Development of mass spectrometry-based carbene footprinting strategies for the study of protein structure and interactions

Lucio Manzi

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

July 2017

Abstract

Protein interactions are crucial for the survival of living organisms. The possibility of mapping the contact surfaces between proteins and their interacting partners is fundamental to understanding the mechanisms involved in the process. For these reasons techniques able to provide structural information on a short time scale and employing small amounts of material are sought after. The work reported in this thesis explores the use of carbene-based labelling in combination with mass spectrometry for protein footprinting and its applications in the study of protein structure and interactions.

Studies on the efficiency and selectivity of a novel water-soluble photo-activated probe revealed its superior properties in comparison with diazirine-based reagents previously described for the same application. Using this methodology, the contact surface of the complex between lysozyme and NAG5, a carbohydrate substrate, was accurately mapped. The same technique was successfully employed to shed light on the structural change occurring to USP5, a large multi-domain deubiquitinating enzyme, upon its binding to diubiquitin. The use of carbene footprinting in combination with other biophysical techniques allowed to characterise the spatial arrangement of domains located at a module junction in the large multi-modular gladiolin polyketide synthase paving the way for future efforts by synthetic biologists to hijack the chemistry of this antibiotic-producing multiprotein enzyme to produce novel active compounds.

The possibility of using carbene-based footprinting to gain insight into the structure of integral membrane proteins was also explored. The probe introduced in this work exhibited peculiar labelling properties when activated in the presence of a detergent-solubilised membrane protein. The reagent selectively reacted with portion of the protein in contact with detergent molecules showing potential to elucidate the quaternary structure of multimeric membrane proteins.

Acknowledgements

I would like to express my utmost gratitude to Prof.Neil Oldham for giving me the opportunity to work on this project and I am immensely thankful for his support, guidance and patience. The friendly and approachable interactions with him and his willingness to listen to my ideas with an open mind were essential to build my confidence as a scientist and to help me develop critical thinking.

I am thankful to past and present members of the Oldham group, in particular Dr.Daniel Scott for his help with the expression, purification, activity assays of USP5 and for many interesting tennis matches and to Dr.Matthew Jenner, for his support with the PKS work and for his great interest in the italian language. I am also indebted to Nkazi Tshuma and Dr. Jose Afonso for many fruitful discussions and for their support whenever I was "disappointed by science". I would also like to thank Jed Bellamy-Carter for his help with any computer-related issue and for conditioning me into locking my computer when leaving my desk to avoid pranks. I am thankful to you all for creating a fun and stimulating environment to work in. I am also grateful to Dr.John Moses and Dr.Andrew Barrow for synthesising some of the photoreagents that were essential for the work detailed in this thesis.

I would like to specially mention Dr.Roberto Mosillo because without his help I probably would have not been able to start my studies.

I want to give special thanks to Alyss for her constant love, support and reassurance, especially during the most difficult moments. She had an enormous part in the completion of this work. Finally and most importantly I would like to thank my parents, without their constant encouragement and support none of this would be possible.

Publications

Carbene footprinting accurately maps binding sites in protein-ligand and protein-protein interactions.

***Lucio Manzi**, Andrew S. Barrow, Daniel Scott, Rob Layfield, Timothy G. Wright, John E. Moses, Neil J. Oldham*

Nature Communications, 7, 13288 (2016)

Mechanism of Subunit Interaction at Ketosynthase-Dehydratase Junctions in trans-AT Polyketide Synthases

*Matthew Jenner, Simone Kosol, Daniel Griffiths, Panward Prasongpholchai, **Lucio Manzi**, Andrew S. Barrow, John E. Moses, Neil J. Oldham, Jozef R. Lewandowski, Gregory L. Challis*

Nature Chemical Biology, Manuscript Under Review

Carbene footprinting reveals binding interfaces of a multimeric membrane spanning protein

***Lucio Manzi**, Andrew S. Barrow, Jonathan T.S. Hopper, Colin Kleanthous, Carol V. Robinson, John E. Moses, Neil J. Oldham*

Manuscript in Preparation

List of Abbreviations

ACP	Acyl Carrier Protein
AD	Sodium 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoate
AT	Acyl Transferase
CCS	Collisional Cross Section
CD	Circular Dichroism
CEM	Chain Ejection Model
CI	Chemical Ionisation
CID	Collision Induced Dissociation
CMC	Critical Micellar Concentration
CoA	Coenzyme A
CRM	Charge Residue Model
CSD	Charge State Distribution
DDA	Data Dependent Acquisition
DEBS	6-deoxyerythronolide B Synthase
DH	Dehydratase
DP	Docking Peptide

DUB	Deubiquitinase
ECD	Electron Capture Dissociation
EDTA	Ethylenediaminetetraacetic Acid
EI	Electron Impact Ionisation
EL	Extent of Labelling
ER	Enoylreductase
ESI	Electrospray Ionisation
ETD	Electron Transfer Dissociation
FAS	Fatty Acid Synthase
FPOP	Fast Photochemical Oxidation of Protein
FT-ICR	Fourier Transform Ion Cyclotron Resonance
HDX	Hydrogen-Deuterium Exchange
HPLC	High Performance Liquid Chromatography
IDP	Intrinsically Disordered Protein
IEM	Ion Evaporation Model
IgG	Immunoglobulin G
IMP	Integral Membrane Protein
IMS	Ion Mobility Mass Spectrometry
IRMPD	Infra-Red Multi-Photon Dissociation
ITC	Isothermal Titration Calorimetry
KR	Ketoreductase

KS	Ketosynthase
LC	Liquid Chromatography
MALDI	Matrix Assisted Laser Desorption Ionisation
MS	Mass Spectrometry
NAG5	Penta-N-acetylchitopentaose
NMR	Nuclear Magnetic Resonance
OG	Octyl Glucoside
PKS	Polyketide Synthase
PPant	Phosphopantetheine
Q-TOF	Quadrupole-Time Of Flight
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
TWIG	Travelling Wave Ion Guide
UBA	Ubiquitin Associated Domain
UBD	Ubiquitin Binding Domain
USP5	Ubiquitin Specific Protease 5
UV	Ultraviolet

Contents

1 Introduction **1**

1.1 Protein structure 1

1.1.1 Levels of protein structure 2

1.2 Methods for studying protein structure 3

1.2.1 X-ray crystallography 3

1.2.2 Nuclear Magnetic Resonance (NMR) 4

1.3 Structural mass spectrometry 5

1.3.1 Ion mobility mass spectrometry 5

1.3.2 ESI charge state distribution 6

1.3.3 Hydrogen-Deuterium exchange 7

1.3.4 Covalent labelling 9

1.3.4.1 Amino acid-specific labelling 10

1.3.5 Hydroxyl radical footprinting 11

1.3.5.1 Generation of hydroxyl radicals 11

1.3.5.2 Hydroxyl radical reactivity 14

1.3.5.3 Protein interactions studied by hydroxyl radical footprinting 15

1.3.5.4 Fast Photochemical Oxidation of Proteins (FPOP) 16

1.3.6 Carbenes 18

1.3.7 Aryldiazirines as photoactivable reagents 19

1.3.8 Carbene footprinting . . . 21
1.3.8.1 Diazirine gas . . . 21
1.3.8.2 Methylene footprinting coupled with scintillation detection . . . 23
1.3.8.3 Methylene footprinting coupled with mass spectrometry detection . . . 24
1.3.8.4 Water-soluble diazirine footprinting . . . 26
1.4 Mass Spectrometry . . . 31
1.4.1 Ionisation techniques . . . 32
1.4.1.1 Electrospray ionisation . . . 32
1.4.2 Mass Analysers . . . 36
1.4.2.1 Ion trap analysers . . . 36
1.4.2.2 FT-ICR analysers . . . 38
1.4.3 Tandem mass spectrometry . . . 41
1.4.3.1 Collision Induced Dissociation . . . 42

2 Materials and Methods 45

2.1 Materials . . . 45
2.1.1 Buffers and reagents . . . 45
2.1.2 Characterisation of AD reactivity . . . 45
2.1.3 Activity assays . . . 46
2.1.4 Protein footprinting . . . 46
2.2 Instruments . . . 46
2.2.1 Thermo Scientific LTQ FT-ULTRA . . . 46
2.2.2 Waters SYNAPT HDMS . . . 47
2.3 Methods . . . 48
2.3.1 Photochemical labelling . . . 48

2.3.1.1 Lysozyme . . . 48
2.3.1.2 C335 USP5 . . . 49
2.3.1.3 DH-ACP . . . 50
2.3.1.4 DH-ACP/^{DH}DP . . . 50
2.3.1.5 OmpF . . . 50
2.3.1.6 Characterisation of AD reactivity . . . 50
2.3.1.7 Labelling efficiency . . . 51
2.3.1.8 Amino acid selectivity . . . 51
2.3.2 Sample preparation . . . 51
2.3.2.1 Bottom-up analysis . . . 51
2.3.2.1.1 SDS-PAGE . . . 51
2.3.2.1.2 Protein digestion . . . 52
2.3.2.2 Intact proteins . . . 53
2.3.2.2.1 ZipTip clean up . . . 53
2.3.2.3 Photolysis products . . . 53
2.3.2.4 Amino acid selectivity . . . 53
2.3.3 Protein digests LC-MS analysis . . . 54
2.3.3.1 nanoHPLC . . . 54
2.3.3.1.1 Lysozyme and USP5 . . . 54
2.3.3.1.2 DH-ACP and OmpF . . . 55
2.3.3.2 Mass spectrometry . . . 55
2.3.4 AD efficiency . . . 56
2.3.5 Structural integrity assessment . . . 56
2.3.5.1 Lysozyme activity assay . . . 56
2.3.5.2 USP5 activity assay . . . 56
2.3.5.3 Circular dichroism . . . 57

2.3.6 Data analysis . . . 57
2.3.6.1 Determination of the Extent of Labelling on intact proteins 57
2.3.6.2 Peptide level analysis . . . 58
2.3.6.3 Residue level analysis . . . 58

3 Introducing a novel photoreagent for carbene footprinting: reactivity characterisation and mapping of the lysozyme-NAG5 binding interface 60

3.1 Introduction . . . 60
3.2 Results and discussion . . . 63
3.2.1 Design of a new footprinting probe . . . 63
3.2.2 Characterisation of AD reactivity . . . 64
3.2.3 Labelling efficiency of AD . . . 65
3.2.4 Selectivity towards amino acids . . . 68
3.2.5 Structural integrity studies . . . 70
3.2.6 Lysozyme-NAG5 footprinting . . . 75
3.3 Conclusions . . . 84

4 Investigating the binding interface and conformational flexibility of the USP5-diubiquitin complex by carbene footprinting 86

4.1 Introduction . . . 86
4.1.1 USP5 . . . 88
4.2 Results and discussion . . . 90
4.2.1 Structural integrity studies . . . 90
4.2.2 Peptide level analysis . . . 92
4.2.2.1 Proteolytic Digestion . . . 92
4.2.2.2 USP5-Diubiquitin Footprinting . . . 94
4.2.3 Residue level analysis . . . 97
4.3 Conclusions . . . 102

5 Mapping the contact surface across a KS/DH split module in the Gladiolin trans-AT Polyketide Synthase. 104

5.1 Introduction . . . 104

5.1.1 Polyketides . . . 104

5.1.2 Polyketide Synthases . . . 105

5.1.3 Modular polyketide synthases . . . 106

5.2 Results and discussion . . . 109

5.2.1 Bioinformatic identification of ^{DH}DP . . . 109

5.2.2 Biochemical characterisation of the KS(^{DH}DP)/DH junction . . . 110

5.2.3 Mapping the KS//DH split module junction interface . . . 112

5.2.3.1 KS//DH peptide level analysis . . . 112

5.2.3.2 KS//DH residue level analysis . . . 115

5.2.3.2.1 DH binding surface . . . 115

5.2.3.2.2 ^{DH}DP binding surface . . . 117

5.2.4 Mapping the DH/ACP interaction . . . 119

5.2.4.1 DH/ACP peptide level analysis . . . 119

5.2.4.2 DH//ACP residue level analysis . . . 121

5.2.4.2.1 DH binding surface . . . 121

5.2.4.2.2 ACP binding surface . . . 123

5.2.4.2.3 DH-ACP modelling . . . 124

5.3 Conclusions . . . 126

6 Carbene footprinting reveals binding interfaces of a multimeric membrane spanning protein 129

6.1 Introduction . . . 129

6.1.1 MS-based techniques for the study of MPs . . . 130

6.1.2 OmpF . . . 131

6.2 Results and discussion 133
6.2.1 Footprinting of OmpF in OG micelles 133
6.2.1.1 Peptide level analysis 133
6.2.1.2 Residue level analysis 136
6.2.1.3 Carbene footprinting of OmpF in OG micelles using Photoleucine 138
6.2.2 Characterisation of AD reactivity in OG 139
6.2.2.1 AD distribution in OG micelles 139
6.2.2.2 Reaction products of AD in OG solutions 141
6.2.3 Conclusions 144

CHAPTER 1

Introduction

1.1 Protein structure

The vast majority of cellular processes occurring in living organisms are performed by biomolecular assemblies such as nucleic acids, proteins and polysaccharides. Proteins are the working entity of the cellular machinery and they are responsible for maintaining the homoeostasis in living organisms. To do so they control a variety of processes, such as metabolism, energy conversion, signalling and defence. Proteins synthesize new molecules and they are integral in their degradation. In the classic view, newly formed protein chains must fold into a defined biologically active structure (native) to be capable of exerting function and they must be able to maintain their three-dimensional structure in a biological environment. This paradigm is -to some extent- challenged by intrinsically disordered proteins, which show high levels of structural flexibility. Native protein conformation is determined by the specific amino acid sequence of each polymeric chain; the thermodynamic hypothesis states that the native structure represents the protein/solvent arrangement with the lowest possible free energy[1]. Misfolding may cause aggregation and misfunction leading to inflammatory processes that may cause several conditions[2,3]. The vast majority of protein tasks in living cells are mediated by proteins interacting specifically with other proteins or small molecules. Given the close relationship between protein complexes and their struc-

ture, elucidating the three-dimensional features of these assemblies has been object of extensive research during the past years, as comprehensive knowledge of protein interactions is vital to understand their physiological function and their role in diseases.

1.1.1 Levels of protein structure

Proteins are composed of small organic molecules called amino acids and their structural levels can be classified in four categories. The linear sequence of amino acids by which a protein is composed is referred to as *primary structure*. The denomination also includes post-translational modifications. Non-covalent interactions between amino acids result in the formation of structural motifs, generally stabilised by intramolecular hydrogen bonds. The most commonly occurring ones are the α-helix and the β-sheet. These structural motifs are collectively referred to as *secondary structure* elements. The *tertiary structure* is the three-dimensional arrangement of the amino acids in a polypeptide chain. Their relative position in space is driven by various types of non-covalent interactions, including salt bridges, hydrophobic effect and hydrogen bonding. The same forces can drive the assembly of more than one polypeptide chain. The spatial arrangement of the different chains is referred to as *quaternary structure*. Protein-protein interactions are mostly driven by the hydrophobic effect. This phenomenon involves the burial of hydrophobic side chains at the interface between the two interacting proteins, minimising their contact with the hydrophilic environment of the bulk solvent. Water molecules tend to aggregate around hydrophobic regions of proteins resulting in the formation of shells of ordered water molecules. The collapse of such structures introduces entropy to the system resulting in a negative Gibbs free energy value which implies that the process is thermodynamically favoured[4].

1.2 Methods for studying protein structure

Several methods have been used to investigate the structure of proteins and to map their interactions. Each method present advantages and disadvantages therefore the choice of one over another depends on the applications and/or availability. The most common methods are briefly described in the sections below.

1.2.1 X-ray crystallography

X-ray crystallography is the gold standard for the high-resolution determination of macromolecule three-dimensional structure. The method relies on the generation of an electron density map derived from the scattering pattern of X-ray beams caused by a protein crystal (Fig. 1.1). X-ray wavelengths are similar to the length of interatomic distances (approx. 1Å) therefore the technique is able to achieve atomic-resolved structures. Once obtained, the electron density map is used to build a molecular model which is then "refined" by modifying several parameters to achieve the best agreement between the observed reflection amplitudes (F_{obs}) and the reflections calculated from the atomic model[5] (F_{calc}).

The structure of myoglobin was the first protein to be solved by X-ray crystallography in 1958[6]. Since then, the pace of structure solving has vastly increased, mainly due to the technological advancements of the last 50 years. The Protein Data Bank is the current repository for protein structures and as of January 2017, of the 116,258 deposited, over 100,000 have been determined by X-ray crystallography. The major limitations of X-ray crystallography derive from the protein crystallisation process. Some proteins can prove difficult to crystallise because of the high conformational flexibility possessed by some regions of their sequence. This issue can be overcome by truncating such domains but the resulting protein fold may be perturbed and unfaithful to its native structure. Other proteins, such as transmembrane proteins, have their native

conformation stabilised by interactions with other biomolecules that cannot be easily segregated in a highly ordered crystal.

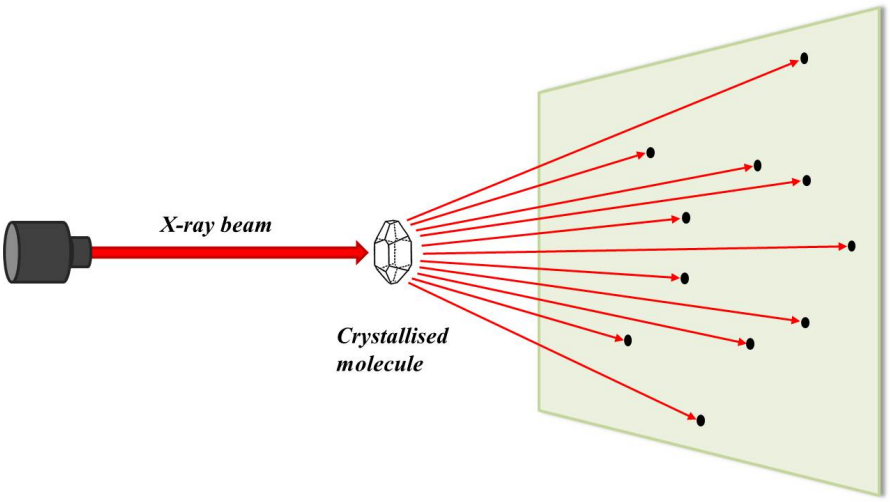


Figure 1.1: Schematic representation of the principle behind structure determination by X-ray crystallography. An X-ray beam is diffracted by a protein crystal and the scattering pattern is used to derive an electron density map which is then used to estimate the relative positions of the atoms in the crystal[7].

1.2.2 Nuclear Magnetic Resonance (NMR)

A significant number of high resolution structures deposited in the Protein Data Bank has been determined by NMR spectroscopy. This technique is based on the determination of distances between nuclei that carry magnetic dipoles such as ^{1}H, ^{13}C, ^{15}N, ^{19}F . The structure of the protein is modelled using such distance constrains in order to build a high resolution structure of the bio-macromolecule. This is based on multidimensional correlations such as COSY, NOESY and HMQC experiments.

NMR spectroscopy shares its main limitation with X-ray crystallography; as mentioned before, proteins with a high degree of flexibility may co-exist in several native conformation and the analysis of such systems can be challenging although ensemble structures can be prduced. Also high concentration of sample is needed to obtain good quality data and some protein native conformation can be affected in such conditions. The technique can also prove challenging for the structural determination of proteins larger than 40 kDa. Multidimensional magic-angle-spinning correlation NMR (MAS

NMR) allows for the determination of torsion angles and interatomic distances of membrane protein in the presence of lipid bilayers[8]. Such constraints have been used to obtain high resolution structures of a variety of membrane proteins including G protein-coupled receptors[9], proton and potassium channels[10–12] and bacterial outer membrane proteins[13–15].

1.3 Structural mass spectrometry

Since conformational shifts do not involve changes in mass-to-charge (m/z), the use of mass spectrometry for structural studies on proteins may seem counter-intuitive. Nonetheless there are a large number of MS-based techniques that allow for the structural investigation of biomacromolecules.

1.3.1 Ion mobility mass spectrometry

Ion mobility mass spectrometry (IMS) is a useful tool to measure protein collisional cross section in the gas phase and give insight into protein shape. Using carefully optimised conditions, ESI allows the retention of the native structure during the transfer of the protein from the solution to the gas phase. The effective survival of the native conformation in the gas phase is still debated but it is supported by strong evidence such as the preservation of specific non-covalent protein-protein and protein-ligand interactions[16–18]. However, it has also been shown that caution is needed when predicting structures of proteins with non globular folds based on collisional cross section (CCS) measurements[19]. In IMS classically ions are accelerated by a weak electric field through a inert gas. Given that the drift time depends on the ratio between the CCS of the analyte ion and its charge, it can be used to calculate its collisional cross section. Folded proteins present a smaller CCS compared to unfolded conformers[20]. The introduction of travelling wave-based ion guides (TWIGs) allowed IMS measurements on widely available Q-TOF mass spectrometers leading to a substantial increase in the

number of users. CCS measurements on TWIGs systems is not straightforward as it is on traditional drift tube machines but several calibration methods have been developed to overcome the issue[21–23]. Several applications have demonstrated the ability of IMS to study the conformation of protein and their complexes including the calcium-dependent conformational shift of calmodulin[24], the equilibrium between biologically relevant compact and extended conformation of cytochrome P450[25] and the large multiprotein GroEL complex[26].

IMS is also currently routinely used to gain insight into the quaternary structure of large macromolecular assemblies. CCS measurements of intact large multiprotein complexes provide only a limited amount of information because a single CCS value can be attributed to many different spatial arrangements of the various subunits of the assembly[27]. Nevertheless, the behavior of subcomplexes can be analysed in order to achieve more detailed information. The use of harsher conditions on labile multiprotein complexes results in the production of a series of subcomplexes composed by different subunits characterised by stronger interactions. These smaller complexes can be produced in solution using mildly denaturing conditions or in the gas phase using various types of ion activation. The analysis of the various subcomplexes provides insight into the connectivity of native complexes and in combination with IMS data has been used as restraints for the generation of accurate models of large multi-subunit protein complexes[28,29].

1.3.2 ESI charge state distribution

When proteins are subjected to positive ion mode ESI, multiply protonated ions are produced. It has been shown that the number of protons largely depends on the conformation of the protein in solution[30]. Tightly folded proteins generate spectra with low protonation states and narrow distributions whereas spectra of unfolded proteins show high charge states and much wider distributions. This different behaviour has been widely used to monitor conformational changes of proteins in solution[31–35]. The

rationale behind this phenomenon is related to the different mechanism of gas phase ion formation between compact folded proteins and unfolded proteins. Several studies suggest that compact proteins follow the charge residue model (CRM) of gas phase ion formation when subjected to ESI[36–38]. This topic is discussed in more detail in section 1.4.1.1. Briefly, according to the CRM model gas phase protein ions are formed via evaporation of small aqueous droplet to dryness. Such a model agrees with the observation that the number of charges found on compact protein ions is similar to the number of charges of a water droplet the size of a protein[38]. On the other hand computational studies suggest that ESI of unfolded proteins proceeds via the chain ejection model (CEM)[39]. According to such model, the unfolded protein protrudes from the water droplet adopting a conformation with the lowest electrostatic energy content. Such conformation favors the migration of protons resulting in a accumulation of charge on the emerging polypeptide tail. This phenomenon occurs on a time scale much shorter than droplet evaporation, thus explaining why unfolded proteins produce more intense MS signals and higher charge states[35].

1.3.3 Hydrogen-Deuterium exchange

One of the most widely used MS-based structural methodology and one of the first covalent modification technique to be developed is hydrogen/deuterium exchange (HDX).

When a protein is exposed to a solvent containing D_2O the backbone amide hydrogen atoms are replaced by deuterium atoms. Each exchange event causes a mass shift of $\sim$ 1 Da therefore such events can be tracked using mass spectrometry. O-H, N-H and S-H side chains hydrogen atoms are also exchanged with deuterium atoms but under typical quenching and digestion conditions and LC analysis the deuteration is lost due to back exchange with hydrogen atoms[40]. Areas of the protein that are fully exposed to the solvent exhibit a burst kinetic of exchange whereas areas of the protein that are buried or involved in hydrogen bonds show much slower exchange rates.

A standard HDX experiment starts with the incubation of the protein in D_2O buffered to the desired pH. The deuterium exchange-in step is quenched after a selected time by cooling the solution to 0 °C and lowering the pH to 2.5. The intact protein can be analysed as such in order to monitor the total deuterium uptake or digested before the analysis to achieve higher-resolution information. Proteolysis must be performed with a protease active in quenching conditions (acid-active). The most common protocols involve the use of pepsin immobilised on a column which allows the proteolytic step to be performed online with reverse phase LC separation. Digestion in high pressure conditions has been shown to enhance proteolysis efficiency[41,42]. MS analysis is then used to evaluate the extent of deuteration in the different segments (Fig.1.2).

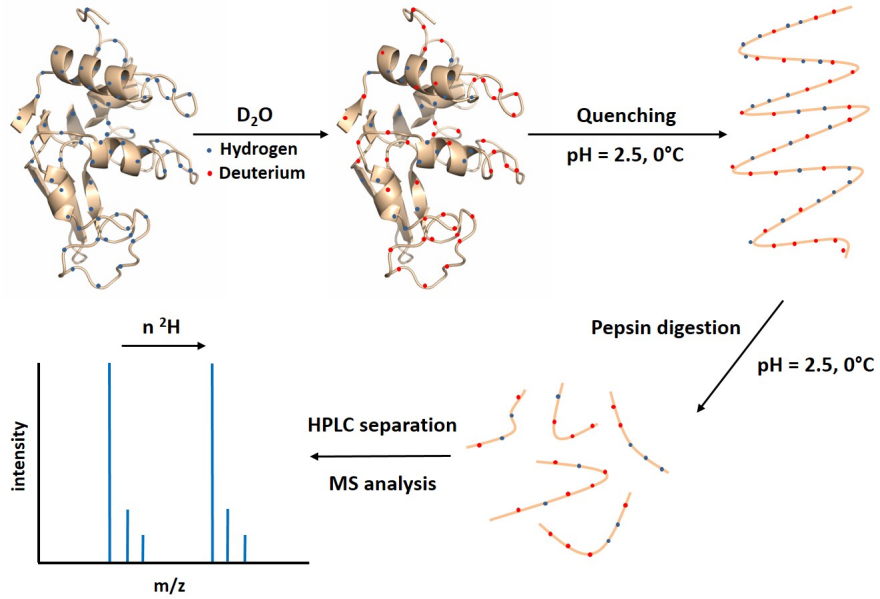


Figure 1.2: General H/D exchange workflow. The protein of interest is incubated with deuterated water leading to the exchange of some of the amide backbone hydrogen atoms with deuterium. The exchange is then quenched by lowering the pH and the sample temperature. The protein is then digested with an acid-resistant protease such as pepsin and analysed by reversed-phase chromatography. The HPLC eluent is connected to the inlet of a mass spectrometer where the peptides are analysed according to their mass-to-charge ratio.

Until recently the highest resolution achievable by HDX was obtained by observing the difference in deuteration extent of two or more randomly overlapping pepsin proteolytic peptides[43–46]. Collision induced dissociation (CID) fragmentation is routinely used to locate modification sites on peptides by analysing the characteristic product

ion spectra generated in ion trap mass spectrometers. Such experiment cannot be performed to locate sites of hydrogen/deuterium exchange since several studies proved that collisional activation causes H/D scrambling leading to major perturbation in the modification pattern[47]. This issue can be largely overcome using electron capture dissociation (ECD), an alternative gas-phase ion activation method. Using ECD H/D "scrambling" has been demonstrated to be less probable due to the non-ergodic nature of the electron-based fragmentation method[40,48].

Protein conformation, dynamics and interaction in solution are nowadays routinely investigated using HDX in combination with ESI-MS. HDX proved successful in locating binding sites and investigating changes in the tertiary structure[49–51]. Intrinsically disordered proteins (IDPs) and membrane proteins (IMPs) have always been object of extensive research, as their role in several pathological conditions but such studies remain challenging. Nonetheless HDX proved successful in the study of conformational switching events caused by ligand binding in IDPs[52,53] and the deuteration behaviour of IMPs in a native-like environment has been characterised which will hopefully lead to more accurate structural information of the latter[54,55]

1.3.4 Covalent labelling

Covalent labelling approaches to study protein structure share their basic principle with HDX since both techniques produce a structure-related mass shift in the protein. Strictly speaking the exchange between a hydrogen atom and a deuterium atom is a covalent modification but it is highly labile. The main difference of this approach compared to HDX methods is the stable nature of the covalently labelled product which makes any possibility of scrambling or back-exchange virtually non-existent[56]. To obtain such information, covalent labelling techniques use reagents that either target specific amino acids or several of them. The reagents used for covalent labelling target amino acid side chains making the structural information obtained complementary to the data obtainable by HDX methods[57] which, as mentioned before, target backbone

amide.

The covalent labelling approach relies on the different reactivity of amino acid side chains towards a certain reagent, implying that solvent exposed amino acids are more prone to react than buried amino acids, which react slowly or not at all[56]. Protein-ligand or protein-protein binding events affect the accessibility of some amino acids therefore their reactivity towards the label will be lower or greater; the detection of such change in reactivity implies the amino acids are in an area of the protein involved in ligand binding or in a region subjected to a conformational shift.

1.3.4.1 Amino acid-specific labelling

A wide range of residue-specific reagents have been used for the structural investigation of proteins. Usually the chemistry employed in this labelling involves well characterised organic reactions, and the corresponding mass shift induced by the addition of the reagent is located and quantified using LC-MS based methodologies. Among the 20 proteinogenic amino acids, only 8 have been used to investigate protein structure via amino acid specific labelling[56]. Examples include the reaction between arginine and vicinal dicarbonyl compounds such as phenylglyoxal. 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) specifically reacts with the carboxylic groups of aspartic acid and glutamic acid, and the acylation of lysine residues by acetic anhydride. A more inclusive list is shown in table 1.1. This approach has been used successfully to investigate the structure of protein-ligand[58,59] and protein-protein complexes[60,61], to analyse the surface structure of biomolecules[62,63], to obtain information about metal-binding sites[64] and to map disulfide bonds[65,66].

