Study of the Desmoplakin Protein using Nuclear Magnetic Resonance Spectroscopy and its Interaction with Gold Nanoclusters using Atomic Force Microscopy

Penélope Rodríguez Zamora

A thesis submitted for the degree of
Doctor of Philosophy

Nanoscale Physics Research Laboratory
School of Physics and Astronomy

THE UNIVERSITY OF BIRMINGHAM

October 2013

Study of the Desmoplakin Protein using Nuclear Magnetic Resonance Spectroscopy and its Interaction with Gold Nanoclusters using Atomic Force Micrsocopy

by Penélope Rodríguez Zamora

Nanoscale Physics Research Laboratory
School of Physics and Astronomy
THE UNIVERSITY OF BIRMINGHAM

Abstract

Desmoplakin is a cytolinker protein that establishes connections with the cell cytoskeleton and to anchoring junctions at the cell membrane, providing the adaptable structural rigidity within the skin and heart cells required to accommodate shear forces and maintain tissue integrity. While the C terminal of desmoplakin, composed of three plakin repeat domains and a linker domain, is in charge of intermediate filament binding, the N terminal end, consisting of a plakin domain (PD), is responsible for plaque attachment and interaction with other desmosomal components. Such critical interactions have been studied but important structural and dynamical aspects of desmoplakin have remained unknown.

In this thesis a research study is presented that combines three physical techniques for the study of desmoplakin, namely Nuclear Magnetic Resonance Spectroscopy and Small Angle X-Ray Scattering (for investigating the desmoplakin C terminal end) and Atomic Force Microscopy (for studying the interaction of the desmoplakin N terminal end with size-selected clusters on surfaces).

Regarding the desmoplakin C terminal end, the structure of the linker domain - that is conserved in six plakin family members - has been resolved by nuclear magnetic resonance spectroscopy, exhibiting an unprecedented fold that contains a pair of regular and irregular subdomains, with the latter offering a pair of basic residues that recognise acidic residues on helical intermediate filament proteins such as desmin and vimentin. Its monomeric state is evident by small angle X-ray scattering, and, in combination with desmoplakin plakin repeat domains, it enhances the desmoplakin interaction with the

intermediate filaments. This ensures the adaptable meshwork of contacts that helps to prevent irreversible damage to desmosomes and the cell cytoskeleton upon exposure to mechanical stress.

Regarding the desmoplakin N terminal, the plakin domain immobilisation on a surface of graphite decorated with size-selected gold clusters was performed to address the interaction of this protein with cluster-functionalised surfaces towards the study of its mechanical and other physical properties. The gold clusters were selected to contain 55 or 147 atoms so that their dimensions were smaller than those of single plakin domain molecules, and also in order to achieve maximum stability since these sizes represent magic numbers for gold clusters. Images from atomic force microscopy operating in contact mode have revealed that, despite having been tagged with two cysteine residues, the plakin domain does not establish a covalent bond with the gold nanoclusters. On the other hand, tapping mode atomic force microscopy provides evidence of enhanced weak adsorption of desmoplakin plakin domain to the clusters, increasing the interaction between the protein and the graphite surface. This mode allowed the characterization of desmoplakin plakin domain protein molecules in the liquid phase. The protein dimensions are compared with predicted values based on the small angle X-ray scattering and crystallographic studies, identifying the height of the protein features with the lowest dimension of the crystallised segment of desmoplakin plakin domain that includes SR3-SR6. Images of the protein are obtained with sub-nanometer vertical resolution demonstrating the use of the AFM as a tool for protein imaging and measurement of single isolated protein molecules, as well as protein complex formation, in this case the growth of quasi-2D protein islands.

With the aim of improving the technique of immobilisation of single biological molecules with metal nanoparticles, a new technique of cluster immobilisation has been developed using small metal clusters (Au_{20}) to create channels in a graphite substrate. These channels function as defects on the surface of the graphite to anchor soft-landed clusters with the potential to bind the biomolecules. The method is tested with atomic force microscopy and compared with previous clusters immobilisation techniques using argon ion defects. A full characterisation of soft-landed gold clusters with 923, 561, 309 and 147 atoms has been completed with high-resolution, non-contact atomic force microscopy. Contact mode atomic force microscopy experiments have demonstrated stronger immobilisation of the gold clusters on the support by using this new technique than by the use of its predecessors.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Richard E. Palmer for giving me the opportunity to undertake a PhD at the University of Birmingham as a member of the NPRL group, and for all his ideas, guidance and support. I would also like to thank my co-supervisors Dr. Martyn Chidgey and Prof. Michael Overduin for opening the doors to the Laboratory of Structural Biology and NMR at the School of Cancer Sciences, and for their teaching and patience. I am also grateful to CONACyT for funding my PhD.

I would like to thank everyone at the Nanoscale Physics Research Laboratory, with special thanks to: Dr. Gueorgui Barreto and Dr. Tian-Luo Pan for their training in SPM techniques; Dr. Feng Yin, Dr. Vahideh Habibpour and Ahmed Abdela for producing the clusters for my experiments. Also, my PhD experience would not have been the same without the presence of my fellow students, past and present *nano-friends* with whom I spent nice times: Mim, Javi and Rich (the bio-team!); Alina, Dongsheng, Martin, Ivy and Max (my generation), Dave, Chris, Thibault, James and Tom (the experienced), Karl, Ruth, Ray, Charlotte, Lu, Wei, Dongxu, Will, Scott, Dogan, Caroline and Nan (the freshers).

My thanks are also extended to everyone at the Laboratory of Structural Biology and NMR, expressly to Dr. Caezar Al-Jassar for initiating the collaboration of this project and for providing most of the structural biology laboratory instruction; Dr. Mark Jeeves for NMR spectra acquisition help and NMR teaching; Dr. Timothy Knowles for NMR data analysis assistance; and Dr. Claudia Fogl for NMR titration experiments assistance and production of protein mutants in collaboration with Arvinder Malli. Everyone in G64 was very welcoming and helpful when I started in Cancer Sciences, so I am also very grateful to my colleagues and friends Piv and Nazia; and Pooja, Jas and Sandya, the best technicians I know. Also thanks to James Bowen from Chemical Engineering for all his AFM advice.

I wish to thank Daniel, for his constant support, for being by my side in the good and bad times and for making the colours of my life more vivid.

My heartfelt thanks to my Dad and all my family, Carlos and specially my beloved siblings Yael, Charlie and Karlita, who, despite the distance, always put a smile on my face.

Finally, this thesis is dedicated to my Mother and my Grandma, to whom I owe everything, and who gave me the love and courage to achieve my goals in life.

Contents

Abbreviations

AFM Atomic Force Microscope

ADR Ambiguous Distance Restraint

ARIA Ambiguous Restraints for Iterative Assignment

ARVC Arrythmogenic Right Ventricular Cardiomyopathy

AUC Analytical Ultracentrifugation

BC Desmoplakin Plakin Repeat Domains B and C including Linker Domain

BMRB Biological Magnetic Data Bank

BPAG1 Bullous Pemphiogid Antigen 1

CD Circular Dichroism

CM-AFM Contact Mode Atomic Force Microscope

CSI Chemical Shift Index

DPlink Desmoplakin Linker Domain

EM Electron Microscopy

ECM Extracellular Matrix

FID Free Induction Decay

GFP Green Fluorescent Protein

GST Glutathione S-transferase

HOPG Highly Ordered Pyrolytic Graphite

HRP Horseradish Peroxidase

HSQC Heteronuclear Single Quantum Coherence

IDP Inner Dense Plaque

IF Intermediate Filament

IPTG Isopropyl β-D-1-thiogalactopyranoside

I-TASSER Iterative Threading Assembling Refinement

LB Luria Broth

MACF1 Microtubule-Actin Cross-linking Factor

MD Molecular Dynamics

MSA Molecular Surface Area

NC-AFM Non-contact Atomic Force Microscope

NOE Nuclear Overhauser Effect

NOESY Nuclear Overhauser Effect Spectroscopy

NMR Nuclear Magnetic Resonance

NPRL Nanoscale Physics Research Laboratory

NSOM Near-field Scanning Optical Microscopy

NTA Nitrilotriacetic Acid

ODP Outer Dense Plaque

OSM Human Oncostatin M

PD (Desmoplakin) Plakin Domain

PDB Protein Data Bank

PPlink Periplakin Linker Domain

PR Plakin Repeat

PRDs Plakin Repeat Domains

PSPD Position-Sensitive Photodetector

RF Radio Frequency

RMS Root Mean Square

SAMs Self-Assembled Monolayers

SAXS Small Angle X-Ray Scattering

SDS-PAGE Sodium Dodecyl Sulfate Polycrylamide Gel Electrophoresis

SOC Super Optimal Broth

SPM Scanning Probe Microscopy

SR Spectrin Repeats

STM Scanning Tunnelling Microscopy

TEM Transmission Electron Microscopy

TM-AFM Tapping Mode Atomic Force Microscopy

TOCSY Total Correlation Spectroscopy

ToF Time of Flight

Chapter 1

Introduction

1.1 Physics Meets Biology in Understanding Protein Structure and Function

Every day there is a new topic in which physics meets biology; the two disciplines overlap to improve the knowledge of the underlying physical mechanisms and structures of biological systems. Physicists working in biology look to explore the most basic elements of biological systems and the most complex processes of life. Only by reaching this level of understanding, will our endeavours be able to decipher the fundamental mechanisms of living organisms.

Of great interest in this field, due to their crucial role within most physiological activities, is the study of proteins. It is not necessary to point out the importance of proteins, since it is well known that they are vital constituents of cells, function as enzymes and hormones, and their malfunction has been proposed to be the responsible of more than 98% of human diseases [1]. For example, mutations in one protein, p53, result in its inactivation, characteristic of more than half of all cancer cases [2]; abnormal protein aggregation is behind many neurodegenerative diseases, such as Alzheimer disease and Creutzfeld-Jakob disorder [3]; tissue related pathologies and cardiac malfunctions are often also a result of mutated proteins [4]. This list of relevant phenomena associated with proteins could extend for several more pages, denoting the undeniable importance of proteins in disease processes. However, complete understanding of protein structure and function still represents a challenge for present day scientific techniques.

The conformational behaviour of proteins represents a complex physical problem given the many degrees of freedom of the constituent atoms. Atoms interact with surrounding water molecules, with other proteins and with other biological macromolecules. In this picture we can see that every particular detail dictates the difference between proteins resulting in a non-universal behaviour. However, proteins also share certain basic characteristics: a core with mainly hydrophobic amino acids; alpha helices and beta sheets as secondary structures; a certain degree of flexibility but at the same time the ability to establish robust interactions [5].

By putting all this together a significant inference arises, that protein structure determines protein function. From this, the importance of investigating structure and interactions in a simultaneous manner to elucidate protein function can be concluded. Both aspects lead to the understanding of biological function and as direct consequence also to the comprehension of non-functionality within disease.

Several methods for protein structure determination have been developed over the last few decades, gaining insight in protein function at a molecular level. The method that has contributed the most to elucidate the protein spatial conformation is X-ray crystallography, providing protein maps with the highest resolution [6]. Thousands of protein structures have been resolved with crystallographic data using X-ray diffraction, but many proteins have not yet been crystallised due to diverse obstacles, of which the main one is probably the high degree of flexibility of some proteins. Moreover, crystallographic techniques may not reveal some protein-protein interfaces, leaving the door open for other techniques to fill the gap.

Like X-ray crystallography, all the other techniques that have been shedding light on protein structure during the last decades (such as nuclear magnetic resonance spectroscopy, small-angle X-ray scattering, atomic force microscopy and electron microscopy among others) are based on physical principles. The reason for this lies in the fundamentals of matter and energy, which are the same for everything in the universe, including living organisms. In this way, an understanding of proteins as matter/energy elements leads to the development or adaptation of powerful computational and mathematical tools that enhance research progress. Furthermore, physics is also able to attack the problem from a distinct point of view, providing a unique perspective [7].

Complex biological systems, such as those involving proteins, need the conjunction of many scientific disciplines. The complementary expertise of disciplines including biology, physics, chemistry and engineering is necessary to approach proteins as an area of study, and to gain insights into protein behaviour that could lead to the development of more applications in biotechnology and medicine.

1.2 About this Work

The work presented in this thesis incorporates the research carried out on one protein using different physical techniques to explore three different aspects of its structure and function. The subject of this study is a human desmosomal protein called desmoplakin. The challenges that this protein exhibits are of diverse nature, from resolving the structures of some of its domains, which remain unresolved, to understanding the specific interactions in which it is involved in essential physiological processes including cell adhesion. Only by the integration of contrasting techniques that contribute to unveil different aspects of it, it will be possible to conceive the whole picture from which desmoplakin forms part of the cell junctions within the tissues of human body.

The techniques used to investigate desmoplakin structure/function were Nuclear Magnetic Resonance (NMR) spectroscopy, complemented with Small Angle X-ray Scattering (SAXS), and Atomic Force Microscopy (AFM). Each one of these techniques has enormous advantages in protein research, which will be briefly described in the following sections of this chapter, followed by an explanation of the structural and functional characteristics of desmoplakin, as well as its importance. Finally, at the end of this chapter a section will describe the specific and global implications of this research work.

Chapter 2 presents the materials and methods used for obtaining the results included in this research work.

Chapter 3 details the use of NMR and SAXS techniques to determine the structure of the desmoplakin linker domain (DPlink) and to generate an envelope for the desmoplakin plakin repeat domains B and C including the linker domain (BC construct).

Interactions of the desmoplakin linker domain, its plakin repeat domain C and the periplakin linker domain (PPlink) with intermediate filaments vimentin and desmin are explored by NMR and presented in a comprehensive review in *Chapter 4*.

Immobilisation experiments of the desmoplakin plakin domain (PD) with metal size-selected nanoclusters on graphite are depicted in *Chapter 5*.

Finally, and as direct consequence of the experience acquired during *Chapter 5* work, *Chapter 6* presents the development of a novel technique of metal clusters immobilisation towards the improvement of protein immobilisation using this approach.

1.3 NMR Spectroscopy and SAXS for Protein Structure Determination and Protein-Protein Interactions Studies

Unlike crystallography, that could be considered a static technique, NMR has access to structure and at the same time dynamics of proteins in solution, being able to explore flexible domains or possible multiple configurations in a native environment. However, unique determination of the global shape and orientation remains as a challenging task for NMR alone, especially for multi-domain proteins or relatively large molecular assemblies. SAXS, on the other hand, specialises in overall size and shape constraints, demonstrating a clear complementarity with NMR spectroscopy. Validation of NMR structures has already been performed by comparison between them and the correspondent envelope generated from SAXS. By resolving the ambiguities of the orientations of the individual domains it is possible to discriminate between similar structural conformations [8]. While NMR spectroscopy provides the atomic detail, SAXS is not restricted by flexibility or molecular weight, being able to analyse large proteins whose smaller domains can be studied by NMR.

1.3.1 Atomic-level Protein Structure and Interactions by NMR

Along with X-ray crystallography, NMR spectroscopy is a technique for determination of protein structure and molecular kinetics. With NMR it is possible to obtain 3D images of the biological structures in solution with atomic resolution, which has the advantage

that the protein is in a native-like environment. In addition, this technique is capable of giving information regarding the dynamics of the system, like reaction mechanisms, molecular motion and protein-protein interactions [9]. NMR spectroscopy does not damage protein samples, allowing the signals to be detected for long periods of time if necessary. The variety of NMR applications in the field of protein structural biology is wide, including the determination of biochemical properties and functional groups [10], the determination of dissociation constants between interacting biomolecules [11] and the detection of intermolecular contacts and hydrogen bonds [12], among many others.

An important limitation of NMR spectroscopy is the size of the protein being studied. The method is limited to proteins with molecular weights less than 40-60 kDa [13], the main reason being the increase of linewidths proportional to the molecular weight, since larger bodies have a slower tumbling in solution, causing a reduction in the efficiency of magnetisation transfer through bonds [14]. The other reason is the high complexity of the acquired data that makes it a time-consuming and challenging technique.

1.3.1.1 Protein structure determination

There are several widely used methods that provide information about protein structure. Spectroscopic methods reveal the atomic nuclei and bond information using specific resonance frequencies and transition energies found by promoting the interaction between the molecule and radiative energy of certain type. In the case of NMR spectroscopy different data sets resolve the uncertainties of distances and angles needed to reconstruct the 3D structure of a protein [15].

A degree of probability is inherent in the structure calculations. Hence it is necessary to consider two or more possible conformations. However, NMR spectroscopy reaches such a high resolution level that constitutes an independent check tool of high accuracy for X-ray crystallography, being able to determine intrinsically unstructured regions, partial protein folding and in some occasions being able to detect errors in structures produced by the crystallography method [16]. Another of the advantages of NMR over X-ray crystallography is that the former is free of artifacts that may occur in crystallisation as a result of intermolecular contacts that may influence the local conformation of a protein [17].

All of the above, along with its ability to estimate time scales in intramolecular motions and to characterise conformations of bound ligands [16], are advantages for the use of NMR rather than X-ray crystallography for protein structure determination under certain circumstances such as structure and function elucidation of novel domains.

1.3.1.2 Protein-protein interactions

NMR spectroscopy is unrivalled in determining protein-protein interactions and intermolecular dynamics for small (<20 kDa) proteins. Specific interactions between proteins are a fundamental subject of study that provides indispensable information about vital processes such as signal transduction, gene expression and cellcell interactions [18, 19].

The resolving power of NMR spectroscopy makes possible to observe chemical kinetics in interactions processes, since it detects changes in chemical environments that exist as consequence of exchange mechanisms [15]. Because of its ability to resolve multiple states [19], NMR spectroscopy is undisputedly a method that gives high accuracy results in ligand binding and macromolecular complex assays. Nonetheless it should be complemented with the additional informational about the protein structures and interaction, to provide a very complete and validated understanding of proteins.

1.3.2 Scope of SAXS in Ascertaining Protein Morphology

A variant of X-ray scattering is small-angle X-ray scattering (SAXS). At small angles (normally between 0.1 and 5 degrees), the scattering is mainly elastic scattering and yields structural information on the sample based on the diffraction effects produced by the interference of coherent scattering [20]. Crystallisation is not necessary in SAXS studies, an advantage that makes it more versatile regarding sample preparation. On the other hand, due to the random orientation of the proteins in solution the information provided by SAXS studies is not as detailed as that provided by X-ray crystallography.

Protein structure determination techniques such as NMR and X-ray crystallography provide high-resolution solutions but require well-structured and rather rigid systems. Proteins with a high level of flexibility or with disordered loops and domains represent challenging subjects for such techniques. Naturally unfolded proteins can function as regulators in important processes such as gene expression and the cell cycle, thus there

is a need of an alternative technique able to handle proteins in this classification and for proteins without a fixed tertiary structure in general [21].

SAXS specialises in macromolecules or complexes in solution and is able to cope with disordered systems. It is a technique capable of reconstructing 3D models, with less resolution than NMR but covering more cases including large systems. Even if a detailed solution for a protein structure is not achieved, a low-resolution model for the tertiary structure can reveal key functions such as binding domains and hinges. Thus SAXS is a powerful method for the analysis of non-crystalline states of bio-macromolecules. But this does not mean SAXS should be classified simply as a low-resolution technique, on the contrary, for large-scale structures, SAXS is a high precision method, and large distances often act as very tight constraints on protein structure refinement [22]. A major advantage for SAXS is the speed of data collection and sample characterisation [23]. While NMR data collection takes several weeks and full assignment plus structural calculations can take several months, SAXS takes a few min of data collection and around a week for resolving an acceptable envelope. Given this convenience, SAXS is commonly also used for rapid screening of samples under varying conditions [24].

1.3.2.1 X-ray scattering at small angles in molecules

Briefly, SAXS is a technique based on the interaction between X-ray radiation and the atoms of the sample of analysis. From the particle point of view, a number of the photons of the X-ray beam directed at the sample are deviated from their original trajectory by the electrons present in the solution, in both solute and solvent. The difference in electron densities between these two components of the sample is often called excess scattering length density [23, 25]. In this way, SAXS data is based on a slight contrast that can be identified by the subtraction of the scattering yield by a blank sample from the protein sample. The observable signal is the scattering intensity I, and is recorded as a function of the scattering angle. But the relation normally studied is between I and the measure of the directional momentum change that the photons undergo, or momentum transfer q, rather than simply the scattering angle [25], therefore the function $I(q)$. (More detail in this regard is presented in Chapter 3, Section 3.3.1).

In a monodisperse protein sample all proteins have exactly the same structure and have random orientation, therefore the scattering intensity is proportional to the single particle scattering averaged over all orientations, resulting in an isotropic (radially symmetric) distribution [21]. The smaller the angle of scattering, the longer the distance that can be characterised by the data. This is the reason why for SAXS the resolution is classified in terms of the smallest angles that can be measured, opposite to crystallography in which the resolution is determined in terms of the highest angle data measured, which provides the smallest distances that can be resolved [26]. *Ab-initio* shape determination techniques then pass from the uni-dimensional scattering patterns to the three-dimensional protein structures by creating a model that fits the data [27].

1.4 AFM for Studies of Protein Interaction with Surfaces

AFM is a well-established analytical, imaging and micro-manipulation tool [28] and, even though the resolution achieved by AFM is considerably lower than that achievable by other techniques such as X-ray crystallography and NMR spectroscopy, it has the advantage of being able to address and manipulate single proteins in bio-friendly physiological conditions [29]. Outside their native conditions, proteins tend to denature with changes to their conformational and dynamic properties [3]. There is a broad range of biophysical applications of AFM, including imaging techniques [30] and force spectroscopy [31]. The ability to image the topography of a biological molecule at sub-nanometre resolution, while simultaneously revealing intrinsic properties such as flexibility and surface potentials, makes AFM a tool of undeniable value for the study of biological samples [32].

1.4.1 Imaging Proteins with AFM

AFM can be used for both topographic imaging and manipulation of biological structures under physiological conditions. Samples undergoing such studies range from macromolecular assemblies [28] to individual biomolecules.

In the first case, imaging membranes with AFM has proved to be a high quality method that is capable of resolving individual molecules within an assembly in great detail [30]. Characteristics exclusive to the proteins incorporated in molecular layers make them

highly suitable for AFM imaging. Properties such as the orientation of the molecule with respect to the substrate, the stabilisation of lateral movement within the molecules as consequence of the tight packing of the array, facilitates image processing, allowing sub-nanometre resolution to be achieved within AFM imaging [33, 34].

In contrast, in the case of individual isolated proteins with molecular weights of tens to hundreds of kDa imaging at such a high resolution becomes a big challenge [29, 35]. The substructures are difficult to resolve and are usually recognised through the high correlation between molecular weight and molecular volume (correlation coefficient of 0.994) [36]. Isolated proteins have a tendency to present themselves as globular shapes in AFM imaging. This makes different proteins, protein complexes and other biological materials difficult to identify based on shape recognition alone (Fig. 1.1). For proteins in membranes, the typical AFM spatial resolutions reported range from 0.5 nm to 5 nm for lateral measurements, and around 0.1-0.2 nm for vertical measurements in contact mode, [37] while for tapping mode a lateral resolution between 1 and 1.5 nm and similar vertical resolution has been reported [38]. On the other hand, the minimum diameter of contact restricts the lateral resolution for individual molecule measurements to several nanometres [39]. The reason for these high resolution images on membranes despite the tip radii normally used is the existence of a nanometre-sized protrusion at the end of the tip, which enables fine structural details to be imaged [34, 40].

For soft samples in general, factors to consider in image acquisition include topography of the sample, tip geometry, sample viscoelastic properties [32] and the way the molecule is anchored to the support.

For weakly bound proteins, and other very fragile structures, tapping mode AFM is an appropriate method as it minimises the lateral force applied to the sample while scanning [34]. When the biomolecules are strongly attached to the support, contact mode AFM is not only able to image them but can also measure the forces involved in the interaction between the support and the molecule or the forces of the molecule dynamics itself. The principal biophysical applications in which AFM has been proved to be a competent tool, in addition to imaging, include: protein unfolding force [41], molecule stretching [31], dissection of isolated DNA [37], metal-chelating peptide bonds [42] and protein-protein interactions in real time [43].

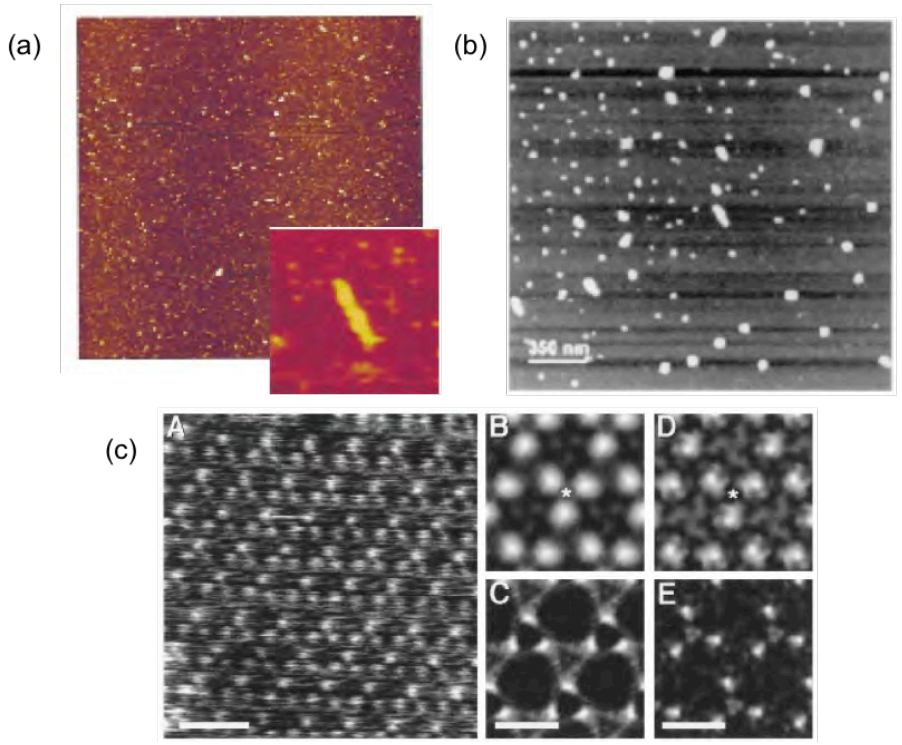


FIGURE 1.1: (a) Tapping mode AFM topographies of beta-amyloid on mica, acquired *in situ*. Larger image corresponds to a scan size of 13.5 μm x 13.5 μm and the smaller image is a digital zoom in of the first image with a size of 1μm x 1μm [29]. (b) AFM topography of Sup35 peptides on mica recorded in air [35]. (c) A, B, C, E are examples of tapping mode AFM topographs of the extracellular purple membrane surface. D was recorded in contact mode at applied force of 100 pN. Scale bars correspond to 10 nm (A) and 5 nm (B-E) [30].

1.4.2 Protein Immobilisation Methods

In order to take full advantage of AFM as an instrument for protein experimental analysis, it is necessary to combine it with an effective surface immobilisation technique that maintains single proteins tethered and stable. In this way, AFM allows us to obtain information from a single molecule. AFM studies of proteins require high stability of the sample, thus proteins in packed layers are normally used [34]. This stretches the limitation of protein research because it overlooks important characteristics at the single molecule level, without intermolecular interaction interference.

The immobilisation of single-molecule proteins has become an important challenge for the field of nanotechnology. This is largely due to the development of biosensors and microarrays, which have gained relevance in the fields of biology and medicine [44]. In recent years, different immobilisation methodologies have been developed in an attempt to provide optimal protein-surface binding conditions with a low unspecific background.

Protein immobilisation has been performed by means of many different techniques and depending on the desired result, some would be more adequate than others. However, in the majority of cases, the aim is to study the properties of the biomolecules and the dynamics of biochemical reactions, therefore keeping the biological capabilities active after the immobilisation is a goal shared by nearly all methodologies. It is also often necessary to prevent a misorientation of the protein, since with a random orientation there is the possibility of blocking the availability of the active binding sites, causing a loss of enzymatic activity. This results in low signal-to-noise ratios in protein assays [45].

In general terms, the majority of immobilisation methods are based on two principles, those that operate by physisorption, and those that operate by chemisorption. The forces involved in physisorption are intermolecular forces that cause a minimal perturbation of the electronic states of the species implicated in the interaction [46]. This is the reason why physisorption is not specific and not very strong. Chemisorption, on the other hand, is characterised by chemical specificity. In this case, the forces involved are valence forces that create ionic and covalent bonds, which are much stronger than the non-specific interactions in physisorption [47].

Some established methods for protein immobilisation based on physisorption and chemisorption are adsorption by hydrophobic and electrostatic interactions [48–50], covalent bonding [51, 52] and via metal ion chelator surfaces [53], among others. Immobilisation methods mediated by each interaction have advantages and disadvantages. The strong attachment and the specification are the main benefits of chemisorption, while physisorption methods may be simpler, less expensive and in some cases retain high catalytic activity [50].

1.4.3 Immobilisation of Proteins with Size-Selected Clusters

The immobilisation of proteins by size-selected atomic clusters is a route towards the study of single protein characteristics and the potential development of biochips and biological microarrays [54]. The pinning technique allows substrate surfaces to be decorated with nanoclusters of a specific size depending on the requirements of the experiment [55]. Though still in its infancy, the method promises several advantages over homogeneous surfaces, such as anchoring, orientation and non-denaturation of single, isolated protein

molecules [56, 57]. In particular, gold clusters provide a platform for the immobilisation of proteins containing available cysteine residues, via the covalent thiol bond existing between the gold and the sulphur atoms in the residue [58]. Not all proteins contain suitable cysteine residues, however, it is now possible, using genetic engineering, to produce proteins that include cysteine residues on their surface.

1.4.3.1 Size-selected metal clusters

On the macroscale, changes in size do not modify the nature of matter. On the other hand, on the true nanoscale (i.e. 1-10 nm), even a minuscule change in size, a few atoms more or less, can make a difference; the properties of matter can vary drastically from one size to another. That is why the ability to select sizes of clusters opens a whole range of possibilities [59].

An atomic cluster is a set of atoms aggregated to form a structure that is between single atoms and solid bulk metals. The characterisation of metallic clusters has been the subject of a tremendous amount of research. From interatomic interactions to the interaction with other clusters and surfaces, the study of metallic clusters has brought insight into many fundamental properties of matter such as electronic structure and interatomic forces [60]. Metallic clusters have also proven to have a extent variety of applications due to the different properties (electrical, optical and magnetic properties among others) that are dependent on their size.

Gold clusters are the most stable metal nanoparticles [61], and have been the target of several investigations regarding their physical and chemical properties, synthesis, assembly and applications [61]. One of the applications of gold nanoclusters is to use them as immobilising agents for proteins containing cysteine amino acids. This approach not only anchors the protein, but also controls its orientation, isolation and minimises the risk of denaturation.

In previous studies a number of proteins containing cysteine residues have been immobilised with gold nanoclusters, e.g. GRoEL, from the chaperonin protein family [62], GFP (green fluorescent protein), and OSM (human oncostatin M) [54]. In all cases, the availability of cysteine residues, as well as the sulphur availability within the cysteines, has demonstrated to be essential to achieve a successful attachment to the gold clusters.

For instance, GFP contains two cysteine residues, but both are embedded within the molecule volume (one completely and the other partially), precluding the bonding between protein and cluster, as it was shown experimentally with Au_{26}, Au_{55} and Au_{70} [63]. Conversely, OSM, having five available cysteines presented strong binding to Au_{40} [58]. HRP molecules (horseradish peroxidase), even though they don't contain free cysteine residues, they have been attached to gold clusters due to the four disulphide bridges that they contain [56]. Although this method is restricted to proteins containing available cysteines it can be further improved by using thiol functional groups as a linking bridge to other specific molecules [64]. Moreover, the immobilisation of proteins using gold nanoclusters presents many advantages over other methods, satisfying four key points of the concept of immobilisation of proteins:

Anchoring: Gold clusters are relatively inert and relatively hydrophilic, thus protein molecules are naturally attracted to their surface [65]; coupled with chemical bonding in the presence of cysteine residues on the proteins, this generally leads to successful immobilisation.

Orientating: Given that the cysteine tags may be located in one specific site on the protein, unique binding sites determine their orientation. Orientation is therefore, in principle, the same in every case, so we can talk of an oriented arrangement of the proteins on the surface.

Isolating: Smaller clusters ensure that only one protein will attach to each cluster and by selecting their density it is possible to avoid lateral interaction between neighbour proteins. Nevertheless, the potential binding of more proteins must be kept under consideration. If agglomeration of proteins is suspected, presenting islands of proteins rather than single molecules, in principle it is possible to remove the weakly bound proteins by washing the sample, thus obtaining single protein elements in each cluster.

Non-denaturing: The probability of the protein becoming denatured is considerably reduced since the cluster size is much smaller than the size of the protein. Also, since experiments can be performed in native conditions, it is possible to keep the integrity of the biological sample.

Other advantages of this method include that no other chemical processes, (which can result in high noise levels), are needed and that it can be managed to high material usage

efficiency [61]. In view of all of this it is straightforward to conclude that size-selected clusters on surfaces are a feasible and interesting alternative approach to immobilising proteins for different purposes.

As stated, different sizes of clusters suit different requirements, so the choice of cluster size for immobilising proteins is a fundamental issue. Cluster sizes must be chemically stable so as not to affect interaction with the sulphur. For gold clusters, the chemical stability can be explained by seeing the clusters as dense packs of atoms; considering those as spheres whose each atom is surrounded by twelve more atoms. By taking one single atom, as a starting point, the first cluster that it is possible to form is one composed of 13 atoms, and the second of 55, since the second layer consists of 42 atoms. This can be calculated following the rule $10n^2 + 2$, where n is the number of layer. Thus, the sizes from the third layer to the eighth are 147, 309, 561, 923, 1415 and 2057 [61]. On this basis, these clusters have been called magic-atom-number metal clusters [66, 67]. This asseveration is mainly based on the fact that, unlike in other cases, they seem to be inert to the presence of atomic oxygen, a behaviour that is ascribed to their closed-shell structure [68, 69].

1.4.4 AFM Characterisation of Size-Selected Metal Clusters

Nanostructured surfaces have demonstrated novel properties, with characteristics which promote a progress in research into the fundamental physics basis of their development [70]. The technological applications of metal nanoparticles supported on different substrates go beyond the biological applications. They include catalytic and electrocatalytic processes (with a varying activity that depends on the geometrical and electronic characteristics of the nanoclusters, which is in turn function of the element and size of the clusters) [71, 72]; plasmonic device development [73]; and single electron transistor prototyping [74].

AFM, as an SPM technique, has been able to perform nanoparticle manipulation under different schemes and in diverse systems. Some examples of its capacity, apart from biomanipulation, include nanotribology studies and nanowire shear stress measurements [75]. In the case of nanoparticles on graphite, AFM has been used to measure the static-kinetic transition of the nanoclusters occurring when the lateral force exerted by the AFM tip is greater than the particle-substrate adhesion force [76].

AFM experiments into protein immobilisation with metal nanoclusters require clusters with two characteristics: that they are smaller than single protein molecules but visible for AFM, and that they are strongly immobilised in a substrate. On previous occasions these two characteristics seemed to be irreconcilable, necessitating the use of STM for the characterisation of the clusters before protein deposition [56, 58]. In the wake of such conditions, demand grew to be able to pre-characterise the clusters and then characterise the cluster-protein system with the same instrument as a priority requirement for the improvement of AFM studies of proteins immobilised by clusters, leading to an independent investigation into how to visualise immobilised metal clusters with less than 1000 atoms with a room temperature non-vacuum AFM.

A combination of deposition techniques for gold size selected clusters with different numbers of atoms and the introduction of a completely new element playing an unprecedented role in the immobilisation of clusters formed the core idea of a novel technique that enhances the characteristics of the graphite-gold cluster system for AFM characterisation and posterior utilisation in protein immobilisation experiments.

1.5 Desmoplakin

Desmoplakin is a plakin protein that acts as a cytoplasmic component of desmosomes. In order to understand the relevance of desmoplakin, it is necessary to first describe both the characteristics that define the plakin family, and desmosome structure and function.

1.5.1 The Plakin Family

Desmoplakin is a member of the plakin family of cytolinkers. The plakins are defined as proteins that associate with junctional complexes, such as desmosomes and hemidesmosomes, and the cytoskeleton [77]. The family includes eleven members: plectin, the bullous pemphiogid antigen I group (BPAG1e, BPAG1a and BPAG1b), desmoplakin (I and II), envoplakin, periplakin, the microtubule-actin crosslinking factor group (MACF1a and MACF1b) and epiplakin (Fig.1.2) [77]. Although each one of the members of the plakin family was discovered under different circumstances there were certain characteristics that identified them as similar proteins, the first one of them being their localisation to intermediate filaments (IFs) [78]. Structural similarities also indicated the existence

of a common factor: all of them contain either a plakin domain or a plakin repeat domain, or both. In plectin, BPAG1, desmoplakin and envoplakin the amino terminal head domain consists of a plakin domain built of an antiparallel bundle of alpha helices, while the carboxyl terminal tail domain consists of a variable number of plakin repeat domains (PRDs), normally designated A, B and C, with a predominant alpha helical secondary structure and a linker domain that tends to be between PRD B and C. The rod domain consists of heptad repeat structural motifs.

All the members of the family, except epiplakin, share the plakin domain, in the N-terminal end. On its part, the linker domain, which structure remained unknown until now, is a region highly conserved among six members of the family: plectin, BPAG1e, desmoplakin I and II, envoplakin and periplakin (Fig. 1.2) [78, 79]. In general, comparison between sequences of the members of the family reveals high homology. For example, identity between desmoplakin and plectin are 32% and 49% for their N and C-terminal head and tail domains respectively [80].

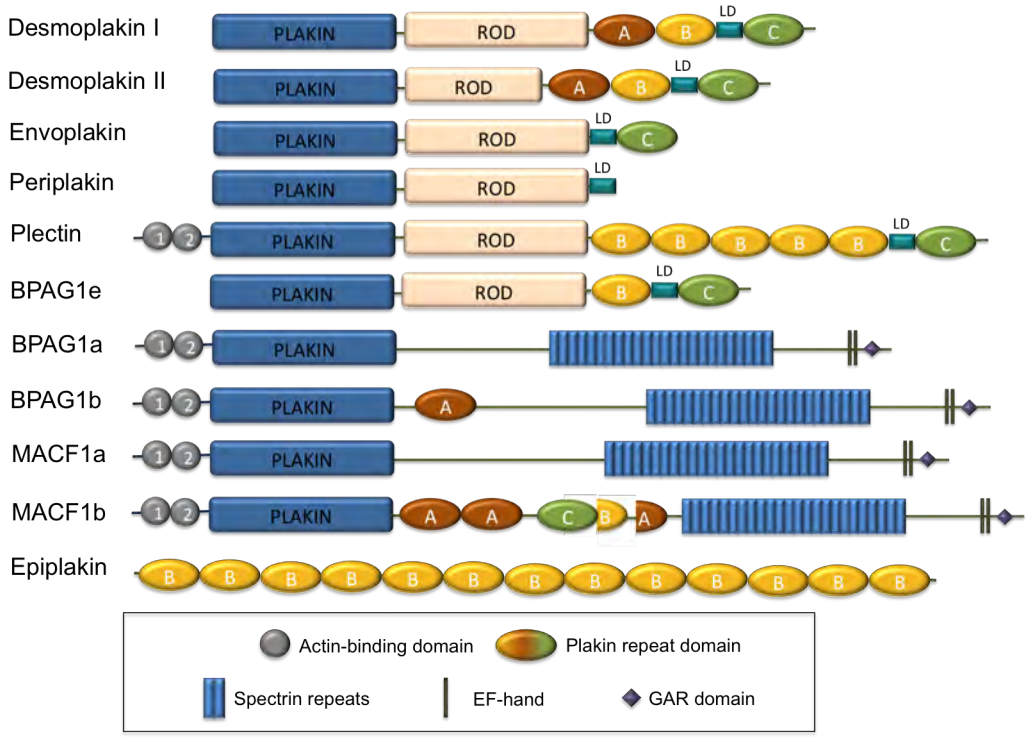


FIGURE 1.2: Schematic of the proteins members of the plakin family. All the members, except for envoplakin, include a plakin domain, reason why this family of proteins is called the plakin family. The linker domain (LD) is a region highly conserved among six members, including desmoplakin, periplakin and plectin. Apart from their structural similarities, the plakins also have in common their localisation to intermediate filaments [78]. Adapted from [81].

Apart from connecting desmosomes and hemidesmosomes, and associating with intermediate filaments, the plakins also bind to microfilaments, microtubules and members of the armadillo family [82], making them important in several biological processes.

1.5.2 The Role of Desmoplakin in Desmosomes

1.5.2.1 Cell junctions

All animals and other multicellular organisms need cell junctions to hold their cells together and to give them structure and strength. It is not difficult to understand the importance of such junctions by comparing the properties of a cell and tissue: cells are deformable and flimsy, while tissues can be strong, firm and rigorously ordered. This radical and essential difference between the unit and the assemblies relies on cell junctions.

There are two types of cell junction, cell-extracellular matrix (ECM) junctions and cell-cell junctions [83]. The extracellular matrix is a complex network that underlies epithelial tissues and acts as a supporting framework. It is formed of water, proteins and polysaccharides with a specific physical, topological and biochemical composition for each tissue [84]. Cell-matrix junctions are multiprotein organisations that provide mechanical attachment between cells and the extracellular matrix whereas cell-cell junctions provide attachment between neighbouring cells. Cell-cell junctions are divided into three types: occluding junctions, anchoring junctions and communicating junctions. Of these, we are especially interested in anchoring junctions, whose function is to link cytoskeletal proteins from two neighbouring cells, providing structural cohesion. There are four different types of anchoring junctions: desmosomes, adherens junctions, hemidesmosomes and focal adhesions (Fig. 1.3) [85].

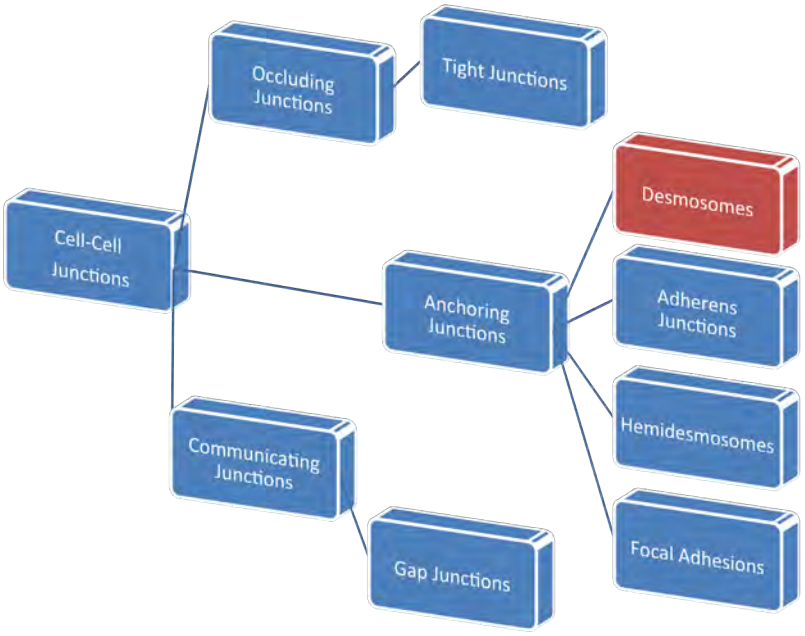


FIGURE 1.3: Classification of cell-cell junctions. There are three types of cell-cell junctions: occluding junctions, anchoring junctions and communicating junctions. The anchoring junctions can be in turn classified in four different kinds: desmosomes, adherens junctions, hemidesmosomes and focal adhesions [85].

1.5.2.2 Desmosomes

Desmosomes (derived from the Greek *desmo* which means bound, and *soma* which means body) are disc shaped structures of up to 1 μm across, consisting of a set of different proteins that interact to build a complex. These complexes are located at the lateral edges of adjacent cells and act as junctions, maintaining mechanical and structural stability of tissues. Desmosomal adhesion is particularly important in tissues subject to mechanical stress, for instance the heart and the skin. Experimental evidence suggests that desmosomes are involved in modulation of signalling pathways by regulating the availability of signalling molecules, especially plakoglobin [86]. Moreover, desmosomes, and particularly desmoplakin, are required for the formation of the blastomeres in the early stages of embryonic development [87].

The relevance of desmosomes can be understood if we look at the structure and function of this cellular complex. Two adjacent cell membranes are directly joined by the cadherins desmoglein and desmocollin, membrane spanning proteins of desmosomes. Other desmosomal proteins occupy part of the intracellular plaques of each of the adjacent cells (Fig. 1.4). Adjacent to the membrane is the outer dense plaque (ODP) (Fig. 1.4), 15-20

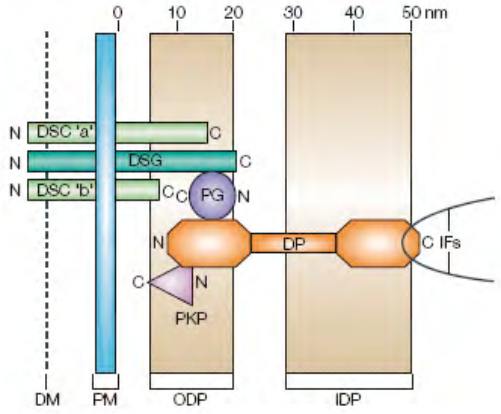


FIGURE 1.4: Desmosome map showing the dimensions of each one of its components, the relative position between them from the plasma membrane (PM) and their orientation (indicated by the N and C-terminals for each protein in the diagram). The N-terminal ends of desmoglein (DSG), desmocollin 'a' and desmocollin 'b' (DSC 'a' and DSC 'b' respectively) start in the extracellular dense midline (DM), while their C-terminal ends are in the outer dense plaque (ODP). In this region there is also plakoglobin (PG), plakophilin (PKP) and the N-terminal end of desmoplakin (DP). The C-terminal end of desmoplakin is located in the inner dense plaque (IDP) connecting with intermediate filaments (IFs). From reference [88].

nm thick [89], which allocates the cadherin C-tails, plakoglobin, plakophilin and the N-terminal end of desmoplakin (Fig. 1.5). The rest of the desmoplakin protein, namely the rod domain and the C-terminal end domain, is located in the inner dense plaque (IDP), where the latter domain associates with a network of intermediate filaments.

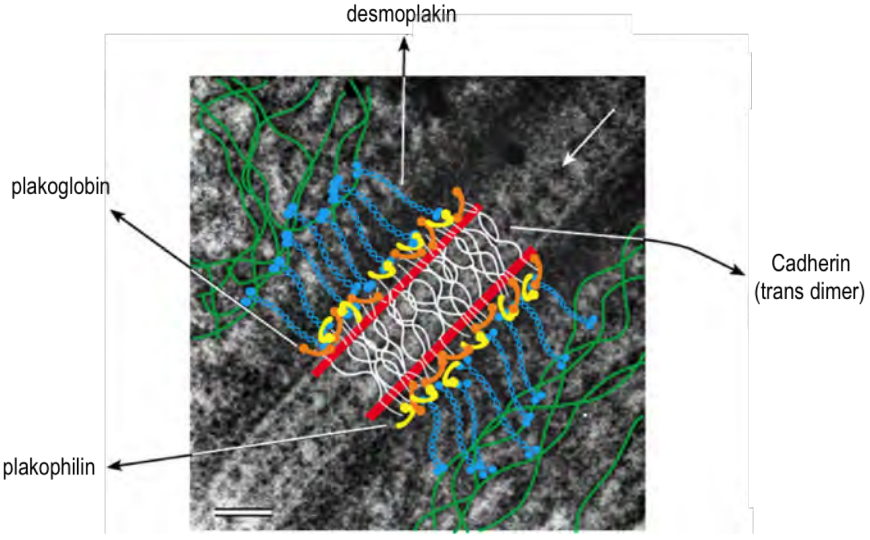


FIGURE 1.5: Protein components of the desmosome superimposed on an electron micrograph. Each desmosomal protein is depicted in a different colour. The membranes of neighbouring cells are shown in red; between them there are the cadherin molecules (white), mediating intercellular interactions. Plakoglobin (orange) and plakophilin (yellow) are located adjacent to the membrane in the outer dense plaque. The N-terminal end of desmoplakin, in blue, is located in the outer dense plaque whereas its C-terminal end associates with the intermediate filaments (green). Adapted from reference [90].

Plakoglobin and plakophilin are two armadillo proteins, which is a family of proteins characterised by the presence of a central Arm repeat domain [87]. The plakoglobin N- and C-terminal domains have approximately 100 residues each, while about 300 residues constitute the N-terminal domain of plakophilin. The latter interacts with the components of the ODP including desmoplakin.

1.5.3 Desmoplakin Structure and Function

There are two possible forms for this macromolecule, desmoplakin I and desmoplakin II, the first being almost two times longer than the second [91]. The difference in length is due to the lack of a 599-residue region in the rod domain of desmoplakin II. Although both of them can be found in all epithelial tissues, no desmoplakin II proteins are found in the heart [82]. Given the above facts, this study focuses only on desmoplakin I, which from now on will be called desmoplakin.

Desmoplakin has a molecular mass of 331.773 kDa, equivalent to 5.5 x 10^{-19} g [92] and calculated without experimental error from the amino acid sequence in the ExPASy Proteomics Server [93]. It is composed of 3 domains, two globular end domains, the N- (amino) and C- (carboxy) terminals ends, with molecular weights of 101.2 and 101.8 kDa respectively [80], connected by a rod domain that forms an alpha-helical long coiled coil with predicted flexibility that may enable the desmoplakin molecule to bend and adopt different possible configurations (Fig. 1.6) [90].

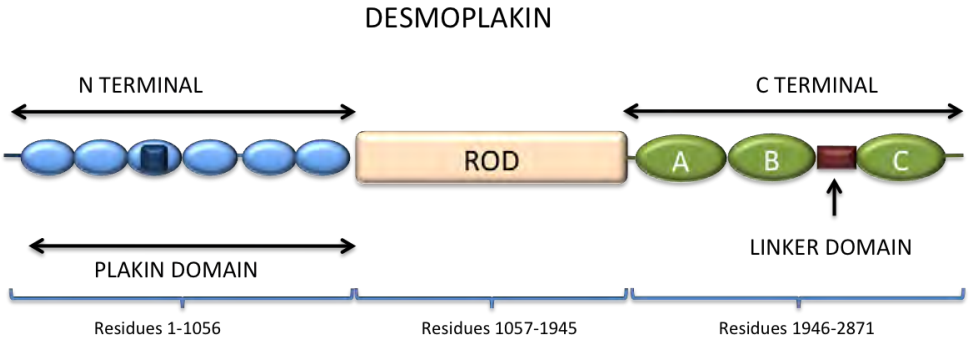


FIGURE 1.6: Structure of desmoplakin. The N-terminal head domain comprises the plakin domain, which can be sub-divided into a number of smaller spectrin repeats (SR) domains and a 180 residue domain at the extreme N-terminus which is predicted to be unstructured. The C-terminal tail domain is comprised of three plakin repeats *A*, *B* and *C*. Adapted from [94].

Electron tomography has been carried out on desmoplakin in an intact desmosome, predicting that the total length of desmoplakin is around 40 nm under native conditions, with N- and C-terminal domains having a globular shape with a diameter between 5 nm and 10 nm each [90]. On the other hand, rotary shadow electron microscopy images of denatured desmoplakin show the central rod of about 130 nm connecting the two globular domains, and a total length of 180 nm under these conditions [91] (Fig. 1.7). This studies of desmoplakin also showed that the end domains are varied from globular (with a diameter of 16 nm), to oblong, extended linear structures (with a diameter of 32 nm) [91]. A comparison with the dimensions given by the tomographic studies suggests that the rod domain is folded up within the IDP [90] and also the existence of other folded regions under native conditions rather than true globular domains [91].

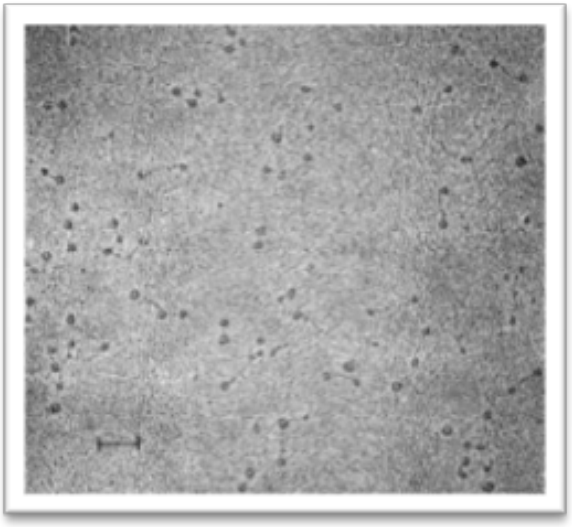


FIGURE 1.7: Rotary shadow electron microscopy images of purified desmoplakin. The sample (with a concentration of 10 μg protein/mL, sprayed on freshly cleaved mica, dried under vacuum) was shadowed at an angle of 5 ° with platinum, floated in copper grids, viewed in a Phillips 300 electron microscope and photographed at a magnification of 50,000. This image shows the rod domain (average length 130 nm) and the end regions with a globular configuration. Bar = 100 nm. From reference [91].

Since desmoplakin extends from the inner plaque to the outer plaque, it interacts with most desmosomal components, and therefore plays an essential part in linking the membrane to the cytoskeleton. Comparing the experimentally obtained dimensions of the protein in its native environment with the dimensions it has in non-native conditions, it is reasonable to assume that desmoplakin, under native conditions is folded in order to provide extensibility [91], useful in shear stress situations to which the tissue could

be subjected. This feature indicates that desmoplakin flexibility could confer flexibility upon the whole desmosome.

1.5.3.1 The N-terminal end of desmoplakin

The N-terminal end of desmoplakin, comprising 1056 residues, contains the characteristic plakin domain (residues 180-1022) common to most of the plakin family members. The plakin domain consists of six tandem spectrin-like repeats (SR) and an inserted SH3 domain [95] (Fig. 1.8). A SR3 to SR6 fragment from the plakin domain, including the SH3 domain, has been crystallised. The study found an elongated rigid rod structure with approximate dimensions of 18×4×4 nm [96] (Fig. 1.9 (a)).

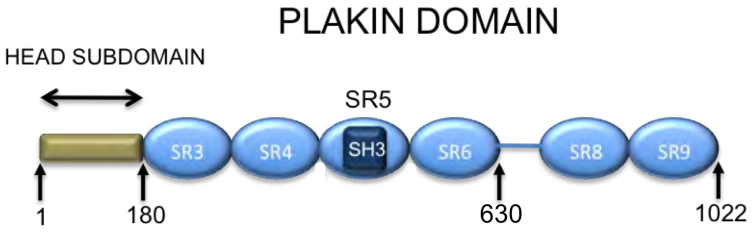


FIGURE 1.8: Schematic of plakin domain spectrin repeats (SR). SR5 has a putative Src homology 3 (SH3) domain embedded within its structure. The plakin domain starts at residue 180 and finishes at residue 1022. Residues 1-180 at the extreme N-terminus are predicted to be unstructured. Adapted from [97].

This fragment of the desmoplakin PD has been fitted to a SAXS envelope generated for the complete PD molecule that shows an L-shaped conformation with maximum diameter of 33.5 nm, consisting of a long arm of 24 nm and a short arm of 17.9 nm (Fig. 1.9 (b)) [98]. This is in contrast with the 5-10 nm diameter globular domain reported by electron tomographic studies of desmoplakin in an intact desmosome [90], but within the dimensions proposed for the different conformations of the domain by electron microscopy studies of the protein [91].

The N-terminal end is responsible for plaque attachment, and for the interaction with plakophilin and other components of the ODP. It also may be responsible for clustering of cadherins at the membrane, a critical function during desmosome assembly [90].

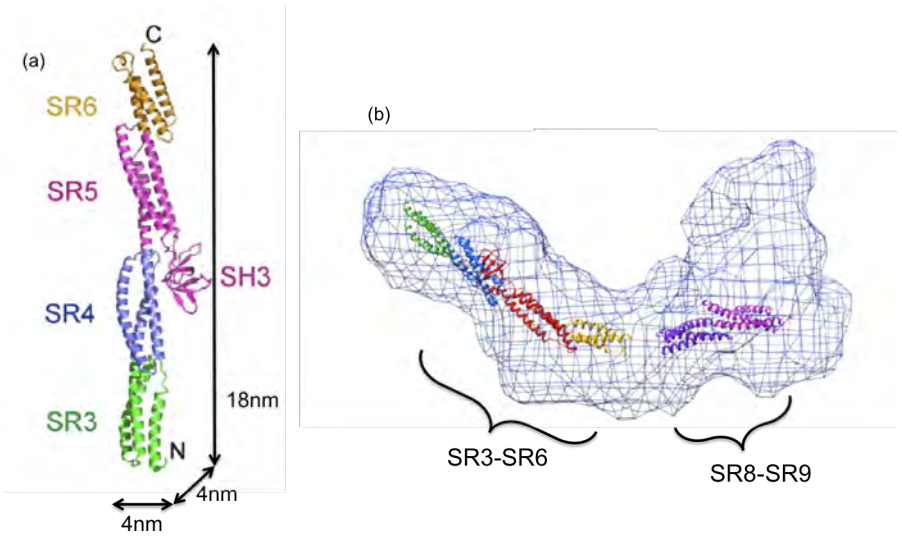


FIGURE 1.9: (a) Ribbon representation of the structure of desmoplakin plakin domain segment that comprises residues 175-630 (SR3-SR6, including SH3), obtained by X-ray crystallography. The approximate dimensions of the segment are 18×4×4 nm. From reference [96]. (b)Three-dimensional model of plakin domain envelope based on experimental SAXS data, revealing an L-shape conformation with a long arm of 24 nm and a short arm of 17.9 nm and a maximum diameter of 33.5 nm. From reference [98].

1.5.3.2 The C-terminal end of desmoplakin

X-ray studies have shown the C-terminal end of desmoplakin (residues 1946-2871) comprises three plakin repeat domains (PRDs A, B and C) equivalent to 862 amino acids (residues 1960- 2822) [80, 94] (Fig. 1.10). Of these, B and C have been characterised by X-ray crystallography. The results are in agreement with computer-aided analysis, which suggested that the carboxyl terminus contains three homologous regions that fold into a globular conformation stabilised by intra-chain ionic interactions [99]. Different theories about the configuration of the PRDs have been proposed, from which probably the most popular for desmoplakin is the one by Choi *et al.* who suggest a "beads-on-a-string" arrangement [94], while, for the case of plectin, Janda *et al.* proffer an architecture in a circular arrangement with the first 5 PRDs surrounding the 6th PRD in an antiparallel fashion [100].

The C-terminal end of desmoplakin displays a structural unit based on 4.6 copies of a 38-residue plakin repeat (PR) [80] that forms each of the three plakin repeat domains [99]. Plakin repeats have a periodic distribution that intercalates acidic and basic residues, the same configuration that exists in vimentin (an intermediate filament protein described in Section 1.5.4.1). This was one of first findings that suggested the existence of

ionic interactions between the desmoplakin C terminus and intermediate filaments [101]. Plakin repeats are also found in the carboxyl terminus of plectin, a plakin protein that is known to bind to IFs [102].

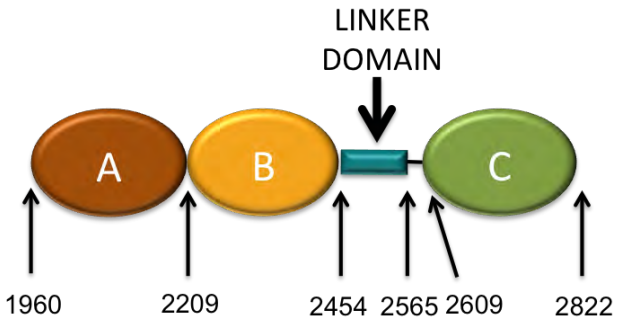


FIGURE 1.10: Schematic of C-terminal end of desmoplakin. It comprises three plakin repeat domains equivalent to 862 amino acids (residues 1960- 2822) [80, 94]. The plakin-repeat domains are usually called A, B and C. Also present is a linker domain. The region between the linker domain and the PRD C has been referred as a spacer region.

1.5.4 Interactions with Intermediate Filaments

Desmoplakin binds the intermediate filaments through its C-terminal domain [101]. The IFs are one of three components that constitute the cell cytoskeleton. The cytoskeleton of eukaryotes confers flexibility and shape upon cells and on the other hand transports different components within them. The other two elements of the cytoskeleton are microfilaments and microtubules that work as cell motility agents and as re-distributors of cell components respectively [103]. IFs are the most stable of the three components, and are found primarily in cells that must resist shearing stresses, providing strength to the cell [103]. They are anchored by the desmosomes to the plasma membrane of several cell types [104] and collaborate with cell-cell junctions to generate transcellular networks integrating individual cells into tissues [105].

A particular feature of the IF cytoskeleton that differentiates it from the other cytoskeletal fibres is its heterogeneous nature. IFs are classified in six types with different properties, among which the most common are the keratins (types I and II). Other examples include vimentin and desmin (type III), the neurofilaments (type IV) and the nuclear lamins (type V) [103]. However, they all share some structural features: a central rod domain with secondary structure mostly alpha helical and two non-alpha helical domains classified as the head and tail domains. The rod domain is sub-divided in four

sections with heptad periodicity, coiled coils 1A, 1B, 2A and 2B, connected by linkers [106]. These coiled coil dimers, which are polar in nature, constitute the basic structure of IFs and it has been proved that under physiological conditions these tend into anti-parallel tetramers, provoking loss of polarity [105].

1.5.4.1 Vimentin and desmin

Although vimentin and desmin share structural characteristics, they have been found in different places in the human body, performing diverse tasks. Vimentin is more widely distributed, being considered the cytoskeletal component responsible for maintaining cell integrity [107], while desmin is restricted to smooth and cardiac muscle [108]. As mentioned, vimentin and desmin are classified as type III IFs, having a particular structure in common. The head domain of IF type III is responsible for filament assembly and dimer-dimer interactions, while the tail domain is in charge of integration of the filaments and interactions with other proteins and organelles [109]. As to the rod domain, it represents the fundamental building block that assemblies the IF structure [105].

In a primary structure analysis of vimentin and desmin it is possible to see an intercalation of acidic and basic amino acids with a slight prevalence of acidic residues in both rod domain amino acid sequences, providing stability to the coiled coil dimers [108]. The exact region of the C-terminal of desmoplakin that binds desmin and vimentin has not yet been identified. Ascertaining the specificity of this binding is primordial for the understanding of the integration of the cells within the tissue, which has repercussions for the function of the entire human body. The correct binding of IFs with desmosomal proteins provides a stress-resistant scaffolding vital for tissues such as heart muscle. Skin diseases are often the result of compromised tissue integrity, generally a consequence of an improper anchor between IFs and plakins, leading to severe skin fragility [110].

Mutations in desmin have been implicated in 80 human diseases [106]. Malfunction of desmin can cause muscular and cardiac dystrophies being able to cause fatal damage to the heart [105]. For its part, over-expression of vimentin has been related to various epithelial cancers including prostate cancer, gastrointestinal tumours, breast cancer, malignant melanoma and lung cancer [111].

1.5.5 ARVC and other Diseases Related to Desmoplakin Mutations

Mutations in desmosomal proteins can result in loss of cell adhesion in tissues subjected to stretching forces. Lack of functional desmosomes implies a long list of serious health issues such as thickening of the epidermis because of reduction in cell adhesion [112]; skin blistering caused by degeneration of keratinocytes; and reduction in hemidesmosomes with altered stability resulting in muscular dystrophy phenotype [113].

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a type of cardiomyopathy characterised pathologically by the replacement of cardiomyocytes by fibrofatty tissue. The fact that it tends to happen in the right ventricle gives the name to the disease [114]. ARVC is distinguished for causing ventricular arrhythmias that result in cardiac death, especially in young competitive athletes. The prevalence of ARVC is thought to be 1 in 5000 individuals [114].

Desmoplakin was the first gene implicated in autosomal dominant ARVC. After analysis, a heterozygous mutation (2034insA) resulted in a premature chain termination that caused the truncation of the intermediate filament-binding site of desmoplakin in the C terminus of the protein. The phenotype associated was left ventricular arrhythmogenic cardiomyopathy. A missense mutation (S299R) in the N-terminal domain of desmoplakin was discovered to cause suppression of a putative phosphorylation site that binds to plakoglobin, resulting in classic ARVC. Other mutations in desmoplakin (R1775I, R1255K and the splicing mutation c.423 to 1G<A) lead to chest pain and myocardial enzyme increase [4]. Another example is a mutation that involves pemphigous-like skin syndromes and woolly hair as well as ARVC. This mutation (G2375R) in the C-terminal end of desmoplakin is closely related to the protein's ability to bind IF [115].

Mutations in desmoplakin can also be the cause of the Carvajal syndrome, disease characterised by wolly hair, cardiomyopathy and striate palmoplantar keratoderma (PPK). PPK is a skin disease characterised by thickening of the skin, most commonly on the palms of hands and soles of feet[4, 116].

What increases the critical character of understanding ARVC and the desmoplakin mutations related to its pathogenesis is the possibility that a considerable number of sudden deaths that have been attributed to myocarditis are in fact caused by early stages of arrhythmogenic cardiomyopathy [4].

1.6 Global and Specific Aims of this Work

Given the importance of desmoplakin as a vital component of desmosomes, a molecular underpinning of its structure and function is essential for a better understanding of its role in human diseases.

On this basis, the first general aim of this work focuses on the C-terminal end of desmoplakin and intends to shed light on its interactions with intermediate filaments, paying especial attention to the linker domain given its proposed key role in IF binding. Thus, the first specific aim of this investigation is to reveal the atomic structure of the desmoplakin linker domain using NMR spectroscopy and SAXS. Knowledge of that structural information will provide functional information about the C-terminal domain of this plakin protein. Considering the hypothesis raised in previous publications of that the binding properties are not exclusive of the linker domain, but shared by the different elements of the desmoplakin C-terminal tail, a second objective is to visualise the arrangement of the three plakin repeat domains of desmoplakin, including the linker domain, in an envelope generated by SAXS. In this way, more evidence on how the C-terminal end of desmoplakin is organised can be obtained to either confirm or disprove previous theories on the arrangement of domains within the C-terminal tail. A third aim is to ascertain the actual nature of desmoplakin-IF binding by finding the precise interaction sites. To this end a series of NMR titration experiments will be carried out between the desmoplakin linker domain and vimentin and desmin. NMR titrations with the periplakin linker domain and vimentin constitute a comparative exercise planned with the intention of complementing the analysis carried out on desmoplakin. The experiments with mutant desmoplakin and periplakin linker proteins will also be performed to identify residues important for the interaction with intermediate filaments as another particular aim of this work.

Beyond its interactions with other proteins, it has been considered of interest to explore the behaviour of desmoplakin on functionalised substrates, constituting this the second general aim of this research work, which focuses on the N-terminal end of desmoplakin.

In the frame of the development of emerging techniques that combine metal nanoparticles in medical research at the molecular level, here the possibility of creating an array of isolated desmoplakin plakin domain molecules is explored. The first particular goal is to

generate an immobilisation state of desmoplakin plakin domain isolated molecules on a graphite surface mediated by the presence of gold size-selected clusters. The achievement of this objective will be tested by non-contact and tapping AFM modes in ambient and liquid phases respectively, the latter using buffer solution to emulate the native conditions of the plakin domain protein. The strength of the immobilisation of the desmoplakin plakin domain on graphite pre-decorated with gold nanoclusters is aimed to be tested with contact mode AFM in native conditions with the intention of investigating a new route for the physical analysis of the desmoplakin PD and lay a basis for a potential force spectroscopic measurements of its mechanical properties.

The study of desmoplakin also served as a promoter for the last aim of this work: the development of a novel method for size-selected clusters immobilisation, directed to the improvement of protein immobilisation platforms. It is intended to prevent soft-landed clusters from diffusing, resulting on a graphite substrate decorated with stable immobilised gold clusters that have not lost their morphological properties and also that are amply above the AFM resolution. Such nanostructured platforms could also have potential applications in electronics, catalysis and plasmonic devices.

Chapter 2

Materials and Methods

This chapter presents the methodologies used during this research, starting with the laboratory techniques for protein production and physicochemical analysis. This is followed by a description of the instrumentation and data acquisition procedures associated with each one of the physical techniques (NMR, SAXS and AFM) employed in this work for the study of the desmoplakin protein structure and interactions. Finally, the methods related to the size-selected clusters are briefly explained, including the functioning of the cluster source and the deposition techniques.

2.1 Constructs

The constructs used in this work are shown in Table 2.1 .

All desmoplakin domains constructs were provided by Dr. M. Chidgey. Constructs encoding full length vimentin and desmin were obtained from ShineGene (Shangai, China). Details of the vectors used can be found in Appendix A.1.

Construct Name	Residues	Estimated Molecular Mass	Vector	Tag
Desmoplakin Plakin Domain	180-1022	99118 Da	pGex6p-1	GST and 6x His
Desmoplakin Linker Domain	2454-2565	12935 Da	pGex6p-1	GST
Desmoplakin Linker E2495K-C2497R	2454-2565	12935 Da	pGex6p-1	GST
Desmoplakin Linker S2526K-Q2527K	2454-2565	12935 Da	pGex6p-1	GST
Desmoplakin PRD C	2609-2822	23270 Da	pProEX HTc	6x His
Desmoplakin PRDB-Linker-PRDC	2209-2822	63947 Da	pProEX HTc	6x His
Periplakin Linker Domain	1646-1756	12870 kDa	pGex6p-1	GST
Periplakin Linker K1687E-R1689E	1646-1756	12870 kDa	pGex6p-1	GST
Periplakin Linker K1717E-K1718E	1646-1756	12870 kDa	pGex6p-1	GST
Periplakin Linker R1655E-R1656	1646-1756	12870 kDa	pGex6p-1	GST
Vimentin full length	1-466	53520 Da	pET21a+	None
Desmin full length	1-470	53535 Da	pET21a+	None

TABLE 2.1: Table with specifications of each protein that appears in this work. The plasmids were provided by Dr. Martyn Chidgey in every case except for vimentin and desmin proteins, which were synthesized by ShineGene.

2.2 Protein Expression

2.2.1 Transformation of BL21(DE3) Cells

Fully sequenced clones were transformed into competent *E. coli* strain BL21 cells by adding 1 μL of circular DNA to 20 μL of the competent cells. The cells were incubated for 20 min on ice, after which they were subjected to a heat shock at 42 °C for 30 sec. Afterwards 80 μL of Super Optimal Broth (SOC) media was added and the cells were incubated at 37 °C and 220 rpm for 1 hour. Bacteria were plated onto Luria Broth (LB) agar plates with ampicillin (50 μg/mL) grown at 37 °C for 12 h.

2.2.2 Overnight Growth of Starter Cultures and Large Scale Expression

For starter cultures a colony was picked into 10 mL of LB medium containing ampicillin (50 μg/mL) and grown at 37 °C at 220 rpm for 12 h. For large-scale cultures the starter culture was added to 1 L of LB medium or Minimum M9 media containing Nutrient

Mix for ^{15}N labelled proteins in presence of ampicillin (50 μg/mL) at 37 °C at 220 rpm. After reaching an optical density value between 0.4 and 0.6 (measured by absorption spectrophotometry at 550 nm with a BioMate3 spectrophotometer) the temperature was decreased to 18 °C. Cells were then induced with 1 mM isopropyl-β,D-thiogalactopyranoside and grown for further 12 h.

Details of the buffer solution compositions mentioned in this and following sections can be found in Appendix A.2.

2.3 Protein Purification

2.3.1 Harvesting Cells and Mechanical Cell Lysing

Cells were harvested by centrifugation at 4000 rpm at 4 °C for 20 min in an Avanti J-20XP centrifuge with JLA 8.1 rotor and re-suspended in 20 mM sodium phosphate (pH 7.4) with protease inhibitors (Roche Complete). Cells were lysed with an Emulsiflex (Avestin) according to the manufacturers' instructions, and lysates were cleared by centrifugation at 25000 rpm for 45 min at 4 °C in a Avanti J-25 centrifuge with a JA25.50 rotor, and filtration with 0.2 μm filters.

2.3.2 Purification of GST-tagged Proteins including Removal of the GST Tag

Bacterial lysates were loaded onto a 5-mL GST-trap column (GE Healthcare) previously equilibrated with 5 column volumes of PBS buffer. Eluted GST fusion proteins were loaded onto the column over the course of 12 h, using a loop system with a pump to recirculate eluted material to the top of the column. Using this procedure the majority of the GST fusion proteins in the lysate was bound to the column. The GST protein was then eluted from the column with glutathione (26 mM). GST fusion proteins were cleaved with PreScisson protease (GE Healthcare). 1 μL of PreScisson was added to the protein per each mL of eluted protein (normally around 20 mL). The sample was then dialysed in 2 L of 20 mM HEPES pH 7.5 (buffer solution) at 4 °C for at least 12 h.

To remove the GST tag, cleaved GST fusion proteins were loaded on to a 5 mL pre-equilibrated Hi-Q Sepharose ion exchange column. After washing with buffer solution

proteins were eluted using an AKTA prime system and a linear salt gradient, starting with low salt buffer and finishing with high salt buffer. Ion-exchange columns are populated with negatively charged beads. All proteins with positive or partially positive surface charge bind to the anionic groups when the sample is loaded on to the column. By applying a salt gradient (i.e. increasing the ionic strength), a competition between the Cl^- ions and the bound proteins is created and proteins with the lowest net charge are eluted first from the column. In this way, proteins with different charge at the same pH are eluted in different elution fractions. In some cases GST was removed following PreScission cleavage by passing PreScission cleaved proteins through a GSTrap column.

2.3.3 Purification of His-tagged Proteins

Following lysis, bacterial lysates were loaded onto a 5-mL His-trap column (GE Healthcare) previously equilibrated with 5 column volumes of equilibration buffer. The sample was loaded onto the column over the course of 1 h using a loop allowing the histidine tagged proteins to bind to the Ni^{2+} charged beads inside the column via metal ion affinity chromatography. Proteins were washed with washing buffer and eluted with elution buffer (see buffer compositions in Appendix A.2). Fractions containing the required protein were dialysed with TEV protease (1.7 mg/mL per each mL of sample) at room temperature for at least 12 h.

