and an additional 1 ns of production run was performed. To attempt to identify many different binding configurations, several different geometries were used for each peptide sequence. The best geometries obtained from the simulations described in Chapter 3 were used as initial configurations for these simulations.

Once again, to analyse the results both structural and energetic metrics were used. Binding affinity was quantified in terms of normalised interaction energy E_N, where the interaction energy between the peptide and the CNT was divided by the number of atoms in the peptide. Structural metrics include the average ring-CNT distance (calculated as the distance between the ring centre-of-mass and the CNT surface), ring-tilt relative to the surface normal (calculated as described in Chapter 3, Section 3.3), average *end-to-end* peptide distance, and peptide *y-shift*. Again, the average end-to-end peptide distance was calculated by determining the distance between each carbon of each

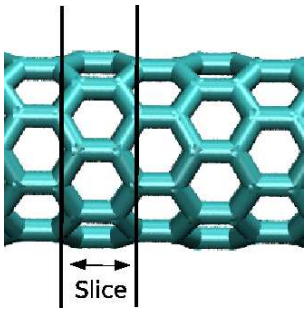

Figure 6.4: Schematic of a CNT slice.

of the two terminal methyl groups (blocking groups) and averaging this distance over the production run. The peptide y-shift was calculated by determining the y-component of the peptide centre-of-mass for each frame and calculating the distance between this y-component for the first frame of the production run, and each subsequent frame of the production run. The y-shift was then averaged over the production run. Since the distance was calculated by taking the square root of a squared distance, the y-shift is always a positive value. Large y-shifts indicate that the peptide is far from its initial position, while a y-shift close to zero indicates that the peptide spent the trajectory close to its initial position. These data do not give any information about the direction in which the peptide is moving. In addition, the *maximum displacement* is defined as the maximum distance travelled by the peptide along the y-axis (the maximum distance in the y-shift profile). In order to estimate the peptide mobility on both the ideal and the defective CNTs, the peptide residence time was calculated by choosing the α-carbon of the residue 6 in each peptide sequence and calculating the time it spent, on average, in a 2 Å slice, which was defined by incrementing $+2.0$ Å from the origin along the y-axis. This was done by arbitrarily picking 10 time origins and calculating the time it took to move 2 Å away from that origin. In Figure 6.4 is a schematic of a CNT slice. In addition,

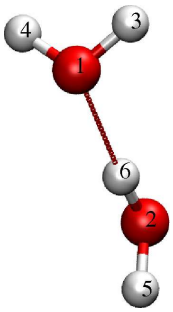

Figure 6.5: Schematic of the water-dimer geometry used in the force-field validation.

the three closest peptide atoms to each of the 4 OH-defects was identified as a function of time. This was done by calculating the distance between all the peptide heavy atoms (C, N, O) and each of the four OH defects. Only the distances less than 8 Å were considered, since this is the value of the non-bonded cutoff used in these simulations. In order to look for the first, second and third atoms that spent more time close to each of the defects, the frequency at which each peptide atom was the closest to the defect was then determined. This frequency was calculated by counting the number of times each heavy atom was the closest and dividing by the number of frames of the production run.

6.4 Results

6.4.1 OH defect validation

In order to validate the extension of the force-field to describe OH defects, the absolute pair VDW and electrostatic energies between H and O were compared for several configurations of a water–dimer and a CNT with one OH defect and a water molecule (water–CNT+OH) in exactly the same configuration. In Figures 6.5 and 6.6 are shown one of the geometries of the water-dimer studied and the correspondent geometry of

a water-CNT+OH, respectively. The water–CNT+OH configuration was obtained by optimising the water–dimer and then replacing one of waters with the CNT+OH defect. The same was done for a benzene–water complex and a benzene–CNT+OH system. In this case, the benzene–water complex was optimised and then the water molecule was replaced with the defect. In Figures 6.7 and 6.8 are shown the geometries for the water–benzene and benzene–CNT+OH systems, respectively. The absolute pair VDW and electrostatic energies along with the distance between the atom pair for both the water–dimer/water–CNT+OH and water–benzene/benzene–CNT+OH are given in Tables 6.2 and 6.3. Comparison of the absolute pair energies and pair distances for the water–dimer and the water–CNT+OH (Table 6.2), show that the trends in the distance and energy of each atom pair are favourable overall, despite the fact that the absolute pair energies and distances do not exactly agree. However, it was not expected for them to necessarily agree; for example, these data show that the water binds less strongly to the defect than to another water, which is in agreement with what was expected since in the case of the defect the oxygen is attached to a carbon atom instead of another hydrogen atom such as in the case of water. In addition, the defect does not include a dipole, quadrupole or polarisation. In the case of the water–benzene and benzene–CNT+OH (Table 6.3), comparison of the absolute pair energies and pair distances show that there is reasonable qualitative agreement, despite the fact that some of the distances and energies being considerably different, *e.g.* the pairs H_{14}—C_6 and H_{14}—C_2 for the VDW energies and H_{14}—C_2 and O_{13}—C_5 for the electrostatic energies. These data suggest that benzene binds more strongly to the defect (in isolation) than to a water molecule. Furthermore, the total interaction VDW and electrostatic energies in the case of the water–dimer/water–CNT+OH and water–benzene/benzene–CNT+OH are also in good agreement, giving us confidence to use this force-field in the simulation

of OH defective CNTs and peptides. In addition, the interaction energy of both a water molecule and a benzene molecule at an ideal CNT and at a CNT with one OH defect was calculated. These data is shown in Table 6.4. Again, this was done for the water and benzene in exactly the same geometry both at the defective and ideal CNT. To accomplish this, first the water/benzene at the defective CNT was optimised and the interaction energy was calculated by running a single point energy calculation on the resulting geometry. Then, the defect was removed and the "ideal" interaction energy was calculated again by running a single point energy calculation of that geometry. As expected, the interaction in the case of the water–CNT+OH is stronger than in the ideal CNT, since the OH group attracts the water molecule more then the ideal CNT. In the case of the benzene–CNT+OH the interaction is slightly stronger then in the ideal case. These data suggest that the OH defect model is suitable to be used in the simulations of peptides and defective CNTs, since it is qualitatively in good agreement with the atom pair distances and energies of a water–dimer and a water–benzene complex and the interaction of water/benzene molecule at the defective CNT is stronger than at the ideal CNT.

6.4.2 Petide–defective CNT simulations

In Figure 6.9 is a snapshot of the best trajectory of B3Y interacting with one of the OH defects in the spread CNT. The normalised interaction energies, E_N averaged over the two distinct lowest binding trajectories for the peptide sequence B1, B3, B1Y and B3Y are presented in Table 6.5 for the ideal CNT, the bunched defects CNT and the spread defects CNT, along with the standard error of the mean (SEM), being defined by the standard deviation of the sample means. These results suggest that the inclusion of OH defects on the CNT exterior does not cause any major difference in the binding affinity

Table 6.2: Absolute pair VDW and electrostatic energies (kcal/mol) and pair distances (Å) for the H and O in the water–dimer and in the water–CNT+OH system.

VDW					
water–dimer			water–CNT+OH		
Atom pair	Actual distance	Energy	Atom pair	Actual distance	Energy
O_1—O_2	2.8920	0.1757	O_1—O_2	2.8683	0.2216
O_1—H_5	2.0213	2.2705	O_1—H_5	2.0113	2.3762
O_1—H_6	3.3277	−0.0260	O_1—C_5	3.8163	−0.0900
O_2—H_3	3.2770	−0.0273	O_2—H_3	3.2596	−0.00277
O_2—H_4	3.2771	−0.0273	O_2—H_4	3.2492	−0.0279
H_3—H_5	3.7857	−0.0025	H_3—C_5	4.2592	−0.0117
H_3—H_6	2.4584	−0.0091	H_3—H_6	2.4686	−0.0097
H_4—H_5	3.7858	−0.0025	H_4—C_5	4.3047	−0.0110
H_4—H_6	2.4585	−0.009	H_4—H_6	2.4298	−0.0071
Total:	2.3425 kcal/mol			2.4369 kcal/mol	
Multipoles					
O_1—O_2	2.890	30.8120	O_1—O_2	2.8683	35.4456
O_1—H_5	3.3800	−13.2630	O_1—C_5	3.8163	−10.7135
O_1—H_6	1.9355	−28.1722	O_1—H_6	1.9269	−32.3869
O_2—H_3	3.3251	−12.9191	O_2—H_3	3.3083	−14.4984
O_2—H_4	3.3253	−12.9187	O_2—H_4	3.2970	−14.5527
H_3—H_5	3.8965	5.5275	H_3—C_5	4.3102	4.5578
H_3—H_6	2.4363	9.6714	H_3—H_6	2.4514	11.5496
H_4—H_5	3.8966	5.5273	H_4—C_5	4.3595	4.5001
H_4—H_6	2.4364	9.6708	H_4—H_6	2.4056	11.7822
Total:	−6.064 kcal/mol			−4.3162 kcal/mol	
	− 3.7215 kcal/mol			**−1.8793 kcal/mol**	

Table 6.3: Absolute pair VDW and electrostatic energies (kcal/mol) and pair distances (Å) for the H and O in the water–benzene and in the benzene–CNT+OH system.

benzene–water			benzene–CNT+OH		
		VDW			
Atom pair	Actual distance	Energy	Atom pair	Actual distance	Energy
H_7—H_{14}	3.7138	−0.0057	H_7—H_{14}	3.8368	−0.0046
O_{13}—H_{14}	4.1608	−0.0158	O_{13}—H_{14}	4.2353	−0.0142
H_{14}—C_1	3.0867	−0.0079	H_{14}—C_1	3.2146	−0.0224
H_{14}—H_{12}	3.4931	−0.0081	H_{14}—H_{12}	3.7593	−0.0052
O_{13}—H_{12}	4.0444	−0.0187	O_{13}—H_{12}	4.1167	−0.0168
H_{14}—C_6	3.0408	0.0010	H_{14}—C_6	3.1589	−0.0177
O_{13}—C_6	3.5954	−0.0987	O_{13}—C_6	3.6839	−0.0973
O_{13}—C_1	3.6813	−0.0974	O_{13}—C_1	3.7635	−0.0934
H_{14}—H_8	3.6247	−0.0065	H_{14}—H_8	3.7279	−0.0055
O_{13}—H_8	4.2966	−0.0130	O_{13}—H_8	4.3588	−0.0119
H_{14}—C_2	3.0264	0.0044	H_{14}—C_2	3.0255	−0.0046
H_{14}—H_{11}	3.6508	−0.0063	H_{14}—H_{11}	3.5653	−0.0072
O_{13}—H_{11}	4.0330	−0.0190	O_{13}—H_{11}	4.1240	−0.0167
H_{14}—C_5	2.9314	0.0347	H_{14}—C_5	3.0255	0.0046
O_{13}—C_5	3.6037	−0.0987	O_{13}—C_5	3.6895	−0.0971
H_{14}—H_9	3.4692	−0.0084	H_{14}—H_9	3.5357	−0.0075
O_{13}—H_9	4.3109	−0.0128	O_{13}—H_9	4.3693	−0.0118
H_{14}—C_3	2.9157	0.0414	H_{14}—C_3	2.9459	0.0290
H_{14}—H_{10}	3.4011	−0.0094	H_{14}—H_{10}	3.7593	−0.0052
O_{13}—H_{10}	4.1870	−0.0152	O_{13}—H_{10}	4.1167	−0.0168
H_{14}—C_4	2.8665	0.0663	H_{14}—C_4	2.9459	0.0290
Total:	−0.2938 kcal/mol			−0.3933 kcal/mol	
		Multipoles			
H_7—H_{14}	3.7412	2.4852	H_7—H_{14}	3.8685	3.1385
O_{13}—H_7	4.2110	−4.4527	O_{13}—H_7	4.2849	−4.9095
H_{14}—C_1	3.0352	−3.0202	H_{14}—C_1	3.1679	−3.7808
H_{14}—H_{12}	3.5048	2.6944	H_{14}—H_{12}	3.7943	3.2001
O_{13}—H_{12}	4.0920	−4.6827	O_{13}—H_{12}	4.1635	−5.0617
H_{14}—C_6	2.9942	−3.0632	H_{14}—C_6	3.1152	−3.8700
O_{13}—C_6	3.5954	5.3295	O_{13}—C_6	3.6839	5.8504
O_{13}—C_1	3.6813	5.1238	O_{13}—C_1	3.7635	5.6667
H_{14}—H_8	3.6239	2.6234	H_{14}—H_8	3.7306	3.2569
H_{14}—C_2	2.9563	3.2244	H_{14}—C_2	3.0756	−3.9393
H_{14}—H_{11}	3.6846	2.5153	H_{14}—H_{11}	3.5768	3.4007
O_{13}—H_{11}	4.0804	−4.6228	O_{13}—H_{11}	4.1709	−5.0526
H_{14}—C_5	2.8708	−3.3302	H_{14}—C_5	2.9654	−4.1517
O_{13}—C_5	3.6037	5.3894	O_{13}—C_5	3.6895	−5.8358
H_{14}—H_9	3.4447	2.8403	H_{14}—H_9	3.5099	3.4639
O_{13}—H_9	4.3647	−4.3442	O_{13}—H_9	4.4221	−4.7474
H_{14}—C_3	2.9563	−3.2244	H_{14}—C_3	2.8652	−4.3712
H_{14}—H_{10}	3.3825	2.8839	H_{14}—H_{10}	3.7943	3.2001
O_{13}—H_{10}	4.2381	−4.4972	O_{13}—H_{10}	4.1635	−5.0617
H_{14}—C_4	2.7858	−3.6480	H_{14}—C_4	2.8652	−4.3712
Total:	−10.2248kcal/mol			−12.304 kcal/mol	
	−10.52			**−12.70**	

Table 6.4: Interaction energy (kcal/mol) for a water–CNT+OH, water–ideal CNT, benzene–CNT+OH and benzene–ideal CNT systems.