The major limitation of site-specific labelling is the small amount of information that can be drawn because of the limited number of sites targeted. This issue can be partially addressed by using more than one type of labelling reagent[67]. Reagents commonly used react with slow kinetics and can perturb the native structure and dynamics of the

protein under examination[68].

Amino acid	*Reagent*
Arginine	Phenylglyoxal, p-hydroxyphenylglyoxal, 2,3-butanedione
Carboxylic acids	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)
Cysteine	Iodoacetamide, N-Alkylmaleimides, IAEDANS, iodoethanol
Histidine	Diethylpirocarbonate (DEPC)
Lysine	Acetic anhydride, maleic anhydride, N-hydroxysuccinimide
Tryptophan	Koshland's reagent, N-bromosuccinimide
Tyrosine	Tetranitromethane, iodine, N-acetylimidazole

Table 1.1: Amino acid specific reagents used in protein footprinting

1.3.5 Hydroxyl radical footprinting

In the last 15 years hydroxyl radicals have been widely used for protein footprinting. Given its Van der Waals radius which is similar to that of water, such a reactive intermediate can provide accurate information about accessibility of solvent exposed residues. Upon generation, hydroxyl radicals readily react with protein side chains and the mass shift observed by MS can be used to determine the accessibility of modified sites[69–72].

1.3.5.1 Generation of hydroxyl radicals

Several methods of generating hydroxyl radicals for protein footprinting have been explored, each presenting advantages and disadvantages. Fenton's reagent is a common chemical method to generate hydroxyl radicals and it has been used to study protein structure[73,74](Fig.1.3 A). This reagent uses transition metals such as ferrous ion (II) to decompose hydrogen peroxide (H_2O_2) to hydroxyl radicals. The reagents needed for the experiment are low cost and easy to acquire however the addition of EDTA is

needed in order to increase the solubility in water of metal ions and the presence of the additive may lead to perturbation of the protein native fold and interfere with the radical reactivity[75]. Furthermore the long exposure time needed can allow undesired side reactions. Hydroxyl radicals can also be generated by radiolysis of water, without the use of any additives, using X-rays[76], electron pulse radiolysis[77] or γ rays[78], but the decomposition of water associated with the process causes a large variation in pH which may lead to structural perturbation if the solution is not strongly buffered. Furthermore, radiolysis by γ rays needs long exposure times therefore it is affected by the same problems as Fenton's reagent (i.e. long exposure times)[75]. Electron and X-ray radiolysis possess the advantage of generating radicals in microseconds, allowing the sample to be rapidly exposed to the labelling intermediate minimising the extent of secondary reactions. However a major drawback of such methodologies is the use of ionising radiation and the restricted access to synchrotron light.

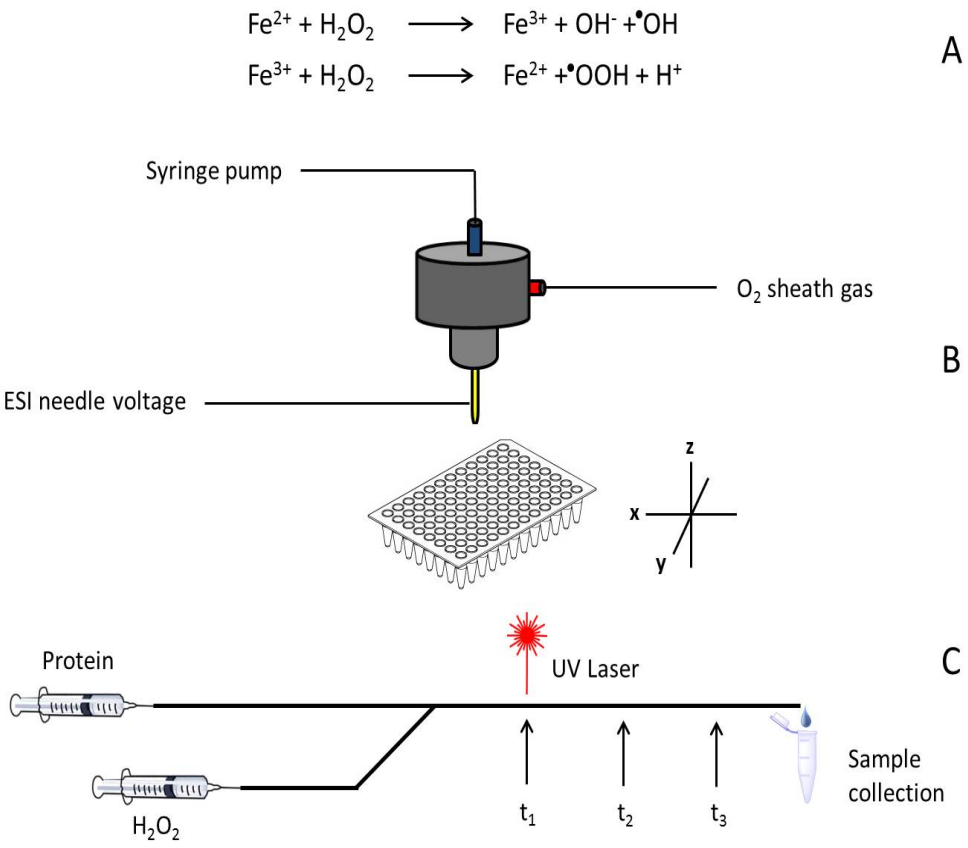


Figure 1.3: Methods developed to generate hydroxyl radicals in solution; **(A)** Fenton's reagent; **(B)** Electrical discharge within an electrospray source; **(C)** Fast Photochemical Oxidation Of Protein (FPOP)

Electrical discharge has also been used to produce hydroxyl radicals from water and molecular oxygen[79]. The device used consists of a standard ESI source operating with a needle voltage higher than that used during normal MS analysis (6-8 kV) and with oxygen gas replacing nitrogen as the nebulizer gas[72] (Fig. 1.3 B). Another popular approach for the generation of hydroxyl radicals is the photolysis of hydrogen peroxide (Fig. 1.3 C). With this method radicals are produced quickly by laser irradiation at 250 nm and the use of scavengers leads to short ·OH lifetimes[80], which carries the same advantages mentioned before for water radiolysis-based methods (i.e. short exposure times). Despite its proven effectiveness one of the drawbacks of this type of experiment is that hydrogen peroxide is added to the protein sample before irradiation, potentially

causing oxidation of readily oxidised residues prior to the photolysis. Such problem is usually minimised by reducing the pre-photolysis exposure time[75].

1.3.5.2 Hydroxyl radical reactivity

Hydroxyl radicals are reactive towards several amino acid side chains and the modifications obtained are covalent and therefore stable under normal proteolysis/analysis conditions. The relative reactivity of the amino acids side chains follows the order[81]: $Cys > Met > Trp > Tyr > Phe > His > Leu \sim Ile > Arg \sim Lys \sim Val > Ser \sim Thr \sim Pro > Gln \sim Glu > Asp \sim Asn > Ala > Gly$. Backbone C-$\alpha$ amide bond reactivity towards the labelling agent is much lower than that of side chains. The broader reactivity of the reactant allows for relatively high resolution footprinting and it is one of the main advantages of this approach compared to amino acid-specific labelling. Oxidation occurs through various mechanisms on the different functional groups located on the amino acid side chains and may lead to several products. Sulfonic acid, sulfinic acid and serine are the observed oxidation products of the highly reactive cysteine[82]. Methionine oxidation can generate a sulphoxide or a sulphone, respectively corresponding to the addition of one or two oxygen atoms. Saturated carbons are modified via hydrogen abstraction (mechanism shown in Fig. 1.4) whereas aromatic side chains can undergo hydroxyl addition[82]. Mass shifts of +16 Da or +14 Da have been reported for less reactive amino acids such as lysine, arginine, serine,threonine and leucine but they are rarely oxidised in protein when more reactive residues are accessible[72]

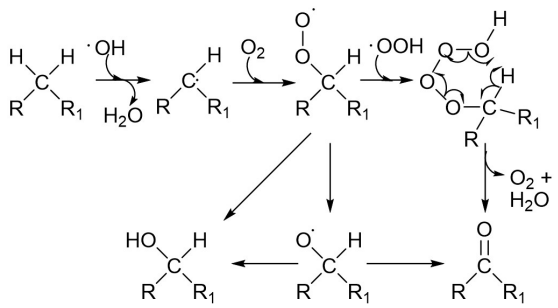


Figure 1.4: Reactions involved in oxidation of hydrocarbon side chains

1.3.5.3 Protein interactions studied by hydroxyl radical footprinting

As mentioned before, when proteins interact with other biomolecules, some areas of the protein become inaccessible to the solvent. Therefore, some of the residues that were solvent exposed in the unbound protein will be buried and their reactivity towards hydroxyl radicals will be significantly lowered. If the experiment is performed in a differential fashion (i.e. protein in solution with and without the binding partner) monitoring labelling changes in residues allows for the mapping of the binding interface. This approach has been used to study the interacting region of a variety of protein-peptide complexes including the calcium-dependant interaction between calmodulin and melittin[83]. This study showed how the extent of oxidation in melittin increased at pH 4, confirming the exposure of reactive residues, consequent to the ph-dependant dissociation of the calmodulin-melittin complex. Areas within the hydrophobic channel of calmodulin were found to show a decrease in the extent of oxidation at pH 7 confirming the role of such region in the binding interface.

The first use of hydroxyl radical footprinting to study a large protein-protein complex was published by Goldsmith et al. in 2000[79]. The calcium-dependant binding of actin to a subunit of gelsolin was investigated by the authors using synchrotron radiation to generate hydroxyl radicals. The oxidation rate of two peptide from gelsolin was monitored when the experiment was performed in the presence and in the absence of actin. The peptide expected to be involved in the interaction showed a large decrease in the

oxidation rate whereas the reaction rate of the peptide used as a control was unaffected confirming the effectiveness of the methodology.

1.3.5.4 Fast Photochemical Oxidation of Proteins (FPOP)

Recently Gross and co-workers developed FPOP (Fast Photochemical Oxidation of Proteins), an hydroxyl radical footprinting methodology that confines the exposure of protein to the labelling intermediate to a microsecond window[69]. Radicals are produced using a KrF excimer UV laser, which provides a high flux of light. Glutamine is included as a scavenger to limit the timescale of oxidation. The protein in the sample is irradiated only once through a fused silica reaction window using a flow system synchronised with the laser pulse frequency.

A very interesting feature of such methodology is its apparent ability to label protein faster than most conformational changes, which implies that FPOP is not affected by the structural perturbation that may occur upon slow oxidation[84]. Given its $\sim 1\mu s$ reaction time, the technique has been successfully applied to investigate the solvent-accessibility changes accompanying submillisecond protein folding[85].

FPOP has recently been used to follow the conformational changes and the aggregation of amyloid β_{1-42}, a protein involved in the insurgence of Alzheimer's disease[86]. Amino acid residue-level structural information was provided and some intermediate steps occurring during the oligomerisation process were detected. The results provided new insights on the self-assembly of Aβ and proved that the technique can be used to guide the design of drugs directed to inhibit amyloid aggregation.

FPOP has also proved successful in probing the structural differences in the interactions between membrane proteins and micelles or amphipols[71 87](see Fig. 1.5). It was shown that amphipols interact with membrane proteins through more extensive intermolecular contacts providing a rationale for the higher stability of membrane proteins in such an environment.

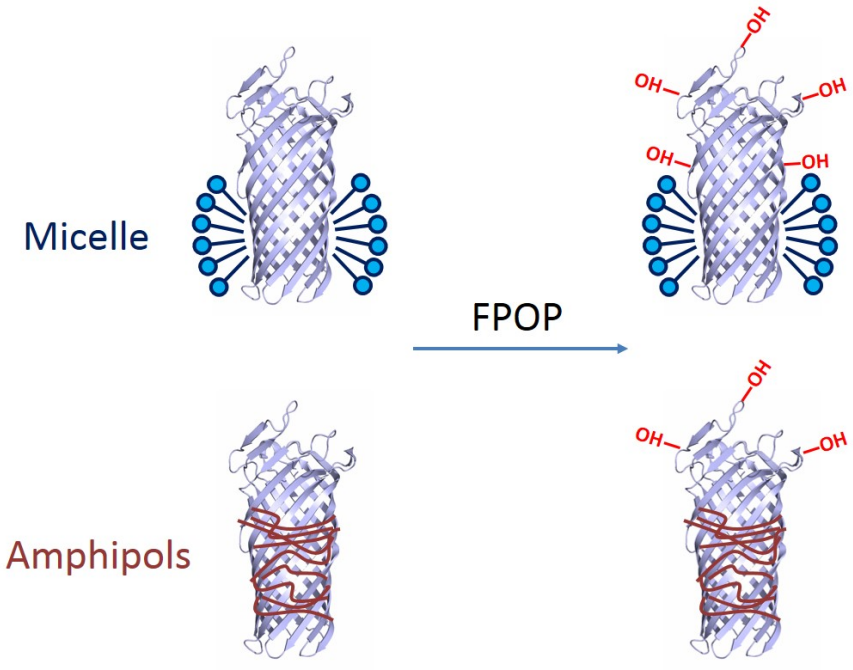


Figure 1.5: FPOP experiments on integral membrane protein solubilised in micelles and amphipols suggested that the latter contact proteins more extensively[71]. Adapted from PDB 1I78

FPOP has been also applied to ·OH labelling in live cells and the results show promise for the study of protein interactions in their native cellular environment and offer the potential to investigate proteins that are difficult to analyse in vitro[88].

Recently, it has been suggested the oxidation obtained through the use of FPOP may not be as fast as 1 μs, but it could extend over tens of milliseconds[89]. Vahidi and Konermann used the bleaching of reporter dye cyanine-5 to monitor the time frame of oxidation and the results suggest glutamine scavengers, routinely used in FPOP experiments to quench hydroxyl radicals and shorten the exposure time, generate metastable secondary radicals. These species cause oxidation long after hydroxyl radicals have disappeared slowing down the oxidation of the reporter dye. The slower nature of the technique may limit its usefulness in investigating sub-millisecond folding experiments and opens debate on the vulnerability of protein structure under oxidative conditions on such a time scale. Although very intriguing, it must be noted that the results were obtained using a small organic molecule and, even though cyanine-5 and similar

are often used to track the exposure of proteins to radicals in solution, the use of a dye may not reflect accurately protein oxidation kinetics in FPOP.

1.3.6 Carbenes

Carbenes are molecules containing a neutral carbon atom with two unshared valence electrons. They are highly reactive species and the large majority have lifetimes <10 ns[90]. The parent species is named *methylene* and it has general formula $:CH_2$. The two unshared valence electron of a carbene can exist in a singlet state, when the two are paired, or in a triplet state, when the two electrons are unpaired. The two states can be distinguished using their reactivity towards double bonds to form cyclopropane derivatives. The addition of a singlet carbene to *cis*-2-butene results in the *cis* isomer of cyclopropane since the two electrons react simultaneously or one very shortly after the other. Triplet carbenes instead generate a mixture of *cis*- and *trans*- 1,2 dimethylcyclopropanes when reacted with *cis*-2 butene (Fig. 1.6). This phenomenon is attributed to the formation of a covalent bond with one of the unpaired electron with one of the electron of the double bond. At this stage the other unpaired electron possesses the same spin as the remaining electron from the double bond and therefore cannot form a covalent bond very rapidly. For the spin to be inverted there is the need for some collision process and during this time the C-C bond can freely rotate leading to the generation of the mixture of geometric isomers.

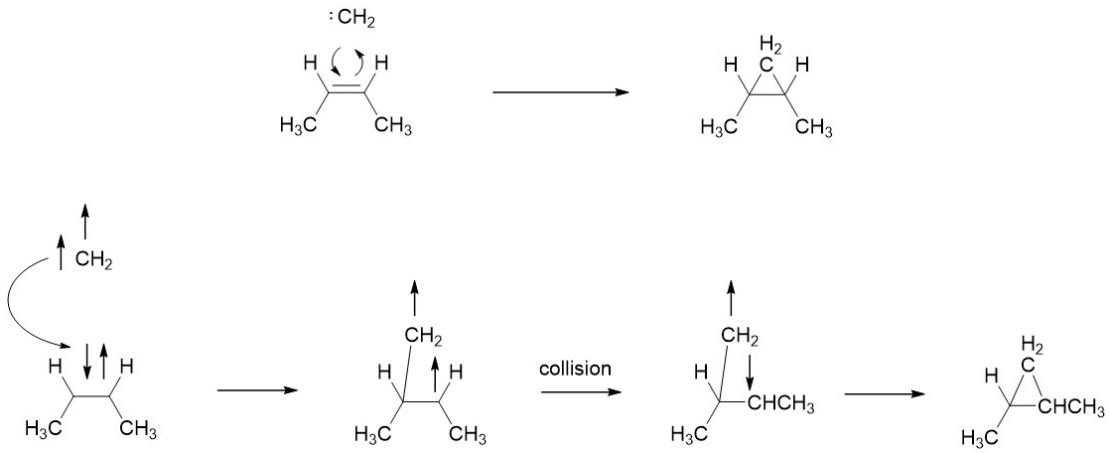


Figure 1.6: Addition of singlet (top) and triplet (bottom) carbene to 2-butene. The reaction can be used to distinguish between the two states of the carbene.

Depending on how they are generated and their molecular structure, carbenes can be in the triplet or singlet state. Usually the triplet state is the ground state but in many cases the difference of energy between the two state is relatively low (about 8 to 10 kcal mol^{-1}) therefore the conversion between the two states is common. An example of this phenomenon is given by methylene, which is generated as a singlet species and generally reacts as such before its decay to the ground state. There are several ways of producing carbenes among which the most common are α elimination of halo-alkanes and photolysis of double bonds.

1.3.7 Aryldiazirines as photoactivable reagents

The possibility of activating a reagent using light is a powerful tool for the investigation of biological systems. An ideal photoactivable reagent should be chemically stable before exposure to light, activate upon irradiation at wavelengths benign to biomolecules, and generate an extremely reactive intermediate[91,92]. Diazirines are a class of compounds that have been widely used for this purpose, mainly in photoaffinity labelling studies[93,94]. A reagent suited for such technique is a ligand specific to target the protein under examination which includes a photolabile moiety able to generate a covalent bond with the protein upon activation. The protein cross-linked with the ligand can be

subjected to proteolysis and mass spectrometric analysis to locate and identify its binding site. The photolabel can also include a purification tag which allows to perform the labelling experiment on whole proteomes without pre-purification of the protein of interest[95].

Diazirines possess a characteristic strained three membered ring containing two nitrogen atoms connected via a double bond that is responsible for their reactivity towards UV light. When a diazirine is photoactivated with long wavelength UV light, it generates a carbene that rapidly reacts with its surroundings, which in the case of a successful photoaffinity labelling experiment include the binding site of the protein under examination (Fig. 1.7).

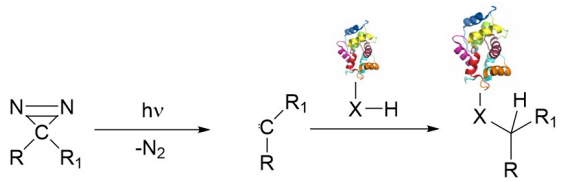


Figure 1.7: Upon near-UV irradiation, diazirines generate carbenes which react with the protein surface.

In addition to the desired highly reactive carbene, the linear diazo intermediate may also be generated[96] (Fig. 1.8). Such intermediates are considerably less reactive than carbenes and may diffuse away from its formation site and react as an electrophile or as a carbene (in the case of further irradiation) at a different location leading to potential errors in the data analysis[97].

Following studies reporting that aromatic diazirines yield a higher carbene-to-linear-diazo-isomer ratio upon photoactivation[97–99], Knowles and co-workers introduced, in the 1970s, an aryl-substituted diazirine[100–102]. In 1980, Brunner and co-workers modified the reagent incorporating a trifluoromethyl group on the diazirine carbon in order to lower the reactivity of the diazo isomer towards nucleophiles[97].

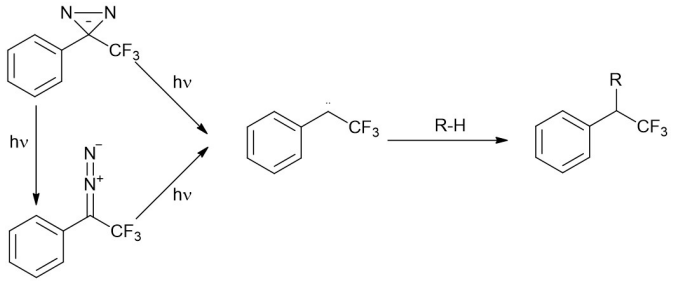


Figure 1.8: Products generated by photolysis of aryldiazirines.

As mentioned above, when a carbene is generated via photolysis of an aliphatic diazirine it can exist in two states, singlet and triplet. The singlet state is the excited state with an energy difference of 8 to 10 kcal/mol compared to the most stable state[103–106]. Carbenes generated from an aryldiazirine however show a reduced energy difference between ground and excited state (2 to 4 kcal/mol)[107] leading to a very rapid equilibration between the two states. Several studies have in fact shown that phenyl diazirines mostly react through the formation of the singlet carbene[108–110]. Such reactivity is advantageous for photolabelling since carbenes in the excited state insert more rapidly than triplet carbenes in a variety of bonds including O-H, N-H and C-H[111].



Figure 1.9: Orbital representation of two of the possible states in which carbenes exist. Singlet carbenes feature a pair of electrons in a p-orbital, triplet carbenes present two unpaired electrons.

1.3.8 Carbene footprinting

1.3.8.1 Diazirine gas

The use of a carbene-based footprinting strategy was first proposed by Richards and co-workers in 2000[112]. Methylene :(CH_2) seemed the ideal candidate for several rea-

sons; this carbene is easily synthesized in the singlet state, it is extremely reactive and unselective, it inserts in single bonds and adds to double bonds, and it is approximately the size of a water molecule (Fig. 1.10) implying that it may probe the same surface area accessible to it.

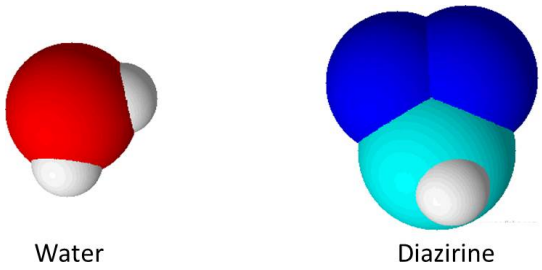


Figure 1.10: Size comparison between a water molecule and a diazirine molecule

Singlet methylene reacts with one of the atoms in its molecular cage in a picoseconds time window and, given its intersystem crossing rate being several orders of magnitude slower, it does not convert to the triplet state[112 113]. The nature of the labelling is covalent, thus there are no back reaction issues to be considered. In the early work of Richards and co-workers, methylene was generated by photolysis of diazirine gas dissolved in a variety of solvents. Diazirine gas was synthesized following the route indicated in fig. 1.11.

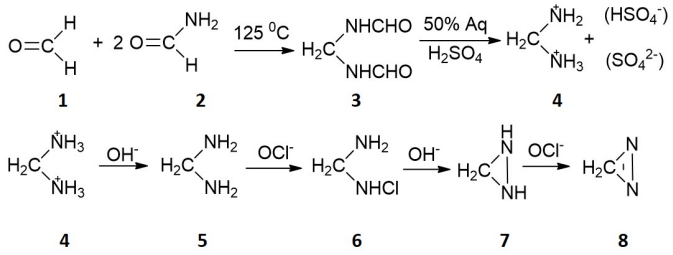


Figure 1.11: A 1:2 mixture of paraformaldehyde **(1)** and formamide **(2)** is heated to 125 ^{0}C for 12-18 hours and then allowed to rest at room temperature giving crystals of methylene bisformamide **(3)**. **3** is then dissolved in 50% aqueous sulfuric acid and allowed to stand at 4 ^{0}C to yield cystals of methylenediamine sulfate **(4)**. Salt **4** is reacted with aqueous sodium hypochlorite to form diazirine gas **(8)**. **5**, **6** and **7** are the presumed reaction intermediates

The authors characterised the photolysis products and explored the ability of the reagent to label a model protein.

Although promising, this approach proved to be limited by the low solubility of the gas in water, which lead to low labelling yields and subsequent difficulties in the analysis. A higher pressure of gas could help overcome this issue however literature references warn against explosions and flash burns when manipulating diazirine gas at high pressure.

1.3.8.2 Methylene footprinting coupled with scintillation detection

Following the work of Richards, Delfino and co-workers further explored the possibility of using diazirine gas to investigate the topography of the surface and inner space of proteins[114]. A similar photolysis apparatus was used to start the reaction but a tritiated version of the reagent ($^3H - DZN$) was employed to label the substrates. Incorporation of methylene in the model protein was followed by radioactivity measurement facilitating the detection of the very low abundance modified products resolving one of the issues that arose in the previous application.

Using UV-CD, the authors proved that methylene labelling does not perturb the secondary and tertiary structure of the protein, which is always a significant concern in this kind of studies. The ability of the technique to probe the accessible surface area was then tested on a model protein, α-lactoglobulin, in its native fold and in a partially denatured conformation. The overall extent of labelling was proved to be 25% higher in the partially unfolded conformer, in agreement with the larger surface area calculated for the denatured protein. The authors also performed a per-peptide quantitation of the modified species that confirmed the extremely low specificity of the methylene carbene and improved the resolution of the technique.

The same research group successfully applied the technique in the structural elucidation of the binding surface of a protein-protein complex[115]. The interaction between hen egg white lysozyme and the monoclonal antibody IgG_1 was mapped obtaining

results in good agreement with previous cystallographic literature. The study also confirmed the non-specific nature of the label and the ability of diazirine gas to probe protein cavities in addition to the already established possibility of probing protein surfaces.

Tritiated diazirine gas labelling in combination with scintillation detection also proved useful to follow conformational transitions in proteins[116]. The technique made possible the experimental measurement of the accessible surface area of the native state of BL-βL as well as partially folded conformers and enlightened subtle aspects of complex transitions occurring between folding states.

1.3.8.3 Methylene footprinting coupled with mass spectrometry detection

Mass spectrometry detection of methylene modified peptides was introduced in 2011 by Delfino and co-workers and it represented an important step forward for the methylene footprinting methodology[117]. The tritiated version of the reagent, despite its proven success, possesses some features that may impede the general use of the technology. Radiotracers are known to be difficult to handle and expensive; furthermore a large amount of protein is needed for the analysis as a consequence of the low specificity radioactivity of the labelled reagent. The use of high-resolution and extremely sensitive mass spectrometry easily overcomes these issues. The authors validated the novel detection approach using a selection of model proteins and also introduced a new photolysis device for the chemical modification of protein with diazirine. The new device used a continuous flow of diazirine gas through the protein solution and photolysis of the reagent allowing for the control over the extent of modification of proteins. Diazirine gas is generated in a syringe mixing methylene diammonium sulfate with an alkaline solution of sodium hypochlorite and bubbled into the sample contained in a quartz cuvette (described in fig.1.12).

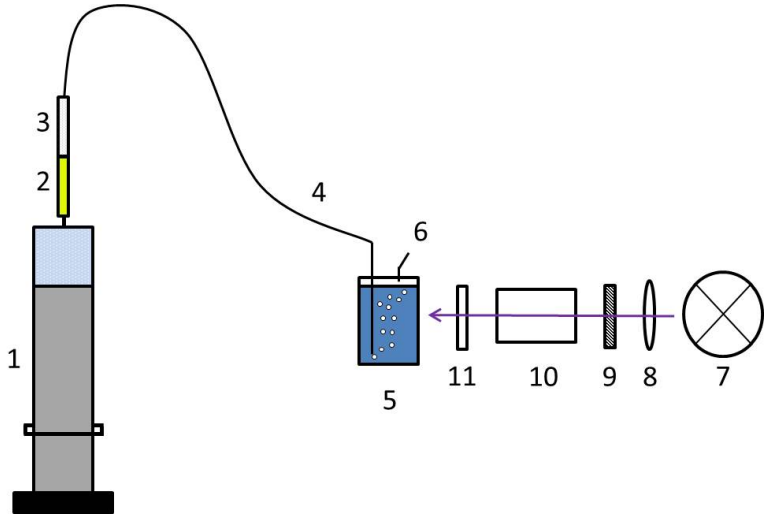


Figure 1.12: Photolysis apparatus for the labelling of protein with diazirine gas[117]. **(1)** Glass syringe containing diazirine gas; **(2)** online sodium bisulfite and **(3)** citric acid filter; **(4)** diazirine line into the protein solution; **(5)** sample contained into a quartz cuvette; **(6)** gas outflow line; **(7)** UV light source; **(8)** condenser lens; **(9)** automated shutter; **(10)** water IR filter; **(11)** cut-off filter λ<300 nm

.

The photolysis reaction is started using a Hg/Xe arc source equipped with a cut-off filter to avoid protein damage. The reaction is carried in a steady-state regime by matching the photonic flux to the syringe input rate.

The mass spectrometric approach resulted suitable to determine the extent of modification of the model proteins and confirmed the usefulness of the diazirine gas-based technique in differentiating between different conformers, as proved in previous studies using the tritiated version of the reagent and scintillation detection.

The use of tandem mass spectrometry allowed carbene modification to the be located to the single amino acid level, thereby increasing the resolution of the methodology significantly.

In 2015 Delfino and co-workers further proved the usefulness of their footprinting approach combined with mass spectrometry[118]. Using the same set-up described above, the large conformational shift occurring when calmodulin binds to Ca^{2+} was successfully investigated, thus confirming the utility of the methodology in dissecting aspects of a physiological structural change. Furthermore the technique was employed to shed

light on the contact interface of a protein-peptide system; the use of MS/MS fragmentation achieved the goal of obtaining high-resolution information about the solvent accessibility of the protein in both its unbound and bound state.