2.3.4 Purification of Vimentin and Desmin

The purification protein for both intermediate filaments (IFs), vimentin and desmin, was the same. Following large scale expression, the soluble fraction was discarded, keeping the insoluble fraction containing IF protein. The insoluble fraction was re-suspended in Triton buffer and then centrifuged at 13500 g for 15 min. The result was new soluble and insoluble fractions, from which the insoluble fraction was again re-suspended in Triton buffer followed by centrifugation at 13500 g for 15 min three times. Finally, the pellet was washed with a washing buffer that did not contain Triton X-100. The washed pellet was re-suspended in solubilisation buffer and left on a rocker for 12 h at 4 °C to solubilise. Then, the sample was centrifuged at 15000 g for 30 min. This time, the protein was in the soluble fraction. In order to obtain folded pure IF, it was necessary

to dialyse the protein in dialysis buffer step wise with different concentrations of urea: 4M, 2M and 0M.

2.4 Biochemical Analysis of Proteins

2.4.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Protein samples were analysed by using Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE). Precast gels (Bio-Rad) were used at 180 V for 35 min in 2-(N-morpholino)ethanesulfonic acid (MES) SDS running buffer. 5 μL of loading dye (Laemmli 2x buffer, Sigma) was added to 5 μL of samples prior to running the gel. After electrophoresis, the gels were stained with Simply Blue Safe Stain (Life Technologies) for 1 h and then washed with water.

2.4.2 Protein Concentration

Protein samples were concentrated using Vivaspin or Amicon concentrators of the appropriate molecular weight cut off depending on the size of the protein (3000 or 10000 MWCO). Samples were centrifuged using an Eppendorf Centrifuge 5810R (4000 g at 4 °C for 10 min) repeatedly until the desired concentration was reached. Between each repetition samples were carefully pipetted up and down in order to avoid precipitation of the protein. Protein concentration was measured indirectly via the optical density of the sample (concentration in M = (Abs_{280nm} x dilution factor) / (extinction co-efficient ($M^{-1}cm^{-1}$) x path length (cm)) using a UV spectrophotometer at an absorption of 280 nm.

2.5 Biophysical Analysis of Proteins

2.5.1 Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC) is a technique based on the application of an effective centrifugal force to macromolecules in solution that tend to sediment as a function of time. There is a direct proportionality between the sedimentation coefficient and the

molecular weight and an inverse proportionality with the frictional coefficient given by the viscosity of the solution.

In order to determine the multimeric state of the proteins, AUC was carried out. The studies were conducted by Mrs Rosemary Parslow at the Birmingham Biophysical Characterisation facility (University of Birmingham). Samples, dialysed in PBS buffer, were centrifuged at 40000 rpm with the absorbance scans taken at 280 nm. Three concentrations of each sample are used for the equilibrium runs. Centrifugation was performed using a Beckman XL-1 analytical ultracentrifuge (Beckman Coulter). Sedimentation profiles were analysed with a macromolecular size and shape distributions software that uses sedimentation velocity AUC (SEDFIT)[117].

2.5.2 Circular Dichroism

Circular dichroism (CD) is an optical technique based on the difference in absorbance of right and left polarised light depending on the symmetry of the medium that is traversed by circularly polarised light. If the body has a structural asymmetry, the absorption of the two different states (right and left) will not have the same amplitude due to the resultant differences in velocities and wavelengths. Protein secondary structure can be predicted as an average of the entire molecular population, determining percentages of alpha helical, beta sheet and random coil structures due to their distinctive shape and magnitude of CD spectrum. Alpha helices show two minima at 208 nm and 222 nm, beta sheets show a single minimum at 217 nm and random coil structures reflect a minimum in 204 nm [118].

To confirm the folded state of the proteins and to investigate secondary structure, CD analysis was carried out. The studies were conducted by Dr. Raul Pacheco-Gomez from the School of Biosciences (University of Birmingham). Samples, dialysed in 10 mM sodium phosphate buffer and at a concentration of 0.1 mg/mL, were introduced in a JASCO-J810 sprectrometer. Spectra were collected at 190-260 nm and analysed with DichroWeb server [119], using the CDSSTR analysis programme.

2.6 Nuclear Magnetic Resonance Methods

The basic principles of the nuclear magnetic resonance spectroscopy technique will be covered at the beginning of Chapter 3.

2.6.1 Sample Preparation

Protein preparation for nuclear magnetic resonance spectroscopy includes labelling with N^{15} or/and C^{13} depending on the kind of experiments to be performed. High levels of purity (more than 90%) are required, tested by SDS PAGE, and a minimum concentration between 50 μM and 500 μM depending on the protein molecular weight and also on the experiment. Samples in 20 mM HEPES pH 7.5 are loaded on 5 mm diameter NMR tubes with 10% D_2O. A final total volume of 600 μL is required to cover at least 5 cm of the tube and avoid end effects once inside the spectrometer. Tubes with sample need to be measured with a provided sample depth gauge so the sample position coincides with the receiver coil area of the spectrometer.

2.6.2 Instrument Characteristics

NMR spectrometers consist of 4 main elements: a superconductive magnet, a number (12 to 32) of shim coils, a radio frequency (RF) coil (transmitter and receiver) and a triple resonance cryogenic probe. Inside the spectrometer, the sample is positioned into the probe between the poles of the magnet in proximity to the transmitter RF coil and surrounded by the receiver RF coil. The shim coils are located strategically around the sample in the magnet to affect different components of the magnetic field (Fig. 2.1).

The spectrometers used were the Agilent 600 MHz, 800 MHz and 900 MHz at the Henry Welcome Building NMR facility at the University of Birmingham (Fig.2.2).

2.6.3 Levels Set Up

After the sample has been loaded into the spectrometer and thermal stability has been reached, the levels of lock, shimming and tuning must be optimised in the control programme, which in this case was the Agilent computer software package VnmrJ 3.

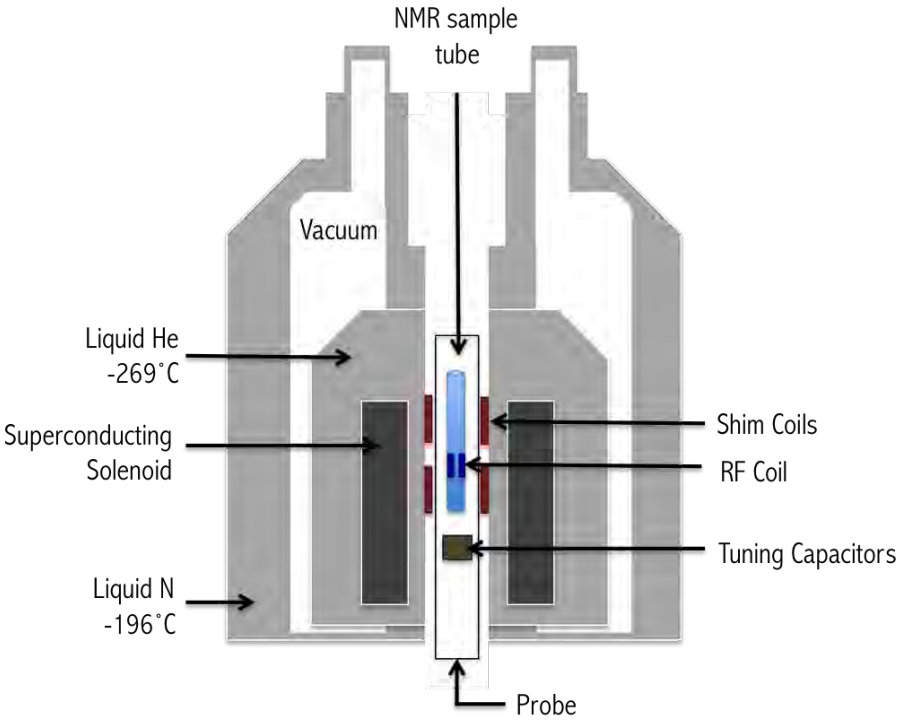


FIGURE 2.1: Schematic representation of a nuclear magnetic resonance spectrometer. The superconductive solenoid or magnet produces the magnetic field B_0 used in the NMR experiments. The shim coils homogenise the magnetic field produced by the superconductive magnet. The RF coil produces the magnetic field that rotates the spins by 90 °C or 180 °C from the B_0 direction. Liquid nitrogen and liquid helium are used to cool down the superconducting magnets to cryogenic temperatures during operation. Adapted from reference [120].

FIGURE 2.2: NMR 900 MHz spectrometer, at the Henry Wellcome Building, School of Cancer Sciences, University of Birmingham.

The lock is a system that ensures a constant magnetic field in time, compensating the possible magnetic drift due to external interference. This system is in essence an independent spectrometer within the main NMR spectrometer that is tuned with the deuterium frequency (at the specific magnetic intensity flux in question) and detects the deuterium signal from the D_2O component of the sample. Fluctuations of the magnetic field are reflected in changes in the resonance frequency of the deuterium, which is continuously monitored, therefore a re-calibration of the magnetic field can be performed in real time. The magnetic field needs to be not only continuous, but also homogeneous in order to limit the variation of the field across the sample. For this there is the shimming system that consists of several coils located in different positions and directions with respect the magnetic field around the sample. Any field inhomogeneity in the sample area (measured by proton density map of the sample) causes a magnetic gradient that can be compensated by the respective gradient produced by the set of shimming coils.

Lock and shimming are related, since homogeneity of the field is directly proportional to the lock signal intensity, therefore, both parameters are optimised alongside. The lock level is first adjusted from the computer programme control to a minimum constant value by making changes in field strength, phase, gain and power. Then an automated acquisition of gradient shim in Z dimension (with previously mapped shims) is carried out, followed by a final readjustment of the lock level. After this process the magnetic field of the spectrometer should be continuous and homogeneous for the sample during data acquisition [121].

In order to improve signal-to-noise ratio tuning and matching the probe is necessary for each of the active isotopes. Tuning is the correction made to the resonant circuit frequencies of the probe to coincide with the resonant frequencies of each of the active decoupled nuclei. Matching is the change in the impedance of the resonant circuit to modify the amplitude of the frequency peaks. Both regulations are made directly in the spectrometer using a knob normally in the bottom part of it.

As in any other spectroscopy technique, in NMR spectroscopy it is necessary to define the signal background and minimise it as low as possible. In liquid NMR the solvent is normally water, which produces a strong signal detectable by the system given its high concentration of protons (110 M in pure H_2O). Considering that regular concentrations

of sample are lower than 1 mM, it is not difficult to conclude that the water signal must be suppressed in order to detect the signal coming from the sample. In order to achieve this, water resonant frequency is selectively saturated with a low power RF pulse causing saturation, which in this case means balance the two possible spin states of the water protons. There being no difference in spin states, no NMR signal is produced by water molecules upon exciting the system and ideally the only signal detected will be the one coming from the sample. Finding the best saturation frequency is also conducted from the control programme in the computer [122].

2.6.4 Data Acquisition

For data acquisition the parameters related to pulse sequence settings need to be adjusted, such as acquisition time, pulse width, relaxation delay, spectral width, number of data points and total scans requested.

The acquisition time is the length of time needed to acquire each free induction decay (FID) signal, which is the NMR observable result of the non-equilibrium magnetisation precessing of the nuclear spin about the magnetic field. The time scale in this case is sec, so it is long enough to avoid truncation artifacts. The pulse width is the time length, in the order of μsec, of the excitation RF pulse necessary to detect maximum signal from a desired nucleus in a given sample, therefore it has to be calibrated for every new sample according to the pulse power and the probe tuning. This is done by making an array of pulse width data for a single resonance, and determining the value that corresponds to a 90° pulse, which is the one that provides maximum signal. The relaxation delay is the time duration between FIDs required to allow spin populations to return to their original state. It is applied at the beginning of each FID pulse to avoid saturation states that result in little or no signal. The spectral width is the spectral observation window that incorporates all resonances. This constraint must be anticipated based on the expected positions of the resonance peaks depending on the spectrum to produce. Ideally it is corrected to encompass just as much spectral space as necessary in order to avoid unnecessary noise, but at the same time evading peak folding. The number of complex data points (np), acquired during acquisition time, will be proportional to the spectral width (sw) times the acquisition time (at): $np = 2 \times sw \times at$; while the dwell time ($dw$) is the time interval between the acquisition of individual data points, inversely

proportional to sw: $dw = 1/(2\times sw)$. Finally, the total number of scans is the number of repetitions to perform in order to obtain a high resolution resultant spectrum [123].

Details of spectra acquisition can be found in Appendix B.1.

2.6.5 NMR Titrations for Binding Experiments

In general terms the methodology can be described in three steps: first the acquiring of an HSQC spectrum of the initial ^{15}N labelled host protein, that will provide the resonance values for the free conformation. Then the gradual addition of the external component (unlabelled protein) to the system, (initial ^{15}N labelled protein) and the acquiring of a new HSQC that provides with the resonances of the bound conformation system. Finally the observation of any perturbation due to the presence of the ligand in the host system by comparing the first and last HSQC spectra [124].

Host proteins are always labelled with ^{15}N and loaded in an NMR tube at an initial concentration of 100 μM and an HSQC spectrum is acquired in each case for a ratio 1:0 of host:ligand. A second NMR tube is prepared with a sample containing the host proteins at 100 μM and the ligand of choice at 100 μM for a ratio 1:1 (molar ratio will depend on the specific proteins present in the sample). Different concentrations are used depending of the desired ratio, but a minimum concentration of 100 μM for the host protein is always ensured. The use of two tubes instead of one avoids dilution factors to considerate during analysis. All titration experiments of this research were performed with a 600 MHz NMR spectrometer.

2.7 Small Angle X-Ray Scattering Methods

2.7.1 Sample Preparation

In each case, purified protein (tested by SDS PAGE) was divided in three samples with different concentrations chosen in a range between 1 mg/mL and 10 mg/mL. Proteins were dialysed into 20 mM HEPES pH 7.5, 100 mM NaCl. The dialysis buffer is also used in SAXS experiments for background subtraction. The importance of having three samples with different concentrations lies in the fact that different concentrations have inherent advantages and disadvantages that will be explained in Section 2.8.3.

2.7.2 Instrument Characteristics

The X-ray beam used in SAXS experiments is produced with a synchrotron light source, a particle accelerator that uses varying orthogonal magnetic and electric fields to focus and accelerate charged particles (electrons, positrons and protons) in a constant radial trajectory.

Synchrotron light sources are conformed by three main stages: production of particles to accelerate, pre-acceleration and storage. Production of electrons is generated from a metal cathode by applying a high voltage under vacuum resulting in thermodynamic emission. The particles move in direction of an anode in bunches produced by the accumulation caused by an on/off grid positioned in their way. Two accelerators compose the pre-acceleration process: a linear accelerator, also called linac, and a small synchrotron, commonly called booster synchrotron. Bunches containing billions of electrons, with an energy around 90 keV, are introduced in the linear accelerator of tens of meters length, and accelerated by a series of oscillating electric potentials caused by the tuning of the gradient of the electric field to the speed of the electron bunch: the accelerator is divided in drift tubes with increasing length that oscillate from positive to negative so the bunch of electrons is at all points after a negative field that repels it and before a positive field that attracts it (Fig. 2.3).

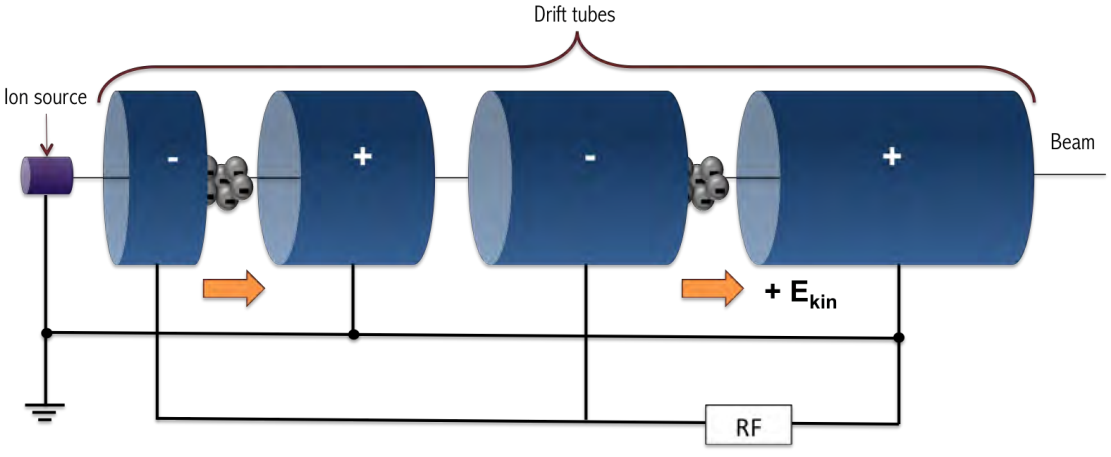


FIGURE 2.3: Scheme of the linear particle accelerator principle. A linear accelerator, or linac, is part of the pre-acceleration process in the synchrotron light source used for SAXS experiments. It consists of drift tubes with increasing length that oscillate from positive to negative to accelerate a bunch of electrons by positioning them at all points after a negative field that repels them and before a positive field that attracts them. Electrons increase their kinetic energy from 90 keV to tens of MeVs [125]

Located at one third of the linac there is a tungsten target to which the accelerated electrons are directed, provoking a reaction that creates positron-electron pairs. Positrons resulting from this reaction are further accelerated until the end of the linac and passed to the second stage of acceleration in the booster synchrotron. At this stage, positrons have low energy (50 - 100 MeV) and are not very well focused, therefore the need of a "boost". The booster synchrotron consists in two straight sections, in which positrons are accelerated with a RF voltage source, and two semi-circular curves, in which the acceleration is performed by a number of dipole bending magnets. The magnetic field produced by these magnets needs to be increased synchronously with the particles energy in order to maintain a fixed radius of curvature. Extracted positrons from the synchrotron booster are injected in the storage ring at energies around 4 GeV.

Finally, the positron beam enters the storage ring in which is kept under vacuum conditions. The storage ring, a toroidal structure, follows the same principles as the booster synchrotron except that it does not provide acceleration to the particles. Instead of that, the purpose of the RF voltage source is to replace the energy loss (approximately 3.5 MeV per turn) during the beam lifetime (the duration of time that the beam remains in the storage ring, typically 10-18 h), maintaining the energy constant. The magnetic field varying in time keeps the radial trajectory using the dipole magnets, while the application of quadrupole magnets focus the beam [125].

SAXS experiments were carried out with the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) as part of the European Molecular Biology Laboratory (EMBL) Hamburg Outstation. DORIS III (in operation until 2012) works at an energy of 4.45 GeV and a beam current of 140 mA of positrons. The radio frequency that it uses is 499.6 MHz with a RF voltage of 7.2 MV. Positrons are generated from the reaction of accelerated electrons in a tungsten target and then accelerated in the linear accelerator of 70 m long Linac II. Accumulation of positrons is done in the PIA accelerator (Positron Intensity Accumulator). Synchrotron radiation produced by the positrons traveling at relativistic velocities in the curved path of the storage ring is collected by one (X33) of the 36 beam-line stations and subsequently passed through an optical arrangement that selects the X-ray range of the spectra and focus the beam.

2.7.3 Data Acquisition

A monochromatic beam of X-rays with a wavelength of 0.15 nm was focused and directed to three samples of the protein in question at different concentrations along with a sample buffer, producing a scattering pattern recorded at room temperature by a single-photon counting pixel detector (PILATUS 500 K). High concentration samples tend to present interference generated by interparticle effects, leading to aggregation/oligomerisation problems. On the other hand, high concentration also has the advantage of giving better signal, while diluted samples result in rather weak signal. Therefore it is important to perform the experiments with three samples at different concentrations, to identify any concentration dependent behaviour and to improve the result by merging the valid data at different concentration obtaining an average intensity curve.

Data acquisition is made by using a graphical interphase designed for the BioSAXS beamLine X33 at EMBL Hamburg. The information received by this interface is generated by an automated pipeline that processes the data immediately after collection. The analysis actions that the pipeline perform go from executing an averaging of the two-dimensional images acquired by the detector to a one dimensional scattering curve that is afterwards analysed; subtracting the background scattering from the sample; determining the final key parameters as molecular mass, radius of gyration (R_g) and maximum diameter (D_{max}) (the two latter explained later in Section 3.3.1); and performing *ab initio* shape determination [126].

2.8 Atomic Force Microscopy Methods

The basic principles of the atomic force microscopy technique will be covered at the beginning of Chapter 5.

2.8.1 Sample Preparation

The protein subjected to atomic force microscopy experiments was the human desmoplakin plakin domain (PD). Cloning and purification of the protein was carried out by Dr. Martyn Chidgey. Briefly, the protein (residues 180-1022) was tagged at the N-terminal end with two cysteine residues by site directed mutagenesis of the coding

sequence. The 2×Cys tagged PD protein was expressed in *E.coli* using the pGex vector system (see vectors in Appendix A.1) and purified following the GST-fusion protein protocol, reaching a purity >95%. The purified protein was dissolved in 20 mM tris HCl, 100 mM KCl, pH 7.5 buffer, and it was stored at $-80\,^{\circ}$C to prevent degradation, and thawed for each AFM experiment.

Reaching the adequate concentration becomes critical since higher concentrations will result in several layers of protein covering the clusters completely, thereby giving no opportunity for single attachments events. Conversely, too low concentrations will result in not enough proteins bound to the clusters for AFM imaging. The stock solution concentration of plakin domain after purification was 2.2 μM (0.22 μg/μL) and it was diluted in buffer solution to obtain a concentration of 2.2×10^{-4} μg/μL, in order to work with protein low concentrations and promote a cluster/protein attachment of one to one.

The protein was left at room temperature for 2 h in order to reach thermal equilibrium and after that, 15 μL of the diluted protein was pipetted onto the graphite surface in the middle of the sample. The drop was left on incubation for 2 h, enclosed with a glass petri dish to minimise evaporation. Then the sample was carefully immersed in 1 M KCl with stirring for 30 min. In solution, potassium chloride dissociates into K^+ and Cl^- ions surrounded by water molecules, in this way KCl behaves as charge shielding and therefore as a modulator of electrostatic interactions between the proteins. Because of this property, it is used as a reagent for washing weakly bound molecules from the graphite surface [62]. The last step of sample preparation differs depending on the measurements to be done. In the case of dry measurements the sample was rinsed with copious distilled water and then dried with nitrogen gas for 3 min. For the *in situ* measurements the sample was also washed with distilled water and, without letting it dry, a 15 μL drop of buffer solution was placed on the region where the protein was deposited. The sample was then immediately taken to the AFM.

For experiments performed in mica the sample preparation protocol differs slightly since high concentration of PD must be used to ensure fixing of the protein to the surface. To this end, the concentration used was 0.22 μg/μL (original stock solution). Samples were rinsed with distilled water without washing with KCl. The volume of the drop remains the same, i.e. 15 μL, and it was also placed in the centre of the mica sample

(a square of 1 cm^2), although in this case the drop spreads on the surface, covering it almost completely due to the hydrophilic nature of mica.

2.8.2 Instrument Characteristics

The AFM equipment used for the dry measurements was a XE100 NSOM (*Park Systems*) (Fig. 2.4) with high resolution commercial non-contact cantilevers, coated with diamond (*Park Systems* 910M-SSS-NCHR) with a length of 125 μm, a mean width of 30 μm, a tip curvature radius of <10 nm, a resonant frequency between 260 and 410 kHz and a spring constant of 21-78 N/m. The AFM equipment for the liquid measurements was a XE100 (*Park Systems*), also with commercial silicon contact cantilevers (*Park Systems* 910M-NSC36) with similar dimensions as above, a full tip angle of 30 °, a resonant frequency between 105 and 230 kHz and a spring constant in the range 0.45 - 5 N/m with a typical value of 1.75 N/m. The regular scan rate and servo gain parameters for all images were around 1 Hz and 1 respectively. All the images were acquired using 512 pixels, processed with XEI software (*Park Systems*).

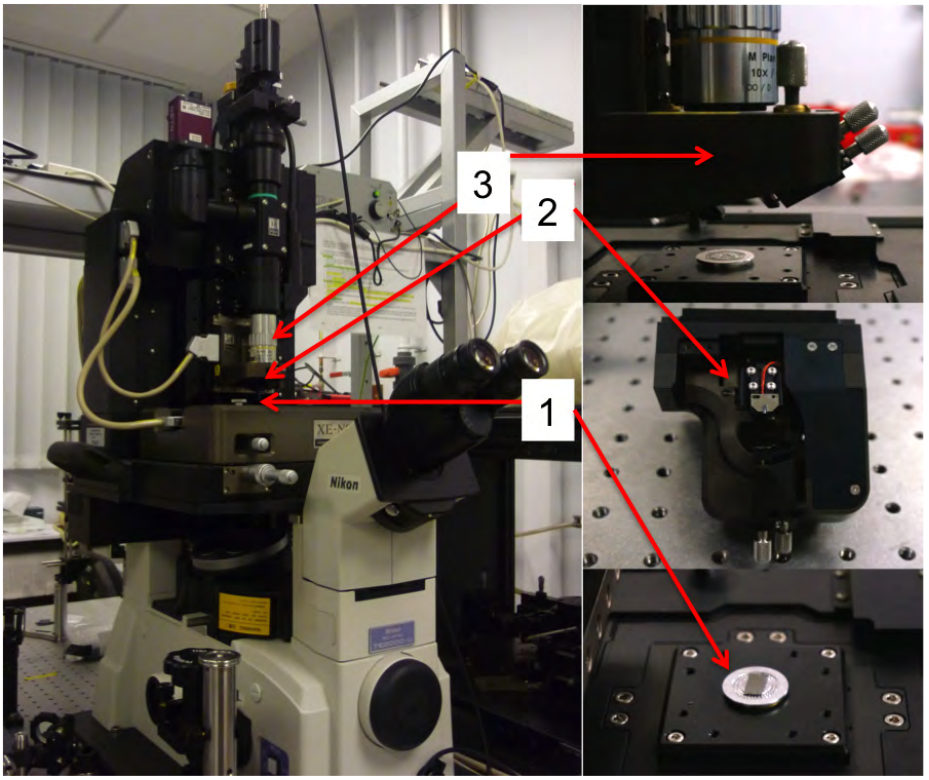


FIGURE 2.4: The AFM XE-100 NSOM (*Park Systems*) in the Nanoscale Physics Research Laboratory, University of Birmingham. The locations of (1) the sample, (2) cantilever and (3) probe head are indicated in the image.

The minimal experimental vertical error is determined by the resolution of the vertical scanner movement, normally <1 Å[127]. The lateral error is experimentally established for each individual tip by the calibration using a test structure (TGX11P *Ultrasharp*), which comprises a chessboard-like array of square pillars (Fig. 2.5).

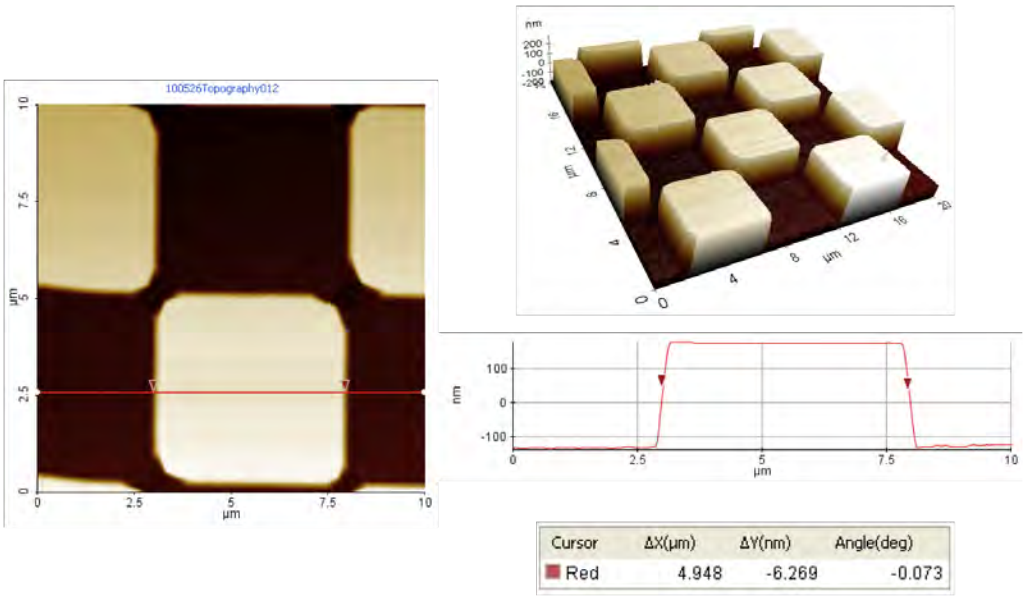


FIGURE 2.5: AFM topography of the calibration grating. The TGX11P test structure is an array of square pillars with sharp edges formed by the (110) crystallographic planes of silicon. The pitch uncertainty is within ±0.25%. Δx indicates the width of one of the columns as 4.948 μm when the theoretical value is 4.9 μm (TGX Series, TGX11P test structure, MikroMasch).

This calibration is based on the tip curvature radius at the height corresponding to the pillars height. Considering the conical shape of the tip, the actual lateral error is taken as the magnitude of the tip curvature radius at the end of the tip, i.e. 10 nm. Another limiting factor for lateral measurements is the number of pixels used per image. Since a number of 512 pixels was used for all topographies, the lateral resolution depends on the size of the topography, in this way, sub-nanometer resolution could be obtained only for scan sizes below 1 μm.

Apart from the tip radius intrinsic error in lateral measurements, other two uncertainties are also involved in the data analysis associated with both, lateral and vertical measurements. One of them is associated with the data handling methods of the software tools; and the other one is the statistical error, given by the standard deviation, which varies for every sample and it will be indicated in each case when required in chapters 5 and 6.

2.8.3 AFM Imaging

In order to acquire topographic information from a sample, the AFM tip is placed in close proximity to the sample surface. A laser bounces off the top of the cantilever onto a position-sensitive photodetector (PSPD) depending on the cantilever deflection as shown in Figure 2.6.

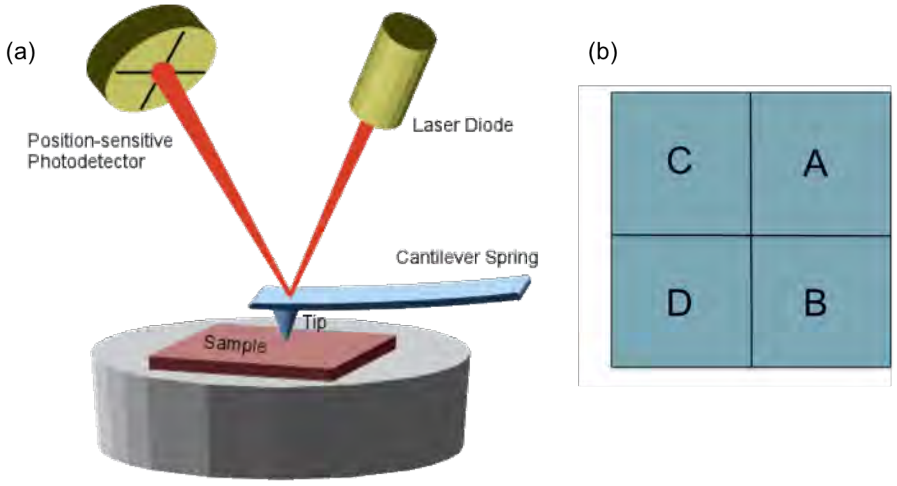


FIGURE 2.6: Representation of the AFM operation. (a) The laser is focussed on the back of the cantilever and the reflected laser is detected by a position-sensitive photodetector (PSPD). The angle of the reflected laser depends on the cantilever position, resulting in a beam-bounce detection system. (b) Quad-cell PSPD. AFM uses a photodetector divided in four quadrants, A through D to sense the deflection of the cantilever while it is scanning the sample surface. The signals from quadrants A and C are added, and from that value the sum of quadrants B and D is subtracted. Adapted from reference [128].

As the AFM tip scans the sample, topographical features on its surface result in changes of the position of the laser spot on the photodiode detector. The detector is part of a feedback loop employed to maintain a constant deflection (or constant force) between the tip and the sample by adjusting the voltage applied to the scanner. Depending on the mode on which the AFM is working, the feedback loop can also maintain a constant oscillation amplitude or frequency.

An AFM can be operated in three basic different modes, contact mode (CM), non-contact mode (NC) and tapping mode (TM), depending on the interaction between tip and sample (Fig. 2.7). In contact mode the scanner is moved vertically at each surface point *(X, Y)*, maintaining a constant setpoint deflection by the feedback loop between the cantilever and the sample. In this way the applied force is constant. Information acquired about the vertical distance at each data point generates a topographic data set.

This is then processed by a computer, which outputs an image of the sample surface. In the non-contact AFM mode the feedback loop maintains a constant frequency (the drive frequency) by vertically moving the scanner depending on the *(X, Y)* position until reaching a setpoint frequency. This mechanism keeps the average tip-sample distance constant as well, maintaining the tip close to the sample (tens of nanometers) during the scanning, with a small oscillation amplitude (tens of angstroms). In order to avoid contact between the tip and the surface, and considering the constant vibration that the cantilever is subjected to, the spring constant must therefore be higher than the spring constant found in a contact mode cantilever, normally varying from 20-75 N/m. The increased stiffness prevents the cantilever from being pulled into contact with the sample surface, but at the same time reduces the intensity of the signal acquired from it. In tapping mode the feedback loop is used to keep constant the RMS (root mean square, a statistical measure of the magnitude of any varying quantity) of the oscillation signal coming from photodiode detector. Therefore the oscillation amplitude is maintained at a constant setpoint value by the Z scanner and, as result, the tip-sample distance as well [129].

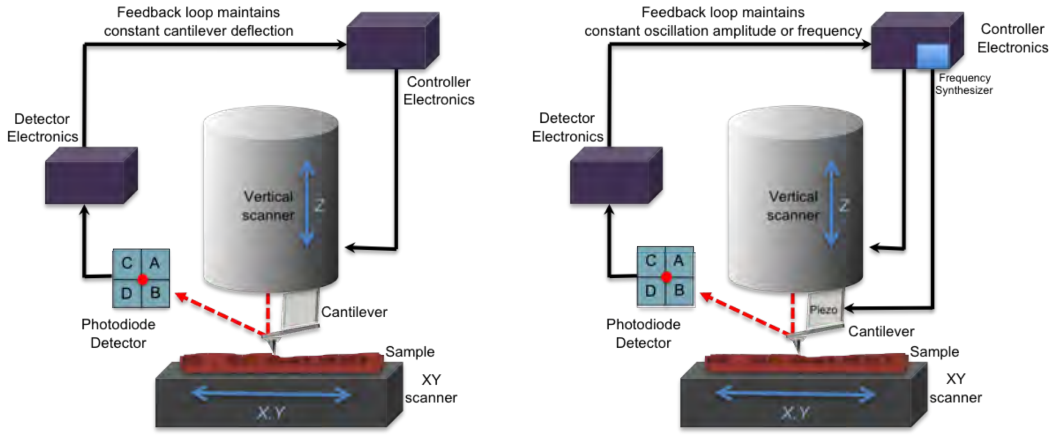


FIGURE 2.7: Diagram of AFM modes: (a) contact, (b) non-contact and tapping mode. In the first case the reflected laser goes from the cantilever base to the photodiode detector. The measurement of the deflection signal is acquired and then sent to the controller electronics to complete the feedback loop that indicates the retraction to the Z scanner. For the non-contact mode the feedback loop maintains constant either the amplitude or the frequency. The measurement of the oscillation amplitude or frequency is sent to the control electronics. The frequency synthesizer in the controller electronics sends the signal directly to the piezo on the cantilever. Finally, in the tapping mode, the detector electronics measures RMS of amplitude signal. The feedback loop maintains the oscillation amplitude constant. Adapted from [129].

The AFM imaging process also varies depending on the mode used. In contact mode AFM the parameter to tune is the deflection force of the cantilever, which is represented by the setpoint in the AFM feedback controls. The lowest force must be chosen while keeping the tip-sample distance balanced, watching not to over lower it causing the loss of stable engagement on the surface.

In non-contact mode, the drive frequency must be selected according to resonant frequency of the cantilever (slightly higher) corresponding to a setpoint amplitude of approximately 85% of the free resonant amplitude. By sweeping the frequency response curve of the cantilever the system can show the resonant frequency and identify it as a peak on a frequency *vs.* amplitude plot (Fig. 2.8). The amplitude associated to the voltage applied to the piezo system that drives the cantilever vibration is determined by the drive amplitude, which is normally chosen to be around 1 V. High amplitude values result in a stronger interaction, which produces a stronger signal, but with the risk of the tip touching the sample and suffer damage. Low amplitudes do not have this inconvenience, but result in a decrease of sensibility.

On the frequency *vs.* amplitude plot, the setpoint value is displayed as a line that represents the cantilever oscillation amplitude that is kept constant by the feedback loop whilst scanning. In non-contact mode, the setpoint ratio is proportional to the tip-sample distance. The effective resonant frequency decreases as the tip gets closer to the sample surface (more detail about this can be found in Section 5.1.1.2), which is represented by a shift of the resonant curve in the frequency *vs.* amplitude plot of Figure 2.8. Each resonant frequency curve intersects the drive frequency value at different amplitudes. This intersection point defines a setpoint value that is compared to the previous one, generating an error signal.

This process continues during the scanning. Every time there is a change in the height of the sample surface, making the tip-sample distance smaller, the resonance curve of the cantilever shifts, causing a change of the amplitude with respect to the drive amplitude, which should match with the setpoint value, generating then a difference of error signal. The system takes this signal as the feedback voltage necessary to retract or lower the Z scanner in order to increase or decrease the tip-sample distance again and shift the resonant curve back in its original value. The opposite happens when the change in the

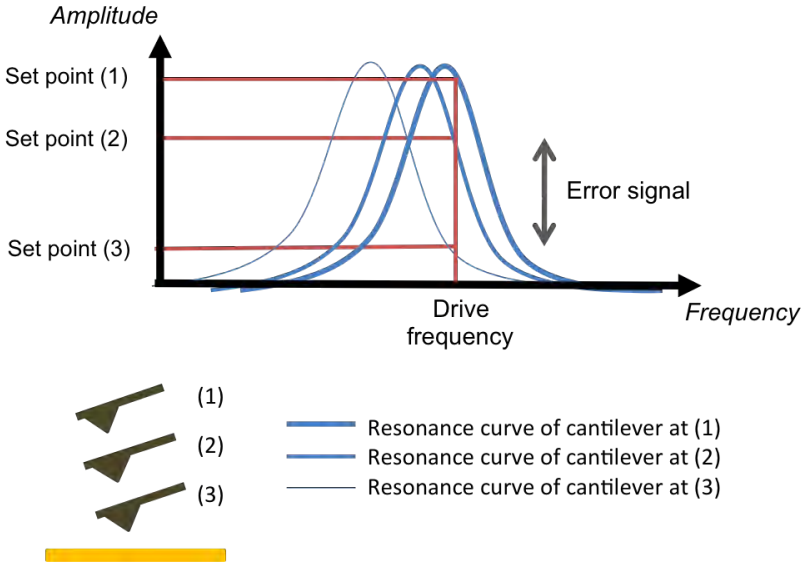


FIGURE 2.8: Resonance curves of cantilever depending on the distance to the sample surface. The corresponding setpoints indicate the error signal that allows the feedback loop system to establish a new Z distance. When the tip gets closer to the sample surface the resonance curve is shifted to the left in the plot. The new matching point between the chosen drive frequency and the shifted curve determines the new setpoint which gives the error signal by being compared with the previous setpoint. The chosen frequency for non-contact mode AFM operation must be slightly higher than the resonance frequency. This is in order to be able to detect different changes in the amplitude signal depending on the contraction or expansion of the tip with respect the sample surface. Adapted from [130].

height of the sample makes the tip-sample distance larger. This is why the error signal is proportional to the topography signal.

All the AFM topographic images presented in this thesis were acquired at room temperature, in non-contact mode for dry measurements and in both contact and tapping mode for measurements in liquid phase. In the case of liquid measurements, a drive frequency of 1/3 of the ambient drive frequency value is selected. Another particularity is that, since there is a droplet of buffer solution on top of the sample, the approach of the tip cannot be completed due to superficial tension forces exerted on the droplet surface. In order to avoid this situation and also to minimise the occurrence of entrapped air at the cantilever, a 15 μL droplet is placed on the cantilever (a special liquid probe head is used for every *in situ* measurement to protect the AFM electronics) so this droplet interacts firstly with the other droplet on the surface thereby forming one single liquid body, enabling the cantilever to approach the sample surface through the buffer solution.

2.8.4 AFM Artifacts

Since the dimensions of the features to image are very often of the same order of magnitude as the dimensions of the AFM tip, the images are actually a convolution of the shape of the features and the probe geometry. When the features are much smaller than the tip, artifact problems arise.

The presence of a feature is detected due to a change in height of the sample surface, causing a difference in the oscillation amplitude of the cantilever and inducing a mismatching of the amplitude setpoint that in turn generates an error signal. The response of the system is not immediate at the beginning or at the end of the feature, which results in a broadening of the feature in the topography. The same effect happens for holes but in the opposite way: concave features are imaged with a narrower diameter than they really are as a consequence (Fig. 2.9).

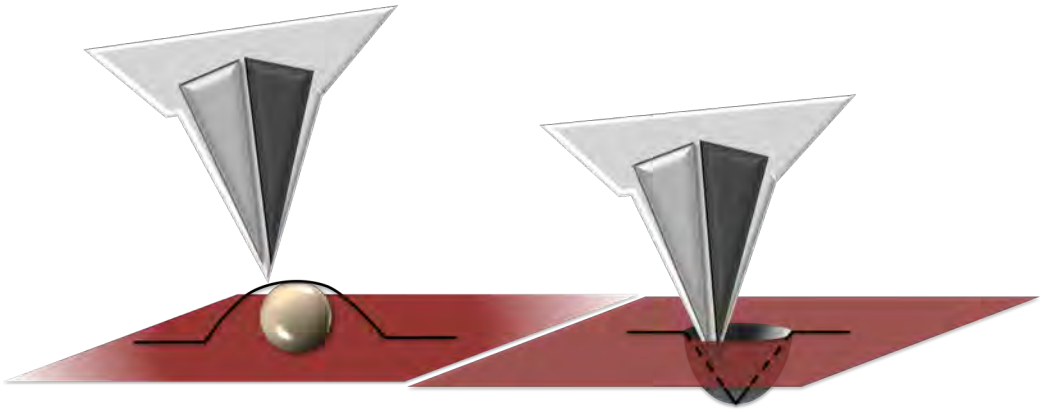


FIGURE 2.9: Common AFM artifact caused by angles difficult to reach for the probe, causing broadening for features and narrowing for depressions. The sidewall angle of the AFM tip determines the highest level of steepness that a surface profile side can have in order to be accurately imaged.

It can be noted that this kind of artifact is mainly due to acute angles existing between the probe and the sample. Even if the tip is sharp enough, the scanning is not capable of reaching angular regions. This is the reason why, with AFM, lateral measurements are less reliable than height measurements, which provide more accurate information about the topography of the sample. Possible artifacts in AFM height measurements are due to deformation of the sample features by the action of the tip, especially for soft samples.

2.9 Production of Size-Selected Clusters

2.9.1 Cluster Source

Size-selected nanoclusters consisting of 55, 147, 309, 561 and 923 gold atoms were used during this experiment. These nanoparticles were generated in the cluster source (G12) at the Nanoscale Physics Research Laboratory, University of Birmingham (Fig. 2.10), which works with a RF magnetron plasma sputtering system and a gas condensation process [131] in combination with a time-of-flight (TOF) mass filter [132].

The generation of size-selected clusters consists of three stages: cluster production, cluster beam focusing and acceleration, and mass selection. Each one of them is associated with a section of the clusters source.

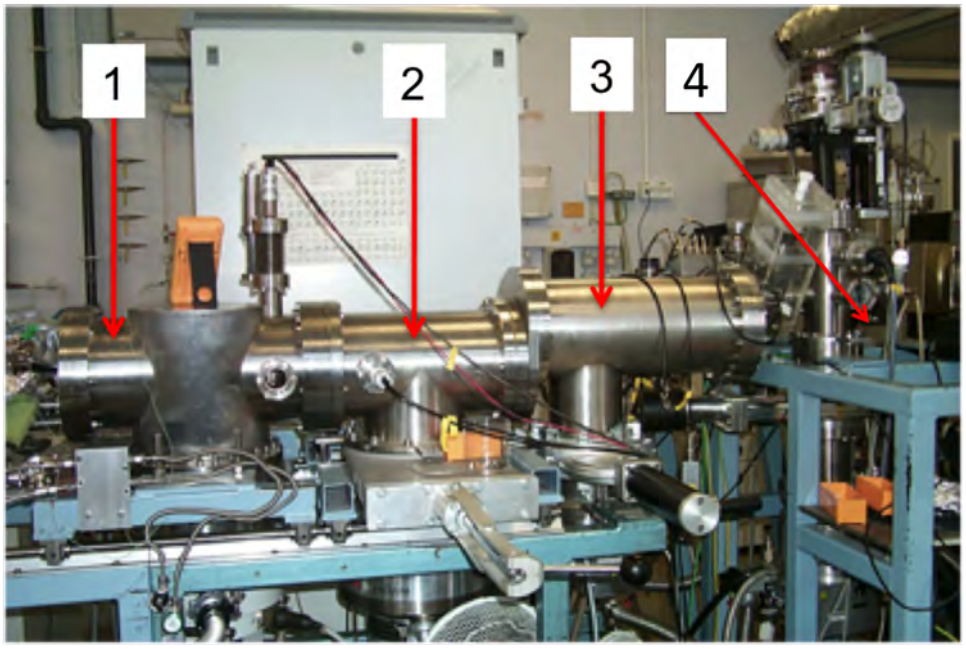


FIGURE 2.10: The G12 cluster source in the NPRL Laboratory, University of Birmingham. The sections corresponding to the sputtering and cluster formation (1), beam focusing and acceleration (2), mass selection (3) and deposition (4) are indicated in the image. This cluster source was used to produce the size-selected gold clusters for graphite samples decoration (by self-pinning, implantation or soft-landing), as well as the argon ion sputtering treatment.

In the first section there is the condensation chamber, in which by applying a radio frequency high voltage, Ar plasma is ignited near to a Au target. As a consequence, the target is negatively charged, and develops a negative self-bias voltage inducing an electric field that causes sputtering of hot Au atoms due to the bombardment of Ar

ions. The atoms are consequently condensated in Au clusters with a broad distribution of sizes (2-10^5 atoms each) by the action of helium gas injected with a typical pressure of 0.2 mbar. In the second section there is a set of einzel lenses with a negative potential that focus the ionized beam. In the middle of the einzel lenses there is a set of deflection plates that adjust its trajectory into the section three by the action of an electric field transverse to its direction that disperses the beam in the horizontal plane according to the mass of the clusters. When the beam enters the third section, it is subjected to an acceleration pulse caused by the momentary raising potential of the bottom plate (bottom region), which is connected to a high frequency voltage switch. With this acceleration all the ions gain the same momentum (we assume all clusters are equally charged with the charge element 1.602×10^{-19} C) and therefore all ions with the same mass will have the same velocity shift directed upwards to the middle region. When all the ions of the same mass are in the middle region they are prevented from going further upwards by the action of a second pulse with the same amplitude, this time coming from the upper plate, which also is connected to a high frequency voltage switch. In this way, the ions with different masses are transmitted in function of the time difference between pulses: the time that it takes the ions to travel from the first to the third region varies, so the second pulse can be applied with a big enough time lapse after the first one so as to position the mass selected beam (among the parallel beams with specific masses created) at the required height of the aperture at the end of the chamber. This principle makes the mass of the transmitted ions independent of the ion beam energy.

The mass resolution is determined by the ratio of the exit aperture width and the lateral displacement and in our case is $m/\Delta m = 25$ and the transmission is around 60%.

At the end of this section a size-selected clusters beam reaches an einzel lens that focuses it into a high vacuum chamber with a pressure of 10^{-8} mbar for cluster deposition (Fig. 2.11).

2.9.2 Deposition Techniques

The purpose of the gold clusters on graphite is to immobilise proteins creating a surface decorated with isolated single molecules attached to it. It is logical then to want the clusters to be immobilised as well. With the so-called soft landing technique, the clusters tend to diffuse on the surface and tend to agglomerate on surface defects, making them

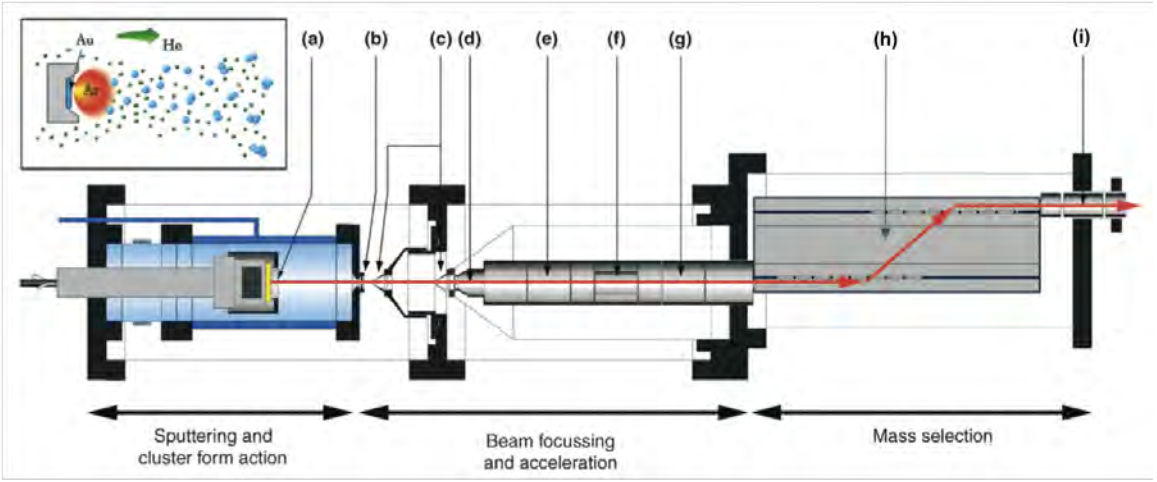


FIGURE 2.11: Diagram of a lateral time of flight size selection clusters source. (a)Magnetron gun with the gold target, (b) adjustable-diameter nozzle, (c) electrostatic skimmers, (d) extraction high voltage lens, (e) first einzel lens, (f) deflection plates, (g) second einzel lens, (h) TOF mass filter acceleration region, (i) third einzel lens. From reference [58]

useless for the objectives of this work. Therefore, a different deposition technique must be implemented.

The deposition energy determines the interaction between the deposited clusters and the substrate in metal-covalent systems, which is the case of gold clusters on graphite. Depending on the effect desired, the impact kinetic energy E is controlled by a bias voltage applied to the sample stage and can be tuned from high to low in order to implant ($E > 50$ eV/atom), pin (10 eV/atom $< E <$ 50 eV/atom), or soft-land ($E < 10$ eV/atom) the clusters on the substrate surface [58]. The soft landing technique preserves the morphology of the clusters, but as it was mentioned, it is not useful in our application. On the other extreme, the implantation technique keeps the clusters immobilised but buried in the surface layer, so they are deformed [133] and not available for binding the proteins. The pinning method is in the middle, suppressing the diffusion but not disabling the clusters' binding properties.

Pinning of clusters on a surface is possible via a defect created on the substrate as a consequence of the cluster impact. In this case, the defect is the displacement of a surface carbon atom from the graphite. The pinning threshold energy therefore is the energy that makes a carbon atom change its equilibrium configuration on a site in the regular array of the graphite as a result of the collision. The pinning energy $E_{pinning}$ is a linear function of the cluster size [133] and it is associated with the clusters' kinetic energy

E_k necessary to remove a carbon atom. Given the principle of conservation of energy in elastic collisions between a massive body (m_1) and a light body (m_2) the transferred energy E_T is given by:

$$E_T = \frac{4E_k m_1 m_2}{(m_1 + m_2)^2} \tag{2.1}$$

Considering the cluster as being the massive body, we have that $m_1 = NM_{Au}$, where N is the number of atoms in the cluster, and considering the carbon atom as the light body, we have $m_2 = \mathrm{M}_C$. In the limit where $m_1 >> m_2$, rearranging gives:

$$E_{pinning} = E_k = \frac{NM_{Au}E_T}{4M_C} \tag{2.2}$$

This formula was tested experimentally and with computational aids [55], showing the linear dependence and concluding that the binding of the cluster in the point defect is stronger than the interaction between the cluster and the surface. This allows the creation of a monodispersed array of stable and available gold clusters on a graphite support.

For experiments described in Chapter 5, the Au clusters are deposited on highly orientated pyrolytic graphite (HOPG) by the cluster pinning method using a beam energy of 23 eV per atom with a current of 150 pA in both Au_{55} and of Au_{147} cases. For the experiments in Chapter 6 the deposition energy is specified within that chapter in every case. The graphite samples were freshly cleaved shortly before their introduction in the vacuum. The density of clusters is controlled by the beam current and the deposition time (80 s). This is because the clusters are considered as single charged particles, so, by knowing the total charge Q given by:

$$Q = It \tag{2.3}$$

we can calculate the number of clusters deposited per sample as:

$$\text{Number of clusters} = \frac{Q}{q} \tag{2.4}$$

where q is the elementary charge with a value of 1.602×10^{-19} C.

Graphite was chosen as surface support for the gold clusters because of two reasons. The first one is its flat nature that permits SPM scanning, the second one is the weak interaction between this substrate and the protein. This ensures that, with the appropriate concentration, most of the molecules are removed for lateral diffusion after having been washed out and only a few proteins remain on the graphite surface defects. In this way it is possible to attribute the attached proteins as being only onto the gold clusters. Additionally, the graphite surface is chemically inert in air and water.

Chapter 3

Desmoplakin Linker Domain Structure Determination by NMR Spectroscopy and SAXS

As the structure of the desmoplakin linker domain was determined by NMR, a brief introduction of this technique is given at the beginning of this chapter, including the theoretical principles and the basics of backbone and side-chain assignment. This is followed by the experimental results obtained from the biophysical characterisation of the desmoplakin linker domain, leading to the determination of its tertiary structure and corresponding analysis. This chapter also includes the experimental results obtained from SAXS experiments on the desmoplakin linker domain and the desmoplakin plakin repeat domains B and C construct, followed by an analysis and discussion.

3.1 Nuclear Magnetic Resonance Spectroscopy

3.1.1 Physical Principles

In the presence of an external magnetic field B_z, with strength B_0, the degeneracy of nuclear spin states in a system is lifted, a phenomenon known as Zeeman effect. A nuclear spin will experience an interaction which energy is represented with the Hamiltonian:

$$\hat{H}_1 = -\gamma B_0 \hat{I}_z \tag{3.1}$$

where I_z is the z component of the nuclear spin angular momentum operator and γ is the gyromagnetic ratio, a property of the nucleus. Being the eigenvalue equation for I_z with the nucleus wave eigenfunction Ψ_m:

$$\hat{I}_z \Psi_m = m\hbar\Psi_m \tag{3.2}$$

the corresponding eigenvalue equation for the hamiltonian gives us the energy levels of a single spin in a magnetic field:

$$\hat{H}_1 \Psi_m = -m\hbar\gamma B_0 \Psi_m \tag{3.3}$$

where m takes values between -I and I in integer steps. For fermions, $I = \frac{1}{2}$, therefore $m = \pm\frac{1}{2}$. Thus there are two possible energies for the system

$$\mathrm{E}_\alpha = -\tfrac{1}{2}\hbar\gamma\mathrm{B}_0 \text{ and } \mathrm{E}_\beta = \tfrac{1}{2}\hbar\gamma\mathrm{B}_0$$

Therefore, knowing that $E = h\nu$, the frequency ν of the photon with energy equivalent to the difference of energy between the states α and β $\Delta\mathrm{E}_{\alpha\rightarrow\beta}$ is:

$$\nu_{\alpha\rightarrow\beta} = \gamma B_0/2\pi \tag{3.4}$$

which is defined as *minus* the Larmor frequency ν_0 (if the gyromagnetic ratio of nuclei is positive)[134] (Fig. 3.1 (a)). In terms of rad/s the Larmor frequency is $\omega_0 = 2\pi\nu_0$, which can be also interpreted as the precession frequency of the nuclei magnetisation vector μ about the external magnetic field axis [134] (Fig. 3.1 (b)), determined by -$\nu_{\alpha\rightarrow\beta}$.

Given that the magnitudes of B_0 in nuclear magnetic resonance spectrometers are between 1 T (equivalent to a ΔE with frequency of 42.6 MHz) and 22.3 T (equivalent to a ΔE with frequency of 950 MHz), and that the gyromagnetic ratios $\gamma/2\pi$ go from 4.314 MHz/T (for ^{15}N) to 42.58 MHz/T (for ^{1}H), the Larmor frequency is in the radio frequency range of the electromagnetic spectrum, with wavelengths that go from 30 m

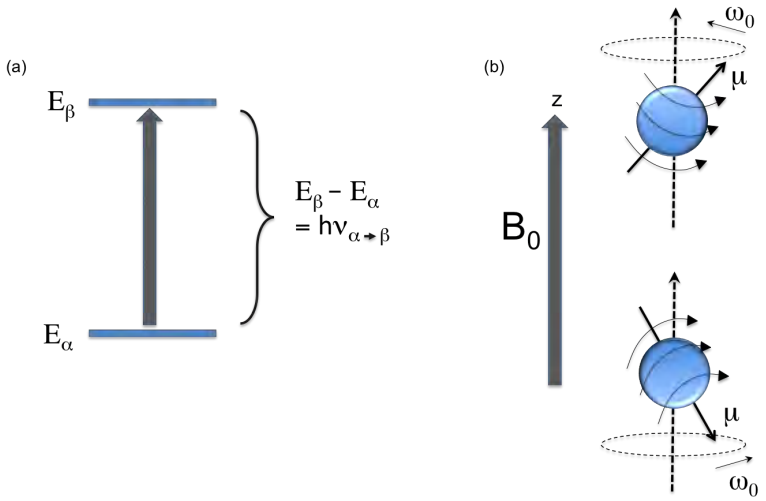


FIGURE 3.1: The Larmor frequency interpreted quantum mechanically and classically. (a) The Larmor frequency of the spin can be defined as the frequency associated to the quantum energy of transition $\Delta E_{\alpha \to \beta}$ between the two possible spin states for $I = \frac{1}{2}$. (b) The Larmor frequency $\omega_0 = 2\pi\nu_0$ can also be visualised as the precession frequency of the nuclei magnetisation vector about the external magnetic field axis with direction z and magnitude B_0.

down to 40 cm. As a consequence, by irradiating with RF pulses in magnetic resonance with the nuclei spins, it is possible to manipulate the changes from one state to another. Initial populations of the energy levels are unequal, with the state of lower energy being slightly more populated, following a Boltzmann distribution. The RF pulse excites the nuclei with lower energy (α state) to the state of higher energy (β state) (Fig. 3.1 (a)). During relaxation nuclei return to thermal equilibrium, causing a RF signal to be emitted with energy equal to the difference in energy between states α and β, which is detected.

Isotopes with an odd number of protons and/or of neutrons have a nonzero overall spin, thus the only nuclei sensible to nuclear magnetic resonance. The most common used nuclei are ^{1}H, ^{15}N and ^{13}C, with natural abundances of 99.985%, 0.37% and 1.108% respectively [135].

3.1.1.1 Chemical shift

The presence of rotating electrons in the molecule causes a shielding effect from the external B_z by also creating a counteracting magnetic field B_e. Therefore, the effective magnetic field B_{eff} has a lower magnitude than B_0. Since electron density distribution around each nucleus in a molecule is different depending on the neighbouring nuclei,

B_{eff} will vary depending on the chemical environment, and as a consequence, there will be a shift in the NMR frequency ν, called the chemical shift $\Delta\nu$, that determines particular NMR signals for each nuclei at each specific chemical environment. Higher electron densities induce a higher shielding effect in which case the NMR frequency is shifted upfield, corresponding to a lower chemical shift. In the opposite case we have a NMR frequency downfield shifted, corresponding to a higher chemical shift [136].

3.1.1.2 Spin-spin coupling

Two bonded nuclei experience a coupling effect due to an existing mutual disturbance of the valence electrons in the molecule. From Equations 3.1 and 3.4 it is possible to write the Hamiltonian for the weak interaction of two spins, considering the scalar coupling term between them J_{12} (in terms of Hz by dividing everything by 2π) :

$$\hat{H}_2 = \nu_1 \hat{I}_{z1} + \nu_2 \hat{I}_{z2} + \hat{J}_{12} \hat{I}_{z1} \hat{I}_{z2} \tag{3.5}$$

The coupling term splits the energy levels in the 2-spins system and therefore the resonances of the two doublets of the NMR spectrum corresponding to two coupled spins centred at the Larmor frequencies of spins one and two (ν_1 and ν_2 respectively), with a separation defined by the coupling constant J_{12} (Fig. 3.2). This constant is independent of the magnetic field and is measured in Hz, ranging between 115 and 250 Hz (usually 125 to 160 Hz) for ^{13}C-^{1}H, 90 to 100 Hz for ^{15}N-^{1}H, and 0-20 Hz for ^{1}H-^{1}H J constants [136]. In the same way, the NMR signal can show triplets quartets or quintets depending on the number of connected nuclei and the number of intervening covalent bonds. The splitting pattern of the signal and the relative intensities can be predicted by a binomial expansion described by the Pascal's triangle [137].

Since spin-spin coupling contains information about distances and angles between nuclei in a molecule, magnetisation transfer between nuclei through covalent bonds becomes a powerful tool for ascertaining molecular structure.

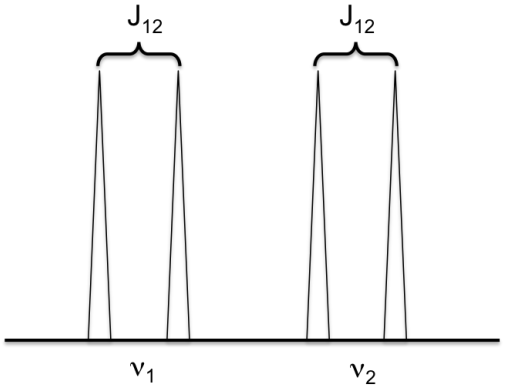


FIGURE 3.2: J-coupling or spin-spin coupling. In a two spin system the nuclei experience a coupling effect that is reflected in the resonances of the two doublets in the NMR spectrum. The doublets are centred at the Larmor frequencies of spins 1 and 2 (ν_1 and ν_2), with a separation defined by the coupling constant J_{12}.

3.1.2 Backbone Assignment

3.1.2.1 Heteronuclear single quantum coherence

In order to determine the structure of a protein, the first stage to accomplish is residue identification. The first and basic spectroscopy method to carry out to characterise a protein is the heteronuclear single quantum coherence (HSQC) experiment [138].

For this experiment the magnetisation is transferred from ^{1}H to a heteronucleus in this case is ^{15}N, (but it could also be ^{13}C), so single-labelled (enriched with ^{15}N) protein must be produced. The HSQC spectrum of a protein is a fingerprint map of the residues that contain a proton coupled to a ^{15}N nucleus (amide groups) in the peptide bond, which is all, except proline that contains an imino group. Each residue of the protein corresponds to an observable peak in the HSQC spectrum that is located at specific chemical shifts in each dimension. In general, a well-dispersed pattern of peaks in a HSQC spectrum is an indicator of a folded protein in solution, but moreover, HSQCs also give immediate qualitative information about the protein secondary structure: amino acids belonging to alpha helical structures experience an average upfield shift of 0.39 ppm with respect the random coil value, while beta strands tend to move downfield an average of 0.37ppm [139]. At the same time, an agglomeration of peaks in the central region of the spectrum is a sign of significant unstructured elements in the molecule [140] that form the random coil of the protein.

3.1.2.2 Triple resonance NMR experiments for sequential assignment

A HSQC spectrum is not enough to assign a protein since the two dimensions that constitute it do not reveal sufficient information about the individual amino acids in the molecule. For this, a third dimension is needed. Triple resonance NMR experiments use the coupling of three atomic nuclei: ^{1}H, ^{15}N and ^{13}C, therefore double-labelled protein must be produced to perform these experiments by enriching the growing medium with ^{13}C and ^{15}N. A double-labelled protein constitutes basically a continuous spin system from which the backbone can be seen as a long series of scalar coupled nuclei [141].

With ^{13}C^{15}N-labelled proteins, sequential assignment using heteronuclear coupling can be obtained by combining the information of different triple resonance experiments that achieve magnetisation transfer via diverse pathways through well-resolved one-bond J-couplings from nucleus to nucleus. Six different triple resonance experiments are considered basic for unambiguous backbone assignment [142–145], a combination of which could normally be HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB and HN(CO)CACB. The experiments are related in couples of intra-residue (i) and inter-residue (i *-1*) magnetisation transfer having partial information that complement each other in order to identify a sequence of residues. In practice, complementary spectra are overlapped to distinguish between peaks corresponding to i or i *-1* residues. If peaks occur in both complementary spectra it means they belong to the i *-1* residue, and if they occur only in one of them then they belong to the i residue. By localising determined succession in the amino acid sequence of the protein it is possible to assign residue by residue following a chain of associations. To initialise the assignment the distinctive chemical shifts of Gly, Ala, Ser and Thr are used.

Details of magnetisation transfer of these experiments and the ones mentioned in following sections can be found in Appendix B.2.

3.1.3 Side Chain Assignment

Once the backbone assignment has been completed, the side chain of the protein must also be assigned. For this, a series of experiments not restricted to directly coupled spins in neighbouring nuclei should be implemented. This type of experiments includes

through-bond experiments, as the ones used for backbone assignment, but also *through-space* experiments.

3.1.3.1 Through-bond NMR experiments: Total correlation spectroscopy

Total correlation spectroscopy (TOCSY) is a method for correlating all protons within a given spin system. Even distant protons can be correlated as long as there are couplings between every intervening proton in the system. TOCSY is classified as a homonuclear through-bond correlation method since the magnetisation transfer occurs between same type of nuclei connected by one or few bonds through J-coupling.

In order to pass from the backbone assigned residues to a side chain assignment different experiments need to be acquired from the same sample of protein (or another sample of the same protein with higher concentration). The arrangement of experiments varies depending on the specific needs of the situation. One possible and effective arrangement consists of two TOCSY experiments: H(C)CH-TOCSY and CTOCSY(CO)NH, and another through-bond experiment: HA(CACO)NH.

H(C)CH-TOCSY is the experiment that gives the majority of the side chain information. A spectrum is obtained for each side chain carbon frequency that shows a vertical strip of all the side chain protons resonances (Fig. 3.3). The resonance peaks can be identified by knowing the characteristic chemical shifts of each type of proton for each amino acid with help of the Biological Magnetic Resonance Data Bank (BMRB) [146].

By synchronising the HA(CACO)NH experiment that contains the chemical shift values for each backbone proton and the associated $^{13}C_{\alpha}$ and $^{13}C_{\beta}$, with the H(C)CH-TOCSY experiment, assignments on the former lead to each corresponding carbon plane on the latter experiment (Fig. 3.3). In order to recognise the respective planes for $^{13}C_{\gamma}$, $^{13}C_{\delta}$ and $^{13}C_{\epsilon}$ we use CTOCSY(CO)NH.

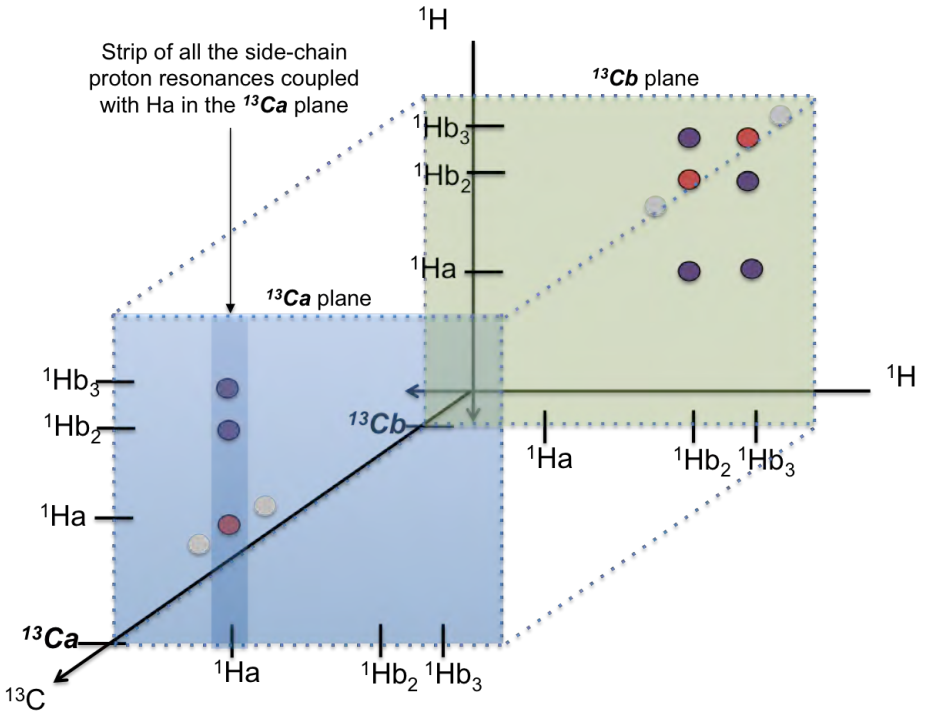


FIGURE 3.3: Schematic representation of two planes of an HCCH spectrum. The blue plane corresponds to a C_α (or Ca) plane while the green plane corresponds to a C_β (or Cb) plane (for the case of a residue with two Hb atoms). Peaks in red are peaks that belong to a diagonal, i.e. Ha peaks in the Ca plane and Hb peaks in the Cb plane. Once the Ha-Ha peak has been identified in the Ca plane, identification of Hb peaks in the Cb plane becomes easy since their chemical shift values will belong to the vertical strip indicated by the Ha-Ha peak in the Ca plane.

3.1.3.2 Through-space NMR experiments: The nuclear Overhauser effect

Resolving the structure of a protein requires information about the correlations of protons not only connected by bonds, which is the case of the scalar couplings. Determining which nuclei are close in space is used to achieve a solution structure of the protein.

Dipole-dipole interaction between the nuclei spins is the phenomenon that provides this information: two nuclei in close proximity are influenced by the external magnetic field but also by the magnetic dipole moment of the other nuclear spin. Reversion of the direction of the magnetic dipole of one of the nuclear spins (decoupling), or annulation of it (saturation by population equalisation), causes a through-space magnetisation transfer to the other nuclear spin. The corresponding rate of transmission is known as cross relaxation rate. This mechanism of relaxation depends on the distance between interacting nuclei r according to r^{-6} [141].

A change in intensity of one NMR resonance as result of spin state transition by cross relaxation is known as the Nuclear Overhauser Effect (NOE) [147]. Given the distance dependence factor, NOEs are used to generate distance restraints for three-dimensional structure determination of proteins, these being derived from the dipole-dipole interactions between protons within the molecule. The experiments related to this effect are the NOESY (Nuclear Overhauser Effect Spectroscopy) experiments, that determine the NOE peaks corresponding to protons close in space with a maximum inter-nuclei distance of 5 Å[148].

3.2 Structure Determination of Desmoplakin Linker Domain by NMR Spectroscopy

3.2.1 Pre-characterisation with Biophysical Analysis Techniques

The desmoplakin linker domain was purified following the GST fusion protein protocol described in Section 2.3.2. The protein contained 5 extra residues (Gly, Pro, Leu, Gly, Ser) at the beginning of the N-terminal end for the PreScission protease cleavable GST tag and the BamH1 restriction site.

3.2.1.1 Determining purity by SDS-PAGE

The purity level of purified desmoplakin linker domain (DPlink) was assessed by SDS-PAGE. A band corresponding to one single species in the correct range of molecular weight between 10 kDa and 15 kDa can be observed in Figure 3.4. This range corresponds to the molecular weight predicted for the protein of 12.9 kDa. High purity ($> 95\%$) is demonstrated as no other bands are seen in isolated fractions.

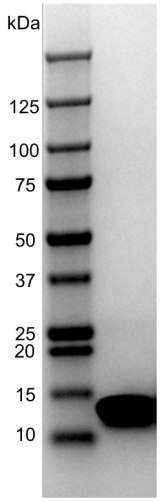


FIGURE 3.4: Purity of the desmoplakin linker domain (right lane) after GST cleavage and ion exchange purification tested by SDS PAGE with molecular weight markers (left lane). The band, that corresponds to one single species, is in the correct range of molecular weight between 10-15 kDa for the DPlink protein predicted to be 12.9 kDa. The purity level reached was > 95%.

3.2.1.2 Determining monomeric state of the desmoplakin linker domain

Sedimentation velocity AUC was performed to assess the multimeric state of the desmoplakin linker domain in solution. The result of the experiment showed a single monomeric species with peak at 12.34 kDa (Fig. 3.5), in agreement with the sequenced derived molecular weight of the protein.