System	Interaction Energy (kcal/mol)
water-ideal CNT	-1.95
water-CNT+OH	-4.57
benzene-ideal CNT	-1.79
benzene-CNT+OH	-2.97

Table 6.5: Normalised interaction energies, E_N (kJ mol^{-1}atom^{-1}), and the standard error of the mean (SEM) of E_N, averaged over the two distinct trajectories of lowest energy for each aptamer, for the peptide—CNT interface for the ideal CNT, the 4 OH bunched defects and the 4 OH spread defects.

	Ideal CNT		**Bunched defects**		**Spread defects**	
Aptamer	E_N	SEM	E_N	SEM	E_N	SEM
B1	-1.35	0.01	-1.29	0.07	-1.36	0.01
B3	-1.32	0.02	-1.35	0.02	-1.35	0.02
B3Y	-1.17	0.05	-1.18	0.04	-1.12	0.06
B1Y	-1.19	0.01	-1.15	0.07	-1.21	0.01

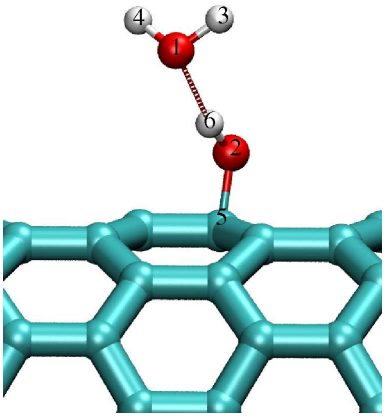

Figure 6.6: Schematic of the water–CNT+OH geometry used in the force-field validation.

between the ideal and defective CNTs. As found for the ideal CNT, in the case of the bunched defects and spread defects CNTs, the original sequences B1 and B3 show the greatest affinity to the CNT compared to the mutant sequences. However, there is no significant difference between the normalised interaction energies of the same peptide sequence in each of the cases (ideal CNT, bunched and spread defect CNTs).

In order to investigate if the binding geometry of the peptide sequences was affected by the inclusion of defects on the CNT, the ring-to-surface distances and the ring-tilt angles were calculated. Figures 6.10 and 6.11 show the distribution profile of the centre-of-mass to CNT surface distances for H, W and Y for the best trajectories of B1, B3, B1Y and B3Y for the bunched defect and spread defect cases, respectively. Profiles for the same peptide sequences adsorbed at an ideal CNT can be found in Chapter 3 in Figures 3.4 and 3.5. The differences in the binding geometry between the original and the mutant sequences at both the ideal and the defective CNT (both bunched and spread) are similar. In the case of the original sequences, B1 and B3, all tryptophans and one histidine mantain relatively close contact to the CNT surface, while

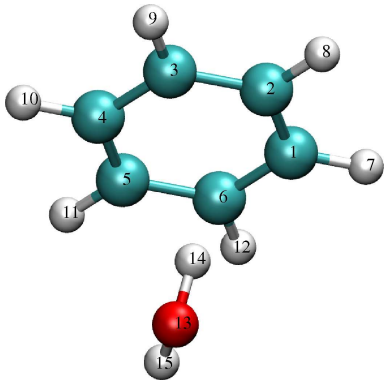

Figure 6.7: Schematic of the benzene–water geometry used in the force-field validation.

for the peptide mutants, B1Y and B3Y, configurations were adopted where only one or two rings were relatively close to the CNT surface. Figures 6.12 and 6.13 show the average ring-tilt angle and fluctuation in this angle (calculated as the root-mean-square deviation of the tilt angle) presented as a bar chart for the original sequences B1 and B3 and the mutated sequences B1Y and B3Y for the bunched and spread defects CNTs, respectively. The same bar charts for the ideal CNT can be found in Chapter 3 in Figures 3.7 and 3.8. These ring-tilts reinforce the information from the ring-to-surface distance profiles, emphasising that tryptophan residues always adopt a flat orientation in the CNT surface while tyrosine can also adopt configurations where the ring is perpendicular to the CNT surface, as for instance, in the case of Y6 in B3Y for both bunched and spread defects (Figures 6.12 and 6.13, respectively). Because the indole ring in the tryptophan residues maintain a flat orientation, they are able to get closer to the CNT surface. On the other hand, it was concluded in Chapter 3 that tyrosine is more flexible (since it only has one 6-membered aromatic ring) and adopts several configurations relative to the surface; therefore it does not get as close to the CNT surface as tryptophan. With the exception of ring-tilts of B3Y in the spread case, the presence of OH defects on the CNT

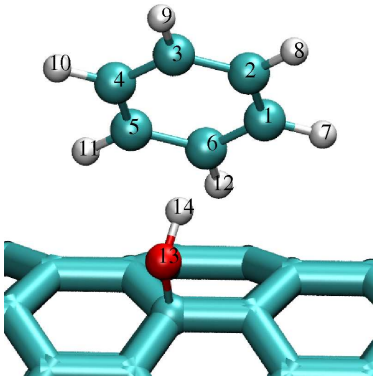

Figure 6.8: Schematic of the benzene–CNT+OH geometry used in the force-field validation.

exterior does not seem to affect either the ring-to-surface profiles or the ring-tilt-angles, as evidenced by the similar profiles to the case of the ideal CNT. However, for B3Y in the spread CNT, there is an increased number of rings with an orientation around 90° (3 rings against 2 in the case of the ideal CNT). Later in this Section an attempt to correlate this result with other data will be presented.

To better understand if the OH defects have any effect on the geometry and mobility of the peptide sequences when adsorbed on the CNT, the peptide end-to-end distance, the y-shift along with the maximum peptide displacement and the peptide residence time were calculated. In Table 6.6 is the average end-to-end peptide distance and respective fluctuation, the average y-shift and fluctuation, the maximum peptide displacement and the average residence time (τ) for the ideal CNT, the bunched case and the spread case. With the exception of B3Y in the bunched defects, the peptide end-to-end distance is greater for the original sequences, B1 and B3, then for the mutated sequences, B1Y and B3Y, for all cases (ideal and defective CNTs). However, with the exception of B3 and B3Y in the bunched case, there is no significant difference

Table 6.6: Peptide end-to-end distance (Å) and respective root-mean-square-deviation (rmsd), y-shift (Å) and rmsd, peptide maximum displacement (Å) and peptide residence time (τ/ps) on the ideal, bunched and spread defect CNTs.

Aptamer			**Ideal CNT**			
	End-end	**rmsd**	**y-shift**	**rmsd**	**Maximum displacement**	τ
B1	27.32	1.39	26.30	8.92	55.90	7.45
B3	27.04	2.85	23.40	5.38	48.07	6.30
B1Y	13.34	3.46	9.34	6.71	29.23	4.80
B3Y	16.69	2.26	3.84	2.52	11.20	4.70
Aptamer			**Bunched**	**defects**		
	End-end	**rmsd**	**y-shift**	**rmsd**	**Maximum displacement**	τ
B1	27.38	1.73	12.14	9.05	31.10	2.30
B3	20.80	3.89	8.53	6.49	28.58	2.70
B1Y	16.75	3.20	6.76	3.01	13.33	4.65
B3Y	36.74	2.61	7.52	5.45	20.87	2.55
Aptamer			**Spread**	**defects**		
	End-end	**rmsd**	**y-shift**	**rmsd**	**Maximum displacement**	τ
B1	27.12	1.34	8.91	6.74	23.24	2.05
B3	21.21	5.54	5.37	4.65	18.04	2.85
B1Y	13.87	2.16	5.64	3.86	14.68	2.90
B3Y	12.79	2.24	6.20	4.01	17.89	2.40

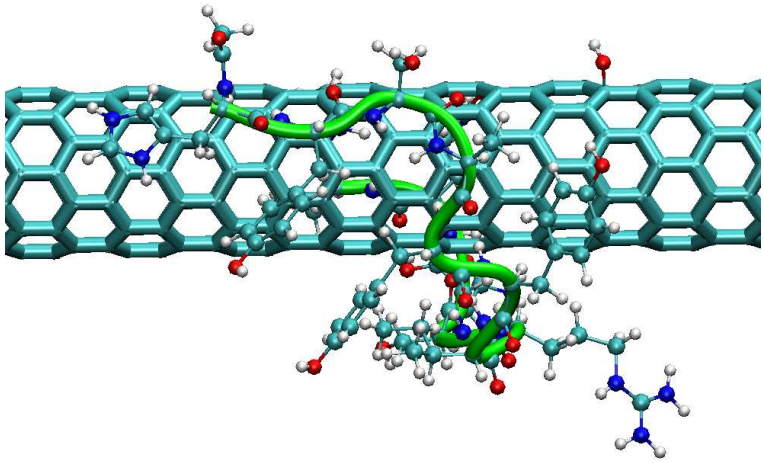

Figure 6.9: Snapshot taken from the trajectory of the B3Y peptide interacting with the CNT with 4 OH spread defects. The peptide backbone is highlighted.

in the peptide end-to-end distance between the ideal and the defective CNTs. In the case of the y-shift and maximum displacement (which give the maximum distance the peptide moved along the y-axis), it is clear from Table 6.6, that B1 and B3 on the ideal CNT have a large y-shift and a large maximum displacement, indicating that these peptides travelled a further distance compared to the peptides with the defective CNTs. However, the y-shift of the mutated sequences, B1Y and B3Y, do not show such distinction between the ideal and defective CNTs. But in the case of the maximum displacement of the mutated sequences, there are differences between the ideal and the defective CNTs. In the case of B1Y at both bunched and spread defect CNTs, the maximum displacement is smaller than in the case of the ideal CNT. On the contrary, the maximum displacement of B3Y at the defective CNTs is greater than in the case of the ideal CNT. The average peptide residence times (τ) help to understand the effect of

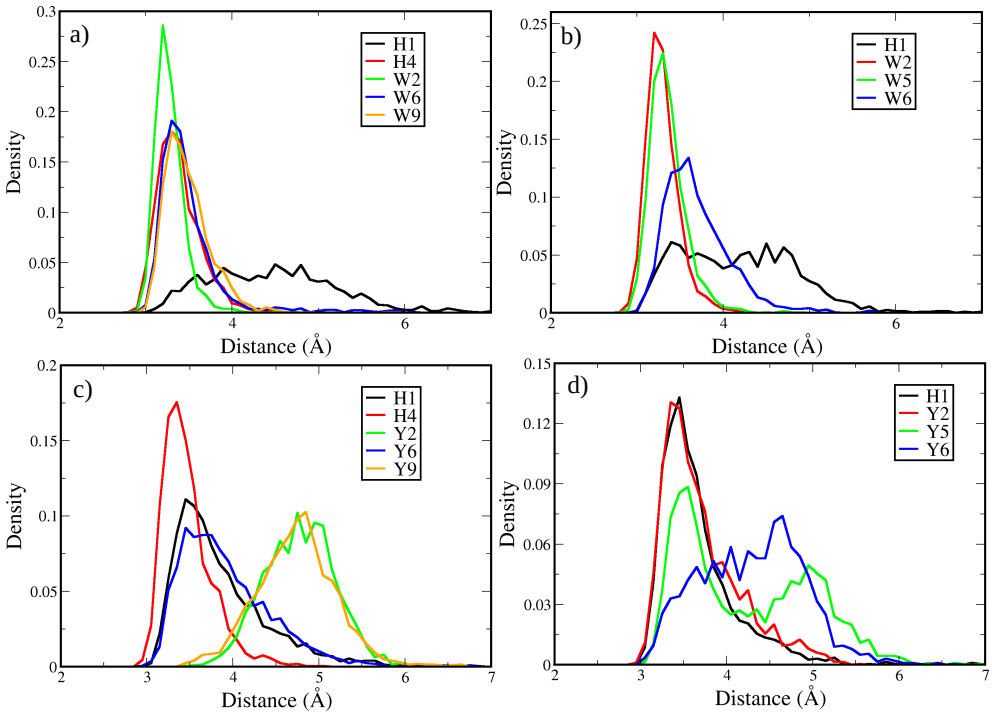

Figure 6.10: Distribution profile of the distance from the nanotube surface to the ring centre-of-mass for all the aromatic residues H, W, Y and F for the peptide sequences (a) B1, (b) B3, (c) B1Y and (d) B3Y interacting with a CNT with 4 OH defects bunched together.