1.3.8.4 Water-soluble diazirine footprinting

Despite its proven efficacy and reliability, the use of diazirine gas for protein footprinting has generated limited interest since its first appearance in 2000[112,114–116]. The main issues arising from the use of such reagent is its low solubility in water and therefore the low yield obtained for modified folded proteins, and also the difficulties in handling the gas[119]. The introduction of mass spectrometry detection improved the overall sensitivity of the methodology but the need of a complex apparatus to synthesise the reagent on-site and its explosive nature makes diazirine gas unappealing and prevented its widespread as footprinting reagent.

Photo-Leucine, a water soluble amino acid functionalised with a diazirine ring, was proposed in 2005 by Suchanek et al. for crosslinking experiments[120]. The reagent was developed to be incorporated in the primary structure of proteins when cells are grown in an environment where leucine is absent (Fig.1.13).

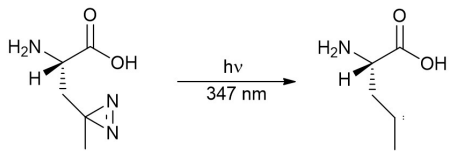


Figure 1.13: Photoleucine activation by UV light photolysis

The use of such water-soluble diazirine-modified leucine overcomes the need of synthesising the reagent on-site and does not need particular storage conditions.

Photolysis of photoleucine using UV light yields a singlet carbene which, as any other carbene intermediate, immediately reacts with its molecular cage with a reaction rate close to the diffusion limit. As mentioned before, carbenes are potentially reactive to-

wards every amino acid in a protein but the reaction most likely to occur in water is the formation of the alcohol[120]. The first use of photoleucine as a potential footprinting reagent was proposed by Schriemer et al. in 2011[121]. The labelling procedure was tested on two model proteins, myoglobin and calmodulin, and the reaction yield was assessed by mass spectrometry. In order to describe the behaviour of the reagent and compare the extent of modification obtained to previous diazirine gas-based labelling the authors proposed a simple kinetic model. A labelling event occurs when a photolytic event is coincident with a collision between the protein and the reagent; this process is described by the unimolecular reaction

$$\mathrm{A} \xrightarrow{h\nu} \mathrm{B} \tag{1.3.1}$$

where A is photoleucine and B is the carbene produced by photolysis and reacted with the protein in its molecular cage. Given that the rate of photolysis is the same for the reagent in the bulk solvent and in proximity of the protein surface, diffusion from the bulk to the molecular cage during irradiation is negligible. Given these conditions the reaction is well described by pseudo-first-order kinetics:

$$\ln\left(\frac{[B_t]}{[B_{max}]}\right) = -kt \tag{1.3.2}$$

where $[B_t]$ is the concentration of carbene that has reacted with the surface of the protein at the time t and $[B_{max}]$ is the maximum concentration of protein-quenched carbene obtainable. The maximum reaction yield is described by the following equation[114]:

$$max.avg.yield = C\left(\frac{SA_{protein}[protein]}{SA_{water}[water]}\right)\left(\frac{[diazirine]}{[protein]}\right) \tag{1.3.3}$$

where $SA_{protein}$ and SA_{water} are respectively the surface area of the protein and water. The C factor represents the deviations from the model where C=1 provides the best agreement for a labelling experiment on a folded protein using diazirine gas[116]. Using equation 1.3.3 the extent of modification calculated on calmodulin and myoglobin resulted respectively 390 times and 60 times higher than the estimated yield of methylene labelling (Fig.1.14).

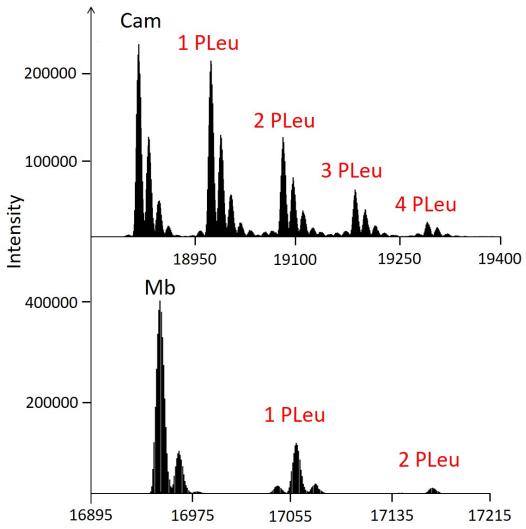


Figure 1.14: Deconvoluted mass spectra of photoleucine labelled calmodulin (top) and myoglobin (bottom). xPleu indicates the number of label inserted per protein. Data represents photolysis at 349 nm with a Nd:YLF laser for 16 seconds of 100 mM photoleucine

The increase in extent of modification can be explained analysing equation 1.3.3; part of it can be linked to the possibility of performing the labelling reaction in the presence of high concentration of diazirine, which is not possible when using diazirine gas given its poor solubility. The aforementioned C factor for the two proteins largely deviates from the ideal value of 1 (3.2 for calmodulin and 0.6 for myoglobin) suggesting differenent behaviour of the soluble diazirine compared to the mostly non-specific reactivity of diazirine gas. The deviations from the C factor can be attributed to the higher or lower affinity towards certain surface residues that photoleucine appears to have. In the case of photoleucine the diazirine moiety is installed on a larger structure therefore it has the potential to be distributed unevenly over the protein surface based on the intrinsic affinity of the probe towards different side chains. Being a modified amino

acid, photoleucine is zwitterionic at neutral pH therefore its potential partitioning over different areas of the protein is influenced by variation in pH and ionic strength[121]. The nature of the labelling is therefore less unspecific than the one obtained using diazirine gas but it has the advantage to be potentially targeted towards certain areas of the protein by introducing different chemical functionalities on the reagent resulting in a fine tuning of its surface activity.

Subtle modifications of the structure of photoleucine were used by Schriemer and co-workers to confirm the concept of "surface activity"[68]. Given their short lifetime it is safe to assume that diffusion of the carbene intermediate is negligible and therefore the label distribution over the protein surface reflects the distribution of the reagent before activation. The use of a de-aminated version of photoleucine (4,4-azipentanoic acid) confirmed that the modification distribution is driven by the affinity of the reagent prior to activation. Both photoleucine and 4,4-azipentanoic acid have been tested on bovine brain calmodulin and the use of ETD fragmentation has resulted successfully in pinpointing the modification on single residues[68](Fig.1.15).

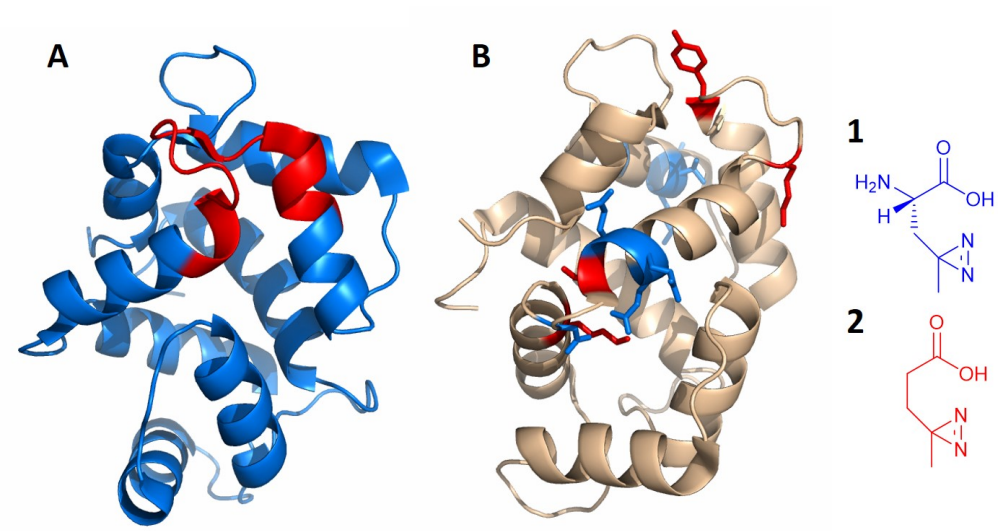


Figure 1.15: Crystal structure of bovine brain calmodulin in compact form (pdb:1PRW). Areas preferentially labelled with reagent **1** and **2** are respectively coloured in blue and red. Areas in wheat were not modified.**(A)** Modification distribution observed at peptide level resolution. **(B)** Modification distribution observed at residue level resolution.

As mentioned above, photoleucine labelling deviates from the ideal model in which no affinity towards specific residues is assumed and in fact it shows a pronounced bias towards acidic residues. Such deviation is likely to be introduced by the pre-concentration at the protein surface driven by electrostatic interaction between the reagent amino group and the carboxylic moieties on glutamate and aspartate residues. This hypothesis is confirmed by the lack of bias towards such residues of 4,4-azipentanoic acid, a reagent identical to photoleucine but lacking the amino group that is potentially involved in the electrostatic interactions with acidic residues. The pre-concentration hypothesis is also confirmed by labelling experiments at different ionic strengths[68]; the overall yield of the modification is generally suppressed when performing the photolysis in solutions with high ionic strength but the presence of a few areas with increased labelling indicate that surface partitioning governed by electrostatics is not the only interaction that drives the process.

Recently photoleucine-based protein footprinting has been successfully implemented in a FPOP-like flow system[122]. Like hydroxyl radical footprinting performed in such way, carbene labelling in a flow system presents the advantage of using many laser shots but with limited exposure per shot. The reaction is confined in a small spot allowing to use low energy laser pulses to obtain good modification yields.

1.4 Mass Spectrometry

Mass spectrometry (MS) is a powerful analytical tool used in a vast number of scientific applications. A mass spectrometer is an instrument able to generate gas-phase ions and analyse them according to their mass-to-charge ratio (m/z). The first mass spectrometer was developed by J.J. Thomson in 1912 and it was used to obtain mass spectra of several gases. Using this technique Thomson also discovered metastable ions and isotopes 20 and 22 of neon[123]. In 1918 A.J. Dempster developed the electron ionisation source and the first magnetic sector mass spectrometer with direction focusing. A significant technological improvement was provided by the work of F.W. Aston which in 1922 allowed the measurement of isotopes of non-radioactive elements[123,124]. Since the early days of mass spectrometry, the technique has become invaluable for many scientific disciplines, from elemental analysis to the investigation of complex biomolecular assemblies.

All mass spectrometers share a common basic layout; the three essential components are the ion source, the mass analyser and the detector (Fig. 1.16). Depending on the design, the ion source may be held under vacuum or at atmospheric pressure[125] whereas the mass analyser and the detector must be held under vacuum. Usually the instrument is controlled by a computer which is also used to process data and produce mass spectra.

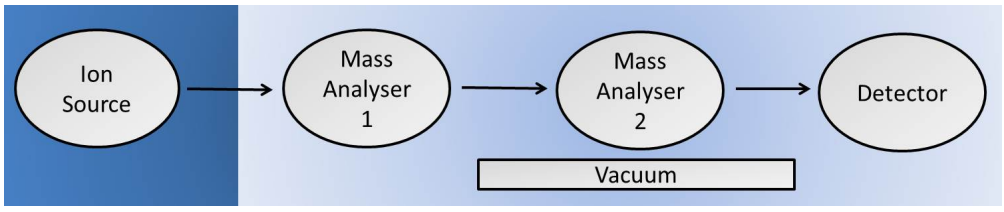


Figure 1.16: Basic diagram of a typical mass spectrometer equipped with two mass analysers between the ion source and the detector. Depending on the ionisation technique, the ion source may be placed in a vacuum region.

1.4.1 Ionisation techniques

As mentioned above, mass spectrometers are instruments able to detect and analyse gas-phase ions. Therefore, many ionisation techniques able to produce such species from samples of different nature have been developed. Electron impact ionisation (EI), chemical ionisation (CI), matrix-assisted laser desorption ionisation (MALDI), electrospray ionisation (ESI) are some of the most popular techniques but many others have been developed in order to tackle different analytical issues. Given its extensive use in this work, ESI will be the only technique to be described further in this thesis.

1.4.1.1 Electrospray ionisation

The success of electrospray ionisation (ESI) begun when Fenn et al. showed that proteins generate multiply charged gas-phase ions, which allows their detection using mass spectrometers with a limited *m/z* range[126]. An important feature of ESI is its ability to ionise and desolvate a large range of analytes without imparting an excess of internal energy. The gentle nature of this ionisation process allows the detection of quasimolecular ions and, with carefully optimised conditions, the preservation of non-covalent interactions occurring in solution[127,128]. An electrospray is generated applying a strong electric field, under atmospheric pressure, to a liquid flowing through a capillary tube. The application of a difference of potential, generally 3-6 kV, between the capillary and the counter-electrode, generates a field in the order of $10^6\,\mathrm{V\,m^{-1}}$ which induces an accumulation of charge at the liquid surface. This unusual excess of charge at the end of the capillary causes elongation of the drop at the end of the capillary; when the repulsion between charges becomes greater than the surface tension of the liquid, the shape of the drop changes into a Taylor cone and results in ejection of a fine spray of charged liquid[129] (see Fig. 1.17).

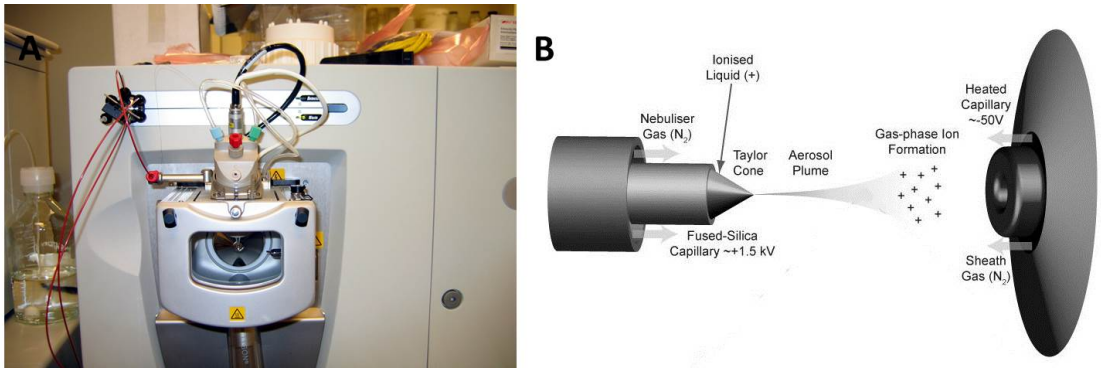


Figure 1.17: (A) Thermo LTQ ESI source. **(B)** Schematic representation of the electrospray ionisation process

The Rayleigh equation describes the relationship between charge and surface tension in the spray droplets:

$$q^2 = 8\pi^2\epsilon_0\gamma D^3 \tag{1.4.1}$$

where q is the charge, ϵ_0 is the permittivity of the environment, γ is the surface tension and D is the diameter of a supposed spherical droplet. As the droplets travel, pushed by the electric field, the solvent evaporates, a process assisted by a flow of heated inert gas in some source designs. The charge of the droplet remains constant, which increases the charge density and therefore Coulombic repulsion within the droplet. When the Rayleigh limit is exceeded, the repulsion between the charges overcomes the surface tension producing a new Taylor cone from which several smaller droplets are released[130]. Coulombic fission was widely believed to be the mechanism behind the formation of an electrospray but in 1994 Gomez and Tang produced photographic data of heptane droplets dividing in partial contrast with the Coulombic fission hypothesis (shown in Fig.1.18)[131]. Heptane droplets generated from an ESI source resulted in mechanically deformed, a phenomenon that actually reduces Coulombic interaction (Fig. 1.18). Small droplets were shown to be ejected from the "tail" of the deformed parent droplet carrying approximately 15% of the charge and 2% of the mass. After multiple

cycles of droplet shrinking and smaller droplets ejection, an individual ion is released into the gas-phase.



Figure 1.18: Mechanically deformed heptane droplet generated from an electrospray source [131]

Nevertheless the general mechanism behind gas-phase ion formation is still a subject of debate and currently there are two prevailing models: the charged-residue model and the ion evaporation model [36](Fig. 1.19). The charged-residue model (CRM) was proposed by Dole and co-workers in 1968 and according to his theory, the droplets experience a series of several cycles of evaporation and Coulombic fission until only a single molecule of the analyte is contained in the droplet and the charges are transferred directly to it [37]. The ion evaporation model (IEM) was proposed by Iribarne and Thompson in 1976 and, in contrast with CRM, suggests that charged analyte ions can evaporate directly from the surface of the droplet when it reaches a certain radius larger than the one suggested by Rayleigh equation [132]. Both models are supported by experimental evidence [38,132–135]. These studies suggest that when the analyte is relatively small the IEM model is favoured whereas the ionisation of large molecules, such as proteins, proceeds via CRM model.

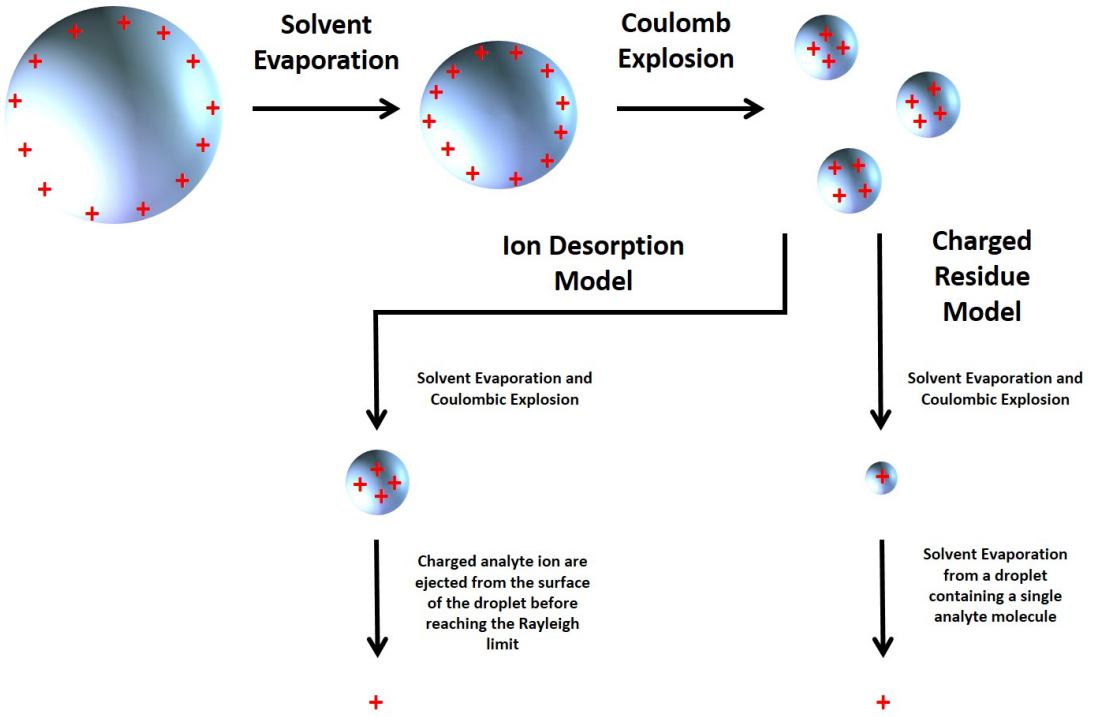


Figure 1.19: Cartoon representation of IEM and CRM mechanisms of gas-phase ion formation from a highly charged droplet

A third mechanism has been recently proposed for the electrospray ionisation of unfolded proteins. The chain ejection model (CEM)[39] predicts that electrostatic repulsion and hydrophobicity lead to the ejection of the extended unfolded protein from the droplet surface followed by charge equilibration between the droplet and the protruding polypeptide[35].

Charge migration is responsible for the high charge states observed in unfolded proteins. This process occurs much faster than complete solvent evaporation, which is the basis of the chain ejection model. A simplistic scheme of the model is shown in Fig.1.20

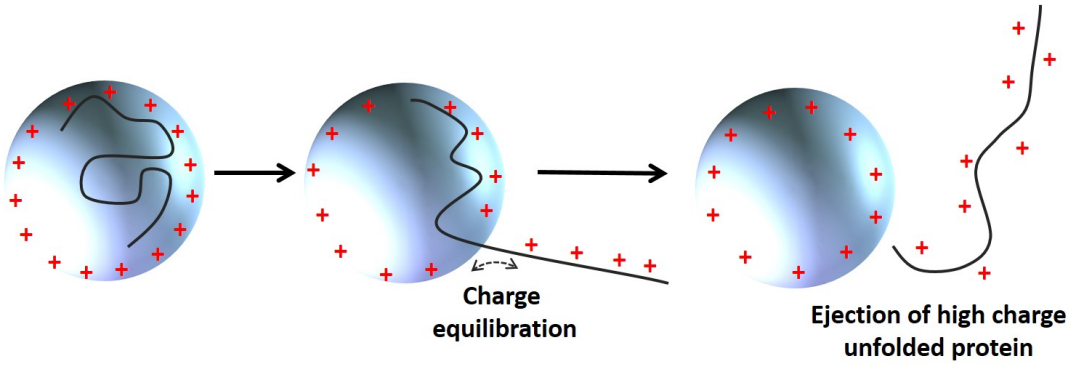


Figure 1.20: Cartoon representation of CEM mechanism proposed for ESI of unfolded proteins. Protein is depicted in black.

1.4.2 Mass Analysers

Once the analytes have been ionised and desolvated, they must be separated by an analyser according to their *m/z*. All mass analysers use electric and/or magnetic fields to achieve analytes separation. The differences between different types of mass analysers lie in the way such fields are exploited to obtain ion separation. Scanning mass analysers include magnetic sector instruments and quadrupole instruments; such analysers start with a mixture of ions having different *m/z* and, using an electromagnetic field, they allow transmission of different ions according to their mass-to-charge ratio along a time scale[136]. Trapped-ion mass analysers include ion traps, FT-ICRs and Orbitraps and they all operate by storing ions and manipulating them using DC and RF electric fields. For the purpose of this thesis only 2D ion trap and FT-ICR mass analysers will be discussed in detail given their extensive use in this work.

1.4.2.1 Ion trap analysers

The ion trap uses an RF quadrupolar field to trap ions in two or three dimension (2D or 3D ion trap). The first ion trap to be developed was the 3D ion trap in 1960 and it is composed of a circular electrode with two ellipsoid caps on the top and the bottom. This analyser can be imagined as a quadrupole bent over to form a loop. In quadrupole

analysers the potentials on the rods are set to allow the transmission of ions with a certain mass-to-charge ratio; in contrast in ion traps ions of various masses are stored in the same space and ejected according to their m/z [125].

In the 2D ion trap ions are confined in the radial dimension by a quadrupolar field and in the axial dimension by a static electric field at the ends of the trap. The application of an RF-only potential on the rods allows the ions to oscillate along the xy plane while they travel along the z axis. An additional DC voltage is applied to the end of the quadrupole in order to trap the ions. The ion cloud is pushed towards the centre of the quadrupole by such voltages and cooled by collision with an inert gas. Focusing along a line rather than along a central point allows linear ion traps to have a higher ion trapping capacity and the ability of storing a larger number of ions before space charging effects occur. These features result in an increased sensitivity and a wider dynamic range of the linear ion trap compared to the 3D ion trap.Trapped ions ejection can occur along the axis of the trap (axial ejection) or perpendicular to its axis (radial ejection).

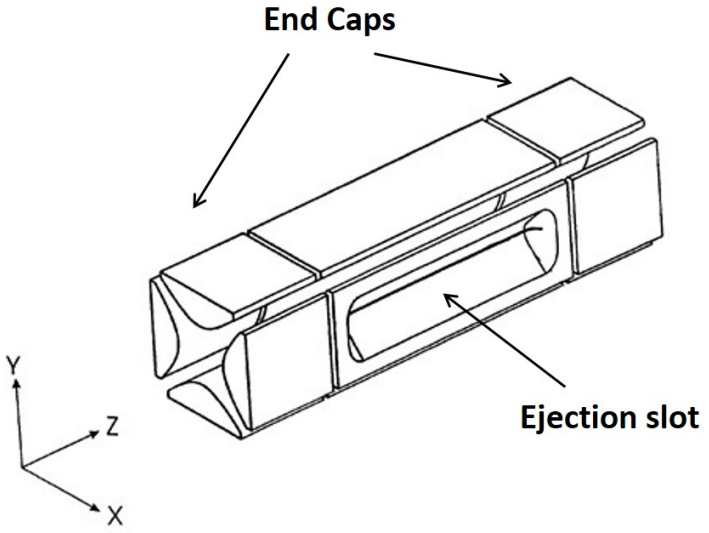


Figure 1.21: Schematic representation of the basic design of a two-dimensional linear ion trap

1.4.2.2 FT-ICR analysers

The kinetic energy of an ion accelerated by a potential difference is described by the following equation:

$$E_k = \frac{mv^2}{2} = zV_s \tag{1.4.2}$$

Where m is the mass of the ion, z is its charge, v its velocity and V is the applied difference of potential. If the ion is exposed to a magnetic field, B, perpendicular to its velocity it experiences a force of which the magnitude is given by equation 1.4.3

$$F_M = zvB \tag{1.4.3}$$

The trajectory of the ion is curved by the magnetic field into a circle with radius r. The centrifugal force of the ion is equilibrated by the magnetic force and the phenomenon is described by equation 1.4.4

$$zvB = \frac{mv^2}{r} \tag{1.4.4}$$

Combining equation 1.4.2 and 1.4.4, the following equation is obtained:

$$\frac{m}{z} = \frac{r^2B^2}{2V_s} \tag{1.4.5}$$

If the magnetic field is intense enough and the velocity of the ion is low, the radius of the trajectory is reduced and the ions can be trapped within the magnetic field. The ions still travel along a circular trajectory described by equation 1.4.4 with a circular frequency

$$\nu = \frac{v}{2\pi r} \tag{1.4.6}$$

Therefore the angular velocity of the ion is

$$\omega_c = 2\pi\nu = \frac{v}{r} = \frac{z}{m}B \tag{1.4.7}$$

It is evident from equation 1.4.7 that for an ion with a given z/m ratio at a fixed B value, the radius of its trajectory increases with its velocity and the ion is expelled from the cell if the radius becomes larger than that of the cell. The m/z of the ion can be determined by measuring its cyclotron frequency. One of the methods by which such frequency can be determined is by using complex waves and Fourier transforms[125] (Fig. 1.22).

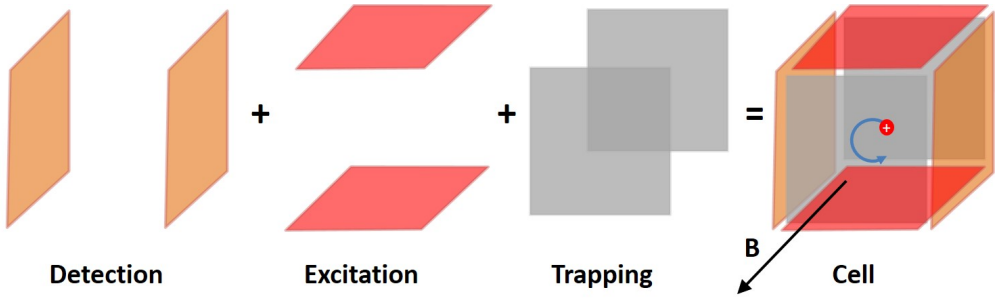


Figure 1.22: Schematic representation of the basic design of a FT-ICR cell. Modified from http://www.massspecpro.com/mass-analyzers/fourier-transform-ion-cyclotron-resonance

If ions trapped in cyclotron motion absorb RF electromagnetic radiation that has the same frequency as ω_c, they experience an increase in their kinetic energy, which causes

the radius of their trajectory to increase. Ions travelling close to the cell detector plates induce a perpendicular image current that can be detected. All the ions of a certain *mass* and excited by the same energy follow the same trajectory with the same frequency, therefore they are forced to travel as tight packets, since the presence of an ion close to one detecting plate generates an image current opposed to another ion close to the opposite plate and that results in a total null induced current. In Fourier transform mass spectrometry all the ions in the cyclotron are excited simultaneously by a broad frequency pulse that puts all the ions in phase and on a larger orbit closer to the detector plates. It has been demonstrated that ions excited by an AC irradiation with the same energy and the same V_0 potential applied during the same time move along the same orbit regardless of their *m/z* (eq.1.4.8).

$$r = \frac{V_0 T_{exc}}{B_0} \tag{1.4.8}$$

Even if excited ions of different *m/z* travel along the same orbit, they exhibit their own characteristic cyclotron frequency. As mentioned before, the current induced by the different ions can be measured and the complex wave detected can be transformed from a free ion decay (FID) time-dependent function into a frequency-dependent function using the Fourier Transform (FT). The frequencies observed are then related to an particular *m/z* values thus generating a mass spectrum[125] (Fig. 1.23).