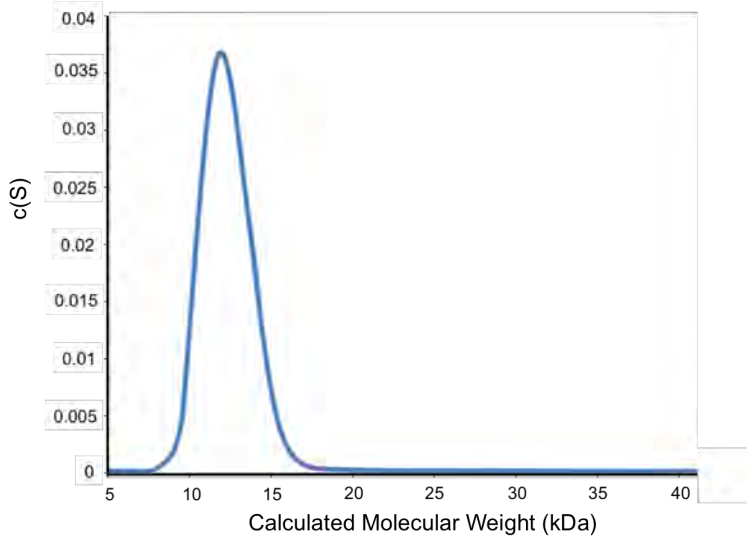
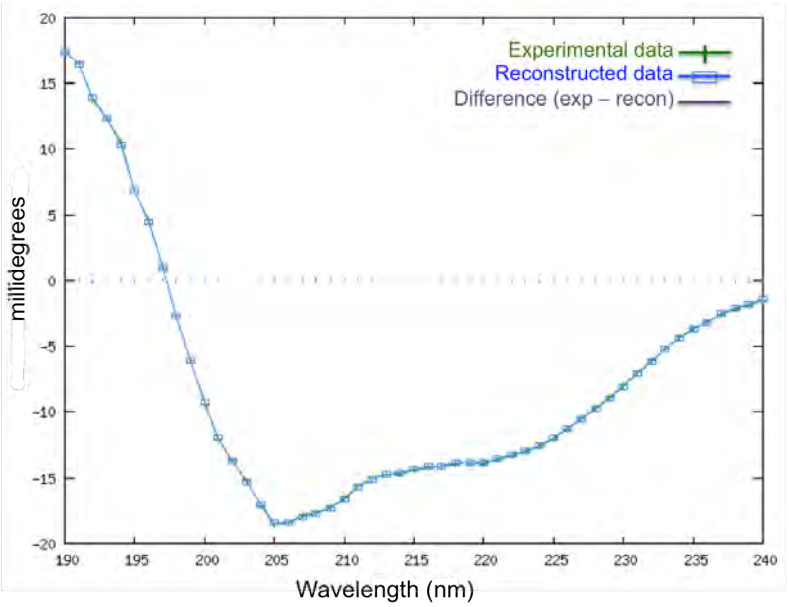


FIGURE 3.5: Analytical ultracentrifugation (AUC) results for desmoplakin linker domain analysed with SEDFIT programme. The desmoplakin linker domain at 5 μM was analysed by AUC in 20 mM HEPES (pH 7.5) and 100 mM NaCl, at 20 °C, with a wavelength of 280 nm. The results indicate the presence of only one species in a monomeric state with a molecular weight of 12.34 kDa.

3.2.1.3 Folded state and secondary structure prediction

Circular dichroism analysis of the sample showed a folded state and also predicted relative proportions of secondary structure of the protein according to the different population elements in the sample.

Results obtained from CD experiments of desmoplakin linker domain revealed a positive signal around the wavelength of 190 nm and a negative signal in the range of 200-240 nm, meaning an absorption of right and left circularly polarised light respectively. Wavelength minima at 206 nm and 220 nm coincide with the established minima that characterise a highly alpha helical structure (208 and 222 nm) [149]. The negative band between 210 nm and 220nm and the positive (but with low values) band between 195 nm and 200 nm indicate the presence of small amount beta-sheet. Finally, the negative band around 200 nm is a sign of random coil in the protein.



Calculated Secondary Structure Fractions

Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
0.29	0.19	0.07	0.05	0.14	0.28	1.02

FIGURE 3.6: Circular dichroism (CD) analysis results for the desmoplakin linker domain. Experimental data (green) is in high agreement with reconstructed data (blue) based on the data base used by the analysis programme, denoting reliable results. The table shows the percentages corresponding to each possible secondary structure, demonstrating a predomination of alpha helices constituting the protein. The analysis was made using the DichroWeb server [119].

Identification of overall secondary structure has been achieved by spectral deconvolution (Fig. 3.6), demonstrating quantitatively that the protein is predominantly alpha helical (48%) while beta strands only represent a 12% and turns in the protein are predicted to be 14% of the structure. The analysis of CD results also indicates that there is a high presence of disordered structure in the protein.

3.2.2 Residue Identification using Backbone NMR Experiments

In order to elucidate the structure of the desmoplakin linker domain by NMR it is necessary first of all to obtain a high resolution HSQC spectrum. HSQC was acquired on the DPlink at 200 μM using an Agilent 600 MHz NMR spectrometer (Fig. 3.7).

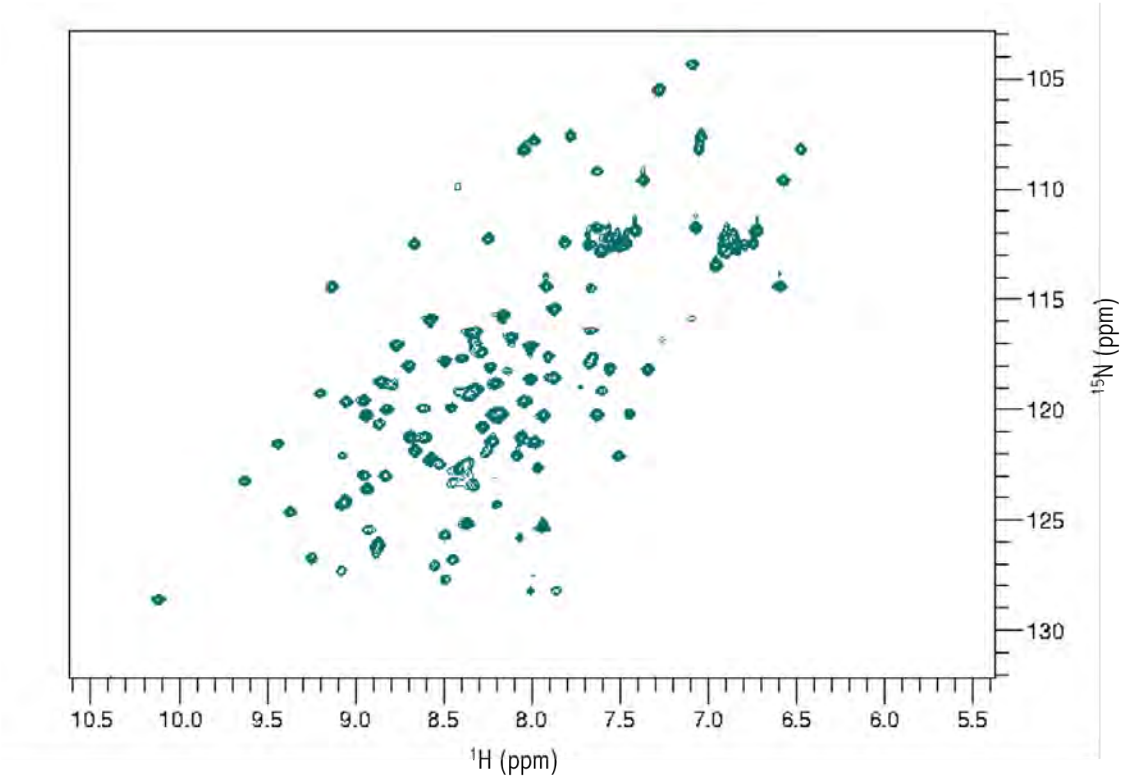


FIGURE 3.7: HSQC spectrum of desmoplakin linker at 200 μM in 20 mM HEPES buffer pH 7.5, acquired with an Agilent 600 MHz NMR spectrometer at 25 °C. The spectrum shows a well dispersed distribution of 98% of the expected peaks, indicative of a folded protein. The side-chain resonances of glutamine and asparagine residues, are at relatively low chemical shifts in both dimensions (right top corner of the spectrum), whereas the tryptophan yields a peak at relative high values of chemical shifts in ^{1}H and ^{15}N (bottom left side of the spectrum). The backbone peaks of glycine residues are shifted upfield in ^{15}N. Data processing was done using NMRPipe software.

The peak distribution of the spectrum is well dispersed between 6.4 ppm and 10.2 ppm in the direct dimension, confirming a folded state of the DPlink protein. The number of peaks expected follows the equation:

$$Peaks_{HSQC} = Aminoacids_{Total} - Pro + Trp + 2(Asp + Gln) \tag{3.6}$$

given that Pro does not present peak in the HSQC spectrum and on the other hand that Trp, Asp and Gln have amide groups in the side chain with 1, 2 and 2 protons respectively (the latter two resulting in two peaks in the spectrum at the same N frequency). Therefore 138 peaks are expected (117 residues in total, 2 Pro, 1 Trp, 3 Asp and 8 Gln), from which 135 are visible in the spectrum, giving a 97.8% of the expected peaks. The peak distribution of the spectrum suggests a structure containing a combination of alpha helical and beta strands.

Once the quality of the HSQC spectrum was assessed, a set of six triple resonance NMR experiments (listed in Section 3.1.2.2) was acquired on the same sample to complete sequential backbone assignment of the DPlink. A total of 112 residues were assigned following the proceeding previously described, representing the 95% of the amino acids of DPlink (Fig. 3.8).

Identification of the remaining 5 residues was not possible due to lack of signal in some of the experiments. This problem is thought to be consequence of a state of flexibility inside the protein, especially in its N-terminal end, where most of the problematic assignments were found.

Data analysis was performed using CCPN Analysis software [150].

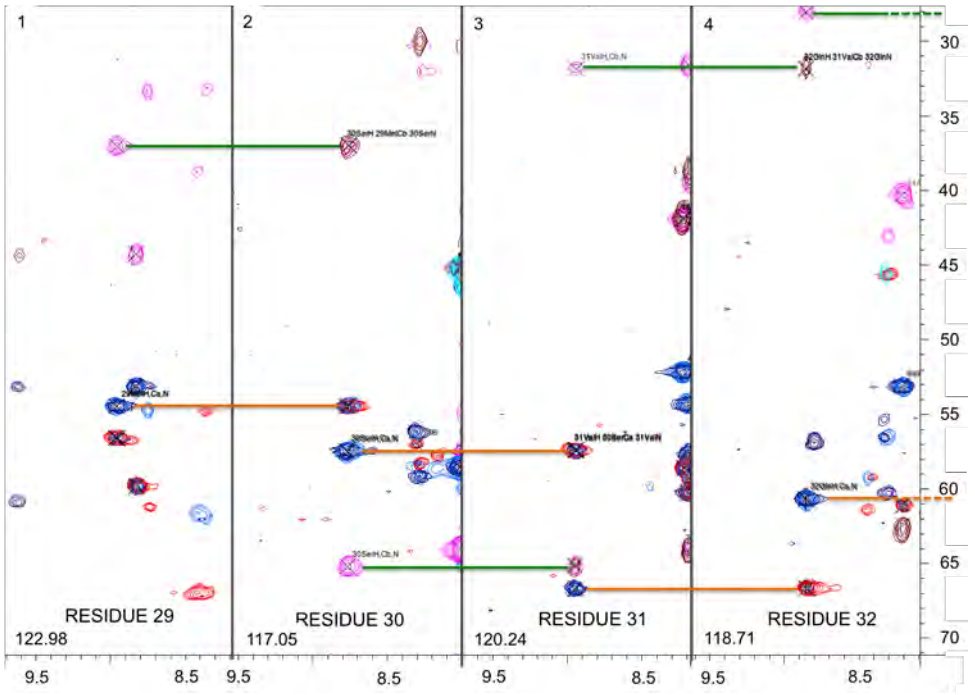


FIGURE 3.8: Example of typical backbone assignment panel viewed with the CCPN Analysis software. The software allows the visualisation of all 6 experimental spectra (HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB, HN(CO)CACB) overlapped, showing the resonance peaks corresponding to each experiment in different colours: blue = Ca (*i* and *i-1* residues) from HNCA, red = Ca (*i-1* residue) from HN(CO)CA, magenta= Ca and Cb (*i* and *i-1* residues) from HNCACB, brown = Ca and Cb (*i-1* residue) from HN(CO)CACB. Each strip (labelled 1 - 4) shows the same ^{1}H-^{13}C dimension (x axis and y axis respectively) at different chemical shift values of ^{15}N. This is in order to identify the peak resonances of sequential amino acids in the backbone. In this example, the orange lines join Ca peaks from residues *i* and *i-1* that belong to associated experiments HNCA and HN(CO)CA (recognised for having the same ^{13}C chemical shift value), whereas green lines join Cb peaks from HNCACB and HN(CO)CACB experiments.

3.2.3 Side-chain Assignment

TOCSY experiments were carried out for the assignment of the desmoplakin linker domain side chain as described in Section 3.1.3.1. In this case a higher concentration of the protein sample was required (785 μM). TOCSY experiments enabled the identification of side chain carbons, while for carbons in aromatic rings side chains an aromatic-NOESY HSQC experiment was performed (Fig. 3.9).

Data analysis was performed in this case using SPARKY analysis software [151].

Some corrections to backbone assignment were done during the side chain assignment giving a total of 100 residues assigned with a higher level of accuracy than the backbone assignment, representing a 85.4% of the amino acids of the protein. Residues towards

the N-terminal end were unassigned after side-chain information was available due to high ambiguity in the discernment between amino acids of the same nature. In other words, because sequential assignment in the backbone was impossible to obtain for this region of the protein, it was not possible to identify with total certainty for example what Arg-associated-peak corresponded to what Arg-position in the sequence.

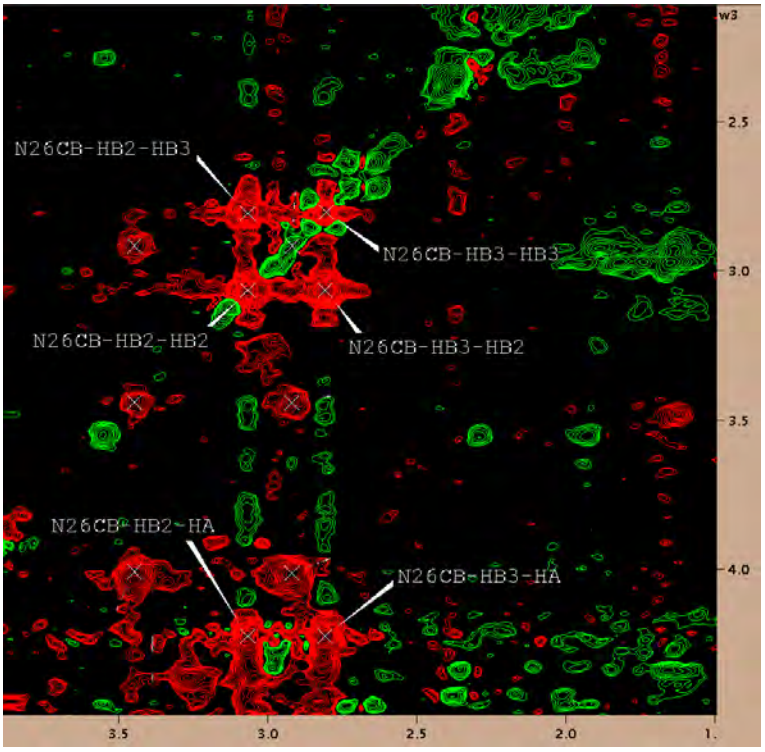


FIGURE 3.9: Example of HCCH spectrum for side chain assignment of the desmoplakin linker domain performed using SPARKY software. This spectrum corresponds to the Cb plane of the N26 amino acid in the desmoplakin linker sequence with the x axis being the Cβ coupled proton dimension and the y axis being the TOCSY dimension that shows the side chain protons. It is possible to observe that the Hb_2-Hb_2 and Hb_3-Hb_3 peaks belong to the diagonal of the spectrum (labelled as N26CB-HB2-HB2 and N26CB-HB3-HB3), while the resonance peaks of coupled hydrogens are located in the vertical strips at 3.1 and 2.8 ppm for Hb_2 and Hb_3 respectively. This image exemplifies the Cb plane of Figure 3.3.

Nonetheless, given the number of successfully assigned residues in the protein it was decided to continue with the structure determination process by assigning NOE peaks and in this way finding the inter-protein distance restraints that will define the protein architecture.

3.2.4 Structural Calculations

3.2.4.1 Dihedral angles

Protein three-dimensional structure is determined by the sequence of its constituent amino acid residues and by the relative orientations between these. The second part of this information can be obtained from the dihedral angles of the protein.

A dihedral angle is a geometrical term for the angle between two planes. In the protein backbone, there are three dihedral angles from which one, ω, involves the backbone atoms $C\alpha$-C'-NH-$C\alpha$, specifying the $C\alpha$-$C\alpha$ distance, and is restricted to 180 ° by the planarity of the peptide bond. The other two, ϕ and ψ, parameterise the relative orientation of the groups on both sides of the $C\alpha$ atom, i.e. C' and N respectively [153].

The desmoplakin linker domain dihedral angles were calculated using the TALOS empirical prediction system [154]. TALOS principle is based on the high correlation between chemical shifts and protein secondary structure and it uses a combination of the assigned five chemical shifts ($^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, $^{1}H_\alpha$ and ^{15}N) to deliver a quantitative prediction for the dihedral angles ϕ and ψ, and also for the protein secondary structure (Fig. 3.10).

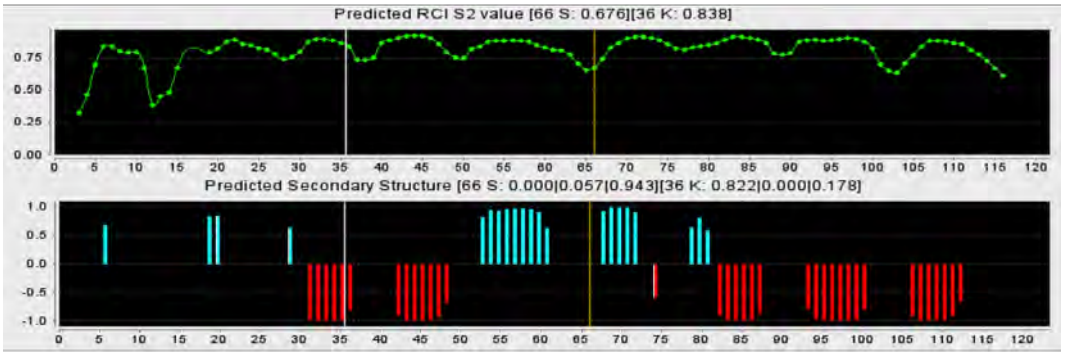


FIGURE 3.10: TALOS graphical output for the desmoplakin linker domain secondary structure prediction. On top the S2 order parameter gives an indication of local dynamics in the molecule. In this case it was calculated using the random coil index (RCI). Values close to one indicate rigidity, while lower values indicate dynamic regions. The plot shows a high rigidity level (S2 $>$ 0.75) for the DPlink protein with few exceptions, including a dynamic region (S2 $<$ 0.5) in the N-terminal end of the molecule. On the bottom the blue bars represent beta strands and the red bars represent alpha helices in the positions for secondary structure prediction according to the amino acid sequence indicated by the numbering in the x axis.

3.2.4.2 NOE restraints

The NOESY HSQCs used to provide the spatial configuration of the DPlink residues, were a ^{15}N resolved NOESY and an aliphatic ^{13}C NOESY.

An average of 10 NOEs per residue were identified, some of them resulted considerably low intensity. The number of NOEs per residue was considered acceptable according to the completeness parameter that compares the expected NOEs number with the experimentally obtained number [152].

3.2.4.3 H bonds and H-D exchange NMR experiment

Intramolecular hydrogen bonds play a vital role in structure determination by providing additional stabilising restraints.

Hydrogen bonds are electromagnetic attractive interactions between a hydrogen and an electronegative acceptor atom. In the case of the protein backbone, these bonds form between the amide hydrogens and the carbonyl oxygens. The space between amino acids participating in a hydrogen bond is a decisive factor in the secondary structure determination. If the spacing between two residues linked by a H bond is of four positions an alpha helix takes place, if the spacing involves two strands of alternating residues, the result is a beta sheet [155]. The backbone amide groups of one strand form hydrogen bond patterns with the backbone carbonyl groups of adjacent strands. Planar inter-strand H bonds take place in antiparallel arrangements, conferring the strongest inter-strand stability. The orientation of parallel strands results in a lightly less stable arrangement of a non-planar H bond network.

H-D exchange can occur because of the natural exchange of protons existing between the H of the amide groups of the protein in solution and the protons of the solvent. If the solvent is initially H_2O and is changed to D_2O the exchange will occur between deuterium atoms and the amide hydrogens. The exchange rate will give a reflection of hydrogen bonding in the molecule, since atoms participating in H bonds will be slower to exchange with deuterons. As D (or ^{2}H) is an isotope of H, it contains an extra neutron in its nucleus, consequently modifying its total spin, so instead of having spin $\frac{1}{2}$, it has a spin of 1. Thus NMR spectroscopy differentiates these nuclei easily. The experiment consists of the acquisition of an HSQC of the protein in water (where

resonances have been previously fully assigned) and, after changing the solvent to D_2O by lyophilisation (freeze-drying), the recording of consequent SOFAST-HMQC (Flip-Angle Short-Transient heteronuclear multiple quantum coherence [156]) experiments at different times during H-D exchange. The NMR peak intensity will diminish as a function of time for those resonances corresponding to residues that have performed an H-D exchange, whereas signal for residues that are involved in H bonds will remain visible for longer, indicating a stable H-bond [157].

The desmoplakin linker domain was exposed to an H-D exchange with SOFAST-HMQCs acquired every 10 min for a total of 100 min. A comparison of the original HSQC with the protein in water and the SOFAST-HMQC acquired 100 min after H-D exchange is shown in Figure 3.11. A significant number of peaks remain after 100 min in D_2O, demonstrating a considerable content of H bonds, arguing against protein generalised high flexibility.

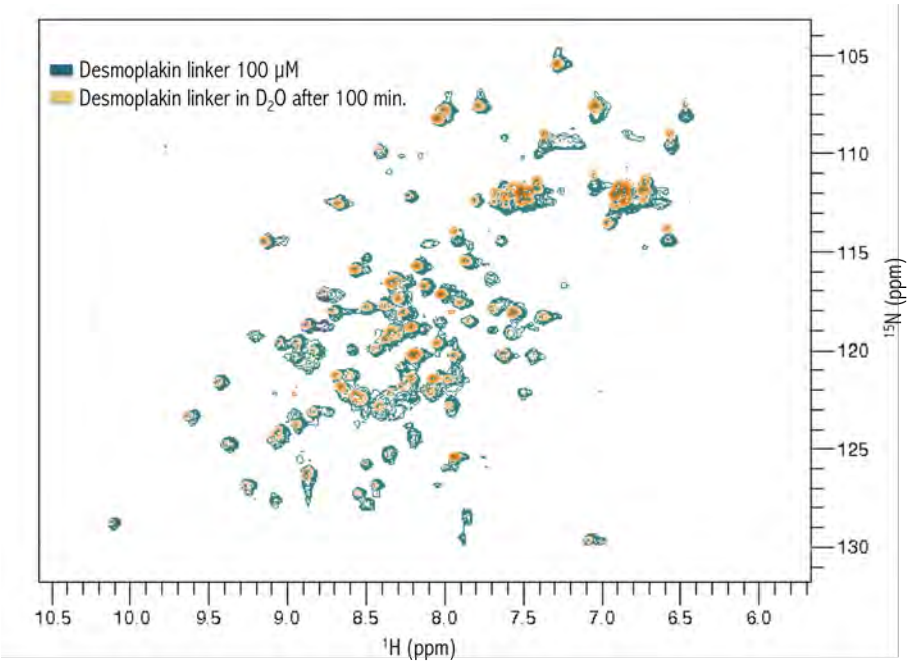


FIGURE 3.11: HSQC spectra of desmoplakin linker at 200 μM in 20 mM HEPES buffer at pH 7.5 and no salt, acquired with a 800 MHz NMR spectrometer. The overlapped spectra display in teal the resonances for the protein in water solvent and in yellow-orange the resonances for the same protein but after 100 min in D_2O. It is possible to identify a significant number of peaks present in both spectra, indicating a considerable content of H bonds.

After identification of residues that form H bonds, such restrictions are included along with the dihedral angles for structure calculations.

3.2.4.4 Structure calculations: ARIA

The programme used for NOE automated assignment and structure refinement was ARIA (Ambiguous Restraints for Iterative Assignment) [158].

ARIA works with an iterative algorithm process based on the extraction of all the possible information provided by NOEs that have been assigned manually, which is enough to define an initial overall fold of the molecule. From these initial NOEs some can be peaks that have been clearly identified but not unambiguously assigned. The correct identification of peaks is critical to ensure a convenient development of the calculations, since the iteration zero of the programme is based on this information, and all the following iterations will be determined by it. A second part of information necessary for the calculation process is generated by ARIA using a standard peak-picking algorithm to select the peaks in the NOE spectra and create a peak list from which the programme obtains distance restraints after a process of assignment and calibration. Iteration zero calculates an initial ensemble of structures based on the manually selected peak list and using simulated annealing to deal with any ambiguities, adding an ambiguous distance restraint (ADR) for every NOE to incorporate the information from ambiguous peaks. From this ensemble the structures with the lowest total energy are selected to be analysed for restraint violations and, after removing all systematic violations, to proceed with the following iterations, which continue until no significant change in the structure is detected by the programme. Finally, the programme delivers a set of calculated structures in pdb format, assigned distance restraints and a list of any kind of rejections classified as such during the calculation process [159].

In order to run ARIA for the calculation of the desmoplakin linker domain a series of inputs were provided, that included the following: the sequence of the protein; the chemical shifts peaks lists of ^{15}N, ^{13}C and aromatic peaks; the distance tolerances; the number of runs of refine steps (4000 or 40000 depending on the case); and the necessary constraints dictated by the dihedral angles provided by TALOS.

As a final set of constraints the hydrogen bonds, identified in the H-D exchange experiments, were incorporated only when the structure had been partially solved as we need to know the acceptor.

3.2.5 Structure of Desmoplakin Linker Domain

After completion of all amino acid assignments and inclusion of NOE, dihedral angles and H bonds restrictions a tertiary structure for the desmoplakin linker domain was obtained using the ARIA calculation program.

The desmoplakin linker structure consists of a short alpha helix **H1** of only 5 residues, followed by two anti-parallel alpha helices **H2** and **H3** comprising both 7 residues. Following this there is a three-stranded beta sheet with two of the strands, **S1** and **S2**, forming a long beta hairpin comprising 19 residues with the first strand shorter than the second one by one amino acid, and the third beta strand, **S3**, with 5 residues. Three more alpha helices follow, **H4**, **H5** and **H6** consisting of 7, 9 and 8 resides respectively (Fig. 3.12).

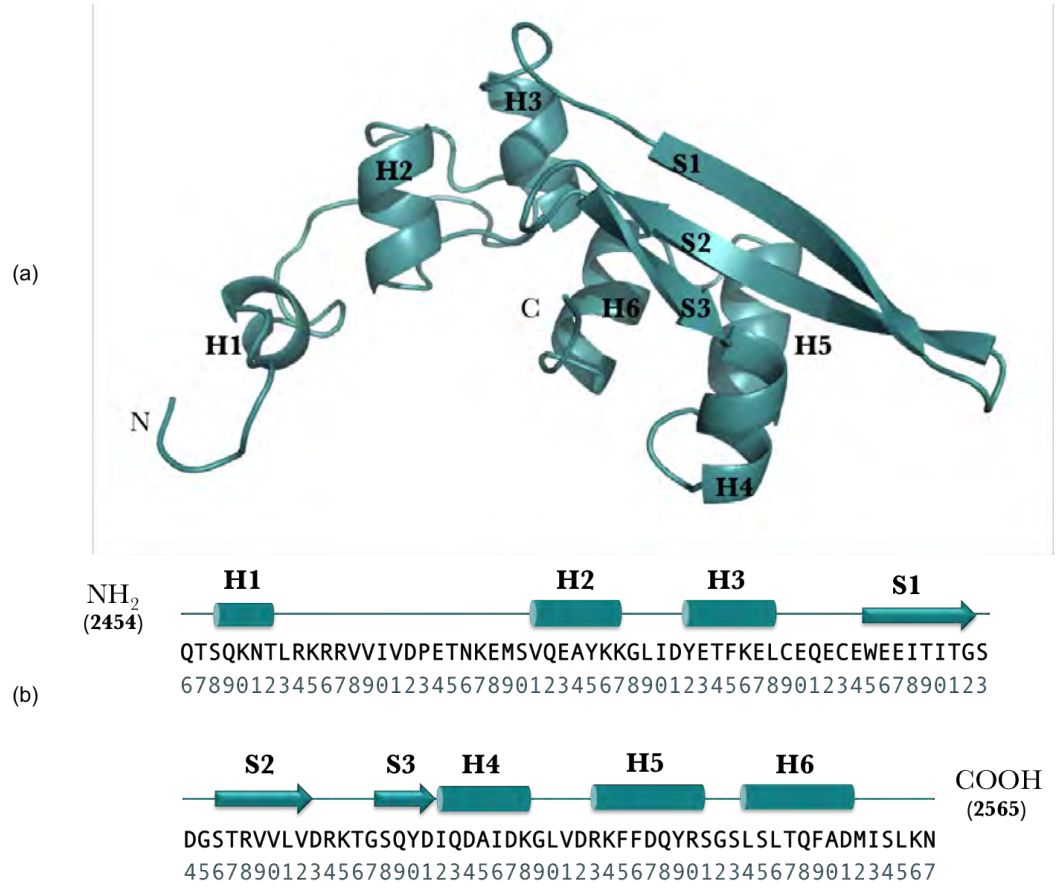


FIGURE 3.12: Desmoplakin linker structure obtained by NMR spectroscopy. (a) Tertiary structure of the desmoplakin linker domain determined using ARIA consisting of a pair of regular (towards the C-terminal end of the protein) and irregular domains (towards the N-terminal end of the protein). Alpha helices are identified as H and the beta strands as S. (b) Secondary structure of the desmoplakin linker domain aligned with its amino acid sequence. The alpha helices represent 38.3% of the secondary structure, whereas beta strands represent only 17.8% of the protein structure.

For simplicity, the residue numbering will follow the one shown in Figure 3.12 (b), with desmoplakin residue 2454 corresponding to residue 6 of desmoplakin linker and so on. Numbering starts with 6 due to the 5 residues included at the beginning of the N-terminal end of the NMR structure of DPlink (Gly, Pro, Leu, Gly, Ser) for the PreScission protease cleavable GST tag and the BamH1 restriction site.

A feature that stands out is that the 4 residues (R14-R17) constituting the basic cluster in the N-terminal region of the protein belong to a large loop and are solvent exposed, 3 of them with the same orientation, available for interaction with other proteins. Following them, also in the loop, 2 acidic residues separated in one position (D22 and E24) are oriented in the same direction, promoting possible ionic interaction with other proteins (Fig. 3.13 (a)).

As well as the mentioned residues, most of the charged amino acids (R,K positively charged and D,E negatively charged) are exposed on the surface of the protein as expected from their hydrophilic properties. These are K10 in **H1** and K36 in **H2**; E50, in the loop between **H3** and **S1**; D64 in the loop between **S1** and **S2**, at the opposite extreme from the basic cluster in the N-terminal end; K88 at the end of **H4**, pointing convergently with K10; and finally K94 and D111 in the concave side of the protein oriented in the same direction. Ionic interactions within the protein are inferred from the juxtaposition of amino acids bearing opposite electric charges. **H2** forms an ionic bond between residues E33 and K37; the loop between **H3** and **S1** contains a second glutamic acid, E54, that interacts with R74, which belongs to the loop between **S2** and **S3**, providing stability to the three-stranded beta sheet; and the arginine R68 in **S2** binds D81 at the end of **S3**. It can be postulated that the important ionic bonds are between E57 in **S1** and R100 in **H5** which is vital for the overall fold of the protein in a globular structure, and the bond established between D73 and K116, setting the position of the C-terminal end of the protein and the three-stranded beta sheet (Fig. 3.13 (b)).

There is a packing of hydrophobic residues in the core of the protein as expected, especially including residues from the three-stranded beta sheet and the three last alpha helices, but also with contributions from **H2** and **H3** with one tyrosine each. The protein structure shows two subdomains, a regular subdomain (towards the C-terminal end) that has greater packing and a well defined core, and a irregular subdomain (towards

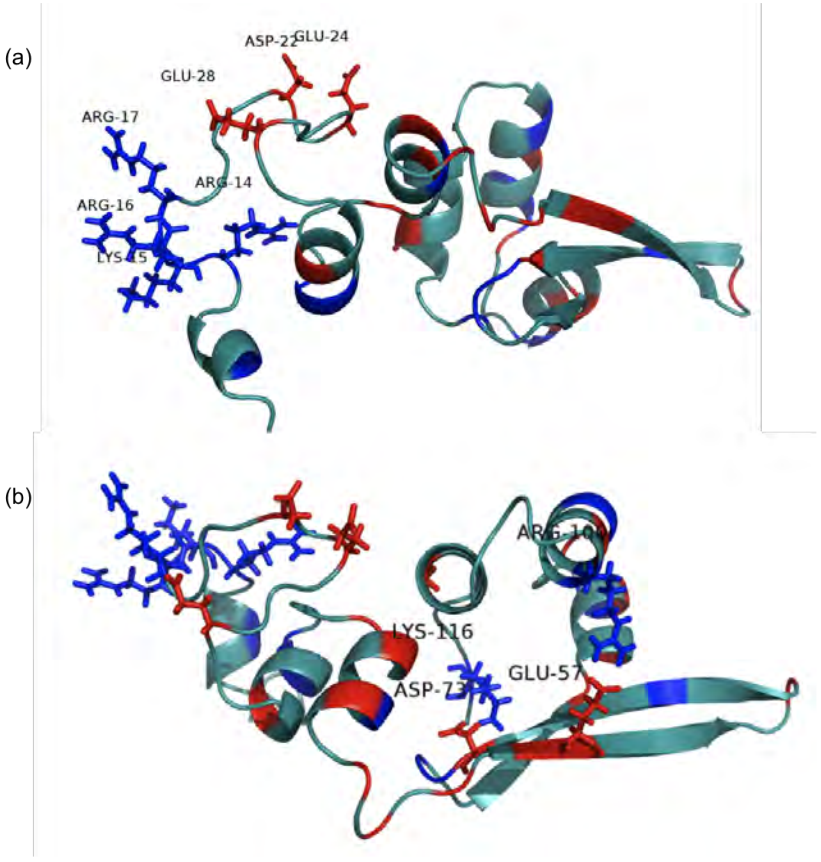


FIGURE 3.13: Different views of the structure of the desmoplakin linker domain showing specific features that characterise it. Basic and acidic residues are highlighted in blue and red respectively. (a) Cluster of basic amino acids in the N-terminal end of desmoplakin linker domain constituted by exposed residues in a large loop and followed by three acidic amino acids also available for inter-protein interactions. (b) Critical ionic bonds suit in place the three-stranded beta sheet of the protein to give the globular fold to the protein.

the N-terminal end) that exhibits a less dense packing of hydrophobic residues with interactions between **H2**, **H3** and also in the loop between **H1** and **H2**, that are not buried inside the protein core (Fig. 3.13 (c)).

All the glycines are located in loops of the protein, providing flexibility to these regions and enabling the required ϕ and ψ angles for tight turns, especially defining starts and ends of the beta sheets. The large N-terminal end loop, on the other hand, does not contain any glycine and instead it includes a proline, amino acid that has the opposite effect to glycine, i.e. it provides rigidity within the structure due to its particular conformational rigidity.

3.2.6 Plakin Repeat Motif Homology

A conserved motif has been identified in the structures of the plakin repeat domains of desmoplakin and other proteins of the plakin family. The 38 amino acid motif, simply called the plakin repeat (PR), has four variations and each one consists of two beta strands in a hairpin structure comprising 11 residues followed by two antiparallel alpha helices comprising 8 and 14 residues respectively [94] (Fig. 3.14 (a)).

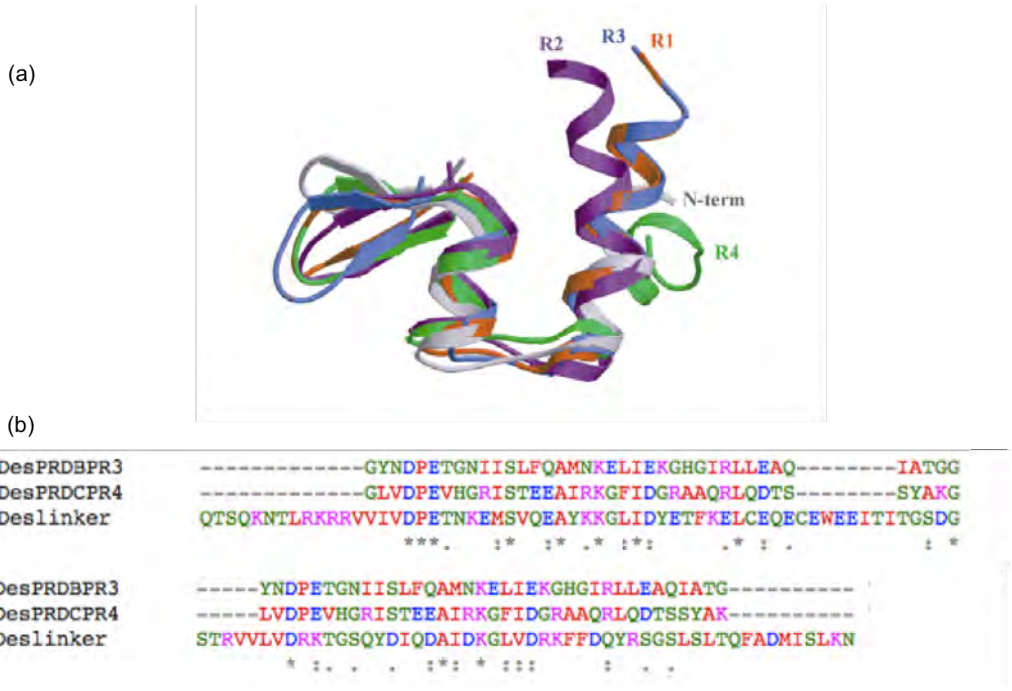


FIGURE 3.14: Plakin repeat motif structure and alignment with desmoplakin linker domain (a) Ribbon representation of the plakin repeat structure, overlaying the four variations of plakin repeats of the desmoplakin PRD B. This 38 amino acid motif is conserved in the structures of the plakin repeat domains of desmoplakin and other proteins of the plakin family. From reference [94]. (b) Alignment of the desmoplakin linker amino acid sequence with two repetitions of the sequence corresponding to the desmoplakin PRD B-PR3 and two repetitions of the desmoplakin PRD C-PR4. The alignment shows the high homology between DPlink and the PR structure suggesting that the linker domain could be composed of two repetitions of the PR canonical motif [94].

A secondary structure prediction for the DPlink protein was derived from the alignment of it with the fourth PR motif of the desmoplakin PRD C and the third plakin repeat motif of the PRD B [160] (Fig.3.14 (b)). This prediction considers the high homology existing between the PRs and the desmoplakin linker sequences, and it is based on the crystal structures of the PRD B and C of desmoplakin. According to this prediction, the linker domain would be composed of two copies of the plakin repeat motif (PR1 and PR2), a structure that has demonstrated well defined characteristics, such as packing

of conserved hydrophobic residues that form hydrophobic cores and ionic interactions between repeats.

NMR data shows partial agreement with the theory of the DPlink being formed of two plakin repeat motifs [160]. The highest agreement between the prediction based on the PR substructure was exhibited in the regular subdomain of the DPlink, which contains a secondary structure that diverges slightly from a PR motif (which would be PR2 according to the prediction) in the length of the strands and helices that constitute it. Nevertheless, the last alpha helix forming the structure of the linker, **H6**, does not have a place in the predicted PR formation (Fig. 3.15). Moreover the loop section in between the supposed PR1 and PR2, according to the homology prediction, is occupied with a long beta strand, the first one of the three-stranded beta sheet of DPlink, demonstrating secondary structure for the previously debated disordered segment of the protein [160].

Analysing the regular subdomain of the DPlink protein in more detail in comparison with the PR motif structure, it is possible to find more subtle disagreements that are critical for the overall shape of the protein. In the case of the PR motif repetition structure in the desmoplakin PRDs, the ionic bond formed between the aspartic acid in position 42, according to the PRDs alignment in Figure 3.15, implicates the lysine in position 57. In the same alignment with DPlink, this numbering corresponds to residues D73 and K88 of the desmoplakin linker, but as shown in Section 3.2.5, the ionic bond that D73 establishes is actually with residue K116 (Fig. 3.13 (b)). As consequence the position of the beta strands with respect the following alpha helices (**H4** and **H5**) differ from the positions that they adopt in the PR motif (Fig. 3.14 (a)).

Although a majority of the hydrophobic interactions occurring after the beta sheet in the NMR structure involve residues from alpha helices **H4** and **H5**, the alpha helical anti-parallel arrangement characteristic of the PR motif is arguable between them, while such symmetry is found between **H5** and **H6**.

Passing to the more divergent case of the irregular subdomain of the DPlink structure, it is possible to see in Figure 3.15 that the secondary structure is only conserved in the case of the alpha helices, admittedly due to the conservation of most of the hydrophobic residues that constitute them, although the lengths and positions vary slightly. In this region, the NMR structure of the linker exhibits two alpha helices, **H2** and **H3**, of 7 amino acids each, while the analogous helices for the PR1 contain 8 and 14 residues

respectively. In both structures the alpha helices can be divided in two regions, one of which participates in hydrophobic interactions between them and the other interacts by hydrophobic residues with the following beta sheets [94].

The lack of the beta hairpin at the beginning of the alignment is obvious evidence of a difference between the NMR structure and the prediction based on the sequence homology with the desmoplakin PRDs. A careful analysis reveals that, although the aspartic acid and the lysine are conserved in positions 22 and 37 (or 4 and 19 in the PRDs numbering according to alignment in fig. 3.15), these residues do not form an ionic bond, instead, they are oriented in opposite direction to each other pointing towards the solvent, being D22 one of the most exposed residues available for inter-protein interactions, as described in Section 3.2.5 and depicted in Figure 3.13 (a). Presumably, the lack of ionic bond at this position is one of the main reasons for the lack of the beta hairpin. Another possible cause is the replacement of the highly conserved glycines (among the PRDs PR substructures [94]) in position 1 for the PRDs sequences in the alignment of Figure 3.15, by a valine in the desmoplakin linker sequence (equivalent residue number 19). As it has been pointed out, in the DPlink structure, the glycines define the start and ends of beta sheets. The same behaviour is followed by the desmoplakin PRDs, where 60% of the beta strands present in PRD B and PRD C start with a glycine, while in the specific case of PRD B PR 3 and PRD C PR 4 (the ones that align best with the linker domain shown in Figure 3.15) all beta strands have a glycine at the beginning of their sequences. A third contribution to the absence of the beta hairpin in the N-terminal end of the linker domain is the few hydrogen bonds present in the region, only 2, in comparison to 8 that stabilise the beta hairpin of the three-strand beta sheet. Nonetheless, the considerable number of hydrophobic amino acids in the N-terminal end loop, as well as the shape in which the loop is assembled suggest the possibility of a non-resolved existent beta hairpin comprising residues V18 to M29 (possibly due to unassigned residues in this region as explained in Section 3.2.3), completing in this way the first PR motif in the desmoplakin linker structure.

In conclusion, the second half of desmoplakin linker domain has a topographically similar structure to the PR motifs with a beta sheet packed against a helix-turn-helix motif, but there are constitutive differences between both structures such as an extra alpha helix and variations in length. On the other hand the first half of the protein also resembles a PR but it does not contain a beta hairpin, making them structurally distinct. This

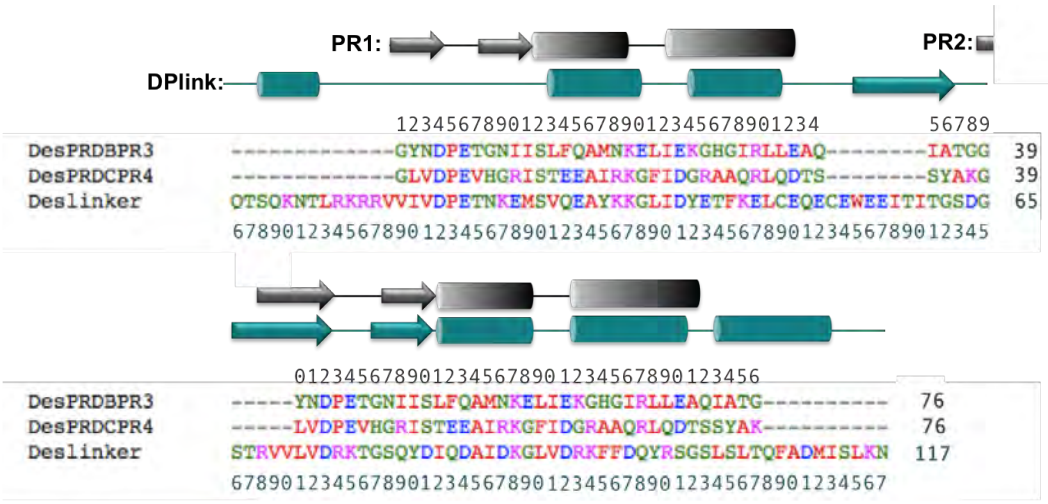


FIGURE 3.15: Alignment of the desmoplakin linker domain secondary structure determined by NMR spectroscopy (shown in teal) with the secondary structure of the desmoplakin plakin repeat canonical motif found in PRD B and PRD C (shown in grey). Despite the high homology between the amino acid sequences of the desmoplakin linker and the PRDs B and C, the secondary structure of DPlink obtained in this work does not resemble completely the predicted structure based on two repetitions of the PR canonical motif. Especially in the case of the first subdomain of DPlink, which does not include a beta hair pin between **H1** and **H2**, as expected from the alignment with PR1. Numbering is shown on top of the alignment in grey for both PRDs and on bottom of the alignment in teal for DPlink.

result is in contrast with the fact that the first half of the desmoplakin linker sequence holds higher homology with the PRDs in comparison with the second half, therefore it would be expected the first half of DPlink to approximate a PR structure.

3.3 Desmoplakin Linker Domain and BC Construct Morphology Determination by SAXS

3.3.1 Scattering Intensity Distribution

As previously described in Section 1.3.2.1, after exposing the sample to the X-ray beam in SAXS experiments, the scattered intensity is recorded. The intensity is a function of the momentum transfer q of each particle, which is in turn a function of the scattering angle 2θ and the incident radiation wavelength λ [26]:

$$q = 4\pi sin\theta/\lambda \tag{3.7}$$

Each particle in the molecule contributes to the scattering profile depending on their positions and orientations. By integrating this information it is possible to obtain a intensity distribution curve that is related to the morphology of the molecule. This is the rough SAXS data; from here different procedures can take place to perform an interpretation and an eventual model of the protein.

In order to obtain more information from the intensity profile, a Fourier transform can be applied to the intensity distribution function to obtain the interatomic distance distribution function of the scattering particle *P(r)*, where r refers to the interatomic vector length within a molecule. In this way, the intensity distribution function *I(q)* at zero angle ($q = 0$), which is proportional to the mass and volume of the protein [25], is given by:

$$I(0) = 4\pi \int_0^{Dmax} P(r)dr \tag{3.8}$$

Where D_{max} is the maximum distance present in the scattering distribution, reflecting the size of the molecule, and for which $P(D_{max}) = 0$. Thus, the extra information obtainable from this function resides in the fact that it reflects the internal symmetry of the protein and the domain structure [26].

While D_{max} reflects the size of the protein more than the shape, there is a complementary parameter that refers directly to shape of the molecule, changing for proteins with the same size but different shape. This parameter is called radius of gyration, R_g, which is calculated as the root mean square distance of the molecule's elements from its centre of gravity, weighted by their scattering densities [25]. The radius of gyration describes the molecule mass distribution.

R_g can also be calculated from *P(r)*, but it is more usually taken from the Guinier approximation for small values of q [25]:

$$I(q) = I(0)exp(\frac{-q^2R_g^2}{3}) \tag{3.9}$$

${R_g}^2/3$ is the slope value for the function $\log(I(q))$ *vs.* q^2 (Guinier plot). If the Guinier plot shows a straight line for the region $q^2 \sim 0$, that is indicative of the absence of aggregation or polydispersity. An overestimation of R_g is often the result of attractive

inter-particle interactions, while repulsive inter-particle interaction causes this parameter to be underestimated. Therefore the Guinier plot is often used as a first examination step for the analysis of SAXS data [23, 26].

For globular proteins the range of momentum transfer for which the Guinier approximation is valid is $qR_g < 1.3$, while for elongated proteins that limit is smaller [25].

In summary, behaviour of the intensity distribution function *I(q)*, Guinier region linearity and the interatomic distance distribution function *P(r)* are essential elements of an initial sample characterisation in SAXS experiments, and that will be evaluated for the cases of the desmoplakin linker domain and the B-linker-C domains construct in the following sections.

3.3.2 Desmoplakin Linker Domain Morphology Determination by SAXS

3.3.2.1 Small angle scattering results

Small angle scattering experiments were carried out on desmoplakin linker after its monomeric and monodispersed state was confirmed with AUC analysis (Fig. 3.5).

A full set of experimental data was completed, based on the buffer subtracted scattering curve acquired at the DESY beam line X33 (Fig. 3.16). Three different concentrations were used, as specified in Section 2.8.1, resulting in complete convergent scattering curves. Data were processed with the PRIMUS programme package [161] for manipulation of experimental SAXS data. The intensity distribution curve (Fig. 3.16 (a)) shows a discernible flat region for q values close to zero, which is a reflection of a mono-dispersed sample [26], allowing further analysis.

The linearity of the Guinier region corroborated the good quality of the DPlink sample (Fig. 3.16 (b)), i.e. no aggregation of the sample and no inter-particle effects that could interfere in the data analysis [23]. The radii of gyration (R_g) and the maximum particle sizes (D_{max}) were obtained using GNOM (a small-angle scattering data processing programme) [162]. The values of these parameters were R_g= 1.68 ± 0.17 nm, demonstrating a globular protein and D_{max}= 5.91 nm. A final criterion of analysis for this first stage of data processing is the behaviour of *P(r)*. As a real spatial representation of the scattering data, *P(r)* displays the characteristics of the molecule shape [23]. For

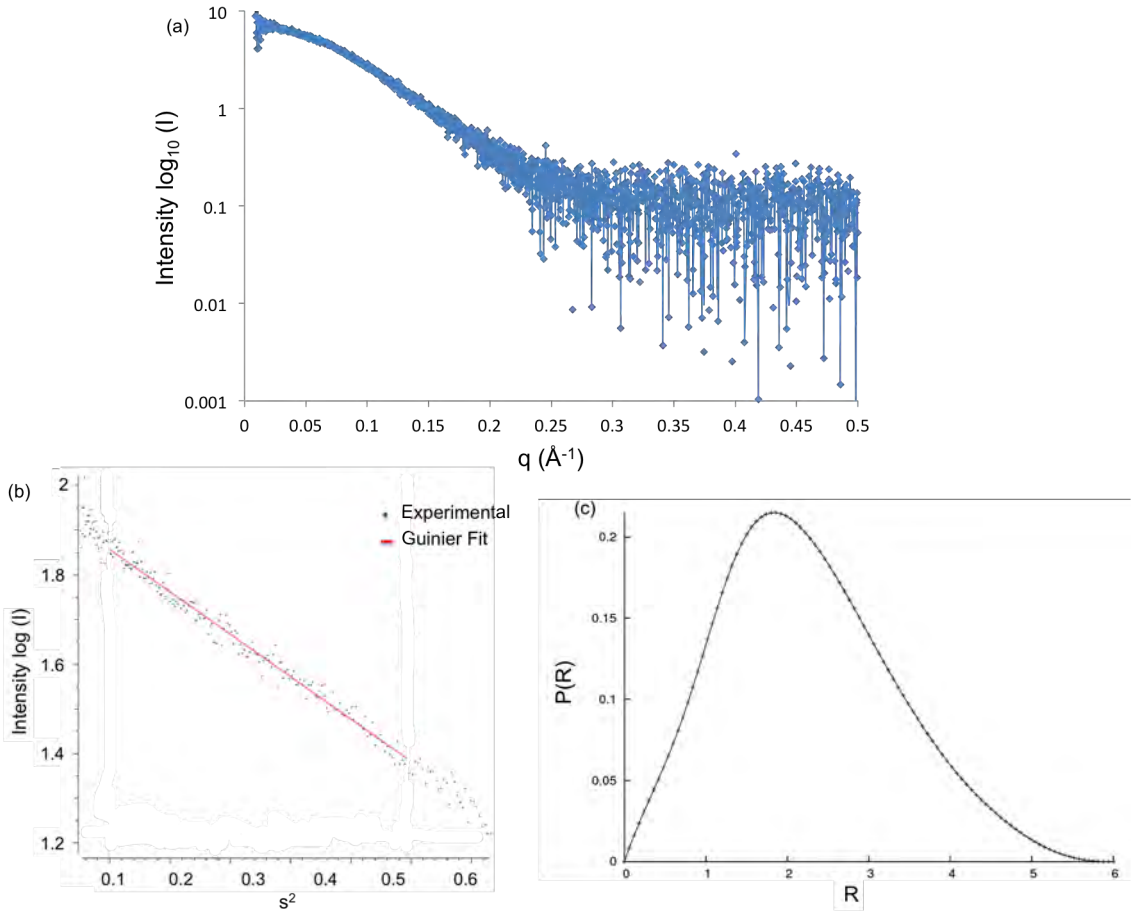


FIGURE 3.16: Scattering data of desmoplakin linker domain sample. (a) Intensity distribution curve of DPlink. q = $4\pi\sin\theta/\lambda$, where θ is the scattering angle, and λ is the wavelength of the X-ray. The X-ray wavelength used in this case was λ=0.15 nm. (b) Guinier fit plotting the scattering intensity I in logarithmic scale against the modulus of the momentum transfer s, i.e. the variable q but with different units (nm instead of Å) [23]. The linearity of the plot indicates absence of aggregation or polydispersity. (c) The interatomic distance distribution function exhibiting a bell-shaped profile with maximum diameter (D_{max}) of 5.91nm.

desmoplakin linker domain, the bell-shaped profile manifests itself as a generally globular monomeric protein. The fact that the maximum of the distribution is shifted to lower values of r indicates the globular shape is not spherical, so a more complex shape can be expected.

As mentioned before, $P(D_{max})$= 0, but the way in which the function tends to zero can reveal possible artifacts that have an effect on the D_{max} value. If the region of the curve around D_{max} shows an abrupt approach then it is normally a consequence of inter-particle effects. If, on the contrary, it crosses the x axis, it means aggregation of the sample. In the case of desmoplakin linker the plot of $P(r)$ demonstrates a concave

approach to the zero region, which exhibits a sample in good conditions for SAXS analysis and simply reflects the non-hard surfaces distinctive of a protein [26].

3.3.2.2 Modelling

In order to extract the relevant information from the scattering data obtained, ten low-resolution *ab-initio* models were generated using DAMMIF (a rapid *ab initio* shape determination programme for SAXS experiments [27]), and averaged in a representative model as an envelope of the protein.

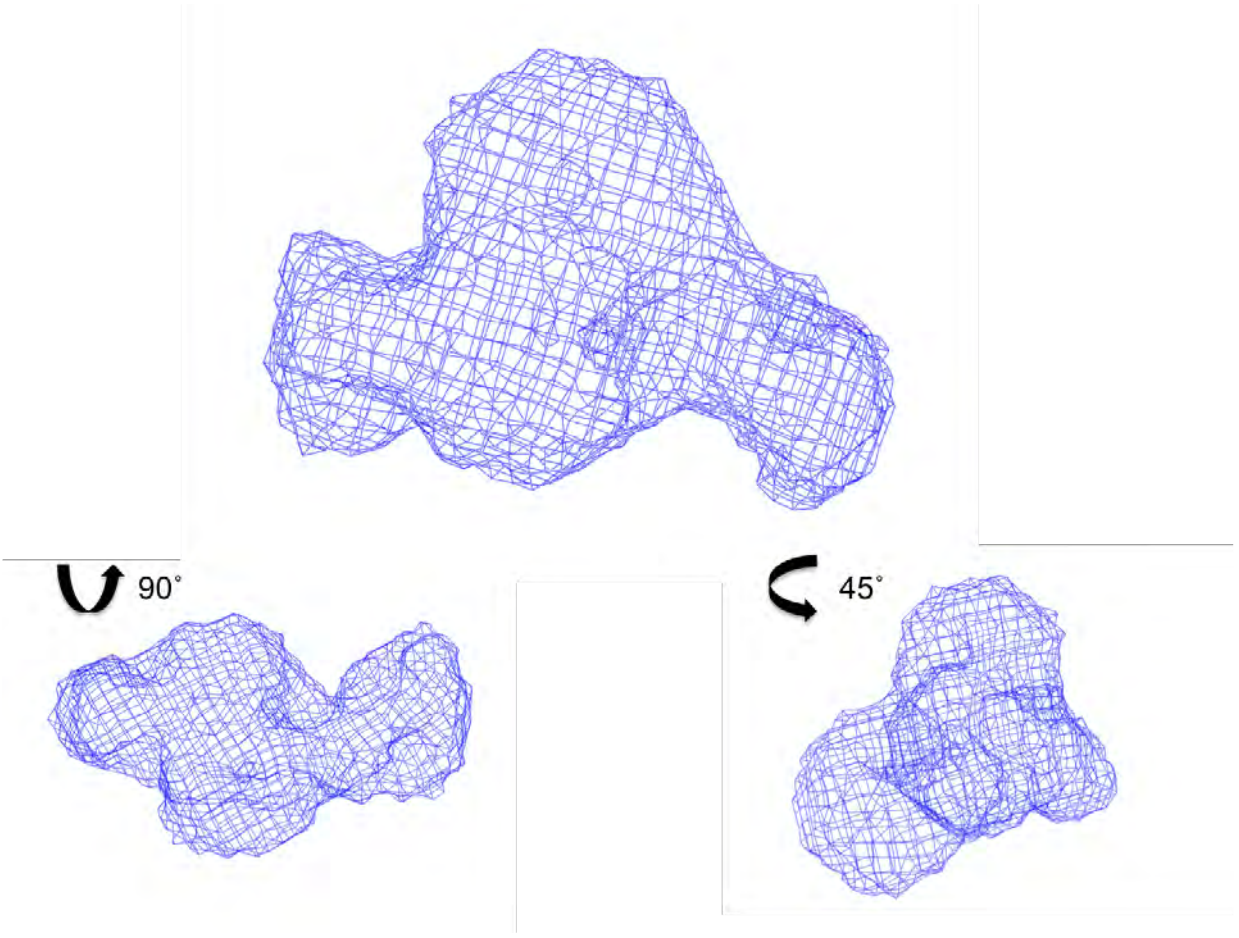


FIGURE 3.17: Envelope of desmoplakin linker domain produced from SAXS data using DAMMIF *ab initio* modeling calculations and visualised with Pymol in orthogonal views and at a 45° angle. The envelope shows a monomeric state with a semi-globular shape as predicted by the interatomic distance distribution function of the scattering particle *P(r)*.

The envelope reveals a semi-globular structure with a quasi-triangular shape, as expected from the *P(r)* profile, suggesting a monomeric state in solution for the desmoplakin linker domain with a level of rigidity that confirmed the D_{max} and R_g values. This result is

not completely unexpected, given the high population of H bonds found in the H-D exchange experiments, which indicated a stable fold in Section 3.2.4.3. A certain degree of flexibility is also denoted by the superimposition of the different models generated by DAMMIF before averaging (data not shown). This is also in agreement with the NMR results for the N-terminal end of the linker domain, which was demonstrated to be highly flexible, causing difficulties for the assignment of residues contained in that region of the molecule.

3.3.2.3 Comparison with NMR structure

The SAXS data acquired for the desmoplakin linker domain was used to validate the NMR structure.

A theoretical scattering distribution is calculated from the desmoplakin linker domain atomic structure generated by NMR, this is evaluated against experimental scattering data from the same molecule. This is carried out with the calculation programme CRYSOL [163]. This programme is based on the spherical harmonics approximation to calculate scattering intensity from atomic models or known atomic structures. By providing the atomic coordinates, CRYSOL produces a theoretical solution scattering curve that can be fit to an experimental scattering distribution. This analysis shows a high agreement between the experimental scattering curve associated with the envelope of the desmoplakin linker domain and the NMR atomic structure, with a χ^2 value of 0.94, as shown in Figure 3.18.

A manual docking was also carried out by overlapping the SAXS envelope of DPlink and the NMR structure, obtaining a good fit with a low level of spatial discrepancy (Fig. 3.19). The quasi-triangular shape of the SAXS envelope is confirmed by the atomic structure, allowing a clear identification of correlation between regions within the molecule.

There are only two regions of the atomic NMR structure that do not fit perfectly inside the SAXS envelope. One of these regions corresponds to the N-terminal end first 2 residues Glu and Thr, plus the preceding 5 residues at the beginning of the N-terminal end that do not form part of the linker domain but constitute the PreScission protease

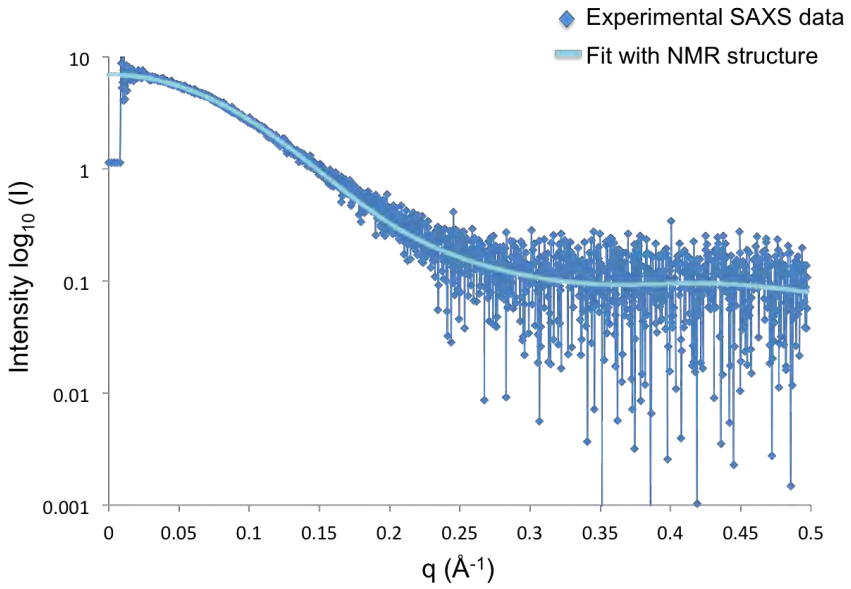


FIGURE 3.18: Fitting of the theoretical scattering distribution calculated from the DPlink NMR structure with scattering intensity curve result of the experimental SAXS results generated by CRYSOL. The SAXS data is shown in blue and the CRYSOL fit is shown in cyan. The Chi value of the fit is 0.94.

cleavable GST tag and the BamH1 restriction site, G1-S5 in the NMR structure numbering. The other region that is visibly outside of the envelope is the large loop, also in the N-terminal end of the protein, which has been here classified as highly flexible. The intrinsic flexibility of this region that comprises residues T12-S30, equivalent to residues 2460-2478 in the protein sequence, possibly caused a high level of discrepancy between the 10 models generated by DAMMIF, leading to the exclusion of such region from the final average rigid model. This result, confirming the flexibility of the N-terminal end of DPlink could imply that it posses important binding properties, a hypothesis that will be discussed in Chapter 4.

Apart from this intrinsic incompatibility between a flexible region of DPlink structure and the generated envelope that is built based on a series of rigid body movements and rotations, the positions and orientations of all the other portions of the NMR structure were accommodated within the SAXS envelope.

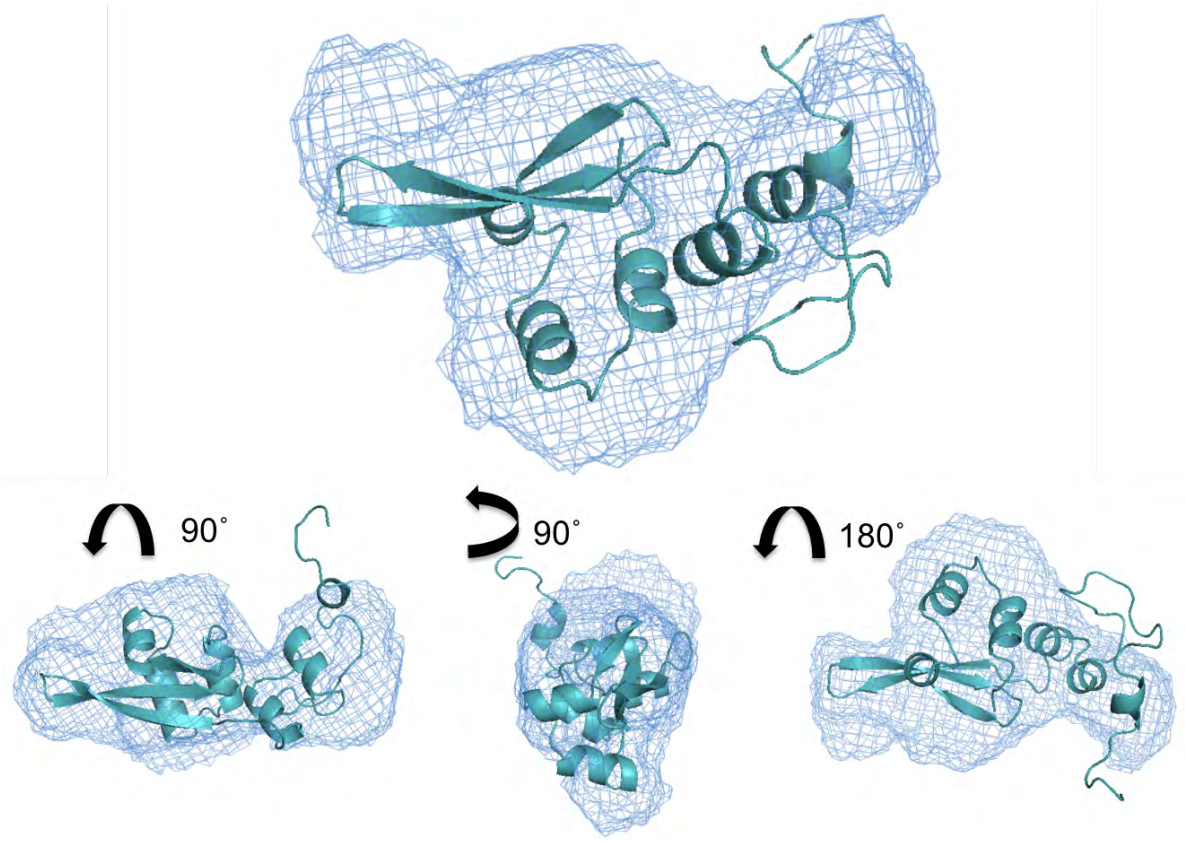


FIGURE 3.19: Manual docking of the SAXS envelope obtained for the desmoplakin linker domain and the atomic structure generated by NMR for the same protein. The level of discrepancy is minimum, margined to two regions. The first region includes the five residues that do not belong to the DPlink sequence (Gly, Pro, Leu, Gly, Ser, from which Gly, Ser are the BamH1 site) plus the first two residues at the beginning of the N-terminal end. The second region comprises residues T12 to S30.

3.3.3 Desmoplakin B-linker-C Construct Morphology Determination by SAXS

As described in the introduction of this thesis in Sections 1.5.3.1 and 1.5.4, it can be seen how the desmoplakin linker domain forms an essential part of the C-terminal end of desmoplakin, with apparent critical binding properties with the IF type III. But it has also been suggested by experimental data that DPlink does not act individually as the only binding agent, but it collaborates with the other domains in desmoplakin to perform the necessary interaction with other proteins. Such domains are likely to be the desmoplakin plakin repeat domains B and C, that precedes and follows the linker domain respectively. Given such circumstances, it was considered relevant to investigate the construct formed by desmoplakin PRD B, linker domain and PRD C (BC construct)

by SAXS, in order to ascertain its configuration as a multidomain system and shed some light into the organisation of the C-terminal end of desmoplakin.

The BC construct has a molecular weight of 63.9 kDa, exceeding the typical size limitations of NMR spectroscopy for proteins structure determination. Because of this, and also because the anticipated highly dynamic character inherent to multidomain protein complexes, SAXS experiments represent a promising technique, capable of dealing with flexible macromolecules in solution.

3.3.3.1 Small angle scattering results

The desmoplakin PRD B-linker-PRD C (BC) construct was purified following the His-tagged protein purification protocol described in Section 2.3.3. The purity of the BC construct, that includes the desmoplakin PRD B, the linker domain, the spacer region, the PRD C and the His tag, was assessed by SDS-PAGE (Fig. 3.20).

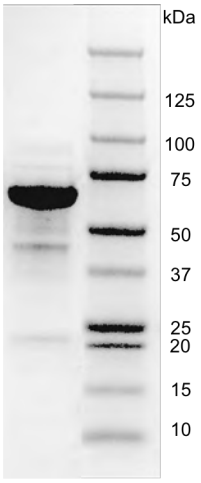


FIGURE 3.20: Purity of the desmoplakin BC construct (left lane) after His-tag purification tested by SDS PAGE with molecular weight markers (right lane). The band, that corresponds to one single species, is in the correct range of molecular weight between 50 kDa and 75 kDa for the BC predicted to be 63.9 kDa. The purity level reached was $> 90\%$.

After confirming the high quality of the sample in terms of purity and monodispersity, SAXS experiments were carried out on the desmoplakin BC construct.

In the same way as for DPlink, a full set of experimental data acquisition was conducted at the DESY beam line X33 (Fig. 3.21). The intensity distribution curve (Fig. 3.21 (a)) did not show a clear flat region for $q \sim 0$, but an examination of the Guinier plot

associated with the BC data (Fig. 3.21 (b)) demonstrated linearity of the function $I(q^2)$ for $q^2 \sim 0$, indicating no aggregation of the sample and no inter-particle effects. The radii of gyration (R_g) and the maximum particle sizes (D_{max}) using GNOM, revealing the values of R_g= 4.44 ± 0.35 nm, and D_{max}= 15.56 nm. Contrary to DPlink, the *P(r)* curve of the BC construct does not present a bell-shaped profile, but it rather yielded a plot with multiple local maxima, characteristic of a multidomain complex, due to intra and inter sub-domain distances [23]. As expected, the *P(r)* profile also behaves different to a globular protein curve, shifting its overall maximum away from $D_{max}/2$ to lower values of *q*. The concave approach of the curve to the D_{max} value is a final confirmation of the good quality of the protein discarding any possible doubts from earlier examinations about aggregation or inter-particle effects (Fig. 3.21 (c)).

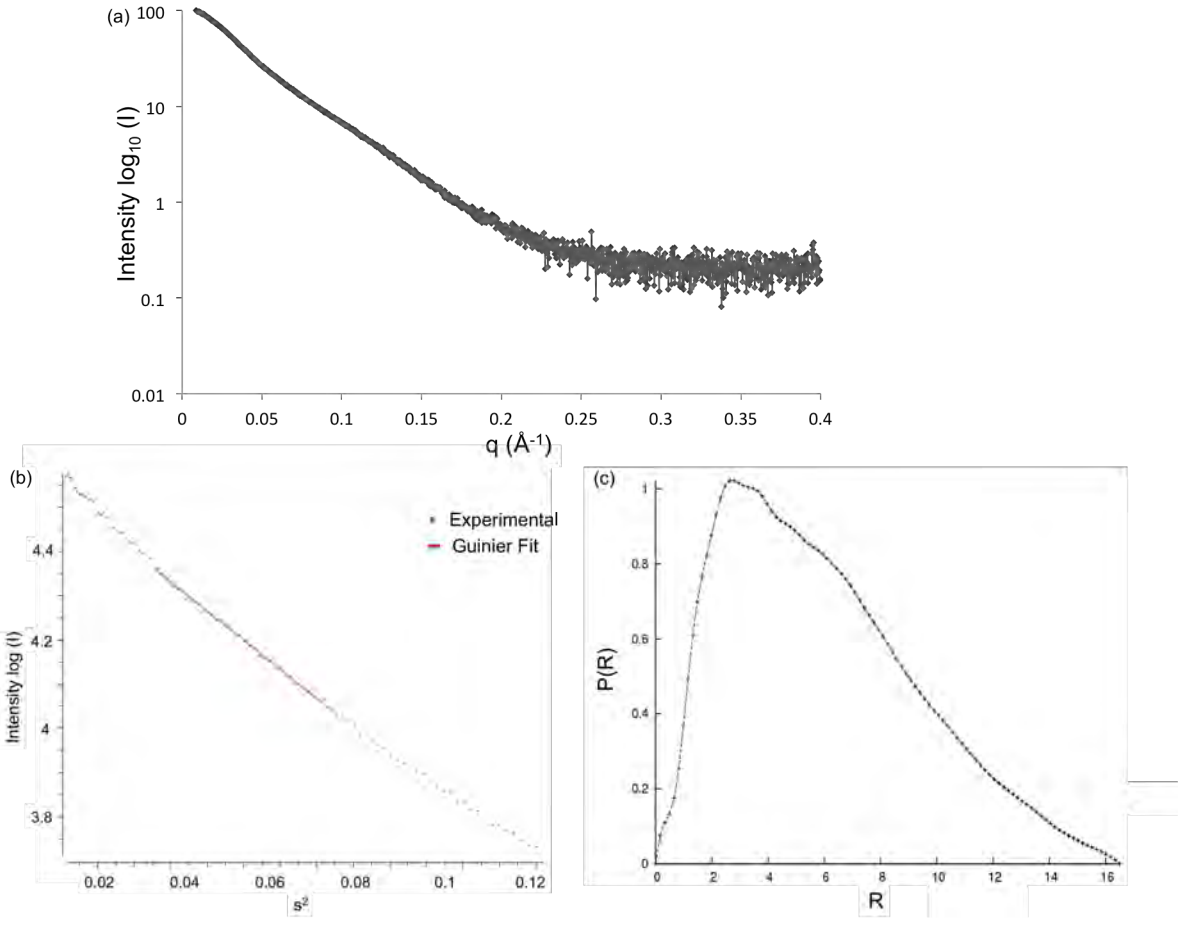


FIGURE 3.21: Scattering data of desmoplakin BC construct sample. (a) Intensity distribution curve of the BC construct The X-ray wavelength used in this case was λ=0.15 nm. (b) Guinier fit plotting the scattering intensity I in logarithmic scale against the modulus of the momentum transfer *s*. The linearity of the plot indicates absence of aggregation or polydispersity. (c) The interatomic distance distribution function exhibiting a bell-shaped profile with maximum diameter (D_{max}) of 15.56 nm.

3.3.3.2 Modelling

Ten low-resolution *ab initio* bead models were generated and one average and optimised final envelope was calculated by DAMMIF. The envelope reveals an elongated structure with an L-shape very similar to the desmoplakin plakin domain envelope obtained by SAXS [98], and as expected from the *P(r)* profile, it shows what could be classified as 4 subunits.