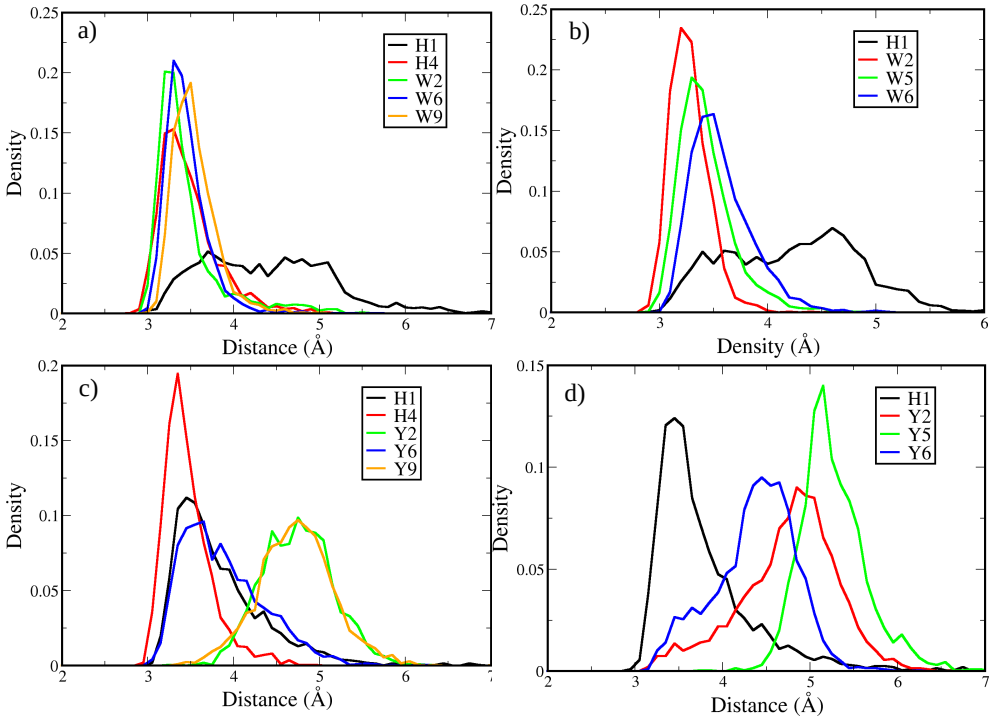

Figure 6.11: Distribution profile of the distance from the nanotube surface to the ring centre-of-mass for all the aromatic residues H, W, Y and F for the peptide sequences (a) B1, (b) B3, (c) B1Y and (d) B3Y interacting with a CNT with 4 OH defects spread over the tube.

the OH defects on the mobility of the peptide adsorbed onto the CNT. First these data show that for the ideal CNT, the peptide residence times of the original sequences, B1 and B3, is greater than in the case of the mutated sequences. Second, these data show that the peptide residence times are slightly larger in the case of the ideal CNT than in the cases of the defective CNTs. In some cases, the residence time is approximately half (or more) the value in the ideal CNT. The relatively small residence time of B3Y in the spread CNT (2.40 ps) might be correlated to the fact that there is a larger number of aromatic rings in a perpendicular orientation to the CNT surface (the ring-tilt angles in Figure 6.13). However in the case of B1 at both bunched and spread CNTs the

residence time is also small even though the ring-tilts do not seem to be affected. These data suggest that the presence of OH defects on the CNT exterior increases the mobility of the peptides. However, it seems that there is no correlation between the maximum displacement and the residence time. Yet, it is rather surprising that the maximum displacement as well the peptide residence time for B1, B3 and B1Y at the ideal CNT is large, indicating that the greater the residence time the further the peptide travels. However, inspection of the y-shift as a function of the trajectory, reveals the reason for this apparent inconsistency. As an example, in Figure 6.14 the y-shift as a function of the trajectory for B1 and B3 at the ideal CNT is shown. When the y-shift is 0 Å, it means that the peptide is at its initial position. Thus, these plots show that in the case of the ideal CNT, the peptide travelled a further distance because it moved without returning to its initial position for a longer period than in the case of the spread defects CNT. On the other hand, in the case of B1 and B3 in the spread defects CNT, the peptide is more unstable, and thus it travels backwards and forwards due to its greater mobility, explaining the smaller maximum displacement in the case of the defective CNTs.

In order to further characterise these systems, the identities of the peptide heavy atoms that interacted with the OH defects (the three closest peptide atoms within the 8.0 Å cutoff to each of the OH defects) during the production run were determined. In Table 6.7 the three closest peptide atoms to each of the OH defects for each peptide, the residue to which that atom belongs and whether or not it is a side chain atom (S) or a backbone atom (B) is summarised. The OH defects are distinguished by OH-1, OH-2, $\cdots$ and the closest atom is presented first, then the second and third closest, respectively. In addition, the frequency (expressed as a percentage of how often a certain atom was within the cutoff) is also given. The B1Y peptide at the bunched defects CNT do not have any atoms interacting with the OH defects during the whole production

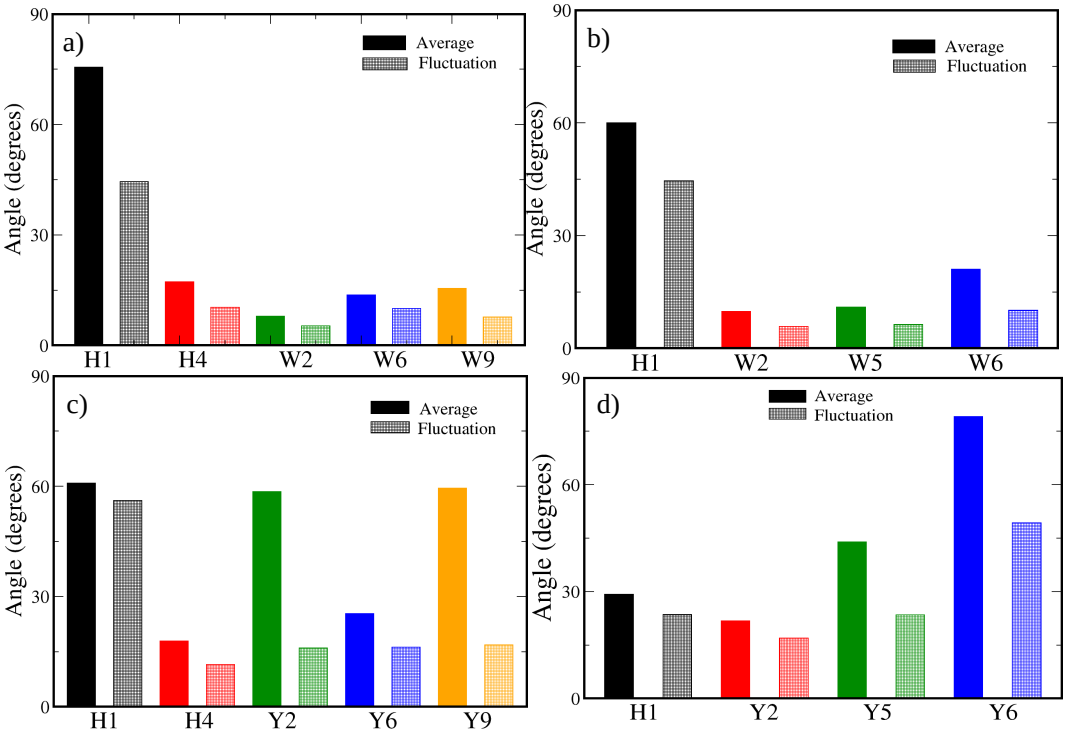

Figure 6.12: Average and fluctuation of the ring-tilt angle (relative to surface normal) at the peptide-nanotube interface for all aromatic residues in the original sequences (a) B1 and (b) B3 and the mutated sequences (c) B1Y and (d) B3Y for the case of the OH defects bunched together. A value of approximately 0° indicates an orientation parallel with the surface.

run. This is because the OH defects were located at one site and the peptide was interacting with the CNT in a different part of the nanotube. Interestingly, this peptide has an average residence time of 4.65 ps, comparable to the residence time of B1Y at an ideal CNT, giving further evidence that the mobility of the peptide is affected by the presence of the OH defects. Overall, the main residues involved in the interaction with the OH defects are histidine, tryptophan, lysine and serine. Alanine and arginine are also involved in the interaction in the case of B3Y on the spread defects CNT. In the case of the bunched defects, it seems that the peptides that have more atoms interacting with

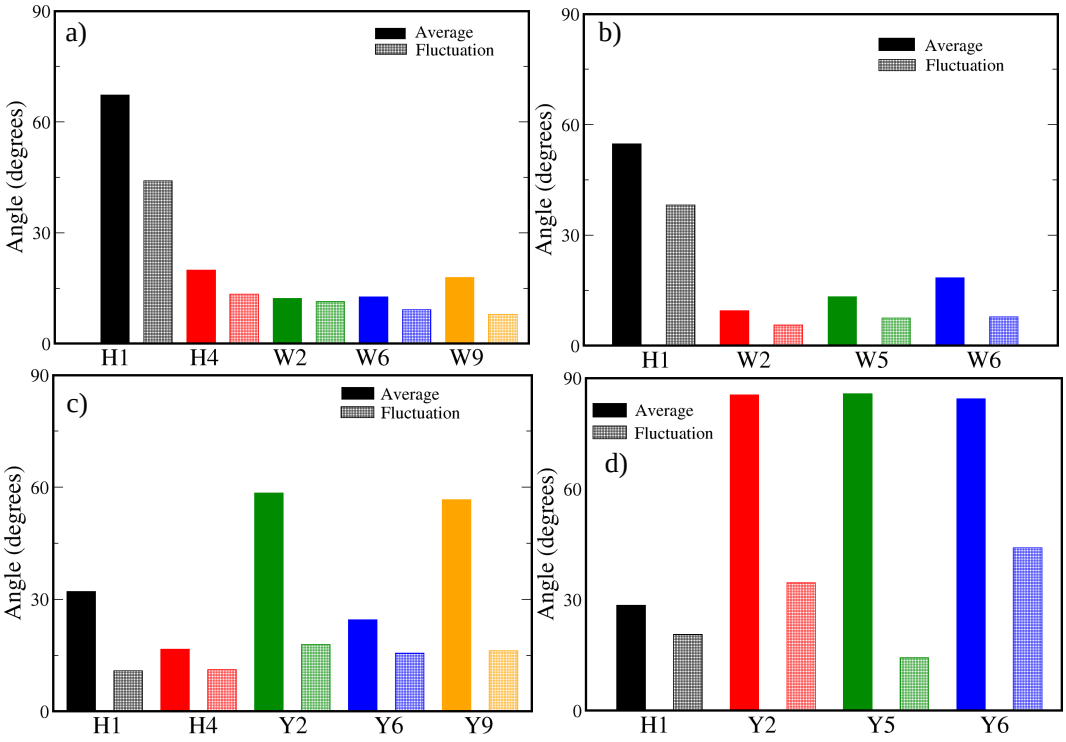

Figure 6.13: Average and fluctuation of the ring-tilt angle (relative to the surface normal) at the peptide-nanotube interface for all aromatic residues in the original sequences (a) B1 and (b) B3 and the mutated sequences (c) B1Y and (d) B3Y for the case of the spread OH defects. A value of approximately 0° indicates an orientation parallel with the surface.