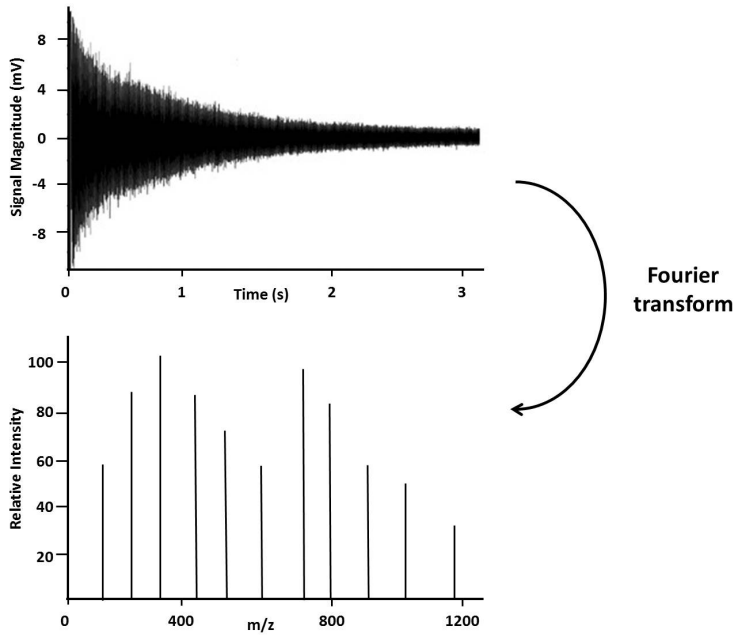


Figure 1.23: Signal magnitude as a function of time is transformed into relative intensity as a function of frequency and therefore into a intensity to m/z spectrum using the Fourier transform

The resolution of Fourier transform mass spectrometers is linked to the disappearance of the signal caused by ion-molecule collisions occurring in the cell. To overcome this limitation, the cell must be held at high vacuum (10^{-10} torr)

1.4.3 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) is a technique which involves the isolation and subsequent activation of a precursor ion leading to its dissociation into product ions and neutral fragments. Such experiments can be classified in two categories: MS/MS in space and MS/MS in time (Fig. 1.24). In tandem mass spectrometry in space the experiment is performed in two physically separated mass analysers whereas in MS/MS in time is carried out in trapping devices where ions can be stored and fragmented at different times in the same space. A common example of the former category is the triple quadrupole mass spectrometer, in which the second quadrupole is operated in RF mode only and it is used as the collision chamber. Hybrid machines such as the QTOF, which combines a quadrupole with a time-of-flight analyser, are also often used

to carry out these kind of experiments. Analysers commonly used to perform MS/MS in time are ion traps, orbitrap and FTICR. These analysers can be programmed to perform different steps of the experiments such as ion isolation and activation at different times. A major difference between ion traps and FTICR analysers is that analytes are expelled from the trap in order to be detected when using a trap whereas FTICR analysers are non-destructive, and ions undergoing subsequent reactions can be monitored continuously[125].

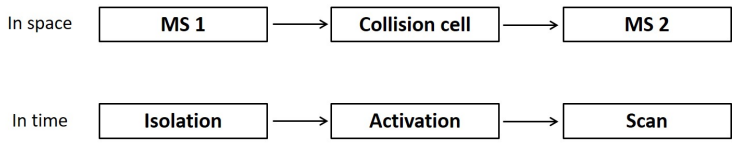


Figure 1.24: Comparison of a product ion scan performed by a space-based and a time-based instrument

1.4.3.1 Collision Induced Dissociation

The most common fragmentation method used in tandem mass spectrometry experiments is collision induced dissociation (CID), in which precursor ions are induced to collide with molecules of inert gas leading to the conversion of a fraction of their kinetic energy into internal energy. This excess energy results in the unimolecular decomposition of the excited ion,which results in fragmentation by bond cleavage. According to the Rice-Ramsperger-Kassel-Marcus (RRKM) theory, the rate of dissociation of an excited ion is much slower than its rate of excitation, therefore the energy deposited is redistributed with equal probability between all the vibrational modes of the analyte ion. These conditions lead to the preferential cleavage of the weakest bond of the ion and, for the same reason, ions with a larger number of atoms need a higher energy deposition in order to achieve fragmentation[125].

Proteins or peptides subjected to CID commonly generate fragment ions of the y- and b-series and to a lesser extent a- and i-series ions (fig. 1.25). The fragments obtained depend on the size of the peptide, the amino acid composition, the charge state of the ion and the collision energy used. Fragment ion series can be used to identify

the protein by searching against sequence databases or by manual *de novo* sequencing where the genome of the product is not sequenced.

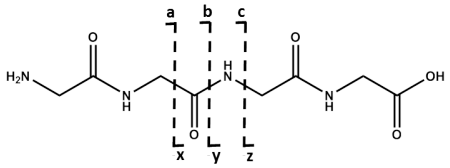


Figure 1.25: Main peptide fragmentation pathways in tandem MS experiments

The "mobile proton" model, introduced by Wysocki and Gaskell explains the observed peptide fragmentation by CID[137]. Singly charged peptides mainly dissociate by charge directed reactions[138–142]. Larger peptides can be protonated at several sites including terminal amino groups, amide nitrogen, amide oxygen and side chain groups. In some cases protonation is energetically favoured at one or more functional groups such as the guanidinium group of an arginine residue which can be found at the C-terminus of tryptic peptides. In such peptides the mobilisation of the proton to less favoured sites requires a large amount of energy. There is also a second class of peptides with several protonation sites accessible in a similar energy range[143]. Molecular orbital calculation showed that protonation on the amide oxygen leads to strengthening of the amide bond but is thermodynamically favoured. On the other hand, protonation of the amide nitrogen results in a substantial weakening of the peptide bond but it is not thermoynamically favoured[140,141]. The "mobile proton" model resolves this issue stating that protons that are not sequestered by a basic amino acid side chain migrate to different sites upon excitation and the fragmentation observed results from the competition between different fragmentation pathways[144]. The majority of the sequence ions are generated following the $b_x - y_z$ peptide fragmentation pathway shown in figure 1.26[143].

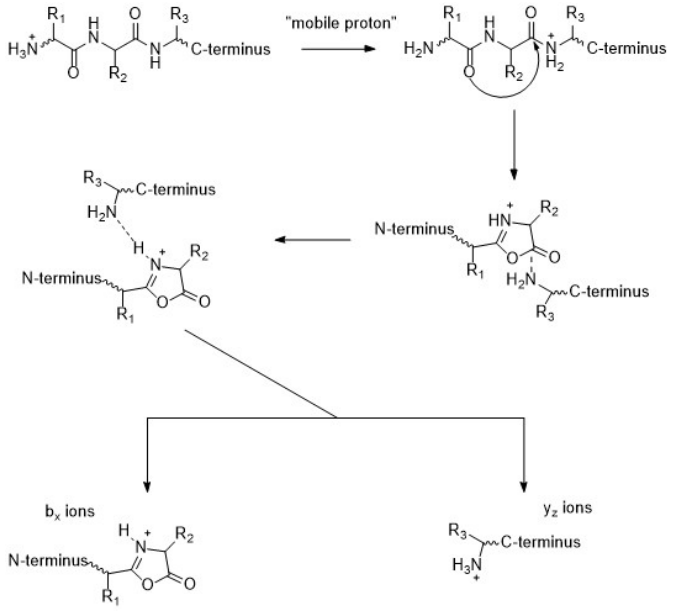


Figure 1.26: Scheme of the $b_x - y_z$ peptide fragmentation pathway

Such fragmentation pathway is started by the mobilisation of one of the non-sequestered protons to the nitrogen atom of a peptide bond. The oxygen of the N-terminal neighbouring residue attacks as a nucleophile the carbonyl group of the protonated amide bond resulting in the formation of a protonated oxazolone species and a leaving C-terminal fragment[145–148]. Under low-energy collision conditions the complex between the oxazolone species and the leaving C-terminal fragment can rearrange resulting in the formation of a proton-bound dimer of the two fragments. The proton affinity of the two species directs the formation of the b- or y- ion after the dimer collapses[149].

CHAPTER 2

Materials and Methods

2.1 Materials

2.1.1 Buffers and reagents

All solutions were prepared using Ultrapure water (18.2 MΩ · cm) obtained from a Millipore water purification system. HPLC grade acetonitrile (VWR Chemicals) and analytical grade formic acid (Fisher Scientific) were used in all the mobile phases and ZipTip buffers. Photoleucine was purchased from Thermo Fisher Scientific and Sodium 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoate (AD) was designed and synthesised by Dr. Andrew Barrow (University of Nottingham). Trypsin and AspN were obtained from Promega (Southhampton, UK). Dithiotreithol (DTT), Iodoacetamide, were purchased from Sigma-Aldrich (Poole,UK).

2.1.2 Characterisation of AD reactivity

Hen egg white lysozyme, calmodulin, equine cytochrome C, horse heart myoglobin, melittin, ubiquitin from bovine erythrocytes and octyl-glucoside were purchased from Sigma-Aldrich (Poole,UK).

2.1.3 Activity assays

Lys48 poly-ubiquitin ladder, wild-type USP5 and Lys48 diubiquitin were purchased from Boston Biochem (Cambridge, MA, USA). Lyophilised cells of *Micrococcus lysodeikticus* ATCC were obtained from Sigma Aldrich (Poole,UK).

2.1.4 Protein footprinting

Hen egg white lysozyme and penta-N-acetylchitopentaose (NAG5) were purchased from Sigma-Aldrich (Poole, UK). C335A-USP5 was expressed and purified by Dr. Daniel Scott (University of Nottingham). DH, DH-ACP and ^{DH}DP were expressed and purified by Dr. Matt Jenner (University of Warwick). OmpF was provided by Prof. Colin Kleanthous.

2.2 Instruments

2.2.1 Thermo Scientific LTQ FT-ULTRA

All the analysis of intact proteins from AD efficiency studies and all the samples derived from a bottom-up workflow (protein ID and footprinting) were performed on a LTQ FT-Ultra mass spectrometer. The machine is a hybrid Linear Ion Trap-Fourier Transform Ion Cyclotron Resonance mass spectrometer and offers the possibility to perform CID, IRMPD and ECD fragmentation experiments. A schematic of the machine is shown in figure 2.1. A series of advanced ion guides are positioned between the ion source and the 2D ion trap and between the latter and the FT cell to provide high ion transmission efficiency. Precursor ion isolation and fragmentation can be performed in the linear ion trap and combined with high resolution MS/MS data acquisition in the FT ICR. For all the experiments detailed in this thesis the mass spectrometer was equipped with a nanospray ion source. Fused-silica PicoTip emitters (New Objective,

Woburn, MA) were used to provide continuous-flow nanospray. Xcalibur 2.0 software was used to control the instrument and to process the data.

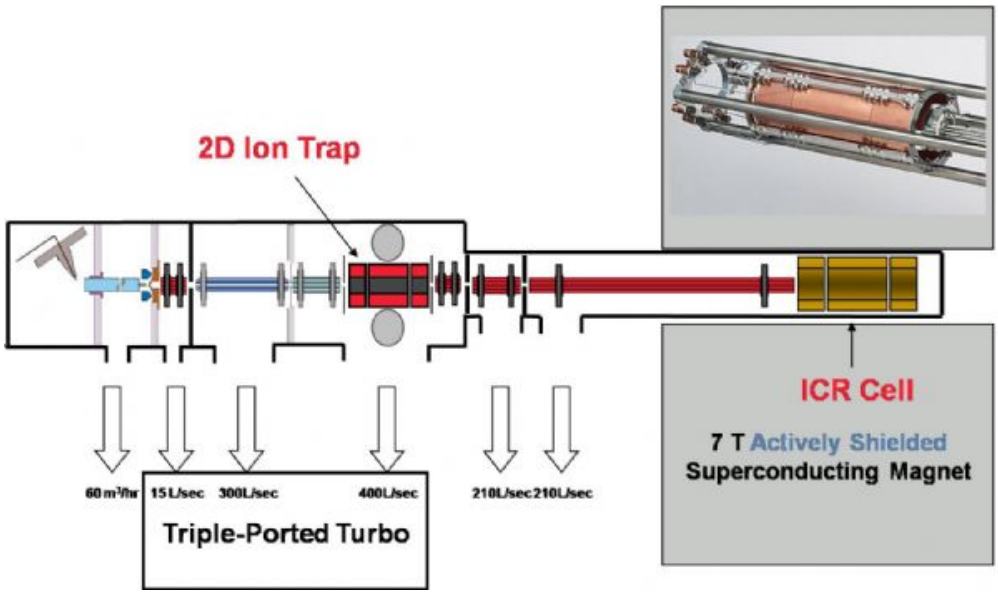


Figure 2.1: Representation of the layout of a Thermo LTQ FT Ultra mass spectrometer.

2.2.2 Waters SYNAPT HDMS

Mass spectrometric analysis performed on intact ubiquitin and photolysis reaction products for the characterisation of AD reactivity were performed on a Waters SYNAPT HDMS mass spectrometer equipped with a nanospray ion source. The instrument is a hybrid quadrupole-ion mobility-orthogonal acceleration TOF mass spectrometer. A schematic of the machine is shown in figure 2.2. The quadrupole allows mass selection before entering the TWIMS cell for separation according to gas-phase ion mobility and a transfer ion guide is used to convey the separated ions in the TOF region for mass analysis. MS/MS experiments can be performed by accelerating the ions in the trap region causing them to collide with an inert buffer gas leading to collisional induced dissociation. Masslynx 4.1 software was used to operate the mass spectrometer and for data processing.

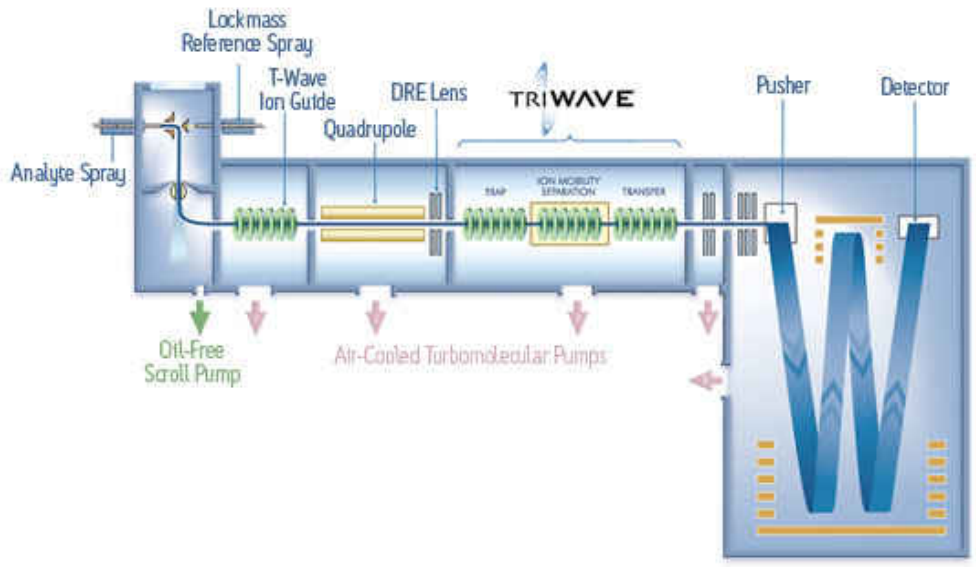


Figure 2.2: Representation of the layout of a Waters Synapt mass spectrometer. Adapted from https://genome.duke.edu/cores-and-services/proteomics-and-metabolomics/synapt-g1-high-definition-mass-spectrometer

2.3 Methods

2.3.1 Photochemical labelling

2.3.1.1 Lysozyme

An aqueous solution of lysozyme (160 μM) in 25 mM ammonium acetate was mixed with an equal volume of a 20 mM solution of the diazirine labelling reagent in 20 mM Tris /150 mM NaCl, to give a final label concentration of 10 mM. The mixture was left equilibrating for 5 minutes at room temperature. Aliquots (6 μL) of this solution were placed in X100 crystal clear sample vials (Thermo Fisher Scientific). In order to avoid enzymatic cleavage of the pentasaccharide substrate, NAG5 was added to the mixture, to a final concentration of 80 μM, just before snap-freezing the samples in liquid nitrogen (77 K). Freezing was found to be important for efficient and reproducible labeling. For lysozyme footprinting in the absence of substrate, the NAG5 solution was substituted by an equal volume of buffer. The labelling reaction was initiated by irradiation

(16 s) of the mixture using a Spectra Physics Explorer 349 laser (an actively Q-switched Nd:YLF laser operating at 349 nm, with a repetition frequency of 1000 Hz, and a pulse energy of 125 μJ, Newport, Didcot, UK, Fig. 2.3). The laser beam was directed into the open top of the sample vial using a small 45°mirror. All samples were generated in triplicates.

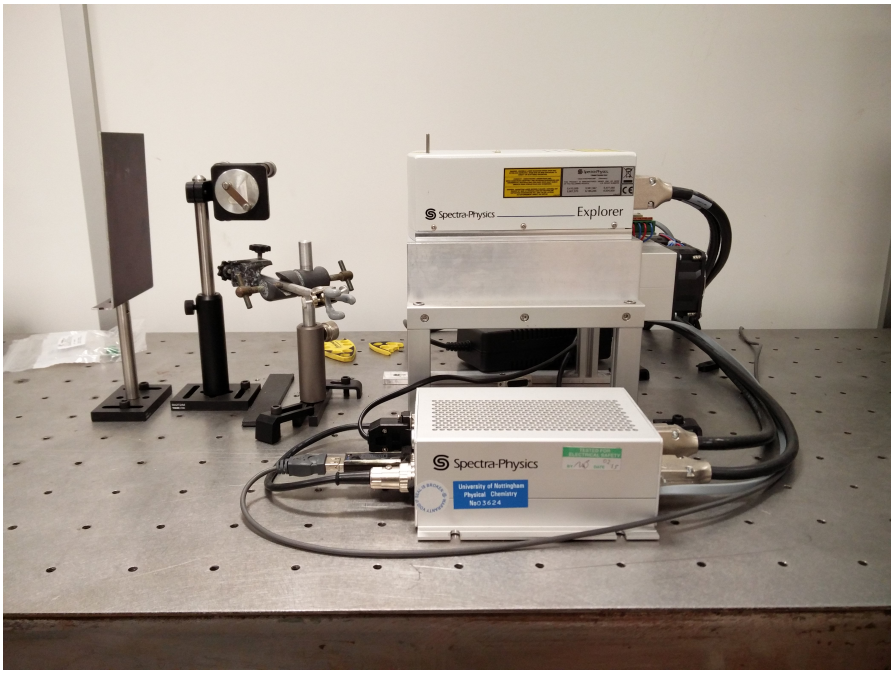


Figure 2.3: Spectra Physics Explorer 349 laser set up used for the photolysis experiments described in this thesis.

2.3.1.2 C335 USP5

6 μL aliquots of a solution of C335A USP5, Lys48 di-ubiquitin (30 μM each, 10 mM ammonium acetate) and AD (in 20mM Tris /150mM NaCl, to a final concentration of 10 mM) in tapered X100 crystal clear vials were snap-frozen and irradiated at 349 nm as described for lysozyme. Di-ubiquitin was substituted by an equal volume of buffer in footprinting experiments on unbound USP5. All samples were generated in triplicates.

2.3.1.3 DH-ACP

A 60 μM solution of DH, ACP or DH-ACP in 20 mM Tris, 150 mM NaCl was mixed with an equal volume of a 10 mM solution of AD in the same buffer to a final concentration of respectively 30 μM and 5 mM. The same protocol described for lysozyme was then used to start the labelling reaction. All samples were generated in triplicates.

2.3.1.4 DH-ACP/^{DH}DP

A 40 μM solution of DH-ACP in 20 mM Tris, 150 mM NaCl was mixed with an equal volume of a ^{DH}DP solution of the same concentration (or buffer for unbound experiments). A 100 mM solution of photoreagent in the same buffer was added to the mixture to a final concentration of 5 mM. The same protocol described for lysozyme was then used to start the labelling reaction. All samples were generated in triplicates.

2.3.1.5 OmpF

An aqueous solution of OmpF (32 μM) in 1% octyl-glucoside and 50 mM ammonium acetate was mixed with an equal volume of a 100 mM solution of AD or 200 mM solution of photoleucine (both in 20 mM Tris/150 mM NaCl, pH 7.4) to a final concentration respectively of 50 mM and 100 mM. The photolysis reaction was conducted in the same conditions described for the labelling of lysozyme. All samples were generated in triplicates.

2.3.1.6 Characterisation of AD reactivity

6 μL aliquots of a 10 mM solution of AD in 20 mM Tris, 150 mM NaCl in absence of any proteinaceous material were irradiated as described for lysozyme. For the study of the reaction products of AD in presence of OG, the a 20 mM solution of the reagent was

mixed with an equal volume of 2% octyl glucoside and irradiated following the same procedure.

2.3.1.7 Labelling efficiency

Solutions of Calmodulin, Myoglobin, Cytochrome C, Ubiquitin or Melittin (70 μM in 20mM Tris /150mM NaCl) and photoleucine (100 mM in 20 mM Tris /150mM NaCl), or aryldiazrine (10 mM in 20mM Tris /150 NaCl) were snap-frozen and irradiated at 349 nm for 4 seconds in the case of aryldiazirine (except calmodulin, which required only 1 second), and 16 seconds in the case of photoleucine. Samples were diluted to 5 μM with mobile phase A (H2O/acetonitrile 95/5 + 0.1% formic acid) and analyzed directly by LC-MS.

2.3.1.8 Amino acid selectivity

A 125 μM solution of amino acid standard mix (Sigma Aldrich) in water was mixed with the same volume of a 20 mM solution of AD to a final concentration of 62.5 μM and 10 mM respectively. For the control experiment AD solution was substituted with 20 mM Tris, 150 mM NaCl buffer. The mixture was irradiated following the same protocol described for lysozyme.

2.3.2 Sample preparation

2.3.2.1 Bottom-up analysis

2.3.2.1.1 SDS-PAGE To produce the resolving gel, 6 ml of the resolving gel stocks (compositions described table 2.1) were mixed with 30 μL of 10% APS solution and 6 μL of TEMED before pouring the mixture between the glass plates to catalyse polymerisation. A water/isopropanol 50/50 solution was used to cover the surface of the gel. Once polymerised, isopropanol was removed. To produce the stacking gel, 2 ml of

stacking gel stock were mixed with 20 μL of 10% APS solution and 2 μL of TEMED and the mixture was poured on top of the resolving gel. Loading buffer was added to protein samples, typically 1.6 μL to 5 μL sample, which were heated for 5 min. A 180 V potential was applied to the system for 90 minutes to allow protein migration. PageRuler Unstained Protein Ladder (Thermo Scientific) was used as a standard to approximate the size of proteins within the samples. Gels were incubated for 1 h in Coomassie Blue stain and destained using a water/methanol/acetic acid 50/40/10 solution.

Resolving Gel Compositions (ml)				
	7 %	**10 %**	**12 %**	**15 %**
H2O	51	41	34	24
1.5 M Tris-HCl, pH 8.8	25	25	25	25
20 % (w/v) SDS	0.5	0.5	0.5	0.5
Protogel	23	33	40	50

Stacking Gel Composition (ml)	
H2O	61.5
0.5 M Tris-HCl, pH 6.8	25
20 % (w/v) SDS	0.5
Protogel	13.4

Table 2.1: Composition of resolving and stacking gel solutions employed for SDS-PAGE separation

2.3.2.1.2 Protein digestion Following separation by SDS-PAGE (12-15 % acrylamide), protein bands were excised, reduced (DTT, 10 mM in 100 mM ammonium bicarbonate), alkylated (iodoacetamide, 55 mM in 10 mM ammonium bicarbonate) and digested at 37 °C with either trypsin (1:15 protease/protein ratio in 10 mM ammonium bicarbonate) or AspN (1:70 protease/protein ratio in 5 mM Tris-HCl) for 16 hrs[150]. Formic acid (0.5 μl) was added to the supernatant (50 μl) to inactivate the proteases. The resulting solutions were analysed by LC/MS without further dilution.

2.3.2.2 Intact proteins

2.3.2.2.1 ZipTip clean up C_4 ZipTips were used to purify intact labelled ubiquitin from byproducts of the photolysis reaction. The samples obtained were used in the characterisation of AD reactivity. The ZipTip was washed with two 10 μl aspirations of 50% MeCN, followed by five 10 μl aspirations of H_2O/0.1% TFA solution. The protein sample was then loaded onto the ZipTip column by 15-20x 10 μl aspirations of the sample, each time injecting back into the sample vial. The loaded protein sample was then separated from salts and byproducts of the photolysis reaction by 15x 10 μl aspirations of 0.1% TFA, followed by elution of the sample into 8 μl H_2O/80% MeCN/0.1% TFA

2.3.2.3 Photolysis products

Following laser irradiation, samples of AD photoactivated in presence and absence of OG were diluted 50 fold with water/acetonitrile 95/5 and the mixture was separated on a Agilent 1100 series HPLC equipped with a Waters Sunfire column (3.5 μm, 2.1x30 mm). The mobile phases A and B consisted of 95/5 water/acetonitrile (v/v) and 5/95 water/acetonitrile (v/v), respectively, and both contained 0.1% formic acid. Analytes were separated using a linear gradient starting from 20% mobile phase B and reaching 90% B in 5 minutes followed by column equilibration. Products eluting from the column were monitored using UV absorbance at 254 nm and identified on a Waters Synapt mass spectrometer equipped with an ESI source operated in negative ion mode. A 3.0 kV voltage was applied to the capillary, the sampling cone and extraction cone offsets were respectively 15.0 and 4.0 V. A 200 L/hr flow of nitrogen gas held at 250 °C was employed to help the desolvation of ions.

2.3.2.4 Amino acid selectivity

Following the labelling procedure, the mixture was diluted 20 fold with acetonitrile. The amino acids were separated on a Agilent 1100 series HPLC equipped with a Phe-

nomenex Kinetex HILIC column (2.6 μm,50 X 2.1 mm). The mobile phases A and B consisted of 95/5 water/acetonitrile and 5/95 water/acetonitrile respectively and both contained 0.1 % formic acid. Products were eluted using the gradient described in table 2.2. Products were detected on a LTQ FT Ultra mass spectrometer equipped with a ESI source operated in positive ion mode. A 1.5 kV voltage was applied to the outlet capillary to initiate the electrospray and a 15 V potential was applied to the ion transfer capillary which was held at 200 °C. The tube lens value was set at 100 V.

Time (min)	% phase A	Flow Rate (ml/min)
0	5	0.5
5	15	0.5
8	25	0.5
8.01	35	0.5
10	35	0.5
10.01	5	0.5
21	5	0.5

Table 2.2: Gradient profile employed to elute reaction products of amino acids with photoactivated AD.

2.3.3 Protein digests LC-MS analysis

Peptides generated from protein digestion were separated using a Dionex U3000 nanoHPLC and detected on a Thermo LTQ FT Ultra mass spectrometer.

2.3.3.1 nanoHPLC

2.3.3.1.1 Lysozyme and USP5 Samples were loaded onto a C4 Pepmap300 trap column and separated using a in-house packed WP C4 column (0.075x150 mm, 5 μm particle size, 300 pore size) equipped with a 30 μm i.d. Picotip emitter (New Objective). The mobile phases A and B consisted of 95/5 water/acetonitrile (v/v) and 5/95 water/acetonitrile (v/v), respectively, and both contained 0.1% formic acid. The samples

(1 μl) were injected in load-trapping mode. Peptides were eluted using a 20 min linear gradient of mobile phase B from 0% to 70% followed by 5 minutes at 90% B and column re-equilibration.

2.3.3.1.2 DH-ACP and OmpF Samples were loaded onto a C18 Pepmap300 trap column and separated using a capillary Thermo Scientific C18 Pepmap300 column (0.075x150 mm, 5 μm particle size, 300 pore size) equipped with a 30 μm Picotip emitter (New Objective). The mobile phases A and B consisted of 95/5 water/acetonitrile (v/v) and 5/95 water/acetonitrile (v/v), respectively, and both contained 0.1% formic acid. The samples (1 μl) were injected in load-trapping mode. Peptides were eluted using a 30 min linear gradient of mobile phase B from 0% to 55% followed by 5 minutes at 90% B and column re-equilibration.

2.3.3.2 Mass spectrometry

The LTQ FT Ultra mass spectrometer was equipped with a nESI source through which a 1.7 kV voltage was applied to the Picotip emitter. The inlet capillary of the mass spectrometer was held at 275 °C with a tube lens value of 145 V. For the identification of peptides the mass spectrometer was operated in data-dependent acquisition mode on the 5 most intense ions per each survey scan. Ions were isolated in the linear ion trap within a window of 2 Th and subjected to collision induced dissociation. Charge-state rejection of +1 ions was employed. A dynamic exclusion list was employed to prevent multiple isolation of the same ion; the repeat count was set at 3 with a window duration of 45 seconds and an exclusion duration of 360 seconds. The data were submitted to Bioworks software (Thermo Fisher Scientific) and searched against a custom database including the sequence of the protein of interest. In order to quantify the fraction modified of each peptide the samples were separated and analysed using the same conditions but with the mass spectrometer operating in full scan mode instead of DDA. Residue-level location and quantitation of the labelling was achieved perform-

ing CID experiments at a nominal energy of 15.0 on singly-modified peptides isolated in the linear trap within an 8 Th window. The activation time was set at 30 ms with an activation-Q value of 0.250.

2.3.4 AD efficiency

Intact labelled ubiquitin samples used for the characterisation of AD efficiency were analysed on a Waters Synapt equipped with a static nanospray source after a ZipTip desalting step. The nESI source was held at 1.4 kV, the temperature was set at 50 °C and the sampling cone voltage was set at 70 V.

2.3.5 Structural integrity assessment

2.3.5.1 Lysozyme activity assay

Cell suspensions of Micrococcus lysodeikticus ATCC (Sigma-Aldrich, Poole, UK, 1 ml, 0.01% in 60 mM potassium phosphate buffer, pH 6.2) were equilibrated in a Multiskan Go microplate spectrophotometer (Thermo Scientific) at 25 °C for 5 minutes with or without the presence of 10 mM aryldiazirine . A solution of Lysozyme (50 μl, 0.1 mg ml^{-1}) was added to each suspension, and the decrease in absorbance at 450 nm was measured every 60 seconds for a total of 10 minutes. As a control, a cell suspension without the addition of either lysozyme or AD was similarly monitored.

2.3.5.2 USP5 activity assay

A solution of USP5 (Boston Biochem, Cambridge, MA, 1 μl, 1.25 μg/μl) was incubated with Lys48 poly-ubiquitin ladder (Boston Biochem, 50 ng) in 50 μL of TBS-DTT (50 mM Tris, 150 mM NaCl, 10m M DTT) with or without aryldiazirine (10 mM) overnight at 37 °C whilst shaking. Proteins from each incubation were directly added to 20 μL gel-loading buffer (0.15 M Tris, 8 M Urea, 2.5% (w/v) SDS, 20% v/v glycerol, 10% (w/v)

2-mercaptoethanol, 3% (w/v) DTT, 0.1% (w/v) bromophenol blue, pH 6.8 (HCl)) and resolved on 5-20% SDS PAGE, western blotted and visualized with anti-ubiquitin (VU-1, Life sensors, Malvern, PA) for the presence of poly-ubiquitin disassembly.