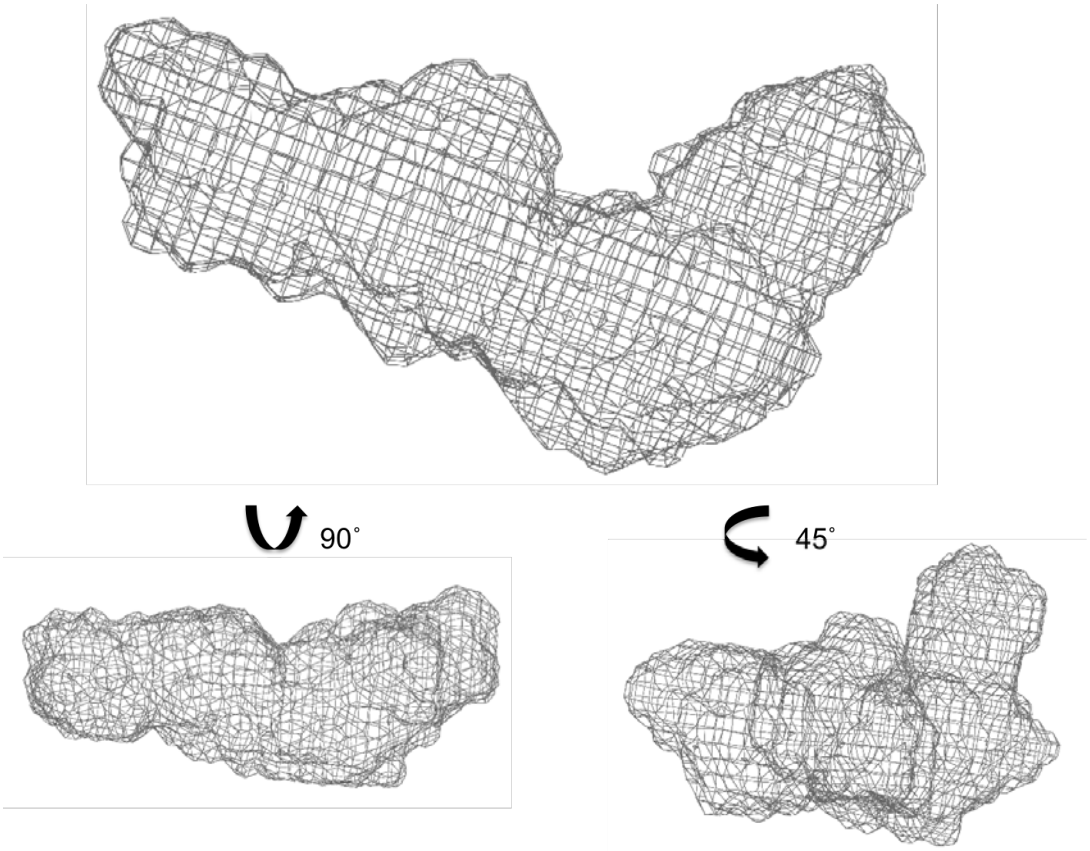


FIGURE 3.22: Envelope of desmoplakin BC construct (including desmoplakin PRD B, C and linker domain) produced from SAXS data using DAMMIF *ab initio* modeling calculations and visualised with Pymol in orthogonal views and at a 45° angle. The envelope shows a structure with elongated shape as predicted by the interatomic distance distribution function of the scattering particle *P(r)*. It can be noticed it is constituted by four domains likely corresponding to the PRD C, the linker domain and the spacer, the PRD B and possibly the His Tag + rTEV protease cleavage site.

The 4 subunits are likely to correspond to the PRD C, the linker domain and the spacer, the PRD B and possibly the His Tag + rTEV protease cleavage site. The envelope showed in Figure 3.22 is experimental evidence that supports the theory reported by Choi *et al.* [94] which suggested that multiple sets of PRDs have an architecture analogous to beads on a string in the case of the C-terminal end region of desmoplakin. Such

arrangement has been the subject of different hypothesis trying to explain how the structure of the plakins C-terminal end allows binding with cytoskeletal proteins. For example, Janda *et al.* have proposed a circular arrangement for the C-terminal end region of plectin, with the first 5 PRD surrounding the 6th PRD in an antiparallel fashion and the linker domain between PRD 5 and 6 being exposed for IF binding [100]. This theory resembles more the head domains globular structures visualised in rotary shadowing electron microscopy images taken for plectin and desmoplakin [90, 100], which show a C-terminal domain of 9 nm-diameter.

On the other hand, the model "beads-on-a-string" proposed by Choi *et al.* suggests an elongated assemble of the PRDs that belong to the desmoplakin C-terminal end, based on a supposedly proteolytic sensitivity of the terminal and its analogous nature to proteins with tandemly repeat domains given the PR motif mentioned in Section 1.5.3.1 [94]. According to this theory, the arrangement of the plakin repeat domains in a string would allow each PRD to interact with IF individually, alluding to the IF binding preferences to be spread in the terminal instead of uniquely found in the linker domain.

3.3.3.3 Comparison with individual desmoplakin domain structures

Crystal structures of desmoplakin plakin repeat domain B and C have been determined already [94] and can be used to build a multidomain model.

A fit of the crystal structures of individual desmoplakin PRDs was done with the SAXS envelope of the BC construct. In the case of the desmoplakin linker domain, the atomic structure elucidated here was used. For the case of the spacer region between the linker domain and the PRD C an I-TASSER model was generated in order to get a better fit with the envelope morphology. In the same way, the short amino acid sequence corresponding to the His Tag + rTEV protease cleavage site was submitted to the I-TASSER server in order to generate a model of the structure, which, although small, is not negligible.

By using the SAXS envelope as a guide, manual docking of the 5 different elements was performed to fit the envelope. PRDs B and C were first located within the envelope, giving them priority for being accepted crystal structures. Then the linker domain structure obtained in this work was placed in between them and finally the empty

spaces were filled with the spacer and cleavage site sequence models. The result can be appreciated in Figure 3.23.

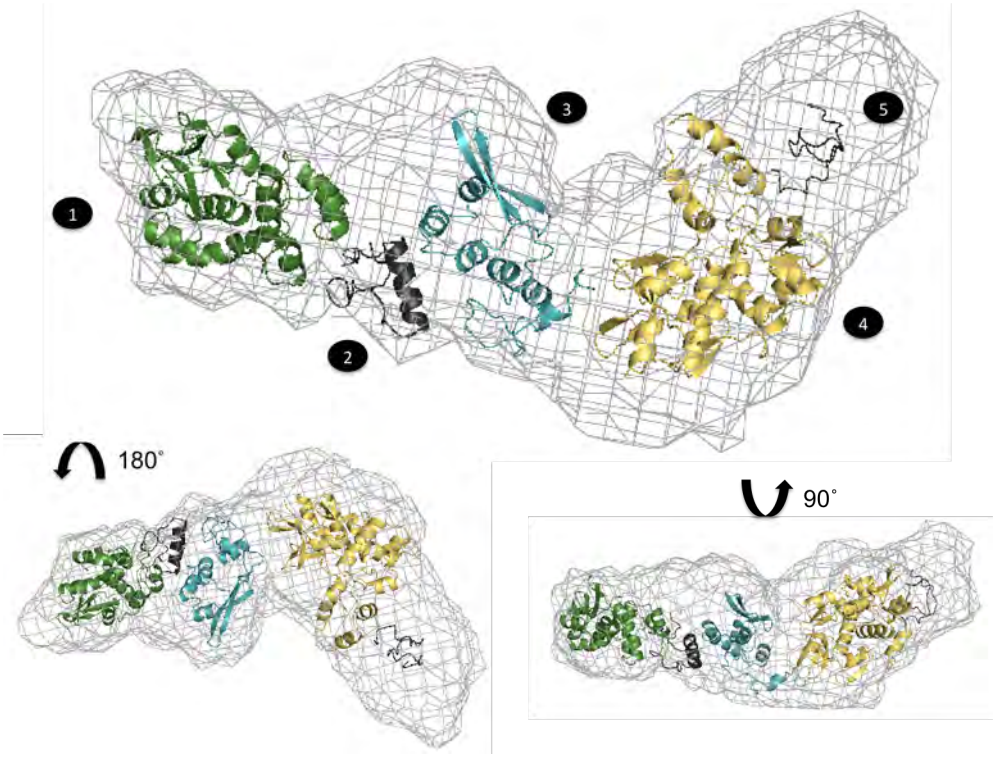


FIGURE 3.23: Envelope of desmoplakin B-linker-C construct produced from SAXS fitting the crystal structures of desmoplakin individual plakin repeat domains B and C and the NMR structure of desmoplakin linker domain (obtained in this work) by manual docking using Pymol. The numbers indicate the following: 1. desmoplakin PRD C; 2. spacer; 3. desmoplakin linker domain; 4. desmoplakin PRD B: 5. His-tag. The orthogonal views and the rotation of 180° reveal a good spatial agreement, with no regions of any domain being in an external position, although the envelope shows empty spaces, especially after the His Tag + rTEV.

As it can be seen, manual fitting of the desmoplakin C-terminal domains into the SAXS envelope demonstrates good agreement for a visual first inspection. All domains can be fitted in the envelope with no regions of any domain being in an external position. However, the envelope shows empty spaces, especially after the His Tag + rTEV protease cleavage site sequence.

With the purpose of having a more quantitative than qualitative analysis, a hypothetical rigid structure was generated by the conjugation of the 5 elements in the positions and orientations according to the manual docking using the envelope as a guide. A theoretical scattering curve was then generated using CRYSOL, to compare it with the experimental SAXS data. The result delivers a regular fit with χ value of 1.87, indicating that a rigid approximation for the BC construct does not represent accurately the state

of the complex in solution. On the other hand, the fact that the fitting accomplishes a χ value lower than 2 suggests that the proposed configuration is commonly adopted by the C-terminal end of desmoplakin within the standards of its inherent flexibility (Fig. 3.24).

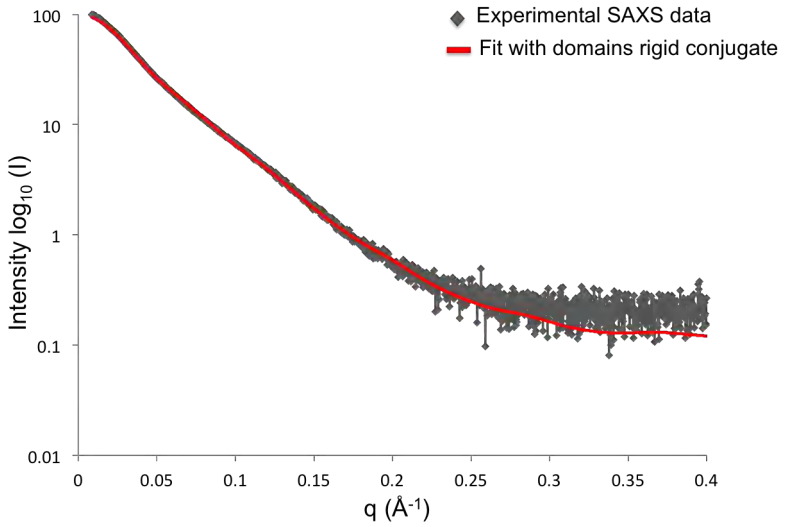


FIGURE 3.24: Fitting of the theoretical scattering curve corresponding hypothetical rigid structure generated by the conjugation of desmoplakin PRDs B and C, linker domain, spacer and His-tag as shown in Figure 3.23 with the experimental SAXS scattering curve of the desmoplakin BC construct generated by CRYSOL. The SAXS data is shown in grey and the CRYSOL fit is shown in red. The Chi value of the fit is 2.06.

Although the conformation presented in Figure 3.24 is the one that gave the lowest χ value from a series of combinations of manual docking, an inverted arrangement of the domains within the envelope (i.e. with the PRD C fitted in the right end of the envelope and the PRD B in the left section of the envelope with respect the orientation depicted in Figure 3.24) obtained a minimum χ of 2.06, demonstrating similar validity to the presented arrangement. In pursuance of a better fit, and considering a certain degree of flexibility should be conferred to the BC construct structure, a second approach was tested in which the spacer region between desmoplakin linker domain and PRD C was left free to adopt any sterically allowed conformation. This structure, located in the semi-dynamic regime, maintains the PRD B and the linker domain together as a fixed rigid domain on one side and the PRD C as another rigid domain on the other side, being flexible in the middle. In this case the fit was slightly better with a χ value of 1.44, demonstrating that the conferred degrees of freedom had a positive input in the representation of the desmoplakin BC construct conformation (Fig. 3.25).

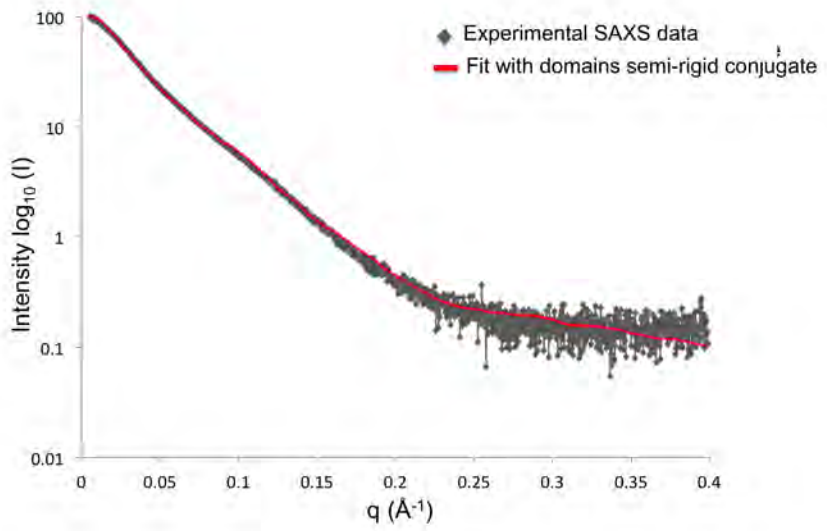


FIGURE 3.25: Fitting of the theoretical scattering curve corresponding hypothetical rigid structure generated by the conjugation of desmoplakin PRDs B and C, linker domain and His-tag as shown in Figure 3.23 but with a flexible spacer, with the experimental SAXS scattering curve of the desmoplakin BC construct generated by CRYSOL. The SAXS data is shown in grey and the CRYSOL fit is shown in red. The Chi value of the fit is 1.44.

As concluded in Section 3.2.5, the linker domain also contains flexible regions, contributing in this way to the dynamic nature of the BC construct.

3.4 Overview of Chapter 3

With the aim of providing supporting evidence of the accuracy of the NMR structure of the desmoplakin linker domain presented in this work, a comparison of it with the models provided by diverse protein structure prediction programmes can be performed.

Different structure predictions, CSI (Chemical Shift Index) [164] and PSIPRED (protein structure prediction server) [165], demonstrated a high agreement with the plakin repeat motif homology based secondary structure prediction for residues V31-K36, A85-K88 and K94-Y99 to be alpha helices and residues V69-V72 to be beta strands. Doing a comparison of the percentages of alpha helices and beta strands given by the mentioned methods with the PR's homology and the CD predictions for DPlink secondary structure 3.6, it is possible to see that CD shows a much higher population of helical structures and a slightly lower population of beta sheets (Table 3.1).

Secondary Structure Prediction	Alpha Helix content (%)	Beta Strand content (%)
CSI	16.0	14.5
PSIPRED	29.9	15.3
PR's homology	33.3	14.5
CD	48.0	12.0

TABLE 3.1: Table showing the respective percentages of alpha helix and beta strand structures for the DPlink as predicted by four different methods: NMR chemical shifts data based prediction using CSI; sequence based prediction using PSIPRED; prediction based on the homology with the PR motif found in the desmoplakin PRDs B and C; and circular dichroism (CD) prediction. The NMR structure has a 38.8% of alpha helix content whereas beta strands only represent the 17.8% of the protein secondary structure.

The secondary structure percentages result of the NMR structure determination of the desmoplakin linker domain indicate that the alpha helices represent a 38.8% of the protein structure, whereas beta strands represent a 17.8%. These results are closer to the PR motif homology described in Section 3.2.6, suggesting that, despite the discrepancy found between a model formed by a succession of PR subunits and the experimental results from NMR spectroscopy, the desmoplakin linker domain shares PR homology with the desmoplakin PRDs to certain extent.

The iterative threading assembly refinement (I-TASSER) server [166] was also used to generate a model of the desmoplakin linker domain from the protein sequence only, to compare with the NMR resulting structure. I-TASSER uses the empirical approach of protein threading [167] (used to produce models of proteins that do not have homologous proteins in the Protein Data Bank (PDB), but do have the same fold as proteins of known structures) to provide predictions for the secondary and tertiary structures of proteins. The I-TASSER server revealed a predicted model for the tertiary structure of the desmoplakin linker domain shown in Figure 3.26. This model produces the highest C-score equal to -1.049, in a range of [-5,2], where 2 is the best result achievable, according to convergence of the structure assembly simulations.

As it can be noticed from Figure 3.26, the I-TASSER prediction for the protein structure also shows only one beta hairpin, in agreement with the NMR structure and in contrast to the two predicted by the PR motif-based model. Differences between the I-TASSER model and the NMR structure are observed, in that the lengths of the alpha helices are different, however the positions coincide with the calculated structure, testifying in

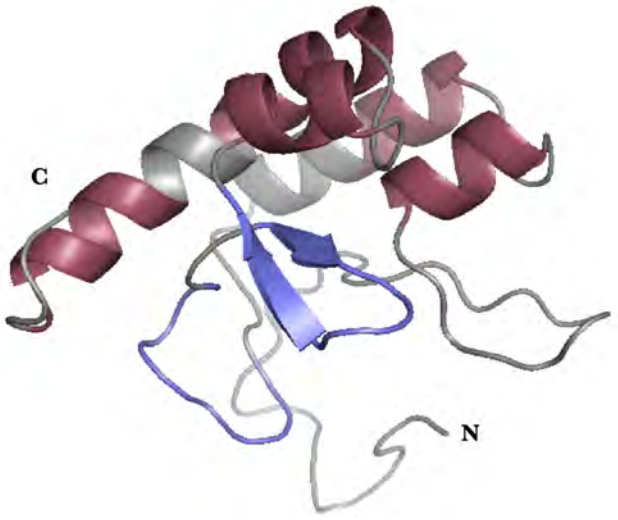


FIGURE 3.26: I-TASSAR model for desmoplakin linker domain based on the amino acid sequence of the protein with NMR resulting secondary structure highlighted: alpha helices in magenta and beta sheet in blue. This model, the one with the highest C-score (-1.049) of the five results given by the I-TASSER platform, was generated using 10 different templates, has demonstrated to have the highest similarity with the NMR structure.

this way as to the reliability of I-TASSER as a modelling method and simultaneously working as a confirmation of our results.

Regarding the desmoplakin PRD B -linker-PRD C construct, the elongated shape of the corresponding SAXS envelope and its D_{max} of 15.56 nm are features in agreement with the "beads on a string" theory suggested by Choi *et al.*, but two things must be considered, that the presence of the PRD A could be a decisive factor in the architecture of the PRDs and that the isolation of the C-terminal end from the rest of the desmoplakin protein could cause re-arrangements in the configuration of the PRDs. Therefore other theories should not be discarded completely. By the attempt presented here to ascertain the architecture of the desmoplakin BC construct, it was intended to shed light on how its structural integrity can be maintained in the presence of continuous shearing forces and extensive mechanical stress. In particular, by studying interdomain distance distributions, and therefore obtaining enough information about structural organisation, it can be possible to understand how protein activity is modulated. Regulation of associated biological processes [21] can also be better understood at a general level, including IF binding, a desmoplakin important function which specificity will be addressed in the following chapter.

Chapter 4

Interaction of Desmoplakin with Intermediate Filaments

This chapter focuses on the experimental results obtained from the NMR titration experiments carried out between desmoplakin and periplakin linker domains and the intermediate filaments type III. The results and corresponding analysis are preceded by a brief description of previous results relevant to the subject obtained by different research groups.

4.1 Interaction of the Plakin Family of Proteins with Intermediate Filaments

Members of the plakin family of proteins associate with intermediate filaments (IFs) [80]. The region of plectin responsible for binding vimentin is located in the carboxyl terminal tail of the protein. It consists of 50 amino acids, residing in the region of plectin that was initially designated repeat domain 5 of the protein [168]. This region was later designated as the fifth PRD B of plectin and the plectin linker domain [78] (after the linker domain was established as a distinctive and highly conserved element of some plakins) (Fig. 4.1). The first 7 of the 50 amino acids are found in the fifth PRD B while the following 43 amino acids are part of the linker domain (Fig. 4.2).

Since the majority of the residues required for IF binding reside in the linker domain of plectin, and since this domain is a region of high homology in plectin, BPAG1e, desmoplakin I and II, envoplakin and periplakin (as mentioned in Section 1.5.1), the linker domain has been of interest regarding IF binding, in particular vimentin binding [168]. Because of this, the IF binding properties associated with the C-terminal tails of the plakin proteins in general have been often attributed to the linker domain [77, 100, 104, 169]

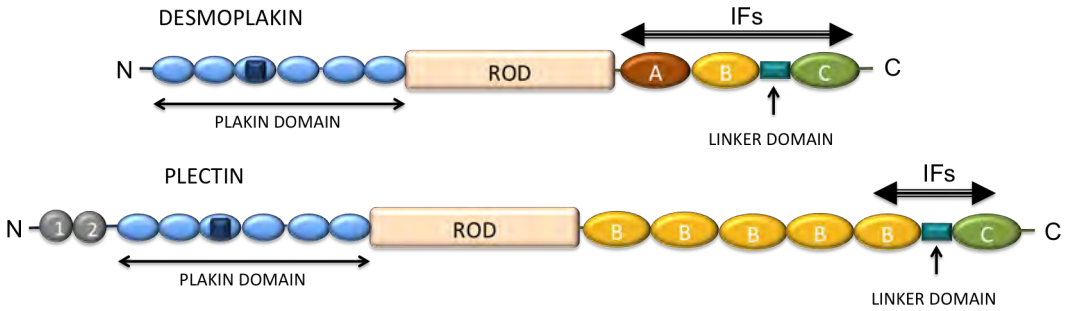


FIGURE 4.1: Schematic representations of plectin and desmoplakin. IF binding sites in plectin have been determined to be in the linker domain, between the 5th PRD B and the PRD C of the protein. In the case of desmoplakin, identification of such binding sites has been more challenging. So far it has been accepted that such binding sites are within the C-terminal tail of the protein. Due to the high homology of the linker domain of the two plakins, the desmoplakin linker domain seems to be a strong candidate for containing such binding sites.

Experimental support for this hypothesis includes BPAG1e, another plakin, which reveals differences in the linker region with desmoplakin, a fact that has been alleged to be the reason why BPAG1e does not interact with either vimentin or K8/K18 keratins, while desmoplakin does [170]. Furthermore, periplakin, a plakin family member whose C-terminal tail consists exclusively of a linker domain, associates with IFs [169, 171, 172]. In order to investigate the binding property of the linker domain, several experimental studies have been performed on the C-terminal end of desmoplakin and periplakin.

4.1.1 Interaction of Intermediate Filaments with Desmoplakin C-terminal Tail

In early experiments, solution binding assays showed that the C-terminal tail of desmoplakin directly interacts with the amino terminal end of epidermal keratins (type II IF), but experiments exploring the association with vimentin (type III IF) gave negative results [102]. The significance of this is not clear as there are numerous examples in which

desmoplakin interacts with type III IF. For example, desmoplakin interacts with desmin in the intercalated discs in cardiac muscle, and with vimentin in the meningeal cells and follicular dendritic cells [173, 174].

In later experiments, the interaction of desmoplakin C-terminal tail with IF type III, vimentin and desmin, was demonstrated [104], albeit with a lower level of binding than with the keratins (binding was at a level about 50% of that of epidermal keratin K1) [104]. Yeast two-hybrid analysis of the C-terminal tail of desmoplakin gave results consistent with the hypothesis that the linker domain has a key role in the interaction of desmoplakin with vimentin [168]. Within the 50 amino acids sequence reported by Nikolic *et al.*, the same group pointed out a cluster of four amino acids in plectin as being essential for vimentin binding [168]. These four residues are R4277-K-R-R4280. In an alignment of plectin with desmoplakin it possible to locate this cluster within the beginning of the linker domain. The equivalent residues in desmoplakin are R2462-K-R-R2465 (Fig. 4.2).

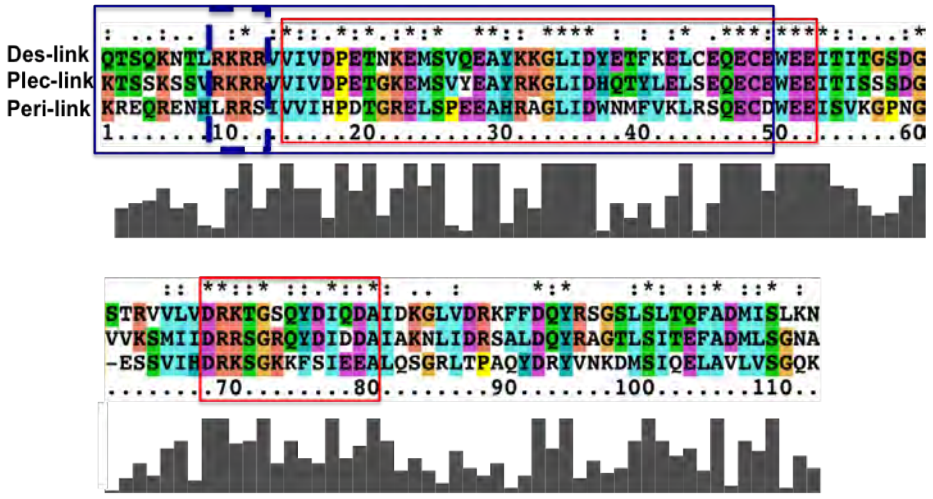


FIGURE 4.2: Sequence alignment of the periplakin, plectin and desmoplakin linker domains performed using Crysol. The red boxes indicate the 2 homologous boxes that, according to Karashima and Watt (2002), are vital for vimentin binding to desmoplkin. The blue box includes 36 of the 50 amino acids that were identified as the plectin binding domain by Nikolic *et al.* (1996). The dashed blue box contains the four residues in plectin essential for vimentin binding [168], and the equivalent residues in desmoplakin and periplakin.

The highly positively charged character of this cluster suggests that the nature of desmoplakin-vimentin binding could be predominantly electrostatic, as supported by the negative results for binding between vimentin and mutant polypeptides containing

neutral amino acids instead of the RLRR cluster [168]. Further experiments were performed by Choi *et al.* [94] to test the ability of the desmoplakin plakin repeat domains (PRDs) to bind vimentin in an individual fashion and as constructs comprising different combinations of PRDs. Co-sedimentation assays demonstrated weak binding at a 1:1 ratio for each individual PRD and for the construct made of PRDs A and B, but for constructs comprising PRDs B and C and PRDs A, B and C the binding to vimentin was comparatively stronger [94].

A key result of Choi *et al.*, was that the construct consisting of desmoplakin PRD B and the linker domain does not bind more strongly than the individual PRDs. Contrasting this result with another result from the same investigation which indicates that the BC construct binds to vimentin more strongly than the AB construct it is possible that the linker domain acts as a required element that confers the flexibility and extensibility to the multiple PRD domains that enable vimentin binding, although it does not possess binding properties as an individual domain. This conclusion is supported by Karashima and Watt, who explored the interaction of periplakin and envoplakin with keratins and vimentin by co-localisation and overlay assays, showing the importance of the C-terminal tail of periplakin in IF binding of both, periplakin and envoplakin [172].

Two homologous boxes have been identified by Karashima and Watt by sequence alignment of periplakin, envoplakin, desmoplakin, plectin and BPAG1 [172]. Deletion of either of the two homologous boxes abolished the ability of the C-terminal tail of periplakin to bind vimentin. This suggests that these two sequence segments in the linker domain are sufficient for vimentin binding. One box (box 1) overlaps the 50 amino acids sequence identified by Nikolic et al (Fig. 4.2). Choi *et al.* also reported the existence of positively charged grooves in the surfaces of desmoplakin PRD B and C and proposed that the residues in those regions interact electrostatically with the vimentin coiled coil domain. This hypothesis is based on manual docking of a dimeric vimentin fragment and the PRDs, the structures of which have been determined by X-ray crystallography [94] (Fig. 4.3).

The basic groove contains residues R2309, R2366 and R2385 in the case of PRD B and K2668, R2725 and R2744 in the case of PRD C. The three residues that belong to each PRD are distributed in the sub-structures PR 2, 3 and 4 within the PRD and they can be aligned with the desmoplakin linker (DPlink) as shown in Figure 4.4.

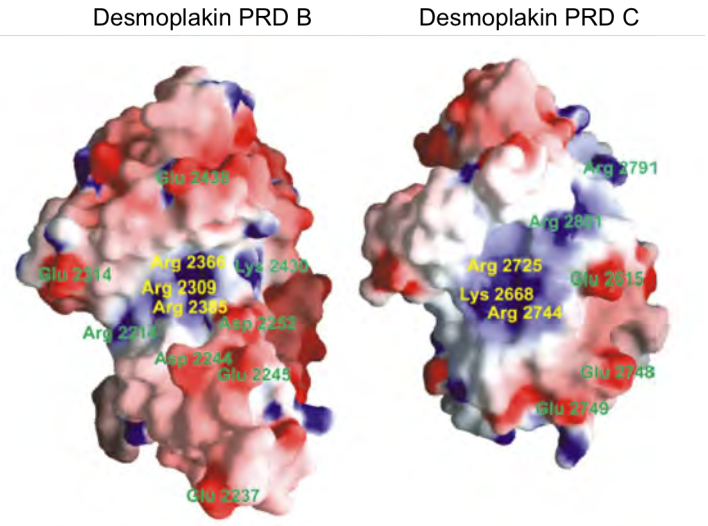


FIGURE 4.3: Electrostatic surface representation of desmoplakin PRDs B and C. Blue represents positively charged regions while red represents negatively charged areas. Conserved amino acids in all plakin family PRDs B and C are shown in yellow and coincide with a conserved basic groove on both domains. From reference [94].

It is worth pointing out that in a sequence alignment between the PRD B and C based on the plakin repeat (PR) sub-structure mentioned in Section 1.5.3.1, it is possible to discern that the residues in the basic grooves are conserved through plakin repeats [94]. This is in agreement with the earlier finding of an acidic-basic residue periodicity in the plakin repeat copies along PRD A, B and C of desmoplakin, that matches the periodicity of the 1B rod domain of IF and therefore suggests again the possibility of an electrostatic interaction between the two proteins [101].

Regarding desmin, experiments including cell transfection, yeast two-hybrid and *in vitro* binding assays have confirmed that the linker domain of desmoplakin binds desmin [175].

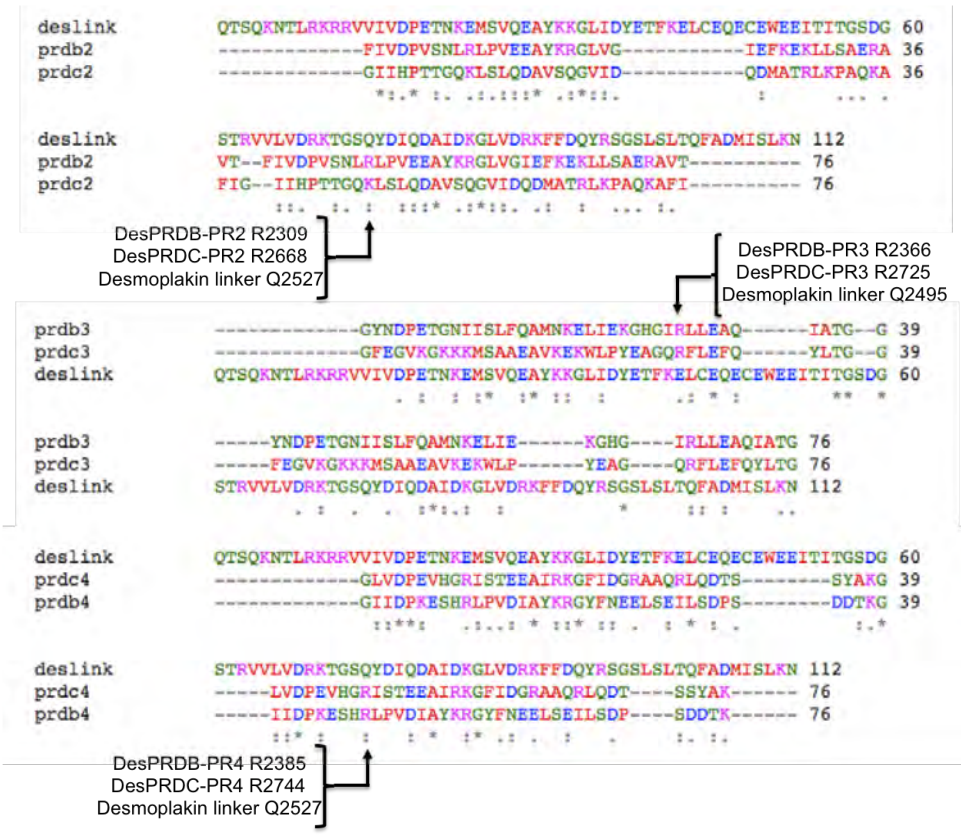


FIGURE 4.4: Sequence alignment of the desmoplakin linker domain (DPlink) with desmoplakin PRD B plakin repeats 2, 3 and 4 and desmoplakin PRD C plakin repeats 2, 3 and 4. Desmoplakin residues that have been identified as belonging to the positively charged groove of PRD B and PRD C, and are suggested to associate with vimentin are indicated with arrows [94].

4.2 NMR Spectroscopic Investigation of Protein Binding

4.2.1 Protein Binding Monitored by Changes in Chemical Shifts and Spectrum Line Broadening

Interaction of a protein with a ligand, (e.g. another protein), can be monitored by NMR based on the physico-chemical changes upon complex formation. Such changes are represented by resonance variations that can be detected in an HSQC spectrum.

When a ligand binds to a protein, the nature of the protein can be affected, which can be reflected in changes of chemical shifts. Also, the molecular weight of the resulting complex increases modifying the dynamics of the free protein in solution, i.e. molecular rotational and translational motions determining the rate constants of the protein. This has an effect on the relaxation times, which are related to the NMR spectrum signal width. Therefore, for a macromolecule or for a small molecule bound to a ligand, the rate

constants increase causing the NMR signals to get broader. If the line broadening effect is too strong the NMR signal essentially disappears [141, 176]. By knowing the assignment of the residues in a labelled protein it is possible to identify which specific residues are the ones experiencing a higher level of perturbation and therefore ascertaining what residues are involved in binding.

Given the high level of specificity with which NMR parameters reflect conformational changes in proteins (to the atomic level), this technique gives more structural information than other spectroscopic techniques and other popular methods used to determine binding between proteins, e.g. co-sedimentation assays and pull-down assays.

4.2.2 Intermediate Filaments as Binding Ligands

The type III intermediate filament proteins vimentin and desmin have molecular weights of approximately 53 kDa, which exceeds the limits for NMR detection as mentioned in Section 1.3.1. This means that any complex formation between vimentin or desmin with any other protein, even if that protein has a low molecular weight, will be evident in a NMR HSQC spectrum of the sample containing both species. Taking advantage of this fact, we used NMR HSQC spectra to reveal the interactions between a variety of proteins and desmin and vimentin, knowing that any interaction with vimentin or desmin will result in line broadening on the HSQC spectrum of the interacting proteins.

NMR titration experiments were carried out to investigate the *in vitro* interaction between vimentin and desmin and plakin proteins. The following ^{15}N labelled wild type proteins were used: desmoplakin PRD C (for titration with vimentin and desmin), desmoplakin linker domain (for titration with vimentin and desmin) and periplakin linker domain (for titration with vimentin). Titration experiments for vimentin were also conducted with ^{15}N labelled mutants of desmoplakin linker domain (DPlink) and mutants of periplakin linker domain (PPlink) that were selected based on a criteria described later in this chapter. Samples for titration were prepared as stated in Section 2.6.5. In all desmoplakin and periplakin linker domains cases titrations were performed using ratios of 1:0, 1:1 and 1:2 for the linkers and the corresponding IF receptively, from which here are presented the results of the ratio 1:2 for each case. Titrations of the desmoplakin PRD C domain with both IF were performed using ratios of 1:0 and 2:1

(in favour of desmoplakin PRD C) only, from which here are presented the results of the ratio 2:1 for each case.

Vimentin and desmin have been purified following the purification protocol described in Section 2.3.4, while all the desmoplakin and periplakin protein domains, including the mutants, have been purified following the GST-tagged protein purification protocol.

4.3 Interaction between Intermediate Filaments and Desmoplakin Explored with NMR

4.3.1 IF Interaction with Desmoplakin Repeat Domain C

As detailed in Section 4.1.1, the PRD C, as the end domain of desmoplakin C-terminal tail, has been identified as a strong candidate for a site of IF binding [177]. However, results reported by others indicate that this individual domain is essential but not sufficient for desmoplakin-IF binding activity [170].

We have carried out binding experiments of desmoplakin PRD C with vimentin and desmin. In order to present the difference in HSQC spectra before and after the addition of IF type III to desmoplakin PRD C, both spectra are overlapped, so broadened peaks or chemical shift changes are evident. In every case a well-dispersed HSQC spectrum was acquired for desmoplakin PRD C alone, confirming its folded nature. For desmoplakin PRD C and desmin experiments the ratio 2:1 (in favour of desmoplakin PRD C) resulted in line broadening of 63.3% of the peaks present in the HSQC spectrum of the desmoplakin domain (fig 4.5).

In the case of desmoplakin PRD C and vimentin experiments the effect was similar to that with desmin but more drastic, causing line broadening of 85.21% of the visible peaks for this domain (Fig. 4.6). This is in agreement with other findings that implicate the desmoplakin tail in IF binding [175].

The majority of the PRD C residues affected by the presence of desmin and vimentin are in the HSQC region that corresponds typically to beta strands, suggesting that regions of the desmoplakin PRD C with this secondary structure are involved in desmin binding, possibly involving the basic amino acids K2668 and R2744, the beta sheet residues that

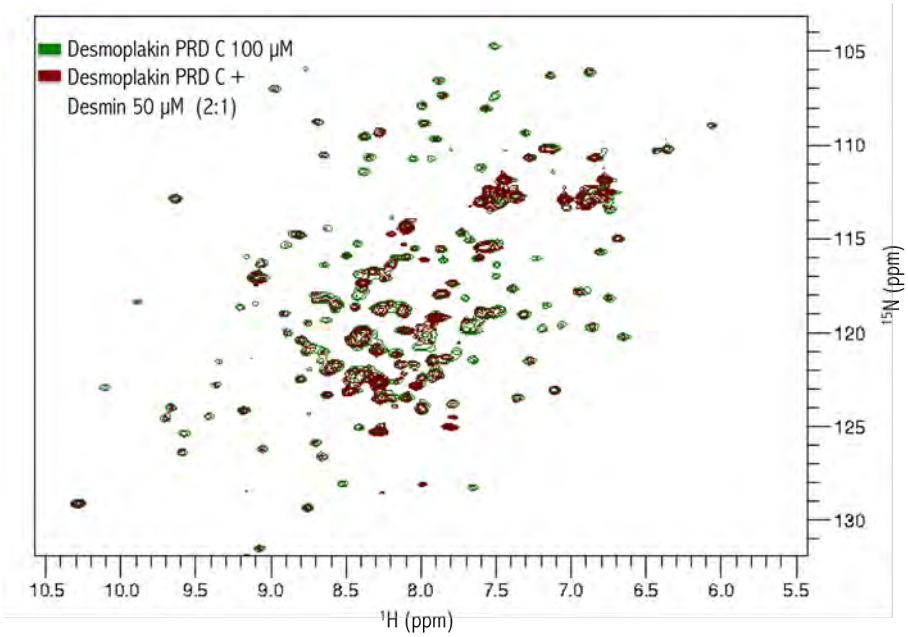


FIGURE 4.5: Overlay of the HSQC spectrum of desmoplakin plakin repeat domain C with and without desmin. In green: HSQC spectrum of desmoplakin plakin repeat domain C at 100 μM in 20 mM HEPES buffer. In magenta: HSQC spectrum of desmoplakin plakin repeat domain C at 100 μM with desmin at 50 μM.

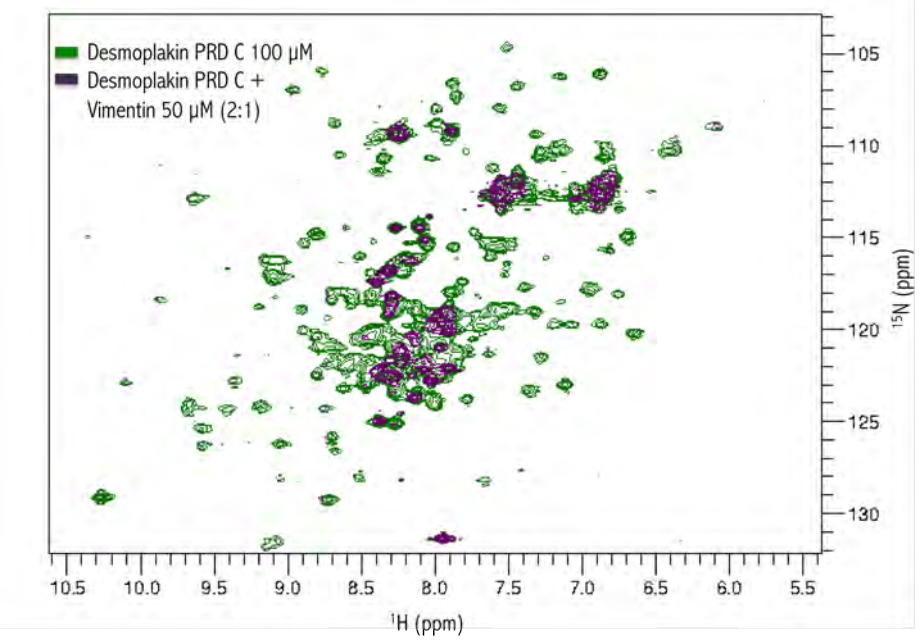


FIGURE 4.6: Overlay of the HSQC spectrum of desmoplakin plakin repeat domain C with and without vimentin. In green: HSQC spectrum of desmoplakin plakin repeat domain C at 100 μM in 20 mM HEPES buffer. In purple: HSQC spectrum of desmoplakin plakin repeat domain C at 100 μM with vimentin at 50 μM.

belong to the conserved electrostatic groove mentioned in Section 4.1.1. In the case of the experiments with vimentin, interaction took place over a wider area, also involving residues in the spectrum regions associated with alpha helices, as can be seen by the

lack of purple peaks on the overlapped HSQC in the central area of the spectrum. Thus residue R2725, the only alpha helix residue residing in the positively charged conserved groove, should be also considered as a residue possibly involved in vimentin binding. Both for vimentin and for desmin, the peaks corresponding to the asparagine and glutamine side chains of desmoplakin PRD C seem to remain considerably strong after addition of the IF.

Deletion of desmoplakin PRD C has been suggested to decrease the interaction between desmoplakin and IF by 50% [104]. According to the results presented here there is evidence of binding between desmoplakin PRD C and vimentin as well as desmin. Therefore, one could conclude the following possibility: desmoplakin PRD C binds IF type III as an individual domain, but it plays a secondary role in binding between desmoplakin and IF type III in physiological conditions.

4.3.2 IF Interaction with Desmoplakin Linker Domain

It is possible that the linker domain of desmoplakin plays a unique and critical role in intermediate filaments binding. Based on the experimental evidence that a combination of various different domains belonging to the C-terminal tail of desmoplakin bind strongly to vimentin if, and only if, they contain the linker domain [170], we decided to examine in more detail the specific interaction between intermediate filaments and the desmoplakin linker domain by NMR.

A well-dispersed HSQC spectrum was obtained for the desmoplakin linker (Fig. 3.7), indicating its folded nature.

Desmoplakin linker and desmin binding HSQC experiments were performed with a ratio 1:2 (in favour of desmin). The spectra showed all the original peaks of the linker domain except for three, the peaks corresponding to residues K2542, L2554 and D2559 (equivalent to residues 94, 106 and 111 in the structure numbering), all of them in the C-terminal end of the linker domain (Fig.4.7). The average loss of intensity was 15.94%, demonstrating an extremely weak interaction between the desmoplakin linker domain and this intermediate filament protein.

The same experiment was repeated with vimentin. The HSQC spectrum shows similar results as for desmin, with no loss of signal for any peaks with the exception of those

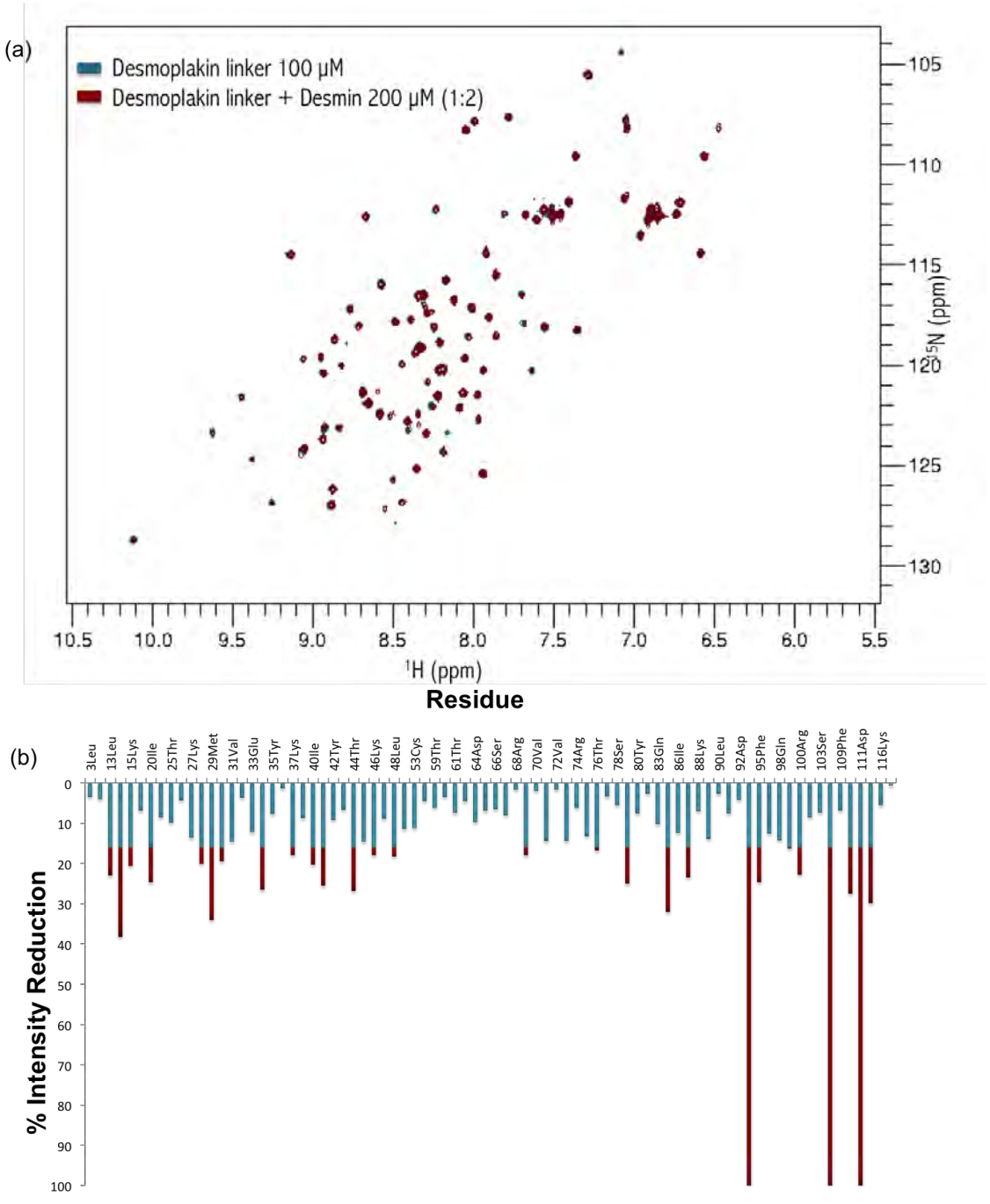


FIGURE 4.7: (a) Overlay of the HSQC spectrum of desmoplakin linker domain with and without desmin. In teal: HSQC spectrum of desmoplakin linker at 100 μM in 20 mM HEPES buffer. In magenta: overlapped HSQC spectrum of desmoplakin linker domain with desmin at 200 μM. (b) Histogram plotting the percentage of loss of intensity (in volume) of peak signal for each amino acid of the DPlink in the presence of desmin. Magenta sections of the columns indicate residues that have lost a higher level of intensity than the average of 15.94%.

corresponding to residues G2454 and T2555 (equivalent to residues 6 and 107 in the structure numbering) (Fig. 4.8). A high reduction in intensity is seen for residues S2456, R2464 and D2470 (equivalent to residues numbers 8, 16 and 22 in the structure), all of them are found towards the N-terminal end of the linker domain (Fig. 4.19). The histogram of intensity loss corresponding to the HSQC of DPlink in presence of vimentin reveals an average reduction in intensity of 34.89% compared to the histogram corresponding to DPlink in the presence of desmin (Figures 4.19 and 4.20).

This reduction, twice the reduction demonstrated for desmin, suggests that the desmoplakin linker domain is more prone to interact with vimentin than with desmin, although both exhibit weak or marginal apparent affinities. The evidence of weak interaction between the DPlink and vimentin is in agreement with the results published by Fontao *et al.*, who found co-localisation of the isolated linker domain with vimentin in cell transfection experiments [170]. This research group has classified the linker domain of desmoplakin as a major IF binding site and the PRD C as a minor binding site. However, the results here presented demonstrate a stronger interaction of the IF with the PRD C than with the linker as individual domains (comparing 15.94% reduction in the linker spectrum against 63.3% in the PRD C spectrum for desmin and 34.89% reduction in the linker spectrum against 85.21% in the PRD C spectrum for vimentin). Placing this evidence in the context of the multiple experiments performed on binding between desmoplakin and the IF type III, it is possible to say that binding activity of individual domains does not fully or accurately explain the interaction between desmoplakin and vimentin and desmin. Consequently, although individual association of the PRD C with vimentin and desmin has been proved stronger than individual association of the linker domain with the same IF proteins, this does not contradict the results by Fontao *et al.* [170].

Following the hypothesis that the linker domain is mediator of IF binding, it is important to note the discrete number of DPlink amino acids that exhibit chemical shift peaks broadening in the presence of desmin and vimentin. In the case of desmin, only one of the DPlink amino acids that undergo line broadening, K2542, is positively charged, another, D2559, is negatively charged, and the last one, L2554, is neutral but with hydrophobic side chain. This variety of characteristics reflects the complexity of the interaction between desmin and desmoplakin, which involves polar and non-polar interactions [94]. As for vimentin, a weak interaction can be established from the NMR results. The

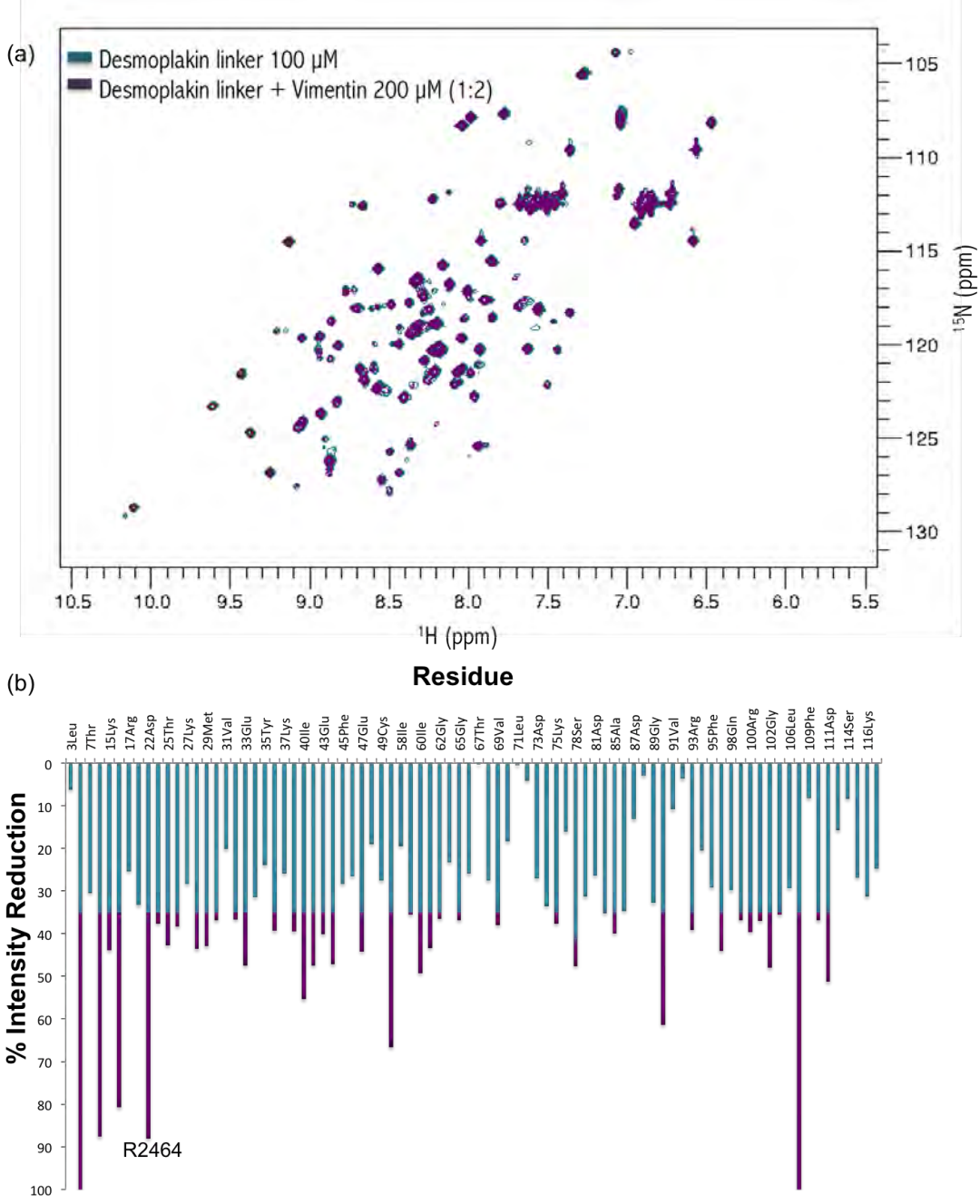


FIGURE 4.8: (a) Overlay of the HSQC spectrum of desmoplakin linker domain with and without vimentin. In teal: HSQC spectrum of desmoplakin linker at 100 μM in 20 mM HEPES buffer. In purple: overlapped HSQC spectrum of desmoplakin linker domain with vimentin at 200 μM. (b) Histogram plotting the percentage of loss of intensity (in volume) of peak signal for each amino acid of the DPlink in the presence of vimentin. Purple sections of the columns indicate residues that have lost a higher level of intensity than the average of 34.89%. Residue R2464 presents itself as a strong candidate for being the main promoter of weak vimentin binding.

residues involved in such interaction are of polar and non-polar nature, similarly to the case of titrations with desmin. G2454, S2456 and T2555 are neutral amino acids, R2464 is positively charged, while D2470 is negatively charged. Faced with this diversity of residues implicated in binding, it would be possible that some of these residues exhibit reduced intensity due to different factors other than direct interaction with vimentin. In particular, the peak corresponding to T2555 is weak in the HSQC acquired from desmoplakin linker wild type (DPlink-WT) alone, possibly explaining its disappearance in a slightly more viscous environment, i.e. upon addition of vimentin to the sample.

Keeping that in mind, we can still address the salient occurrence that none of the involved residues mentioned above are conserved within the plakin repeats of desmoplakin PRDs (Fig. 4.4), denoting the exceptional character of the N-terminal end of the linker with respect the PRDs. The perturbed residues instead map to the highly flexible and divergent terminal sequence, as is evident from the results described in Section 3.2.5. Belonging to this region of DPlink, a strong candidate for being the principal promoter of the weak interaction found is residue R2464, given its basic nature and conserved position.

4.4 Interaction of Vimentin and Periplakin Linker Domain Explored with NMR

The periplakin linker was chosen for further vimentin titration experiments due to its position in the C-terminal end of the protein, resembling the position of the PRD C in desmoplakin, and therefore being a potential binding agent for IF.

The HSQC spectrum in orange corresponding to PPlink alone in Figure 4.9 shows the folded nature of this protein. The overlapped spectrum in purple in the same figure is the HSQC spectrum of PPlink acquired in the presence of vimentin. A reduction of 59% in the number of the peaks appearing in the spectra and significant intensity reduction of 93% in the peaks underscores the clear binding activity between the PPlink domain and vimentin (Fig. 4.9). The average reduction of intensity is 85.45% (Fig. 4.31).

Although the desmoplakin linker is more conserved through different species than periplakin linker (Fig. 4.12), the latter has been shown to play an individual role as a binding region

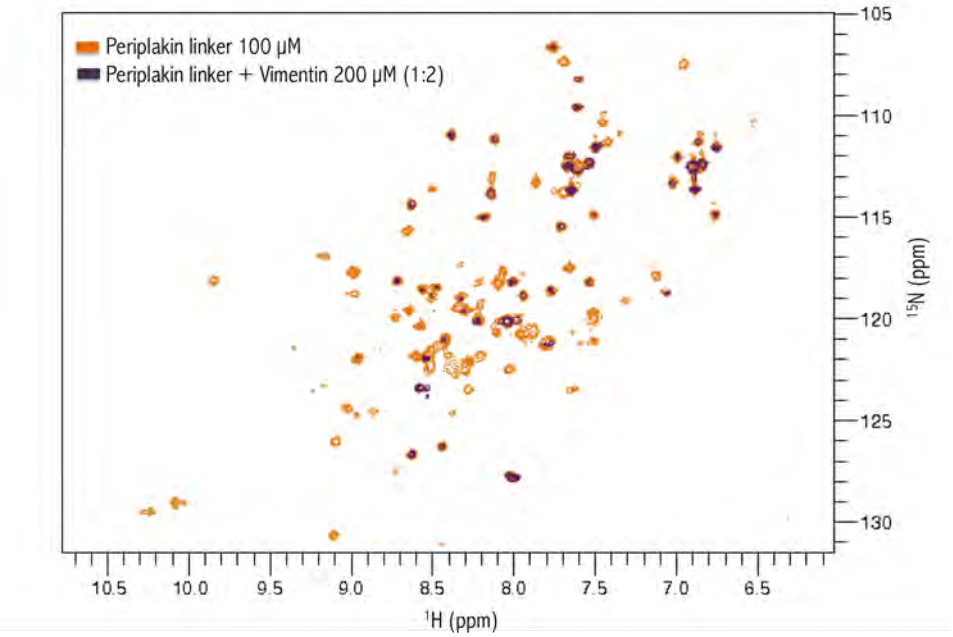


FIGURE 4.9: Overlay of the HSQC spectrum of periplakin linker domain with and without vimentin. In orange: HSQC spectrum of periplakin linker at 100 μM in 20 mM Tris buffer. In purple: HSQC spectrum of periplakin linker domain with vimentin at 200 μM.

for vimentin. The C-terminal end of periplakin is comprised of the linker domain only, as mentioned in Section 1.5.1. Since the C terminus of plakins mediate association with IFs, this structural feature provides reasonable explanation why all the periplakin IF binding properties rely on its linker domain. The similar percentages of intensity reduction in presence of an excess of vimentin for periplakin linker (85.45%) and desmoplakin PRD C (85.21%), as shown in Figures 4.9 and 4.6 respectively, suggest the importance of C terminus ends of plakins in relation with IF binding.

4.5 Interaction of Vimentin and Desmoplakin Linker Domain Mutants

The desmoplakin linker domain alone does not bind desmin and has a weak interaction with vimentin IF. It also has been demonstrated that the periplakin linker domain does bind vimentin with an average reduction of intensity of the 85.45% under the same conditions in which desmoplakin linker domain shows only 34.89% intensity reduction. Given these results it is possible to conclude that the periplakin linker domain possesses certain properties that provide it with the ability to bind vimentin. Therefore, based on

the following criteria, we aimed to find the residues of the sequence that are related to or confer such properties.

4.5.1 Criteria Considered to Assess Amino Acids Involved in Vimentin Binding

The following points were considered in order to select suitable mutations for desmoplakin and periplakin linker domains.

I. Sequence alignment of Desmoplakin and Periplakin linker domains

Given the high homology between the linker domains of periplakin and desmoplakin it is acceptable to hypothesise that the binding properties of periplakin linker reside on those amino acids that differ from desmoplakin linker. On that basis, a sequence alignment of the desmoplakin linker domain and the periplakin linker domain was carried out with the purpose of finding key residues that differ between both proteins. First we looked for binding candidates in periplakin linker, i.e. basic amino acids such as arginine and lysine found together within 3 residues creating a patch that could recognise acidic residues in the electronegative patch found in the alpha helix of vimentin rod domain 2B [178] (Fig. 4.10).

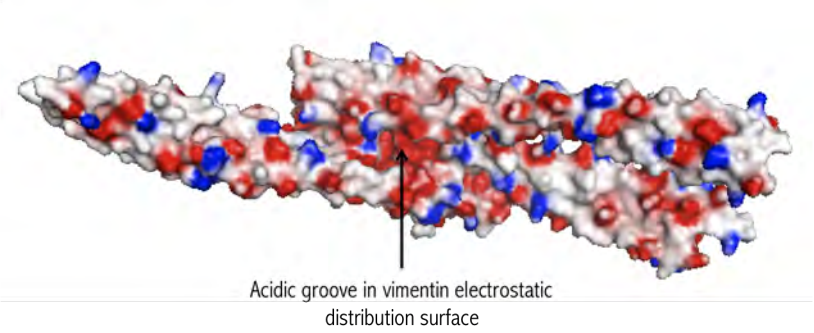


FIGURE 4.10: Electrostatic surface distribution of tetrameric vimentin (segment 2B of the rod domain with PDB entry 1GK4 [178]) processed using Delphi. The intercalation of positive (blue) and negative (red) residues is explicit, with a visible groove of acidic residues at the centre of the structure. The scale is -7.0 to 7.0.

Then we looked for the equivalent residues in desmoplakin linker that were not basic amino acids. In this way we identified two pairs of residues for mutation in each linker protein. There were periplakin linker residues K1687/R1689 and K1717/K1718, and the equivalent residues in the desmoplakin linker E2495/E2497 and S2526/Q2527 (Fig. 4.11).

II. Desmoplakin secondary structure potential binding sites

Based on the high homology of their primary structure, a similar secondary structure for the two linker domains is feasible. On this premise, exposed residues of the desmoplakin linker domain structure were taken into consideration, as well as their periplakin linker equivalents according to the sequence alignment, as potential binding sites for vimentin. Considering that flexible regions often play an important role in binding, special attention was paid to residues towards the N-terminal end of the protein since the NMR structure of desmoplakin linker showed a high degree of flexibility there. Taking point *(I)* into consideration we proposed one more pair of residues for mutation in periplakin linker: residues R1655/R1656, which are equivalent to residues R2464/R2465 in the desmoplakin linker (Fig. 4.11).

III. Plectin binding sites and homologous boxes among plakins

Previous experimental results have suggested specific groups of amino acids in plectin are responsible for vimentin binding, as described in Section 4.1.

Exploiting the high homology between plectin and the linker domains of desmoplakin and periplakin (70% identity) [170], a sequence alignment of all the mentioned proteins was performed (Fig. 4.2) with the aim of identifying the amino acids in the linkers that correspond to those in plectin that are involved in binding vimentin (the basic amino acid residue cluster RLRR observed by Nikolic *et al.*). On the other hand, the existence of two homologous boxes within the linkers has been identified to be directly involved in plakins association with vimentin, as explained in Section 4.1.1 (Fig. 4.2), so here we have considered the mutations to be included in such regions (Fig. 4.11).

IV. Amino acids conservation through species

High conservation of basic amino acids through species, paying special attention to those ones which are together within 3 residues (Fig. 4.12).

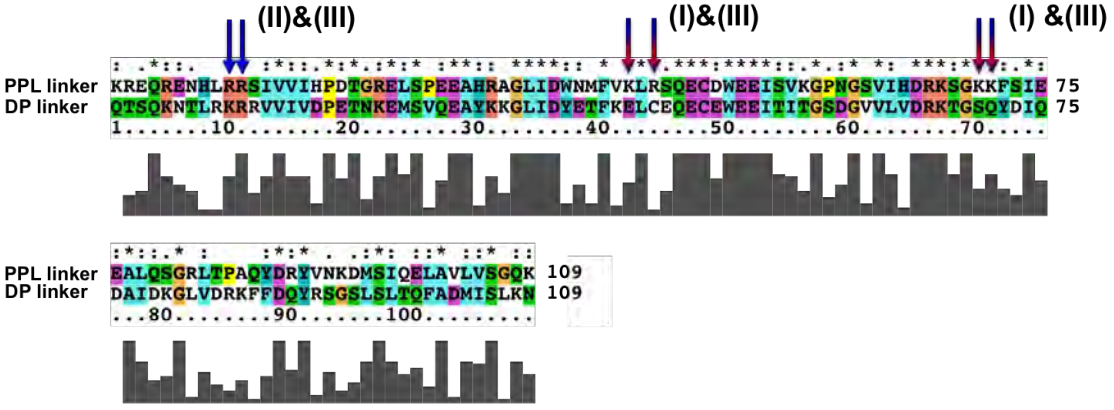


FIGURE 4.11: Sequence alignment of periplakin and desmoplakin linker domains done using Crysol. The blue/red arrows indicate the basic amino acids in periplakin linker that could be binding sites for vimentin and the equivalent non-basic residues in desmoplakin linker (criteria point *I*). These positions also lie within the homologous boxes indicated in criteria point *III*. The blue arrows indicate the exposed basic amino acids in desmoplakin linker that form part of the flexible region in the N-terminal end of the domain and that also correspond to basic amino acids (together within 3 residues) in periplakin linker (criteria point *II*). Also, these residues are the central two residues of the highly positive charged amino acid cluster in plectin indicated in criteria point *III*.

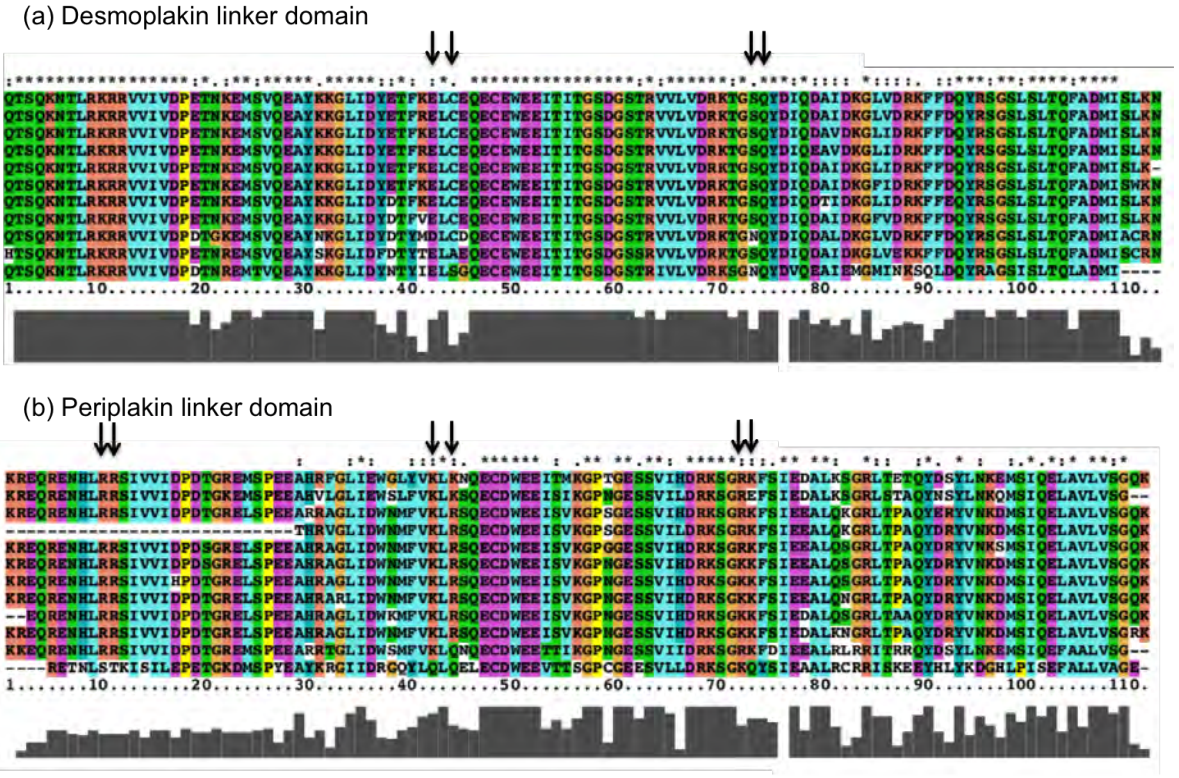


FIGURE 4.12: Alignment of the desmoplakin an periplakin linker domains through different animal species showing the mutations carried out in each protein. (a) Alignment of the DPlink domain through different animal species showing high conservation. (b) Alignment of the PPlink domains through different animal species. Species are Homo Sapiens, Equus Caballus, Anolis Carolinensis, Felis Catus, Canis Lupus, Sus Scrofa, Bos Taurus, Monodelphis Domestica, Ailuropoda Melanoleuca and Gallus Gallus.

As result of the combination of the listed criteria, two double mutants for desmoplakin linker and three double mutants for periplakin linker were produced in order to introduce IF binding in DPlink and to compromise IF binding in PPlink. The mutations selected were the following. Based on criteria points *(I)* and *(III)*: Desmoplakin, residues E2495 and C2497 were mutated to a lysine and arginine respectively, producing DPlink double mutant E2495K-C2497R; desmoplakin, residues S2526 and Q2527 were both mutated to lysines, producing DPlink double mutant S2526K-Q2527K; periplakin residues K1687 and R1689 (aligned to DPlink residues E2495 and C2497) were mutated both to glutamic acids, producing PPlink double mutant K1687E-R1689E; periplakin residues K1717 and K1718 (aligned to DPlink residues S2526 and Q2527) were mutated both to glutamic acids, producing periplakin linker double mutant K1717E-K1718E. Finally, following criteria points *(II)* and *(III)*: Periplakin residues R1655 and R1656 were both mutated to glutamic acids, producing periplakin linker double mutant R1655E-R1656E (Fig. 4.13 and 4.14).

(a)

DPlink structure (NMR)	Desmoplakin sequence	PPlink structure (I-TASSER)	Periplakin sequence	Desmoplakin PRD B	Desmoplakin PRD C
K15	K2463	R10	**R1655**		
R16	**R2464**	R11	**R1656**		
E47	**E2495**	K42	**K1687**	R2366	R2725
C49	**C2497**	R44	**R1689**		
S78	**S2526**	K72	**K1717**		
Q79	**Q2527**	K73	**K1718**	R2309 R2385	K2668 R2744

(b)

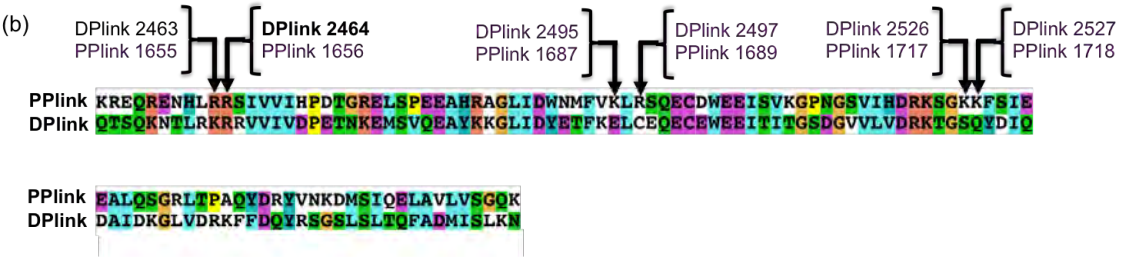


FIGURE 4.13: (a)Table of the desmoplakin and periplakin linkers residues chosen for mutation (in bold purple). Equivalent residues in electrostatic patches of desmoplakin PRDs B and C [94] (in yellow and green respectively), according to sequence alignment with plakin repeats sub-structures as shown in Figure 4.4. The column *DPlink structure (NMR)* refers to the corresponding numbering in the protein structure generated by NMR assignment and presented in Chapter 3. The column *PPlink structure (I-TASSER)* indicates the numbering according to the model created by I-TASSER. (b) Sequence alignment of the desmoplakin and periplakin linker domains indicating the residues chosen from mutations of table in (a).

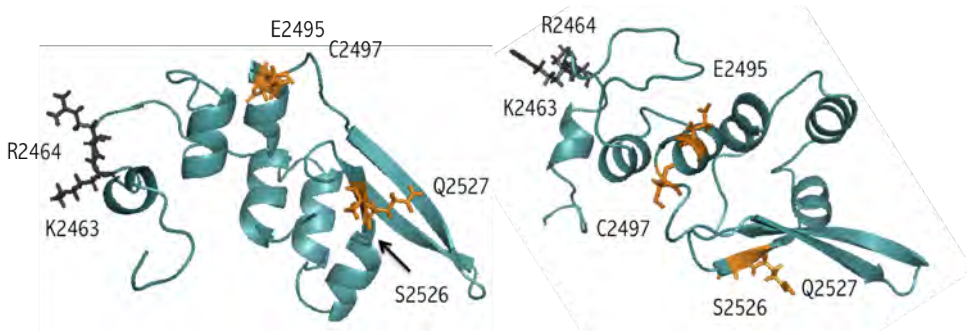


FIGURE 4.14: Two views of desmoplakin linker domain structure by NMR highlighting in orange the mutated residues (by charge reversal to basic amino acids) and in black the residues potentially responsible for weak binding between DPlink and vimentin. All highlighted residues have been also mutated in periplakin linker.

4.5.2 Desmoplakin Linker Domain Mutants

The first desmoplakin linker mutant produced to increase its ability to bind vimentin was the DPlink double mutant E2495K-C2497R. Its HSQC spectrum resulted well-dispersed (Fig. 4.15, spectrum in light-blue) showing that the mutant protein is correctly folded. An HSQC spectrum was then acquired in the presence of vimentin (Fig. 4.15 spectrum in purple).