the OH defects have the smaller residence times. In addition, B3Y, which has slightly fewer atoms interacting with the OH groups than B1, has a slightly larger residence time. Again, these data provide evidence that the presence of OH defects interacting with the peptide atoms can affect the mobility of the peptide. At first sight, it may seem that in the case of the spread defects, there is no clear reason for the differences in the residence times. However, a closer inspection of the data in Table 6.7 helps us to understand the origin of such smaller residence times. Despite these peptides (in the spread defects) having fewer atoms in total interacting with the OH defects, the

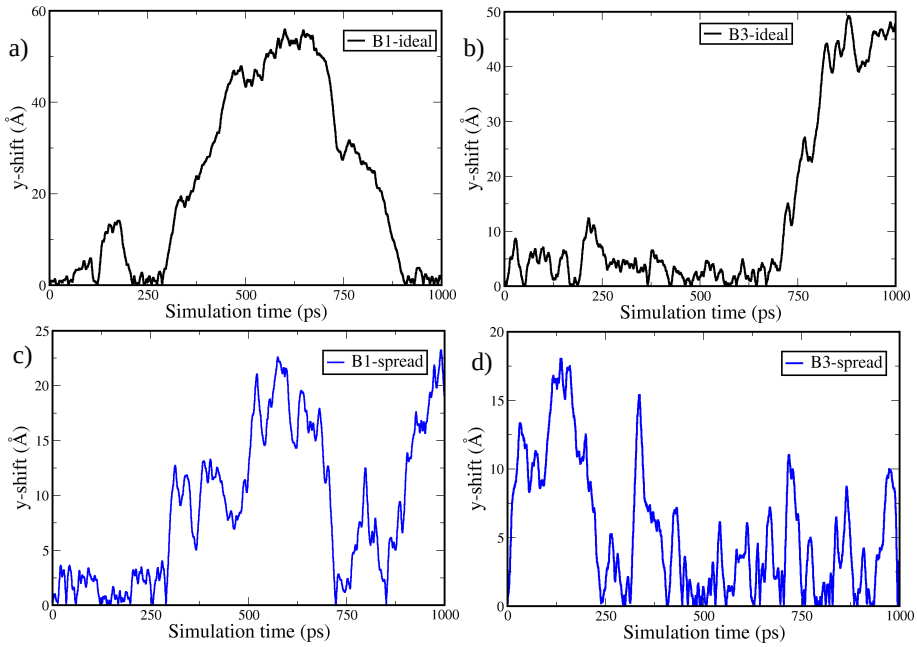

Figure 6.14: y-shift distance as a function of time for (a) B1 and (b) B3 at the ideal CNT and (c) B1 and (d) B3 at the spread defects CNT.

frequency at which it occurs was higher, meaning that these atoms spent more time interacting with the OH defects. For instance, B3Y in the case of the spread defects has one peptide oxygen atom that interacted 12% of the time with OH-1. On the other hand, the highest frequency in the case of B3Y in the bunched defects is 3.0% (carbon interacting with OH-2). So, even though B3Y in the spread defects has fewer atoms interacting with the OH defects, the residence time is comparable to that of B3Y in the bunched defects case (2.40 vs 2.55 ps , respectively), due to the fact they interact with the defects for a longer time. Similarly, despite B1 in the bunched case having more peptide atoms interacting with the OH defects, the frequency is much lower than in the case of B1 in the spread defects (4% vs 11.5%, respectively), which possibly explains the lower residence time for B1 in the spread defects (2.30 and 2.05 ps, respectively).

Table 6.7: The three closest peptide atoms (within the 8.0 Å cutoff) to each of the OH defects during the production run, for both the bunched case and the spread case. — means that no atom was found within the cutoff distance. For each OH defect site the three closest atoms, the residue (*Res.*) to which it belongs to, the *type* (sidechain *S*, or backbone *B*) and the frequency (%) are given.

Bunched defects													
Pept.	Site	Atom	Res.	Type	%	Atom	Res.	Type	%	Atom	Res.	Type	%
B1	OH-1	C	**H1**	*S*	1.5	N	**H1**	*S*	1.4	N	**K3**	*S*	1.0
	OH-2	C	**H1**	*S*	4.0	N	**H1**	*S*	2.0	N	**K3**	*S*	1.3
	OH-3	C	**H1**	*S*	3.0	C	**H1**	*S*	2.0	N	**H1**	*S*	2.0
	OH-4	N	**H1**	*S*	4.0	C	**H1**	*S*	1.6	C	**H1**	*S*	0.75
B3	OH-1	—				—				—			
	OH-2	—				—				—			
	OH-3	C	**S12**	*S*	0.6	O	**S12**	*S*	0.1	—			
	OH-4	C	**S12**	*S*	0.55	—				—			
B1Y	OH-1	—				—				—			
	OH-2	—				—				—			
	OH-3	—				—				—			
	OH-4	—				—				—			
B3Y	OH-1	C	**S12**	*S*	0.3	—				—			
	OH-2	C	**S12**	*S*	3.0	—				—			
	OH-3	C	**S12**	*S*	2.2	O	**S12**	*S*	0.4	N	**S12**	*S*	0.1
	OH-4	C	**S12**	*S*	1.4	O	**S12**	*S*	0.05	N	**S12**	*S*	0.05
Spread defects													
Pept.	Site	Atom	Res.	Type	%	Atom	Res.	Type	%	Atom	Res.	Type	%
B1	OH-1	—				—				—			
	OH-2	—				—				—			
	OH-3	N	**H4**	*S*	11.5	C	**H1**	*S*	9.5	C	**H4**	*S*	7.8
	OH-4	C	**W9**	*S*	11.3	C	**W9**	*S*	3.0	C	**K3**	*S*	2.6
B3	OH-1	C	**S12**	*S*	7.7	C	**W6**	*S*	2.2	O	**S12**	*S*	1.5
	OH-2	C	**S12**	*S*	13.0	C	**W6**	*S*	1.5	N	**S12**	*S*	0.15
	OH-3	—				—				—			
	OH-4	—				—				—			
B1Y	OH-1	—				—				—			
	OH-2	—				—				—			
	OH-3	N	**H1**	*S*	7.6	C	**H1**	*S*	6.0	N	**H1**	*S*	0.3
	OH-4	—				—				—			
B3Y	OH-1	O	**S3**	*B*	12.0	C	**A4**	*S*	3.0	C	**R8**	*S*	2.0
	OH-2	—				—				—			
	OH-3	—				—				—			
	OH-4	—				—				—			

6.5 Discussion

As pointed out before, this work represents the first attempt at studying the effect of disrupting the CNT ideality on the binding of peptides onto the CNT surface. These findings suggest that the presence of defects has a small effect on the binding affinity of peptides to CNTs. However, these data suggest that the peptide mobility is increased by the presence of OH defects on the CNT. It is not surprising that these OH defects only present a physical barrier to the peptide adsorption, since the distance that the defect group extends from the surface is small (compared for example with a COOH group), such that it causes only a minor perturbation. This is in agreement with previous studies where the heterogeneity of graphite was modelled by either fixing single water molecules [216] or by attaching different functional groups (carbonyl, hydroxyl, and carboxyl groups) [213; 216; 215] to the surface. From the study of Do *et al.* [216] it was concluded that the adsorption of water that resulted from the presence of carbonyl and hydroxyl groups was very small and that the adsorption due to the carboxyl groups was comparable to that of two fixed water molecules. These differences were attributed to the distance that each group extends away from the surface, which is in agreement with the weak "adsorbing power" of the OH defects used in this study. In addition, recent studies also agree with this finding [216], in that defects which extend further from the surface, such as a COOH group, can have a major influence on the adsorption of molecules [215; 272]; certainly more than the OH groups have. The weak attraction of this OH defect to polar molecules can be inferred from the weak interaction energy between the defective CNT and, for example, a water molecule. As shown in Table 6.4, the interaction energy between a water molecule and an ideal CNT, calculated with this force-field is -2.20 kcal/mol, while the interaction energy for a water molecule in the same geometry and a CNT with one OH defect is -4.57 kcal/mol. In addition, as

shown in Table 5.1 of Chapter 5, the position of the first peak of the density profile (in the ideal CNT) is 3.16 Å for oxygen and 2.96 Å for hydrogen. Since the distance that the OH defect projects from the surface is 1.36 Å (C—O distance Table 6.1), it is clear that the water molecules cannot approach close enough to engage with the defect.

Even though it seems that the OH defects have not significantly affected the peptide-CNT interaction energy, in this study our defect concentrations corresponded to the concentration in high-quality CNTs, which are thought to contain on average one defect per 4μm [237]. However, these "almost pure" CNTs are very expensive, and in many applications lower-quality CNTs are used; therefore the effect of the defects might be greater. To gain an insight into the OH defect density effect, a number of preliminary simulations were also performed with 15 OH spread defects attached to the exterior of the CNT (of the same dimensions to those previously used in this work), for the original sequence B1 and for the mutated sequences B1Y, B3Y and B1F. In this system we therefore have increased the density concentration by almost four-fold. The geometries of the best trajectories of the simulations in Chapter 3 were used as initial configurations for these simulations with 15 OH spread defects. In Table 6.8 the normalised interaction energies, E_N, and the peptide residence time, τ averaged over a 1 ns production run are presented. The original sequence B1 does not seem to be affected by the density of the OH defects, as evidenced by the greater binding affinity, which is comparable to the interaction energy at the ideal CNT (Table 6.5). In the case of the mutated sequences, B1Y and B3Y, the binding affinity is slighly weaker than that reported for 4 OH defects (both bunched and spread), which suggests that the the binding affinity is affected by the presence of OH defects. Even though the mutated sequence B1F, has not been studied in this Chapter, it also seems that the the binding affinity is slightly affected by the presence of the OH defects (see Table 3.3 in Chapter 3). The peptide residence

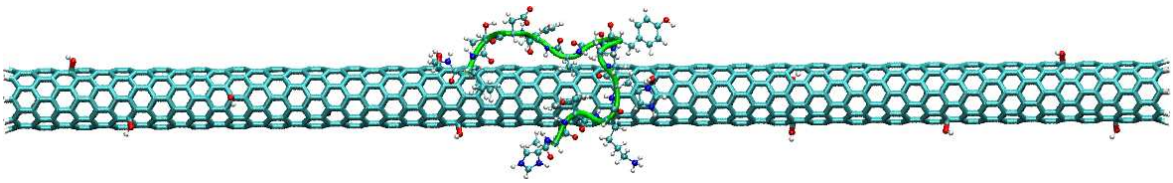

Figure 6.15: Snapshot taken from the trajectory of the B1Y peptide interacting with a CNT with 15 OH spread defects. The peptide backbone is highlighted.

time is also slightly lower for B3Y and B1Y in this case, than in the case of the 4OH defects. However, in the case of B1, the residence time is lower than in the case of the 4 OH bunched defects and higher than in the case of the 4 OH spread defects (Table 6.6). It is stressed that this is just a preliminary study, and further simulations would have to be performed in order to draw conclusions about the consequences of higher OH defect concentrations. However, these data suggest that the density of OH defects affects the binding affinity of peptides to CNTs. In Figure 6.15 a snapshot of the B1Y peptide interacting with a CNT with 15 OH spread defects is shown.