2.3.5.3 Circular dichroism

CD spectra of lysozyme were recorded on a Chirascan CD spectrometer following AD labelling. 400 μL of a 375 μg/mL solution of post-labelling mixture was transferred to a quartz cuvette and its CD spectrum was acquired at a scan rate of 0.5 points per second across a range of wavelengths from 180 nm to 260 nm. The same conditions were employed for the analysis of the unlabelled control protein.

2.3.6 Data analysis

2.3.6.1 Determination of the Extent of Labelling on intact proteins

The method proposed by Delfino et al.[117] was used to calculate the extent of labeling per protein (EL) obtained with the commercial photoleucine and AD . The spectra obtained for each intact labelled protein were deconvoluted using the Xtract algorithm (Thermo Xcalibur software) and the extent of labelling was calculated using the following equation:

$$EL = \sum_{i=0}^{i=n} iI_i \Big/ \sum_{i=0}^{i=n} I_i \tag{2.3.1}$$

where Ii is the sum of the intensities of all the peaks for each species *i* including oxidation products and *i* is the index indicating the number of labels inserted per protein.

2.3.6.2 Peptide level analysis

The method proposed by Jumper et al. was used to evaluate the amount of labelling on each peptide and on single residues[68]. The chromatograms for each singly-labelled and unlabelled peptide were extracted within a range of ± 0.1 m/z and the spectrum for each peak was manually inspected to ensure the sampling of the correct ion only. In the few cases where there was an interfering signal in the same mass window as the selected peptide, a smaller mass tolerance or the second isotope of the peak envelope was selected for the extraction of the ion chromatogram. The peptide fractional modification was obtained calculating the ratio between the peak area of the labelled form and the sum of the peak areas of labelled and unlabelled forms.

$$P = \frac{A_{labelled}}{A_{labelled} + A_{unlabelled}} \tag{2.3.2}$$

The fractional modification values obtained for each peptide derived from bound and unbound proteins were respectively averaged and compared in order to determine the difference in the reaction extent. Differences in the extent of labelling between peptides were considered significant when the p-value obtained from a Student t-test was <0.05

2.3.6.3 Residue level analysis

Modification sites on the peptides were investigated using tandem MS. On each singly-labelled peptide the label is distributed over the sequence and after undergoing fragmentation it will generate labelled and unlabeled fragments. The data has been averaged over all the chromatographic features of the precursor and the ratio between the intensity of the n_i labelled fragment and the sum of the intensities of n_i labelled and n_i

unlabelled allows the calculation of the fractional modification on the ni residue.

$$f.mod(n_i) = \frac{I(n_{i_{labelled}})}{I(n_{i_{labelled}}) + I(n_{i_{unlabelled}})} \tag{2.3.3}$$

The difference in fractional modification between two consecutive fragments multiplied by the fractional amount of labelling on the peptide will give the absolute level of modification on the n_i residue

$$abs.mod. = P[f.mod(n_i) - f.mod.(n_{i-1})] \tag{2.3.4}$$

In the case of the impossibility of detection of a n_i ion, its fractional modification was grouped together with its subsequent fragment. The few fractional modification values of residues that showed an anomalous increase or decrease in respect to the fractional modification curve trend were excluded from the dataset and were grouped together with the value corresponding to the adjacent residue resulting in only a small loss of resolution. The phenomenon appeared mostly on fragments corresponding to cleavage of the peptide bond of serine, threonine and cysteine residues for reasons yet to be investigated. Per-residue modification values from each repeat were averaged and the averages compared between the two states of the protein. Differences in per-residue modification were considered significant when the p-value obtained from a Student t-test was <0.05 and the difference in label/residue was >0.01

CHAPTER 3

Introducing a novel photoreagent for carbene footprinting: reactivity characterisation and mapping of the lysozyme-NAG5 binding interface

3.1 Introduction

As mentioned in Chapter 1, MS-based structural techniques have several advantages over more "classic" structural biology tools such as X-ray crystallography and NMR. HDX, covalent labelling and cross-linking have in common high rapidity and low consumption of analytes. HDX is nowadays widely used for the study of protein complexes but it suffers from two main drawbacks: undesired hydrogen-deuterium back exchange under common bottom-up workflow conditions, and potential H/D scrambling when analytes are subjected to CID MS/MS experiments[47,151]. Amino-acid specific covalent labelling and chemical cross-linking have also proved successful in the mapping of protein interactions. The main concern regarding these two approaches is the relatively low resolution achievable resulting from the selective reactivity of the

reagents employed towards specific amino acid side chains, which in turn leads to low labelling density. Some of the probes commonly employed also exhibit relatively slow kinetics towards amino acid side chains, leading to potential alteration of the native protein structure or dissociation of the complex[152]. Hydroxyl radical footprinting in its FPOP variation is a technique that allows the labelling of several amino acids side chains with very rapid kinetics. The technique has been successfully employed in a number of applications including protein complex mapping, investigation of folding intermediates and in-vivo protein footprinting[86,153,154]. One drawback of such a strategy is the need to expose the protein to hydrogen peroxide, which may lead to unwanted protein oxidation prior to the photo-activation of the reagent. Moreover, the hydroxyl radical reactivity is severely skewed towards certain amino acid side chains, which in some scenarios it may lead to loss of resolution, and also the 266 nm laser used to generate hydroxyl radicals from hydrogen peroxide can potentially damage some of the more susceptible sites of the protein under examination.

Photoactivable diazirines have been widely used for photoaffinity labelling, a strategy in which a specific molecule is functionalised with a diazirine ring and its binding to a target protein is investigated upon photo-activation. As reported in Chapter 1, diazirines have also been used for protein footprinting, a completely different methodology to photoaffinity labelling, in which a single diazirine-based reagent is used to label more or less selectively residues on the surface of the protein rather than using a specific diazirine derivative to target the active site.

Herein we describe the use of a newly designed and synthesised diazirine-based reagent for carbene-based differential protein labelling experiments. This photoreagent proved to be significantly more efficient than the commercially available Photoleucine at substantially lower concentration and it resulted non-disruptive towards the native conformation of proteins. The probe has been successfully used to map at near amino acid resolution the binding surface of the lysozyme-NAG5 complex and the subtle conformational changes associated with the binding event.

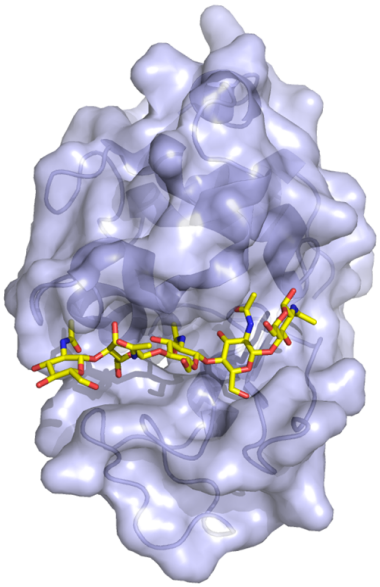


Figure 3.1: X-ray powder diffraction structure of lysozyme bound to NAG5 (ligand rendered as sticks in yellow). Rendered in Pymol from PDB: 1SFB

3.2 Results and discussion

3.2.1 Design of a new footprinting probe

The preferential reactivity of a water-soluble diazirine towards different amino acids has been shown to be dependent on the structure of the probe which influences its partitioning over the surface of the protein prior to activation. As shown by Schriemer *et al.* the presence or absence of an amino group distant from the reaction centre (i.e. the diazirine ring) can steer the selectivity of carbene-based labels towards acidic residues if the diffusion of the reagent is reduced by flash-freezing[152]. In order to obtain a diazirine-based reagent exhibiting improved labelling efficiency, a diazirine ring was installed on the benzylic carbon of sodium benzoate; the aromatic ring provides a non-polar surface for the interaction with hydrophobic regions of the protein and the carboxylic moiety promotes the association with basic amino acids and improves the solubility of the reagent in water. A trifluoromethyl group was also incorporated on the benzylic carbon in order to reduce the reactivity of the diazo-compound generated as a minor reactive intermediate from the photolysis of diazirines[155]. Diazoisomers are less reactive than carbenes and can potentially diffuse away from the site where they are generated leading to misinterpretation of the labelling data. The reduced reactivity provided by the presence of the trifluoromethyl group results in a reduction of unwanted side reactions. The ground spin state of aryltrifluoromethyl carbenes has been shown to be the triplet but it mainly reacts through its low-lying singlet state[156]. This feature allows aryltrifluoromethyl carbenes to insert into a wide array of chemical functionalities including C-H bonds of non-polar amino acid side chains and X-H bonds (where X = O, N, S) found on polar amino acid side chains improving the resolution of the structural information obtainable through protein footprinting[111]. This aryldiazirine (from now referred as AD) (Sodium 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoate) was developed in collaboration with Dr. John Moses and was synthesised by Dr.Andrew Barrow via a modified literature procedure from 4-bromotoluene (Fig.

3.2). The free acid of the molecule has been previously synthesised and used in the development of photoaffinity reagents but never as a footprinting probe.

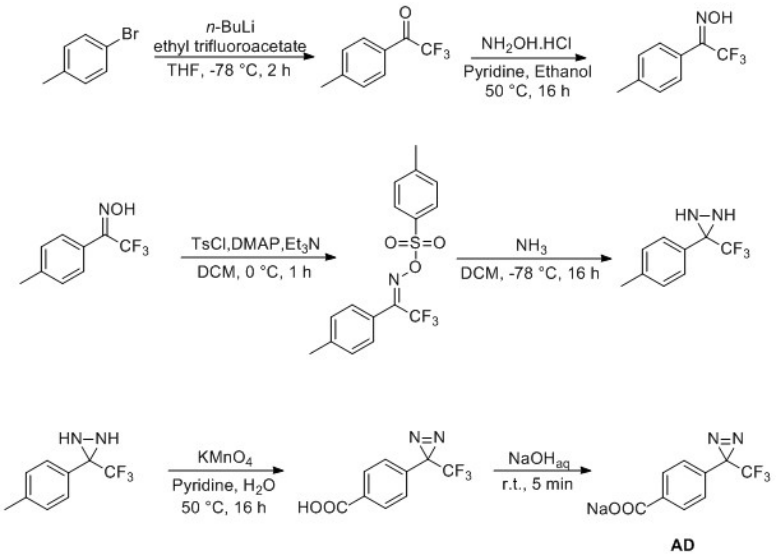


Figure 3.2: Synthesis of Sodium 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoate starting from 4-bromotoluene

3.2.2 Characterisation of AD reactivity

Before testing the ability of the novel reagent to label biological molecules it is useful to characterise its intrinsic reactivity and potentially shed some light on the factors governing the labelling process. In order to do so the photoreagent was irradiated with a UV laser at 349 nm in water. The products were separated by HPLC, detected using both UV absorbance at 254 nm and mass spectrometry (shown in Fig. 3.3). Surprisingly the most abundant product originated from the reaction resulted to be the dimer composed by a molecule of the carbene and a molecule of unreacted reagent. This phenomenon indicates that the reagent tends to non-covalently self-associate in solution in absence of a different substrate such as a peptide or a protein before carbene generation. The connectivity of the two monomers was not completely characterised but there is chromatographic evidence of multiple isomer formation. As expected one of the products revealed to be originated from the insertion of the carbene into a solvent molecule resulting in the formation of the corresponding alcohol. Two additional

peaks were assigned to unreacted AD and to the formation of the corresponding peroxide. The product corresponding to the peak eluting at approx. 4 minutes in Fig.3.3 showed an *m/z* of 339.5 which does not correspond to any of the expected products and will need more experimental data in order to be identified.

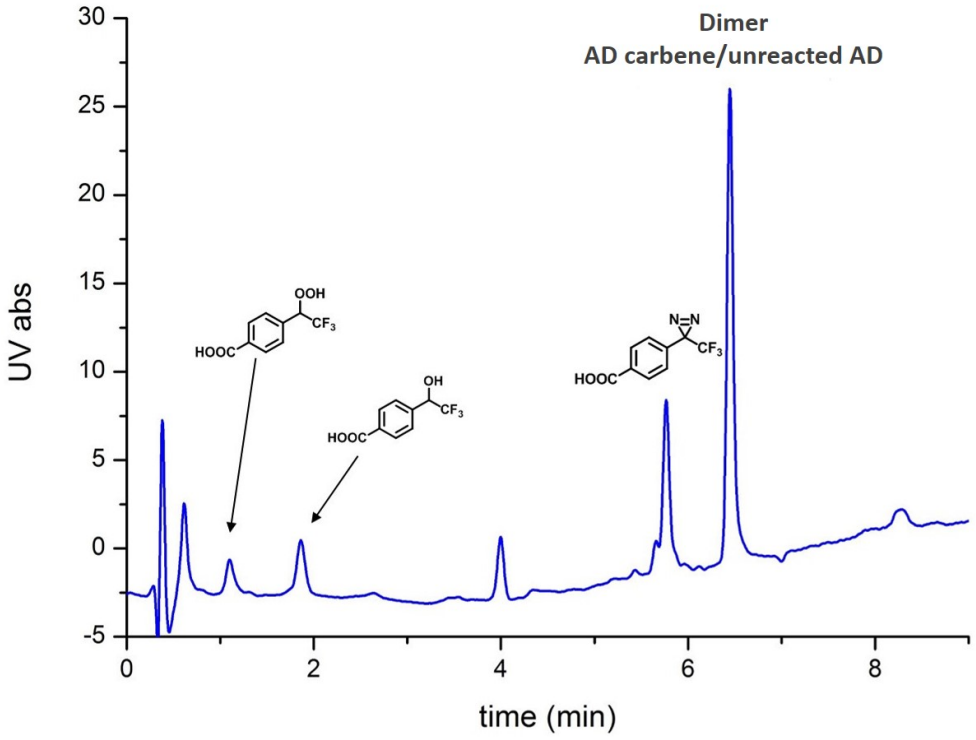


Figure 3.3: HPLC-UV chromatogram of photolysis products of AD in water in absence of protein. The products were identified by mass spectrometry. The most intense signal eluting at 6.3 minutes corresponds to the product of reaction of a molecule of activated AD with a unactivated molecule of the reagent.

3.2.3 Labelling efficiency of AD

In order to characterise the reactivity of the novel reagent towards proteins, photolysis experiments were performed on ubiquitin, as a model in the presence of increasing concentration of the diazirine probe. Irradiation was conducted at -78 °C to avoid diffusion of the reagent away from the protein surface following the findings of Jumper and co-workers[152]. Laser repetition frequency was set at 1000 Hz and the pulse energy was set at 125 μJ. Samples were irradiated for 16 seconds in order to impart in the samples the same energy used by Schriemer et al. for Photoleucine-based protein labelling[152]

sufficient for the conversion of 200 μmol of diazirine. Samples were then desalted using a standard ZipTip protocol (see Chapter 2) and analysed by static nano-ESI. The spectra obtained were deconvoluted and they are shown in Fig. 3.4. The intensities of the signals corresponding to unlabelled protein and to the addition of up to three molecules of the probe to the protein were used to calculate the extent of modification achieved for each reagent concentration.

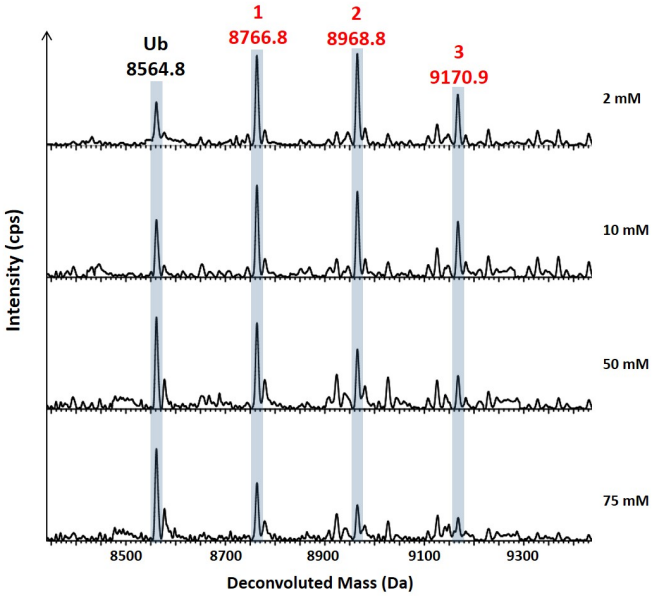


Figure 3.4: Deconvoluted spectra of ubiquitin labelled in the presence of AD at 2 mM, 10 mM, 50 mM and 75 mM concentration. Maximum extent of labelling observed irradiating the protein in the presence of 10 mM AD, the use of higher concentration of reagent resulted in a lower labelling yield.

Photolysis at a concentration of 2 and 10 mM reagent resulted in a product distribution where the most represented species were the singly and doubly labelled; the triply labelled product was also more represented than the unlabelled protein. Surprisingly the product distribution obtained from photolysis using a 50 mM concentration of the probe was more skewed towards the unlabelled and singly labelled protein with the doubly and triply labelled species substantially less represented. When the photolysis was performed at 75 mM concentration the most represented species was the unlabelled protein; the singly labelled protein resulted to be significantly lower than the unlabelled and the doubly and triply labelled species were almost undetectable. The

lower labelling efficiency exhibited by the aryldiazirine at high concentration can be attributed to the self-association of the reagent prior to the reaction and subsequent dimerisation resulting in lower protein labelling. This hypothesis is strengthened by the presence of AD dimer in reaction mixtures irradiated in absence of protein and its disappearance when a polypeptide is added to the mixture (see Fig.3.3).

The labelling efficiency of AD was then compared to that of the commercial photoleucine. In order to do so, a range of model proteins were labelled with both reagents and the extent of modification was calculated as described in Chapter 2. For photoleucine labelling experiments, the conditions used corresponded to the optimal ones described previously in literature to allow complete conversion of the diazirine[152] (100 mM concentration, 16 seconds irradiation on the available setup). A concentration of 10 mM was used in all the AD labelling experiments because although it is possible to achieve the same level of labelling using lower concentration as shown above, it is vital that the probe is present in large excess with respect to the protein. As shown in Fig 3.5, AD exhibited substantially higher efficiency using only 10 mM concentration and 4 seconds irradiation. The higher modification yield resulted constant on all the peptides and proteins tested.

The efficiency of photoleucine was found in most cases to be modest with the exception of calmodulin, which also resulted to be very highly labelled irradiating the sample for only 1 second in the presence of 10 mM AD. Excluding calmodulin, the degree of labelling obtained for the model proteins investigated resulted to be sub-optimal for footprinting experiments.

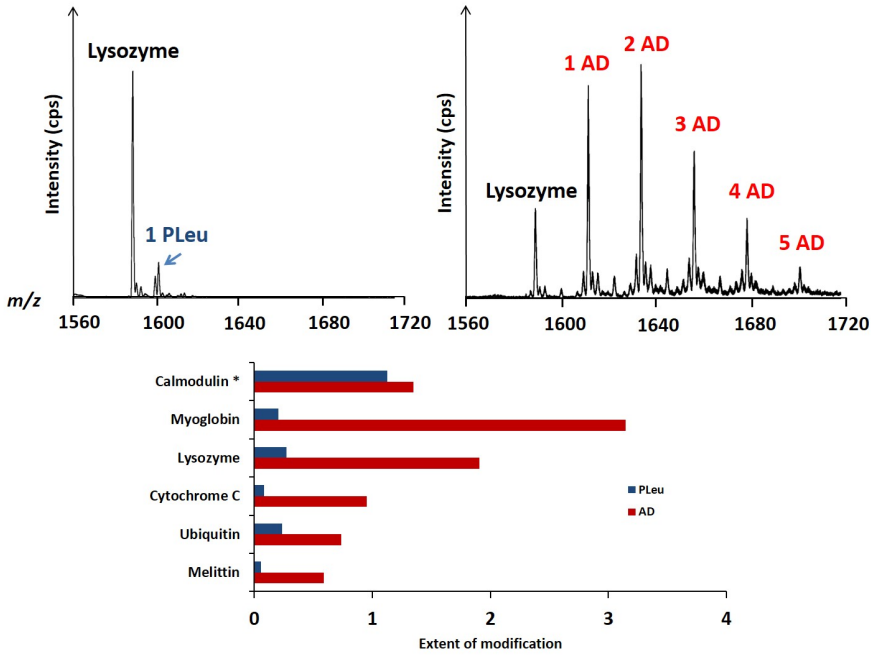


Figure 3.5: ESI mass spectra of the 10+ charge state of lysozyme photolabelled with 100 mM photoleucine (top left) and 10 mM AD (top right). Photoleucine labelling lead to the addition of one molecule of the reagent to approximately 10 % of the protein. Irradiation in the presence of 10 mM AD yielded up to 5 additions of the reagent to lysozyme. The superior labelling efficiency of AD was confirmed performing labelling experiments on various model proteins (bottom). Extent of modification was calculated as described in chapter 2

3.2.4 Selectivity towards amino acids

The postulated preferential reactivity of AD towards hydrophobic and basic amino acid side chains induced by the partitioning of the reagent before photoactivation was also investigated. A mixture composed of thirteen of the natural occurring amino acids was incubated with 10 mM AD, flash frozen and irradiated in the same conditions used for the model proteins employed in the experiments described in the previous section. The mixture was then separated using HILIC chromatography, without any further modification, and the products detected by ESI-MS. The chromatograms corresponding to each unlabelled amino acid were extracted, integrated and each peak area value was compared to the peak area of the corresponding amino acid from the control mixture (identical amino acid mixture irradiated in absence of the photoreagent). A decrease in peak area indicated that part of the unlabelled amino acid reacted with the photo-

reagent revealing the preference of insertion of the label towards single amino acids. Four of the thirteen amino acids exhibited significant depletion, Arg, His, Lys and Met (see Fig.3.6); presenting on their side chains respectively a guanidinium group, an imidazole and an amino group, the first three amino acids exhibit basic properties whereas methionine possess a hydrophobic side chain.

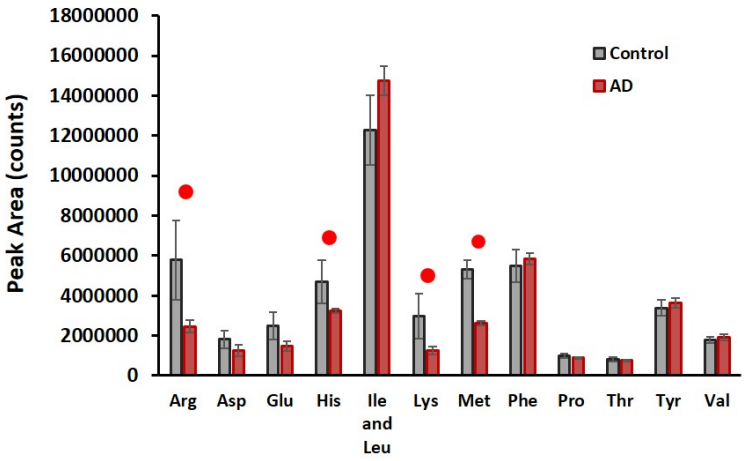


Figure 3.6: Peak area values of integrated extracted ion chromatograms of unlabelled single amino acids post-laser irradiation in the presence and absence of AD (respectively red and grey bars). Decrease in peak area revealed selectivity of AD towards Arg, His, Lys and Met (difference exceeding ±SD, indicated by red dots). Error bars are ±SD, n=3

The depletion of basic and hydrophobic amino acids confirms the hypothesis of skewed selectivity towards such functionalities in slow diffusion conditions. Although no significant labelling was observed for any other amino acids, labelling on different protein residues cannot be excluded given that the selectivity observed is due to the partitioning of the reagent prior activation and not to the intrinsic reactivity of the carbene. Amino acids located in proximity of preferred residues are likely to react with the carbene, depending on the orientation of the reagent in respect to the basic or hydrophobic amino acid. Furthermore hydrophobic amino acid side chains are often involved in the formation of hydrophobic patches to which AD may exhibit selectivity because of the contribution of several non-polar side chains, selectivity not detected in isolated amino acids. The presence of free -NH_2 moieties may also have affected the selectivity causing an overall decrease in differential labelling and potentially difficulties in detecting total

subtle differences. Nevertheless the preferential reactivity observed is still significant considering that each amino acid in the mixture presented a free N-terminus.

3.2.5 Structural integrity studies

Structural information obtained from any covalent labelling methodology is valuable only if the native structure of the protein is preserved before and during the reaction. Once labelling is complete the preservation of the structure is no longer required. Depending on the size and features of the probe used, alteration of the native conformation is likely upon modifcation of the exposed residues and subsequent local change in the surface properties of the protein. If this phenomenon occurs the conformation probed is no longer the correct one and the reliability of the structural data obtained becomes questionable. Several approaches have been described to ensure structural integrity related to covalent labelling techniques including monitoring by circular dichroism (CD), reaction kinetics, activity assays, fluorescence spectroscopy and the use of a limited amount of modification. Circular dichroism is sensitive to global alteration in protein structure but it is insensitive to local changes which in turn can be detected using fluorescence spectroscopy but only if a fluorophore is located in the region of interest. Keeping the number of modifications relatively low can be a reliable method to maintain the native fold unaffected but in the case of multiple potential modification sites, the extent of overall labelling can be extremely low leading to difficulties in the detection of modified residues. Activity assays are very informative methods to assess structural changes but perturbation in activity mostly occurs only if the active site of the enzyme is involved in the modification. A combination of different methodologies seems to be the most effective way to assess the structural integrity of the protein under investigation[56].

In order to ensure that the presence of AD in solution does not perturb the native fold of the protein prior to the UV laser irradiation an activity assay in the presence and absence of the reagent was performed. The activity of lysozyme was monitored through

a well established turbidimetric assay; the cell wall of *micrococcus lysodeikticus* cells is destroyed in the presence of lysozyme resulting in a clarification of the cell suspension and the phenomenon can be followed by measuring light transmission through a suspension of bacteria in buffer.

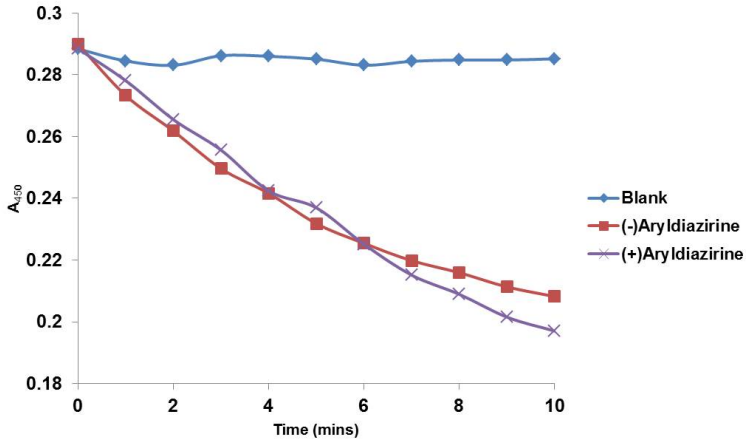


Figure 3.7: Plot of absorbance vs. time for lysozyme activity using Micrococcus lysodeikticus cells showing that the presence of AD has no effect on the activity of the enzyme, and by inference its structure. Control (+ lysozyme, - AD), AD (+ lysozyme, + AD)

The experiment was performed in the presence of 10 mM AD and the activity was monitored over 10 minutes. A control experiment without any reagent added was also performed and both activity curves are shown in Fig. 3.7. No change in activity was detected in the presence of the diazirine, indicating that the structure of the enzyme was not affected by the reagent.

As mentioned above, if the addition of the reagent can occur on multiple sites of a protein there is the possibility that upon the first addition the conformation of the protein may change and subsequent additions of the probe can occur on a protein which is no longer in its native conformation. This point has been considered by Gau et al. for FPOP experiments[70]. Assuming that the probability of modification at any site is only a function of its solvent accessibility and that upon modification there is no structural perturbation, then the probability of each addition to the surface is not dependent to any other labelling event. The probability that a protein is labelled k times is described

by the following:

$$P(k;N,p) = \binom{N}{k} {p_{mod}}^{k}(1-p_{mod})^{N-k} \tag{3.2.1}$$

where N is the number of modification sites and p_{mod} is the probability of modification at each site. As $N \rightarrow 0$ the binomial distribution approaches the Poisson distribution described by the probability function

$$P(k;\lambda) = \frac{e^{-\lambda}\lambda^{k}}{k!} \tag{3.2.2}$$

If the labelling reaction occurs without significant structural perturbation it follows that the distribution of labelled products should be close to a Poisson distribution. In cases where significant unfolding happens during the reaction, the distribution should no longer be described by the Poisson model. In such scenario the resultant distribution deviates from Poisson firstly because the probabilities of modification at each site are not independent and secondly because after unfolding, the newly exposed sites are available for the reaction for a shorter time. This event equals to the labelling of two or more different conformations of the same protein at the same time leading to a distribution that could be approximated to Poisson-like for each conformation but with the overall product distribution deriving from the sum of them deviating from the ideal Poisson distribution.

In order to test if any structural alteration happens during the labelling reaction with the novel aryldiazirine, a range of proteins were labelled using standard conditions (10 mM AD, 16 seconds exposure except calmodulin exposed for 1 second) and the distribution of the labelled products on the intact proteins was analysed. The spectra obtained were deconvoluted using the Xtract algorithm implemented in Xcalibur, the

intensities of each signal corresponding to modified and unmodified protein was plotted as a percentage of the total modification and the curve obtained was fitted to the Poisson distribution.