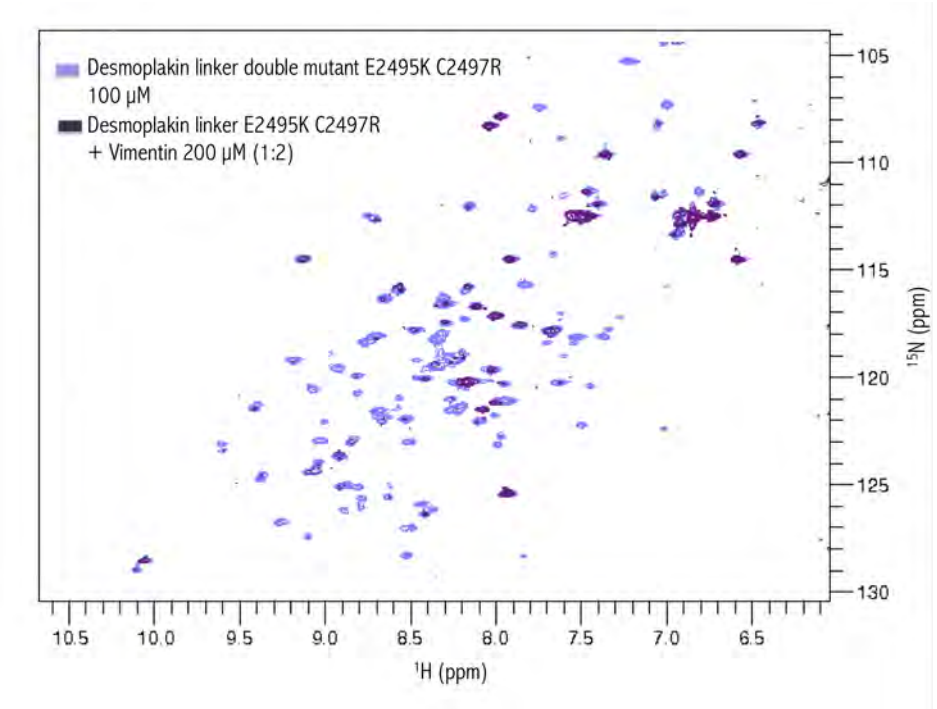


FIGURE 4.15: Overlay of the HSQC spectrum of desmoplakin linker domain double mutant E2495K-C2497K with and without vimentin. In light-blue: HSQC spectrum of DPlink mutant E2495K-C2497K at 100 μM in 20 mM Tris buffer. The mutation results in a folded protein. In purple: HSQC spectrum of DPlink mutant E2495K-C2497K with vimentin at 200 μM.

An intensity reduction of 96.49% from the original spectra and a complete disappearance of 84.52% of the peaks demonstrate binding activity between the DPlink mutant E2495K-C2497R linker domain and vimentin. By superimposing the spectra of desmoplakin linker domain wild type and the mutant E2495K-C2497K (Fig. 4.16) it is possible to observe chemical shift changes in 86% of the peaks in the ^{15}N and ^{1}H dimensions.

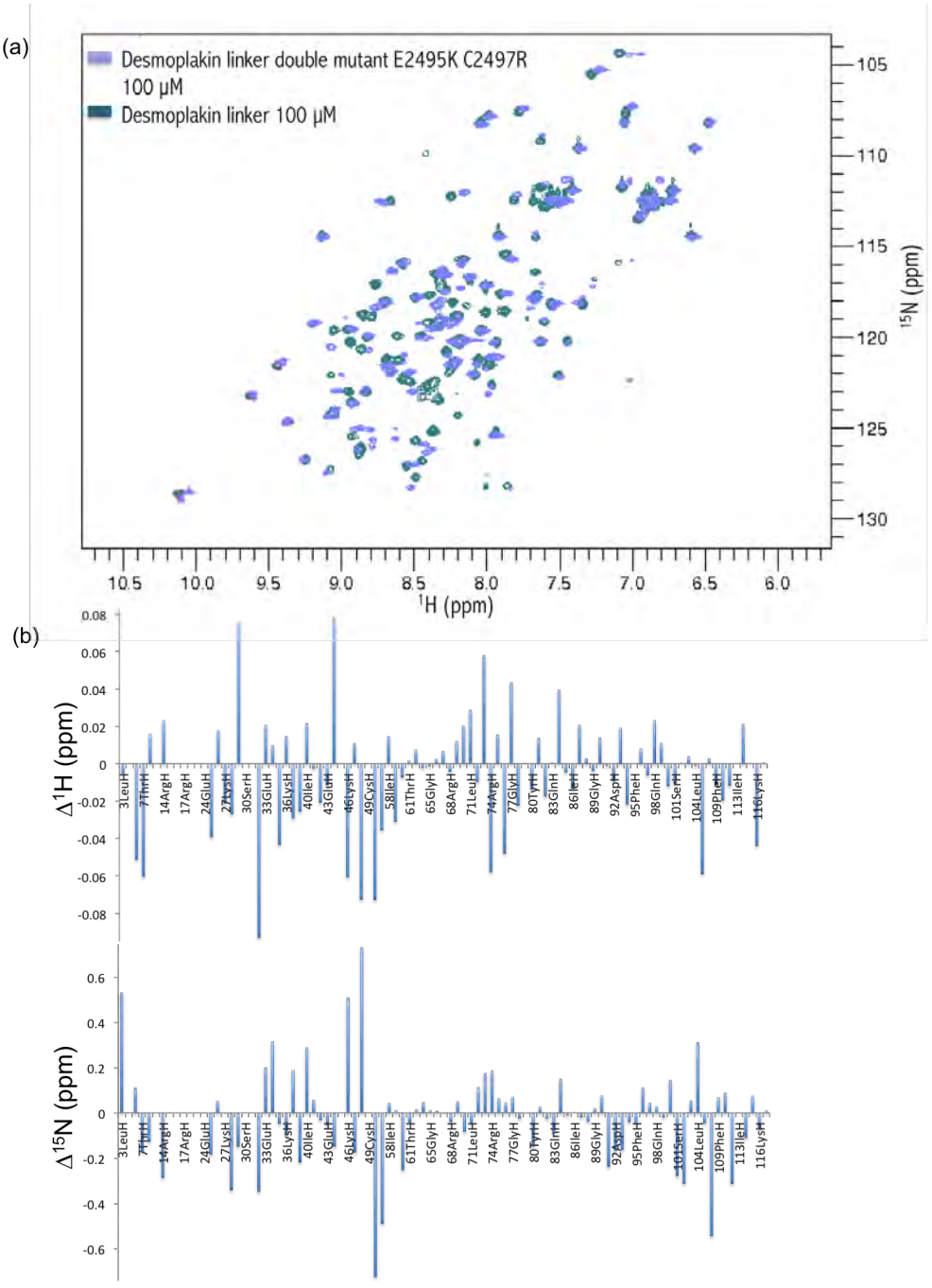


FIGURE 4.16: (a) Overlay of the HSQC spectrum of DPlink mutant E2495K-C2497K (in light-blue) and DPlink-WT (in teal), both at 100 μM in 20 mM Tris buffer. (b) Histograms showing the difference in chemical shifts between desmoplakin linker wild type and desmoplakin linker mutant E2495K-C2497K.

The explanation of such chemical shift modifications as consequence of the mutation could reside in the following: glutamic acid residues are negatively charged, whereas lysines are positively charged amino acids. When mutating residue E2495 into a lysine, the electrostatic environment is locally modified: lysines, due to their basic nature, favour interaction with acidic amino acids, contrary to glutamic acids [179]. Chemical shift changes in the ^{1}H dimension seem to be distributed along the sequence, whereas in the ^{15}N dimension the changes seem to be concentrated around the mutated amino acids section.

The second mutant produced for the desmoplakin linker domain was the double mutant S2526K-Q2527K. The DPlink double mutant E2495K-C2497R was also folded, as confirmed by the dispersion of peaks shown in the cyan spectrum of Figure 4.17. An HSQC spectrum of the S2526K-Q2527K linker mutant was acquired in the presence of vimentin in a ratio of 1:2 in favour of vimentin (purple spectrum in fig. 4.17), resulting in the disappearance of only two peaks and a generalised intensity reduction of 64.06%.

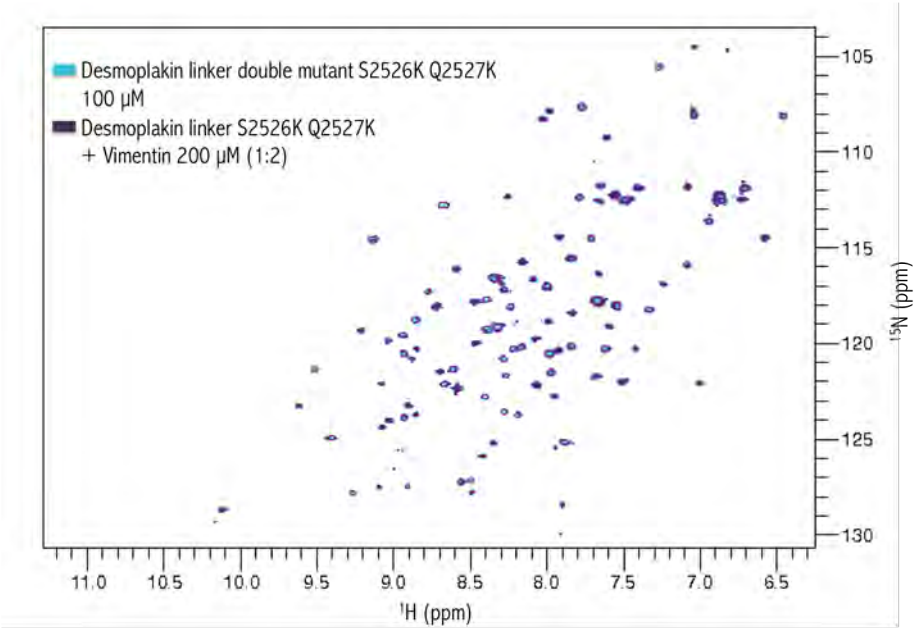


FIGURE 4.17: (a) Overlay of the HSQC spectrum of desmoplakin linker domain double mutant S2526K-Q2527K with and without vimentin. In cyan: HSQC spectrum of DPlink mutant S2526K-Q2527K at 100 μM in 20 mM Tris buffer. In purple: HSQC spectrum of DPlink mutant S2526K-Q2527K with vimentin at 200 μM.

One of the residues that showed 100% intensity loss is R2464, the residue proposed to be the one responsible for the weak binding to vimentin of the linker in isolation. The other residue is D2512, a negatively charged amino acid.

The comparison of the S2526K-Q2527K linker mutant with the wild type of desmoplakin shows considerably lower changes in chemical shifts in contrast with the E2495K-C2497R mutant (Fig. 4.18), reflecting a minor change in the protein structure and faculties.

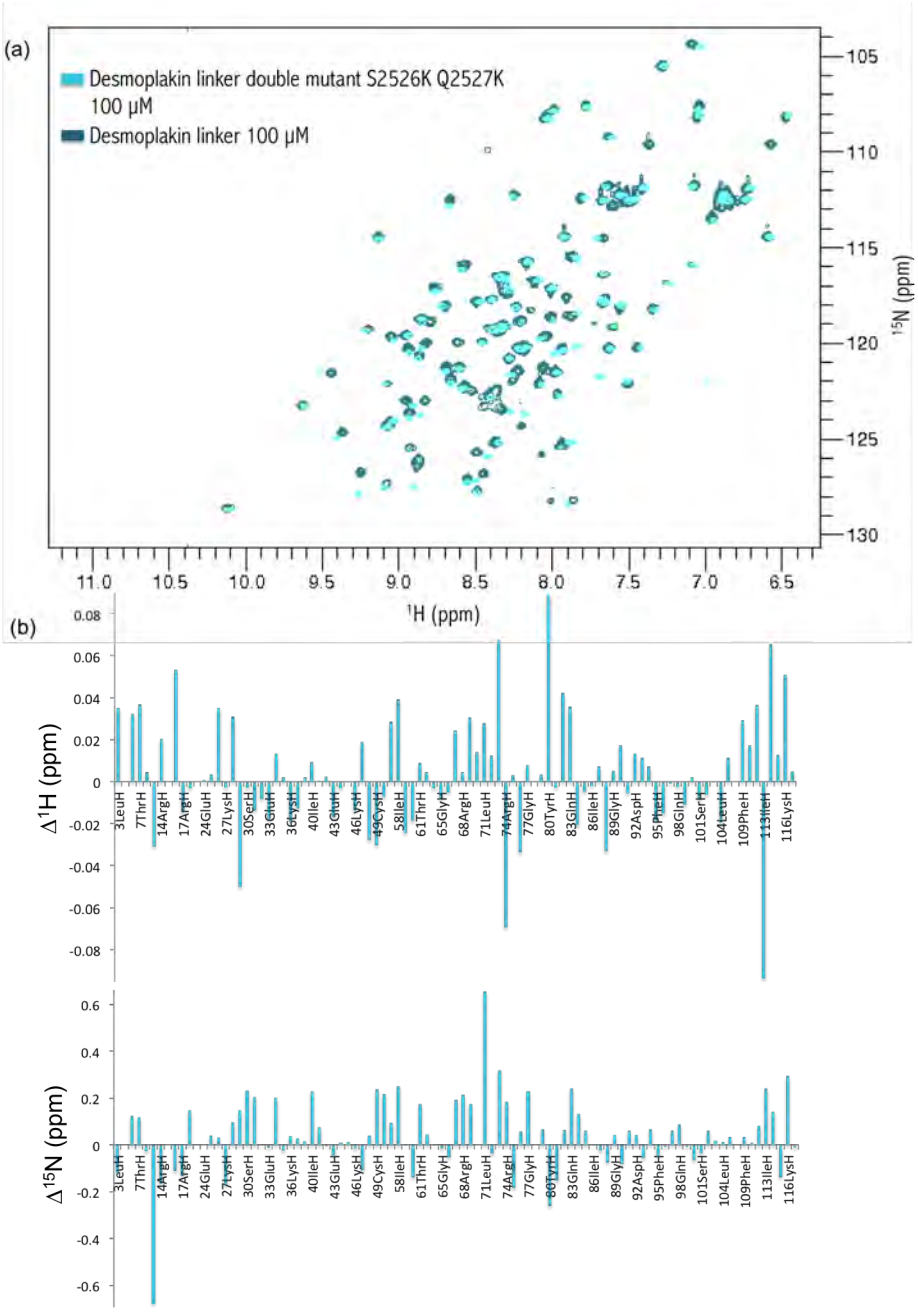


FIGURE 4.18: (a) Overlay of the HSQC spectrum of DPlink mutant S2526K-Q2527K (in cyan) and DPlink-WT (in teal), both at 100 μM in 20 mM Tris buffer. A comparison of both spectra shows less chemical shift changes than the resulting from the E2495K-C2497K mutation.(b) Histograms showing the difference in chemical shifts between desmoplakin linker wild type and desmoplakin linker mutant S2526K-Q2527K.

Identification of sections within the protein that could have been more affected by the mutation is difficult to do given the homogeneous distribution of chemical shift changes

showed in both dimensions of the spectrum.

Figure 4.19 shows the histograms of intensity reduction obtained for DPlink wild type, DPlink mutant E2495K-C2497R and DPlink mutant S2526K-Q2527K aligned so it is easy to identify the same assigned residue in each histogram. Figure 4.20 shows the averages percentages of loss of intensity in each case.

The residues chosen for mutation in desmoplakin linker are of negative or neutral electrostatic nature, and they have been chosen partly because their periplakin linker counterparts are basic amino acids. It is important to point out that one of each of the residues that have been replaced corresponds to a basic amino acid in the alignment with the plakin repeat sub-domains that constitute desmoplakin PRDs B and C (Fig. 4.13 (a)).

Charge reversal of two central amino acids, E2495 and C2497, has a strong consequence in desmoplakin linker structural environment, and these modifications have clearly enhanced its ability to interact with vimentin. More specifically, the introduced basic amino acids seem to have recognised vimentin acidic residues, engaging in this way the complete IF. By identifying these key residues, it is possible to also declare periplakin residues K1687 and R1689 as essential for the binding between this protein's linker domain and vimentin, a conclusion that is supported by the alignment argument described in Section 4.5.1 point *(I)*.

For desmoplakin wild type the peak corresponding to residue E2495 showed a loss intensity of 44%, while the peak correlated to C2497 lost only 27.5% of its intensity due to interaction with vimentin. In the case of desmoplakin double mutant E2495K-C2497R, the peaks corresponding to residues K2495 and R2497 showed a 100% loss of intensity. The results obtained for DPlink mutant S2526K-Q2527K were less drastic, which is reflected in the similarity between the wild type and mutant HSQC spectra shown in Figure 4.18. However, the stronger interaction between the PD mutant and vimentin compared to the wild type case (reflected in 30% more line broadening of the spectrum peaks), suggests that the introduction of basic amino acids in this positions of desmoplakin sequence have also had important repercussions in the protein structure and therefore capacity to interact with vimentin.

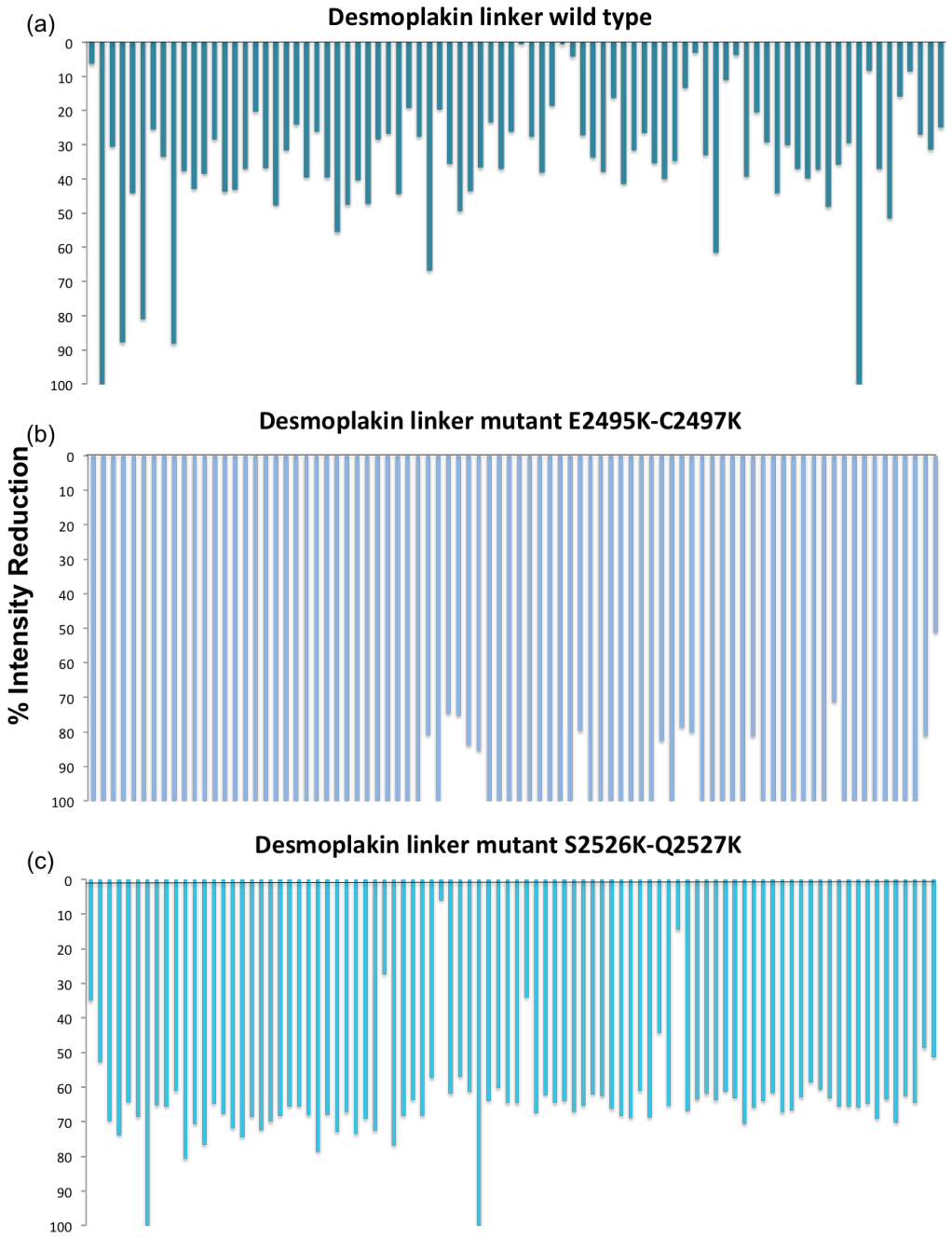


FIGURE 4.19: Histograms showing a comparison of the percentages of intensity reduction (in volume) that were observed in the different desmoplakin linker proteins (all at a concentration of 100μM) in presence of vimentin (at a concentration of 200μM). (a) In teal: desmoplakin linker wild type; (b) In light-blue: desmoplakin linker mutant E2495K-C2497R; (c) In cyan: desmoplakin linker mutant S2526K-Q2527K.

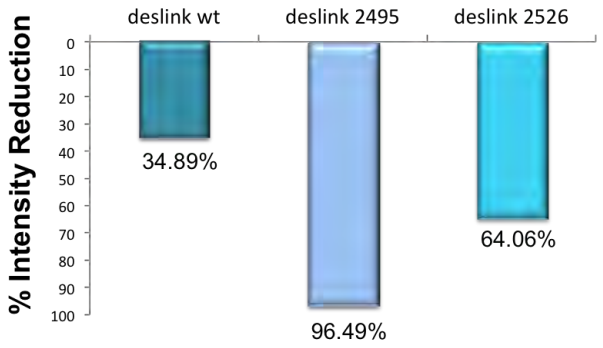


FIGURE 4.20: Histogram showing a comparison of the average percentage intensity reduction (in volume) that was observed for different desmoplakin linker proteins (all at a concentration of 100μM) in presence of vimentin (at a concentration of 200μM). In teal: desmoplakin linker wild type; in light-blue: desmoplakin linker mutant E2495K-C2497R; in cyan: desmoplakin linker mutant S2526K-Q2527K. The charge reversal mutation performed to DPlink residues S2526 and Q2527 increased its interaction with vimentin to slightly less than double the activity for the DPlink WT. In the case of residues E2495 and C2497, the charge reversal mutation resulted in almost three times increase in activity with respect to the DPlink WT, showing a higher intensity reduction that PPlink WT of 85.45%.

Based on the sequence alignment between desmoplakin linker domain and domains B and C of Figure 4.4, we could infer that desmoplakin residues R2366 and R2725, belonging to PRD B and C respectively, are essential amino acids involved in vimentin binding. In the same manner, but playing perhaps a secondary role, desmoplakin residues R2309 and R2385, from PRD B, as well as K2668 and R2744, from PRD C, are also binding elements for vimentin. This would confirm the relevance of the positive groove in PRDs B and C surfaces proposed by Choi *et al.* [94].

Notably, the fact that R2464 appears again as a highly active amino acid with vimentin confirms its role as key mediator of IF engagement for the linker as an independent domain. Moreover, the acidic amino acid D2512 completes the characteristic pair of charged active residues that have been present in the results coming from three different binding experiments: DPlink-WT with desmin, DPlink-WT with vimentin and DPlink mutant S2526K-Q2527K with vimentin (results from the experiment of DPlink mutant E2495K-C2497R with vimentin show too many active residues to distinguish any specificity) (Fig. 4.21).

(a)

Binding experiment	Active charged residues	
	(+)	(-)
DL-WT + Desmin	K2542	D2559
DL-WT + Vimentin	R2464	D2470
DL E2495K-C2497R + Vimentin	R2464 (plus others)	D2470, D2512, D2559 (plus others)
DL S2526K-Q2527K + Vimentin	R2464	D2512

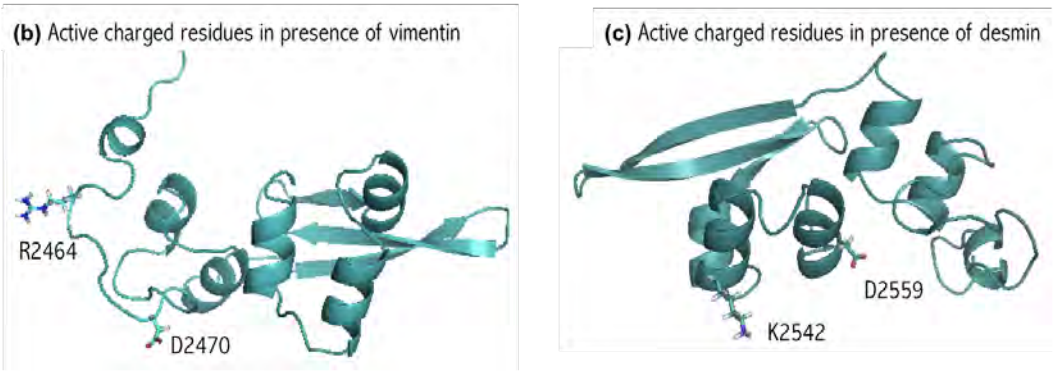


FIGURE 4.21: Constant presence of a pair of positive a negative amino acids as main players in the IF-plakin binding interaction has been detected, suggesting that the association with IF could have an electrostatic character involving positive and negative charges in both, IF and plakin protein. (a) Table showing the charged active residues (residues which chemical shift peaks undergo line broadening or disappear in presence of a ligand) in DPlink-WT in presence of vimentin and desmin and mutants E2495K-C2497R and S2526K-Q2527K in presence of vimentin. (b) Active charged residues in DPlink-WT located in the NMR structure in presence of vimentin. Residues R2464 and D2470 are oriented in the same direction, suggesting that both could act together in binding vimentin. (c) Active charged residues in DPlink-WT located in the NMR structure in presence of desmin. Residues K2542 and D2559, are oriented in the same direction, also suggesting that both could act together in binding desmin.

Amino acids repeatedly shown to be key include a pair of complementary positive and negative residues, suggesting that acidic amino acids in plakins could also lead to association with an alpha helix of vimentin, in concordance with the hypothesis of Choi *et al.* about a shallow negatively charged conserved groove on the surface of desmoplakin PRD C interacting with vimentin [94].

Finally, with the aim of providing with a model that shows the proposed conformation of the complex formed by desmoplakin linker WT and vimentin through weak electrostatic binding, a manual docking of the desmoplakin linker domain structure generated by NMR in this work with vimentin rod domain in its tetrameric state has been performed based on the positive-negative pairs of active residues in presence of vimentin. By positioning the DPlink in close proximity to the acidic grove in vimentin (Fig. 4.10) it is easy to find an electrostatic binding site for residue R2464 (this is with residue

D385 of the *F* repetition of vimentin rod domain 2B) [178], a principal determinant of vimentin binding. Once that binding site was established, a vimentin basic residue of the same repetition (this is K373) fits the position to create an electrostatic bond with reside D2470 of desmoplakin, completing the pair proposed in the table of Figure 4.21 (a). Lastly, desmoplakin residue K2463 resulting position with respect vimentin promotes electrostatic interaction with residue E346 from the *D* repetition of the same rod domain of vimentin (Fig. 4.22).

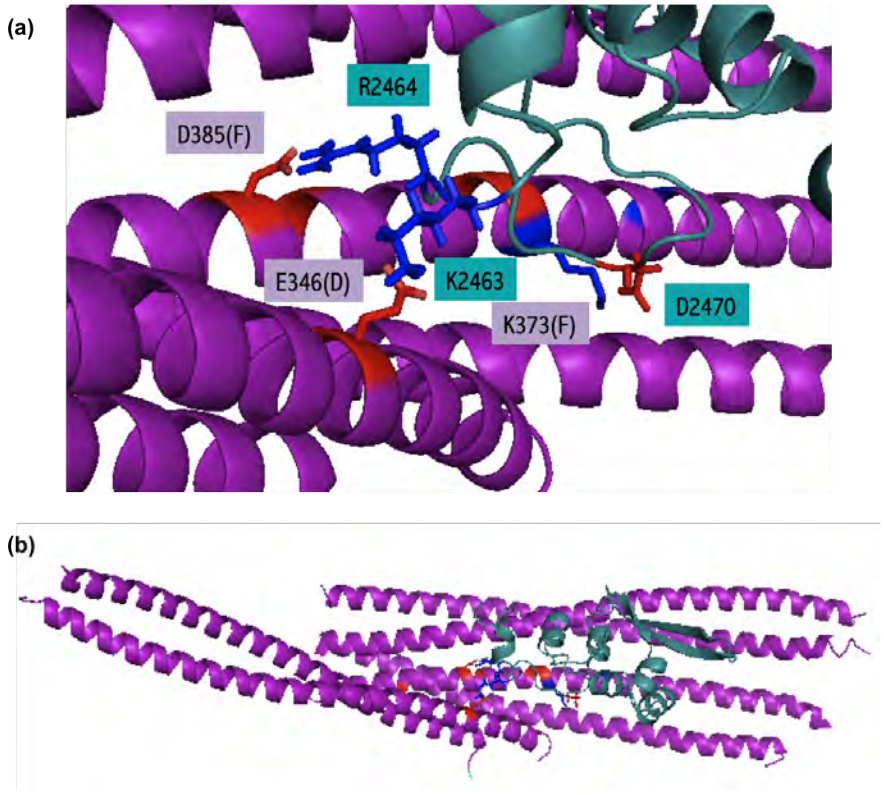


FIGURE 4.22: Manual docking of the desmoplakin linker domain, shown in teal, to vimentin rod domain 2B in its tetrameric state, shown in purple (with PDB entry 1GK4 [178]). (a) A close view to the complex in the proposed binding region. An excellent fit for 3 electrostatic binding sites includes vimentin repetitions D and F and the N-terminal end of desmoplakin. Acidic residues are shown in red and basic residues are shown in blue in both proteins. Interacting residue numbers that belong to DPlink are highlighted in cyan, while vimentin corresponding binding residue numbers are highlighted in purple. This model nominates residues K2463, R2464 and D2470 of DPlink to bind vimentin residues E346 (from the D repetition), D385 and K373 (both from the F repetition) respectively. (b) A complete view of the complex.

Interestingly, the DPlink region suggested to provide the binding sites of vimentin (N-terminal end) does not align with the PR motif structure that forms part of the desmoplakin PRDs, as exhibited by an alignment that highlights the sequence homologies between domains [160].

4.5.3 Periplakin Linker Domain Mutants

In analogy, but aiming the retrieving of binding properties from the linker domain of periplakin, titration experiments were carried out between the mutants of periplakin and vimentin. Periplakin linker double mutant K1687E-R1689E was folded as it is evident from the HSQC spectrum acquired at a concentration of 100 μM, shown in Figure 4.23.

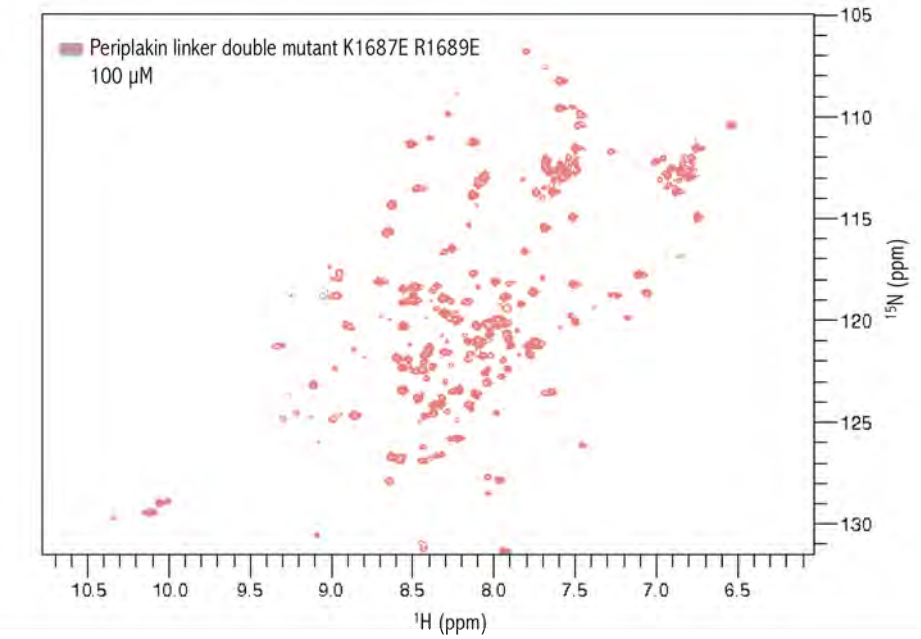


FIGURE 4.23: HSQC spectrum of periplakin linker double mutant K1687E-R1689E at 100 μM in 20 mM Tris buffer. The mutation results in a folded protein.

The addition of 200 μM vimentin to the solution, giving a 1:2 ratio in favour of vimentin, resulted a small effect in the HSQC spectrum, which exhibited almost no broadening of peaks, with an average intensity reduction of 20.76%. Only 6 peaks (out of the 90 original peaks) had % loss of intensity. Therefore, it can be established that no binding activity occurred between the periplakin linker mutant K1687E-R1689E and vimentin (Fig.4.24).

Drastic changes in chemical shifts between periplakin linker wild type (PPlink-WT) and periplakin linker mutant K1687E-R1689E can be seen in the spectra comparison depicted in Figure 4.25. Given the high discrepancy between spectra, chemical shift changes in this case are difficult to resolve, since correspondence between peaks in the WT spectrum are not very easily associated with peaks in the mutant spectrum.

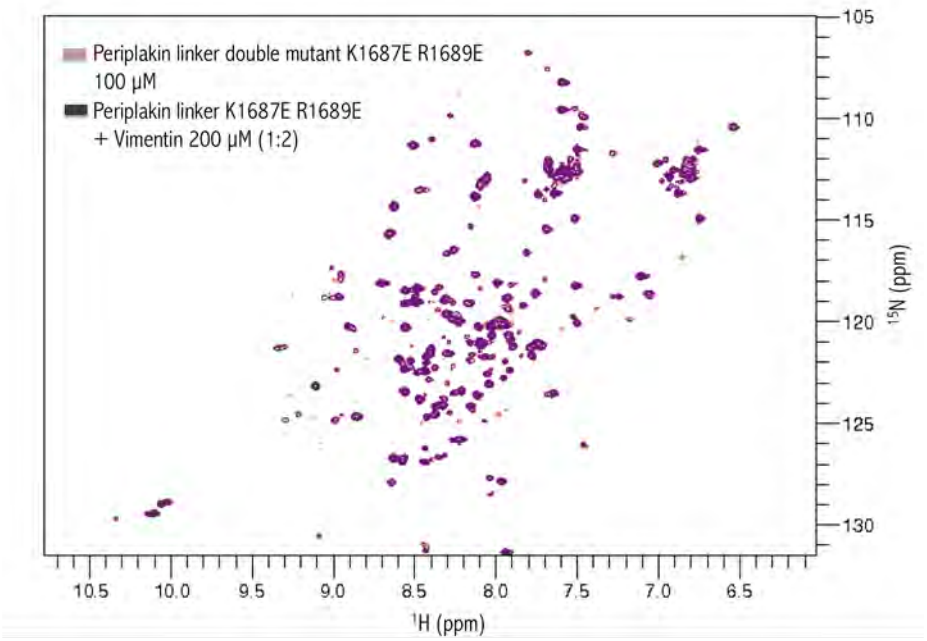


FIGURE 4.24: Overlay of the HSQC spectrum of periplakin linker domain double mutant K1687E-R1689E with and without vimentin. In pink: HSQC spectrum of PPlink mutant K1687E-R1689E at 100 μM in 20 mM Tris buffer. In purple: HSQC spectrum of PPlink mutant K1687E-R1689E and vimentin at 200 μM.

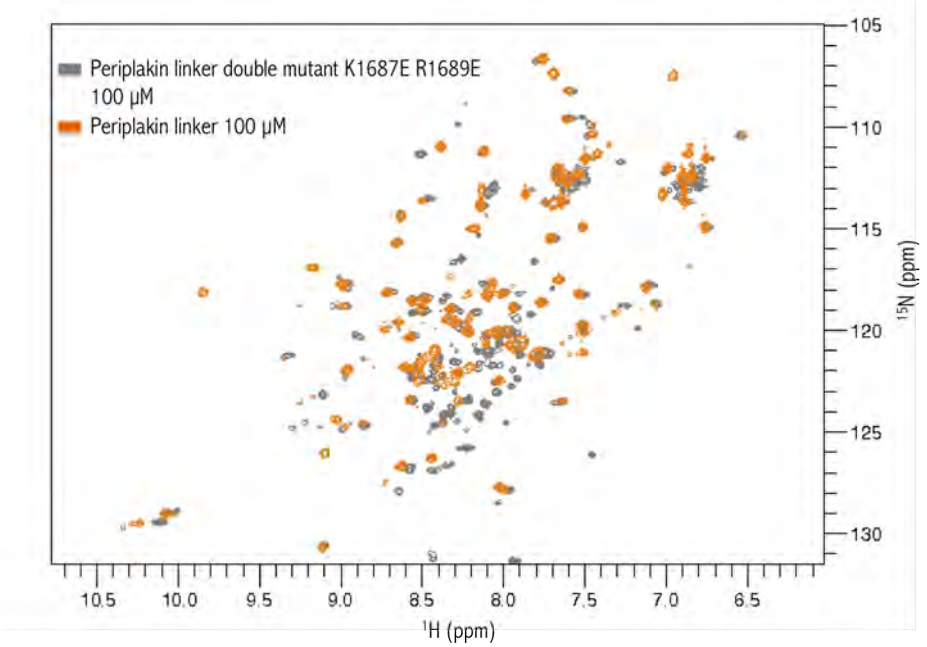


FIGURE 4.25: Overlay of the HSQC spectrum of periplakin linker domain double mutant K1687E-R1689E and periplakin linker wild type. In grey: HSQC spectrum of PPlink mutant K1687E-R1689E at 100 μM in 20 mM Tris buffer. In orange: HSQC spectrum of PPlink wild type at 100 μM in 20 mM Tris buffer. A comparison of both spectra shows drastic changes in chemical shift values for a number of residues result from the mutation.

The second mutant of periplakin planned for titration experiments with vimentin was periplakin linker double mutant K1717E-K1718E. The HSQC spectrum of the mutant protein indicated that the protein was unfolded, with a retraction of peaks to the central area, leaving the lateral regions, corresponding to beta sheet secondary structure, with no peaks. The destabilising effect of this modification could be indicative of a structural key feature contained in residues K1717 and K1718 of periplakin.

Note that the peaks corresponding to the two tryptophan residues of the periplakin linker superimpose in the mutant spectrum, which reflects the same chemical environment for both amino acids in both, opposite to the conditions found for the periplakin linker domain wild type (Fig.4.26).

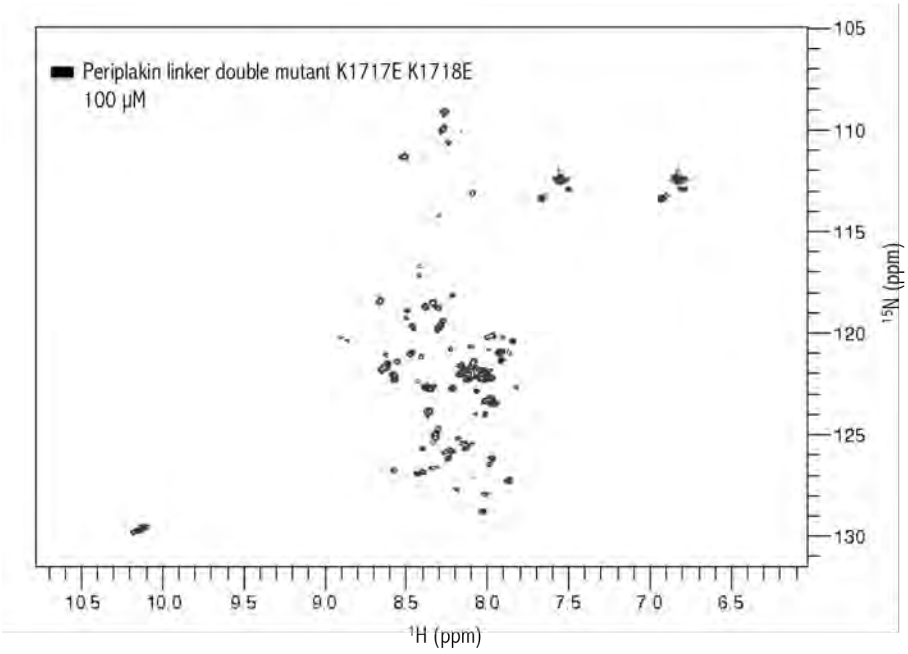


FIGURE 4.26: HSQC spectrum of periplakin linker double mutant K1717E-K1718E at 100 μM in 20 mM Tris buffer. The spectrum corresponds to a typical unfolded protein lacking completely from beta strands.

The last periplakin mutant produced for binding experiments with vimentin was periplakin linker double mutant R1655E-R1656E, which produced folded protein, as evidenced by the HSQC spectrum that shows a well-dispersed distribution of peaks (Fig. 4.27).

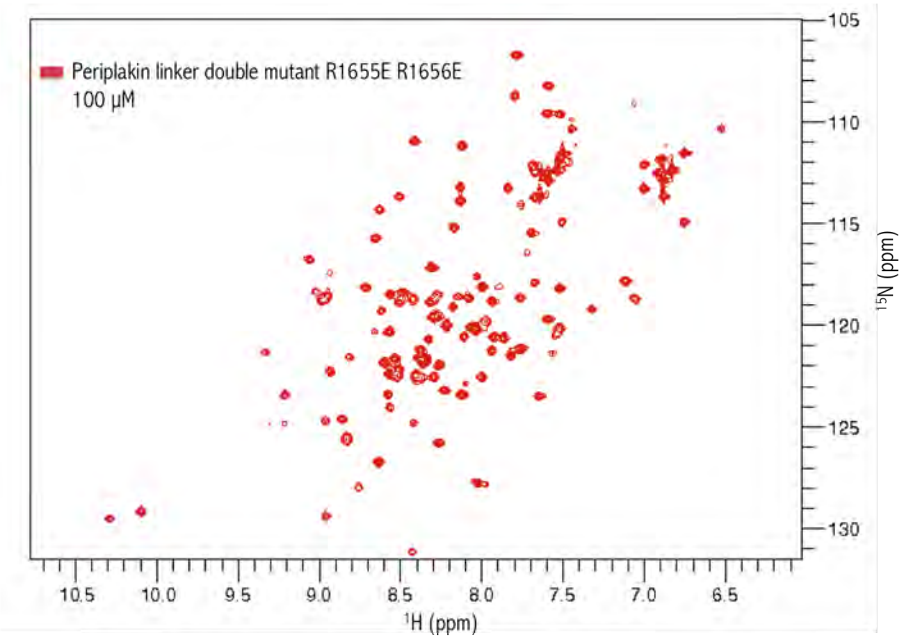


FIGURE 4.27: HSQC spectrum of periplakin linker double mutant R1655E-R1656E at 100 μM in 20 mM Tris buffer. The mutation results in a folded protein.

The overlapped HSQC spectra of periplakin mutant with vimentin at a ratio 1:2 in favour of the latter, demonstrates no binding activity between the two proteins as shown by the retention of all peaks in the spectrum and a negligible reduction of intensity of only 3.89% (Fig. 4.28).

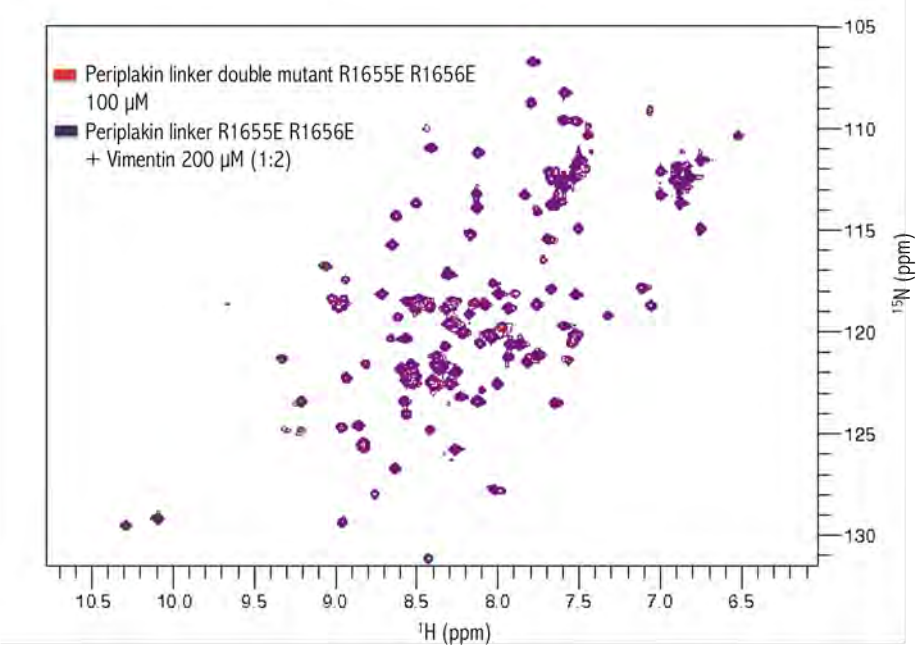


FIGURE 4.28: Overlay of the HSQC spectrum of periplakin linker domain double mutant R1655E-R1656E with and without vimentin. In red: HSQC spectrum of PPlink mutant R1655E-R1656E at 100 μM in 20 mM Tris buffer. In purple: HSQC spectrum of PPlink mutant R1655E-R1656E and vimentin at 200 μM.

As in the case of periplakin mutant K1687E-R1689E, the HSQC spectrum corresponding to periplakin mutant R1655E-R1656E differs considerably from the wild type typical spectrum, revealing a slightly less dramatic change in conformation of the protein due to the mutation (Fig.4.29).

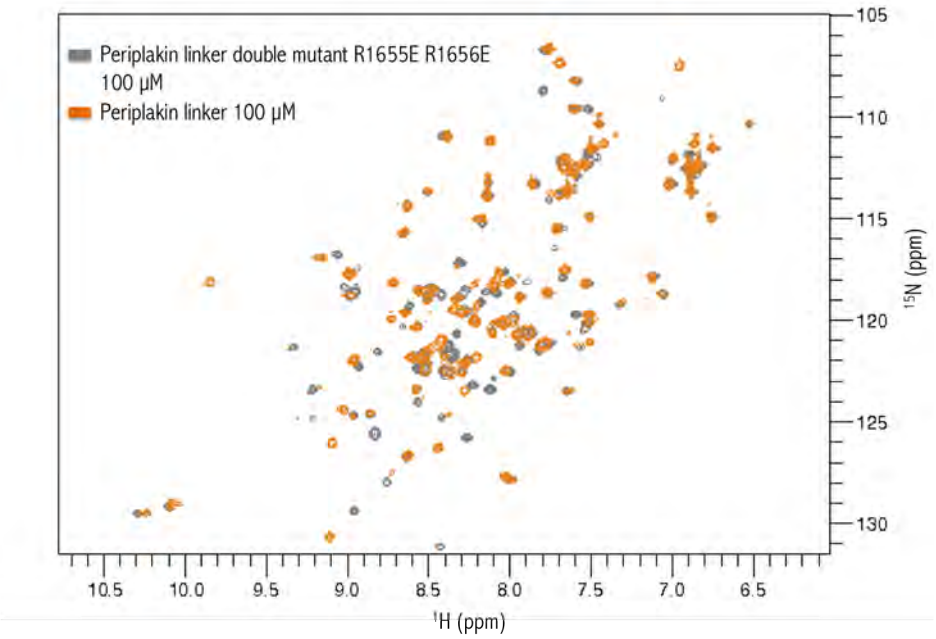


FIGURE 4.29: Overlay of the HSQC spectrum of periplakin linker domain double mutant R1655E-R1656E and periplakin linker wild type. In grey: HSQC spectrum of PPlink mutant R1655E-R1656E at 100 μM in 20 mM Tris buffer. In orange: HSQC spectrum of PPlink wild type at 100 μM in 20 mM Tris buffer. A comparison of both spectra shows changes in chemical shift values for a number of residues result from the mutation.

Although in this case assignment of the peaks of the mutant protein to the wild type peaks is not as difficult to do as in the case of K1687E-R1689E, it is not possible to identify with certainty what regions have been specially affected by the mutation due to the lack of backbone assignment for periplakin linker domain.

Figure 4.30 shows the aligned histograms of intensity reduction obtained for PPlink wild type, PPlink mutant K1687E-R1689E and PPlink mutant R1655E-R1656E so it is easy to identify the same residue in each histogram.

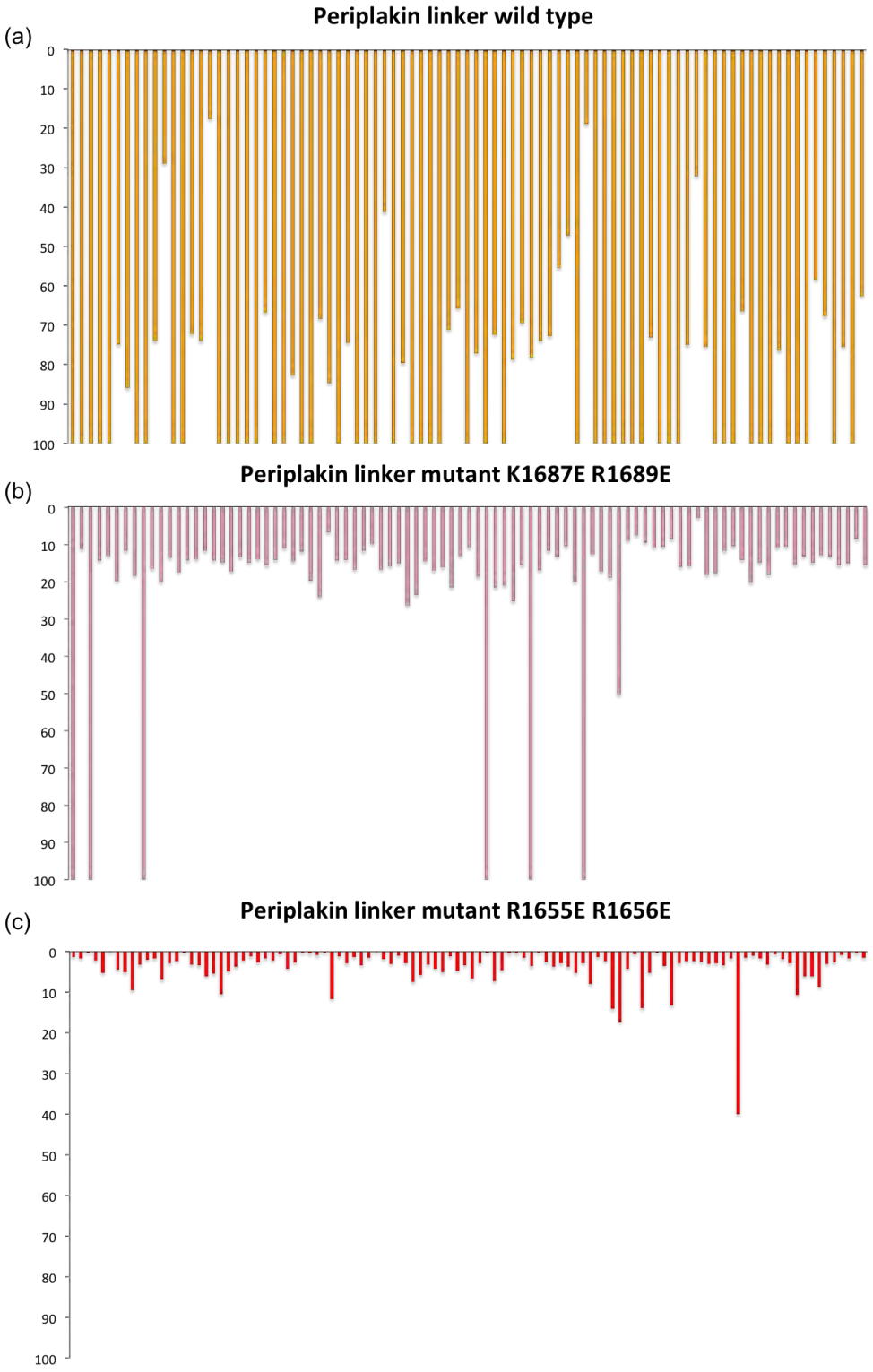


FIGURE 4.30: Histograms showing a comparison of the percentages of intensity reduction (in volume) that were observed in different periplakin linker proteins (all at a concentration of 100μM) in presence of vimentin (at a concentration of 200μM). (a) In orange: periplakin linker wild type; (b) in pink: periplakin linker mutant K1687E-R1689E; (c) in red: periplakin linker mutant R1655E-R1656E.

Backbone data was not collected for periplakin linker domain and therefore no assignment of peaks could be conducted. Due to this limitation it was not possible to determine what residues were active in vimentin binding for any of the samples. In order to be able to compare general intensity variations between samples spectra, an arbitrary numbering has been assigned to the peaks of the periplakin linker domain wild type and the same numbering has been transferred to the mutants. In this way it is possible to align the histograms presented in Figure 4.30. Figure 4.31 shows the averages percentages of loss of intensity in each case.

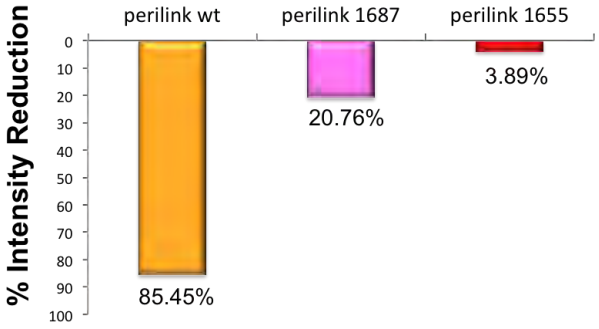


FIGURE 4.31: Histogram showing a comparison of the averages percentages of intensity reduction (in volume) that were observed in different periplakin linker proteins (all at a concentration of 100μM) in presence of vimentin (at a concentration of 200μM). In orange: periplakin linker wild type; in pink: periplakin linker mutant K1687E-R1689E; in red: periplakin linker mutant R1655E-R1656E. The charge reversal mutation performed to PPlink in both pairs of residues inhibited its interaction with vimentin considerably, retrieving it almost completely in the case of residues R1655 and R1656.

The three residues chosen for mutation in periplakin linker are of a positive electrostatic nature. In two of the three cases, PPlink K1687E-R1689E and PPlink R1655E-R1656E, charge reversal in the selected residues resulted, as expected, in reduction and almost lost respectively of the binding activity between the periplakin linker proteins and vimentin.

The modification in the N-terminal end of the protein that led to the production of the PPlink mutant R1655E-R1656E seems to have a more decisive impact on the interaction with vimentin. It should be pointed out the equivalency between these PPlink residues to DPlink residues K2463/R2464 in the alignment between both proteins (Fig. 4.13), which are in the proposed site in DPlink that promotes vimentin binding. It is worth noticing that neither of the two residues involved in this periplakin double mutation (R1655E-R1656E) has an equivalent in desmoplakin PRDs C or B in an alignment with the sub-domain PR structure that constitutes them (Fig. 4.13). In the same way as for desmoplakin linker, outstanding residues are located in the segment of periplakin linker

N-terminal end (Fig. 4.32) that does not resemble the PR sub-structure found in the desmoplakin PRDs according to an alignment that emphasises the similarities between linkers and PRDs sequences [160].

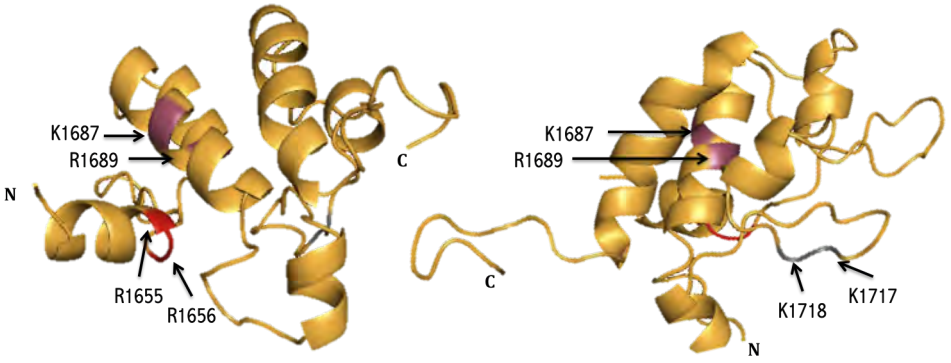


FIGURE 4.32: Two views of the I-TASSER model for periplakin linker protein showing the positions of the mutated residues. Pairs of residues of PPlink that showed a drastic inhibition of its interaction with vimentin after charge reversal (R1655/R1656 and K1687/R1689) are located in the N-terminal of the protein, whereas PPlink residues K1717/K1718, whose charge reversal mutation resulted in an unfolded protein, are located in a loop according to the model.

4.6 Overview of Chapter 4

A summary of the results presented in this chapter is given in Table 4.1.

First of all, regarding the experiments in Sections 4.3.1 and 4.3.2 that compared IF desmin and vimentin effect in desmoplakin PRD C and desmoplakin linker, we can conclude that both, the linker domain and the PRD C of desmoplakin are shown to be more interactive with vimentin than with desmin, in agreement with previous reports [175]. Also, these experiments permit to establish that desmoplakin linker domain amino acid binding sites for desmin are in the C-terminal end of the domain, while its binding sites for vimentin are in the N-terminal end, flexible region of the protein.

From our research with desmin, we can propose specific desmoplakin residues that could be binding sites for this IF: K2542 and D2559, and possibly L2554 if there is a non-polar interaction between plakins and the IF network.

Protein	Vimentin binding	Desmin binding
Desmoplakin PRD C	+	+
Desmoplakin linker WT	+/-	+/-
Desmoplakin linker E2495K-C2497R	+	ND
Desmoplakin linker S2526K-Q2527K	+	ND
Periplakin linker WT	+	ND
Periplakin linker K1687E-R1689E	-	ND
Periplakin linker R1655E-R1656E	-	ND
Periplakin linker K1717E-K1718E	ND	ND

TABLE 4.1: Summary of the binding experiments between desmoplakin and periplakin C-terminal domains and vimentin and desmin. Experiments with desmin were only carried out for desmoplakin PRD C and desmoplakin linker. Experiments with vimentin were carried out for all the proteins shown in the table except PPlink mutant K1717E-K1718E, which resulted unfolded after purification. The + symbol means strong binding activity, the +/- symbol means partial binding activity and the - symbol means no binding activity, ND means not done.

Regarding the analysis that involved only vimentin binding experiments, it has been found that periplakin linker strongly associates with the IFs, exposing a functional difference with its homologue from desmoplakin. By using the charge reversal technique, we were able to retrieve desmoplakin linker's vimentin binding properties and to compromise the vimentin binding activity of the periplakin linker domain. In the case of DPlink proteins that were subjected to site-directed mutagenesis, the successful delineation of the interaction interface by progressively disappearing resonances during titrations with vimentin demonstrated the specific nature of these interactions by comparing them with the neutrality of the DPlink wild type in the presence of vimentin. Opposite results were obtained for PPlink, whose mutants remained practically unaffected by the presence of vimentin. In this way, we found the essential amino acids in desmoplakin and periplakin linkers for vimentin binding: R2464 in the case of DPlink, for a demonstrated weak interaction, and for PPlink R1655-R1656 primarily and K1687-R1689 secondly.

Interestingly, R2464 and R1656 are equivalent amino acids in the linker alignment, and neither of them, nor R1655, align with any residue of desmoplakin PRDs B or C according to the plakin repeat motif sub-structure alignment (Fig. 4.33). This indicates

that the uniqueness of the N-terminal end of the linkers is the characteristic that confers functional importance to the linker domains of plakins, an idea supported by the high degree of conservation of such region through plakin linkers. Special attention must be paid to the fact that the N-terminal end is where more basic residues are conserved (Fig. 4.33) and it is where both linkers seem to interact more strongly with vimentin, alluding to the presumable importance of the acidic groove in vimentin and its interaction with the highly positive charged amino acid cluster highlighted by Nikolic *et al.* [168].

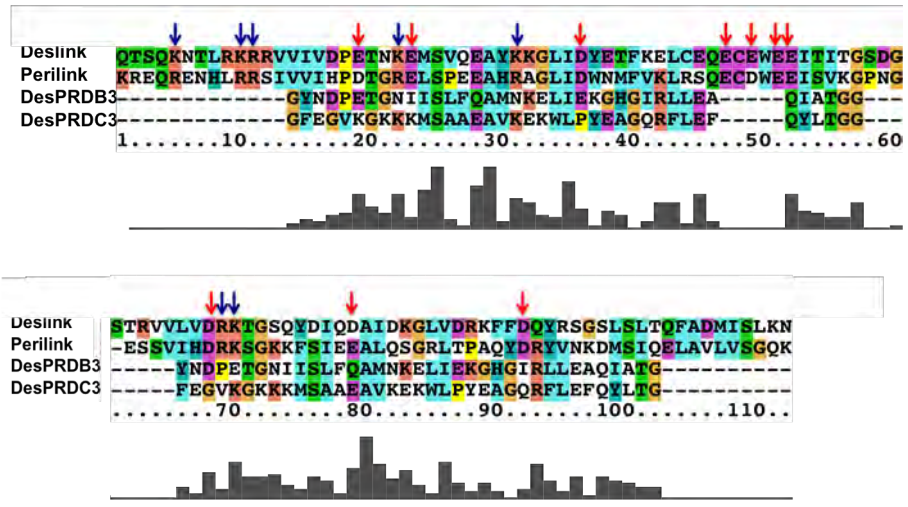


FIGURE 4.33: Alignment of desmoplakin linker (Deslink), periplakin linker (Perilink), desmoplakin PRD B plakin repeat 3 (DesPRDB3) and PRD C plakin repeat 3 (DesPRDC3), showing the conserved positively charged (basic) residues (blue arrows) and negatively charged (acidic) residues (red arrows). The N-terminal end of both linker domains, which does not align with the desmoplakin PRDs, conserves a cluster of basic residues, being the region where both linkers seem to interact more strongly with vimentin.

The results here presented demonstrate that the flexibility of the N-terminal end of the linker domains permits vimentin binding. This irregular sub-domain seems to be the one that provides essential engaging sites for vimentin, not the subsequent regular domain with structure that resembles the PR motif. In this way it can be explained why the linker is necessary to promote vimentin binding, ensuring binding specificity, while the PRDs, constituted mainly by repetitions of the PR motif, act only as secondary participants in this process, in agreement with Lapouge [175]. Since specific binding sites in the desmoplakin linker are essential but not sufficient to engage vimentin, the secondary role of PRDs B and C is also highly relevant, providing additional binding sites, that according to our findings could be residues residing in the positive groove in PRDs B and C surfaces proposed by Choi *et al.* [94].

We can hypothesise that periplakin linker possesses what used to be the original structure of the plakin linkers, including desmoplakin, but due to evolution of the latter, the structure of desmoplakin linker domain changed at the same time that some of its functional properties were transferred to more recent domains, i.e. the PRDs. This would affirm that, while all vimentin binding sites for periplakin reside in its linker domain, in desmoplakin they are distributed among the linker and the PRDs B and C.

Finally, regarding the nature of the identified binding sites, it has been concluded that electrostatic activity is the main regulator of binding between desmoplakin and periplakin linkers with vimentin, but possibly not only bringing together basic amino acids in the plakins with acidic amino acids in vimentin, but also the other way round, given the continuous presence of a negative residue in the vimentin binding site of desmoplakin linker through the binding experiments (Fig. 4.21 (a)), implying a substantially more complex recognition process between plakins and the IF network.

Chapter 5

AFM Study of Non-covalent Immobilisation of the Desmoplakin Plakin Domain by Size-Selected Gold Clusters

As AFM was the technique used to study the desmoplakin plakin domain interaction with gold nanoclusters on a graphite substrate, a brief introduction of this technique is given to begin this chapter, covering physical principles and different AFM modes. To continue, the experimental results of the interaction of the desmoplakin plakin domain with different substrates is presented, including mica, bare graphite and graphite decorated with gold nanoclusters. A summary of the results and discussion finishes the chapter.

5.1 Atomic Force Microscopy

5.1.1 Physical Principles

The behaviour of two non-bonded atoms in close proximity (as in the case of the tip-surface system) as a function of the internuclear separation r can be described by the Lennard-Jones potential [49] (Fig. 5.1):

$$V(r) = \frac{A}{r^{12}} - \frac{C}{r^6} \tag{5.1}$$

where A and C are constants associated with each type of atom that involve the depth of the potential well and the distance at which the inter-particle potential is zero.

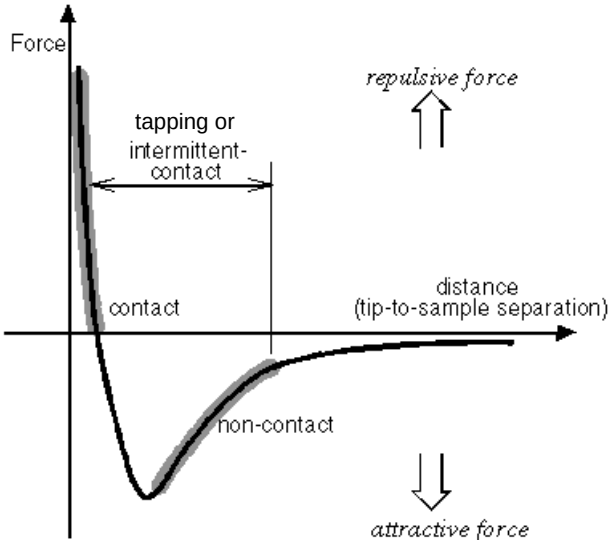


FIGURE 5.1: Interatomic force *vs.* distance curve. The graphic indicates the regions corresponding to contact, tapping (or intermittent contact) and non-contact AFM modes. In the contact mode regime the cantilever is located a few angstroms from the sample surface, where the interatomic force is repulsive. The intermittent-contact region starts when the tip is separated from the surface by about the length of a chemical bond, i.e. a couple of angstroms, when the interatomic force is zero. From this point onwards, the interatomic force is attractive, and when the tip-sample spacing is on the order of tens to hundreds of angstroms the tip has reached the non-contact regime. From reference [180].

As a function of distance, the interatomic force goes from a strong repulsive force, caused by coulomb interactions, represented by the first term in Equation 5.1, to an attractive force, due to van der Waals forces, represented by the second term in the same equation, being zero at a distance of normally 2 Å. Before crossing that point the atoms are still in contact, from that point and on, now in the attractive force regime, the atoms are first in intermittent contact and after passing through an equilibrium point go to the region in which contact is completely lost.

Atomic force microscopy is based on the attractive-repulsive electromagnetic force between the atoms in the sample surface and the atoms in the AFM tip. The force interaction causes the cantilever to deflect following Hooke's law. As the tip scans the

sample, variations in the surface topography cause changes to the force interaction and therefore an angular displacement of the cantilever.

As explained in Section 2.8.3, there are three main modes for AFM imaging: contact mode (CM), non-contact mode (NC) and tapping mode (TM), and this theoretical frame will be briefly explained in the following subsections.

5.1.1.1 Contact mode AFM

In contact mode AFM (CM-AFM) the proximity of the tip to the sample falls in the range in which the repulsive interaction is dominant. Within this range almost any force that attempts to push the atoms closer together will be balanced with the repulsive force, making tip-sample separation unlikely to be compressible, which will result in the bending of the cantilever. The intrinsic force on the cantilever depends on its spring constant and the deflection to which it is subjected. This force is calculated from Hookes law describing a linear elastic spring:

$$F = k_c \delta \tag{5.2}$$

Where δ is the cantilever deflection (nm) and k_c is the spring constant (N/m) [181], ranging normally from 0.03 to 5 N/m, resulting in forces from pN to μN in ambient conditions.

As a consequence of the stiffness of the cantilevers and the forces applied, contact mode AFM can be performed in air and liquid phases. If the measurements are performed in air, the capillary force must also be taken into consideration as part of the tip-surface system. The ambient environment often produces an intrinsic thin water layer, which, in contact with the tip, generates capillary and meniscus forces that depend on the tip geometry, relative humidity, surface tension and contact angles [182].

5.1.1.2 Non-contact mode AFM

The right hand of the curve illustrated in Figure 5.1 represents the regime in which the force interaction becomes attractive and follows van der Waals law. At these distances

the electrons of neighbouring atoms are redistributed, inducing an instantaneous polarisation of the interacting molecules generating dispersion forces [183]. This dipole-dipole interaction induces an attractive force, which is stronger than the previously dominant repulsive force. It is in this regime that the AFM works in its non-contact mode (NC-AFM).

In the same way, as in contact mode, the force between the atoms is the physical response that gives us the topographic information of the surface. In this case, the van der Waals force modifies the amplitude and phase of the cantilever's resonant vibration; thus, we get an effective resonant frequency. The force gradient experienced by the cantilever, i.e. the derivative of the potential-distance curve in Figure 5.1, which changes with tip-sample separation [184], dictates the variation on the effective spring constant (k_{eff}) of the cantilever, which in turn is related to its effective resonant frequency (f_{eff}) as the square root (resolving the equation for the simple harmonic oscillator):

$$f_{eff} = (\frac{k_{eff}}{m})^{1/2} \tag{5.3}$$

where m is the mass of the cantilever. Therefore, changes in the resonant frequency indicate changes in the tip-sample separation. An attractive force acting on the tip decreases the effective spring constant as the tip gets closer to the surface, resulting in a decreasing effective resonant frequency. The cantilever in the non-contact mode AFM is excited at a frequency (drive frequency) slightly above its resonance frequency (100-400 kHz depending on the cantilever), as explained in Section 2.8.3.

Unlike CM-AFM, NC-AFM is not capable of penetrating a liquid layer to image the underlying surface due to the low amplitude of the oscillations, surface tension of the liquid and frequency of oscillation in an aqueous medium (Fig. 5.2). Biological samples must often be measured in aqueous conditions, in order to obtain accurate information in their natural environment. As a result, NC-AFM cannot be used, leaving only two options available: contact and tapping modes. However, since most of the biological samples are soft samples, scanning in contact mode AFM can cause damage and/or deformation due to the force applied by the tip on the surface, often making tapping mode a preferred imaging technique for AFM biological analysis.

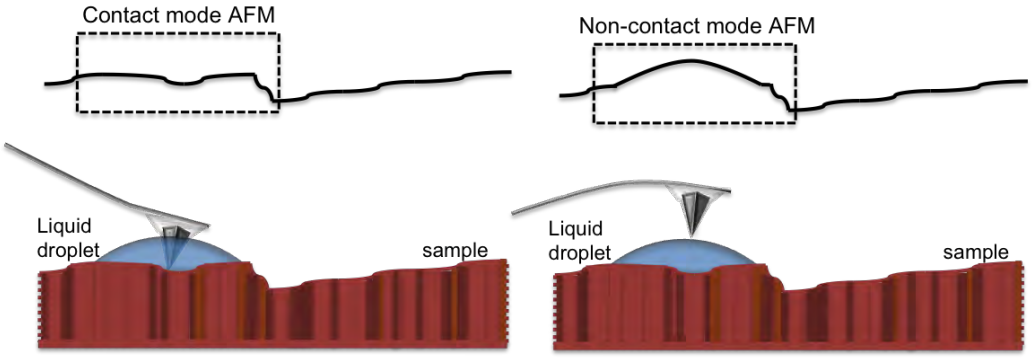


FIGURE 5.2: Illustration of the action of non-contact mode AFM when imaging aqueous samples compared with the corresponding action by contact mode AFM. The contact mode AFM tip is able to image the underlying surface by penetrating the liquid layer, whereas the non-contact AFM tip will image the surface of the liquid layer. Adapted from [185].

5.1.1.3 Tapping mode AFM

Also known as intermittent contact mode, the tapping mode (TM-AFM) was developed to overcome the respective difficulties and limitations that result from contact and non-contact in different situations [186]. This mode is mainly set in the attractive force regime, as with non-contact mode, and also works with an oscillation cycle with a frequency close to the resonant frequency of the cantilever, but with a much higher vibration amplitude (in the range of 20-100 nm compared to the non-contact mode amplitude of 10 nm). The tip gets so close to the sample that it taps its surface during oscillation resulting in less deformation of soft samples than contact mode AFM (Fig. 5.3).

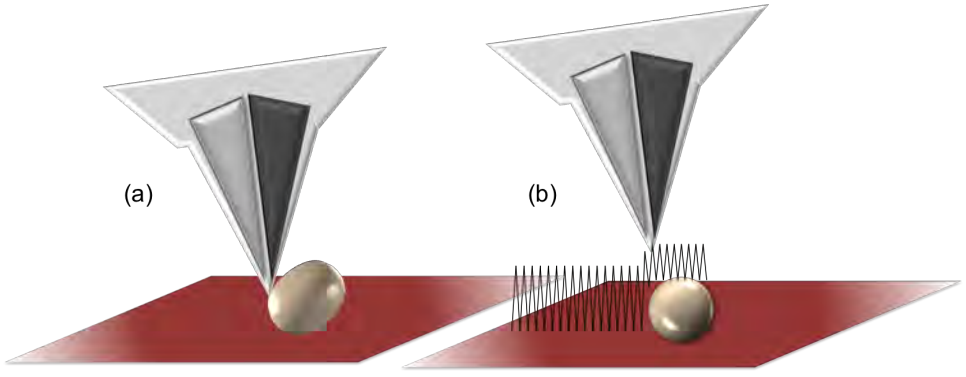


FIGURE 5.3: Representation of the difference between the action of the AFM tip in contact mode (a) and tapping mode (b). In the first case, the soft specimen is exposed to possible deformation, while in the second, deformations of this kind are prevented.

As mentioned previously, tapping mode solves problems in imaging soft biological samples, because it is able to penetrate in a liquid layer, but at the same time almost completely eliminates the lateral frictional force exerted. The spring constant of the tapping mode cantilevers must be higher compared with those of contact mode since the restoring force should be higher than the adhesion force due to the liquid layer. The cantilevers force constants ranges from 10 to 100 N/m. In tapping mode, a balance is achieved in keeping the tip close enough to the sample surface so as to be within the detectable range of the short range forces, and yet still achieve high resolution without the effect of a sticky tip which may arise from constant tip/sample close proximity.

5.2 Preparation of the Desmoplakin Plakin Domain for AFM Experiments with Gold Nanoclusters

As described in Section 2.8.1, the plakin domain protein used in this project was tagged at the N-terminal end with two cysteine residues. 2×Cys tagged proteins have previously been used to bind proteins to gold surfaces [187, 188]. In addition to the two N-terminal cysteine residues, the desmoplakin plakin domain contains 17 cysteines throughout its sequence.