Table 6.8: Normalised interaction energies, E_N (kJ mol^{-1}atom^{-1}) and peptide residence time (ps), averaged over a 1 ns production run for the peptide—CNT interface in the case of the CNT with 15 OH defects spread over the surface.

Aptamer	E_N	τ
B1	-1.38	3.30
B3Y	-1.09	2.20
B1Y	-1.06	2.05
B1F	-0.998	3.10

Another important aspect in the adsorption of peptides onto defective CNTs is the distribution of the defects. From an experimental point of view, it is currently impossible to selectively study the influence of the type, concentration or distribution of

defects, since it would be challenging to isolate each of these variables in the laboratory. This is why molecular simulation constitutes an imperative tool for this study. The work of Seaton *et al.* [213] provides a relevant investigation of the separate effects of density, distribution and type of defects on the water adsorption of activated carbon pores. The distance between the defects plays an important role in water adsorption, as well as the orientation of the OH defects.The optimum orientation of the OH groups was found to be the one in which the hydrogens are pointing towards each other (converging hydrogens) since water molecules can fit between the defects [213]. In addition, the presence of closely located defects causes a stronger hydrophilic effect than isolated defects, even if the total density of isolated defects is higher. Surprisingly, the findings of Seaton *et al.* [213] suggested that the amount of adsorbed water is mainly affected not by the type of defect, but by the number of oxygen atoms on the surface, since the amount of water adsorbed at a COOH group is similar to the amount of water adsorbed at two close carbonyls. Based on these findings, it appears important to investigate the effect of the distribution of defects on the binding affinity of peptides onto CNTs, along with the study of the effect of the defect density.

This previous paragraph leads us to a very important topic which has been subject of much research, namely water. Currently, no implict solvent model can accurately describe microscopic water effects (both structural and dynamic), no matter how good such a model is. While in our previous work [120], comparison of the structural properties of a strong binding peptide at an ideal CNT in both implicit and explicit solvent lead us to conclude that the use of an implicit model in that case was reasonable, we contend that in the case of a defective CNT, the situation might be different since the OH defects might increase the hydrophilicity of the CNT. The consequences of a more hydrophilic surface is that water adsorption is influenced by the presence of the polar groups on the

surface. As discussed in Chapter 5, the adsorption of water at pure graphitic surfaces is weak, however as several studies suggested, at activated carbons the adsorption process can be significantly affected [256; 210; 211; 212; 213; 265; 215]. However, recent studies suggest that in the case of the OH defect, the interaction between water molecules and OH defects is quite weak compared to the water-water interaction, and as a consequence the water molecules (adsorbed in finite numbers) tend to aggregate above the surface instead of being trapped (unless the temperature is low enough) by the OH sites [265]. In addition, in our own studies we have shown that the oxygen density profile around the OH defect (Chapter 5) also indicates a weak structuring around the defect. On the other hand, in the case of the COOH defect, studies show that this group exhibits a stronger affinity for water molecules and is therefore able to attract more water molecules [215]. Even though the previous studies indicate that graphitic surfaces containing OH sites behave similarly to the ideal surfaces [262; 215], none of these studies have simulated liquid water at the interface. This is why further studies should be performed in order to investigate whether or not the inclusion of explicit water will significantly affect our findings.

In terms of the peptide sequences studied in the present Chapter, it was expected that B1Y and B3Y would show a preference for binding at the OH defects, due to the presence of the tyrosine residues that also have an OH included in the phenol group. In addition, B3 has more polar groups than B1, thus B3Y is the most polar of all sequences and therefore, it was expected that it would show a preference for the OH defects. These results suggest that B1Y and B3Y (especially B3Y) in the defective CNTs are more flexible and suggest that the defects might perturb the tyrosines (even if only locally). But, the inclusion of explicit water may alter this. In addition, it could be expected that the original sequences, B1 and B3, would be affected by the presence of

the OH defects in terms of binding affinity, since from Table 6.4 the interaction between benzene and a CNT with one OH defect is slightly stronger than that at an ideal CNT. However, the difference in interaction energy between the two cases is so small (-1.79 vs -2.97 kcal/mol) that at 298 K this will not make a difference. This is why the interaction energies of B1 and B3 at the defective CNTs (Table 6.5) do not reflect such increases in the binding affinity. However, the peptide residence times at the defective CNTs suggest a greater mobility of these peptides than at the ideal CNT. In addition, tryptophan has one of the closest atoms to one of the OH defects for B1 and B3 in the spread case (Table 6.7).

6.6 Summary and Outlook

The binding behaviour of several peptide sequences adsorbed at two defective CNTs were investigated by molecular dynamics simulations. Four OH defects in two different arrangements (bunched together and spread over the CNT) were considered to break the ideality of the CNT. Results show that the binding affinity was not significantly affected by the presence of the OH defects. In addition, it was found that the original sequences, B1 and B3, have the greatest affinity to the CNT regardless of the presence or absence of OH defects. However, results showed a general increase in the peptide mobility resulted from the presence of the OH defects. These findings should give insights for the applications of real CNTs with peptides. In the future it would be interesting to explore the effect of Stone-Wales, vacancy defects and charged defects on the binding affinity of peptides onto CNTs. It would also be interesting to explore the effect of having mixtures of different defects on the CNT and its effects on the peptide binding affinity. In addition, a further study to evaluate the effects of OH defects on the binding affinity to peptides in explicit water is also necessary.

Chapter 7

Conclusions and Outlook

The main aim of this thesis was to study the interactions of peptides with graphitic surfaces in order to aid our understanding of the general mechanisms that govern peptide-inorganic recognition. Specifically, the goal was to investigate peptide-graphitic surface interfaces with a view to establishing the rules of design that give peptide sequences controllable affinity to graphitic surfaces. It is imperative to understand the mechanisms that dictate peptide-inorganic interactions so these can be predicted and controlled for the desired application.

This thesis focused mainly on carbon nanotubes (CNTs) due to their wide range of current and potential applications in the fields of nanotechnology and medicine. For instance the variety of biomedical applications such as drug [72; 73], gene [76] and peptide [74] delivery vehicles makes CNTs an attractive object of research. These applications are, curently limited, due to the fact that CNTs are hydrophobic and are very difficult to disperse. Moreover, the use of CNTs still raises concerns associated with the potential impacts on human health [69] and environmental safety [59]. The ability of peptides to disperse and solubilise CNTs have opened up new prospects for

the implementation of a wide range of applications. In addition, the ability to control the strength of binding similarly gives such promise.

Part of this thesis is based on the experimental work of Wang *et al.* [94] which suggested two peptide sequences, B1 (HWKHPWGAWDTL) and B3 (HWSAWWIRSNQS), both of which having strong affinity for CNTs. These authors showed the important role of the aromatic content (histidine and tryptophan) in the sequences found for the strong binding to CNTs. In our previous work [118], where we performed molecular dynamics (MD) simulations of B1 and B3 and a weak-binder control peptide with a single-walled CNT, we had already confirmed the experimental observations [94] of the relative binding affinity for the sequences studied. In order to study the importance of the aromatic content, mutations of the tryptophan content by the aromatic residues tyrosine and phenylalanine on B1 and B3 were performed. In agreement with recent studies [102; 101], the results presented in Chapter 3 showed that none of the peptide mutants exceeded the original sequences in terms of binding affinity. In addition, none of the mutants reached the lower range of binding affinity of the weak-binder control peptide. These differences in binding affinity were attributed to the greater stability of the indole ring in tryptophan. Furthermore, the interfacial shape was also studied by performing similar investigations with the same peptides adsorbed on a graphene sheet. It was found that a compromise existed between maintaining all the aromatic residues close to the graphene surface and also allowing the non-aromatic residues to approach the surface. This suggests that these peptide sequences not only bind strongly to a graphitic surface, but bind better at a *curved* graphitic surface. Therefore, the results presented in Chapter 3 suggest that these peptide sequences may have been selected for the interfacial shape, as well as interfacial chemistry. In addition, the results presented in Chapter 3 gave a glimpse that peptide composition is an important factor in determining strong binding;

but this might not be the only factor. This idea was pursued in more detail in Chapter 4, where the influence of the peptide sequence on the binding affinity to CNTs was studied. It was demonstrated that the binding affinity of a peptide sequence with fixed composition could be modified by mixing the peptide sequence. These findings show that the sequence plays an important role in controlling the strength of the peptide binding affinity to CNTs and if experimentally tested, it should be detectable. Furthermore, these findings constitute the first step to understanding the design rules for peptide-CNT interfaces. For example, it was found that strong binding may be due to the presence of isolated pair of tryptophans, while weaker binding may be due to the presence of two tryptophan residues intercalated by another residue.

Chapter 5 was dedicated to the study of the behaviour of water at graphitic surfaces – CNTs, graphite and fullerenes. This study was necessary in order to understand if water structuring at the interface would affect the binding affinity of peptides. It was found that water structuring around the graphitic surfaces is weak. In addition, this study aimed to validate the extended polarisable force-field AMOEBAPRO (which includes a description of the non-bonded interactions between water and graphitic surfaces) in order to be able to use it in future simulations of peptides with graphitic surfaces in aqueous solution. Finally, in Chapter 6, the influence of surface defects on the binding affinity between peptides and CNTs was investigated. Again, this study is important since real CNTs are not ideal pristine structures. Understanding how defects affect the binding affinity will help to rationalise the mechanisms that govern the interaction between peptides and real CNTs. Results showed that the presence of OH defects did not significantly affect the binding affinity; however an increase in the peptide mobility was noticed. The weak effect was proposed to be related to the distance that the OH group extends from the surface. For example, a COOH group is proposed

to have a greater effect on the binding affinity, since this group can extend a greater distance from the surface (compared to the OH group) [272].

The work presented in this thesis has helped to understand the nature of peptide-CNT interactions, why composition (in this case the aromatic content) is important for the binding affinity and why the sequence (which gives the peptide conformation) also plays a very important role. These findings highlight the importance of motifs rather than isolated residues, since motifs can modulate the local environment around the key residues that participate in the binding. Therefore, the work described in this thesis opens up new *vistas* into the decoding of the mechanisms that control molecular binding at peptide–inorganic interfaces. These findings should be further explored *via* both simulations and experiments; for instance, the peptide sequences studied here could be experimentally tested and characterised. In addition, the key motifs responsible for strong and weak binding could be isolated and studied. There is also scope for further simulations of these peptides in explicit solvent and for using the knowledge gained in this thesis to design novel peptides with controllable affinity to graphitic surfaces.

Bibliography

[1] M. Sarikaya, C. Tamerler, A. K. Y. Jen, K. Schulten, and F. Baneyx. *Nat. Mater*, 2:577, 2003.

[2] C. P. Ponting and R. R. Russell. *Annu. Rev. Biophys. Biomol. Struct.*, 31:45, 2002.

[3] D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry*. Worth, University of Wisconsin-Madison, 3 edition, 2000.

[4] J. L. Bada and A. Lazcano. *Science*, 296:1982, 2002.

[5] P. Ball. *Nature*, 409:413, 2001.

[6] M. Sarikaya, C. Tamerler, D. T. Schwartz, and F. Baneyx. *Annu. Rev. Mater. Res.*, 34:373, 2004.

[7] S. Brown. *Nature Biotechnology*, 15:269, 1997.

[8] M. Hnilova, E. E. Oren, U. O. S. Seker, B. R. Wilson, S. Collino, J. S. Evans, C. Tamerler, and M. Sarikaya. *Langmuir*, 24:12440, 2008.

[9] H. Heinz, B. L. Farmer, R. B. Pandey, J. M. Slocik, S. S. Patnaik, R. Pachter, and R. R. Naik. *J. Am. Chem. Soc.* , 131:9704, 2009.

[10] R. R. Naik, S. J. Stringer, G. Agarwal, S. E. Jones, and M. O. Stone. *Nat. Mater.*, 1:169, 2002.

[11] R. R. Naik, S. E. Jones, C. J. Murray, J. C. McAuliffe, R. A. Vaia, and M. O. Stone. *Adv. Funct. Mater.*, 14:25, 2004.

[12] U. O .S. Seker, B. Wilson, S. Dincer, I. W. Kim, E. E. Oren, J. S. Evans, C. Tamerler, and M. Sarikaya. *Langmuir*, 23:7895, 2007.