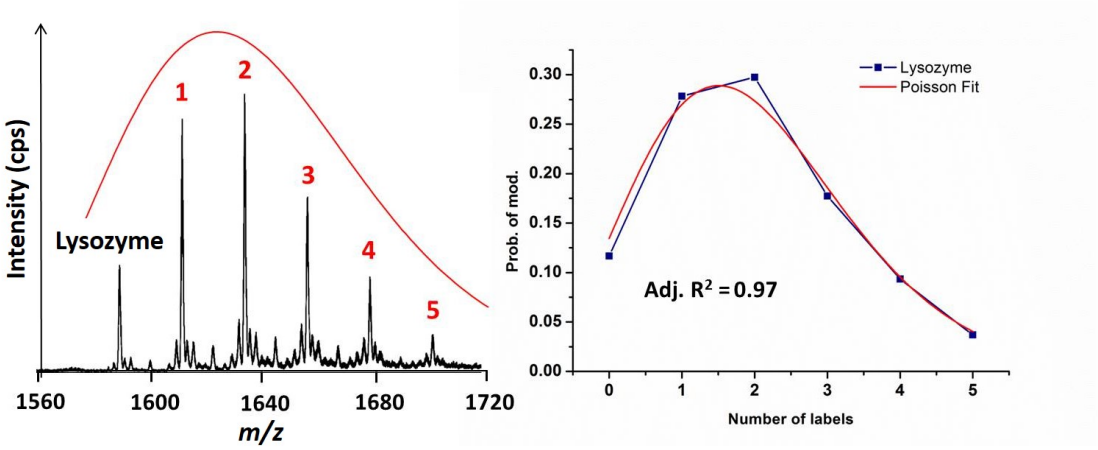
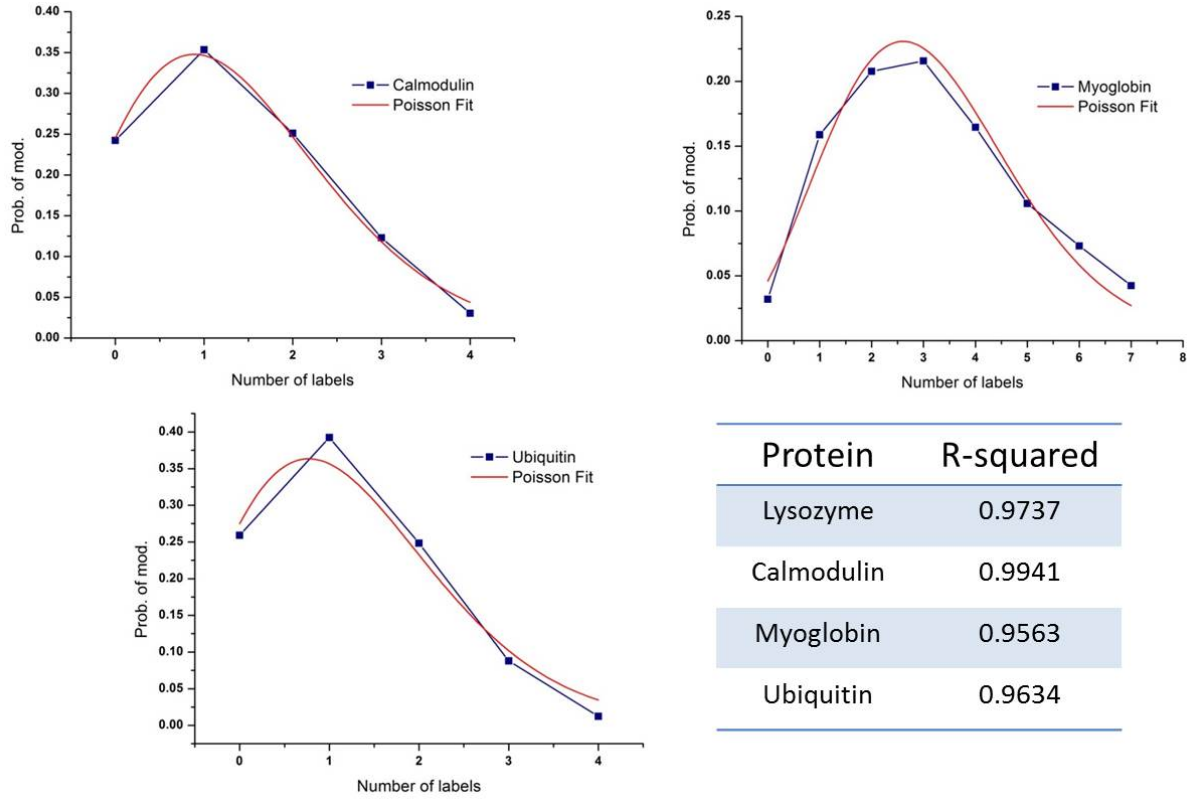


Figure 3.8: (Left)ESI-MS spectrum of the 10+ charge state of AD labelled lysozyme showing up to 5 insertion of AD. The red line indicates the calculated Poisson distribution for 5 random events. (Right) Probability distribution of AD products (blue line) and calculated Poisson fit (red line). The goodness of the fit indicates that no structural perturbation occurs during the labelling reaction.

All the proteins considered showed distribution very nearly Poisson with R-squared values between 0.95 and 0.99. This result suggests that during the labelling process there is no significant protein unfolding occurring. Carbenes have been reported to react on the scale of picoseconds therefore it is possible to assume that most of the labelling reaction occurs faster than the eventual unfolding caused by the presence of the covalent modification.



Protein	R-squared
Lysozyme	0.9737
Calmodulin	0.9941
Myoglobin	0.9563
Ubiquitin	0.9634

Figure 3.9: Probability distribution of AD products (blue line) and calculated Poisson fit (red line) of calmodulin, myoglobin and ubiquitin. The goodness of the fit indicates that no structural perturbation occurs during the labelling reaction

Further evidence of the absence of structural perturbation was provided by circular dichroism analysis performed before and after the labelling process on lysozyme. As shown in figure 3.10 the two profiles are almost indistinguishable and the characteristic absorbance at 210 nm and 230 nm are retained post-labelling indicating that the secondary structure elements are preserved post-labelling. It could be argued that lysozyme is a protein well known for its stability and therefore still folded after a covalent labelling process whereas a different, less stable protein may undergo some conformational changes when treated in this fashion. Unfortunately it was not possible to perform CD analysis on the other model proteins used for the studies mentioned above because the set-up used required large amount of protein material which was not available for the analysis. Nevertheless the absence of variation in the CD spectra

constitute valid evidence of structure preservation for lysozyme after photolabelling with AD.

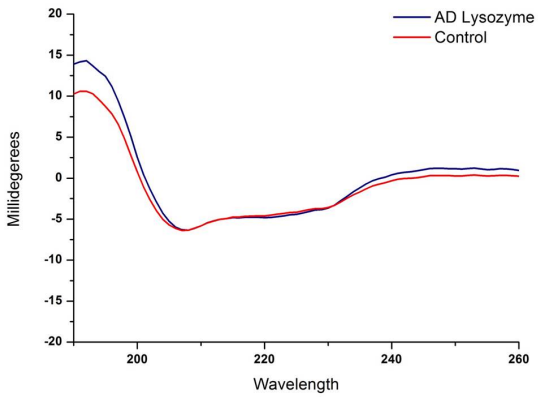


Figure 3.10: Circular dichroism spectra of lysozyme before and after photolysis in the presence of 10 mM AD (respectively red and blue trace). The spectra show no significant difference confirming lysozyme retains its secondary structural elements after labelling.

The evidence collected with the activity assays (more examples in the following chapters) and with the product distribution analysis leads to believe that there is no significant conformational perturbation induced by the presence of the reagent prior to the photolysis or during the reaction confirming that the novel aryldiazirine is suitable for protein footprinting studies.

3.2.6 Lysozyme-NAG5 footprinting

Hen egg-white lysozyme is a 129 amino acids enzyme that catalyses the hydrolysis of β-1-4 glycosidic bonds of bacterial cell wall, consisting of alternating polymers of N-acetyl-glucosamine and N-acetyl-muramic acid. The three-dimensional structure of this enzyme has been elucidated by Blake *et al.* in 1965 using X-ray crystallography, the first enzyme structure to be solved at high resolution[157]. Since then the structure of the protein in complex with various substrate analogues has also been studied[158]. The protein presents a pronounced cleft on its surface composed by six subsites, named A to F, in which carbohydrate residues bind and undergo hydrolysis at defined positions. The complex between Lysozyme and NAG5, a pentasaccharide composed by units of

N-acetyl-glucosamine, was chosen as the model system on which the possibility of using the novel aryldiazirine as a protein footprinting reagent was tested. The structure of the complex between lysozyme and NAG5 was solved in 2004 and it shows how each glycosidic unit binds to a subsite on the surface cleft of lysozyme[158] (Fig.3.13).
As shown in Fig.3.8, photoactivation of 10 mM AD in the presence of lysozyme following flash freezing leads to the addition of up to five molecules of the reagent to the protein. The low temperature at which the experiment is performed prevents diffusion of the probe and also reduces the hydrolytic activity of the enzyme towards NAG5. In order to identify which areas of the protein were labelled and quantify the extent of the modification at the peptide level, the protein labelled in its bound and unbound state was subjected to proteolysis using AspN or trypsin and the resulting peptide mixture, representing 79% coverage of the protein, was analysed using LC-MS.

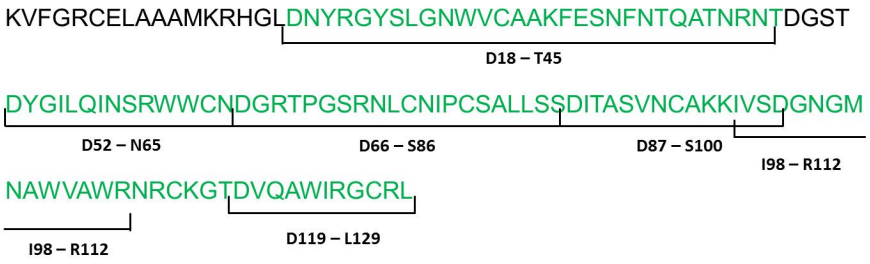


Figure 3.11: Peptide coverage obtained digesting lysozyme in gel using trypsin and AspN. Identified peptides are indicated in green.

Observing the extracted ion chromatograms of the labelled species it is evident how their retention time on reversed phase columns is longer, in agreement with the higher hydrophobicity imparted by the addition of the label. Another striking feature of the labelled peptides is the presence of multiple, partially resolved peaks; this phenomenon is due to the partial co-elution of different positional isomers of the labelled species (see Fig. 3.12). Being extremely reactive, carbenes can potentially react with a large number of chemical functionalities even within the same amino acid residue, leading to a number of chemically similar positional isomers which can be challenging to separate fully by HPLC.

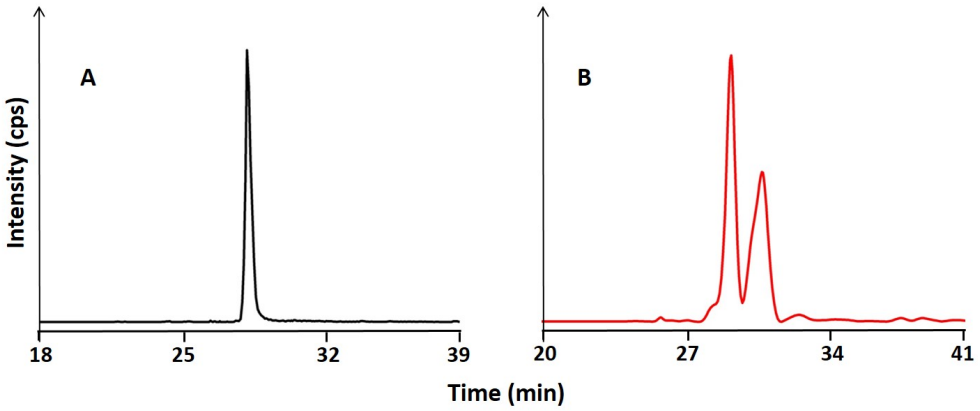


Figure 3.12: Extracted ion chromatograms of unlabelled peptide D52-N65(A) and labelled peptide D52-N65 (B). It is noticeable how the labelled peptide produces a chromatogram with multiple peaks.

In order to improve the structural resolution achievable, each labelled peptide was subjected to collision induced dissociation and the resulting fragmentation products were used to quantify the modification on single amino acids or small regions in cases where the fragmentation yield was suboptimal. Among the peptides obtained following proteolysis with AspN or trypsin, peptide D52-N65 is the most suitable to exemplify the capabilities of the methodology given that it is composed of amino acids with a wide range of polarities and it forms large part of the binding site including residue W62, which has been shown to be necessary for ligand binding. (shown in Fig. 3.13).

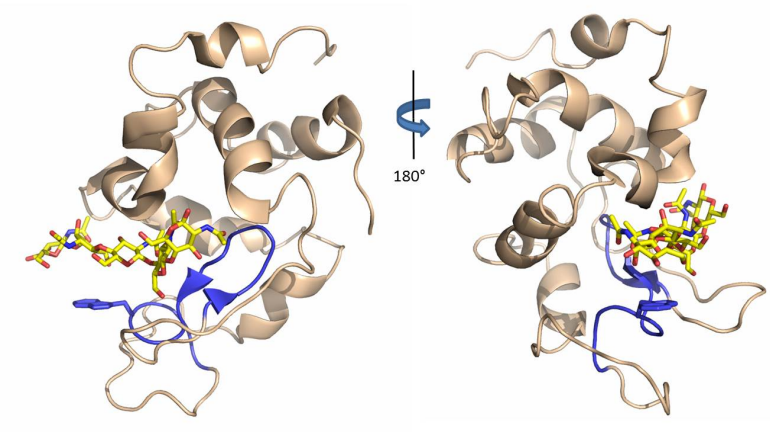


Figure 3.13: Area corresponding to peptide D52-N65 (blue). NAG5 is represented in yellow sticks. The peptide corresponds to large part of the binding site of lysozyme. Rendered in Pymol from PDB: 1SFB

After isolation and CID fragmentation of the ion corresponding to singly labelled D52-N65 the data were treated as discussed in the Methods section and the extent of labelling achieved on single residues or small groups of residues was calculated. As shown in Figure 3.14, the reagent preferentially targeted hydrophobic and basic residues respectively represented by R61 and W62. Such reactivity skewed towards amino acids with these features was expected given that the reagent was provided with a carboxylic moiety and an aromatic ring. As shown by Schriemer and co-workers[152], the preferential targeting of certain residues is driven by partition of the reagent on the surface of the protein before photoactivation. The carboxylic moiety of the reagent is anionic at pH 7.4 and it is likely to be involved in ionic interactions with positively charged side chains whereas the aromatic ring is expected to be involved in hydrophobic interactions with non-polar amino acids. The results show agreement with this model but also targeting of neighbouring residues cannot be ruled out.

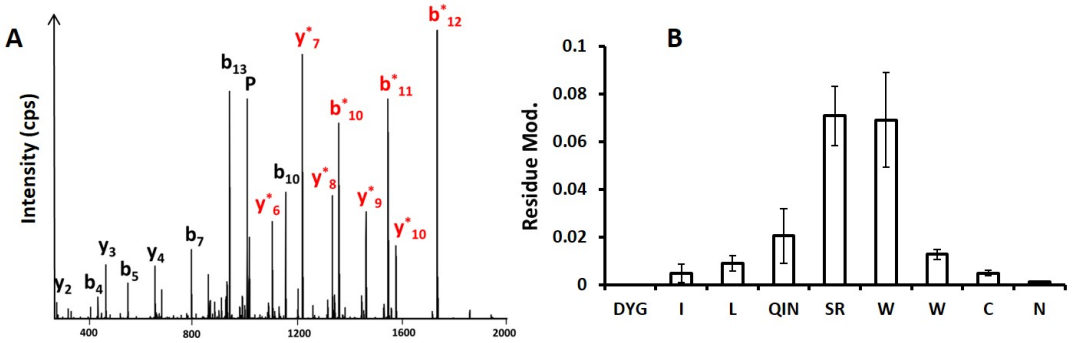


Figure 3.14: A) ESI CID MS/MS spectrum of labelled D52-N65. Labelled and unlabelled fragments are respectively noted in red and black. B) Average number of label per residue on peptide D52-N65. Error bars represent ± SD, n=3

The labelling experiment was performed under the exact same conditions except for the presence of NAG5 in a 1:1 ratio to the protein. The ligand was added moments before snap freezing the mixture in order to stop the turn-over of the enzyme. The dissociation constant of the lysozyme-NAG5 complex has been reported to be approximately 5.5 μM[159] therefore at the concentration of protein and ligand at which the experiment is performed (80 μM) most of the protein is expected to be in its bound form. Observing the chromatographic profile of the labelled form of peptide D52-N65

derived from bound lysozyme it is evident how the features of the chromatogram are different. Such difference is a first indication that the label distribution over the peptide is affected by the presence of the ligand.

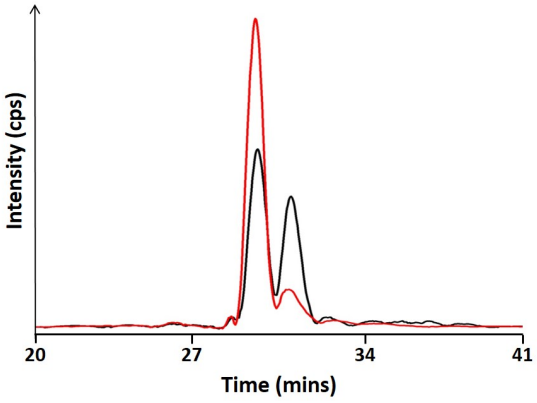


Figure 3.15: NanoHPLC-MS extracted ion chromatograms of labelled peptide D52-N65 derived from unbound (black trace) and lysozyme-NAG5 (red trace). The different chromatographic profile suggests a different distribution of the reagent over the peptide.

The labelled peptide was subjected to CID fragmentation in the same conditions used for the peptide labelled in absence of ligand, the extent of labelling on each residue was evaluated and compared to the per-residue modification levels obtained in absence of ligand. Several residues showed a significant decrease in labelling when the protein was bound to NAG5, a phenomenon caused by the presence of the ligand in the binding site that led to a decrease in accessibility to the reagent. Interestingly the extent of labelling observed on the unresolved segment S60-R61 resulted to be higher in the presence of the ligand (see Fig.3.16A); analysing in more detail the x-ray structures of free lysozyme and of the NAG5 complex it is possible to observe that when the protein is not bound to any ligand, the side chain guanidinium group of R61 is involved in a hydrogen bond with the side chain carboxylate of D48. In the presence of NAG5 the same carboxylate is involved in an ionic interaction with the amino group of the fourth N-acetyl-glucosamine unit of NAG5 leaving the side chain of R61 more accessible to the probe and resulting in an increased extent of labelling (see Fig. 3.16C).

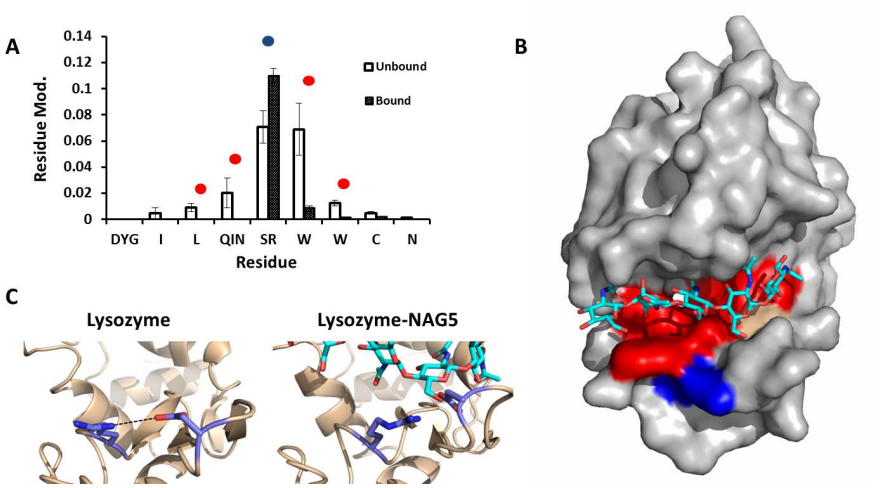


Figure 3.16: Differential footprinting of peptide D52-N65. A) average number of labels per residue in the presence (black bars) and in absence (white bars) of NAG5. Error bars are ± SD, n=3. Red and blue dots indicate respectively significant masking and unmasking on the corresponding amino acid. B) Differential surface plot. Red = significant masking by NAG5, blue = significant unmasking in the presence of NAG5 wheat = no difference in presence/absence of NAG5. Ligand shown as sticks in cyan. C) Details of the region around R61 of HEWL in NAG5-unbound (left, PDB 1LYZ) and -bound (right, 1SFB) states showing the disruption of the intramolecular interactions on NAG5 (cyan) binding and the resulting increased availability of R61 for labelling. Adapted from reference 160

The opposite side of the binding cleft is represented by peptide I98-R112 obtained from tryptic digest. The same CID-based experiment was performed on the labelled version of the peptide originated from the free protein and the protein bound to NAG5. The peptide resulted modified on all residues excluding W111 and the N-terminal R112 (see Fig. 3.17). Modification on amino acids corresponding to protease cleavage sites resulted to be non-existent or very poor in all the peptides observed. The absence of labelling detected on these residues does not imply that such amino acids are not labelled but it is very likely that proteases are not able to cleave the peptide bond of a modified residue making the modification difficult to detect and quantify. When compared with the per-residue labelling obtained on the bound protein, significant masking was observed on several residues of I98-R112, specifically on the section of the active site responsible for the binding of the first 3 NAG units (shown in Fig. 3.17). The residue that underwent the largest decrease in the extent of labelling was M105, which does not appear to be part of the active site, having its side chain located away from the binding cleft in the crystal structure of both bound and unbound lysozyme.

Such a residue, with W28 and W111 forms an area known as the "hydrophobic box" and it has been shown that ligand binding induces subtle structural changes resulting in a more rigid arrangement of the region which is highly likely to cause a decrease in the labelling efficiency at this site when bound to a ligand [161].

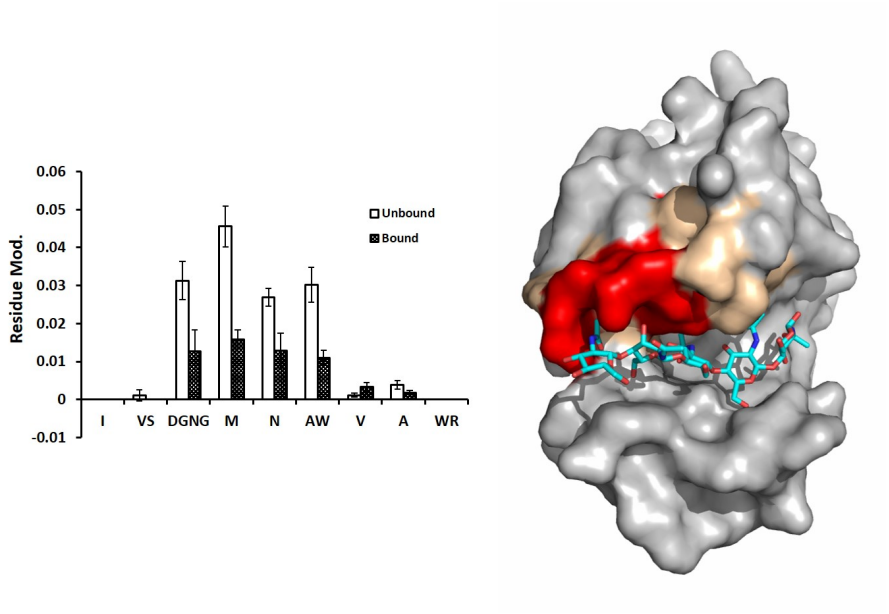


Figure 3.17: Differential footprinting of peptide I98-R112. Left: average number of labels per residue in the presence (black bars) and in absence (white bars) of NAG5. Error bars are ± SD, n=3. Red dots indicate significant masking on the corresponding amino acid. Right: Differential surface plot. Red = significant masking by NAG5, wheat = no difference in presence/absence of NAG5. Ligand shown as sticks in cyan

Peptide D119-L129 corresponds to an area of the protein that is not involved in the formation of the binding site and it is located on the opposite face of the protein. Its location makes D119-L129 ideal to be used as a negative control for this differential footprinting experiment since the extent of labelling on such peptide should not be influenced by the presence of the ligand. The peptide showed preferential labelling on R125 and on an unresolved area between G126 and L129 but pleasingly no labelling differences were observed in the two different states of the protein (see Fig. 3.18).

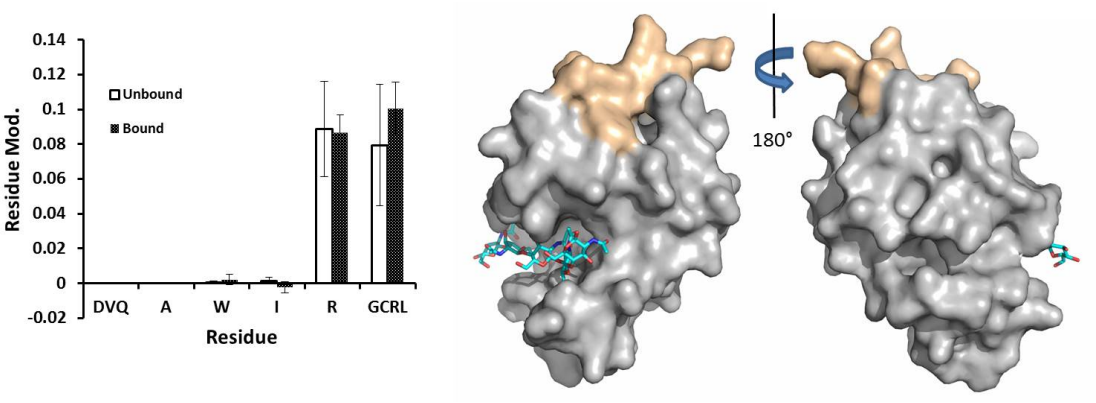


Figure 3.18: Differential footprinting of peptide D119-L129. Left: average number of labels per residue in the presence (black bars) and in absence (white bars) of NAG5. Error bars are ± SD, n=3. Right: Differential surface plot. Wheat = no significant difference, ligand represented in cyan sticks

The same differential approach was used on other three peptides in order to obtain as much structural information as possible and the data obtained, including the residue-level labelling described above are shown in Fig. 3.19. Differences between the two states of the protein were mapped on the crystal structure of the complex in order to have a clearer evaluation of the agreement between the complex structure described in literature and the data obtained using carbene footprinting (see Fig. 3.20)

The differential map (Fig. 3.20) shows clearly that the presence of the ligand reduced the extent of labelling in and around the active site whereas distal regions of the protein were not influenced by the binding event confirming the validity of the carbene footprinting approach to the mapping of protein-ligand complexes with near-residue resolution using AD as photolabelling reagent in combination with mass spectrometry.

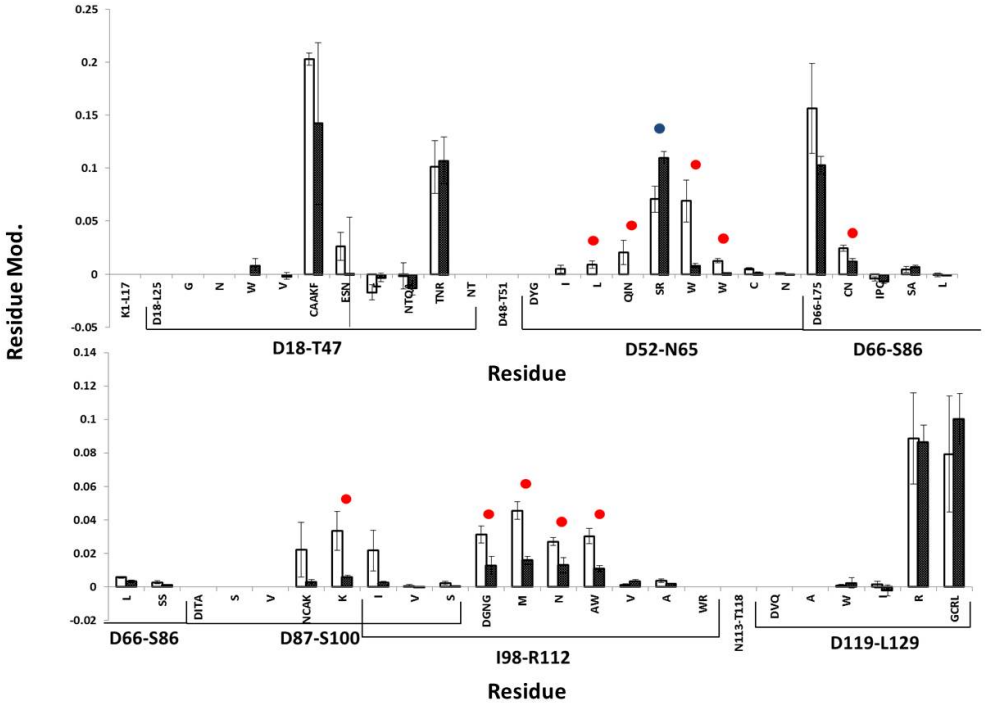


Figure 3.19: Average number of labels per amino acid residue. Black and white bars respectively represent the average number of labels per residue in the presence and absence of NAG5. Error bars are ± SD, n=3. Red and blue dots indicate respectively significant masking and unmasking on the corresponding amino acid. Adapted from reference 160

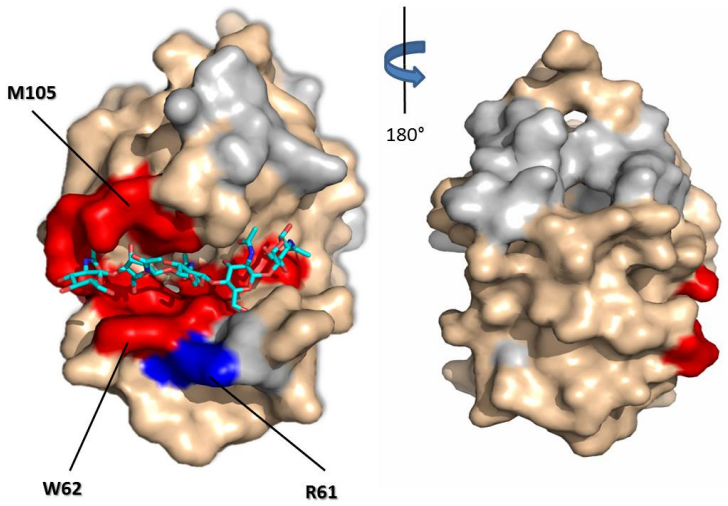


Figure 3.20: Differential surface map of HEWL. Red = significant masking by NAG5, blue = significant unmasking in the presence of NAG5, wheat = no difference in presence/absence of NAG5, grey = no data available. NAG5 ligand shown as sticks in cyan. Adapted from reference 160

3.3 Conclusions

The use of a novel water-soluble aryldiazirine suitable for near-residue resolution mapping of protein interaction sites through differential carbene footprinting in combination with mass spectrometry is here presented.