In order to better understand the interaction of cysteine with gold nanoclusters it is necessary to roughly establish the framework of the fundamental concepts involved in such a process, namely the gold sulphur bond and the molecular surface area (MSA).

5.2.1 Cysteine and the Au-S Bond

The thiol group (SH) of the cysteine side chain is a highly reactive functional group consisting of a sulphur atom and a hydrogen atom. The low electronegativity of these two elements results in a non-polar group, leading to its characteristic hydrophobicity. When the cysteine is available in the protein surface and the sulphur is also available within the cysteine, the thiol group tends to form covalent bond given its ability to donate electrons [189]. The free pair of electrons on the sulphur atom can be shared with other atoms in order to achieve a more stable state by completing their last electronic shells. Since sulphur and gold have similar electronegativities, (2.58 and 2.54 respectively) [190], the

Au-S covalent bond is frequently formed, especially between organosulphur compounds and gold surfaces. Experimental work has found that the force needed to break a Au-S anchor is 1.4 nN at a force-loading rate of 10 nNs^{-1}, concluding that the Au-S bond is stronger than the Au-Au bond [64]. The nature of the Au-S interaction is not completely understood [191]. Multiple experimental and computational works such as Functional Density Theory studies (especially for Au (111)) [64], X-ray Photoelectron Spectroscopy experiments with SAMs [192] and Atomic Force Microscopy measurements [193] have been conducted in order to characterise the dynamics of this interaction. However, none of them have been fully successful. Nevertheless, the Au-S covalent bond has been widely used to link biomolecules to surfaces due to its strength and stability, making possible the chemisorption of proteins containing available thiol groups from, for example, cysteine residues into metals. Two potential obstacles for the thiol group in performing this task are the probability of being buried in the protein, blocking the availability for reaction, and possible oxidation [194].

5.2.2 Molecular Surface Area

One of the most important factors involved in immobilisation procedures is the exposed molecular surface area (MSA) of cysteine residues present in the protein. The molecular surface area is defined as the contact surface area of a probe sphere that rolls along the surface of the protein [195]. An accurate representation of the outer surface of a molecule depends on its size [58]. For small molecules the van der Waals surface of the atoms defines the molecular outer surface, but for larger molecules only a part of the van der Waals surface (contact surface) is exposed on the outer molecular side [62] (Fig. 5.4). In such cases, in order to complete the molecular outer area, it is necessary to apply the concept of MSA, which is related to the probe sphere when it is in simultaneous contact with more than one atom, as is shown in Figure 5.4.

The accessibility of the functional residues in a protein can be quantified by this parameter; the higher the accessibility, the higher the probability of reaction between the residue and the external probe. The probe radius used to calculate MSA is based on the radius of a water molecule, which is 1.4 Å[195]. In our case the external probe role is played by one of the atoms within the gold nanoclusters (with a radius of 1.4-1.7 Å) so, the cysteine residues contained in the protein must be available to chemisorb on the

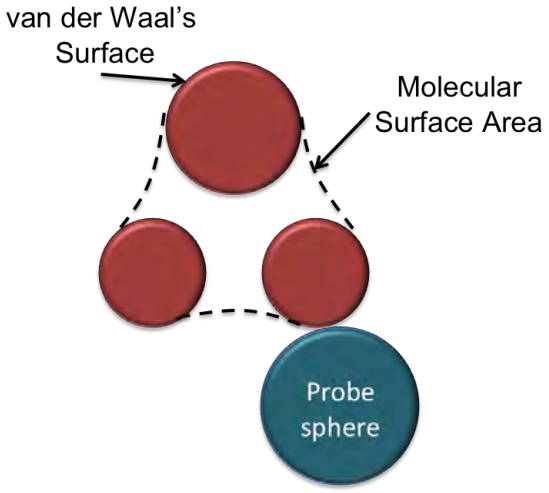


FIGURE 5.4: Representation of molecular outer surface consisting of two parts, the part of the van der Waals surface exposed to a probe sphere, also called contact surface, and the molecular surface area (MSA), also called the re-entrant surface, which is the contact surface area of a probe sphere which rolls along the surface of the protein. Adapted from [62].

gold clusters. Similar experiments have been carried out with proteins, which are known to contain cysteine residues buried inside their structures (GFP) and as a consequence these proteins did not present covalent bond to the gold nanoclusters [63].

5.3 Immobilisation of the Desmoplakin Plakin Domain by Gold Size-Selected Clusters on Graphite

Protein immobilisation techniques have been developed for different purposes, including the complementation of single protein AFM analysis techniques. Mica is one of the substrates most used to immobilise proteins through a process of physisorption promoted by its hydrophilic surface. However, the non-specificity of the interaction is disadvantageous in a number of aspects.

Immobilisation of desmoplakin plakin domain proteins on graphite pre-decorated with gold size-selected clusters attempts to demonstrate the following advantages of this nanostructured substrate over mica: the facilitation of obtaining an array of isolated single PD molecules by specific interaction with gold clusters; a strong interaction through covalent binding, in order to simplify stretching AFM experiments; and also, as a benefit of the covalent bond specificity, the protein orientation.

In the following sections, an AFM study of the desmoplakin plakin domain deposited in different substrates will be described, starting with mica as a general reference, followed by bare graphite HOPG and finally HOPG decorated with gold nanoclusters of 55 and 147 atoms.

5.3.1 Interaction of the Desmoplakin Plakin Domain on Homogeneous Surfaces

5.3.1.1 The plakin domain on mica

The outer region of proteins is highly polar, which enables attractive interaction with charged surfaces [196]. This makes mica a preferred substrate for imaging proteins with AFM, given its favourable properties.

Mica is a mineral with a high tendency to form crystallographic hexagonal structural planes of atoms and ions. Two double layers, $(SiAl)_2O_5$ and $Al_2(OH)_2$, constitute mica layered crystals, creating with their oxygen atoms a net negative charge which is balanced by a layer of cations K^+. These positive ions are exposed when the mica is cleaved, and, in the presence of water, dissociate from the surface resulting in a negative charged surface (-0.0025 C m^{-2}) and therefore are rather hydrophilic [30, 62].

The polarity of the protein core easily forms ionic bonds with the mica surface, so, combining this property with the characteristic flatness of this substrate, AFM imaging of high concentrations of plakin domain could be carried out successfully. A high and homogeneous density of protein was found all over the mica surface as can be appreciated in Figure 5.5 (a). The topographies display features of three main diameters, 17.27±3.04 nm, 28.82±2.90 nm, and 39.90±2.98 nm (Fig. 5.6). Taking into consideration the diameter overestimation intrinsic to measurements due to the tip sectional radius, it is possible to interpret the resulting the tri-modality diameter distribution by comparing the dimensions obtained with those provided by investigations of desmoplakin carried out in dry conditions, mentioned in Section 1.5.3.2.

Given the heterogeneous nature of the features imaged on mica, the diameter was calculated as the diameter of a spherical shape approximation for each feature. Large features, with mainly elongated shapes, have average diameters that can be compared with the

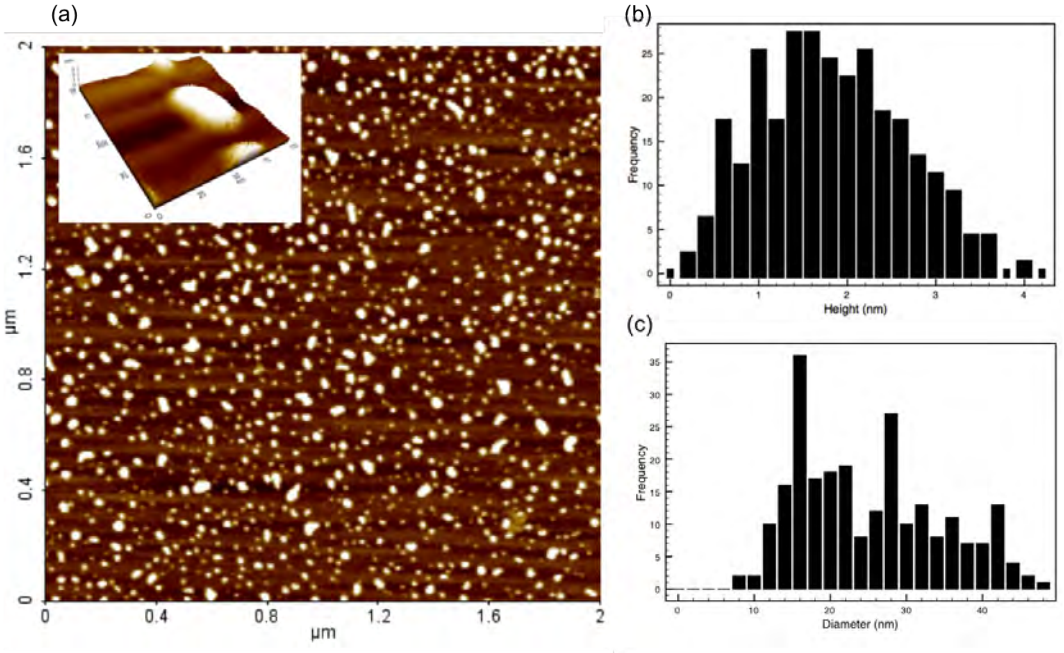


FIGURE 5.5: Desmoplakin plakin domain proteins deposited on mica imaged by non-contact mode AFM in dry phase. (a) Dry phase NC-AFM topography showing homogeneous density of PD molecules (concentration of 0.22 μg/μL) on freshly cleaved mica. The 3D image in the inset shows a scan size of 100×100 nm of one of the features with an elongated shape. (b) Height distribution with mean value is 1.81±0.79 nm. Possible dehydration effect involved, as well as fragmentation after deposition could be reasons for low heights. (c) Diameter distribution showing a tri-modality of the distribution gives three main values, centered at 17.27±3.04 nm, 28.82±2.90 nm, and 39.90±2.98 nm. The overall mean diameter is 25.06±9.41 nm. For the diameter distribution no subtraction of the AFM tip sectional radius was performed.

32 nm diameter extended configuration of the desmoplakin PD previously reported by electron microscopy rotary shadowing images of desmoplakin in dry phase [91]. This is considering the predicted compensation of AFM lateral measurements, given that the tip radius used was 10 nm [35]. Medium sized features, with globular shape, have an average diameter of 19 nm after subtraction of the tip sectional radius, in agreement with the minimum length of the range measured in electron microscopy experiments which corresponds to the round configuration (16 nm) [91]. Small size features display an average diameter of 7 nm after subtraction of the tip sectional radius, which are dimensions under the minimum diameter expected, which suggests possible denaturation of the protein.

The height of large and medium features is very similar, ranging from 2 nm to 2.5 nm in both cases. For the small features, the height mean value is 1.31±0.47 nm. Taking into account all particles, the mean height is 1.81±0.79 nm. This discrepancy with the dimensions (a minimum height of 4 nm, which is the lowest dimension of the desmoplakin

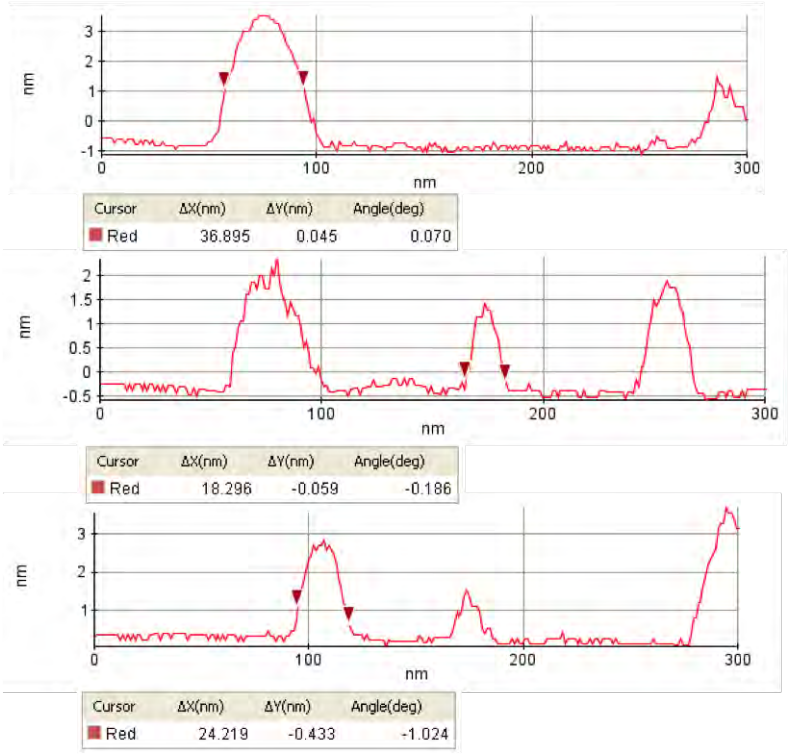


FIGURE 5.6: Profiles corresponding to the three typical sizes of features found in mica after deposition of the desmoplakin PD in dry phase. In the boxes, ΔX indicates the peak width in each case. The diameters displayed lie within the range of values indicated in the distribution of diameters shown in Figure 5.5 (c).

plakin domain segment that has been crystallised (18×4×4 nm) [96]) is probably caused by dehydration of the protein and/or flattening of the protein induced by the interaction of the PD with the mica, enhanced by the dry conditions under which the measurements were conducted.

Experiments on mica were a first approach to imaging the desmoplakin plakin domain with AFM, and were used only as a reference for further experiments on HOPG and on HOPG decorated with clusters.

5.3.1.2 The plakin domain on HOPG

Highly oriented pyrolytic graphite (HOPG) is a type of graphite with a high degree of preferred crystallographic orientation of the vertical axes perpendicular to the surface. The plakin domain with different concentrations was deposited on this substrate and imaged with non-contact, contact and tapping mode AFM in liquid and dry phase.

As previously mentioned, graphite is a hydrophobic material, so in principle its surface presents a repulsive interaction with proteins due to the protein characteristics. However, using the highest (0.22 μg/μL) concentration of PD, and because of conformational changes that the protein suffers when approaching a hydrophobic surface (which lead to an exposition of the hydrophobic core that comes into contact with the surface) it was possible to visualise how PD interacts with bare graphite [62]. The typical scan size for these measurements was 5×5 μm. Some larger scans were also taken in order to see the behaviour of the protein in a relatively extended area. The topographies show how the molecules tend to bind to the graphite steps, accumulating in such regions (Fig. 5.7 (a)). Graphite steps, defects, grain boundaries and terraces act as reactive sites suitable for protein attachment, which is enhanced by using high concentrations of protein. This increases the probability of weak non-specific adsorption between PD molecules and the graphite surface taking place, leading to heterogeneous nucleation. The feature dimensions are larger, compared with the dimensions of the features shown on the mica substrate, suggesting a higher occurrence of protein aggregation. Given the differences in surface chemistry between mica and graphite (hydrophilic and hydrophobic respectively), it can be inferred that each one induces a different adsorption process in the desmoplakin PD, resulting in conformational changes of the protein [197].

The concentration was then reduced to 2.2×10^{-4} μg/μL, and the same experiment was reproduced, resulting in a drastic fall in the number of proteins remaining on the graphite and proving the steps and defects to be attractive regions of few protein molecules, as can be seen in Figure 5.7 (b), which again presents features with dimensions comparable to the largest and medium complexes appearing in mica imaging.

In order to investigate the desmoplakin PD under native conditions, contact and tapping mode were used for liquid phase imaging.

For contact mode, PD with a concentration of 0.22 μg/μL was deposited on bare graphite. In a first CM scan, using a force of 1 nN (proportional to the frictional force exerted by the AFM tip [198, 199]), it is possible to see a high number of horizontal bright streaks (Fig. 5.8), from which it can be implied that the proteins are swept by the AFM tip during the scanning. In a second scan applying the same force, almost all features have been completely removed from the surface region (Fig. 5.8). These results reveal the strong reaction of PD to the action of the contact mode tip. Forces lower than

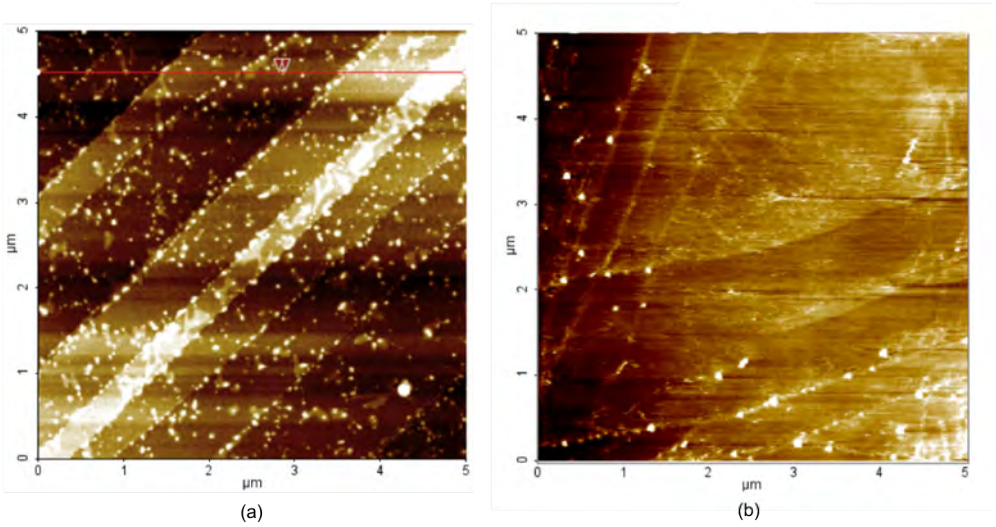


FIGURE 5.7: Desmoplakin plakin domain proteins at different concentrations deposited on bare graphite imaged by non-contact mode AFM in dry phase (a) NC-AFM topography of PD at 0.22 μg/μL on bare graphite, showing multiple graphite steps. It is possible to distinguish the preference of the molecules for the steps, aggregating in such regions. The protein was deposited following the standard protocol described in Section 2.8.1. (b) NC-AFM topography of PD at 2.2×10^{-4} μg/μL on bare graphite surface showing steps and defects, showing how only a few proteins anchored to the graphite steps. The remaining of proteins in the graphite defects after washing the sample is an indicator of weak adsorption interactions between the proteins and the surface exclusively in its step sites. The lack of proteins on the rest of the surface indicates that bare graphite, without steps or defects, does not provide the protein with enough reactive sites to anchor.

this value result in the false engagement of the tip in the surface, so it is possible to say that, even under the action of the contact tip scanning with the minimum possible force, the protein is displaced very easily from the surface of the sample. This demonstrates that PD does not bind to the unmodified bare graphite surface, since it does not provide enough binding reactive sites and the molecules are completely removed by lateral displacement.

For tapping mode AFM, a concentration of 2.2×10^{-4} μg/μL of PD was used for liquid phase imaging (Fig. 5.9), showing no presence of PD proteins on the scanned surface, including the graphite defects. On this basis, it could have been argued that the lower concentration used for this TM-AFM topography, as well as for the NC-AFM topography of Figure 5.7 (b), explains the absence of proteins on the HOPG surface in both cases, but as will be evidenced by repetition of the same experiments using the same concentration but on graphite pre-decorated with gold clusters (following sections) this is not the case. Instead, this is supporting evidence of the lack of binding interaction between PD and the bare surface of graphite.

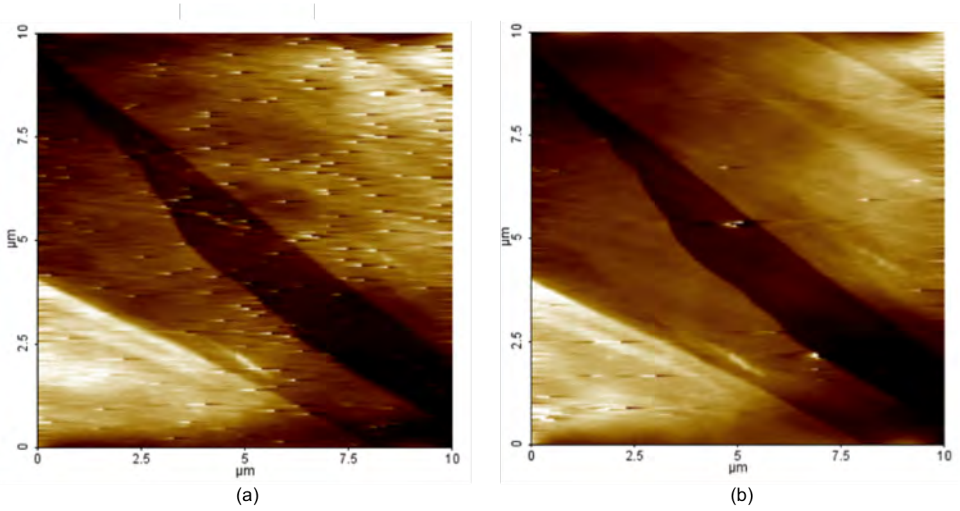


FIGURE 5.8: Comparative liquid phase contact mode AFM images on buffer solution of PD on a bare graphite surface. On this occasion the sample was not washed before imaging in order to keep the highest possible density of protein on the substrate. On the left a strong action of the tip over the features is discernible, sweeping them away. On the right is an image of a second scan in exactly the same region, showing almost no features over the surface. The force applied was 1 nN. The highest protein concentration (0.22 μg/μL) was used for these measurements in order to increase the probability of passive adsorption of the protein by the hydrophobic substrate. This explains why it is possible to see proteins (although they are being swept by the action of the tip in contact mode) in the first scan.

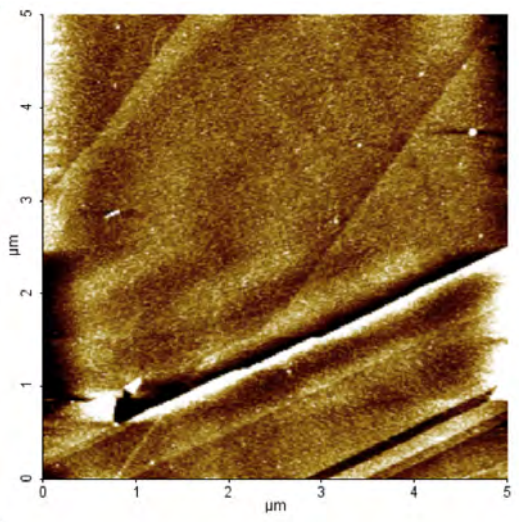


FIGURE 5.9: Liquid phase tapping mode AFM topography image of PD on bare graphite. Almost no features are observable on the graphite surface except for one in the right top. The protein concentration for all tapping mode images was 2.2×10^{-4} μg/μL. Without the stability provided by several monolayers of protein covering the graphite surface (as in the case of liquid phase contact mode measurements of PD at 0.22 μg/μL shown in Figure 5.8), the proteins are displaced by the AFM tip in tapping mode, demonstrating that the graphite surface without modification does not offer binding sites for anchoring the molecule.

5.3.2 Interaction of the Desmoplakin Plakin Domain with Size-Selected Gold Clusters

5.3.2.1 Size-selected gold cluster characterisation with STM

The graphite samples decorated with Au_{55} and Au_{147} produced in the cluster source were imaged with STM and it was possible to corroborate individual gold clusters pinned on the graphite surface with no agglomerations. Samples with Au_{147} had a density of 600 clusters/μm^2 and an average height is 0.44±0.34 nm, which, compared with its theoretical diameter of 1.72 nm (calculated based on the spherical approach, using the bulk lattice spacing of gold, 4.08 Å, and the consequent nearest neighbor distance of 2.88 Å), exposes a deformation of the cluster morphology, expressed as flattening due to the impact on the surface (Fig. 5.10).

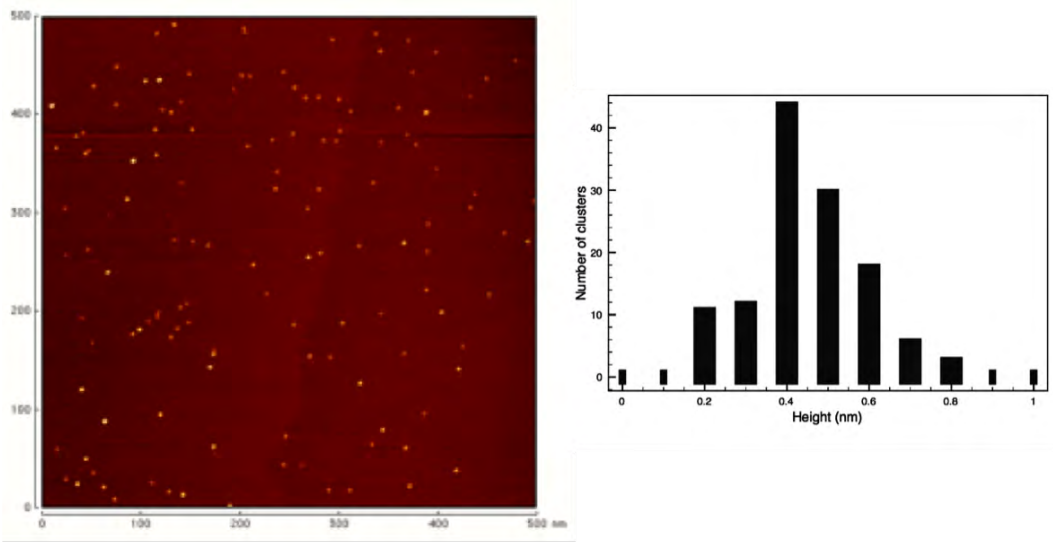


FIGURE 5.10: Au_{147} clusters on graphite. (a) STM image acquired in constant current mode with a current of 100pA, a voltage of 800 mV and a scan size of 500 nm×500 nm. The density of clusters deposited on the graphite (600 clusters/μm^2) is considerably higher compared with Au_{55}. (b) Histogram of the heights distribution of Au_{147} clusters on graphite. The distribution mean value and standard deviation is 0.44±0.34 nm.

Samples with Au_{55} had a density of 200 clusters/μm^2 with an average height of 0.40±0.11 nm, which, compared to the corresponding theoretical diameter (1.14 nm), demonstrates the same deformation effect as for Au_{147} clusters (Fig. 5.11). The images were processed and analysed with the *Image SXM* analysis software.

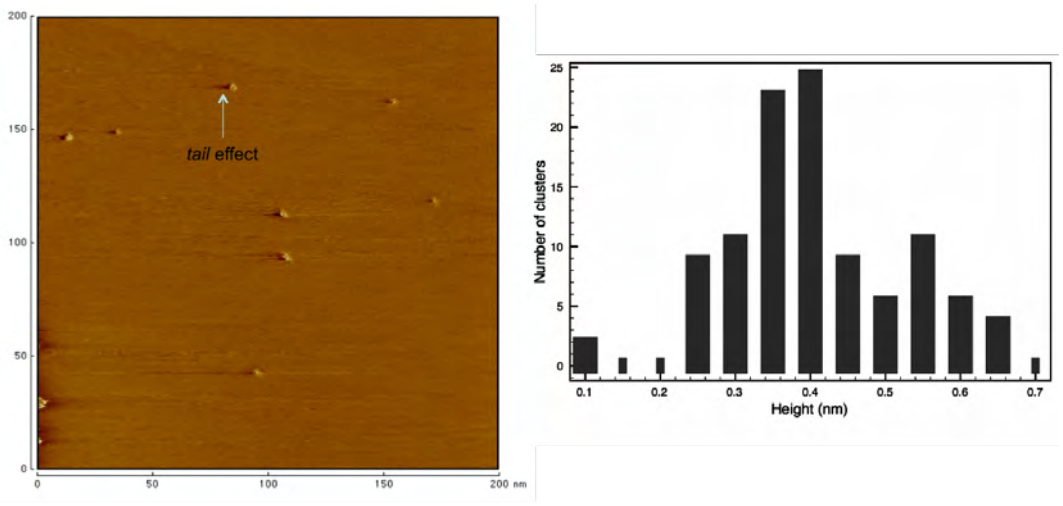


FIGURE 5.11: Au_{55} on graphite (a) STM image acquired in constant current mode with a current of 100 pA, a voltage of 400 mV and a scan size of 200 nm×200 nm. The clusters are randomly deposited on the substrate with a density of 200 clusters/μm^2. (b) Histogram of the height distribution of Au_{55} clusters on graphite. The mean height is 0.40±0.11 nm. Values towards the outer height range are probably due to double-layer clusters resulting from free parameters involved in the deposition procedure such as orientation and direction of the clusters in the gas phase.

The mean distance between deposited clusters measured in STM images for Au_{55} clusters was 150 nm and in the case of Au_{147} was 40 nm all over the surface of the sample (Table 5.1), which is a rectangle of 10 × 5 mm. Several STM scans were carried out on these samples within a period of four months thereby demonstrating that they are stable at room temperature without displacements or agglomerations.

CLUSTERS	MEAN SEPARATION (nm)	HEIGHT (nm)	DIAMETER (nm)
Au_{55}	150 ± 56	0.40 ± 0.1	1.22 ± 0.4
Au_{147}	70 ± 48	0.44 ± 0.3	1.30 ± 0.4

TABLE 5.1: STM characterisation of gold clusters on graphite produced by the NPRL cluster source for the experiments of the present work. The error in every case is ± 0.1 nm, associated with the SXM software measurements. STM error is negligible (0.01Åfor height measurements and 0.1Åfor lateral measurements). The difference between lateral dimensions and height suggests a flattening effect due to the deposition technique.

The histogram of Figure 5.11 shows that in the case of Au_{55} the height distribution exhibits a double normal distribution. A viable explanation for this relies on the cluster deposition procedure. Prior deposition, the clusters in the gas phase have variable configurations that confer them with different orientations once deposited in the substrate.

The different orientation of the clusters leads to pinned clusters with one or two layers, therefore determining their height. In view of this, it could be said that the larger values in the height distribution are the result of double-layered clusters.

5.3.2.2 The plakin domain on size-selected gold clusters in dry conditions

After studying the behaviour of PD in the presence of mica and graphite homogeneous surfaces, the same experiments were performed with the mediation of size-selected gold clusters. The sample preparation followed the protocol described in Section 2.8.1 and the features found in these studies were statistically analysed: height and diameter distribution are presented and discussed. Lateral measurements were subjected to deconvolution by subtraction of the AFM tip sectional radius (10 nm) from the features apparent sectional radius.

15 μL of PD at low concentration of 2.2×10^{-4} μg/μL were deposited on two kinds of HOPG samples, one decorated with Au_{55} clusters, and the other with Au_{147}. A decisive difference was found in the results with respect to bare graphite. Non-contact AFM measurements show how both sizes of gold clusters enhance the binding of the molecule, leading to high density in the remaining features on the surface after the washing procedure, as can be seen in Figure 5.12.

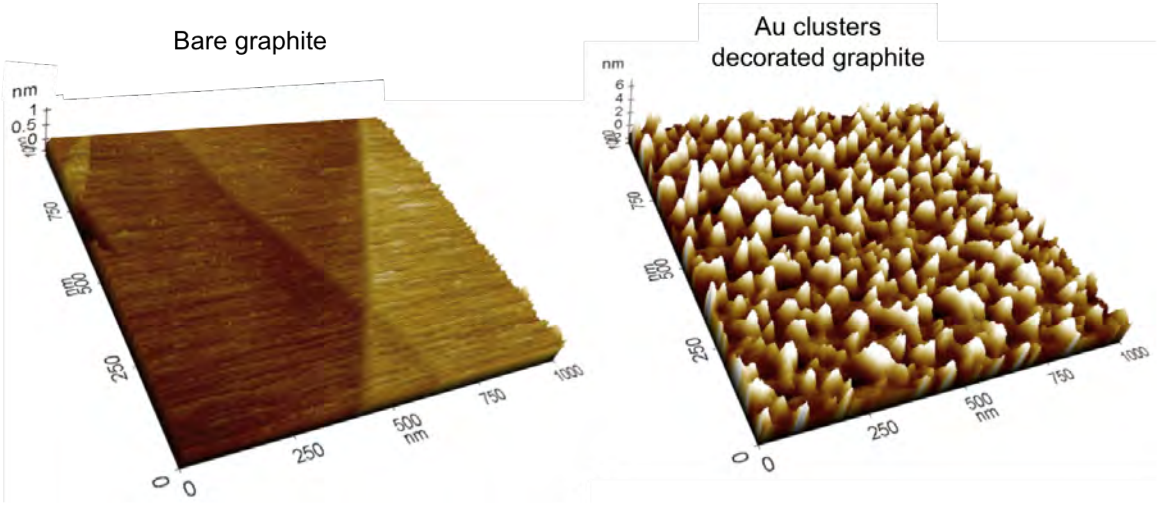


FIGURE 5.12: Comparative dry phase NC-AFM images of the binding of PD with the same concentration in both images, (2.2×10^{-4} μg/μL), on graphite without (left) and with (right) gold nanoclusters (Au_{55}). The action of the clusters over the protein as enhancers of the interaction is evident.

Contrary to the binding on homogeneous surfaces, the size-selected clusters induce a stronger adsorption of the PD protein molecules onto the surface, this possibly being independent of the existence of specific binding sites provided by the gold clusters for a specific part of PD, i.e. the N-terminal containing the two cysteine residues. On the basis of such an idea, given that it is not possible to guarantee that the cysteine residues are the binding site of the desmoplakin PD to the substrate surface, it is not possible to establish whether or not a preferred orientation of single proteins is given after deposition.

In the case of samples decorated with Au_{55} with deposited protein (Fig. 5.13) the diameter histogram reveals a tri-modality (Fig. 5.13 (c)), with a slightly broader size distribution than for mica. The highest frequencies are around the value 27.85±3.49, which can be translated as a prevalence of large-size features that can be associated with the oblong protein configuration found in electron microscopy rotary shadowing images [91]. Sizes over 40 nm may be associated with protein complexes. Taking into account these values, the mean height is displaced to a higher value, i.e. 28.93 nm, with a standard deviation of 9.92 nm. Assemblies of multiple PD are presumably formed through weak interactions between the proteins. For this reason, although the mean distance between clusters was large enough to avoid the overlapping of proteins, the results show protein agglomerations.

The mean height value for PD on Au_{55} clusters is 1.88±0.71 nm, according to the (b) histogram plot of Figure 5.13, very similar to the height dimensions found in the case of deposition on mica (1.81±0.79 nm). Given the dry conditions, it is possible to argue that these low values for protein height are due to factors such as dehydration, fragmentation and/or modifications of the protein structural conformation due to interaction with the surface. Changes in protein structure and dynamics have been related to the reduction of hydration of the molecule, caused by the removal of water [200, 201], when the protein is imaged in a dry environment. The lack of liquid promotes the maximisation of the contact surface of the protein with the graphite substrate. The possible exposition of the hydrophobic core due to the interaction with the graphite also contributes to the increased contact area. Another factor, that cannot be completely ruled out despite the less disruptive tapping mode, is the deformation of soft samples due to the action of the AFM tip. This effect is also evident on the images acquired from Au_{147} samples, reflected in the height distribution with a mean height of 2.74±0.88 nm (Fig. 5.14). The

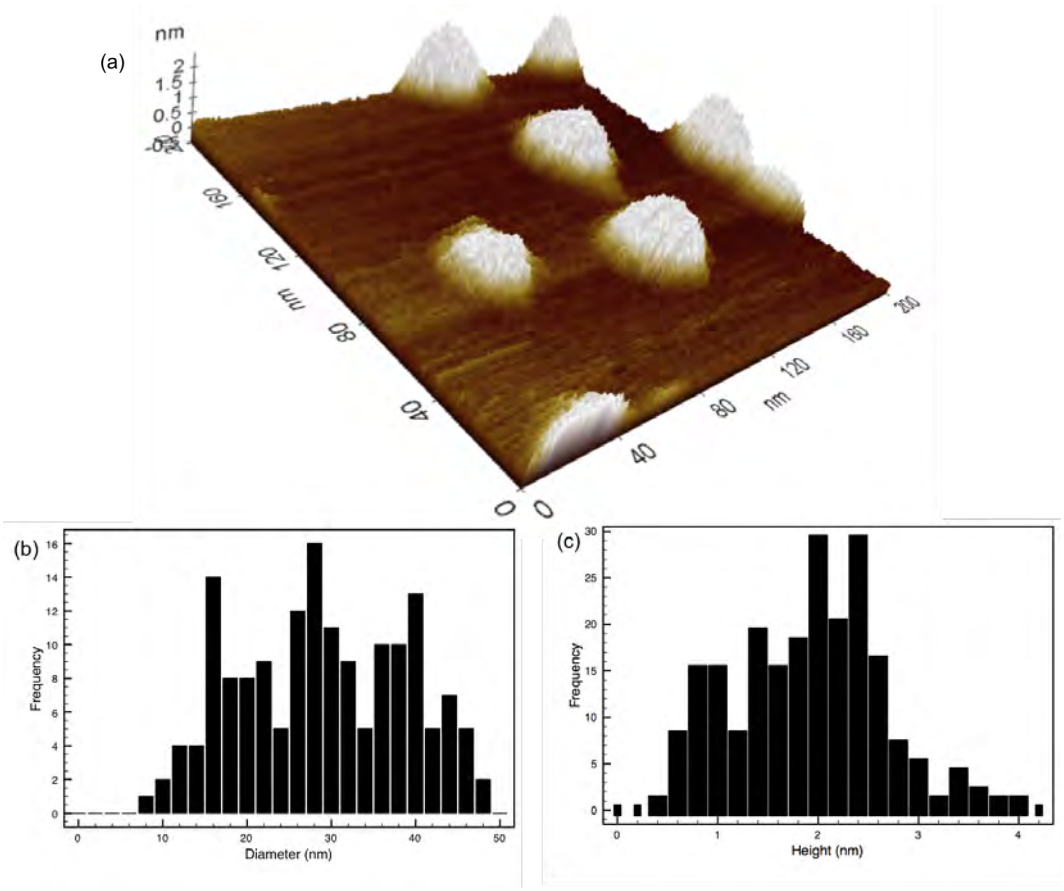


FIGURE 5.13: Desmoplakin plakin domain proteins on graphite pre-decorated with Au_{55} clusters imaged by non-contact mode AFM in dry phase (a) 3D plot of dry phase non-contact mode AFM image of PD on graphite with Au_{55} clusters. (b) Diameter distribution histogram displaying three local mean values very similar to the results from the sample in mica, i.e. 18.18±2.61 nm, 27.85±3.49 nm and 38.96±3.02 nm. The main diameter of the overall distribution is 28.93±9.92 nm. (c) Height distribution histogram that shows higher frequencies for values of 2 and 2.4 nm. The mean value is 1.88 nm with a standard deviation of 0.71 nm.

diameter distribution covers a narrower range of values, in which there is a frequency peak for the value of 16 nm (Fig. 5.14 (b)), indicating a higher density of features associated with medium-size globular molecules than the density of features identified as protein in the extended configuration.

Feature diameters over 40 nm were more rare for the Au_{147} samples. Lower densities of average laterally larger protein complexes were measured by non-contact mode AFM on samples with Au_{55}, pointing to the lower density of clusters observed in these samples with STM in comparison to samples decorated with Au_{147} clusters. With a lower density of clusters, samples decorated with Au_{55} attract more proteins per cluster, leading to molecule aggregation. From the results of experiments on bare graphite it is demonstrated that PD finds the reactive sites during incubation time and binds to them. In

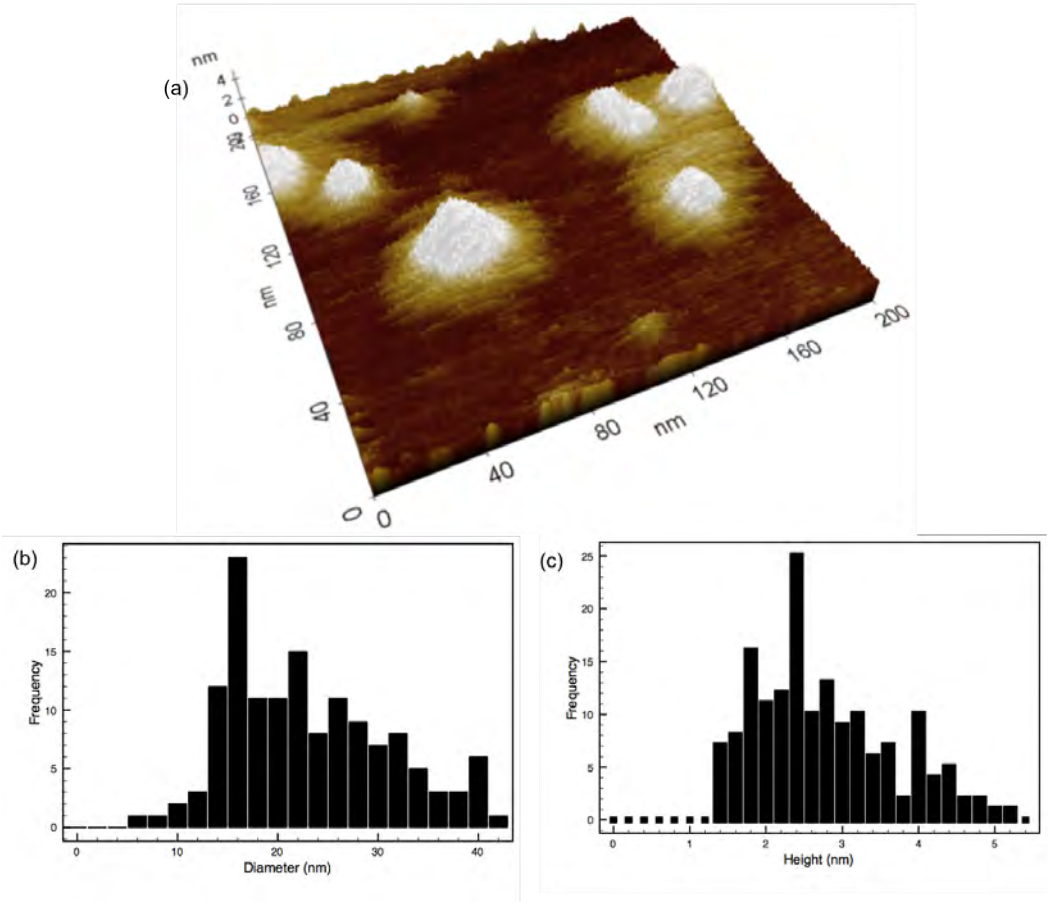


FIGURE 5.14: Desmoplakin plakin domain proteins on graphite pre-decorated with Au_{147} clusters imaged by non-contact mode AFM in dry phase (a) 3D plot of dry phase non-contact mode AFM topography of PD on graphite with Au_{147} clusters. (b) Diameter distribution histogram, in which is not possible to identify the tri-modality previously seen in the diameter distributions. The value of 16 nm has the highest frequency within the diameter distribution, as well as in the cases of PD on mica and PD at 0.22 μg/μL in HOPG. The mean value is 23±10.10 nm. (c) Height distribution histogram with a mean value is 2.74 nm and a standard deviation of 0.88 nm.

modified samples, the clusters play the role of such reactive sites and in the absence of enough of them, more than one protein tends to accumulate near one cluster via weak interactions. On the other hand, a higher density of clusters allows proteins to spread out, and create a binding interaction of higher intensity. The density of features found matches the density of clusters for Au_{55} (200 clusters/μm^2), having in average 152 features/μm^2 for samples decorated with this size of clusters, while for the case of Au_{147} an average density of 167 features/μm^2 was found, which is much lower than the density of clusters per μm^2 (600 clusters/μm^2). The latter ratio seems to be common in previous experiments in the same line of research, using gold clusters with different proteins [62], suggesting that the density of features depends mainly on the concentration of the solution and remains constant for cluster densities varying within the range

of hundreds.

5.3.2.3 The plakin domain on size-selected gold clusters in liquid

In situ (i.e. in buffer solution) AFM imaging of the plakin domain provides a suitable environment for the protein to avoid denaturation. For this reason a liquid phase is the preferred condition to study the interaction of the protein with the nanoclusters by AFM.

Contact and tapping mode AFM were carried out in order to assess two different aspects: contact mode was used to determine whether or not PD is strongly immobilised via the size-selected gold clusters pinned on a graphite surface, while tapping mode was implemented to complete a characterisation analysis of the protein features and determine the presence of single PD molecules. Contact mode measurement results will be discussed first for samples with both sizes of clusters and subsequently for the tapping mode experiments.

CONTACT MODE MEASUREMENTS

Previous experimental results have demonstrated the stability of pinned gold clusters on graphite against lateral diffusion at high temperatures [202] and against repeated scans in contact mode AFM [56, 62]. Such stability allowed to carry out contact mode AFM studies on PD deposited on a graphite surface pre-decorated with gold nanoclusters in order to test how strong is interaction between the protein and the functionalised substrate by subjecting the attached proteins to the action of the scanning AFM tip.

It has been proved [37] that forces greater than 100 pN in contact mode AFM induce, in soft samples, a reversible deformation, whilst smaller forces prevent deformation. The forces applied during this research ranged from 100 pN to 3 nN, so according to this classification they can be catalogued within the reversible-deformation forces. In general, forces around 1 nN were implemented, since more stability and better image quality was obtained this way. Forces below 100 pN were impossible to use since, by applying such forces, the tip-surface interaction resulted in instability, i.e. the signal was lost continuously.

The two comparative tapping mode images in Figure 5.15 present the same region in the sample before and after just one scan in contact mode. The procedure consisted of

scanning once in TM in order to delineate a region, identifying precise features, then scanning a sub-region in CM and afterwards scanning the original region (except for some drift) again in TM in order to see the results of the CM scanning. For the image of Figure 5.15, the scan area of the tapping mode scans was larger than that of the contact mode in order to see the region (a square) that had been swept clear by the tip in contact mode. Despite the drift effect induced by the piezo scanner, it is possible to distinguish the same features in both pictures, revealing a clean scanned area totally free of proteins and an accumulation of material at the borders of the square. The accumulation of material at the edges of the initial contact mode scan is further proof of the sweeping of weakly bound molecules.

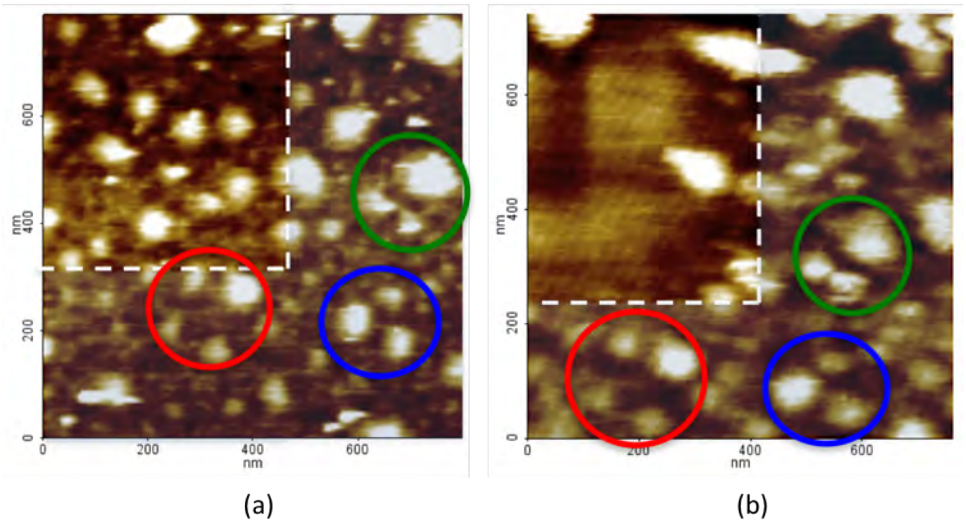


FIGURE 5.15: Comparison between images of PD on Au_{147} clusters before *(a)* and after *(b)* a contact mode scan applying a force of 720 pN. Both images were taken in tapping mode in buffer solution. *(a)* Firstly, a tapping mode scan is performed, showing stable molecules immobilised on the graphite surface. A second scan, but now in contact mode, is performed on the same region but with a smaller scan size, showing a clean surface with no proteins (not shown). *(b)* Finally, a third scan is performed in tapping mode the same region and with the same scan size as the first scan, showing a "window" of clean surface and proteins accumulated at the edges, demonstrating how the PD molecules remain on their original sites after tapping mode scans but not contact mode scans. The circles indicate the features that can be recognised in both images.

In the images of Figure 5.16, the contact mode scan was carried out in an area of the same size as the tapping mode scans but the surface region had a particularity (triangle at the top left) that allowed the images to be related. In this case the second tapping mode image reveals a clean surface with the exception of a couple of slightly visible features remaining on the surface irregularity. The same results were found in samples with Au_{55}.

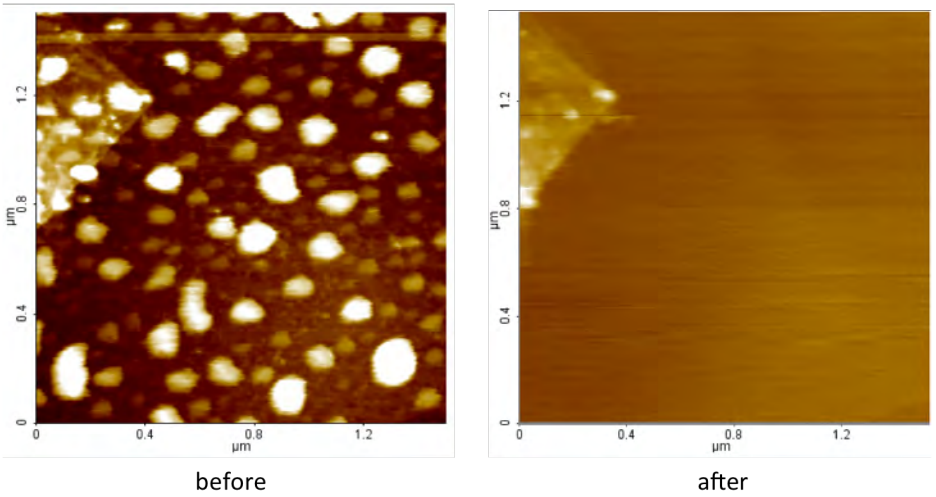


FIGURE 5.16: Comparison between images before and after a contact mode scan applying a force of 900 pN in samples pre-decorated with Au_{55} clusters. Deposition of the protein and imaging process followed the same protocol as before, using a reduced concentration of the PD (2.2×10^{-4} μg/μL). The left image was taken in tapping mode and the right image was taken in contact mode AFM. Both images were taken in buffer solution.

These "before and after" images (Figures 5.15 (a) and (b) and 5.16) exhibit how PD molecules are displaced by the action of the AFM tip in contact mode, demonstrating the lack of covalent bond in the system.

In previous experiments made by the NPRL group [56], proteins with a high number of cysteine residues, such as GroEL with 21 residues, show stability under the stress caused by several scans with the AFM tip in contact mode, due to the larger number of thiolate bonds formed with the gold atoms. In cases like this it is possible to repeatedly scan an area without losing any molecules, proving a strong immobilisation of the proteins performed by the gold clusters. A similar level of stability is found when undertaking the same experiment for proteins with only five Cys tags with different degrees of accessibility, as is the case of OSM [63].

In the case of this research, the degrees of availability of the different cysteines were ignored, since the molecular structure of the desmoplakin plakin domain had not been determined when this investigation took place. The addition of 2 cysteine residues at the beginning of the amino acid sequence of the protein may have not accomplished it purpose, i.e. to ensure the presence of at least one available cysteine in a free thiolate state from the total 19 cysteine residues contained in the protein (17 in the wild type plus two engineered).

The results obtained for the plakin domain are comparable to previous results obtained for GFP [63, 203]. As mentioned in Section 1.4.3.1, immobilisation of GFP by size-selected clusters was attempted by using gold nanoclusters with 26, 40, 55 and 70 atoms pinned in different graphite surfaces. The arrangement of cysteines in this GFP protein comprised two cysteine residues along the backbone of its tertiary structure. The MSA characterisation indicates that one of these cysteines is completely buried within the protein structure, while the other is partially exposed. However, the exposed region of the second cysteine does not include the sulphur of the thiol group within the amino acid. This configuration implies no accessibility of sulphur to covalently bind to the gold atoms of the clusters.

The interaction of this protein with the gold clusters in buffer solution was first studied with contact mode AFM by Prisco *et al.* [63], revealing that GFP does not bind covalently to the gold nanoclusters since the protein molecules are easily swept by the scanning tip. This result demonstrates that the presence of cysteines in a protein does not guarantee its immobilisation by gold via a covalent bond. A second study of the same system was carried out by Lidgi-Guigui *et al.*, but this time using tapping mode AFM [203]. In this case, an enhanced weak interaction between GFP and the Au_{55} clusters was identified, representing hydrophobic interactions between both elements. The lack of covalent bond was corroborated with these studies, showing stable GFP islands under the action of the less disruptive tapping mode AFM tip, but not when scanning in contact mode.

As can be seen, contact mode AFM studies have demonstrated that, under the same conditions, (protein deposited into a graphite surface pre-decorated with gold nanoclusters and imaged in buffer solution) the plakin domain behaves similar to GFP. Based on experimental research on the subject and our results, it is possible to assert that there is no covalent bond between plakin domain and the gold nanoclusters notwithstanding the addition of the cysteine tags to the PD molecule. However, the stability of the protein features observed during tapping mode scan is evidence of a weak non-covalent interaction between the desmoplakin PD protein and the graphite pre-decorated with gold clusters.

TAPPING MODE MEASUREMENTS

The stability of the molecules under tapping mode scanning reveals an enhanced physisorption interaction between the PD molecules and the gold clusters, presumably involving hydrophobic interactions between the cluster-decorated surface and hydrophobic residues on the protein surface [204], similar to the physisorption regime in small molecule-surface interactions [203].

Even though no specific bond is associated with hydrophobic interactions, they play an important role in protein conformation and molecular self-assembly. Despite the common assumption of having the hydrophobic residues exclusively in the core of the protein, the accessibility of hydrophobic groups on protein surfaces has been proved in native proteins [205]. Moreover, gold has demonstrated interactions with hydrophobic polymers mediated by thiol modification of the gold surface, showing a weak affinity that enables the polymers to adsorb to its surface in a non-covalent mode [206]. Based on this evidence, it is possible to state that the hydrophobic interactions do not have to involve a total exposition of the hydrophobic core to the substrate surface [207]. Instead, proteins can be immobilised by non-covalent adsorption via hydrophobic interactions, potentially preserving native activity with only partial conformational changes [208]. In the case of the PD molecules, the suggested enhanced hydrophobic interaction with the gold clusters would not necessarily imply an exposition of the hydrophobic protein core, but the unfolding of the protein or structural changes upon adsorption have not been excluded [209].

The stability of the samples under tapping mode enabled a characterisation of the protein features in liquid phase. For aqueous TM-AFM measurements a contact mode cantilever, with force constant between 0.45 N/m and 5 N/m, (which is much lower than the non-contact cantilever with a spring constant around 40-50 N/m), was used in non-contact mode. This meant oscillating the cantilever at a frequency slightly above the resonance frequency, which in the case of contact-mode cantilevers is also much lower, around 155 kHz, instead of the 250 kHz of a non-contact-mode cantilever. Since the cantilever in this case is much less stiff than it should be for NC-AFM, the effect created is a larger deflection of the cantilever, causing the oscillations to be enough to tap the surface of the sample. In this way an equivalent to the tapping mode recommended for aqueous measurements of soft samples is obtained. Applying this technique, features

from samples with Au_{55} show a considerable increase in size that can be observed when going from dry to liquid phase measurements, both in diameter and height (Fig. 5.17).

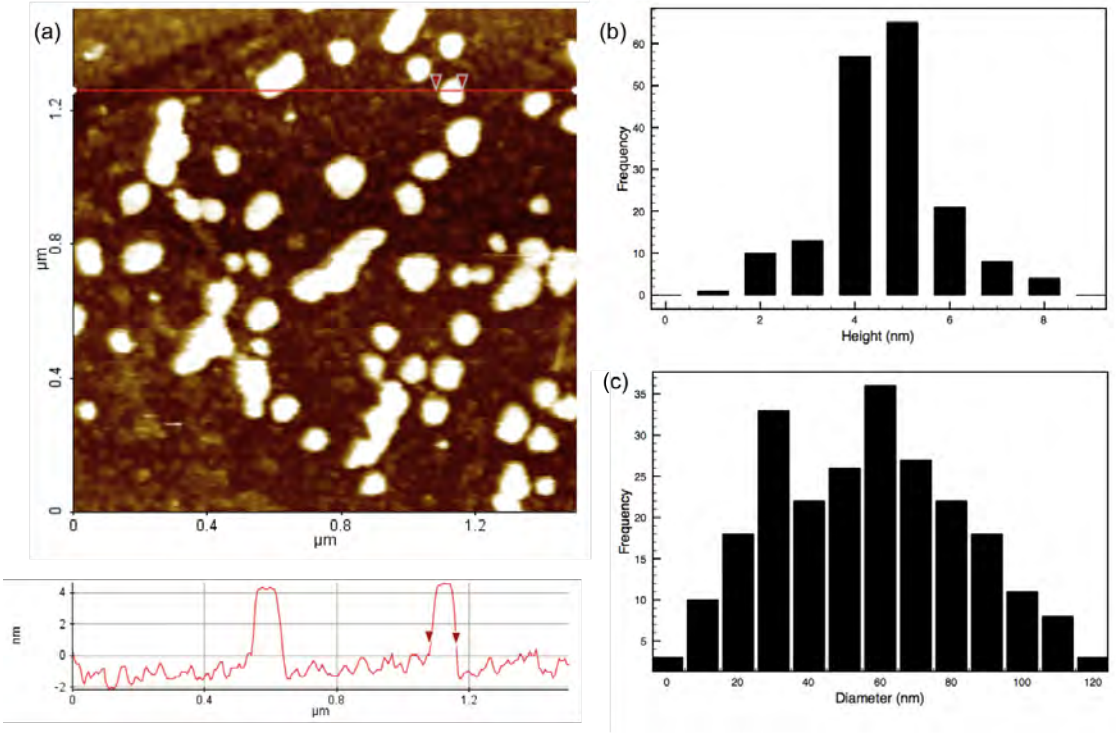


FIGURE 5.17: Desmoplakin plakin domain proteins on graphite pre-decorated with Au_{55} clusters imaged by tapping mode AFM in liquid phase (a) Liquid phase tapping mode AFM topography of PD on graphite with Au_{55} clusters. Some of the features are clearly round-shaped, and others seem to result from the overlapping of two or more round features. The lateral profile gives examples of typical heights of 5±0.1 nm. (b) Height distribution histogram with a mean value is 5.04±1.19 nm. More reliable height results are obtained from measurements in the liquid phase since factors such as dehydration or loss of structure are not considered. (c) Diameter distribution histogram generated considering the tip broadening effect due to the tip curvature radius of 10 nm. The mean diameter is 60.20±28.43 nm as a result of protein aggregation of molecules with a maximum diameter of 33.5 nm according to SAXS results [98]. Tip-broadening effects are also accentuated due to the properties of the imaging medium such as pH and viscosity.

The height distribution shows a mean of 5.04±1.19 nm (Fig. 5.17 (b)), which is in agreement with the width dimensions of the crystallised rod-shaped spectrin-like repeats SR3 to SR6, including the SH3 domain with dimensions of 4×4×18 nm [96]. The narrow height distribution suggests that most features correspond to protein structures of similar orientations, correlating to the lying down of the rods on the surface.

Regarding the diameter distribution, it is difficult to distinguish the characteristic tri-modality that has been appearing in the previous cases, showing a rather wide and fairly homogeneous distribution (Fig. 5.17 (c)). The local peak frequency at 35 nm can be

related to the maximum diameter of 33.5 nm determined for the PD by SAXS studies [98], while smaller features with diameter < 10 nm are possibly the result of few protein degradation products as a result of strong interaction with the graphite on a reduced number of molecules. Larger diameters, including the mean diameter of 60.2 nm and diameter vales > 70 nm presumably show the statistical nature of protein island growth process. It can be hypothesised that the enhanced interaction between one molecule and the graphite surface via the gold nanoclusters induces other particles to find the adsorbed molecule, attach to it via weak interactions, and therefore stop diffusion. The consequent formation of multimeric complexes has been favoured by the viscoelastic properties of soft samples, which contribute significantly to the topographic formation [32].

The occurrence of features attaching together, forming the larger non-globular shaped features that are easy to locate in Figure 5.17 (a), contributes to larger diameters within the distribution (Fig. 5.17 (c)). The results obtained demonstrate the formation of quasi-2D protein islands, with a wide range of diameters but roughly equal height. This implies a molecular growth process that takes place due to lateral diffusion of PD molecules to seed molecules bound to clusters, starting aggregation.

The size increment of features with respect to their dimensions measured in ambient conditions was also observed for proteins immobilised by Au_{147} clusters (Fig. 5.18 (a)) , on a slightly larger scale. The mean height is 5.51 ± 1.34 nm but the distribution has drifted to larger values, between 6 nm and 8 nm (Fig. 5.18 (b)), suggesting that lateral aggregation as only possibility for nucleation is not a direct conclusion in this case.

The diameter distribution also exhibits a local peak at 35 nm displayed on the histogram of Figure 5.18 (c). As in the case of PD deposited on Au_{55} clusters, this value is in agreement with the maximum diameter of 33.5 nm indicated by SAXS results [98]. Values above 70 nm are much less than for proteins immobilised with Au_{55}; in other words, the formation of non-globular features was rather low for samples with Au_{147}, in agreement with the results of dry measurements, so the justification, based on the cluster density, is likely the same, i.e. higher cluster densities result in less protein aggregation.

In both cases, Au_{55} and Au_{147} cluster samples, higher than expected values in height and especially in diameter are possibly explained by electrostatic interactions between the AFM tip and the features depending on the pH and salt concentration of the buffer

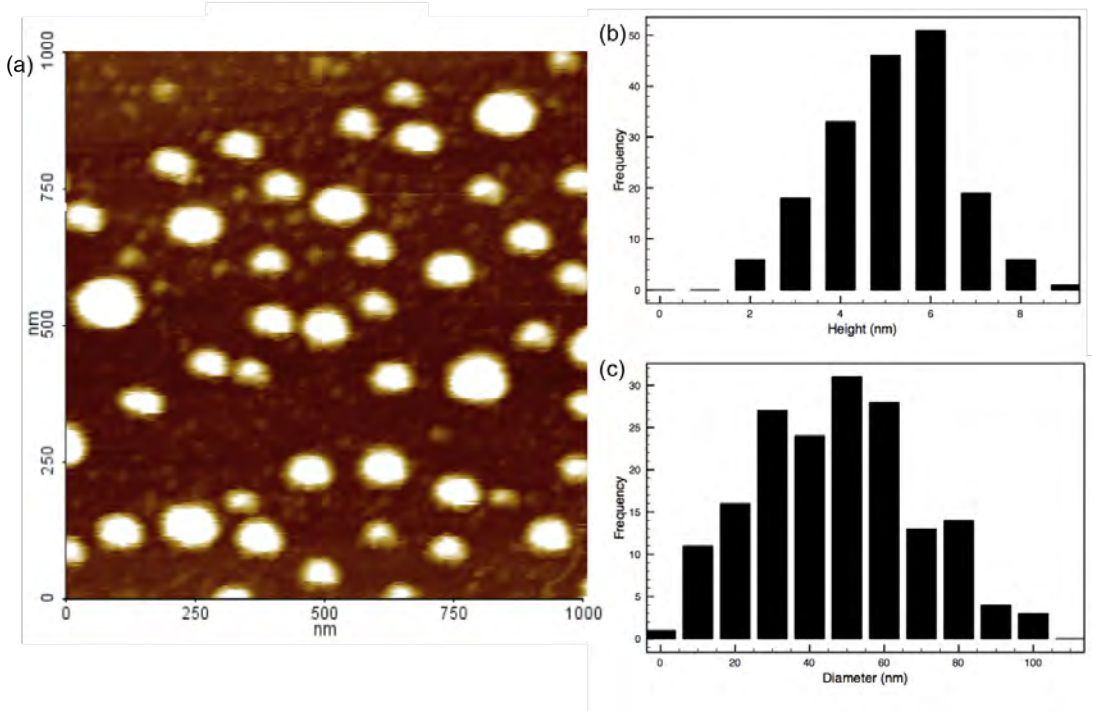


FIGURE 5.18: Desmoplakin plakin domain proteins on graphite pre-decorated with Au_{147} clusters imaged by tapping mode AFM in liquid phase (a) Liquid phase tapping mode AFM topography of PD on graphite with Au_{147} clusters. (b) Height histogram with mean height of 5.51±1.34 nm. (c) Diameter distribution generated considering the tip broadening effect due to the tip curvature radius of 10 nm. The mean diameter is 49.16±21.92 nm, which is lower than that found in the diameter distribution of features in the sample pre-decorated with Au_{55}. This could be indicative of a lower level of aggregation due to a higher density of clusters decorating the surface.

solution [209]. The AFM tip is a silicon surface with isoelectric point (pI) between pH 3 and pH 6 [209, 210], which translates in a negative charge in the buffer solution used for these experiments (20 mM tris HCl, 100 mM KCl, pH 7.5). On the other hand the desmoplakin plakin domain has a pI of pH 4.83, which is also below pH 7.5, and therefore it is also negatively charged. As consequence of this double negative charge interaction, the electrostatic repulsion between tip and the PD protein exceeds the van der Waals attractive forces, resulting in a distortion factor for height and lateral measurements. Regarding height measurements in particular, the better results obtained are plausibly due to the high salt concentration of the buffer solution, i.e. 100 mM KCl. Low ionic concentrations of buffer solutions have been demonstrated to promote exaggerated height measurements of biological samples [209]. On the contrary, sufficient salt concentrations act as a counterpart to the repulsive forces experienced by two equally charged surfaces, due to the reduction of the Debye length (the thickness of the double layer of counter ions formed at the charged surface) [49]. This condition is also thought to have a positive

effect in lateral measurements, but it is offset by the tip broadening effect.

A final consideration is that the desmoplakin plakin domain has been shown to exist in folded, partially unfolded and unfolded forms, under the same physiological conditions [91], which will increase the variation in height measurements. In this case, the difference in the dimensions would be a potential confirmation of denaturation translated into structural conformation modifications and systematic aggregation.

Tapping mode AFM measurements have given evidence that for both cases, samples pre-decorated with Au_{55} clusters and Au_{147} clusters, there is an enhancement of the binding of PD proteins on graphite via the gold nanoclusters. The effective influence of the gold clusters over the protein has been confirmed in native conditions, compared with a bare graphite surface under the same conditions, as exemplified in Figure 5.9 in Section 5.3.1.2. The comparison between the results of tapping mode AFM of bare graphite (the images of which do not show any protein attached to the surface) and tapping mode in graphite pre-decorated with clusters (the images of which reveal protein molecules bond to the surface) demonstrates that gold clusters are able to enhance adsorption of desmoplakin plakin domain proteins to the graphite surface, making possible an array of single proteins and protein multimeric complexes over a non-reactive graphite surface.

5.4 Overview of Chapter 5

During this work, the emerging technique for protein immobilisation by size-selected clusters on surfaces has been demonstrated and exploited in AFM studies of the plakin domain of desmoplakin.

Table 5.2 summarises the results obtained in liquid and dry phases using the three different AFM scanning modes for samples with full and reduced concentrations.

The nature of the interaction between the protein molecules and the gold clusters was explored by comparing bare and modified graphite surface with size-selected gold nanoclusters. At a concentration of 2.2×10^{-4} μg/μL of the protein, the non-modified area provides very few suitable binding sites for immobilisation; the protein only attaches along the line defects of graphite, thus the PD molecules are displaced by both contact and tapping mode AFM scanning. In contrast, the addition of gold nanoclusters to the

		Dry Phase	*Liquid Phase*	
		NC-AFM	TM-AFM	CM-AFM
HOPG	*Stock concentration*	**Aggregates of PD accumulated mainly on defects**	**ND**	**PD displaced by the AFM tip**
	Reduced concentration	**Single PD molecules on defects**	**PD displaced by the AFM tip**	**ND**
HOPG + Au clusters	*Reduced concentration*	**Single PD and PD complexes surface coverage**	**Single PD and PD complexes surface coverage**	**PD displaced by the AFM tip**

TABLE 5.2: Comparative table of the results obtained under the combination of conditions studied throughout this work. Stock concentration means 0.22 μg/μL, reduced concentration is equal to 2.2×10^{-4} μg/μL, ND means not done. Since the aim of the work is to study the possibility of immobilising single PD protein molecules, only reduced concentration of PD was used in samples of graphite pre-decorated with gold nanoclusters (samples with Au_{55} and Au_{147} included). Of special interest are the results in liquid phase as the intention is to study the protein in its native environment. Comparing the results in liquid phase for bare HOPG and HOPG pre-decorated with clusters, it is evident that the presence of the clusters enhances the interaction between the protein and the surface, achieving a surface coverage of single proteins and protein complexes over the cluster-decorated graphite. Tapping mode measurements were only taken for samples with a reduced concentration of protein.

graphite surface leads to the retention of proteins under TM-AFM (but not CM-AFM), and stability to be imaged with non-contact and tapping mode AFM in dry and liquid phases respectively, at low concentrations (2.2×10^{-4} μg/μL). The enhanced binding of the PD molecule to the clusters-decorated surface under physiological conditions (i.e. buffer solution) allows us to explore single protein molecules (associated with features of diameter = ∼35 nm) as well as their participation in protein complexes growth by tapping mode AFM.

Features found with more frequency by TM-AFM in both kind of samples, Au_{55} and Au_{147} clusters, have dimensions of ∼5 nm height and ∼60 nm or ∼35 nm diameter. Height dimensions are close to the width dimensions determined by X-ray crystallography structure determination of a segment of the desmoplakin plakin domain [96] that comprises SR3-SR6 including SH3, and which is reasonably adjudged to the whole plakin domain. Diameters of ∼35 nm are in agreement with the maximum diameter found in SAXS studies into the plakin domain in solution [98], and diameters of ∼60 nm could be therefore associated with dimeric formations of the desmoplakin PD promoted by the interaction with the surface.

It was observed that samples with lower cluster density showed a higher tendency to

protein aggregation, while samples with a higher cluster density showed more features associated with single proteins. From such an effect, and considering previous work done in this regard [62], it can be concluded that the density of clusters must be set in order to minimise the probability of protein aggregation. Low cluster densities generate protein aggregation due to limited number of binding sites, while too high cluster densities can be the cause of the same agglomeration effect but as a consequence of too short cluster-cluster distances, promoting lateral interactions between PD molecules. On the other hand, it was not possible to identify a cluster size-dependence for protein binding strength.

Limited morphological information about the desmoplakin plakin domain was extracted from AFM topographies. Although single molecules were imaged, submolecular structure was not discernible and an L-shaped protein was not addressed. Instead, a globular shape was constant in all kind of measurements, in agreement with electron microscopy studies of the desmoplakin head domains [91] and also in similarity with the plectin head domains [100].

Finally, a non-specific interaction, formed between the cysteine residues and the clusters, was confirmed. The absence of specific tethering suggests the unavailability of the cysteines on the protein surface (alluding to a possible disulphide bridge formed between adjacent cysteines), or the unavailability of sulphur atoms within the cysteine residues themselves.

Chapter 6

AFM Characterisation of Gold Nanocluster Immobilisation via Small Cluster Channels

This chapter is centred in the development of a new technique for the immobilisation of gold nanoclusters on graphite, work motivated by the experience gained during the research corresponding to Chapter 5. The experimental results and analysis are preceded by a brief description of previous methodologies for the immobilisation of metal clusters, including the argon sputtering technique, which will be afterwards compared with the new technique proposed here.

6.1 Metal Cluster Immobilisation: A Nanostructured Surface for Biological Systems

Single protein immobilisation on surfaces decorated with metal clusters has demonstrated in Chapter 5 to be a technique with potential for application in protein arrays and for single biomolecule studies. However, during its development a major difficulty was detected, residing in the impossibility of recognising the clusters with AFM while doing protein immobilisation experiments. The simultaneous detection of the two main elements playing a role in the immobilisation assays, i.e. clusters and proteins, is essential to the better understanding of the interaction between them.

Under physiological conditions for biological samples (i.e. buffer solutions), in order to obtain resolution better than 1-2 nm, the electrostatic environment must be carefully calibrated so as to compensate for short-range (van der Waals) and external forces that act on the AFM tip [30]. This is not always easy to achieve, leading to loss of detail in certain structural features of the proteins. In the particular case of immobilisation of proteins with nanoclusters, these resolution difficulties also result in the non-visualisation of the pinned clusters, whose heights are below 0.5 nm.

As consequence, the success of the functionalisation of graphite substrates has normally been confirmed by measuring the pinned clusters with STM prior to AFM imaging. Once the sample is transferred to the AFM, no clusters can be detected, and their presence is only known because both STM and AFM topographies are carried out in the same region of the sample. Consequently, before protein deposition, AFM topographies show a clean graphite substrate decorated with "invisible" clusters with an unknown location (Fig. 6.1). After protein deposition, AFM images show isolated proteins covering the graphite surface, but their exact relationship with the cluster positions is ignored.

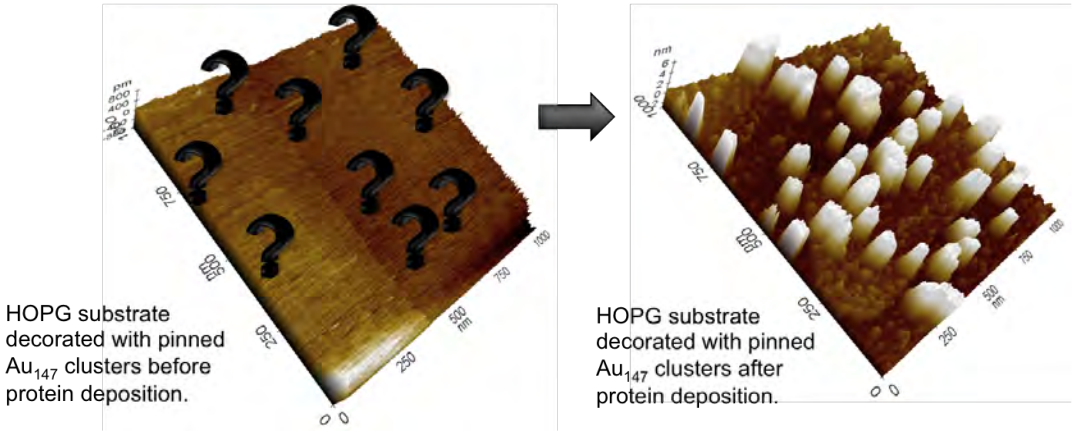


FIGURE 6.1: NC-AFM topographies of HOPG samples decorated with self-pinned Au_{147} clusters before and after protein deposition. Protein immobilisation experiments with AFM have used pinned clusters which are deformed due to the high deposition energies. With a reduced height, the clusters pinned on graphite used to be imaged by STM, prior protein deposition, in order to confirm the cluster-decoration of the graphite sample. After protein deposition, the samples are imaged with AFM, with which protein features are identified, but clusters are not recognised, due to their low heights. Therefore locations of clusters and proteins can not be compared.