[13] W. S. Choe, M. S. R. Sastry, C. K. Thai, H. Dai, D. T. Schwartz, and F. Baneyx. *Langmuir*, 23:11347, 2007.

[14] S. R. Whaley, D. S. English, E. L. Hu, P. F. Barbara, and A. M. Belcher. *Nature*, 405:665, 2000.

[15] B. R. Peelle, E. M. Krauland, K. D. Wittrup, and A. M. Belcher. *Langmuir*, 1:145, 2005.

[16] K. I. Sano and K. Shiba. *J. Am. Chem. Soc.*, 125:14234, 2003.

[17] H. Chen, X. Su, K-G Neoh, and W-S Choe. *Anal. Chem.*, 78:4872, 2006.

[18] C. K. Thai, H. Dai, M. S. R. Sastry, M. Sarikaya, D. T. Schwarwtz, and F. Baneyx. *Biotechnol. Bioeng.*, 87:129, 2004.

[19] M. Gungormus, H. Fong, II W. Kim, J. S. Evans, C. Tamerler, and M. Sarikaya. *Biomacromolecules*, 9:966, 2008.

[20] M. A. Schembri, K. Kjaergarrd, and P. Klemm. *FEMS Microbiol. Lett.*, 170:363, 1999.

[21] C. Tamerler and M. Sarikaya. *MRS Bull.*, 33:504, 2008.

[22] C. Tamerler and M. Sarikaya. *ACS Nano*, 3:1606, 2009.

[23] C. Tamerler, M. Duman, E. E. Oren, M. Gungormus, X. Xong, T. Kacar, B. A. Parviz, and M. Sarikaya. *Small*, 2:1372, 2006.

[24] C. Tamerler and M. Sarikaya. *Acta Biomaterialia*, 3:289, 2007.

[25] C. Tamerler, E. E. Oren, M. Duman, E. Venkatasubramanian, and M. Sarikaya. *Langmuir*, 22:7712, 2006.

[26] U. O. S. Seker, B. Wilson, D. Sahin, C. Tamerler, and M. Sarikaya. *Biomacromolecules*, 10:250, 2009.

[27] C. R. So, C. Tamerler, and M. Sarikaya. *Angew. Chem. Int. Ed.*, 48:5174, 2009.

[28] D. C. Cullen and C. R. Lowe. *J. Colloid Interface Sci.*, 166:102, 1994.

[29] A. Rodger and B. Norden. *Circular Dichroism and Linear Dichroism*. Oxford University Press, Oxford, UK, 1997.

[30] S. Brahms and J. Brahms. *J. Mol. Biol.*, 138:149, 1980.

[31] C. M. Niemeyer. *Angew. Chem. Ind. Ed.*, 40:4128, 2001.

[32] E. E. Oren, C. Tamerler, D. Sahin, M. Hnilova, U. O.S. Seker, M. Sarikaya, and R. Samudrala. *Bioinformatics*, 23:2816, 2007.

[33] H. Dai, W. S. Choe, C. K. Thai, M. Sarikaya, B. A. Traxler, F. Baneyx, and D. T. Schwartz. *J. Am. Chem. Soc.* , 127:15637, 2005.

[34] T. Kacar, M. T. Zin, C. So, B. Wilson, H. Ma, N. G. Karaguler, A. K. Y. Jen, M. Sarikaya, and C. Tamerler. *Biotechnol. Bioeng.*, 103:696, 2009.

[35] C. Tamerler and M. Sarikaya. *Philos. T. Roy. Soc. A*, 367:1705, 2009.

[36] M. T. Zin, K. Leong, N. Y. Wong, H. Ma, M. Sarikaya, and A. K. Y. Jen. *Nanotechnology*, 20:15305, 2009.

[37] S. Brown, M. Sarikaya, and E. Johnson. *J. Mol. Biol.*, 299:725, 2000.

[38] M. Cretich, F. Damin, G. Pirri, and M. Chiari. *Biomol. Eng.*, 23:77, 2006.

[39] Y. Kwon, M. A. Coleman, and J. A. Camarero. *Angew. Chem. Int. Ed.*, 45:1726, 2006.

[40] F. Rusmini, Z. Zhong, and J. Feijen. *Biomacromolecules*, 8:1775, 2007.

[41] A. Watzke, M. Kohn, M. G. Rodriguez, R. Wacker, H. Schroder, R. Breinbauer, J. Kuhlmann, K. Alexandrov, C. M. Niemeyer, R. S. Goody, and H. Waldmann. *Angew. Chem. Int. Ed.*, 45:1408, 2006.

[42] D. G. Castner and B. D. Ratner. *Surf. Sci.*, 500:28, 2002.

[43] U. T. Bornscheuer. *Angew. Chem. Int. Ed.*, 42:3336, 2003.

[44] T. Kacar, J. Ray, M. Gungormus, E. E. Oren, C. Tamerler, and M. Sarikaya. *Adv. Mater.*, 21:295, 2009.

[45] M. T. Zin, H. Ma, M. Sarikaya, and A. K. Y. Jen. *Small*, 1:698, 2005.

[46] J. C. T. Carlson, S. S. Jena, M. Flenniken, T. F. Chou, Siegel R. A, and C. R. Wagner. *J. Am. Chem. Soc.* , 128:7630, 2006.

[47] T. Li, L. Wu, N. Suthiwangcharoen, M. A. Bruckman, D. Cash, J. S. Hudson, S. Ghoshroy, and Q. Wang. *Chem. Commun.*, 20:2869, 2009.

[48] H. Li, J. D. Carter, and T. H. LaBean. *Mater. Today*, 12:24, 2009.

[49] S. Fuchs, S. Ghanaati, C. Orth, M. Barbeck, M. Kolbe, A. Hofmann, M. Eblenkamp, M. Gomes, R. L. Reis, and C. J. Kirkpatrick. *Biomaterials*, 30:526, 2009.

[50] Y. Fang, N. Poulsen, M. B. Dickerson, Y. Cai, S. E. Jones, R. R. Naik, N. Kroger, and K. H. Sandhage. *J. Mater. Chem.*, 18:3871, 2008.

[51] K. Y. Lee and D. J. Mooney. *Chem. Rev.*, 101:1869, 2001.

[52] A. C. Jen, C. Wake, and A. G. Mikos. *Biotech. Bioeng.*, 50:357, 1996.

[53] J. E. Dover, G. M. Hwang, E. H. Mullen, B. C. Prorok, and S. J. Suh. *J. Microbial. Biotech.*, 78:10, 2009.

[54] P. V. Bower, E. A. Louie, J. R. Long, P. S. Stayton, and G. P. Drobny. *Langmuir*, 21:3002, 2005.

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

[56] S. Iijima. *Nature*, 354:56, 1991.

[57] Y. Ando and S. Iijima. *Jpn. J. Appl. Phys.*, 32:107, 1993.

[58] L. C. Qin. *Phys. Chem. Chem. Phys.*, 9:31, 2007.

[59] W. Yang, P. Thordarson, J. J. Gooding, S. P. Ringer, and F. Braet. *Nanotechnology*, 18:412001, 2007.

[60] F. Lu, L. Gu, M. J. Meziani, X. Wang, P. G. Luo, L. M. Veca, L. Cao, and Y. P. Sun. *Adv. Mater.*, 21:139, 2009.

[61] Z. Liu, S. M. Tabakman, Z. Chen, and H. Dai. *Nat. Protoc.*, 4:1372, 2009.

[62] D. J. Cooke, D. Eder, and J. A. Elliott. *J. Phys. Chem. C*, 114:2462, 2010.

[63] H. Wang. *Curr. Opin. Colloid In.*, 14:364, 2009.

[64] W. Huang, S. Fernando, Y. Lin, B. Zhou, L. F. Allard, and Y.-P. Sun. *Langmuir*, 19:7084, 2003.

[65] Y. Han and J. Elliott. *Comp. Mater. Sci.*, 39:315, 2007.

[66] K. A. Williams, P. T. M. Veenhuizen, B. G. de la Torre, R. Eritja, and C. Dekker. *Nature*, 420:761, 2002.

[67] K. Awasthi, D. P. Singh, S. K. Singh, D. Dash, and O. N. Srivastava. *New Carbon Mater.*, 24:301, 2009.

[68] C. A. Poland, R. Duffin, I. Kinloch, A. Maynard, W. A. H. Wallace, A. Seaton, V. Stone, S. Brown, W. Macnee, and K. Donaldson. *Nat. Nanotechol.*, 3:423, 2008.

[69] C. P. Firme III and P. R. Bandaru. *Nanomedicine*, 6:245, 2010.

[70] A. Albini, V. Mussi, A. Parodi, A. Ventura, E. Principi, S. Tegami, M. Rocchia, E. Francheschi, I. Sogno, R. Cammarota, G. Finzi, F. Sessa, D. M. Noonan, and U. Valbusa. *Nanomedicine*, 6:277, 2010.

[71] E. L. Bakota, L. Aulisa, D. A. Tsyboulski, R. B. Weisman, and J. D. Hartgerink. *Biomacromolecules*, 10:2201, 2009.

[72] A. Bianco, K. Kostarelos, and M. Prato. *Curr. Opin. Chem. Biol.*, 9:674, 2005.

[73] Y. Cheng, Q. X. Pei, and H. Gao. *Nanotechnology*, 20:145101, 2009.

[74] D. Pantarotto, J. P. Briand, M. Prato, and A. Bianco. *Chem. Commun.*, 7:16, 2004.

[75] N. S. S. Kam, T. C. Jessop, P. A. Wender, and H. Dai. *J. Am. Chem. Soc.* , 126:6850, 2004.

[76] N. W. S. Kam, Z. Liu, and H. Dai. *Angew. Chem. Int. Ed.*, 45:577, 2006.

[77] N. W. S. Kam, Z. Liu, and H. Dai. *J. Am. Chem. Soc.* , 127:12492, 2005.

[78] M. Zhang, T. Murakami, K. Ajima, K. Tsuchida, A. S. D. Sandanayaka, O. Ito, S. Iijima, and M. Yudasaka. *Proc. Natl. Acad. Sci. U.S.A.*, 105:14773, 2008.

[79] F. Zhou, D. Xing, B. Wu, S. Wi, Z. Ou, and W. R. Chen. *Nano Lett.*, 10:1677, 2010.

[80] S. Liu, Li. Wei, L. Hao, N. Fang, M. W. Chang, R. Xu, Y. Yang, and Y. Chen. *ACS Nano*, 3:3891, 2009.

[81] X. Chen, U. C. Tam, J. L. Czlapinski, G. S. Lee, D. Rabuka, A. Zettl, and C. R. Bertozzi. *J. Am. Chem. Soc.* , 128:6292, 2006.

[82] A. Kis and A. Zettl. *Phil. Trans. R. Soc. A*, 366:1591, 2008.

[83] R. L. Willett, K. L. Baldwin, K. W. West, and L. N. Pfeiffer. *Proc. Natl. Acad. Sci. USA*, 102:7817, 2005.

[84] K. Goede, P. Busch, and M. Grundmann. *Nano Lett*, 4:2115, 2004.

[85] B. R. Pelle, E. M. Krauland, K. D. Wittrup, and A. M. Belcher. *Langmuir*, 21:6929, 2005.

[86] E. E. Oren, C. Tamerler, and M. Sarikaya. *Nano. Lett.*, 5:415, 2005.

[87] J. J. Gray. *Curr. Opin. Struct. Biol.*, 14:110, 2004.

[88] V. Zorbas, A. L. Smith, H. Xie, A. O.-Acevedo, A. B. Dalton, G. R. Dieckmann, R. K. Draper, R. H. Baughman, and I. H. Musselman. *J. Am. Chem. Soc.* , 127:12323, 2005.

[89] T. Hayashi, K. I. Sano, K. Shiba, Y. Kumashiro, K. Iwahori, I. Yamashita, and M. Hara. *Nano Lett.*, 6:515, 2006.

[90] X. Li, W. Chen, Q. Zhan, L. Dai, L. Sowards, M. Pender, and R. R. Naik. *J. Phys. Chem. B*, 110:12621, 2006.

[91] T. Hayashi, K. I. Sano, K. Shiba, K. Iwahori, I. Yamashita, and M. Hara. *Langmuir*, 25:10901, 2009.