The behaviour of the novel reagent under photolysis conditions was investigated and the reaction products in the absence and presence of protein material were identified. The label showed a pronounced preferential reactivity towards the formation of its homo-dimer at -78 °C when no polypeptide was added to the reaction mixture whereas in the presence of a protein the major products resulted to be both homo-dimer and labelled protein. Because of its tendency to dimerise, higher concentrations of the reagent exhibited lower labelling efficiency compared to lower concentration towards model proteins with an optimal concentration value of 10 mM. Tested on a range of model proteins, AD exhibited far superior labelling efficiency compared to that of the commercial photoleucine, a water-soluble reagent previously described for protein footprinting, which resulted to be needed in 10-fold higher concentrations in order to achieve sub-optimal levels of labelling.

The aryldiazirine reagent is here shown to be benign and non-disruptive towards the native fold of proteins. Enzymatic activity proved to be unaffected by the presence of the label implying that the native structure is not altered in such conditions prior to label photoactivation. The distribution of labelled products detected in a standard carbene footprinting experiment resulted to follow a Poisson-like distribution which indicates that only one conformation of the protein is being probed and no significant unfolding occurs during the reaction as demonstrated by Gross and co-workers. Furthermore circular dichroism experiments showed retention of tertiary structure post-labelling in one of the model proteins employed in this work.

The novel probe was successfully employed to accurately map the binding interface between lysozyme and a penta-N-acetylchitopentaose carbohydrate substrate. The struc-

tural data collected resulted in very strong agreement with the current literature and furthermore carbene footprinting revealed fine details such as the disruption of an intermolecular hydrogen-bond upon ligand binding and subtle conformational changes associated with substrate interaction.

In summary the work in this chapter shows that differential carbene labelling with AD provides high quality structural data about the binding site of protein-small molecules interactions.

CHAPTER 4

Investigating the binding interface and conformational flexibility of the USP5-diubiquitin complex by carbene footprinting

4.1 Introduction

Protein ubiquitylation is a post translational modification that plays a fundamental role in a wide variety of cellular processes in eukaryotes. The conjugation of one or several molecules of ubiquitin, a 76 residues highly conserved protein, is responsible for localisation, degradation, interactions with other proteins, enzymatic activity modulation of substrate proteins and DNA repair[162–164]. This post translational modification occurs usually through the formation of an isopeptide bond between the C-terminal glycine carboxylic group of ubiquitin and the side chain amino group of a lysine residue exposed on the surface of the target protein. Ubiquitylation is regulated by three families of enzymes, E1,E2 and E3, as summarised in Figure 4.1. Briefly, the E1 enzyme promotes the activation of the C-terminus of ubiquitin by the formation of a thioester bond

between the C-terminal glycine and the side chain of a cysteine residue in its active site. The E2 conjugating enzymes receives the ubiquitin molecule from E1 and binds an E3 ligase enzyme. The substrate is then transferred on E3 which is then responsible for the conjugation of the ubiquitin moiety to its final substrate. Ubiquitin itself is a substrate of ubiquitylation given the presence of seven lysines on its surface. The ubiquitylation cycle can be performed several times resulting in polyubiquitin chains of different length and topology attached to the target molecule. Different chains have been found to be responsible for specific signalling outcomes. Lys48-linked polyubiquitin chains are recognised by the 26S proteasome and promote the degradation of the substrate to which they are attached[165]. Lys63-linkages have been found to be implicated in several pathways including inflammatory response, regulation of protein synthesis and protein localisation[166–168] whereas Lys11- and Lys29-linked are involved in lysosomal degradation[169]. Other topologies of polyubiquitin chains have been detected but their role is not yet well understood.

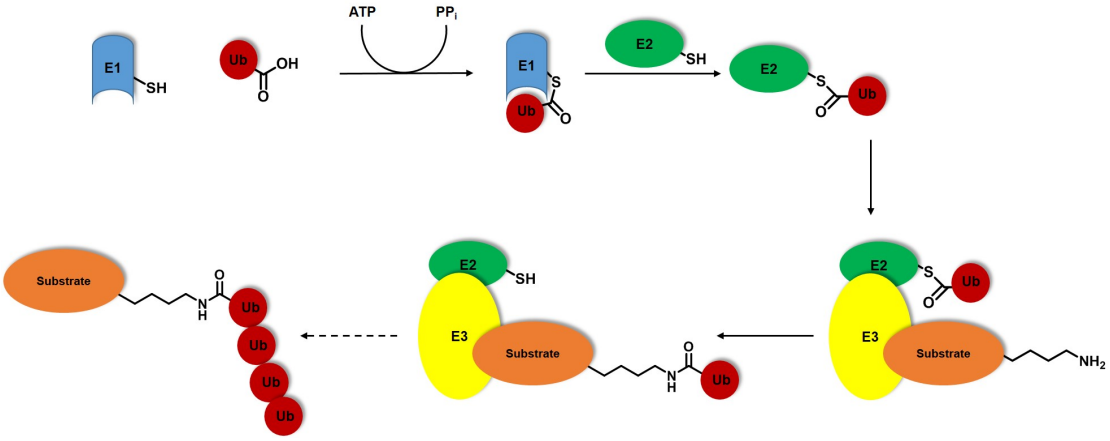


Figure 4.1: Overview of the ubiquitylation reaction mechanism. The ubiquitylation pathway starts with the charging of the E1 enzyme through the hydrolysis of an ATP molecule. A thioester bond is formed between the ubiquitin C-terminus and the E1 active site cysteine. A transthioesterification reaction results in the transfer of the ubiquitin molecule to the E2 active site cysteine. An E3 ligase finally catalyses transfer of the ubiquitin from the active site cysteine of the E2, to a primary amine on a lysine side chain (or protein N-terminus).

4.1.1 USP5

Ubiquitylation, as any other post-translational modification, is reversible. The removal of ubiquitin molecules from modified proteins is carried out by a family of enzymes collectively named deubiquitinases (DUBs)[170]. Enzymes classified as DUBs are responsible for the regulation of the ubiquitin system via hydrolysis of the isopeptide bond between target proteins and ubiquitin molecules and between ubiquitin moieties linked in polyubiquitin chains. Five families of DUBs have been identified with the largest one being the ubiquitin-specific protease (USP) family. A significant USP family member is ubiquitin-specific protease 5 (USP5). USP5, also known as Isopeptidase T or IsoT, is a multidomain protease specialised in the degradation of unanchored (substrate-free) polyubiquitin chains[171], which have been proved to have a regulatory role in several physiological processes such as the regulation of 26S proteasome activity[172] and the NF-κB signalling pathway[173]. Recognition of ubiquitin occurs through non-covalent interactions between the protein and ubiquitin-binding domains (UBDs) of Ub receptor proteins and these motifs are able to interact with different areas of the Ub surface. The first Ub binding domain described is the ubiquitin-associated domain (UBA) which is characterised by a three-helix bundle structure and is known to interact with the Ile-44 centred hydrophobic patch of ubiquitin through a conserved (M/L)G(F/Y) motif. Such a domain is found in many proteins involved in the ubiquitylation pathway including USP5.

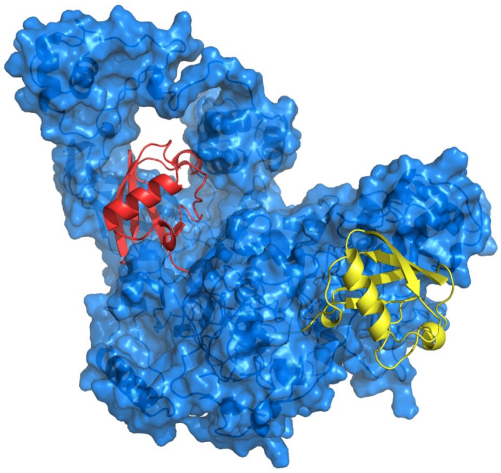


Figure 4.2: Model of the complex between USP5 (blue) and monoubiquitin bound at the catalytic site (red) and the ZnF UBP domain (yellow). Rendered in pymol from PDBs 3IHP and 2G45

The crystal structure of USP5 has been solved recently[174] and it revealed the presence of a catalytic domain and four additional folded domains, two Zinc finger ubiquitin binding domains (ZnF UBP) and two UBAs, connected by flexible loops (see Fig. 4.2). One ZnF-UBP domain has been shown to be responsible for the recognition of the free C-terminus of polyubiquitin chains, making the enzyme selective for the unanchored versions of the chains[175]. The other ZnF-UBP domain is cryptic, but appears to be important for the recognition of ubiquitin at the catalytic site, although it does not display affinity for ubiquitin. Using a combination of mutagenesis experiments and isothermal titration calorimetry (ITC) measurements, the two UBAs have been implicated in polyubiquitin binding only when the chain possesses three or more units of Ub, suggesting that the domains interact with the third and fourth ubiquitin moiety in a polyubiquitin chain[171]. This hypothesis is corroborated by their position revealed in the crystal structure which places them in a distal position to the C-terminal ZnF UBP domain. The catalytic domain interacts extensively with ubiquitin as shown in the X-ray crystal structure of USP5 covalently attached to a ubiquitin-based suicide inhibitor (ubiquitin-bromoethylamine). Such crystal structure raises uncertainties concerning the arrangement of the domains in respect to the predicted model of binding of polyubiquitin chains, in particular the ZnF UBP pointing away from the active site

preventing polyubiquitin from interacting with both regions simultaneously. Such arrangement would disfavour the avidity effect that underlie polyubiquitin recognition by UBD and make substrate binding incompatible with the mode proposed in previous literature[176]. The ZnF UBP is connected to the core of the protein through flexible loop regions therefore there is the potential for a large conformational shift occurring upon substrate binding in order to place the isopeptide bond connecting the first and second ubiquitin moiety in a polyubiquitin chain in the correct orientation for enzymatic cleavage (see Fig. 4.3). AD-based carbene footprinting seemed to be the appropriate tool to elucidate the substrate binding sites of this large multi-domain protease and clarify the questions concerning the X-ray crystal structure.

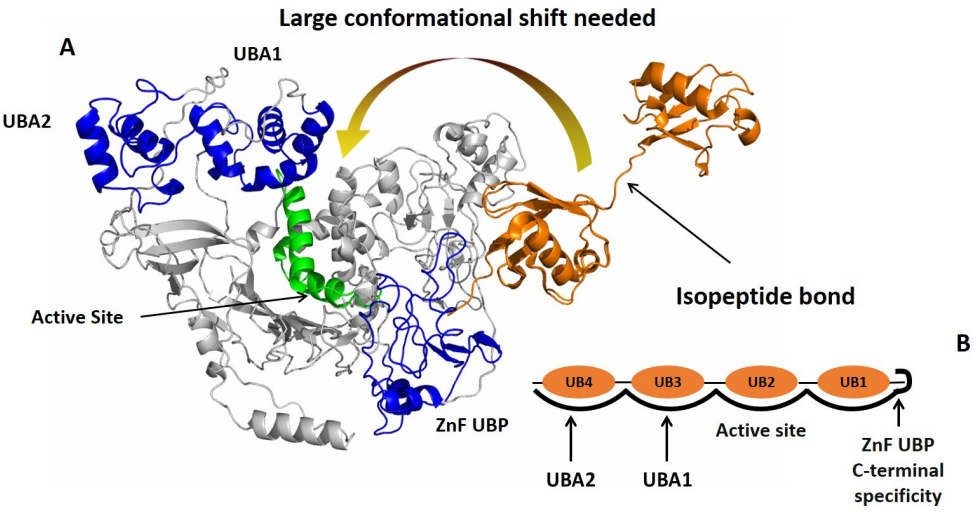


Figure 4.3: A) Model of USP5 (based on PDB 3IHP) with di-ubiquitin (orange, PDB 2W9N) bound. The ubiquitin binding domains are shown in blue, and the catalytic domain in green. The motion required for the ZnF UBP-bound substrate to approach the catalytic site is indicated by an arrow. B) Current model of binding of USP5 to polyubiquitin chains.

4.2 Results and discussion

4.2.1 Structural integrity studies

As discussed in Chapter 3, one of the main concerns when studying protein structure using a covalent label is the possible distortion of the native fold induced by the pres-

ence or the reactivity of the labelling reagent. The presence of the reagent at millimolar concentration before photolysis can potentially induce alterations in the tertiary structure and subsequently lead to errors in the analysis. To demonstrate that the protein retains its native fold in the presence of the labelling reagent an activity assay was first performed. As mentioned previously the biological role of USP5 is to disassemble unanchored poly-ubiquitin chains. To perform this task the protein must exist in its native,active three-dimensional structure. To test the activity, and therefore the integrity of the native fold, USP5 was incubated at 37 °C for 60 minutes with a Lys48 poly-ubiquitin ladder including mono-, di-, tri- and tetra-ubiquitin in the presence of the diazirine (10 mM). The reaction was quenched after 5, 15, 30 and 60 minutes. Samples were resolved by SDS-PAGE, Western blotted and visualised with anti-ubiquitin. In order to obtain more quantitative information on the extent of the reaction at each time-point, densitometric measurements were performed on the bands visualised through Western blotting (Fig. 4.4 C).

As shown in Fig. 4.4B, it is clear that bands corresponding to tetra- and tri-ubiquitin that are visible in the ubiquitin load lane disappear progressively at the same rate in the presence and in the absence of the reagent. The band visible on the bottom corresponds to mono-ubiquitin which is the product of the complete disassembly of poly-ubiquitin chains. Bands corresponding to tetra-ubiquitin are also visible in the 60 minutes lanes, implicating partial hydrolysis of that particular chain length, but the presence of the band in both USP5 in presence and absence of AD suggests that the issue is not related to the presence of the labelling reagent and that the reaction could be taken to completion over a longer time course.

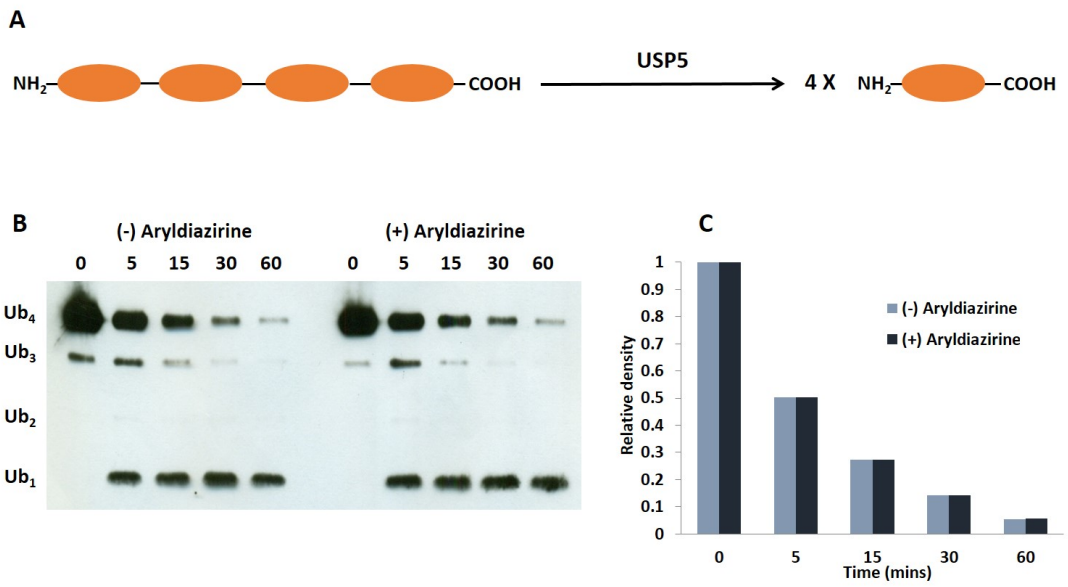


Figure 4.4: A) Hydrolysis of tetraubiquitin chains catalysed by USP5. B) Western blot of tetraubiquitin deubiquitination by WT USP5 over the time course, in the presence and absence of aryldiazirine. C) Densitometric analysis of tetraubiquitin deubiquitination over the time course.

The indistinguishable bars obtained plotting the densitometry measurements (Fig. 4.4 C) also confirm the total absence of reactivity alteration induced by the presence of the reagent. The results prove that in the presence of the diazirine the enzyme is still active and therefore it is safe to assume that the protein exist in solution in its native fold. This finding is significant, as USP5 is a less stable protein than lysozyme, confirming the benignity of the reagent.

4.2.2 Peptide level analysis

4.2.2.1 Proteolytic Digestion

USP5, at ∼100 kDa, is a much larger protein than lysozyme and, as mentioned in the previous section, it contains different domains able to interact with its binding partner. For these reasons the choice of a suitable protease is vital to get the most complete information from the differential footprinting study. A series of protease including trypsin, AspN and GluC were tested in order to assess which protease yielded the best peptide

coverage for the key interaction areas of the enzyme. The digestion products obtained with each protease were analysed by LC-MS/MS operating in data-dependant mode in which each full MS scan was followed by four MS/MS scans where the four most abundant ions were selected for collision-induced dissociation. The spectra acquired were searched using Bioworks Browser as described in Chapter 2. Trypsin proved to be the most suitable protease for the experiment yielding a good number of peptides among which some involved in the formation of the ZnF UBP domain and the catalytic domain. AspN and GluC digests were found not useful for the purpose of the experiment because they yielded a significantly smaller amount of identified peptides and all of them were already covered by the peptides obtained by trypsin digestion. The coverage obtained resulted slightly over 51% and it is showed in Fig. 4.5; although not comprehensive, the detection and identification of peptides corresponding to areas involved in binding was considered sufficient for the purpose of mapping the interaction between the enzyme and its substrate.



Figure 4.5: USP5 tryptic digestion sequence coverage. Identified peptides are indicated by a red arrow

4.2.2.2 USP5-Diubiquitin Footprinting

In order to investigate the binding of diubiquitin to USP5 a non-catalytic mutant of the enzyme in which the catalytic cysteine was mutated to an alanine residue (C335A-USP5) was expressed and purified by Dr. Daniel Scott. This active site mutant is able to form stable complexes with the substrate without hydrolysing its isopeptide bond. C335A-USP5 was labelled with AD in its unbound version and in the presence of one equivalent of K-48-linked diubiquitin. It has been shown in previous literature that the dissociation constant (K_d) of the complex is ≤ 2 nM[171] therefore it is safe to assume that in the conditions in which the labelling experiment is performed the large majority of C335-USP5 is bound to diubiquitin. Samples containing labelled bound and unbound USP5 were firstly analysed by SDS-PAGE, with the dual purpose of establishing that the enzyme was not undergoing any significant cleavage, and of removing all the by-products of the photolysis reaction. Bands corresponding to USP5 were then excised, reduced, alkylated, treated with trypsin and the digests obtained were analysed by nanoLC-MS. Firstly data-dependent acquisition (DDA) was used to identify labelled and unlabelled peptides. The four most intense ions in each surveyor scan were isolated and fragmented using CID and the data obtained was searched against a custom-made database including the sequence of USP5 as indicated in Figure 4.5. All the peptides previously detected were identified in both their labelled and unlabelled forms, excluding E208-R223 and L589-R605 which were only detected in their unlabelled form. Unfortunately, location of the modification was not achievable by database searching since carbenes are reactive towards a wide variety of moieties, leading to only a partial chromatographic separation of each isomer making the software unable to assign the modification to a specific residue. The fractional modification of each identified peptide was assessed using the same procedure adopted for lysozyme (see Chapter 2). Briefly, the fraction modified was calculated as the ratio between the peak area of the extracted chromatogram of each labelled peptide over the sum of the peak areas of the unlabelled and labelled peptide. The values corresponding to peptides derived

from labelling of the enzyme in its bound and unbound state were compared. Being a protein-protein interaction involving a considerably large surface area, it was hypothesised that the peptide level analysis of the USP5-diubiquitin complex could lead to informative results. Five peptides originated from the digestion of labelled USP5 in complex with diubiquitin showed a distinct and reproducible decrease in fractional modification (p-value from Student t-test < 0.05, shown in Fig. 4.6) suggesting that the binding event lead to a decrease in the accessibility of certain areas of USP5.

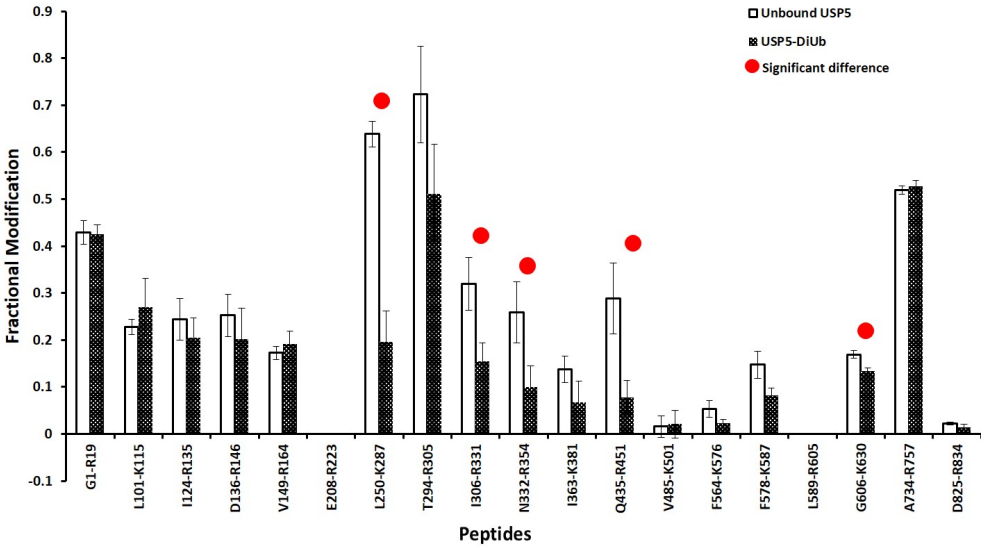


Figure 4.6: Fractional modification (by AD) of USP5 peptides in the presence (black bars) and absence (white bars) of di-ubiquitin. Error bars are ± SD (n = 3) and significant differences (Student t-test, p < 0.05) are highlighted with a red circle

The peptide that exhibited the largest difference in fractional modification ($\Delta_{frac.mod.}$) was L250-K287 ($\Delta_{frac.mod.}$=0.43), a peptide corresponding to a large area of the ZnF UBP domain. As previously alluded to, the ZnF UBP domain is responsible for the recognition of the proximal ubiquitin moiety in poly-ubiquitin chains through their C-terminus. L250-K287 includes an area of the domain involved in the formation of the binding pocket, therefore the observed decrease in labelling can be attributed to the diminished accessibility of the binding site caused by the presence of diubiquitin bound to USP5 through its di-glycine terminal motif. Peptides N332-R354 and Q435-R451 are involved in the formation of the catalytic domain of the enzyme with the actual cat-

alytic residue being part of N332-R354. The decrease in labelling observed for these two peptides ($\Delta_{frac.mod.}$=0.16 and $\Delta_{frac.mod.}$=0.20 for N332-R354 and Q435-R451 respectively) confirms the occupation of the site by the second ubiquitin moiety. I306-R331 is a peptide localised between the aforementioned two domains, an evidence that reinforces the suggested model of binding across the ZnF UBP and the catalytic domain. The fifth peptide that showed a small but significant decreased labelling ($\Delta_{frac.mod.}$=0.03) is G606-K630 which corresponds to a flexible loop distant from the domains involved in substrate binding. Given its flexibility, the diminished accessibility could well be related to the large conformational shift that the enzyme must undergo upon binding of poly-ubiquitin chains. No significant differences in labelling were observed on peptides involved in the formation of the two UBAs, confirming the lack of interaction of these ubiquitin-binding domains with polyubiquitin chains with less than three units of Ub.

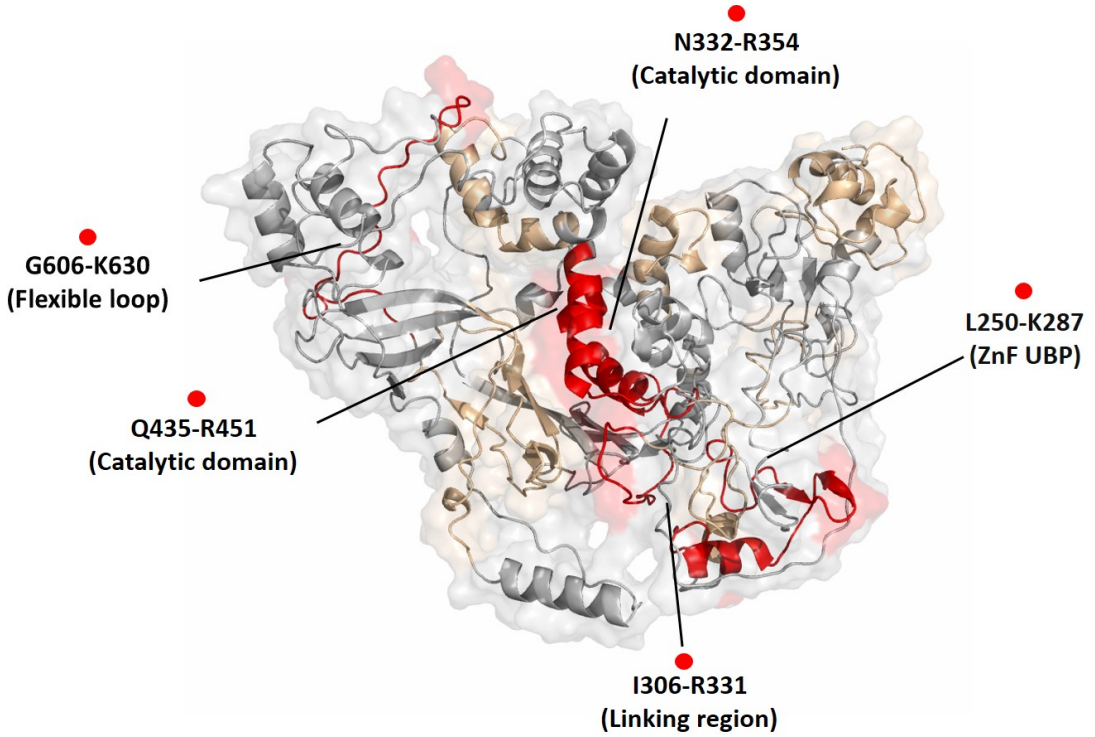


Figure 4.7: Model of USP5 (based on PDB 3IHP) showing the locations of the five peptides (red) masked from labeling by di-ubiquitin binding, and their correspondence to the different areas involved. Wheat and gray colored areas correspond respectively to peptides showing no significant difference in the extent of labelling and undetected peptides

As discussed above, the X-ray crystal structure available seems inconsistent with the previous studies on USP5-polyubiquitin complexes and critically USP5 function. The masking detected on peptides found on both the ZnF-UBP domain and catalytic domain of USP5 upon addition of K48-linked di-ubiquitin provide evidence for an alternative conformation of USP5 where the ZnF-UBP and the catalytic domain are close in space in order to allow the binding of diubiquitin at both sites (see Fig. 4.8). The existence of two distinct conformations of USP5 was previously demonstrated by MS-based techniques, specifically CSD, IM-MS[176], and SAXS[174]. These studies showed that free USP5 in solution exists as a mixture of a compact and a more extended conformation and adopts the compact fold when bound to unanchored polyubiquitin chains. The results altogether strengthen the theory asserting that the conformational flexibility of USP5 is vital for its ability to hydrolyse various topologies of unanchored polyubiquitin chains.

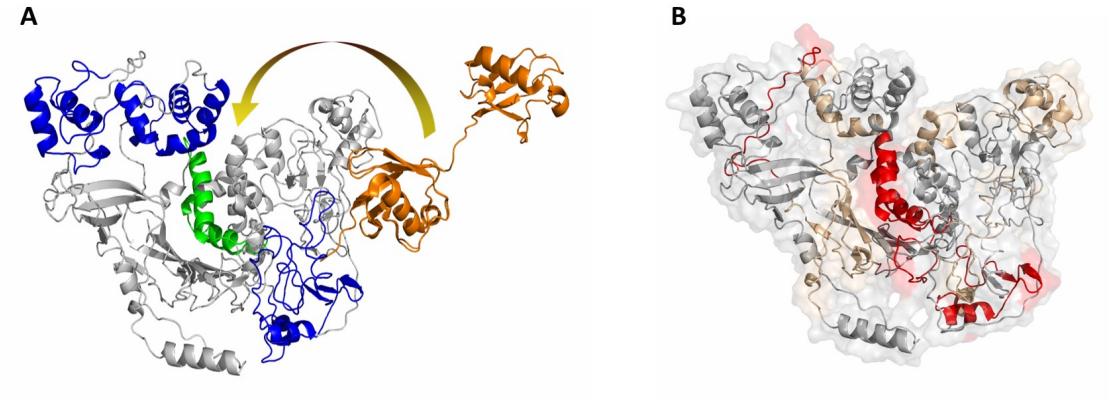


Figure 4.8: A)Model of USP5 (based on PDB 3IHP) with di-ubiquitin (orange, PDB 2W9N) bound. The ubiquitin binding domains are shown in blue, and the catalytic domain in green. The motion required for the ZnF UBP-bound substrate to approach the catalytic site is indicated by an arrow. B) Model of USP5 (based on PDB 3IHP) showing the locations of the five peptides (red) masked from labeling by di-ubiquitin binding, and their correspondence to the ZnF-UBP and catalytic domains.