Improved AFM experiments in protein immobilisation with metal nanoclusters would be facilitated by the implementation of clusters with an extra characteristic apart from being smaller than single proteins and strongly immobilised in the substrate: to be

easily detectable by AFM at room temperature and under physiological conditions as requirement for experiments with biological systems [211].

In the case of the investigation depicted in Chapter 5, the height of pinned clusters (reduced by the deformation caused by high deposition energies) fell below our microscope resolution threshold. Therefore the cluster characteristics of being firmly immobilised and visible for bio-AFM seemed to be irreconcilable, since strong immobilisation had been only achieved using the pinning deposition technique. In order to bring all these characteristics together, we set the task of developing a new cluster deposition technique that allows immobilisation of clusters of larger dimensions, so they can be easily visualised by AFM in liquid or air. This was accomplished via the creation of channels by small cluster implantation. Clusters with less than 1000 atoms, which fulfil the original requisites, have been visualised with AFM in non-contact mode (NC-AFM), making possible an AFM characterisation with high-resolution images. Tapping and contact mode AFM imaging have demonstrated strong immobilisation of such gold clusters in the substrate.

6.2 Metal Clusters in Surfaces

In the absence of external forces, the relative movement of metal clusters on a substrate derives primarily from the cluster/surface interface properties, as existing adhesion or friction [212]. Those properties dictate the kinetics and therefore diffusion rate of the nanoparticles on the substrate. Gold clusters on graphite adsorption and diffusion barrier energies have particularly been investigated via *ab initio* calculations based on the cluster's epitaxial orientation on the substrate [213], and also tested experimentally [214]. In this system it is possible to differentiate between two kinds of states that constitute the diffusion process. One state is referred to as the commensurate interface, described by a nanocluster whose contact facet with the surface is in geometrical agreement with respect to the superficial graphite layer atomic sites angles. Such a cluster is in the energy minimum of the interaction potential that governs the interface system [215]. A state of incommensurate interface would be the second state, in which the nanocluster is displaced from the geometrically matching position with the graphite surface undergoing a low friction configuration that allows it to move over the surface until it finds angle registry again with the graphite structure [212].

Based on the accumulated knowledge of the interface between supported metal nanoparticles and their respective substrates it has been possible to develop diverse techniques of nanomanipulation of clusters on substrates. Deposited nanoparticles can be manipulated by different techniques, such as driving force, or by preferential nucleation sites specific to the substrate in question [216], while immobilisation methods have exploited various of their properties such as their negative surface charge and, in the specific case of gold nanoparticles, their affinity for thiol groups [217].

6.3 Previous Cluster Immobilisation Techniques: Obstacles for AFM Measurement

6.3.1 Pinning Technique

Typical methodology for protein immobilisation with gold clusters uses the pinning deposition technique, as in the case of the experiments reported in Chapter 5.

As explained in Section 2.9.2, in a metal-covalent system, which is the case for gold clusters on graphite, the deposition energy determines the cluster-substrate interaction. Thus, by using a deposition energy lying on the pinning range [58], it is possible to immobilise the clusters at their impact sites, inhibiting diffusion. The result is a nanostructured surface free of cluster aggregation or cluster migration to the natural surface defects. But the high kinetic energy also has less favourable consequences, as it causes deformation of the clusters, reducing their height after deposition (0.44 nm for a gold cluster with 147 atoms measured by STM (Fig. 5.10), when the theoretical diameter for this size of gold cluster is 1.72 nm [218]).

From the deposition categories, only soft-landed gold clusters preserve their morphology, since their height is not affected by the deposition energy, being easily imaged by AFM. However, without a proper anchor, they diffuse on the substrate at a rate of about 10^{-8} cm^2/s (for clusters with hundreds of atoms) induced by the internal vibrations of the clusters and/or the vibrations of the substrate [213, 214, 219], and tend to agglomerate on surface defects, being useless for immobilising proteins.

6.3.2 Defects created by Argon Sputtering Technique

As a solution to the dilemma, surfaces with defects created by argon bombardment appear as an adequate platform that immobilises soft-landed clusters [220]. This methodology consists of two steps: the preparation of a surface with atom defects produced by Ar^+ sputtering and soft-landing deposition of metal clusters. After deposition of clusters at low energies (500 eV), clusters diffuse on the surface until they find an atomic defect site produced by the argon ions sputtering, that works as an anchor which immobilises them. Coverage of soft-landed clusters is half the argon ions coverage, ensuring a dispersed array of clusters and reducing to the minimum the statistically inevitable cluster aggregation for high particle densities.

Another advantage over the pinning method is the flexibility that this technique offers regarding cluster sizes. For large clusters (>500 atoms), the impact energies required for self-pinning are normally an impediment due to the linear dependence between the pinning threshold and the number of atoms in the clusters, as explained in Section 2.9.2. With the use of pretreated surfaces covered with atom defects produced by argon bombardment (called simply *argon defects* from now on), immobilised clusters can contain up to 2700 atoms, as proven experimentally [220].

The stability at room temperature of metal clusters immobilised by this method has been validated by STM measurements. However, there is evidence that the bonding is not as strong as that resulting from the pinning deposition technique, causing the tendency of the clusters to be displaced by an STM tip with slightly higher voltage and current [220].

So, in conclusion, clusters immobilised by argon defects have a clear advantage over the pinning method regarding SPM measurements: that clusters can be soft-landed and thus visible to bio-AFM. On the other hand there is also a big disadvantage: that clusters are not strongly immobilised and thus easily removed from the surface by a scanning tip.

6.4 Small Clusters Channels: A New Method for Clusters Immobilisation

Given the above scenario, a different and new deposition technique, involving the implantation of small clusters, has been developed and implemented.

Although the final purpose of the immobilisation of clusters is in this context the posterior immobilisation of proteins, experiments were performed in air instead of liquid with the purpose of understanding the technique before submitting it as an adequate immobilisation method for proteins. The advantage here resides in the fact that imaging resolution in liquid under regular conditions (non-vacuum, room temperature) is easier to achieve than in a humid environment as it is air, due to the minimisation of tip-sample interacting forces [221]. Therefore, accomplishing high resolution imaging in air also presupposes high resolution in the liquid phase.

6.4.1 Creation of Channels by the Implantation of Small Gold Clusters

For immobilisation of biological material, the intention of implanting metal clusters on graphite is to functionalise a substrate as a preceding step for decoration with immobilising agents for proteins.

Implanted clusters lose their geometrical properties as a consequence of the deformation suffered during the deposition process. The applied kinetic energy is high enough (>50 eV) for the clusters to break the C-C bonds on the surface layers and to become embedded in the substrate. The result in each individual case is a disordered cluster inserted into the graphite and a corresponding tunnel with a characteristic depth [222]. The cluster is embedded in the substrate, reducing availability for interaction with the biomolecules. On the other hand, the tunnel obtained is the basis of a powerful protein immobilisation technique, which is explained in this section.

Prior molecular dynamics (MD) studies into the interaction of size-selected clusters and graphite at different energy depositions have demonstrated that there is a defined implantation behaviour for metallic clusters in the size range N = 20 - 200 implanted on graphite with kinetic energies ranging 0.75 to 6 keV [223]. Above a certain threshold energy deposition, the clusters impact on the substrate and penetrate successive graphite

planes creating a tunnel with a depth that is a function of the kinetic energy and of the cross-sectional area of the cluster. Other MD calculations, supported by experimental results, suggest that in the case of gold, silver and silicon clusters with 7 atoms deposited on graphite at energies ranging between 1 and 5.5 keV the implantation depth ID_{max} follows a linear dependency with the momentum p of the clusters [224]:

$$ID_{max} \propto \frac{p}{N^{2/3}} \tag{6.1}$$

where the momentum p can be calculated in function of the number of atoms N in the cluster and the implantation energy E_k as:

$$p = \sqrt{2m_0 E_k N} \tag{6.2}$$

By understanding this relationship, and considering that the adsorption energy of gold on graphite has a negligible effect on implantation depths (of order 0.45 eV) [225], it is possible to manipulate the depth and size of the pores or defects on the graphite substrate, leading to the possibility of generating applications in single molecule physics, chemistry, and biology [224].

6.4.2 Defects Created by Implantation of Small Clusters as Platforms for Cluster Immobilisation

In this work, the creation of channels in graphite is used to modify the substrate that will interact with certain biological material via larger gold clusters immobilised by such channels. The idea is to anchor the Au_n clusters (with hundreds of atoms) which will immobilise the proteins with channels created by the implantation of smaller Au_{20} clusters (Fig. 6.2). This will inhibit soft-landed Au_n clusters from diffusing, resulting in a graphite substrate decorated with stable immobilised gold clusters which have not lost their morphological properties and are also visible for AFM resolution.

The gold clusters for soft landing (Au_n) were selected to contain n = 147, 309, 561 and 923 atoms so that their dimensions were smaller than those of single protein molecules, and also in order to achieve maximum chemical stability.

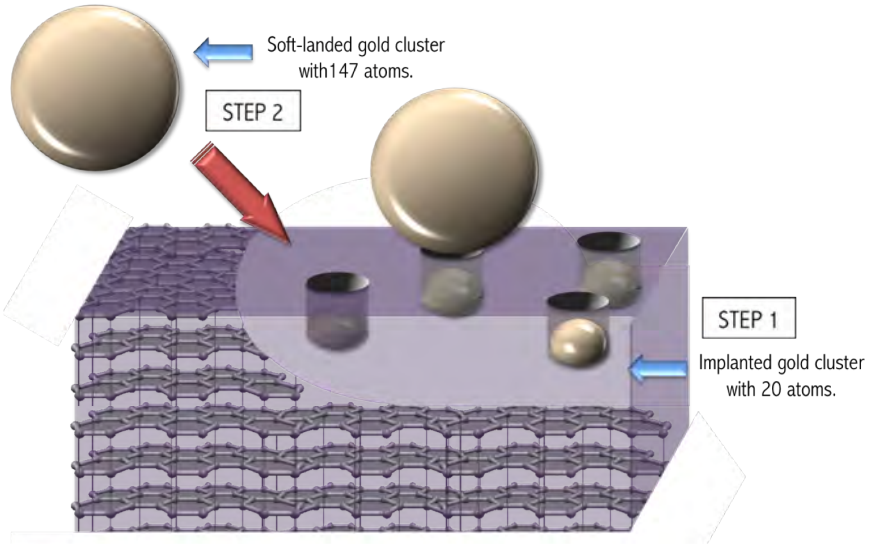


FIGURE 6.2: Illustration of the concept of immobilisation of clusters via the channels created by the implantation of small clusters. STEP 1 refers to the implantation of small clusters in the substrate (Au_{20} into HOPG) at 1.5 keV. The implantation depth will be determined by the kinetic energy of the clusters and the cross sectional area (given by the number of atoms in each cluster). STEP 2 refers to the deposition of larger clusters (Au_n, n=147, 309, 561, 923) using the soft-landing technique, with a specific deposition energy of 0.5 keV.

In synthesis the proposed methodology for protein immobilisation consists of three steps, from which we are only concerned in **I** and **II** in this chapter:

I.- Implantation of small clusters (20 atoms) to create a functionalised graphite surface able to immobilise larger clusters.

II.- Soft-landing of larger clusters (hundreds of atoms) to decorate the graphite surface by being immoblised in a mono-disperse array for protein binding.

III.- Deposition of a volume of protein solution equivalent to a submonolayer to obtain a single molecule protein array in the graphite surface.

In order to test the efficiency of this new technique two different kinds of samples were prepared: graphite (HOPG) treated with argon sputtering, and graphite (HOPG) implanted with Au_{20} clusters. Both were used as substrates for larger gold clusters and their respective performances were explored and compared with AFM in non-contact and contact mode. This evaluation method allowed us to accomplish a detailed examination of the effects of the two different immobilisation techniques for each size of cluster here analysed, as well as the characterisation of gold cluster dimensions with high-resolution for all sizes using the two different immobilisation techniques.

6.5 High Resolution AFM Characterisation of Size-Selected Gold Clusters

6.5.1 Gold Clusters Immobilised by Argon Defects

We prepared graphite samples treated with argon sputtering to create defects on the surface, intended to attach the clusters following soft-landing deposition. Graphite samples were pre-treated with an Ar^+ ion beam at 500 eV produced in the NPRL cluster source at a coverage of 2000 pA·s creating a point defects array. Following the two-step process described in Section 6.3.2, the samples with Ar defects were the substrate for gold clusters with hundreds of atoms.

Size-selected gold clusters with 147, 309, 561 and 923 atoms were produced by the NPRL cluster source and deposited on the prepared substrates at 0.5 keV with approximately half the coverage of the Ar ions using a current of 1000 pA·s. Soon after, each sample was imaged and analysed with atomic force microscopy in non-contact mode at room temperature. In every case the clusters proved to be stable under NC-AFM at room temperature, in a similar way as large silver clusters in other findings [220]. The achieved stability of AFM imaging allowed us to obtain high resolution images of gold clusters with a number of atoms down to 147 (Fig. 6.3).

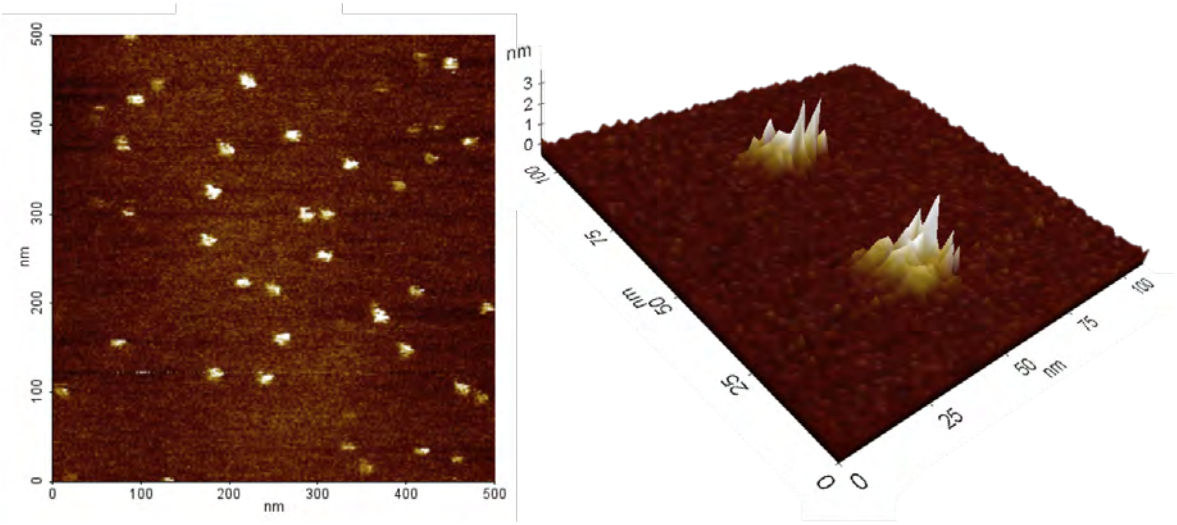


FIGURE 6.3: NC-AFM images of Au_{147} on HOPG decorated with Ar defects with deposition energy of 0.5keV. The average height, 1.60±0.42 nm, is very close to the Au_{147} theoretical diameter: 1.72 nm proving that the so-called soft-landing deposition technique does not deform the original clusters symmetry as the self-pinning technique does due to the high deposition energy. The argon defects stop the gold clusters from diffusing, making them visible and stable under NC-AFM.

High-resolution AFM vertical measurements of cluster dimensions result in grater accuracy than STM measurements, since the later depend on the voltage settings at the time of measurement (it measures the electronic states of the clusters instead of the real height) and this generates deviations from the actual height of the clusters [226] (Fig. 6.4).

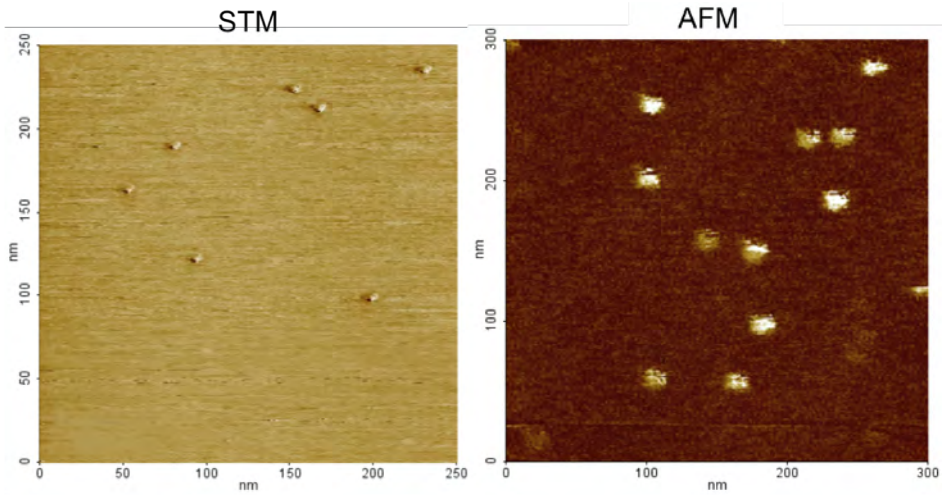


FIGURE 6.4: Comparison of STM (250 μm x 250 μm) and AFM (300 μm x 300 μm) images of Au_{147} on graphite to exemplify high AFM resolution at room temperature and non-UHV. While STM provides higher resolution for lateral measurements, the accuracy for vertical measurements provided by AFM is higher since the latter does not deviate from the real high measurement as a function of the voltage.

Vertical and lateral measurements of the Au_n clusters dimensions were acquired in non-contact mode AFM. The tip broadening effect inherent in AFM lateral resolution, which makes the lateral distances a convolution of the tip radius and the feature diameter, must be considered, while for vertical measurements the minimum experimental vertical error was 0.4 Å(Figures 6.5 and 6.5).

6.5.2 Gold Clusters Immobilised by Small Clusters Channels

Au_{20} clusters were produced by the NPRL cluster source and implanted with an energy of 1.5 keV and a current of 2000 pA·s on the HOPG surface. The implantation of the small clusters creates nano-channels on the graphite substrate with depth dependent on the size and kinetic energy of the Au_{20} clusters and with diameter equal to the cluster diameter. MD simulations have indicated that in the case of silver, with these parameters, the implantation depth ID is about 1.2 nm, in accordance with the linear function $ID \sim E/N^{2/3}$ [223]. We predict a slightly larger implantation depth for gold

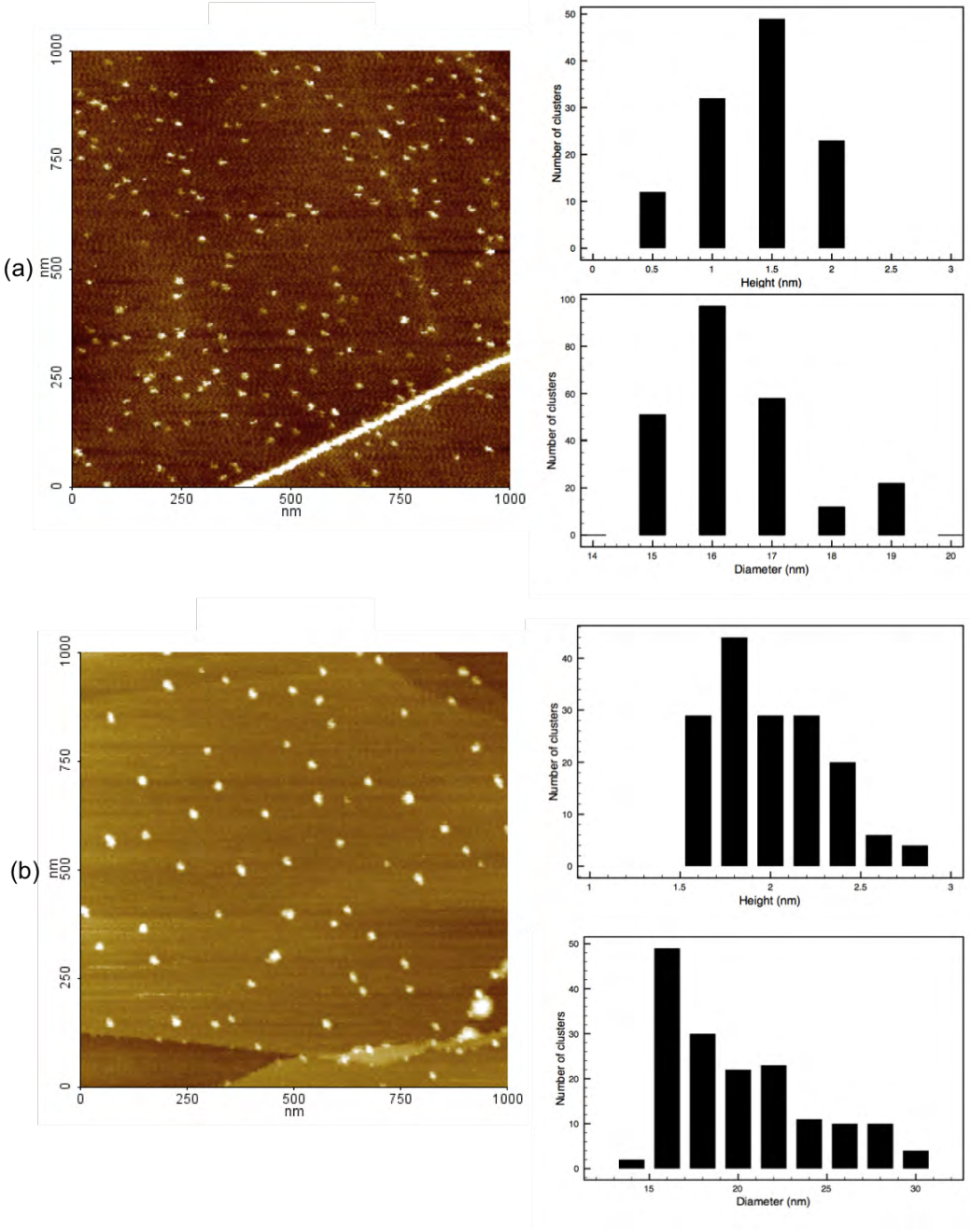


FIGURE 6.5: NC-AFM topographies of size-selected clusters soft-landed at 0.5 keV on HOPG decorated with argon defects. (a) Au_{147}. Average height = 1.60 ± 0.42 nm, average diameter = 16.92 ± 1.08 nm. (b) Au_{309}. Average height = 2.09 ± 0.30 nm, average diameter = 21 ± 3.92 nm. For the diameter distribution no subtraction of the AFM tip sectional radius was performed.

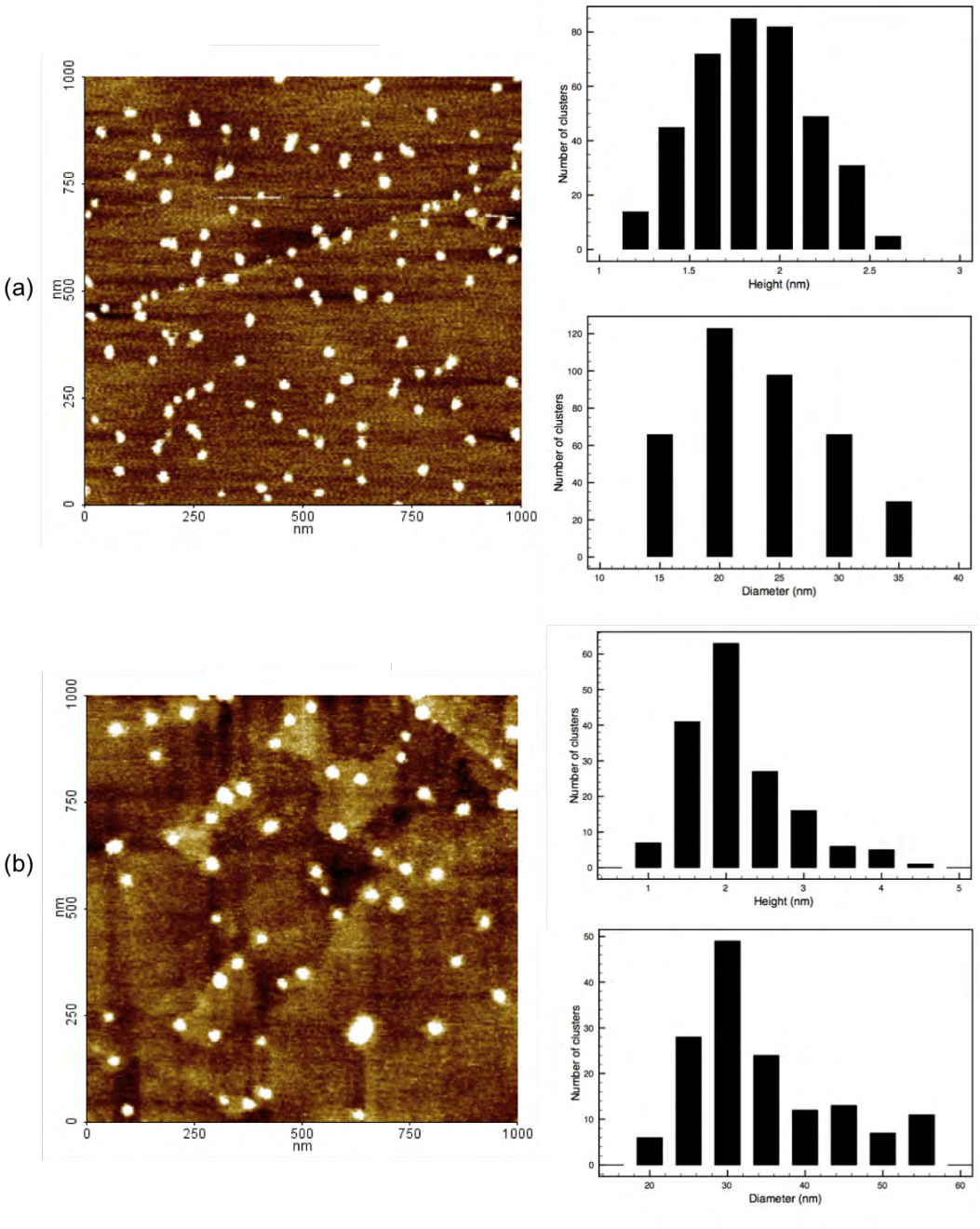


FIGURE 6.6: NC-AFM topographies of size-selected clusters soft-landed at 0.5 keV on HOPG decorated with argon defects. (a) Au_{561}. Average height = 1.95 ± 0.31 nm, average diameter = 26.23 ± 5.65 nm. (b) Au_{923}. Average height = 2.36 ± 0.66 nm, average diameter = 37.23 ± 8.84 nm. For the diameter distribution no subtraction of the AFM tip sectional radius was performed.

clusters with the same number of atoms, given the difference in mass that varies the maximum energy transfer in a binary collision with a C atom at rest. Similar adsorption energy on graphite of gold and silver [225] indicates that the actual implantation depth for Au_{20} is not too far above the mentioned depth given for implanted silver clusters.

Au_{147}, Au_{309}, Au_{561} and Au_{923} clusters were produced by the NPRL cluster source and soft-landed on the pre-treated substrates using the same parameters as for the clusters deposited in the substrates with argon defects, an energy of 0.5 keV in every case, and the coverage was approximately half the coverage of the implanted Au_{20} clusters using a current of 1000 pA·s. The stability of the soft-landed Au_n clusters was proved with NC-AFM. A full characterisation of the cluster dimensions was completed for all cases, paying special attention to the height since that is the measurement that provides more accurate information about the clusters (Fig. 6.7). The results of the NC-AFM characterisation are in the Table 6.1 showing the Au_n cluster heights for clusters deposited on HOPG decorated with argon defects and Au_n clusters deposited on HOPG decorated with channels produced by the implantation of Au_{20}.

Au CLUSTER	THEORETICAL DIAMETER (nm)	REPORTED HEIGHT (nm)	Ar SAMPLES AVERAGE EXPERIMENTAL HEIGHT (nm)	Au_{20} SAMPLES AVERAGE EXPERIMENTAL DIAMETER (nm)
Au_{923}	3.45	3.1	2.36 ± 0.66	2.12 ± 0.53
Au_{561}	2.88	2.6	1.95 ± 0.31	1.84 ± 0.48
Au_{309}	2.30	2.3	2.09 ± 0.30	1.76 ± 0.55
Au_{147}	1.72	1.6	1.60 ± 0.42	1.58 ± 0.42

TABLE 6.1: Comparison between experimental heights obtained from samples with argon defects and samples with small cluster channels and theoretical diameter for different sizes of gold clusters. The theoretical diameter is calculated using the bulk lattice spacing of gold, 4.08 Å, and the consequent nearest neighbour distance of 2.88 Å. Reported height refers to the experimental results published for the height of each different size of cluster under similar conditions [218, 227–229]. Both experimental heights of clusters on argon samples and clusters on samples with Au_{20} are very close to the theoretical and reported height, specially for clusters with a lower number of atoms.

The theoretical diameters were calculated based on the spherical approach, considering that 923, 561, 309 and 147, being geometrical magic numbers for gold clusters, form clusters of exactly 6, 5, 4 and 3 shells respectively. The bulk lattice spacing of gold, 4.08 Å, and the consequent nearest neighbor distance of 2.88 Åwere used for the calculation. The reported diameters were taken from the appropriate and relevant literature [218, 227–229] and we can compare them with the experimental results here presented for

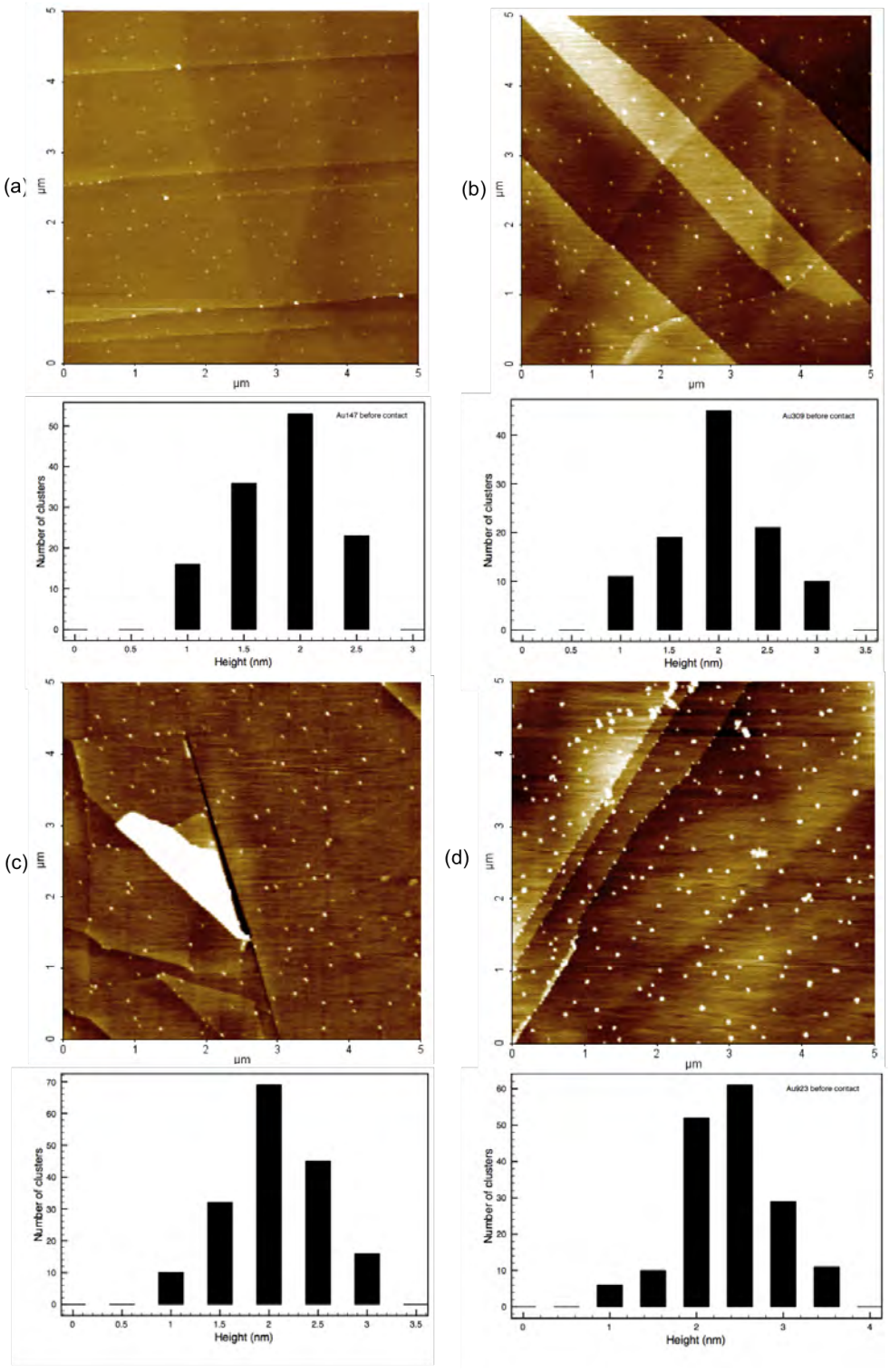


FIGURE 6.7: NC-AFM topographies of size-selected clusters soft-landed at 0.5 keV on HOPG decorated with implanted Au_{20} clusters (with implantation energy of 1.5 keV) (a) Au_{147}. Average height = 1.58 ± 0.42 nm, theoretical diameter: 1.6 nm. (b) Au_{309}. Average height = 1.76 ± 0.55 nm, theoretical diameter: 2.3 nm. (c) Au_{561}. Average height = 1.84 ± 0.48 nm, theoretical diameter: 2.6 nm. (d) Au_{923}. Average height = 2.12 ± 0.53 nm, theoretical diameter: 3.1 nm.

vertical measurements of clusters in the two different kind of substrates, demonstrating that both values are very close to the theoretical and the reported height in every case, especially for Au_{147} clusters. In the other three cases the diameter measured is slightly lower than the reported values, but in all cases the difference is less than 8 Åfor argon samples and 1 nm for Au_{20} samples. This discrepancy can be explained by the way the soft-landed clusters interact with the substrate. Gold clusters deposited on graphite with the soft-landing technique have shown to have a hemispherical shape given the deformation suffered by impacting on the surface. According to molecular dynamic simulation in this kind of systems, a flattening effect of the cluster after impact at low energies is minor but inevitable [133, 230]. The accuracy of the results, with respect to the reported values, decreases with the number of atoms in the Au_n clusters, the Au_{147} average height being closest to the expected diameter, and the Au_{923} average height diverging most from the expected number, both, experimental and theoretical. This behaviour is opposite to what would be expected, since for clusters with a larger number of atoms the cluster-surface interaction energy is lower than the kinetic energy, so larger clusters do not undergo major restructuring and tend to preserve the properties of a free particle to a higher degree [230].

6.6 Contact-Mode AFM Effect on Clusters: Testing Immobilisation Efficiency of Argon Defects vs. Small Clusters Channels

As mentioned before, preventing clusters with preserved geometry from diffusion is fundamental to the development of engineered substrates for protein immobilisation (Fig. 6.8). Therefore, a systematic experimental study using contact mode AFM has been conducted in order to compare previous cluster immobilisation techniques using argon sputtering, with the new technique based on defects created by the implantation of small clusters.

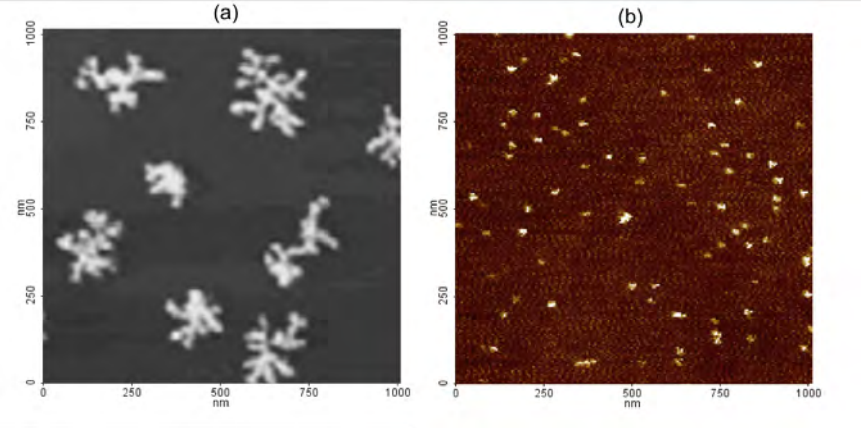


FIGURE 6.8: Comparison between soft-landed small gold clusters deposited in HOPG and pre-treated HOPG. (a) AFM image of Au_{750} cluster deposition on HOPG graphite at room temperature. Clusters diffuse and aggregate forming ramified islands of clusters [231]. (b) NC-AFM 500×500 nm topography of HOPG decorated with Au_{147} clusters soft-landed on argon defects. The defects stop clusters from diffusing and aggregating. This comparison illustrates the efficiency of defects for anchoring clusters.

6.6.1 Effect of CM-AFM on Clusters Immobilised on HOPG with Argon Defects

The strong immobilisation of clusters on a substrate is a condition that can be tested with contact mode AFM. Contrary to non-contact mode AFM, which is based on the vibration of the AFM cantilever close to its resonant frequency, lying in the attractive force regime (distance range of 10 - 150 Åfrom the sample) [184] and resulting in practically no contact with the sample surface (Section 5.1.1.2), contact mode AFM exerts a lateral force on the sample to scan, having an effect on the supported size-selected small clusters, by disturbing their stability when even a minimum magnitude force is applied [232].

In the case of Au_n soft-landed clusters immobilised by argon defects this result has been demonstrated by scanning one sample three times in a conserved region: first a 2×2 μm topography image in non-contact mode AFM was acquired, then a sub-area within the first scanned area was scanned in contact mode AFM (using an applied force of 1 nN), and finally a 2×2 μm scan again in non-contact mode AFM. The first scan resulted in an image that confirms the stability of the clusters under NC-AFM, the outcome of the second scan was a topography that shows no presence of clusters, and the third scan shows the final result: clusters that have not been exposed to CM scan are undisturbed, while clusters in the sub-area subjected to one CM scan have been displaced by the action of the AFM tip. Contact scanning applying a constant force of 1 nN was executed for all

Au_{147}, Au_{309}, Au_{561} and Au_{923} samples with argon defects, showing in every case that only one scan at this force is able to remove all, or almost all clusters from the scanned region (Fig. 6.9), which can be called a "contact window" for future reference.

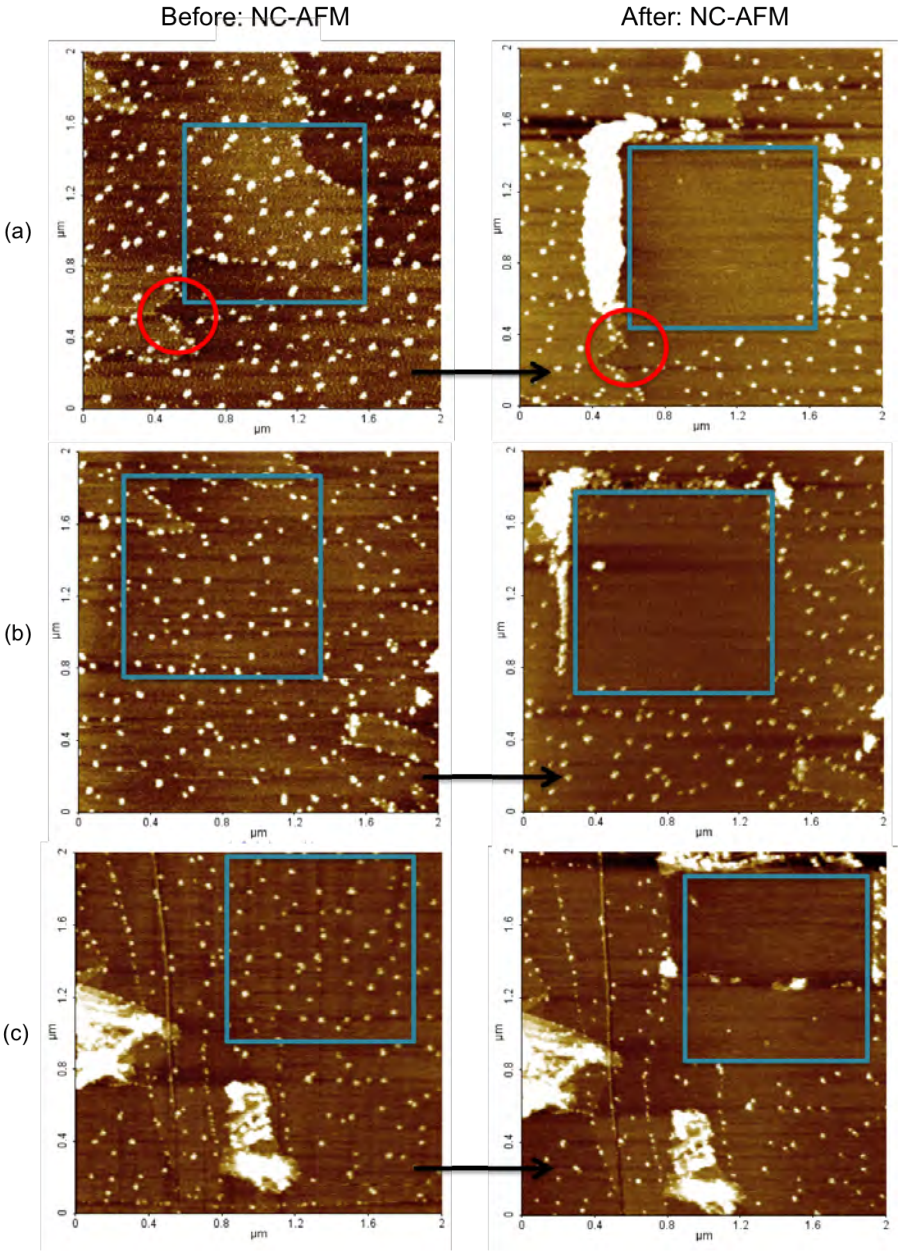


FIGURE 6.9: Left images of (a) Au_{923}, (b) Au_{561} and (c) Au_{309}, show a 2×2 μm scan of each respective sample taken in NC-AFM to confirm the stability of the clusters deposited on graphite decorated with argon defects. Right images show a 2×2 μm topography acquired in NC-AFM after a scan in CM-AFM was performed in a 1×1 μm subregion in the same area. It is clear how the clusters outside the scanning region conserve their positions, while the clusters inside the 1×1 μm area have been removed from their original positions and dragged to the edge of the scanned zone, where they accumulate in a vertical line. This demosntrates that the clusters are not strongly immobilised by the argon defects. The arrow indicates temporal order.

At higher normal forces than the threshold provided by the weak immobilisation effect that the argon defects have on the clusters, the friction between these and the AFM tip is stronger than the interfacial friction existing between them and the HOPG surface [232]. This causes the applied lateral force to become high enough to displace the Au_n clusters from the argon defects. The displacement vector is parallel to the fast scan direction, which is why it is possible to a see large accumulation of dragged clusters specially in the right and left edges of the scan area in contact mode.

6.6.2 Effect of CM-AFM on Clusters Immobilised on HOPG with Channels created by the Small Clusters Implantation

Graphite samples implanted with size-selected gold clusters of 20 atoms were imaged and analysed with NC-AFM and CM-AFM. The outcome was the same for every size: clusters endure both AFM modes.

Non-contact mode was first used to image the samples, corroborating that all sizes of clusters are stably immobilised by the implanted cluster defects after being imaged several times over a long period of time. Contact mode is then used to subject the clusters to an effective applied force of 1 nN and ascertain the strength of the attachment to the modified substrate. During CM-AFM scan no clusters are visible on the HOPG surface. Given that the contact mode scans have been taken in the same regions as the non-contact mode imaging (which have showed high density of clusters), the absence of clusters leads us to think that the clusters have been completely removed from their original sites. But by performing an immediate subsequent NC-AFM imaging of exactly the same area that has undergone contact mode, it is revealed that the clusters have remained in their original positions after being scanned in contact mode. This behaviour was repeated in all cases without exception (Fig. 6.10).

The apparent invisibility of the clusters when scanning in contact mode is a question of resolution, and it explains as follows. As previously described, in order to carry out the experiments presented here it is necessary to pass from non-contact mode to contact mode and again to non-contact mode in a systematic way. These continuous changes of modes would, in principle, imply continuous changes of cantilevers since ideal cantilevers for each mode have different characteristics. Nevertheless, changes of cantilever in between scans bring considerable difficulties when trying to image the

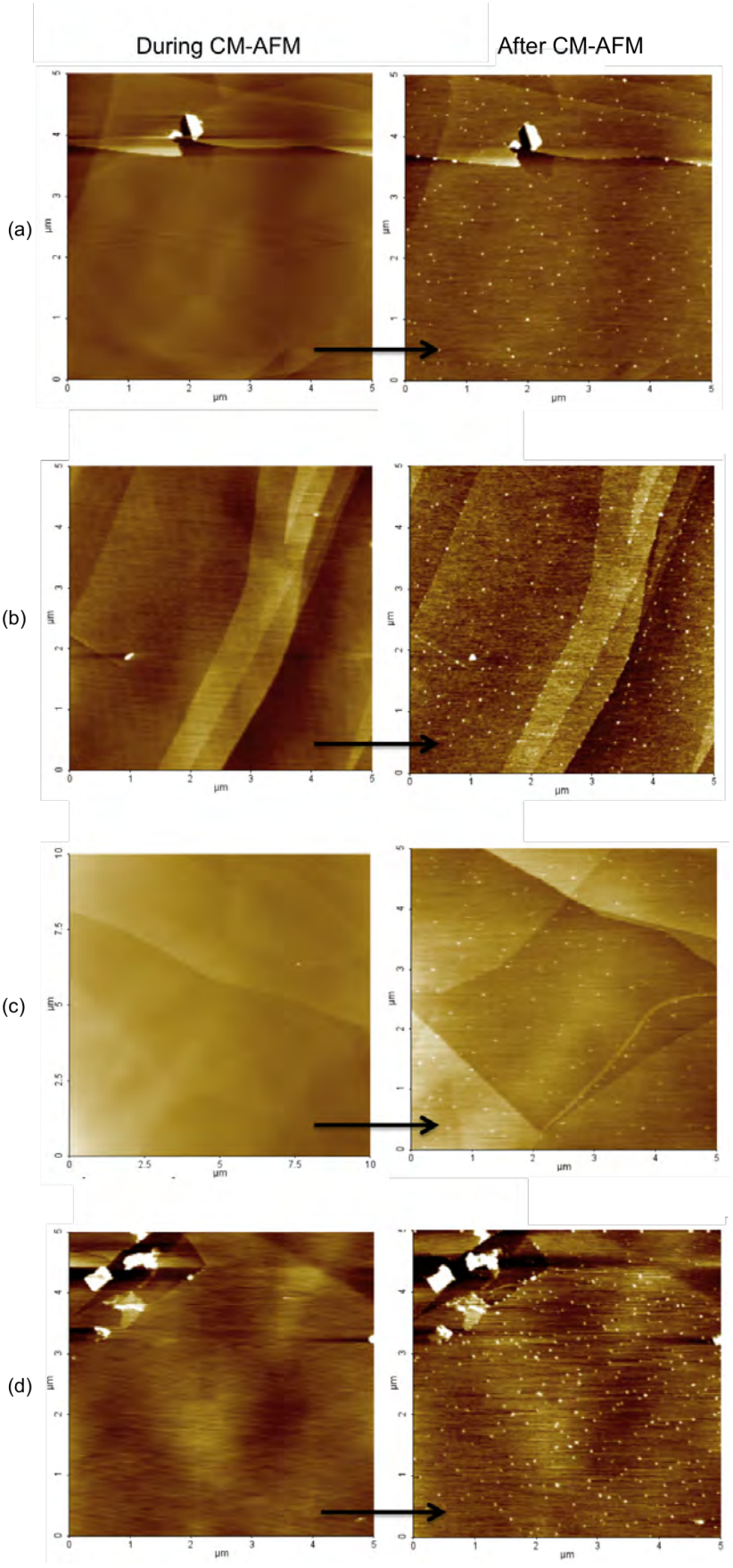


FIGURE 6.10: Result of applying a contact mode scan to the samples. Left images of (a) Au_{923}, (b) Au_{561} and (c) Au_{309}, (d) Au_{147} show topographies of each respective sample taken in CM-AFM. In all cases no cluster is imaged. But after an immediate NC scan in exactly the same region, it is possible to see how a large number of clusters remain in the CM-AFM scanned area ("contact window"), as displayed in the right images for each respective sample. The arrow indicates temporal order.

same area, hence such changes were avoided in the course of these experiments. This means that one type of cantilever had to be chosen for both contact and non-contact measurements.

Given that the studies realised in this work are focused in the final state of the clusters, i.e. after contact mode, priority was given to the final non contact mode imaging that reveals whether the clusters have been removed from their attachment sites by the action of the CM-AFM tip or not. This means that all scans, NC and CM, were performed with a high frequency cantilever (260-410 KHz, as specified in Section 2.8.2) and with a large spring constant (21-78 N/m). These characteristics are optimal for NC mode, but CM-AFM imaging needs probes with small spring constant (around 1 N/m) in order to detect small deflections and therefore to have high resolution. Since a cantilever with large spring constant was used for the contact mode scans, the resolution obtained was not high enough to visualise the clusters during the scan. However the visualisation of the clusters during CM scan was not one of the objectives of this research, whereas the clear identification of the region subjected to CM scan was critically important for enabling posterior NC-AFM analysis of the effect of the CM tip on the clusters. This was achieved by comparing a CM topography with the corresponding subsequent NC image as shown in Figure 6.10. A clean HOPG surface indicates that such a region has been subjected to CM scan and allows a direct comparison with a successive NC mode topography, normally because of a distinctive graphite defect.

After analysing the acquired topographies it is possible to deduce that the channels defects trap the soft-landed Au_n clusters, inhibiting diffusion, following the same principle as the argon defects, but the larger dimensions of the nano-channels, in both depth and diameter, permit the Au_n clusters landed on the surface to be trapped more easily, limiting further diffusion. A possible explanation for this is that the nano-channels give Au_n clusters stability by generating a higher probability of a state of commensurate interface between the cluster and the graphite [215]. The stronger interaction works as a barrier for diffusion that exceeds the surface vibrations and/or the internal vibrations of most of the clusters [213] and moreover, it also exceeds the force exerted by the AFM tip in contact mode during one scan at 1 nN. The force required to shear the cluster-substrate interface is larger than the lateral force needed to shear the cluster-tip interface, having as consequence the permanence of the clusters in their sites after CM-AFM scanning [76]. However, some of the clusters are dragged by the CM-AFM scanning tip, forming

a line of clusters at the edges of the scan area, in the same way as for the argon defects substrates, but with a lower number of occurrences, as shown in the topography of Au_{923} clusters on Au_{20} defects (Fig. 6.11)

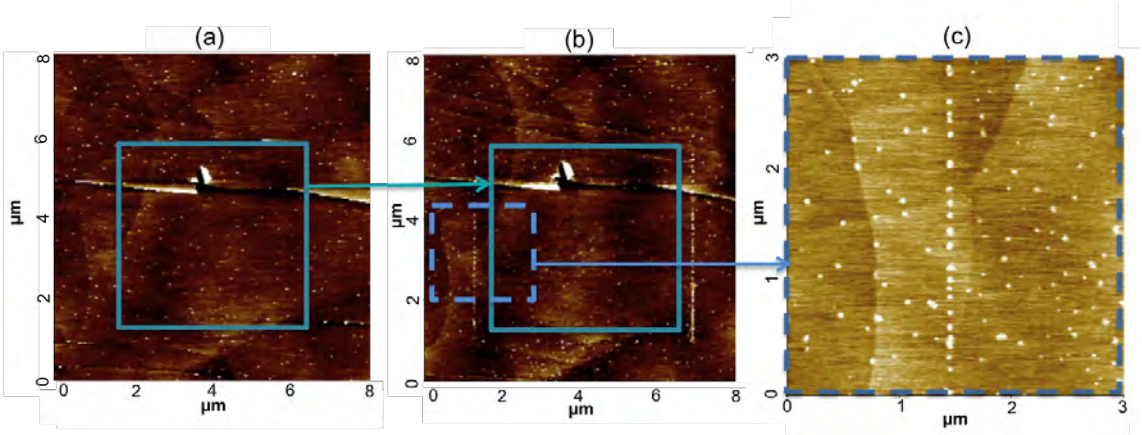


FIGURE 6.11: NC-AFM topographies of Au_{923} immobilised by Au_{20} channels sample acquired (a) before and (b) after contact mode scanning at 1 nN was performed in a sub-region of the NC first scan. In (b) it is possible to see the "contact window" referred previously as a result of the contact mode scan. (c) Close up to topography (b) that includes an area of the sample that did not undergo CM scan, i.e. outside the "contact window" (left) and an area inside the "contact window" (right), divided by a vertical line of clusters that were dragged to the edge of the scanned region by the AFM tip.

Moreover, there are other effects that CM-AFM scanning has on the samples. By analysing the samples before and after a CM-AFM scan in certain region, it is possible to see two effects after contact mode: that the number of Au_n clusters is higher and that their dimensions are smaller. These consequences are exemplified in Figure 6.12 that shows a NC-AFM 10×10 μm scan of a sample with Au_{561} clusters immobilised by Au_{20} defects acquired after a 5×5 μm subregion was scanned in CM-AFM (region correspondent to topography (b) of Figure 6.10).

It is worth noting how a "contact window" results from applying contact mode to a subregion, just as in the case of samples with argon defects. The difference is that Figure 6.12 (a) shows a window full of clusters that have been successfully immobilised by Au_{20} defects, in contrast to the windows formed by contact mode scans of Figure 6.9, that show no remaining clusters inside, due to an insufficiently strong interaction with the argon defects. The edges of the window in Figure 6.12 are formed of dragged Au_n clusters, accumulated at the spatial end of the CM-AFM scan, just as demonstrated before in Figure 6.11. Nevertheless, by performing a smaller scan that includes both regions, inside and outside the "contact mode window" (Fig. 6.12 (b)), it is noticeable that clusters do not seem to be intact after CM-AFM. A clear decrease in height and

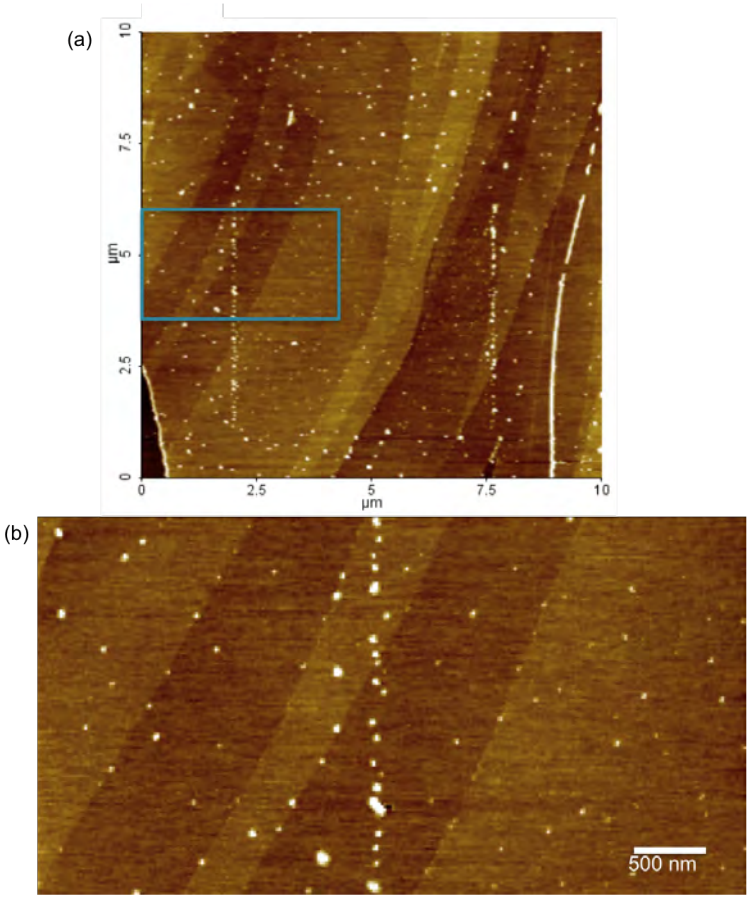


FIGURE 6.12: (a) NC-AFM topography of Au_{561} immobilised by Au_{20} channels sample acquired after contact mode scanning at 1 nN was performed in a 5×5 μm sub-region of the image. It is possible to confirm the presence of clusters inside the 5×5 μm window ("contact window"), evidencing the strong immobilisation of clusters by using the channels technique. (b) Close up to (a) in the highlighted rectangular region to show the visible difference in size and density that exists between the clusters that heve not been subjected to a CM-AFM scan (left side of the image) and the clusters inside the 5×5 μm area that have been subjected to CM-AFM scan at 1 nN (right side of the image).

raise in number has occurred, raising the question of the real effect that the AFM tip in contact mode has over the Au_n clusters, even if they are strongly immobilised in the substrate.

In order to answer this question, a systematic characterisation was carried out to quantitatively determine the differences in clusters density and dimensions between the samples before and after being subjected to one single scan in contact mode AFM applying a force of 1 nN. Since the same effect occurred for all the sizes of clusters (examples of Au_{309} and Au_{147} shown in Figure 6.13), it has been possible to conduct a comparison between samples and conclude whether or not the size of the cluster is a factor that intervenes in the reaction of the cluster to the action of the AFM tip in contact mode.

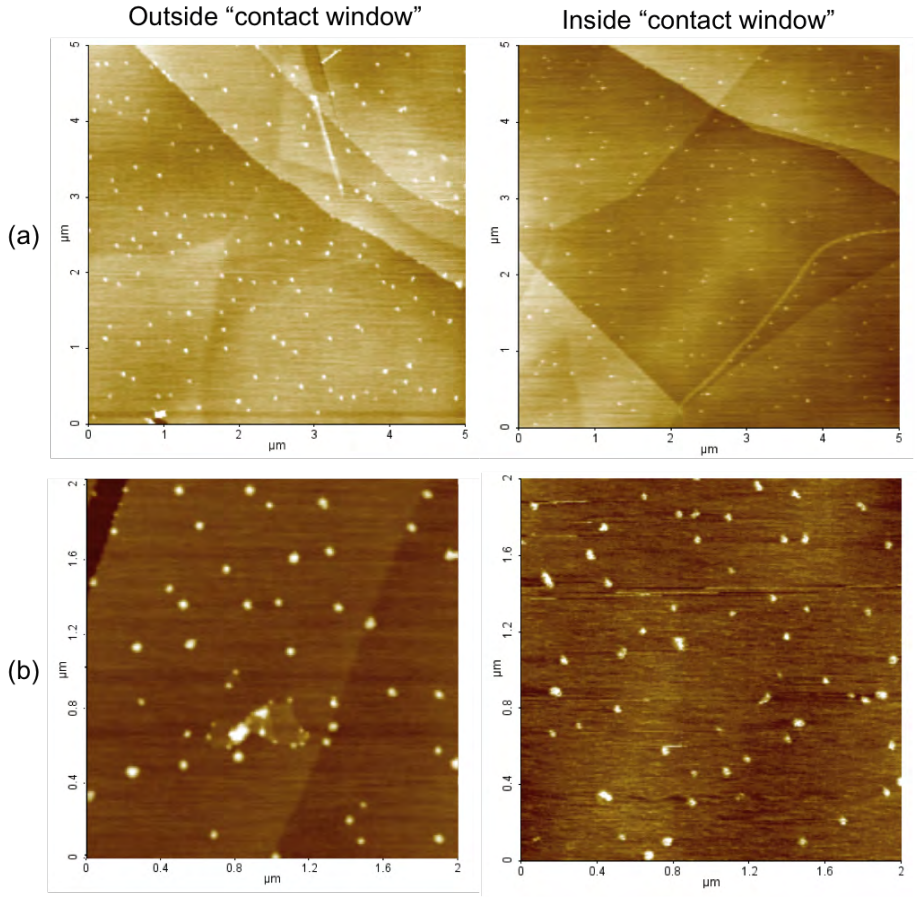


FIGURE 6.13: NC-AFM topography of (a) Au_{309} and (b) Au_{147} immobilised by Au_{20} channel samples acquired before (left images) and after contact mode (right images). *Before* images have been acquired from an external area to the "contact window" whereas *after* images have been acquired inside the "contact window" region. Au_n clusters are visibly smaller and higher in number after CM-AFM. The contact mode scan is always performed at a scanning rate of 1 Hz and with an applied force of 1 nN.

By calculating the ratio existing between the average density of the clusters before contact mode and the average density of the clusters of the same sample after contact mode we found a dependence with the inverse number of atoms in the clusters. The associated plot (Fig. 6.14) shows that change in density is minimal for Au_{147} clusters, with an average of 2.35 clusters/μm^2 more after contact mode scan, and a maximum for Au_{923} with an average of 6.16 clusters/μm^2 more after contact mode scan (Table 6.2). This size-dependent behaviour presumably resides in the size-relation between the soft-landed clusters and the channels created by the implantation of Au_{20}. The dimensions of the defects produced by the implantation of Au_{20} clusters lie between the dimensions of gold clusters with 1 layer (Au_{13}) and 2 layers (Au_{55}), therefore, the size of the defects produced by such treatment is considerably close to the size of Au_{147} clusters, which have only one more layer of atoms [61]. These matching conditions between the dimensions of

the cluster to immobilise and the immobilising agent are thought to be the main reason why Au_{147} is the size of clusters that remain less affected by the action of the contact mode AFM tip.

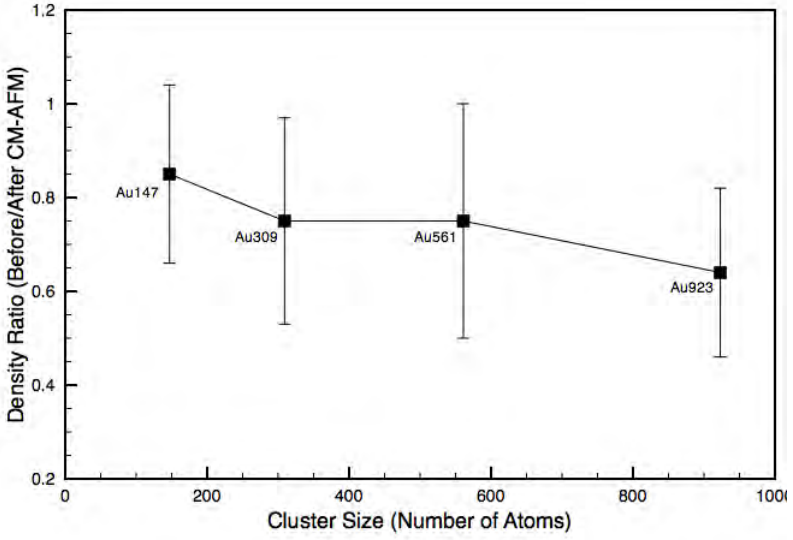


FIGURE 6.14: Plot of the density ratios for the different sizes of Au_n clusters before and after CM-AFM scan at 1 nN. The inverse dependence with number of atoms in the cluster in this case is more defining. Gold clusters with a minor number of atoms tend to remain less affected by the action of CM-AFM tip, causing a minor change in density.

	Before Contact mode AFM measurements	After Contact mode AFM measurements
Sample with implanted Au_{20}	Density (clusters/μm²)	Density (clusters/μm²)
Au_{923}	11.28 ± 2.73	17.44 ± 2.83
Au_{561}	6.5 ± 1.77	8.64 ± 1.80
Au_{309}	7.6 ± 1.87	10.00 ± 1.63
Au_{147}	12.42 ± 1.70	14.77 ± 2.72

TABLE 6.2: Density averages of the different sizes of Au_n clusters before and after CM-AFM. Mean values were obtained from the analysis of several NC-AFM scans in different areas outside and inside the "contact window".

In general, the increase in density for all the samples can be explained by the difference in coverage between Au_{20} and Au_n. Given this difference, approximately half the original defects created by implantation of the small clusters remain free after deposition of the larger clusters. It can be suggested that these free channels trapped the excess of clusters generated after cluster fragmentation caused by the action of the CM-AFM tip.

Regarding changes in cluster heights after a scan in CM, it was found that the average height of the Au_n clusters was reduced in a range that goes from 0.47 nm (for Au_{147} clusters) to 0.79 nm (for Au_{923} clusters). This relationship with the cluster size can be seen in the plot of height ratios for the clusters of all sizes before and after a scan in contact mode-AFM at 1 nN (Fig. 6.15).

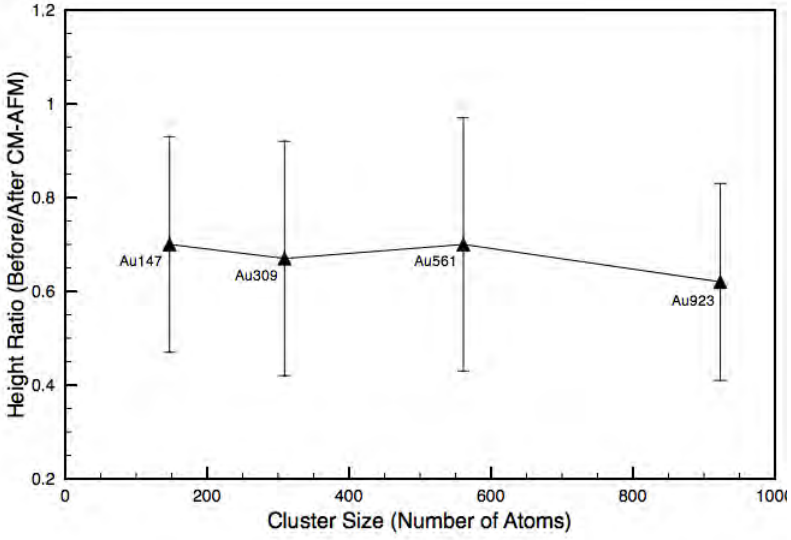


FIGURE 6.15: Plot of the height ratios for the different sizes of Au_n clusters before and after CM-AFM scan at 1 nN. The inverse dependence with number of atoms in the cluster is not defining enough to be able to establish a clear relationship between change in heigh and size of clusters.

A final consequence of the contact mode scan on the Au_n clusters is the homogenisation of the cluster heights. Height distributions of Au_n clusters subjected to contact mode AFM demonstrate a lower standard deviation than the distributions obtained before contact mode. A plausible explanation for this is based on the nature of Au_{20} clusters and how it affects the Au_n initially immobilised by the defects created after their implantation on graphite. Given the tetrahedral geometry of Au_{20} [233], the orientation of the Au_{20} clusters when deposited determines the implantation depth, causing a difference in heights for larger clusters immobilised by such defects. The implantation depth of Au_{20} at 1.5 keV is estimated to be $\sim$1 nm, and the approximate diameters of the cluster itself is also $\sim$1 nm. This matching between the holes depth and the height of the implanted clusters makes even more critical the orientation of the Au_{20} clusters as a determining factor that can have an influence on the heights of Au_{2n} clusters immobilised by the defects.

6.7 Overview of Chapter 6

Thanks to the combination of techniques for cluster deposition on graphite it is possible to visualise soft-landed small gold nanoclusters (down to 147 atoms) with AFM in non-contact mode with sub-nanometre resolution. The combination of the advantages of the self-pinning technique and the soft-landing on Ar defects technique resulted in a more solid cluster immobilisation methodology (Fig. 6.16).

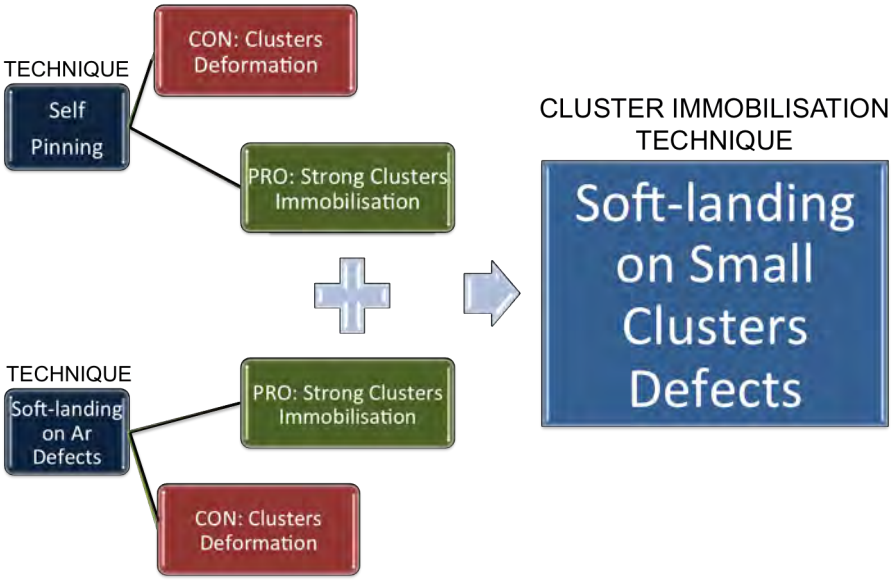


FIGURE 6.16: Scheme of the combination of cluster immobilisation techniques on graphite self-pinning and argon sputtering, showing their respective advantages and disadvantages. The combination of the respective advantages, i.e. the strong cluster immobilisation that the self-pinning technique provides and the lesser deformation of the clusters due to the low deposition energy of the soft-landing on Ar defects, results in the development of the soft-landing on small clusters defects technique, which has been demonstrated to overcome the main disadvantages of the two techniques from which it was derived.

Restrictions of the pinning method for immobilising clusters on surfaces, such as the deformation of clusters during the transition to the supported state and the limitation for large clusters ($N > 500$) due to the linear dependency of the pinning threshold on the number of atoms in the cluster, have been overcome with this new technique. Pre-treatment of graphite surfaces, decorating it with channels created by implantation of small gold clusters (20 atoms), has proved to efficiently immobilise clusters with hundreds of atoms, by partially resisting the force applied by contact mode AFM. The results presented in this work demonstrate the implantation of small gold clusters to be an efficient technique of cluster immobilisation for development of nanostructured surfaces with clusters. The argon sputtering technique also meets the given requirements, but

the cluster immobilisation is not strong enough to resist the lateral force exerted by contact mode AFM, which is an essential requirement for posterior attempts at protein immobilisation.

However, for characterisation purposes, the achieved conditions for both immobilisation methods, argon sputtering technique and the novel cluster channels pre-treatment technique, have enabled a topographical analysis of soft-landed gold clusters on graphite using an instrument (AFM) that has the advantage of also working under physiological conditions for biophysical experiments.

The difference in height found between clusters in samples with argon defects and those in samples with Au_{20} channels can be explained due to the characteristics that differ from one kind of sample to the other. Whereas Ar^+ sputtering produces defects at atomic scale by removing only one carbon from the graphite lattice, the implantation of the small clusters creates nano-channels in the graphite substrate with a depth dependent on the size and kinetic energy of the Au_{20} clusters and with a diameter equal to the cluster diameter. As consequence, Au_n clusters soft-landed on Au_{20} samples have a larger area of interaction with the substrate, incrementing the probability of adopting a hemisphere configuration [133], and therefore slightly decreasing in height compared to the clusters deposited in Ar^+ samples.

In the particular case of gold clusters soft-landed in graphite with implanted Au_{20} clusters, the contact mode AFM studies was revealed to have an effect on the size and density of the clusters, noticed in the reduction of height and in the increment in density. This effect could refer to a possible nano-planarization [234] action of the AFM tip over the clusters, performing as a nanofabrication technique that is able to produce an array of clusters with the same height.

Chapter 7

Conclusions and Future Work

The project presented in this thesis encompasses two distinct aspects of the interactions of the desmoplakin protein as assessed by NMR, SAXS and AFM. One aspect is focused on the interaction with the intermediate filament proteins, associated with the desmoplakin C-terminal region, and the other is the interaction with a nanostructured functionalised surface based on size-selected gold clusters, associated with its N-terminal region.

Since knowing the structure of a protein is key to understanding its function, this project has addressed and achieved the determination of the atomic structure of the desmoplakin linker domain - a protagonist domain in intermediate filament binding - by NMR spectroscopy. A biophysical characterisation of the desmoplakin linker domain has been completed, as well as the backbone and side chain assignments, obtaining the tertiary structure of the desmoplakin linker domain and demonstrating that this domain is a structured and folded protein, by contrast to the disordered tether previously proposed.

Despite the sequence homology of the desmoplakin linker domain with the desmoplakin plakin repeat domains, it has been asserted that, unlike the latter, the linker domain is not formed by canonical plakin repeats. While the C-terminal subdomain of desmoplakin linker has a topologically similar structure to the plakin repeat (PR) motifs, with a beta sheet packed against a helix-turn-helix motif, the N-terminal subdomain of the protein resembles a PR to a lesser extent since it does not contain a beta hairpin, making it structurally distinct. Other constitutive differences between both structures are an extra

alpha helix towards the C-terminal end of the protein and variations in length of helices and strands.

The structural differences found between the desmoplakin linker domain (DPlink) and the desmoplakin plakin repeat domains suggest that their binding functions also differ. The N-terminal region of DPlink is of particular interest, since it exhibits the greatest difference from the PR motif structure and is characterised by a high level of flexibility. It also contains a cluster of basic amino acids which are exposed in a large loop and are followed by three acidic amino acids also available for inter-protein interactions. In combination with the acidic groove of vimentin (a type III intermediate filament), this characteristic of DPlink suggested that the binding activity of desmoplakin and vimentin is of predominantly electrostatic nature. In order to elucidate the nature of the interaction between desmoplakin and vimentin and to narrow down the binding sites within the desmoplakin C-terminal end, NMR titration experiments were carried out with vimentin and various types of desmoplakin and periplakin linkers (PPlink), wild types and selected mutants. The wild type desmoplakin linker domain showed a weak interaction with vimentin, while the wild type periplakin linker showed a stronger interaction. Based on these results, charge reversal mutations of specific residues of DPlink and PPlink were chosen to enhance and retrieve, respectively, the ability to bind vimentin. The results showed that the residues in the N-terminal end of DPlink play a critical role in vimentin binding, confirming the importance of this protein sequence segment, potentially conferred by its unique secondary structure. Further mutations on desmoplakin linker are planned as part of future work to provide more supportive evidence of this already well substantiated functional hypothesis. Similarly, repetition of the NOESY experiments on the DPlink are planned with the purpose of eradicating any ambiguity related to the structure of the N-terminal end of the domain.