[92] N. F. Breen, T. Weidner, K. Li, D. G. Castner, and G. P. Drobny. *J. Am. Chem. Soc.* , 131:14148, 2009.

[93] G. R. Dieckmann, A. B. Dalton, P. A. Johnson, J. Razal, J. Chen, G. M. Giordano, E. Munoz, I. H. Musselman, R. H. Baughman, and R. K. Draper. *J. Am. Chem. Soc.*, 125:1770, 2003.

[94] 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, and A. Jagota. *Nat. Mater.*, 2:196, 2003.

[95] Z. Su, T. Leung, and J. F. Honek. *J. Phys. Chem. B.*, 110:23623, 2006.

[96] S. Brown, T. S. Jespersen, and J. Nygard. *Small*, 4:416, 2008.

[97] L. Zheng, D. Jain, and P. Burke. *J. Phys. Chem. C*, 113:3978, 2009.

[98] Y. Morita, T. Oshugi, Y. Isawa, and E. Tamiya. *J. Mol. Catal. B-Enzym*, 28:185, 2004.

[99] D. Kase, J. L. Kulp III, M. Yudasaka, J. S. Evans, S. Iijima, and K. Shiba. *Langmuir*, 20:8939, 2004.

[100] R. H. Baughman, A. A. Zakhidov, and W. A. de Heer. *Science*, 297:787, 2002.

[101] H.Xie, E. J. Becraft, R. H. Baughman, A. B. Dalton, and G. R. Dieckmann. *J. Pept. Sci.*, 14:139, 2008.

[102] C. G. Salzmann, M. A. H. Ward, R. M. J. Jacobs, G. Tobias, and M. L. H. Green. *J. Phys. Chem. C*, 111:18520, 2007.

[103] R. J. Chen, Y. Zang, D. Wang, and H. Dai. *J. Am. Chem. Soc.*, 123:3838, 2001.

[104] V. Z. Poenitzsch, D. C. Winters, H. Xie, G. R. Dieckmann, A. B. Dalton, and I. H. Musselman. *J. Am. Chem. Soc.* , 129:14724, 2007.

[105] J. S. Evans, R. Samudrala, T. R. Walsh, E. E. Oren, and C. Tamerler. *MRS Bull.*, 33:514, 2008.

[106] J. Zhou, S. Chen, and S. Liang. *Langmuir*, 19:3472, 2003.

[107] M. Yang, P. M. Rodger, J. H. Harding, and Stipp S. L. S. *Mol. Simulat.*, 35:547, 2009.

[108] E. E.Oren, C. Tamerler, and M. Sarikaya. *Nano. Lett.*, 5:415, 2005.

[109] A. A. Mungikar and D. Forciniti. *Biomacromolecules*, 5:2147, 2004.

[110] R. B. Pandey, H. Heinz, J. Feng, B. L. Farmer, J. M. Slocik, L. F. Drummy, and R. R. Naik. *Phys. Chem. Chem. Phys*, 11:1989, 2009.

[111] D. Antypov and J. A. Elliott. *Macromolecules*, 41:7243, 2008.

[112] N. Kantarci, C. Tamerler, M. Sarikaya, T. Haliloglu, and P. Doruker. *Polymer*, 46:4307, 2005.

[113] V. Carravetta and S. Monti. *J. Phys. Chem. B*, 110:6160, 2006.

[114] Y. Cheng, G. R. Liu, Z. R. Li, and C. Lu. *Physica A*, 367:293, 2006.

[115] Z. Yang and Y-P Zhao. *Engineering Analysis with Boundary Elements*, 31:402, 2007.

[116] A. V. Verde, J. M. Acres, and J. Maranas. *Biomacromolecules*, 10:2118, 2009.

[117] A. A. Skelton, T. Liang, and T. R. Walsh. *ACS Applied Materials Interfaces*, 1:1482, 2009.

[118] S. D. M. Tomásio and T. R. Walsh. *Mol. Phys.*, 105:221, 2007.

[119] Z. Su, K. Mui, E. Daub, T. Leung, and J. Honek. *J. Phys. Chem. B.*, 111:14411, 2007.

[120] S. M. Tomásio and T. R. Walsh. *J. Phys. Chem. C*, 113:8778, 2009.

[121] H. D. Holtje, W. Sippl, D. Rognan, and G. Folkers. *Molecular Modeling: Basic Principles and Applications*. Wiley-VCH, 3 edition, 2008.

[122] A. R. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2nd edition, 2001.

[123] H. B. Schlegel. *Modern Electronic Structure Theory*. World Scientific Publishing, Singapore, 1995.

[124] I. R. Levine. *Physical Chemistry*. McGraw-Hill, New York, 4th edition, 1995.

[125] D. A. McQuarrie. *Statistical Mechanics*. Harper and Row, New York, 1976.

[126] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, University of Amesterdam, Netherlands, 2nd edition, 2002.

[127] F. Jensen. *Introduction to Computational Chemistry*. Wiley, University of Southern Denmark, Denmark, 2 edition, 2007.

[128] M. P. Allen and D. J. *Computer Simulation of Liquids*. Oxford University Press, University of Oxford, 1987.

[129] L. Verlet. *Phys. Rev.*, 159:98, 1967.

[130] W. G. Hoover. *Phys. Rev. A*, 31:1695, 1985.

[131] D. J. Wales. *Energy Landscapes*. Cambridge, University of Cambridge, 2003.

[132] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136:864, 1964.

[133] A. J. Stone. *The Theory of Intermolecular Forces*. Oxford, U.K., 1997.

[134] P. Ren and J. W. Ponder. *J. Comput. Chem.*, 23:1497, 2002.

[135] T. A. Halgren. *J. Am. Chem. Soc.* , 114:7827, 1992.

[136] B. Chen, G. Martin, and J. I. Siepmann. *J. Phys. Chem. B*, 102:2578, 1998.

[137] A. J. Stone. *Chem. Phys. Lett.*, 83:233, 1981.

[138] A. J. Stone and M. Alderton. *Mol. Phys.*, 56:1047, 1985.

[139] A. Hinchliffe. *Molecular Modelling for beginners*. Wiley, University of Manchester, UK, 2 edition, 2008.

[140] B. T. Thole. *Chem. Phys.*, 59:341, 1981.

[141] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, Jr., M. H-Gordon, G. N. I. Clark, M. E. Johnson, and T. H-Gordon. *J. Phys. Chem. B*, 114:2549, 2010.

[142] P. Ren and J. W. Ponder. *J. Comp. Chem.*, 23:1497, 2002.

[143] P. Ren and J. W. Ponder. *J. Phys. Chem. B*, 107:5933, 2003.

[144] J. W. Ponder and D. A. Case. *Adv. Protein Chem.*, 66:27, 2003.

[145] T. R. Walsh. *Mol. Phys.*, 106:1613, 2008.

[146] L. Wesson and D. Eisenberg. *Prot. Sci.*, 1:227, 1992.

[147] A. Hirsch. *Angew. Chem. Int. Ed.*, 41:1853, 2002.

[148] Hersam M. C. *Nat. Nanotechnol.*, 3:387, 2008.

[149] N. R. Tummala and A. Striolo. *ACS Nano*, 3:595, 2009.

[150] A. B. Dalton, W. J. Blau, G. Chambers, J. N. Coleman, K. Henderson, S. Lefrant, B. McCarthy, C. Stephan, and H. J. Byrne. *Synth. Met.*, 121:1217, 2001.

[151] M. Zheng, A. Jagota, Semke E. D, B. A. Diner, R. S. Mclean, S. R. Lustig, R. E. Richardson, and N. G. Tassi. *Nat. Mater.*, 2:338, 2003.

[152] M. Zheng, A. Jagota, M. S. Strano, A. P. Santos, P. Barone, S. G. Chou, B. A. Diner, M. S. Dresselhaus, R. S. Mclean, G. B. Onoa, G. G. Samsonidze, E. D. Semke, M. Usrey, and D. J. Walls. *Science*, 302:1545, 2003.

[153] M. Numata, M. Asai, K. Kaneko, A.-H. Bae, T. Hasegawa, K. Sakurai, and S. Shinkai. *J. Am. Chem. Soc.* , 127:5875, 2005.

[154] B. G. Cousins, A. K. Das, R. Sharma, Y. Li, J. P. McNamara, I. H. Hillier, I. A. Kinloch, and R. V. Ulijn. *Small*, 5:587, 2009.

[155] Y. Li, B. G. Cousins, R. V. Ulijn, and I. A. Kinloch. *Langmuir*, 25:11760, 2009.

[156] C. M. Arnett, C. P. Marsh, C. R. Welch, M. S. Strano, J.-H. Han, J. H. Gray, and T. A. Carlson. *Langmuir*, 2009.

[157] P. Petrov, F. Stassin, C. Pagnoulle, and R. Jerome. *Chem. Comm.*, 23:2904, 2003.

[158] J. Zhang, J.-K. Lee, Y. Wu, and Murray R. W. *Nano Lett.*, 3:403, 2003.

[159] A. Kyani and B. Goliaei. *J. Mol. Struc.-Theochem*, 913:63, 2009.

[160] G. Gianese, V. Rosato, F. Cleri, M. Celino, and P. Morales. *J. Phys. Chem. B.*, 113:12105, 2009.

[161] C.-C Chiu, G. R. Dieckmann, and S. O. Nielsen. *Biopolymers*, 92:156, 2009.

[162] C. Rajesh, C. Majumder, H. Mizuseki, and Y. Kawazoe. *J. Chem. Phys.*, 130:124911, 2009.

[163] W. Fan, J. Zeng, and R. Zhang. *J. Chem. Theory Comput.*, 5:2879, 2009.

[164] Y.. Wand and H. Ai. *J. Phys. Chem. B*, 113:9620, 2009.

[165] T. Roman, W. A. Dino, H. Nakanishi, and H. Kasai. *Eur. Phys. J. D*, 38:117, 2006.

[166] C.-C. Chiu, G. R. Dieckmann, and S. O. Nielsen. *J. Phys. Chem. B.*, 112:16326, 2008.

[167] S. R. Friling, R. Notman, and T. R. Walsh. *Nanoscale*, 2:98, 2010.

[168] C. C. Chiu, M. C. Maher, G. R. Dieckmann, and S. O. Nielsen. *ACS Nano*, 2010.

[169] A. Ortiz-Acevedo, H. Xie, V. Zorbas, W. M. Sampson, A. B. Dalton, R. H. Baughman, R. K. Draper, I. H. Musselman, and G. R. Dieckmann. *J. Am. Chem. Soc.* , 127:9512, 2005.

[170] P. Ren and J. W. Ponder. *J. Phys. Chem. B.*, 107:5933, 2003.

[171] J. W. Ponder, P. Ren, R. V. Pappu, R. K. Hart, M. E. Hodgson, D. P. Cistola, C. E. Kundrot, and F. M. Richards. *Software tools for molecular design*. Washington University School of Medicine, version 4.2 edn. edition, 2004.

[172] D. Eisenberg and A. D. McLachlan. *Nature*, 319:199, 1986.

[173] J. H. Walther, R. Jaffe, T. Halicioglu, and P. Koumoutsakos. *J. Phys. Chem. B*, 105:9980, 2001.

[174] A. A. Skelton and T. R. Walsh. *Mol. Simul.*, 33:379, 2007.

[175] J. Marti, G. Nagy, M. C. Gordillo, and E. Guardia. *J. Chem. Phys.*, 124:094703, 2006.

[176] T. R. Walsh. *Phys. Chem. Chem. Phys.*, 7:443, 2005.

[177] S. Mann. *Nature*, 332:119, 1988.

[178] K. I. Sano, H. Sasaki, and K. Shiba. *Langmuir*, 21:3090, 2005.

[179] M. Umetsu, M. Mizuta, K. Tsumoto, S. Ohara, S. Takami, H. Watanabe, I. Kumagai, and T. Adschiri. *Adv. Mater.*, 17:2571, 2005.

[180] K. Nakanishi, T. Sakiyama, and K. Imamura. *J. Bioscci. and Bioeng.*, 91:233, 2001.

[181] A. B. Sanghvi, K. P. H. Miller, A. M. Belcher, and C. E. Schmidt. *Nat. Mater.*, 4:496, 2005.