The conclusions derived from the NMR titration experiments corroborate the theory that the linker domain is not the only element that mediates the desmoplakin binding activity with vimentin; hence the importance of gauging the architecture of the desmoplakin plakin repeat domains (PRDs) B and C in a construct including the linker domain. For this, SAXS experiments were performed in order to generate a molecular envelope that revealed the structural shape of the desmoplakin BC construct.

The crystal structures of the desmoplakin plakin repeat domains B and C were combined

with the NMR structure of the desmoplakin linker domain obtained here to form a model of all three domains fitted in the SAXS envelope of the desmoplakin BC envelope. The fit of the theoretical scattering curve of this model to the SAXS scattering curve of the BC construct revealed a considerable level of flexibility for the desmoplakin C-terminal end that suggests that this structure can be classified as semi-dynamic. Despite the different conformations that the BC construct can apparently adopt due to its lack of rigidity, the SAXS envelope obtained in our experiments is in agreement with the theory that says the desmoplakin C-terminal end possesses an elongated shape, where the domains that constitute it are arranged like "beads-on-a-string", as proposed by Choi *et al.* [94]. In this manner, the SAXS study sheds light on the domain organisation of the C-terminal end of desmoplakin, complementing the NMR information gained for the desmoplakin linker domain and clarifying the cytoskeletal binding preferences of desmoplakin.

As well as its role as an adaptable bridge to the intermediate filaments, the significance of the desmoplakin protein includes its individual properties, in particular its peculiar elasticity, which allows the desmosomes to resist shearing stresses, providing strength to the cell-cell junctions. With the ultimate aim of studying this and other physical properties of desmoplakin at the single molecule level, as well as potentially studying similar properties of other proteins, a platform was developed to enable AFM imaging, and ultimately force spectroscopy, of protein molecules. This platform was built on a nanostructured graphite surface, prepared by the pinning of size-selected gold clusters containing 55 and 147 atoms. The interaction between the desmoplakin plakin domain (PD), containing a 2xCys tag, and this functionalised surface was analysed using AFM in non-contact, contact and tapping modes. Tapping mode AFM in buffer solution demonstrated enhanced non-covalent bonding between the desmoplakin plakin domain and gold nanoclusters, leading to the weak immobilisation of the protein on the surface. It was suggested that this interaction is a physisorption process (versus chemisorption), promoted by hydrophobic interactions between the support and hydrophobic residues on the protein surface [204], independent of the thiol-gold bond initially expected to form between the cysteine and the gold clusters. It was assumed that the Cys tag, or at least the thiol group, is not available for specific binding.

The achieved level of protein immobilisation was not considered strong enough to subject the desmoplakin plakin domain to mechanical studies, such as stretching force measurements. This would require the specific tethering of both ends of the protein, by anchoring

the N-terminal end of the PD to gold clusters through a properly exposed Cys tag and the C-terminal end to a functionalised AFM tip, e.g. with Ni-NTA via a 6×His tag. Nonetheless, the enhancement of the interaction of PD with the graphite surface decorated with the size-selected gold clusters did inhibit protein diffusion and accumulation on graphite defects. This allowed a topographical study of isolated single molecules and protein complexes at low concentrations, due to the stability showed under the AFM non-contact mode and tapping mode in dry and liquid phase respectively. The height and diameter distributions demonstrate the feasibility of obtaining an array of isolated single proteins and also suggests a quasi-2D protein film growth process. Here the immobilised PD molecules lie parallel to the surface, while other molecules diffuse laterally to find nucleation points of molecules bound to the clusters and attach to them via weak interactions to form 2D islands.

Given the recent atomic structure determination of a segment of the desmoplakin plakin domain [96], it is now possible to identify the level of accessibility of the cysteines contained in the protein. This could be used to assess whether or not any of them are exposed and which ones form intramolecular disulphide bonds. This information should be sufficient for the future re-design of a desmoplakin PD protein with cysteines actually available to bind to the gold nanoclusters, thereby enabling AFM force spectroscopy.

In connection with the desmoplakin plakin domain AFM experiments on the nanocluster-treated graphite surface, and in order to improve this prospect for immobilisation and imaging of proteins using this platform, a novel technique for large cluster immobilisation has been developed. The new cluster immobilisation methodology was based on defects created by the implantation of small Au_{20} clusters and shown to improve upon previous immobilisation techniques such as cluster self-pinning and cluster immobilisation at argon sputtering defects. The approach combines the main benefts of the preceding techniques: the strong immobilisation of clusters by pinning deposition and the modest deformation of the cluster morphology by soft-landing the clusters on pre-treated graphite with argon ion defects. The immobilisation of larger Au clusters by the Au_{20} cluster channels permits AFM imaging of the clusters when in the past only STM was able to image self-pinned clusters with reduced height due to high deposition energies, while clusters bound only to Ar ion defects were displaced by the AFM tip in contact mode. This improvement should allow progress of protein immobilisation experiments using nanoclusters, since AFM imaging in the native environment should now be able

to correlate the images of clusters and proteins spatially, allowing for example real-time studies of protein binding and complex formation.

In summary, the interdisciplinary work presented in this thesis combines physics and biology approaches at the nanoscale to advance the characterisation of desmoplakin, an important constituent of desmosomes. This was achieved by gaining insight into the structure of this biomolecule and its interactions with both the cytoskeleton intermediate filaments and structured cluster-decorated surfaces in a physiological environment, the latter contributing to the future development of nanostructured platforms for single molecule biology.

Appendix A

Protein Production

A.1 Vectors

The pGex6p-1 vectors include an ampicillin resistance gene (Ampr), an origin of replication, and the tac promoter (Ptac) which allows inducible expression of the fusion protein using Isopropyl-β-D-Thio-Galactoside (IPTG) (Fig. A.1).

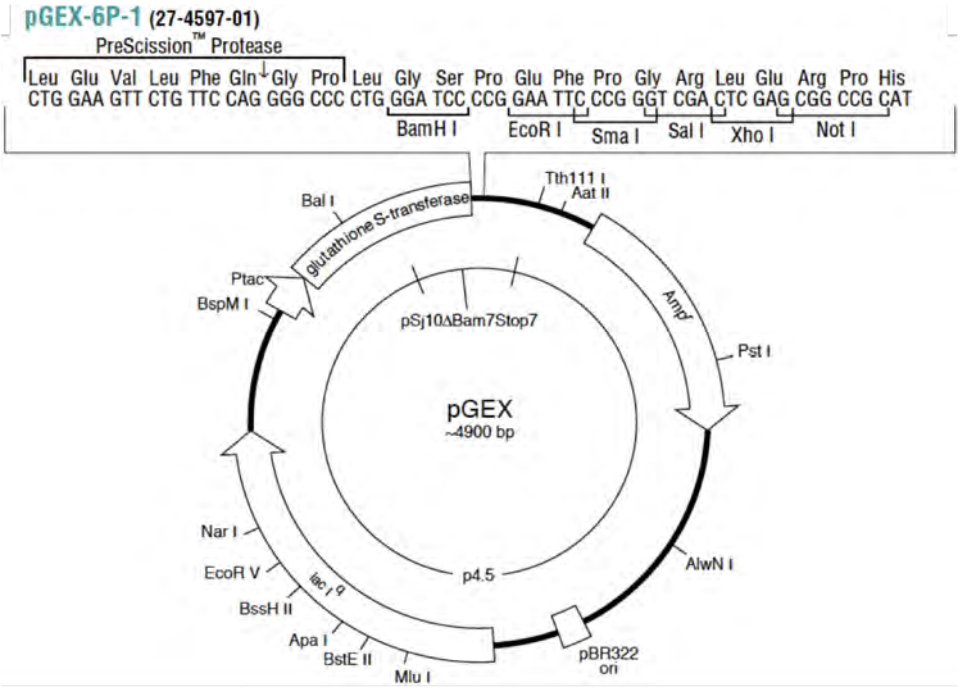


FIGURE A.1: Vector pGEX-6P-1. Bacterial-expression plasmid that contains the glutathione S-transferase (GST) gene, followed by DNA encoding a cleavage site for the enzyme PreScission protease. The PreScission site is followed by a multiple cloning site containing six unique restriction enzyme sites that can be used to clone exogenous DNA into the vector. From reference [235].

The pProEX HTc vector includes an ampicillin resistance gene (Ampr), an origin of replication, and a Trc promoter, which allows inducible expression of the fusion protein with IPTG (Fig. A.2).

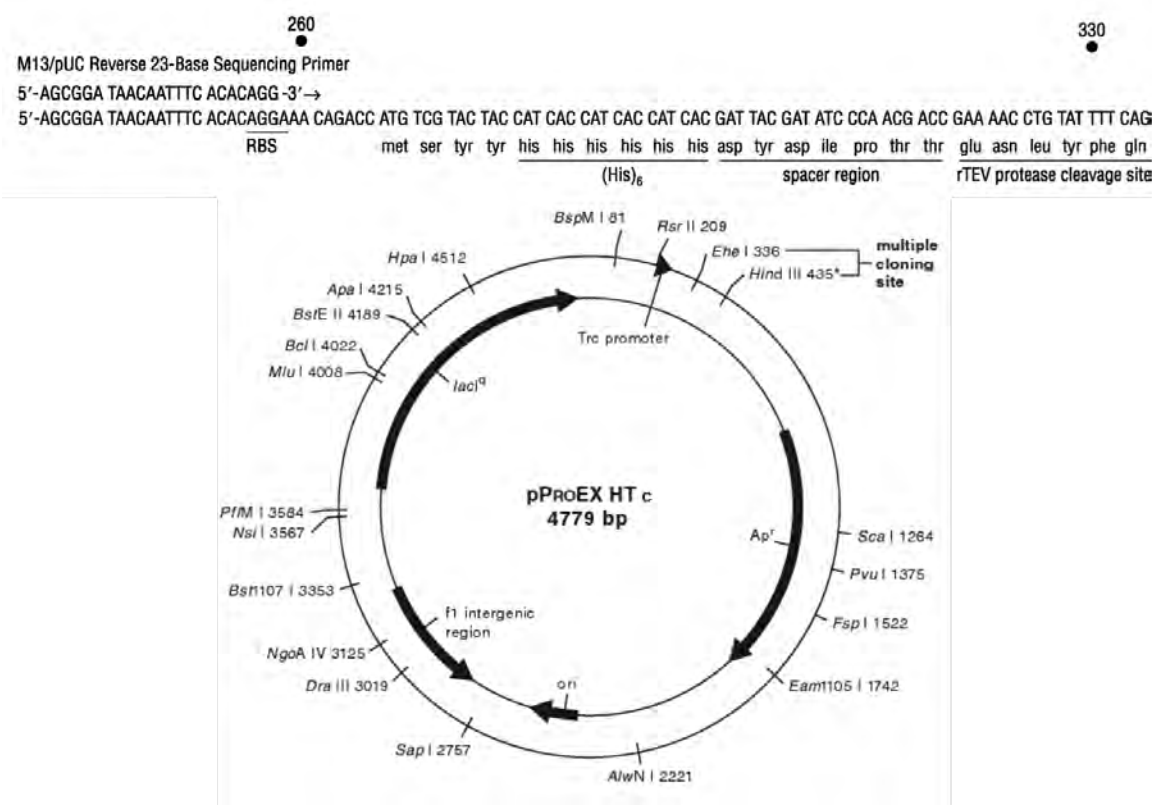


FIGURE A.2: Vector pProEXHTc. Bacterial-expression plasmid that encodes a N-terminal 6xHis tag, followed by a spacer and a Tobacco Etch Virus (TEV) protease cleavage site. The TEV protease cleavage site is followed by a multiple cloning site containing 10 unique restriction enzyme sites for cloning of exogenous DNA. From reference [236].

The pET-21a(+) vector contains an T7 ampicillin resistance gene (Ap), an origin of replication, and a tac promoter for inducible expression of target genes with IPTG (Fig. A.3).

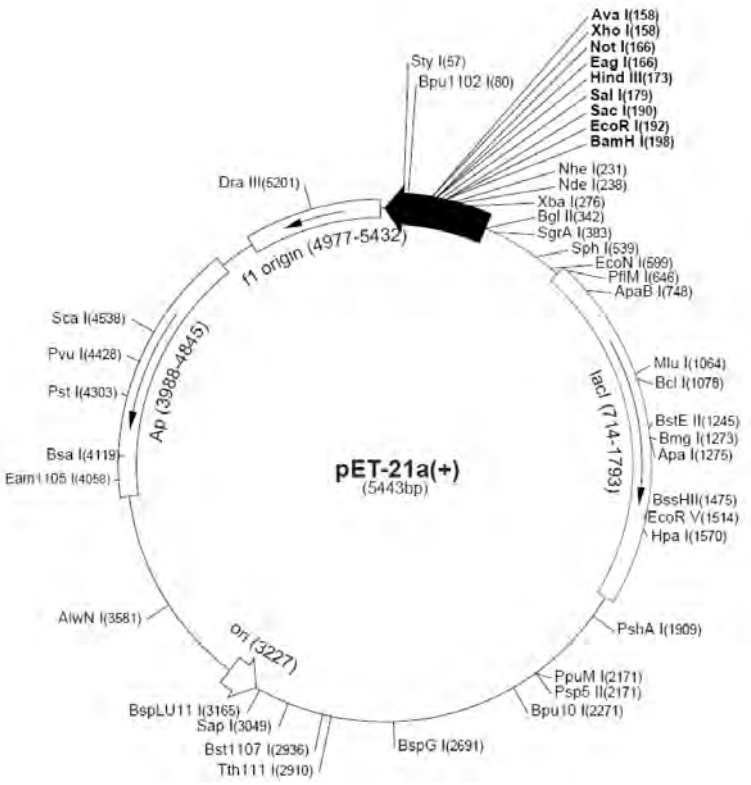


FIGURE A.3: Vector pET-21a(+). Bacterial-expression plasmid that contains a multiple cloning site containing 9 unique restriction sites for cloning of exogenous DNA. From reference [237].

A.2 Buffer solutions

Minimum M9 media (for 1 L): 6 g of Na_2HPO_4, 3 g of KH_2PO_4 and 0.5 g of NaCl.

Nutrient Mix for ^{15}N labeled proteins (for 1 L): 10 μL of 1 M CaC_2, 2 mL of 1 M $MgSO_4$, 400 μL of 1M $FeCl_3$, 1 mL of 20 mg/mL Thiamine, 1 g of $^{15}NH_4Cl$ and 0.5 mL of Metal Mix, containing 0.115 g of $ZnSO_4$, 0.0169 g of $MnSO_4$, 0.029 g of $H3BO_3$ and 0.0175 g of $CuSO_4$ for 100 mL.

Low salt buffer for AKTA salt gradient: 20 mM HEPES pH 7.5, 20 mM NaCl.

High salt buffer for AKTA salt gradient: 20 mM HEPES pH 7.5, 0.5 M NaCl.

Equilibration buffer for His-tagged purification: 50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidiazole.

Washing buffer for His-tagged purification: 20 mM HEPES pH 7.5, 500 mM NaCl, 30 mM Imidazole.

Elution buffer for His-tagged purification: 20 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole.

Triton buffer: Triton X-100 (100 %), 4 M NaCl, 0.5 M EDTA, 10 % Na Azide, 1M Tris pH 8.

Washing buffer with no Triton X-100: 0.5 M EDTA, 10 % Na Azide and 1M Tris pH 8.

Solubilisation buffer: 10 mM Tris-HCl pH 8, 8M Urea.

Dialysis buffer for vimentin and desmin purification: 10 mM Tris-HCl pH 8, 5 mM DTT.

Appendix B

NMR Experiments

B.1 Spectra Acquisition

^{1}H, ^{15}N-HSQC (heteronuclear single quantum coherence) spectra were acquired using the fast-HSQC method and 3919 WATERGATE for suppression of the water signal [238]. Spectra were collected using 2048 data points in the direct dimension, and 256 increments were used in the indirect dimension. The sweep widths used were 6720.34 Hz in ^{1}H and 1635.56 Hz in ^{15}N. States-TPPI was used for quadrature detection.

Backbone assignment spectra were collected on a Agilent DirectDrive spectrometer operating at 600 Mhz in ^{1}H. Pulse sequences were taken from the standard Agilent BioPack library of pulse programs. BEST pulse sequences were used to reduce acquisition time [239]. The inter scan delay was set to 0.4 sec for all BEST experiments. All Spectra were collected using 860 points in the direct dimension and 64 increments in ^{15}N. Sweep widths were 6720.34 Hz in ^{1}H and 1635.56 Hz in ^{15}N. States-TPPI was used for quadrature detection in ^{13}C and EchoAntiecho used for ^{15}N.

For BEST-HNCA and BEST-HNCOCA spectra 128 increments and a sweep width of 4521.77 Hz in ^{13}Ca. For BEST-HNCO and BEST-HNCACO spectra 96 increments and a sweep width of 2414.38 Hz in $^{13}C'$. For BEST-HNCACB and BEST-HNCOCACB spectra 128 increments and a sweep width of 12057.71 Hz for $^{13}CaCb$.

A CCH-TOCSY-NNH spectrum [240] was acquired to correlate backbone amide assignments to side chain carbon atoms. The spectrum was collected on an Agilent 600DD

spectrometer using 1024 points in the direct dimension, 120 increments in ^{13}C dimension and 64 increments in ^{15}N. Sweep widths were 6720.43 Hz in ^{1}H,12058.30 Hz in ^{13}C and 1635.56 Hz in ^{15}N. States-TPPI was used for quadrature detection in ^{13}C and Echoantiecho in the ^{15}N dimension.

For side chain assignments a 3D HCCH-TOCSY spectrum [241] was acquired using an Agilent Inova 800 MHz spectrometer. For spectrum collection 2048 points were used in the direct dimension, 256 increments were used in the indirect proton dimension and 96 increments in ^{13}C. Sweep widths were 8797.01 Hz in ^{1}H and 8045.05 Hz in ^{13}C. The ^{13}C dimension was folded in order to reduce acquisition time and improve resolution. The mixing time of the tocsy experiments is 14 ms. States-TPPI was used for quadrature detection in both indirect dimensions.

A ^{15}N-resolved NOESY-HSQC with WATERGATE for solvent suppression was collected on an Agilent Inova 900 MHz spectrometer with 2048 points used in the direct dimension, 144 increments in the indirect proton dimension and 88 increments in ^{15}N. Sweep widths were 10787.49 Hz in ^{1}H and 2735.23 Hz in ^{15}N. A NOESY mixing time of 100 ms was used. States-TPPI was used for quadrature detection in both indirect dimensions.

Two ^{13}C-resolved NOESY-HSQC spectra were acquired on an Agilent Inova 800 MHz spectrometer. The aliphatic NOESY-HSQC spectrum was collected with 2048 points in the direct dimension, 272 increments in the indirect proton dimension and 96 increments in the ^{13}C dimension. Sweep widths were 8792.17 for ^{1}H and 8045.05 for ^{13}C. The mixing time was set at 100 ms. The aromatic centred NOESY-HSQC spectrum was collected with 2048 points in the direct dimension, 168 increments in the indirect proton dimension and 48 increments in the ^{13}C dimension. Sweep widths were 6397.95 for direct ^{1}H, 8000 for indirect ^{1}H and 6435.52 for ^{13}C. The mixing time was again set at 100 ms.

For titrations a SOFAST-HMQC spectra were acquired [156]. Spectra were collected using 876 data points in the direct dimension, and 256 increments in the indirect dimension. The inter scan delay was set to 0.4 sec. Sweep widths used were 6720.34 Hz in ^{1}H and 1635.56 Hz in ^{15}N. States-TPPI was used for quadrature detection.

B.2 Magnetisation Transfer

^{1}H-^{15}N HSQC: the magnetisation is transferred from the ^{1}H of the NH of a residue to the attached ^{15}N via J-coupling. Evolution of the chemical shift occurs in the ^{15}N nucleus and the magnetisation is transferred back to the proton for FID acquisition.

HNCA.- This correlates the amide ^{1}H and ^{15}N chemical shifts of an amino acid, *i*, with the $^{13}C_{\alpha}$ chemical shift of the same amino acid (strongly) and the $^{13}C_{\alpha}$ resonance of the correspondent *i -1* amino acid (weakly) via N-C_{α} J-coupling.

HN(CO)CA.- Similar to the HNCA, the magnetisation is passed from the ^{1}H of the amide to ^{15}N and then to $^{13}C_{\alpha}$ resonance of the preceding residue, passing through its ^{13}C', in which the chemical shift is not evolved. This experiment complements the HNCA by being selective for the $^{13}C_{\alpha}$ of the correspondent *i -1* amino acid.

HNCO.- This correlates the amide ^{1}H and ^{15}N chemical shifts of an amino acid *I* with the ^{13}C' of the carbonyl resonance of the correspondent *i -1* amino acid.

HN(CA)CO.- Similar to the HNCO, this experiment correlates the amide and ^{15}N chemical shifts of an amino acid *I* with the ^{13}C' of the same residue and also with the ^{13}C' of the *i -1* amino acid, but in both cases the magnetisation transfer passes first through $^{13}C_{\alpha}$ of the corresponding residues via J-coupling.

HNCACB.- This correlates the resonances of the amide and ^{15}N frequencies with those of the *i* and *i -1* residue alpha and beta carbon resonances. In each of the residues a simultaneous magnetisation transfer goes from $^{1}H_{\alpha}$ and $^{1}H_{\beta}$ to $^{13}C_{\alpha}$ and $^{13}C_{\beta}$, respectively, and then from $^{13}C_{\beta}$ to $^{13}C_{\alpha}$. Magnetisation is then transferred to ^{15}N of the *I* residue from both $^{13}C_{\alpha i}$ and $^{13}C_{\alpha i-1}$, resulting in two visible peaks that appear in the same dimension since the chemical shifts evolved simultaneously.

HN(CO)CACB.- Similar to HNCACB, this experiment correlates the amide and ^{15}N resonances of the *i* residue with both the $^{13}C_{\alpha}$, and $^{13}C_{\beta}$ resonances, but in this case only of its preceding residue. Magnetisation transfer follows also the same path except that it passes trough the ^{13}C' of the *i -1* amino acid.

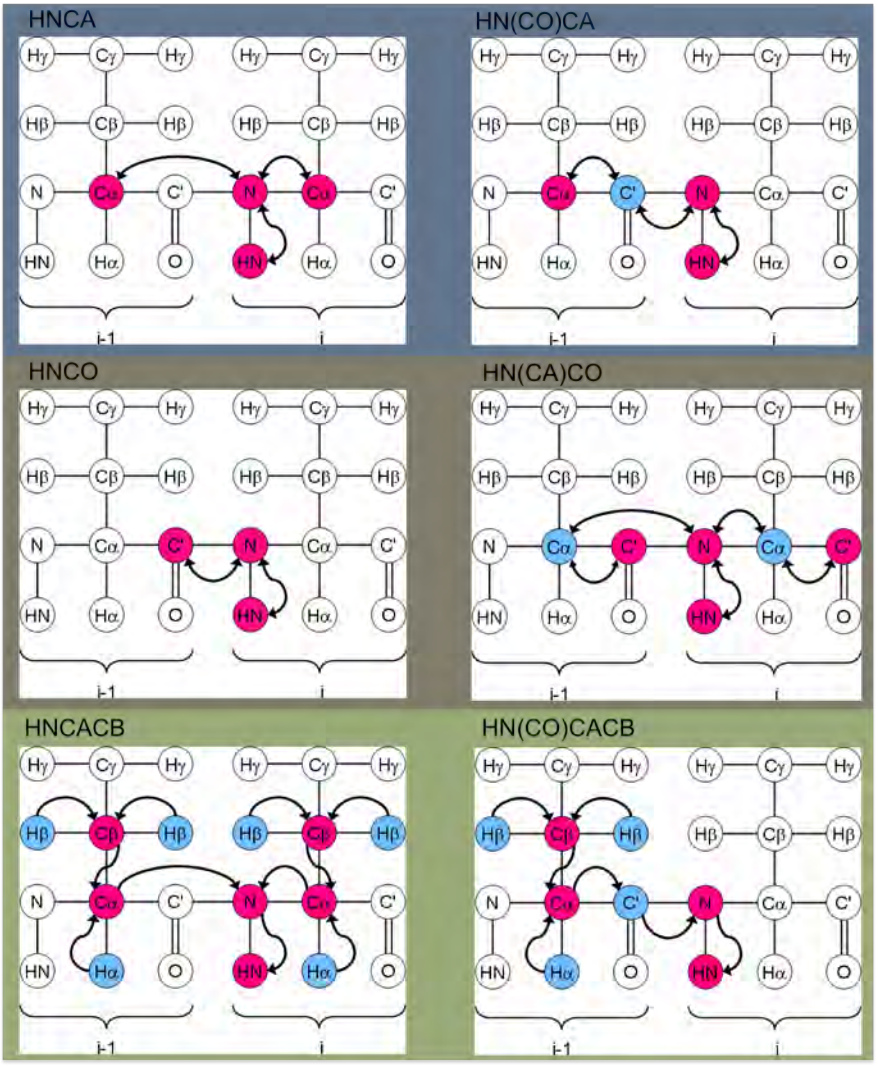


FIGURE B.1: Triple resonance NMR experiments. Letters in parenthesis indicate nuclei involved in the magnetisation transfer but not taken into account for data acquisition, since the chemical shift does not evolve in such nuclei (blue). The name also does not indicate explicitly that the magnetisation transfer starts on a proton and ends on a proton, doing a closed cycle, which is necessary for sensitivity purposes. Taken from reference [242].

HA(CACO)NH.- 3D experiment that links the chemical shifts of the backbone amide with the $^{1}H_{\alpha}$ of the preceding residue . The magnetisation transfer, via J-coupling, goes from $^{1}H_{\alpha}$ of residue *I-1* to $^{13}C_{\alpha}$ and then to the ^{13}C' of the same residue. From there it goes to the NH group of the *I* residue where is detected (fig. B.2(a)). Thus the ^{1}H, ^{15}N peaks of the HSQC are correlated to $^{1}H_{\alpha}$. This gives a starting point for the assignment of the side chain ^{13}C and ^{1}H using TOCSY experiments.

H(C)CH-TOCSY.- The magnetisation transfer occurs from the side-chain ^{1}H nuclei to their attached ^{13}C nuclei. After isotropic ^{13}C mixing, the magnetisation is transferred

back to the side-chain ^{1}H for its detection [243].

H(CC)(CO)NH TOCSY is an experiment that correlates the amide resonances of the residue *i* with $^{1}H_{\alpha}$ and all ^{1}H side chain resonances of the residue *i-1*. This involves starting with excitation of the side chain aliphatic resonances followed by transfer to the attached carbon. A ^{13}C spin lock is applied allowing transfer of magentisation throughout the spin system (all directly connected aliphatic carbons). Magnetisation on the $^{13}C_{\alpha}$ is then transfered via J-coupling to the C' of the same residue then to the amide N and finally amide H group of the next residue which the direct acquisition occurs (fig. B.2(c)).

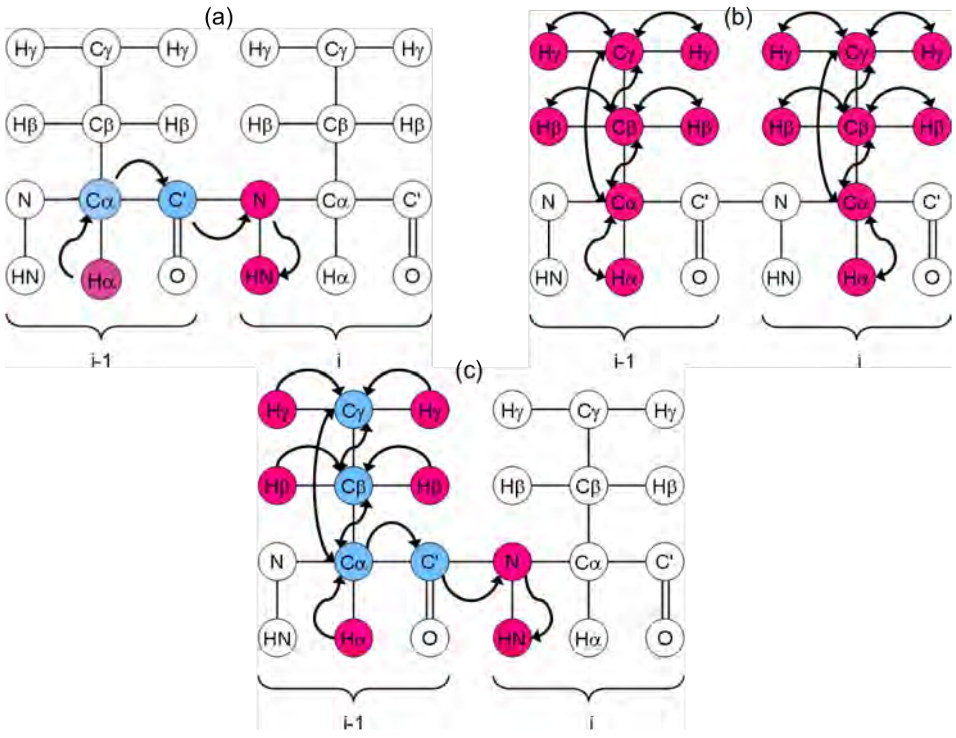


FIGURE B.2: Magnetisation transfer diagrams for the three experiments used in side chain assignment. (a) HA(CACO)NH, (b)H(C)CH-TOCSY, (c) H(CC)(CO)NH TOCSY. Taken from reference [242].

The NOESY-^{15}N-HSQC experiment is a 3D NMR experiment that shows the resonances caused by the NOE existing between one NH group and all other neighboring protons. Each cross peak corresponds to a through-space dipole-dipole coupling and the intensity of the peaks reflects the distance between nuclei. In this case, the magnetisation is exchanged between all protons due to the NOE ^{15}N nuclei and back to ^{1}H for detection [244] (fig.B.3).

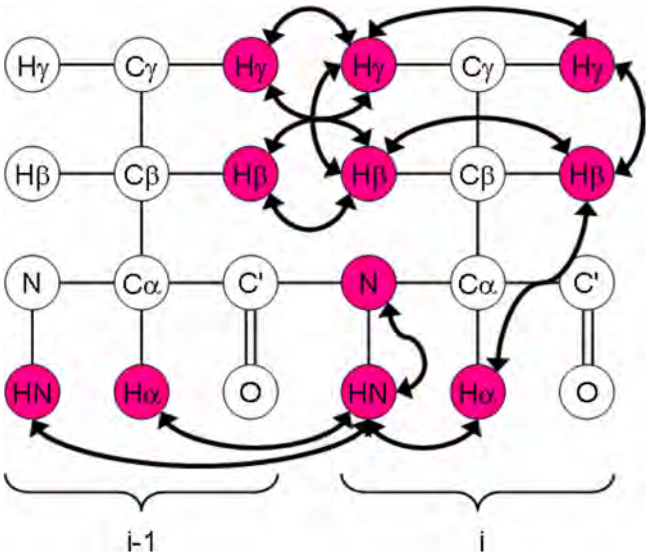


FIGURE B.3: Diagram of NOESY-N-HSQC magnetisation transfer. The amide groups resonances conform the direct dimension, the ^{15}N the indirect dimension and all the protons in spatial proximity to the amide proton constitute the indirect proton dimension.

Bibliography

[1] J Fritz, D Anselmetti, J Jarchow, and X Fernandez-Busquets. Probing single biomolecules with atomic force microscopy. *Journal of Structural Biology*, 119(2): 165–171, 1997.

[2] PH Kussie, S Gorina, V Marechal, B Elenbaas, J Moreau, AJ Levine, and NP Pavletich. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, 274(5289):948–953, 1996.

[3] CM Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, 2003.

[4] M Delmar and WJ McKenna. The cardiac desmosome and arrhythmogenic cardiomyopathies from gene to disease. *Circulation Research*, 107(6):700–714, 2010.

[5] JR Banavar and A Maritan. Physics of proteins. *Annual Review of Biophysics Biomolecular Structure*, 36:261–280, 2007.

[6] E.E. Lattman and P.J. Loll. *Protein Crystallography, A Concise Guide*. Johns Hopkins University Press, 2006.

[7] P Echenique. Introduction to protein folding for physicists. *Contemporary Physics*, 48(2):81–108, 2007.

[8] J Wang, X Zuo, P Yu, IJL Byeon, J Jung, X Wang, M Dyba, S Seifert, CD Schwieters, and J Qin. Determination of multicomponent protein structures in solution using global orientation and shape restraints. *Journal of the American Chemical Society*, 131(30):10507–10515, 2009.

[9] K Wüthrich. Nobel lecture: NMR studies of structure and function of biological macromolecules. *Bioscience Reports*, 23(4):119–168, 2003.

[10] X Li, Q Shen, D Zhang, X Mei, W Ran, Y Xu, and G Yu. Functional groups determine biochar properties (pH and EC) as studied by two-dimensional ^{13}C NMR correlation spectroscopy. *PLOS ONE*, 8(6):e65949, 2013.

[11] L Fielding. NMR methods for the determination of protein-ligand dissociation constants. *Current Topics in Medicinal Chemistry*, 3(1):39–53, 2003.

[12] F Cordier and S Grzesiek. Direct observation of hydrogen bonds in proteins by interresidue 3h J NC'scalar couplings. *Journal of the American Chemical Society*, 121(7):1601–1602, 1999.

[13] GS Rule and TK Hitchens. *Fundamentals of Protein NMR Spectroscopy*. Springer, 2006.

[14] GM Clore and AM Gronenborn. NMR structure determination of proteins and protein complexes larger than 20 kDa. *Current Opinion in Chemical Biology*, 2 (5):564–570, 1998.

[15] A Ambrus. Comparison of NMR and X-ray crystallography as methods of the protein structure determination, 1997.

[16] G Wagner, SG Hyberts, and TF Havel. NMR structure determination in solution: a critique and comparison with X-ray crystallography. *Annual Review of Biophysics and Biomolecular Structure*, 21(1):167–198, 1992.

[17] S Dasgupta, GH Iyer, SH Bryant, CE Lawrence, and JA Bell. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins: Structure, Function, and Bioinformatics*, 28(4): 494–514, 1997.

[18] N Nishida and I Shimada. An NMR method to study protein-protein interactions. In *Integrin and Cell Adhesion Molecules*, pages 129–137. Springer, 2012.

[19] ERP Zuiderweg. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry*, 41(1):1–7, 2002.

[20] V Changizi, MA Oghabian, R Speller, S Sarkar, and AA Kheradmand. Application of small angle X-ray scattering (SAXS) for differentiation between normal and cancerous breast tissue. *International Journal of Medical Sciences*, 2(3):118, 2005.

[21] P Bernadó, E Mylonas, MV Petoukhov, M Blackledge, and DI Svergun. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, 129(17):5656–5664, 2007.

[22] C Neylon. Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. *European Biophysics Journal*, 37(5):531–541, 2008.

[23] HDT Mertens and Dmitri I Svergun. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*, 172(1):128–141, 2010.

[24] C Hamiaux, J Pérez, T Prangé, S Veesler, M Riès-Kautt, and P Vachette. The BPTI decamer observed in acidic pH crystal forms pre-exists as a stable species in solution. *Journal of Molecular Biology*, 297(3):697–712, 2000.

[25] CD Putnam, M Hammel, GL Hura, and JA Tainer. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, 40(03):191–285, 2007.

[26] DA Jacques and J Trewhella. Small-angle scattering for structural biologyexpanding the frontier while avoiding the pitfalls. *Protein Science*, 19(4):642–657, 2010.

[27] D Franke and DI Svergun. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, 42(2): 342–346, 2009.

[28] J Mou, DM Czajkowsky, SJ Sheng, R Ho, and Z Shao. High resolution surface structure of *E. coli* GroES oligomer by atomic force microscopy. *FEBS letters*, 381(1):161–164, 1996.

[29] HKL Blackley, GHW Sanders, MC Davies, CJ Roberts, SJB Tendler, and MJ Wilkinson. *In-situ* atomic force microscopy study of β-amyloid fibrillization. *Journal of Molecular Biology*, 298(5):833–840, 2000.

[30] DJ Müller and A Engel. Atomic force microscopy and spectroscopy of native membrane proteins. *Nature Protocols*, 2(9):2191–2197, 2007.

[31] A del Rio, R Perez-Jimenez, R Liu, P Roca-Cusachs, JM Fernandez, and MP Sheetz. Stretching single talin rod molecules activates vinculin binding. *Science Signaling*, 323(5914):638, 2009.

[32] M Radmacher, RW Tillamnn, M Fritz, and HE Gaub. From molecules to cells: imaging soft samples with the atomic force microscope. *Science*, 257(5078):1900–1905, 1992.

[33] HG Hansma and JH Hoh. Biomolecular imaging with the atomic force microscope. *Annual Review of Biophysics and Biomolecular Structure*, 23(1):115–140, 1994.

[34] S Allen, SM Rigby-Singleton, H Harris, MC Davies, and P O'Shea. Measuring and visualizing single molecular interactions in biology. *Biochemical Society Transactions*, 31(5):1052–1058, 2003.

[35] S Xu, B Bevis, and MF Arnsdorf. The assembly of amyloidogenic yeast sup35 as assessed by scanning (atomic) force microscopy: an analogy to linear colloidal aggregation? *Biophysical Journal*, 81(1):446–454, 2001.

[36] SW Schneider, J Lärmer, RM Henderson, and H Oberleithner. Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. *Pflügers Archiv*, 435(3):362–367, 1998.

[37] D Fotiadis, S Scheuring, SA Müller, A Engel, and DJ Müller. Imaging and manipulation of biological structures with the AFM. *Micron*, 33(4):385–397, 2002.

[38] C Möller, M Allen, V Elings, A Engel, and DJ Müller. Tapping-mode atomic force microscopy produces faithful high-resolution images of protein surfaces. *Biophysical Journal*, 77(2):1150–1158, 1999.

[39] TP Weihs, Z Nawaz, SP Jarvis, and JB Pethica. Limits of imaging resolution for atomic force microscopy of molecules. *Applied Physics Letters*, 59(27):3536–3538, 1991.

[40] A Engel, HE Gaub, and DJ Müller. Atomic force microscopy: a forceful way with single molecules. *Current Biology*, 9(4):R133–R136, 1999.

[41] M Rief, J Pascual, M Saraste, and HE Gaub. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *Journal of Molecular Biology*, 286(2):553–561, 1999.

[42] L Schmitt, M Ludwig, HE Gaub, and R Tampe. A metal-chelating microscopy tip as a new toolbox for single-molecule experiments by atomic force microscopy. *Biophysical Journal*, 78(6):3275–3285, 2000.

[43] MB Viani, LI Pietrasanta, JB Thompson, A Chand, IC Gebeshuber, JH Kindt, M Richter, HG Hansma, and PK Hansma. Probing protein-protein interactions in real time. *Nature Structural & Molecular Biology*, 7(8):644–647, 2000.

[44] JP Cloarec, Y Chevolot, E Laurenceau, M Phaner-Goutorbe, and E Souteyrand. A multidisciplinary approach for molecular diagnostics based on biosensors and microarrays. *IRBM*, 29(2):105–127, 2008.

[45] TW Cha, A Guo, and XY Zhu. Enzymatic activity on a chip: the critical role of protein orientation. *Proteomics*, 5(2):416–419, 2005.

[46] A.D. McNaught and A. Wilkinson. *Compendium of Chemical Terminology*. International Union of Pure and Applied Chemistry, 1997.

[47] N Patel, MC Davies, M Hartshorne, RJ Heaton, CJ Roberts, SJB Tendler, and PM Williams. Immobilization of protein molecules onto homogeneous and mixed carboxylate-terminated self-assembled monolayers. *Langmuir*, 13(24):6485–6490, 1997.

[48] G Haran. Single-molecule fluorescence spectroscopy of biomolecular folding. *Journal of Physics: Condensed Matter*, 15(32):R1291, 2003.

[49] J. Israelachvili. *Intermolecular and Surface Forces*. London: Academic Press, 1992.

[50] M Basri, K Ampon, WM Yunus, CNA Razak, and AB Salleh. Immobilization of hydrophobic lipase derivatives on to organic polymer beads. *Journal of Chemical Technology and Biotechnology*, 59(1):37–44, 1994.

[51] MB Stark and K Holmberg. Covalent immobilization of lipase in organic solvents. *Biotechnology and Bioengineering*, 34(7):942–950, 1989.

[52] G Demirel, MO Çağlayan, B Garipcan, M Duman, and E Pişkin. Formation and organization of amino terminated self-assembled layers on Si (001) surface. *Nanoscale Research Letters*, 2(7):350–354, 2007.

[53] Mark Schena. *Protein Microarrays*. Jones nad Bartlett, 2005.

[54] JA Collins, C Xirouchaki, RE Palmer, JK Heath, and CH Jones. Clusters for biology: immobilization of proteins by size-selected metal clusters. *Applied Surface Science*, 226(1):197–208, 2004.

[55] M Di Vece, S Palomba, and RE Palmer. Pinning of size-selected gold and nickel nanoclusters on graphite. *Physical Review B*, 72(7):073407, 2005.

[56] C Leung, C Xirouchaki, N Berovic, and RE Palmer. Immobilization of protein molecules by size-selected metal clusters on surfaces. *Advanced Materials*, 16(3): 223–226, 2004.

[57] RE Palmer, S Pratontep, and HG Boyen. Nanostructured surfaces from size-selected clusters. *Nature Materials*, 2(7):443–448, 2003.

[58] RE Palmer and C Leung. Immobilisation of proteins by atomic clusters on surfaces. *TRENDS in Biotechnology*, 25(2):48–55, 2007.

[59] John Mongillo. *Nanotechnology*. Greenwood, 2007.

[60] M Moskovits. Metal clusters. *Annual Review of Physical Chemistry*, 42(1):465–499, 1991.

[61] MC Daniel and D Astruc. Gold nanoparticles: assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology. *Chemical Reviews-Columbus*, 104(1):293, 2004.

[62] C. Leung. *New Physical Methods for Characterization of Biological Molecules*. PhD thesis, University of Birmingham, 2005.

[63] U Prisco, C Leung, C Xirouchaki, CH Jones, JK Heath, and RE Palmer. Residue-specific immobilization of protein molecules by size-selected clusters. *Journal of The Royal Society Interface*, 2(3):169–175, 2005.

[64] GM Wang, WC Sandberg, and SD Kenny. Density functional study of a typical thiol tethered on a gold surface: ruptures under normal or parallel stretch. *Nanotechnology*, 17(19):4819, 2006.

[65] Peter Wagner. Immobilization strategies for biological scanning probe microscopy. *FEBS letters*, 430(1):112–115, 1998.

[66] H Zhang, G Schmid, and U Hartmann. Reduced metallic properties of ligand-stabilized small metal clusters. *Nano Letters*, 3(3):305–307, 2003.

[67] M Turner, VB Golovko, OPH Vaughan, P Abdulkin, A Berenguer-Murcia, MS Tikhov, BFG Johnson, and RM Lambert. Selective oxidation with dioxygen by gold nanoparticle catalysts derived from 55-atom clusters. *Nature*, 454 (7207):981–983, 2008.

[68] HG Boyen, G Kästle, F Weigl, B Koslowski, C Dietrich, P Ziemann, JP Spatz, S Riethmüller, C Hartmann, and M Möller. Oxidation-resistant gold-55 clusters. *Science*, 297(5586):1533–1536, 2002.

[69] G Schmid and B Corain. Nanoparticulated gold: syntheses, structures, electronics, and reactivities. *European Journal of Inorganic Chemistry*, 2003(17):3081–3098, 2003.

[70] F Rosei. Nanostructured surfaces: challenges and frontiers in nanotechnology. *Journal of Physics: Condensed Matter*, 16(17):S1373, 2004.

[71] L Bardotti, F Tournus, P Mélinon, M Pellarin, and M Broyer. Mass-selected clusters deposited on graphite: Spontaneous organization controlled by cluster surface reaction. *Physical Review B*, 83(3):035425, 2011.

[72] N Toshima. Metal nanoparticles for catalysis. In *Nanoscale Materials*, pages 79–96. Springer, 2004.

[73] SA Maier, PG Kik, HA Atwater, S Meltzer, E Harel, BE Koel, and AAG Requicha. Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides. *Nature Materials*, 2(4):229–232, 2003.

[74] SI Khondaker, K Luo, and Z Yao. The fabrication of single-electron transistors using dielectrophoretic trapping of individual gold nanoparticles. *Nanotechnology*, 21(9):095204, 2010.

[75] H Xie and S Régnier. High-efficiency automated nanomanipulation with parallel imaging/manipulation force microscopy. *Nanotechnology, IEEE Transactions on*, 11(1):21–33, 2012.

[76] D Dietzel, M Feldmann, H Fuchs, UD Schwarz, and A Schirmeisen. Transition from static to kinetic friction of metallic nanoparticles.

[77] CL Leung, RKH Liem, DAD Parry, and KJ Green. The plakin family. *Journal of Cell Science*, 114(19):3409–3410, 2001.

[78] C Ruhrberg and FM Watt. The plakin family: versatile organizers of cytoskeletal architecture. *Current Opinion in Genetics & Development*, 7(3):392–397, 1997.

[79] MG Mahoney, S Aho, J Uitto, and JR Stanley. The members of the plakin family of proteins recognized by paraneoplastic pemphigus antibodies include periplakin. *Journal of Investigative Dermatology*, 111(2):308–313, 1998.

[80] KJ Green, MLA Virata, GW Elgart, JR Stanley, and DAD Parry. Comparative structural analysis of desmoplakin, bullous pemphigoid antigen and plectin: members of a new gene family involved in organization of intermediate filaments. *International Journal of Biological Macromolecules*, 14(3):145–153, 1992.

[81] M Chidgey. Plakin proteins, hemidesmosomes and human disease. *eLS*.

[82] CL Leung, KJ Green, and RKH Liem. Plakins: a family of versatile cytolinker proteins. *Trends in Cell Biology*, 12(1):37–45, 2002.

[83] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. Cell junctions, cell adhesion, and the extracellular matrix. 2002.

[84] C Frantz, KM Stewart, and VM Weaver. The extracellular matrix at a glance. *Journal of Cell Science*, 123(24):4195–4200, 2010.

[85] J Wegener. Cell junctions. *eLS*, 2003.

[86] J Zhurinsky, M Shtutman, and A Ben-Ze'ev. Plakoglobin and beta-catenin: protein interactions, regulation and biological roles. *Journal of Cell Science*, 113(18): 3127–3139, 2000.

[87] D Garrod and M Chidgey. Desmosome structure, composition and function. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1778(3):572–587, 2008.

[88] S Getsios, AC Huen, and KJ Green. Working out the strength and flexibility of desmosomes. *Nature Reviews Molecular Cell Biology*, 5(4):271–281, 2004.

[89] AJ North, WG Bardsley, J Hyam, EA Bornslaeger, HC Cordingley, B Trinnaman, M Hatzfeld, KJ Green, AI Magee, and DR Garrod. Molecular map of the desmosomal plaque. *Journal of Cell Science*, 112(23):4325–4336, 1999.

[90] DL Stokes. Desmosomes from a structural perspective. *Current Opinion in Cell Biology*, 19(5):565–571, 2007.

[91] EJ O'keefe, HP Erickson, and V Bennett. Desmoplakin I and desmoplakin II. Purification and characterization. *Journal of Biological Chemistry*, 264(14):8310–8318, 1989.

[92] URL `http://www.ncbi.nlm.nih.gov/nuccore/58530839`.

[93] Proteomics. URL `http://expasy.org/tools/protparam.html`.

[94] HJ Choi, S Park-Snyder, LT Pascoe, KJ Green, and WI Weis. Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure. *Nature Structural & Molecular Biology*, 9(8):612–620, 2002.

[95] A Bateman, E Birney, L Cerruti, R Durbin, L Etwiller, SR Eddy, S Griffiths-Jones, KL Howe, M Marshall, and ELL Sonnhammer. The pfam protein families database. *Nucleic Acids Research*, 30(1):276–280, 2002.

[96] HJ Choi and WI Weis. Crystal structure of a rigid four-spectrin-repeat fragment of the human desmoplakin plakin domain. *Journal of Molecular Biology*, 409(5): 800–812, 2011.

[97] H Thomason, A Scothern, S McHarg, and D Garrod. Desmosomes: adhesive strength and signalling in health and disease. *Biochemical Journal*, 429:419–433, 2010.

[98] C Al-Jassar, T Knowles, M Jeeves, K Kami, E Behr, H Bikker, M Overduin, and M Chidgey. The nonlinear structure of the desmoplakin plakin domain and the effects of cardiomyopathy-linked mutations. *Journal of Molecular Biology*, 411(5): 1049–1061, 2011.

[99] KJ Green, DA Parry, PM Steinert, ML Virata, RM Wagner, BD Angst, and LA Nilles. Structure of the human desmoplakins. implications for function in the desmosomal plaque. *Journal of Biological Chemistry*, 265(5):2603–2612, 1990.

[100] L Janda, J Damborskỳ, GA Rezniczek, and G Wiche. Plectin repeats and modules: strategic cysteines and their presumed impact on cytolinker functions. *BioEssays*, 23(11):1064–1069, 2001.

[101] TS Stappenbeck and KJ Green. The desmoplakin carboxyl terminus coaligns with and specifically disrupts intermediate filament networks when expressed in cultured cells. *Journal of Cell Biology*, 116(5):1197–1209, 1992.

[102] PD Kouklis, E Hutton, and E Fuchs. Making a connection: direct binding between keratin intermediate filaments and desmosomal proteins. *Journal of Cell Biology*, 127(4):1049–1060, 1994.

[103] L Machesky R Insall. *Cytoskeleton*. John Wiley and Sons, Ltd, 2002.

[104] Jin-Jun Meng, Elayne A Bornslaeger, Kathleen J Green, Peter M Steinert, and Wallace Ip. Two-hybrid analysis reveals fundamental differences in direct interactions between desmoplakin and cell type-specific intermediate filaments. *Journal of Biological Chemistry*, 272(34):21495–21503, 1997.

[105] H Herrmann, H Bär, L Kreplak, SV Strelkov, and U Aebi. Intermediate filaments: from cell architecture to nanomechanics. *Nature Reviews Molecular Cell Biology*, 8(7):562–573, 2007.

[106] AA Chernyatina, S Nicolet, U Aebi, H Herrmann, and SV Strelkov. Atomic structure of the vimentin central α-helical domain and its implications for intermediate filament assembly. *Proceedings of the National Academy of Sciences*, 109(34): 13620–13625, 2012.

[107] RD Goldman, S Khuon, YH Chou, P Opal, and PM Steinert. The function of intermediate filaments in cell shape and cytoskeletal integrity. *Journal of Cell Biology*, 134(4):971–983, 1996.

[108] E Fuchs and K Weber. Intermediate filaments: structure, dynamics, function and disease. *Annual Review of Biochemistry*, 63(1):345–382, 1994.

[109] H Bär, SV Strelkov, G Sjöberg, U Aebi, and H Herrmann. The biology of desmin filaments: how do mutations affect their structure, assembly, and organisation? *Journal of Structural Biology*, 148(2):137–152, 2004.

[110] AC Huen, JK Park, LM Godsel, X Chen, LJ Bannon, EV Amargo, TY Hudson, AK Mongiu, IM Leigh, and DP Kelsell. Intermediate filament–membrane attachments function synergistically with actin-dependent contacts to regulate intercellular adhesive strength. *Journal of Cell Biology*, 159(6):1005–1017, 2002.

[111] A Satelli and S Li. Vimentin in cancer and its potential as a molecular target for cancer therapy. *Cellular and Molecular Life Sciences*, 68(18):3033–3046, 2011.

[112] JE Cheong, V Wessagowit, and JA McGrath. Molecular abnormalities of the desmosomal protein desmoplakin in human disease. *Clinical and Experimental Dermatology*, 30(3):261–266, 2005.

[113] K Andrä, H Lassmann, R Bittner, S Shorny, R Fässler, F Propst, and G Wiche. Targeted inactivation of plectin reveals essential function in maintaining the integrity of skin, muscle, and heart cytoarchitecture. *Genes & Development*, 11(23): 3143–3156, 1997.

[114] C Gemayel, A Pelliccia, and PD Thompson. Arrhythmogenic right ventricular cardiomyopathy. *Journal of the American College of Cardiology*, 38(7):1773–1781, 2001.

[115] R Alcalai, S Metzger, S Rosenheck, V Meiner, and T Chajek-Shaul. A recessive mutation in desmoplakin causes arrhythmogenic right ventricular dysplasia, skin disorder, and woolly hair. *Journal of the American College of Cardiology*, 42(2): 319–327, 2003.

[116] N Protonotarios and A Tsatsopoulou. Naxos disease and carvajal syndrome: cardiocutaneous disorders that highlight the pathogenesis and broaden the spectrum of arrhythmogenic right ventricular cardiomyopathy. *Cardiovascular Pathology*, 13 (4):185–194, 2004.

[117] PH Brown and P Schuck. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophysical Journal*, 90(12): 4651–4661, 2006.

[118] SM Kelly and NC Price. The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science*, 1(4):349–384, 2000.

[119] L Whitmore and BA Wallace. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, 89(5): 392–400, 2008.

[120]

[121] TDW Claridge. *High-resolution NMR techniques in organic chemistry*, volume 27. Elsevier Science, 2008.

[122] Agilent. 2011.

[123] Instrumentation New Mexico State University, NMSU. URL `http://web.nmsu.edu/~kburke/Instrumentation/NMR_Table.html`.

[124] P Thordarson. Determining association constants from titration experiments in supramolecular chemistry. *Chemical Society Reviews*, 40(3):1305–1323, 2011.

[125] H Nesemann, W Brefeld, F Brinker, W Decking, O Kaul, and B Sarau. DORIS III as a dedicated source for synchrotron radiation. In *Particle Accelerator Conference, 1995., Proceedings of the 1995*, volume 1, pages 195–197. IEEE, 1995.

[126] AR Round, D Franke, S Moritz, R Huchler, M Fritsche, D Malthan, R Klaering, DI Svergun, and M Roessle. Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *Journal of Applied Crystallography*, 41(5):913–917, 2008.

[127] R Howland and L Benatar. *A Practical Guide: To Scanning Probe Microscopy*. Park scientific instruments, 1996.

[128] Greifswald University. URL `http://www3.physik.uni-greifswald.de/method/afm/eafm.htm`.

[129] Digital Instruments Manual. 2000.

[130] Park Systems. 2008.

[131] S Pratontep, SJ Carroll, C Xirouchaki, M Streun, and RE Palmer. Size-selected cluster beam source based on radio frequency magnetron plasma sputtering and gas condensation. *Review of Scientific Instruments*, 76(4):045103–045103, 2005.

[132] B Von Issendorff and RE Palmer. A new high transmission infinite range mass selector for cluster and nanoparticle beams. *Review of Scientific Instruments*, 70 (12):4497–4501, 1999.

[133] SJ Carroll, S Pratontep, M Streun, RE Palmer, S Hobday, and R Smith. Pinning of size-selected Ag clusters on graphite surfaces. *Journal of Chemical Physics*, 113 (18):7723, 2000.

[134] James Keeler. *Understanding NMR spectroscopy*. Wiley, 2011.

[135] CH Yoder and CD Schaeffer. *Introduction to multinuclear NMR*. Benjamin/Cummings Pub. Co., 1987.

[136] PJ Hore. *Nuclear magnetic resonance*, volume 18. Oxford University Press New York, 1995.

[137] FA Carey. *Organic Chemistry*. McGraw-Hill, 2001.

[138] G Bodenhausen and DJ Ruben. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1):185–189, 1980.

[139] DS Wishart, BD Sykes, and FM Richards. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *Journal of Molecular Biology*, 222(2):311–333, 1991.

[140] CT Jones, L Ma, JW Burgner, TD Groesch, CB Post, and RJ Kuhn. Flavivirus capsid is a dimeric alpha-helical protein. *Journal of Virology*, 77(12):7143–7149, 2003.

[141] FM Poulsen. A brief introduction to NMR spectroscopy of proteins, 2002.

[142] LE Kay, M Ikura, R Tschudin, and A Bax. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance*, 89(3):496–514, 1990.

[143] M Wittekind and L Mueller. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-and beta-carbon resonances in proteins. *Journal of Magnetic Resonance. Series B*, 101(2):201–205, 1993.

[144] S Grzesiek and A Bax. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society*, 114(16):6291–6293, 1992.

[145] RT Clubb, V Thanabal, and G Wagner. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ^{1}HN, ^{15}N, and $^{13}C'$chemical shifts in ^{15}N-^{13}C-labelled proteins. 1992.

[146] EL Ulrich, H Akutsu, JF Doreleijers, Y Harano, YE Ioannidis, J Lin, M Livny, S Mading, D Maziuk, and Z Miller. Biomagresbank. *Nucleic Acids Research*, 36 (suppl 1):D402–D408, 2008.

[147] JH Noggle and RE Schirmer. *The nuclear Overhauser effect: chemical applications*. Academic Press New York, 1971.

[148] JW Keepers and TL James. A theoretical study of distance determinations from NMR. two-dimensional nuclear Overhauser effect spectra. *Journal of Magnetic Resonance*, 57(3):404–426, 1984.

[149] DHA Corrêa and CHI Ramos. The use of circular dichroism spectroscopy to study protein folding, form and function. *African Journal of Biochemistry Research*, 3 (5):164–173, 2009.

[150] R Fogh, J Ionides, E Ulrich, W Boucher, W Vranken, JP Linge, M Habeck, W Rieping, TN Bhat, and J Westbrook. The CCPN project: an interim report on a data model for the NMR community. *Nature Structural Biology*, 9(6):416–418, 2002.

[151] TD Goddard and DG Kneller. SparkyNMR assignment and integration software. *University of California, San Francisco*, 2006.

[152] JF Doreleijers, ML Raves, T Rullmann, and R Kaptein. Completeness of NOEs in protein structures: A statistical analysis of NMR data. *Journal of Biomolecular NMR*, 14(2):123–132, 1999.

[153] N Ci. Stercochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7:95499, 1963.

[154] G Cornilescu, F Delaglio, and A Bax. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR*, 13(3):289–302, 1999.

[155] RE Hubbard and M Kamran Haider. Hydrogen bonds in proteins: Role and strength. *eLS*, 1963.

[156] Paul Schanda, Ē Kupče, and B Brutscher. SOFAST-HMQC experiments for recording two-dimensional deteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR*, 33(4):199–211, 2005.

[157] SW Englander and NR Kallenbach. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Quarterly Reviews of Biophysics*, 16(04):521–655, 1983.

[158] M Nilges. Calculation of protein structures with ambiguous distance restraints. automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *Journal of Molecular Biology*, 245(5):645–660, 1995.

[159] M Nilges, MJ Macias, SI ODonoghue, and H Oschkinat. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *Journal of Molecular Biology*, 269(3):408–422, 1997.

[160] AJ Caezar. *Biophysical characterization of the plakin family.* PhD thesis, University of Birmingham, 2012.

[161] PV Konarev, VV Volkov, AV Sokolova, MHJ Koch, and DI Svergun. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36(5):1277–1282, 2003.

[162] DI Svergun. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25(4):495–503, 1992.

[163] D Svergun, C Barberato, and MHJ Koch. CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, 28(6):768–773, 1995.

[164] DS Wishart, BD Sykes, and FM Richards. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, 31(6):1647–1651, 1992.

[165] LJ McGuffin, K Bryson, and DT Jones. The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4):404–405, 2000.

[166] Y Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9(1):40, 2008.

[167] T Akutsu and KL Sim. Protein threading based on multiple protein structure alignment. *Genome Informatics Series*, pages 23–29, 1999.

[168] B Nikolic, E MacNulty, B Mir, and G Wiche. Basic amino acid residue cluster within nuclear targeting sequence motif is essential for cytoplasmic plectin-vimentin network junctions. *Journal of Cell Biology*, 134(6):1455–1467, 1996.

[169] T DiColandrea, T Karashima, A Määttä, and FM Watt. Subcellular distribution of envoplakin and periplakin insights into their role as precursors of the epidermal cornified envelope. *Journal of Cell Biology*, 151(3):573–586, 2000.

[170] L Fontao, B Favre, S Riou, D Geerts, F Jaunin, JH Saurat, KJ Green, A Sonnenberg, and L Borradori. Interaction of the bullous pemphigoid antigen 1 (BP230) and desmoplakin with intermediate filaments is mediated by distinct sequences within their COOH terminus. *Molecular Biology of the Cell*, 14(5):1978–1992, 2003.

[171] S Kazerounian, J Uitto, and S Aho. Unique role for the periplakin tail in intermediate filament association: specific binding to keratin 8 and vimentin. *Experimental Dermatology*, 11(5):428–438, 2002.

[172] T Karashima and FM Watt. Interaction of periplakin and envoplakin with intermediate filaments. *Journal of Cell Science*, 115(24):5027–5037, 2002.

[173] Andrew P Kowalczyk and Kathleen J Green. The desmosome: a component system for adhesion and intermediate filament attachment. *Current topics in Membranes*, 43:187–209, 1996.

[174] KJ Green and CA Gaudry. Are desmosomes more than tethers for intermediate filaments? *Nature Reviews Molecular Cell Biology*, 1(3):208–216, 2000.

[175] K Lapouge, L Fontao, MF Champliaud, F Jaunin, MA Frias, B Favre, D Paulin, KJ Green, and L Borradori. New insights into the molecular basis of desmoplakinand desmin-related cardiomyopathies. *Journal of Cell Science*, 119(23):4974–4985, 2006.

[176] TL James. Fundamentals of NMR. *Department of Pharmaceutical Chemistry University of California San Francisco, CA 94143-0446 USA*, 1998.

[177] TS Stappenbeck, EA Bornslaeger, CM Corcoran, HH Luu, ML Virata, and KJ Green. Functional analysis of desmoplakin domains: specification of the interaction with keratin versus vimentin intermediate filament networks. *Journal of Cell Biology*, 123(3):691–705, 1993.

[178] Sergei V Strelkov, Jens Schumacher, Peter Burkhard, Ueli Aebi, and Harald Herrmann. Crystal structure of the human lamin a coil 2b dimer: implications for the head-to-tail association of nuclear lamins. *Journal of molecular biology*, 343 (4):1067–1080, 2004.

[179] PH Jonson and SB Petersen. A critical view on conservative mutations. *Protein Engineering*, 14(6):397–402, 2001.

[180] SPMagic. URL `http://www.spmagic.com/afm_techniques.htm`.

[181] R Rounsevell, JR Forman, and J Clarke. Atomic force microscopy: mechanical unfolding of proteins. *Methods*, 34(1):100–111, 2004.

[182] Thomas Stifter, Othmar Marti, and Bharat Bhushan. Theoretical investigation of the distance dependence of capillary and van der waals forces in scanning force microscopy. *Physical Review B*, 62(20):13667, 2000.

[183] Shan Zou. *Exploring individual supramolecular interactions and stimuli-responsive polymers by AFM-based force spectroscopy*. PhD thesis, University of Twente, 2005.

[184] Y Martin, CC Williams, and HK Wickramasinghe. Atomic force microscope-force mapping and profiling on a sub 100-å scale. *Journal of Applied Physics*, 61(10): 4723–4729, 1987.

[185] Park Systems. URL `http://www.parkafm.com`.

[186] Q Zhong, D Inniss, K Kjoller, and VB Elings. Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. *Surface Science Letters*, 290(1):L688–L692, 1993.

[187] JR Forman, S Qamar, E Paci, RN Sandford, and J Clarke. The remarkable mechanical strength of polycystin-1 supports a direct role in mechanotransduction. *Journal of Molecular Biology*, 349(4):861–871, 2005.

[188] M Rief, M Gautel, F Oesterhelt, JM Fernandez, and HE Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315): 1109–1112, 1997.

[189] LA Pérez, X López-Lozano, and IL Garzón. Density functional study of the cysteine adsorption on Au nanoclusters. *The European Physical Journal D*, 52(1-3): 123–126, 2009.

[190] Inc. Wolfram Research. URL `http://www.periodictable.com/Properties/A/Electronegativity.al.html`.

[191] MC Bourg, A Badia, and RB Lennox. Gold-sulfur bonding in 2D and 3D self-assembled monolayers: XPS characterization. *Journal of Physical Chemistry B*, 104(28):6562–6567, 2000.

[192] DG Castner, K Hinds, and DW Grainger. X-ray photoelectron spectroscopy sulfur 2p study of organic thiol and disulfide binding interactions with gold surfaces. *Langmuir*, 12(21):5083–5086, 1996.

[193] M Grandbois, M Beyer, M Rief, H Clausen-Schaumann, and HE Gaub. How strong is a covalent bond? *Science*, 283(5408):1727–1730, 1999.

[194] Haschemeyer and Haschemeyer. *A Guide to Study by Physical and Chemical Methods*. Wiley-Interscience, 1973.

[195] B Lee and FM Richards. The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology*, 55(3):379–IN4, 1971.

[196] M Yudkin and R Offord. *A Guide to Biochemistry*. Cambridge University Press, 1971.

[197] KL Marchin and CL Berrie. Conformational changes in the plasma protein fibrinogen upon adsorption to graphite and mica investigated by atomic force microscopy. *Langmuir*, 19(23):9883–9888, 2003.

[198] RW Carpick, N Agrait, DF Ogletree, and M Salmeron. Variation of the interfacial shear strength and adhesion of a nanometer-sized contact. *Langmuir*, 12(13):3334–3340, 1996.

[199] Dirk Dietzel, Tristan Monninghoff, Lars Jansen, Harald Fuchs, Claudia Ritter, Udo D Schwarz, and André Schirmeisen. Interfacial friction obtained by lateral manipulation of nanoparticles using atomic force microscopy techniques. *Journal of Applied Physics*, 102(8):084306–084306, 2007.

[200] SJ Prestrelski, N Tedeschi, T Arakawa, and JF Carpenter. Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers. *Biophysical Journal*, 65(2):661–671, 1993.

[201] G Diakova, YA Goddard, JP Korb, and RG Bryant. Changes in protein structure and dynamics as a function of hydration from ^1H second moments. *Journal of Magnetic Resonance*, 189(2):166–172, 2007.

[202] FENG Yin, Chrisa Xirouchaki, QUANMIN Guo, and Richard E Palmer. High-temperature stability of size-selected gold nanoclusters pinned on graphite. *Advanced Materials*, 17(6):731–734, 2005.

[203] N Lidgi-Guigui, C Leung, and RE Palmer. Weak precursor state binding of protein molecules to size-selected gold nanoclusters on surfaces. *Surface Science*, 602(4): 1006–1009, 2008.

[204] V Silin, H Weetall, and DJ Vanderah. SPR studies of the nonspecific adsorption kinetics of human igG and BSA on gold surfaces modified by self-assembled monolayers (SAMs). *Journal of Colloid and Interface Science*, 185(1):94–103, 1997.

[205] BHJ Hofstee. Accessible hydrophobic groups of native proteins. *Biochemical and Biophysical Research Communications*, 63(3):618–624, 1975.

[206] Élisabeth Lojou and Pierre Bianco. Buildup of polyelectrolyte-protein multilayer assemblies on gold electrodes. Role of the hydrophobic effect. *Langmuir*, 20(3): 748–755, 2004.

[207] Paul Roach, David Farrar, and Carole C Perry. Interpretation of protein adsorption: surface-induced conformational changes. *Journal of the American Chemical Society*, 127(22):8168–8173, 2005.

[208] LG Butler. Enzyme immobilization by adsorption on hydrophobic derivatives of cellulose and other hydrophilic materials. *Archives of Biochemistry and Biophysics*, 171(2):645–650, 1975.

[209] JP Rossell, S Allen, MC Davies, CJ Roberts, SJB Tendler, and PM Williams. Electrostatic interactions observed when imaging proteins with the atomic force microscope. *Ultramicroscopy*, 96(1):37–46, 2003.

[210] TJ Senden and CJ Drummond. Surface chemistry and tip-sample interactions in atomic force microscopy. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 94(1):29–51, 1995.

[211] DJ Müller and YF Dufrene. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. *Nature Nanotechnology*, 3(5):261–269, 2008.

[212] M Tripathi, G Paolicelli, S DAddato, and S Valeri. Controlled AFM detachments and movement of nanoparticles: gold clusters on HOPG at different temperatures. *Nanotechnology*, 23(24):245706, 2012.

[213] P Deltour, JL Barrat, and P Jensen. Fast diffusion of a Lennard-Jones cluster on a crystalline surface. *Physical Review Letters*, 78(24):4597–4600, 1997.

[214] L Bardotti, P Jensen, A Hoareau, M Treilleux, and B Cabaud. Experimental observation of fast diffusion of large antimony clusters on graphite surfaces. *Physical Review Letters*, 74(23):4694–4697, 1995.

[215] R Guerra, U Tartaglino, A Vanossi, and E Tosatti. Ballistic nanofriction. *Nature Materials*, 9(8):634–637, 2010.

[216] L Bardotti, B Prevel, P Jensen, M Treilleux, P Melinon, A Perez, J Gierak, G Faini, and D Mailly. Organizing nanoclusters on functionalized surfaces. *Applied Surface Science*, 191(1):205–210, 2002.

[217] A Vinelli, E Primiceri, M Brucale, G Zuccheri, R Rinaldi, and B Samorì. Sample preparation for the quick sizing of metal nanoparticles by atomic force microscopy. *Microscopy Research and Technique*, 71(12):870–879, 2008.

[218] OD Häberlen, SC Chung, M Stener, and N Rösch. From clusters to bulk: A relativistic density functional investigation on a series of gold clusters Au, n= 6,, 147. *Journal of Chemical Physics*, 106:5189, 1997.

[219] LJ Lewis, P Jensen, N Combe, and JL Barrat. Diffusion of gold nanoclusters on graphite. *Physical Review B*, 61(23):16084, 2000.

[220] F Claeyssens, S Pratontep, C Xirouchaki, and RE Palmer. Immobilization of large size-selected silver clusters on graphite. *Nanotechnology*, 17(3):805, 2006.

[221] DJ Müller, D Fotiadis, S Scheuring, SA Müller, and A Engel. Electrostatically balanced subnanometer imaging of biological specimens by atomic force microscope. *Biophysical Journal*, 76(2):1101–1111, 1999.

[222] DJ Kenny, RE Palmer, CF Sanz-Navarro, and R Smith. Implantation depth of size-selected silver clusters into graphite. *Journal of Physics: Condensed Matter*, 14(8):L185, 2002.

[223] SJ Carroll, PD Nellist, RE Palmer, S Hobday, and R Smith. Shallow implantation of size-selected Ag clusters into graphite. *Physical Review Letters*, 84(12):2654–2657, 2000.

[224] S Pratontep, P Preece, C Xirouchaki, RE Palmer, CF Sanz-Navarro, SD Kenny, and R Smith. Scaling relations for implantation of size-selected Au, Ag, and Si clusters into graphite. *Physical Review Letters*, 90(5):055503, 2003.

[225] P Jensen, X Blase, and P Ordejón. First principles study of gold adsorption and diffusion on graphite. *Surface Science*, 564(1):173–178, 2004.

[226] H Hövel and I Barke. Morphology and electronic structure of gold clusters on graphite: Scanning-tunneling techniques and photoemission. *Progress in Surface Science*, 81(2):53–111, 2006.

[227] S Rapino and F Zerbetto. Dynamics of thiolate chains on a gold nanoparticle. *Small*, 3(3):386–388, 2007.

[228] JC Garcia-Martinez and RM Crooks. Extraction of Au nanoparticles having narrow size distributions from within dendrimer templates. *Journal of the American Chemical Society*, 126(49):16170–16178, 2004.

[229] Z Wang, A Abdela, F Yin, and R Palmer. Three-dimensional atomic structure and surface dynamics of size-selected Au_{923} clusters. *Microscopy and Microanalysis*, 16(S2):1646–1647, 2010.

[230] VN Popok, I Barke, EEB Campbell, and KH Meiwes-Broer. Cluster-surface interaction: From soft landing to implantation. *Surface Science Reports*, 66(10):347–377, 2011.

[231] J Bansmann, SH Baker, C Binns, JA Blackman, JP Bucher, J Dorantes-Dávila, V Dupuis, L Favre, D Kechrakos, and A Kleibert. Magnetic and structural properties of isolated and assembled clusters. *Surface Science Reports*, 56(6):189–275, 2005.

[232] D Dietzel, M Feldmann, C Herding, UD Schwarz, and A Schirmeisen. Quantifying pathways and friction of nanoparticles during controlled manipulation by contact-mode atomic force microscopy. *Tribology Letters*, 39(3):273–281, 2010.

[233] ZW Wang and RE Palmer. Direct atomic imaging and dynamical fluctuations of the tetrahedral Au_{20} cluster. *Nanoscale*, 4(16):4947–4949, 2012.

[234] X Han and YX Gan. Investigation the complex dynamic evolvement mechanism of particle cluster and surface integrity in the chemical mechanical planarization. *The International Journal of Advanced Manufacturing Technology*, 64(1-4):13–22, 2013.

[235] The Chris Hill Lab. URL `https://wasatch.biochem.utah.edu/chris/`.

[236] Invitrogen. URL `http://www.invitrogen.com/search/global/searchAction.action?query=pproex`.

[237] Addgene. URL `http://www.addgene.org/12652/`.

[238] S Mori, C Abeygunawardana, MO Johnson, and PCM Vanzijl. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *Journal of Magnetic Resonance, Series B*, 108(1):94–98, 1995.

[239] P Schanda, H Van Melckebeke, and B Brutscher. Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of the American Chemical Society*, 128(28):9042–9043, 2006.

[240] S Grzesiek, J Anglister, and A Bax. Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of ^{13}C magnetization. *Journal of Magnetic Resonance, Series B*, 101(1):114–119, 1993.

[241] LE Kay, GY Xu, AU Singer, DR Muhandiram, and JD Formankay. A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples of proteins. *Journal of Magnetic Resonance, Series B*, 101 (3):333–337, 1993.

[242] Protein NMR, a practical guide.

[243] A Bax, GM Clore, and AM Gronenborn. $^{1}H^{1}H$ correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *Journal of Magnetic Resonance*, 88(2):425–431, 1990.

[244] D Marion, PC Driscoll, LE Kay, PT Wingfield, A Bax, AM Gronenborn, and GM Clore. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1. beta. *Biochemistry*, 28(15):6150–6156, 1989.