[182] E. O. Oren, R. Notman, I. W. Kim, J. S. Evans, T. R. Walsh, R. Samudrala, C. Tamerler, and M. Sarikaya. *Langmuir*, 26:11003, 2010.

[183] F. Baneyx and D. T. Schwartz. *Current Opinion in Biotechonology*, 18:312, 2007.

[184] B. R. Peelle, E. M. Krauland, K. Dane Wittrup, and A. M. Belcher. *Acta Biomaterialia*, 1:145, 2005.

[185] M. A. Henderson. *Surf. Sci. Rep.*, 46, 2002.

[186] Z. Li and T. Lazaridis. *Phys. Chem. Chem. Phys.*, 9:573, 2007.

[187] E. Osawa. *Perspectives of Fullerene Nanotechnology*. Kluwer, Dordrecht, 2001.

[188] M. S. Dresselhaus, G. Dresselhaus, and P. Avouris. *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*. Springer, New York, 2001.

[189] P. M. Ajayan and T. W. Ebbesen. *Rep. Prog.*, 60:1025, 1997.

[190] R. Taylor. *The Chemistry of the Fullerenes*. World Scientific, Singapore, 1995.

[191] H. Somada, K. Hirahara, S. Akita, and Y. Nakayama. *Nano. Lett.*, 9:62, 2009.

[192] M. C. Gordillo and J. Marti. *J. Chem. Phys.*, 117:3425, 2002.

[193] M. C. Gordillo, G. Nagy, and J. Marti. *J. Chem. Phys.*, 123:054707, 2005.

[194] T. Werder, J. H. Walther, R. L. Jaffe, T. Halicioglu, and P. Koumoutsakos. *J. Phys. Chem. B*, 107:1345, 2003.

[195] G. R. Jenness and K. D. Jordan. *J. Phys. Chem. C*, 113:10242, 2009.

[196] M.C Gordillo. *Phys. Rev. B*, 78:075432, 2008.

[197] M. Rubes, P. Nachtigall, J. Vondrasek, and O. Bludsky. *J. Phys. Chem. C*, 113:8412, 2009.

[198] X. Zhao and J. K. Johnson. *Molec. Simulat.*, 31:1, 2005.

[199] A. Pertsin and M. Grunze. *J. Phys. Chem. B*, 108:1357, 2003.

[200] A. Pertsin and M. Grunze. *J. Chem. Phys.*, 125:114707, 2006.

[201] E. M. Huff and P. Pulay. *Mol. Phys.* , 107:1197, 2009.

[202] M.C. Gordillo and J. Marti. *Phys. Rev. B*, 67:205425, 2003.

[203] F. Moulin, M. Devel, and S. Picaud. *Phys. Rev. B*, 71:165401, 2005.

[204] R. L. Jaffe, P. Gonnet, T. Werder, J. H. Walther, and P. Koumoutsakos. *Molec. Simulat.*, 30:205, 2004.

[205] L. Li, D. Bedrov, and G. D. Smith. *J. Chem. Phys.*, 123:204504, 2005.

[206] L. Li, D. Bedrov, and G. D. Smith. *J. Phys. Chem. B*, 110:10509, 2006.

[207] J. H. Walther, R. L. Jaffe, E. M. Kotsalis, T. Werder, T. Halicioglu, and P. Koumoutsakos. *Carbon*, 42:1185, 2004.

[208] N. Choudhury. *J. Chem. Phys.*, 125:034502, 2006.

[209] D. Ulberg and K. E. Gubbins. *Mol. Phys.*, 84:1139, 1995.

[210] E. A. Muller, L. R. Rull, L. F. Vega, and K. E. Gubbins. *J. Phys. Chem.*, 100:1189, 1996.

[211] E. A. Muller and K. E. Gubbins. *Carbon*, 36:1433, 1998.

[212] C. L. McCallum, T. J. Bandosz, S. C. McGrother, E. A. Muller, and K. E. Gubbins. *Langmuir*, 15:533, 1999.

[213] M. Jorge, C. Schumacher, and N. A. Seaton. *Langmuir*, 18:9296, 2002.

[214] S. Hamad, J. A. Mejias, and S. Lago. *J. Phys. Chem. B*, 108:5405, 2004.

[215] S. Picaud, B. Collignon, P. N. M. Hoang, and J. C. Rayez. *Phys. Chem. Chem. Phys.*, 10:6998, 2008.

[216] G. R. Birkett and D. D. Do. *J. Phys. Chem. C*, 111:5735, 2007.

[217] J. Marti and M.C. Gordillo. *J. Chem. Phys.*, 119:12540, 2003.

[218] T. Werder, J. H. Walther, R. Jaffe, T. Halicioglu, F. Noca, and P. Koumoutsakos. *Nano Lett.*, 1:697, 2001.

[219] N. Markovic, P. U. Andersson, M. B. Nagard, and J. B. C. Petterson. *Phys. Chem. Chem. Phys.*, 3:3667, 2001.

[220] M. J. Bojan and W. A. Steele. *Langmuir*, 3:3667, 1987.

[221] M. A. Hamon, J. Chen, H. Hu, Y. Chen, M. E. Itkis, A. M. Rao, P. C. Eklund, and R. C. Haddon. *Adv. Mater.*, 11:834, 1999.

[222] J. L. Bahr, E. T. Mickelson, M. J. Bronikowski, R. E. Smalley, and J. M. Tour. *Chem. Commun*, 2:193, 2001.

[223] K. D. Ausman, R. Piner, O. Lourie, R. Ruoff, and M. Korobov. *J. Phys. Chem. B*, 104:8911, 2000.

[224] R. S. Ruoff, D. S. Tse, R. Malhotra, and D. C. Lorents. *J. Phys. Chem.* , 97:3379, 1993.

[225] J. Chen, M. A. Hamon, H. Hu, Y. Chen, A. M. Rao, P. C. Eklund, and R. C. Haddon. *Science*, 282:95, 1998.

[226] A. Pertsin and M. Grunze. *J. Phys. Chem. B.*, 108:16533, 2004.

[227] D. Feller and K. D. Jordan. *J. Phys. Chem A*, 104:9971, 2000.

[228] K. Karapetian and K. D. Jordan. *Water in Confining Geometries*. Springer, Berlin, 2003.

[229] S. F. Boys and F. Bernardi. *Mol. Phys.*, 19:553, 1970.

[230] W. L. Jorgensen, J. Chandrasekhar, J. F. Madura, R. W. Impey, and M. L. Klein. *J. Chem. Phys.*, 79:926, 1983.

[231] O. Teleman, B. Jonsson, and S. Engstrom. *Mol. Phys.*, 60:193, 1987.

[232] A. J. Stone. *J. Chem. Theory Comput.*, 1:1128, 2005.

[233] M. J. Frish, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr.and T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota,

R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. *Gaussian 03 - Revision B.05*. Wallingford, 2004.

[234] N. Choudhury and B. M. Pettitt. *J. Am. Chem. Soc.* , 127:3556, 2005.

[235] S. H. Lee and P. J. Rossky. *J. Chem. Phys.* , 100:3334, 1994.

[236] D. Xenides, B. R. Randolf, and B. M. Rode. *J. Chem. Phys.* , 122:174506, 2005.

[237] Y. Fan, B. R. Goldsmith, and P. G. Collins. *Nature*, 4:906, 2005.

[238] M. Bockrath, W. Liang, D. Bozovic, J. H. Hafner, C. M. Lieber, M. Tinkham, and H. Park. *Science*, 291:283, 2001.

[239] M. Ishigami, H. J. Choi, S. Aloni, S. G. Louie, M. L. Cohen, and A. Zettl. *Phys. Rev. Lett.*, 93:196803, 2004.

[240] K. Tunvir, A. Kim, and S. H. Nahm. *Nanotechnology*, 19:65703, 2008.

[241] L. Y. Pin, T. L. Gan, L. M. Hsien, and L. F. Yin. *Chin. Phys. B*, 19:27102, 2010.

[242] Q. Lu and B. Bhattacharya. *Nanotechnology*, 16:555, 2005.

[243] Y. S.-Jun, K. Yong, and L. F.-Shen. *Chin. Phys. Lett.*, 24:2036, 2007.

[244] M. Yang, V. Koutsos, and M. Zaiser. *Nanotechnology*, 18:155708, 2007.

[245] L. G. Tien, C. H. Tsai, F. Y. Li, and M. H. Lee. *Diam. Relat. Mater.*, 17:563, 2008.

[246] C. G. Navarro, P. J. D. Pablo, J. G. Herrero, B. Biel, F. J. G. Vidal, A. Rubio, and F. Flores. *Nat. Matter.*, 4:534, 2005.

[247] J. Yuan and K. M. Liew. *Carbon*, 47:1526, 2009.

[248] E. Faizabadi and A. Bagheri. *Physica E*, 41:1828, 2009.

[249] D. E. Hill, Y. Lin, A. M. Rao, L. F. Allard, and Y.-Ping Sun. *Macromolecules*, 35:9466, 2001.

[250] G. Ke, W. Guan, C. Tang, W. Guan, D. Zeng, and F. Deng. *Biomacromolecules*, 8:322, 2007.

[251] P. Singh, S. Campidelli, S. Giordani, D. Bonifazi, A. Bianco, and M. Prato. *Chem. Soc. Rev.*, 38:2214, 2009.

[252] K. I. Tserpes and P. Papanikos. *Comp. Struct.*, 79:581, 2007.

[253] I. I. Salame and T. J. Bandosz. *J. Colloid Interf. Sci.*, 210:367, 1999.

[254] I. I. Salame and T. J. Bandosz. *Langmuir*, 15:587, 1999.

[255] J. Martu and M.C Gordillo. *J. Chem. Phys.*, 119:12540, 2003.

[256] M. Maddox, D. Ulberg, and K.E. Gubbins. *Fluid Phase Equilib.*, 104:145, 1995.

[257] M. M. Dubinin. *Carbon*, 18:355, 1980.

[258] D. D. Do, S. Junpirom, and H. D. Do. *Carbon*, 47:1466, 2009.

[259] G. R. Birkett and D. D. Do. *Mol. Simul.*, 32:887, 2006.

[260] S. Picaud, P. N. M. Hoang, S. Hamad, J. A. Mejias, and S. Lago. *J. Phys. Chem. B.*, 108:5410, 2004.

[261] F. Moulin, S. Picaud, P. N. M. Hoang, L. Partay, and P. Jedlovszky. *Mol. Simul.*, 32:487, 2006.

[262] F. Moulin, S. Picaud, and P. N. M. Hoang. *J. Chem. Phys.*, 127:164719, 2007.

[263] F. Moulin, M. Devel, and S. Picaud. *J. Quant. Spectrosc. Ra.*, 109:1791, 2008.

[264] B. Collignon, P. N. M. Hoang, s. Picaud, and J. C. Rayez. *Chem. Phys. Lett.*, 406:430, 2005.

[265] S. Picaud, B. Collignon, P. N. M. Hoang, and J. C. Rayez. *J. Phys. Chem. B*, 110:8398, 2006.

[266] B. Gorbunov, A. Baklanov, N. Kakutkina, H. L. Windsor, and R. Toumi. *J. Aerosol Sci.*, 32:199, 2001.

[267] H. P. Boehm. *Carbon*, 32:759, 1994.

[268] N. W. S. Kam, M. O'Connell, J. A. Wisdom, and H. Dai. *Proc. Natl. Acad. Sci. U.S.A.*, 102:11600, 2005.

[269] G. A. Rivas, M. D. Rubianes, M. C. Rodrígues, N. F. Ferreyra, G. L. Luque, M. L. Pedano, S. A. Miscoria, and C. Parrado. *Talanta*, 74:291, 2007.

[270] T. Mirfakhrai, J. D. W. Madden, and R. H. Baughman. *Mater. Today*, 10:30, 2007.

[271] Y. Xiao, X. Gao, O. Taratula, S. Treado, A. Urbas, R. D. Holbrook, R. E. Cavicchi, C. Thomas Avedisian, S. Mitra, R. Savla, P. D. Wagner, S. Srivastava, and H. He. *BMC Cancer*, 9:351, 2009.

[272] T. R. Walsh and S. M. Tomásio. *Mol. Biosyst.*, 6:1707, 2010.