4.2.3 Residue level analysis

Following the interesting results obtained at the peptide level, MS/MS experiments were performed on the peptides that showed significant masking in order to obtain

higher resolution information and gain insights into the residues involved in the USP5-diubiquitin interaction. Peptide L250-K287 corresponds to a large section of the ZnF UBP domain and it exhibited the highest decrease in labelling. The peptide derived from labelling experiments on the unbound and bound protein was isolated and fragmented by CID using the same conditions described in Chapter 3 for the per-residue analysis on lysozyme and the data obtained was processed using the same approach. The calculated per-residue modification obtained for both unbound and bound protein is shown on the histogram in Figure 4.9.

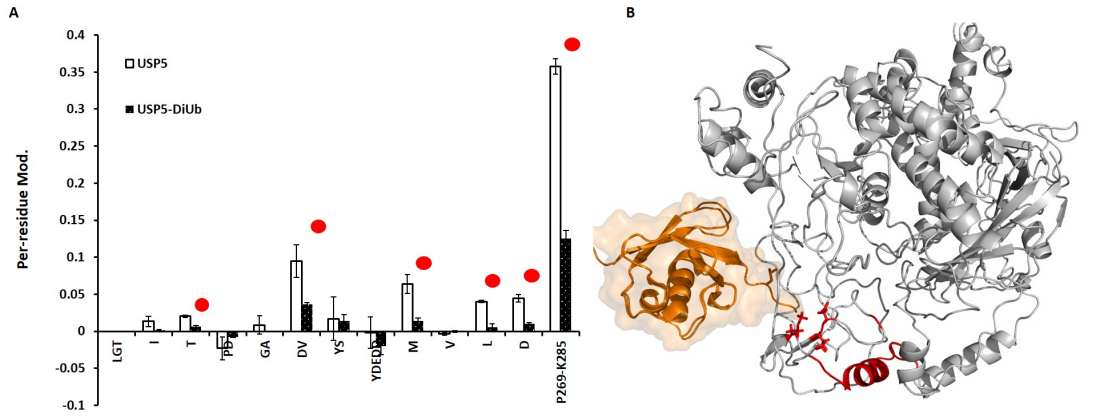


Figure 4.9: A) L250-K287 per residue modification in the presence (black bars) and absence (white bars) of di-ubiquitin. Error bars are ± SD (n=3) and significant differences (student t-test,$p<0.05$) are highlighted with a red circle. B) Model of USP5 (based on PDB 3IHP) with ubiquitin (orange,PDB 2G45) bound at the ZnF-UBP. Residues from L250-K287 showing masking from labelling by di-ubiquitin binding are coloured in red

Unfortunately CID fragmentation did not yield good coverage over the peptide sequence, especially towards the C-terminus, and that caused a loss of resolution on the structural information obtained. Some of the residues that showed significant masking are located within the binding pocket and in close contact with the C-terminus of ubiquitin in the complex (as shown in Fig 4.9). The decrease in labelling on these residues can be related to the direct masking of the binding site caused by the presence of diubiquitin. The largest variation was observed on the area spanning from P269 to K285 but unfortunately no fragments were obtained within this part of the peptide. The peptide is preferentially cleaved at the proline residue therefore it was not possible

to pinpoint single amino acid labelling variations for the area. Based on the observation of the crystal structure of the isolated ZnF UBP in complex with ubiquitin it is not possible to relate the clear and reproducible masking observed on the residues between L277 and the C-terminus with a direct interaction with the ligand. The decrease in labelling is probably due to a large structural rearrangement of the protein that reduces the accessibility of some of the residues of the area. This hypothesis is consistent with the conformational shift associated with the placement of the isopeptide bond of polyubiquitin chains on the catalytic domain.

Peptide Q435-R451 covers large part of the catalytic domain. Apart from the catalytic C336, no detailed information is available on which residue of the domain is vital for the interaction with its substrate. The crystal structure of the protein has been solved in the presence of ubiquitin modified to act as a suicide inhibitor and it shows that the ligand interacts with the catalytic domain over a large area. This information may not be completely accurate given the assumption that the crystal structure does not represent the active conformation of the enzyme. This hypothesis is strengthened by the fact that in the crystal structure the ubiquitin moiety also seems to interact with one of the UBA domains, which is in contrast with the evidence pointing to the interaction of the two UBAs only with the third and fourth ubiquitin molecule in a polyubiquitin chain and not the one interacting with the catalytic domain. Figure 4.10 shows the per-residue modification observed on peptide Q435-R451 in the bound and unbound state.

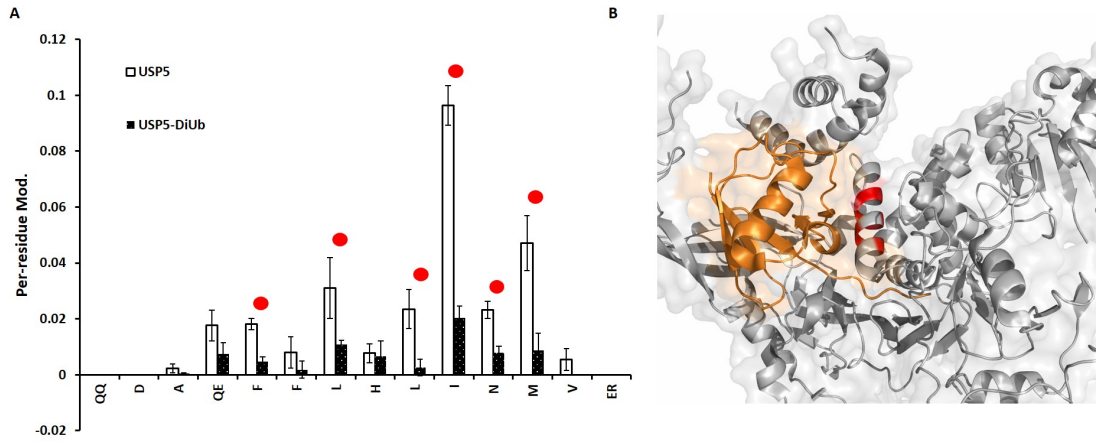


Figure 4.10: A) Q435-R451 per residue modification in the presence (black bars) and absence (white bars) of di-ubiquitin. Error bars are ± SD (n=3) and significant differences (student t-test,$p<0.05$) are highlighted with a red circle. B) Model of USP5 (based on PDB 3IHP) with ubiquitin covalently bound to the active site. Residues from Q435-R451 showing masking from labelling by di-ubiquitin binding are coloured in red

Significant masking was detected on several residues with the highest difference observed on I445 followed by M447,L444, L442, N446 and F440. All these residues are localised on a α-helix which, according to the crystal structure of USP5 covalently bound to ubiquitin, is in close contact to the substrate. The orientation of some of the side chains seems slightly inconsistent with the position of the ubiquitin molecule but, as discussed above, the crystal structure is not an accurate representation of the active conformation of the complex. The absence of difference in the labelling extent on residues F441 and H442 seem to point to a conformation in which the helix is rotated clockwise by approximately 90 degrees placing all the residues showing significant differences in direct contact with the substrate and F441 and H442 away from the interacting surface.

I306-R331 corresponds to a flexible loop located between the ZnF-UBP domain and the catalytic domain. The per-residue analysis narrowed down the masking to a few residues but given the flexibility of the area it is difficult to predict the location of such amino acids in the active conformation of the enzyme but it is probably safe to assume that they are involved in the conformational shift associated with ubiquitin binding (see Fig. 4.11).

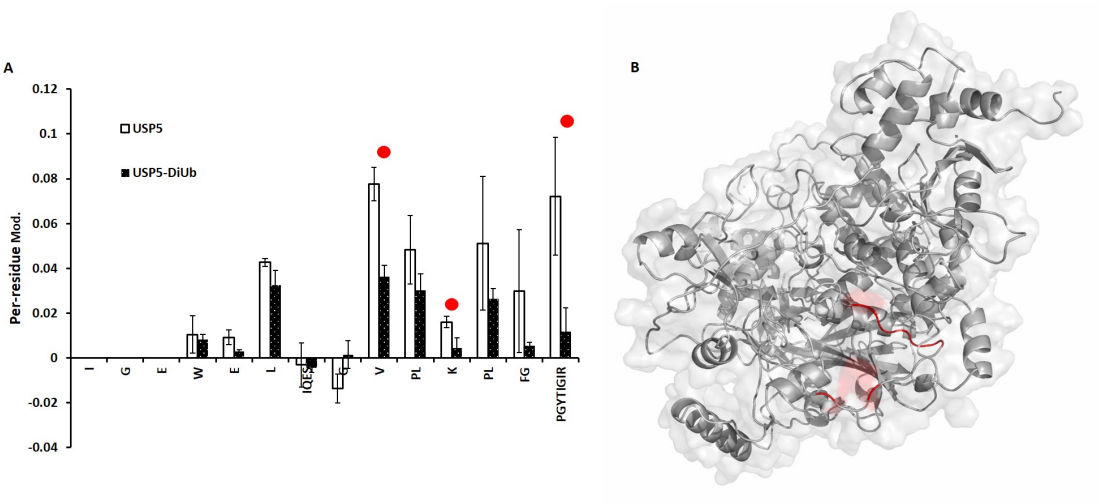


Figure 4.11: A)I306-R331 per residue modification in the presence (black bars) and absence (white bars) of di-ubiquitin. Error bars are ± SD (n=3) and significant differences (student t-test,$p<0.05$) are highlighted with a red circle. B) Model of USP5 (based on PDB 3IHP). Residues from I306-R331 showing masking from labelling by di-ubiquitin binding are coloured in red

In the native protein the catalytic C335 is part of peptide N332-R354 but as mentioned before, for this study the residue was mutated to an alanine residue to prevent the hydrolysis of diubiquitin before the footprinting experiment. Given the presence of the "pseudo-catalytic" residue, more detailed structural information about how the area corresponding to such peptide interacts with the substrate would have been valuable but unfortunately CID fragmentation yielded very few ions making the location and quantitation of the label unreliable.

As mentioned above USP5 presents two other ubiquitin binding domains which are expected to interact with poly-ubiquitin chains composed of three units (or longer) and contribute to the avidity effect that the enzyme shows when in complex with longer length chains. The crystal structure shows that one of the UBAs makes large contact with the ubiquitin attached to the catalytic domain but this interaction would be in contrast with previous studies revealing that the UBAs interact only with the third and fourth ubiquitin moiety of polyubiquitin chains. Peptide A734-R757 covers a large portion of the UBA domain that in the crystal structure seems to be interacting with ubiquitin but it did not show any difference in labelling at the peptide level. In order to further confirm that the UBA is not involved in any interaction with diubiquitin A734-

R757 was fragmented and the per-residue data between bound and unbound protein was compared. As shown in Figure 4.12 no clear masking is observed on any residue confirming that the UBA is not involved in the interaction as previously suggested.

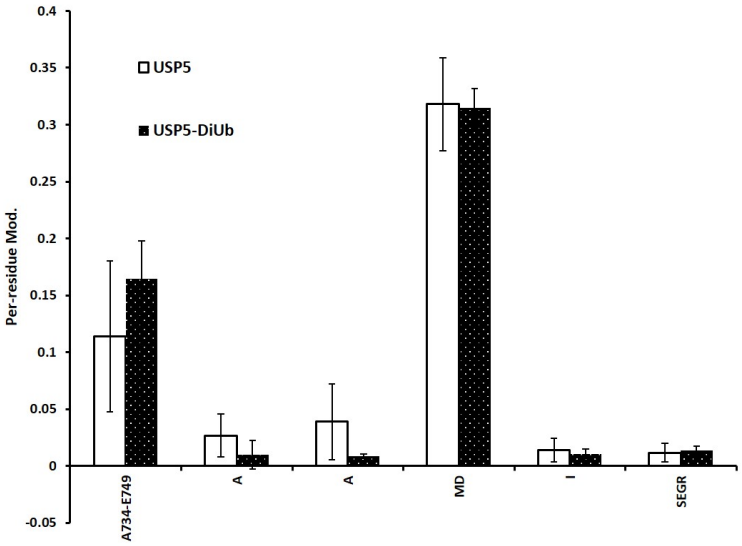


Figure 4.12: A734-R757 per residue modification in the presence (black bars) and absence (white bars) of di-ubiquitin. Error bars are ± SD (n=3), no significant differences observed.

4.3 Conclusions

In summary, the use of the carbene footprinting strategy proved successful in mapping the contact surface of the USP5-diubiquitin interaction, a partially characterised large protein-protein complex.

The novel aryldiazirine was confirmed to be benign towards the native fold of the protein. The activity of the enzyme was tested in the presence of the reagent and was found to be unaffected leading to infer that the structure of the protein is not perturbed by the aryldiazirine.

The decrease in labelling observed in the presence of diubiquitin on peptides corresponding to parts of the ZnF-UBP domain and the catalytic domain pointed to the existence of a distinct active conformation of the enzyme in which the two domains

are in spatial proximity allowing the substrate to be bound at both sites. This result is corroborated by previous IM-MS and SAXS studies which suggested the existence of two distinct conformational states. These two conformations have been shown to correspond to an extended and a more compact conformation, with only the latter being able to interact with unanchored polyubiquitin chains.

Carbene protein footprinting provided novel information about the structure and mechanism of USP5 showing how the methodology can be useful to gain insights into large complex protein-protein systems. The results pave the way to characterise biologically significant DUBs interactions and shed light on the mechanisms underlying the diseases implicated in their misfunction.

CHAPTER 5

Mapping the contact surface across a KS/DH split module in the Gladiolin trans-AT Polyketide Synthase.

5.1 Introduction

5.1.1 Polyketides

Polyketides constitute a class of natural products known for its wide range of pharmaceutical activities including antibiotic (Erythromycin A), anti-tumour (Geldanamycin), immunosuppressant (FK-506) and antifungal (Amphotericin B). Although these products are available from natural sources, obtaining sufficient amounts to test their biological activity can be challenging[177]. For this reason, amongst others, the biosynthetic mechanisms underlying the production of polyketides are subject of extensive research aimed to the exploitation of polyketide-producing enzymes to produce novel bioactive compounds[178,179].

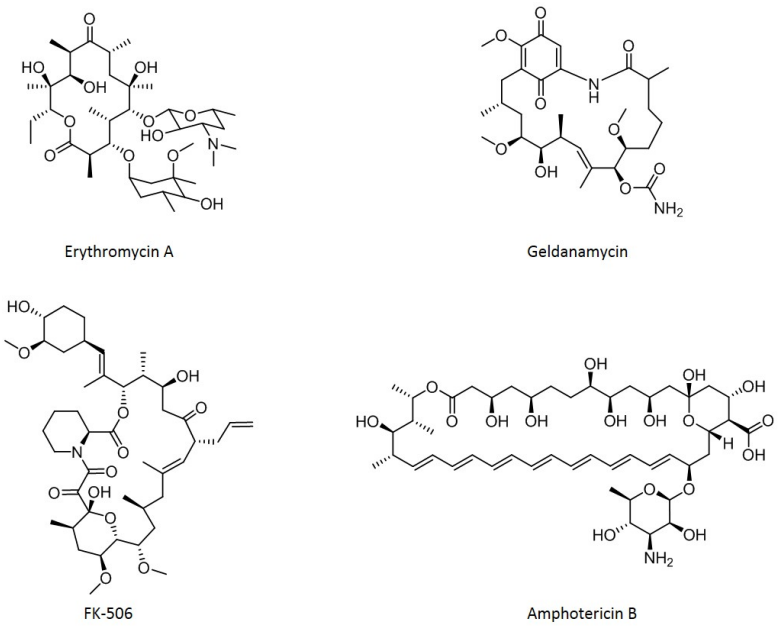


Figure 5.1: Examples of polyketides currently used as medicinal products

5.1.2 Polyketide Synthases

Polyketides are bio-synthesised by large protein complexes collectively known as polyketide synthases (PKSs). These enzymes use simple acyl-CoAs to produce a very diversified range of compounds and follow a biosynthetic logic similar to that of fatty acid synthases (FASs)[180,181]. In both PKSs and FASs the growing acyl chain is tethered to a phosphopantetheine (PPant) arm, in the form of a thioester. The PPant prosthetic group is post-translationally attached to a conserved Ser residue on the ACP, and provides a flexible, dynamic arm to allow diverse chain modifications by multiple domains to take place[182,183]. In order to perform a single round of chain elongation a minimal set of enzymatic domains composed of an ACP, an acyltransferase (AT) and a ketosynthase (KS) is required. This nominal set of domains, in addition to optional reductive domains is referred to as a module. The biosynthesis is started by the AT domain, which is responsible for the transfer of a malonyl-derived unit from a CoA-thioester to the phosphopantetheinyl chain of the ACP. The KS domain is then responsible for the formation of a carbon-carbon bond through a Claisen-like condensation mechanism

between the growing acyl chain and the malonyl-derived extender unit, resulting in a β-keto, elongated product. The presence of optional domains such as ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) directs further modification at the β-keto position leading to the formation of a β-hydroxyl, unsaturated or fully saturated intermediate[183,184] (see Fig. 5.2).

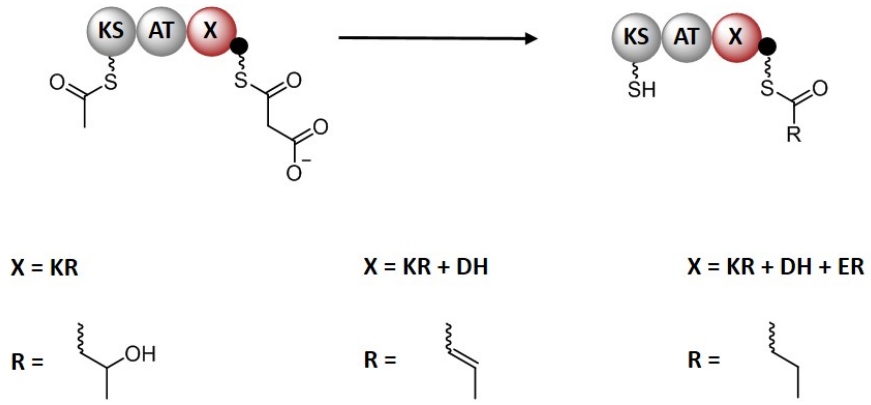


Figure 5.2: The KS domain catalyses the elongation of the growing chain (an acetyl unit in the case shown) through a Claisen-like condensation with a malonyl unit tethered to an ACP. The X represents optional reductive domains able to modify the elongated chain at the b-keto position.

5.1.3 Modular polyketide synthases

PKSs can be classified in three different main groups, type I,II and III and further subdivided on the basis of their architecture[185,186]. For the purpose of this work, the attention will be focused only on modular type I PKSs. This type of enzymes can be considered similar to a molecular assembly line, in which after a round of elongation and eventual optional modification, the growing product is passed from the ACP to the KS of the following downstream module. The chain is transferred from module to module until the final product is complete and released through the action of a thioesterase domain[185]. After its release, the compound is often further modified by additional tailoring enzymes to produce the final bioactive molecule. Usually type I PKSs are composed of more than one polypeptide chain, assembled through non-covalent interactions resulting in complexes with a molecular mass that often exceeds one MDa. Modular type I PKSs can also be grouped into two distinct classes according to whether acyltrans-

ferases domains are integrated into polypeptide chains of the enzymes, *cis*-AT PKSs, or are stand-alone proteins, *trans*-AT PKSs[187,188]. The ordered self-assembly of *cis*-AT PKSs is relatively well characterised and it has been shown that the different polypeptide chains interact through compatible docking domains located at their N- and C-termini. In this type of PKSs the different proteins are disconnected at the end of a given module with the following polypeptide chain presenting the downstream KS domain at its N-terminus[189–191]. A well characterised example of this type of PKS is the erythromycin-producing 6-deoxyerythronolide B synthase (DEBS)[192]. In comparison trans-AT PKSs tend to place their pathway disconnections in the middle of modules, giving rise to a range of split module junctions[193] (e.g. KS/KR, DH/KR, KS/DH, see Fig. 5.3).

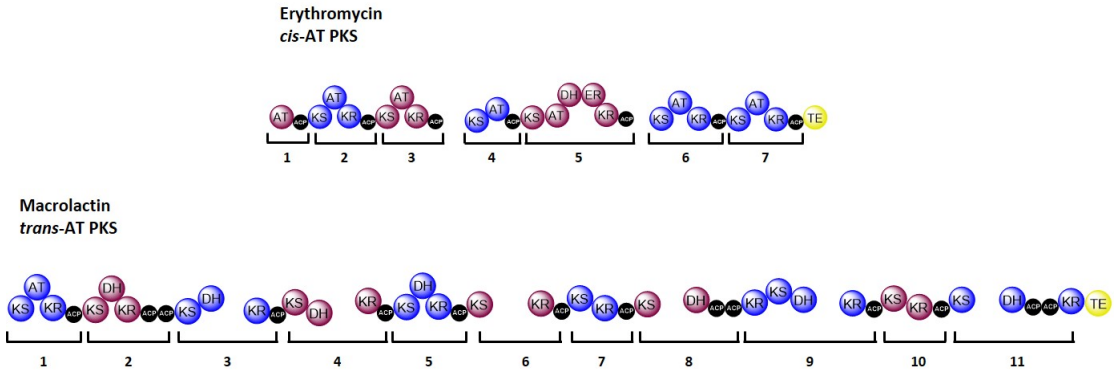


Figure 5.3: Six elongating modules of the cis-AT erythromycin PKS spanning over three polypeptide chains (top). The proteins are disconnected end of modules 2 and 4. The Macrolactin trans-AT PKS is composed by eleven modules over seven proteins (bottom). The disconnections between different polypeptide chains occur within a module giving rise to "split modules".

The assembly of trans-AT PKSs has been shown recently to be directed by pairs of docking domains at the boundaries between ACP/KS, DH/KR and KS/KR in virginiamycin and macrolactin PKSs[194,195]. These N- and C- terminus regions were found to be folded in four-helix bundles upon assembly, and their binding affinity was calculated to be between 0.8 and 10 μM. Another common protein-protein boundary in trans-AT PKSs occurs across a non-elongating KS domain and a DH domain (KS/DH) but the dehydratase does not possess a N-terminal appendage and only a highly disordered region is observed C-terminal of the KS^0 (docking peptide ^{DH}DP)[195]. In these

pathway disconnections the KS domains are usually non-elongating (KS^0), which lack one of the histidines necessary for their decarboxylative catalytic activity. These domains are usually located upstream of a dehydratase domain which introduces a *cis* double bond in the growing product[188].

Using the carbene-based footprinting methodology described in chapter 3, the spatial arrangement of the domains involved in the GbnD4 KS14 and GbnD5 DH split module junction from gladiolin PKS was investigated and is described in this chapter (cartoon representation of the modules and structure of the final product shown in Fig. 5.4). Furthermore, the interaction between the dehydratase domain and its downstream ACP is characterised. The interacting surfaces were successfully mapped revealing two distinct binding areas on the DH for the ACP and the disordered C-terminal peptide of the KS, allowing both domains to bind at the same time. The structural characterisation of the interaction may help future efforts by synthetic biologists to exploit the chemistry of trans-AT PKSs in order to produce novel bioactive compounds.

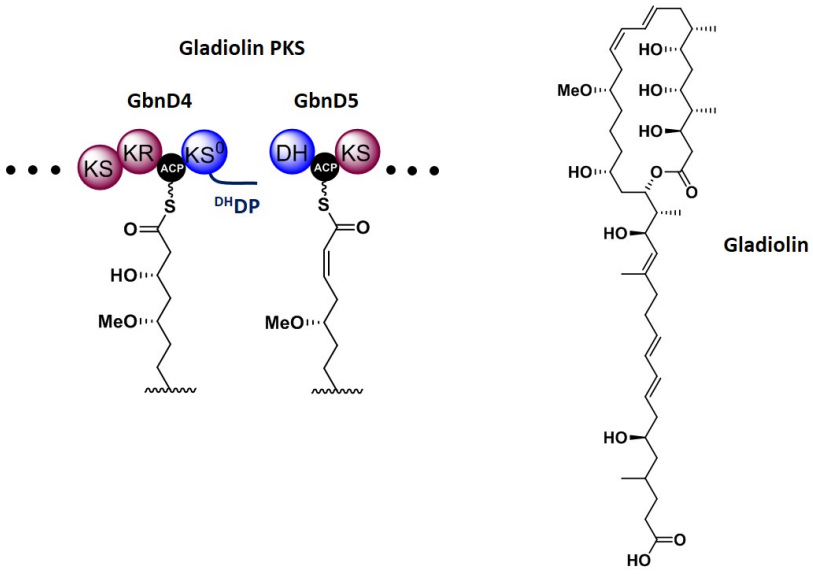


Figure 5.4: Dehydration step catalysed by the dehydratase domain on protein GbnD5 of gladiolin PKS at the boundary with the KS^0 domain on protein GbnD4 introducing a cis double bond at the β-keto position (left). The chemical structure of gladiolin is shown on the right.

5.2 Results and discussion

5.2.1 Bioinformatic identification of ^{DH}DP

The recently discovered gladiolin *trans*-AT PKS presents two modules disconnected at the KS/DH interface, and was therefore used to investigate the contact surface involved in the common interaction mediated by KS/DH boundary. As mentioned in the introduction section, module boundaries in *cis*-AT PKSs and some split module interfaces in trans-AT PKSs have been found to be directed by N- and C- terminal pairs of docking domains. In the case of the KS/DH split module junction however, there is no N-terminal domain appended to the N-terminus of the DH to promote an interaction. The C-terminus of the polypeptide chain corresponding to the upstream domain in the split junction presents an unannotated peptide extending from the terminal KS domain. Using hidden Markov model profiling (pHMM) this type of KS(^{DH}DP)/DH boundary was found in several trans-AT PKSs, however the sequence alignment of the various examples of the docking region were found to lack conservation (see Fig. 5.5A) The regions were predicted to be highly disordered (PsiPred and IUpred servers predictions) and were found to have potential to be involved in a non-covalent interaction (ANCHOR server, Fig. 5.5 B). All bioinformatics studies were conducted by Dr. Matthew Jenner and Dr. Simone Kosol (University of Warwick).

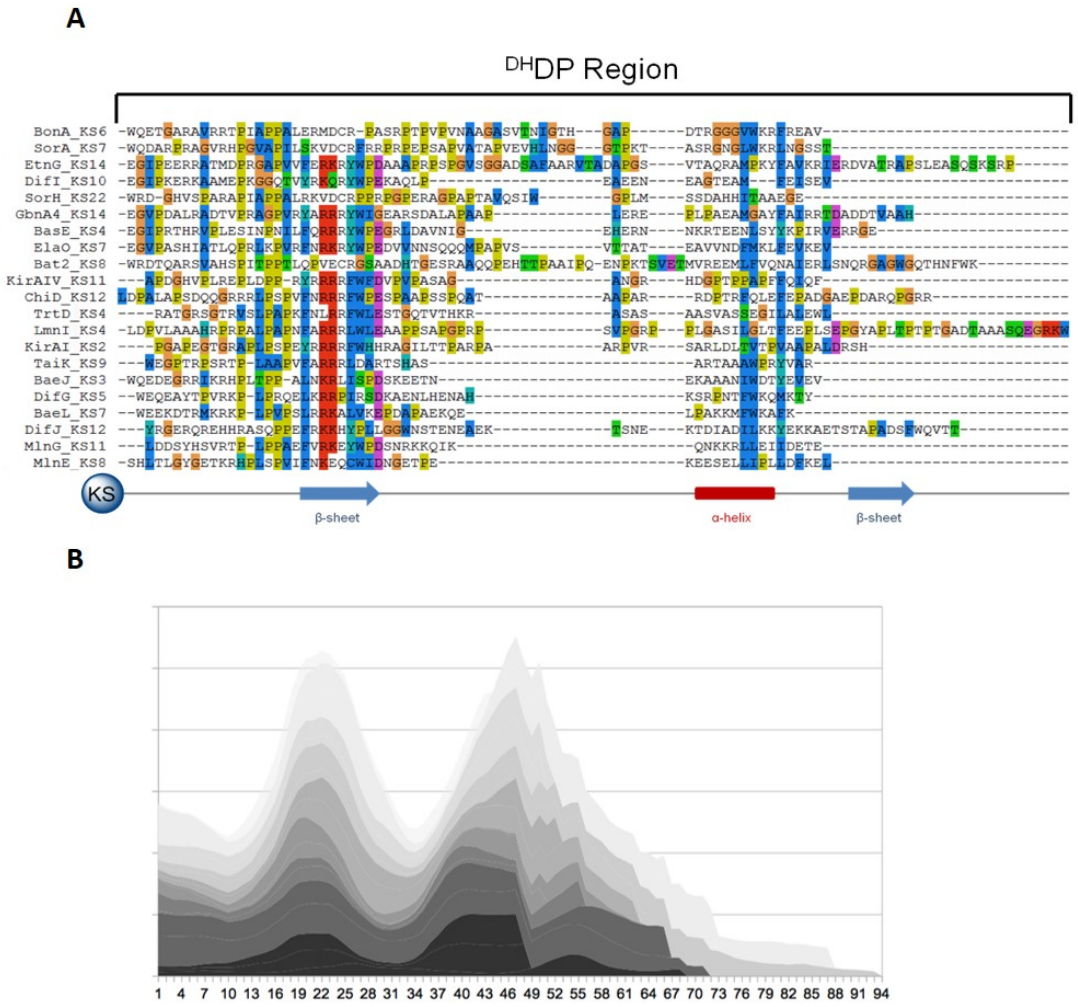


Figure 5.5: A) A ClustalX multiple sequence alignment of all ^{DH}DP regions identified by the pHMM and predicted secondary structure elements. The alignment highlights localized conservation towards the N-terminus, but relatively low sequence conservation towards the C-terminus. Binding site predictions conducted on all ^{DH}DP regions suggest two interaction sites situated around the structured regions.
B) A stacked plot of the added probabilities of binding sites calculated with ANCHOR for all eighteen identified KS(^{DH}DP)/DH regions reveals a distinct two site pattern that is present in most of the ^{DH}DPs. Produced by Dr. Matthew Jenner (University of Warwick)

5.2.2 Biochemical characterisation of the KS(^{DH}DP)/DH junction

In order to biochemically characterise the role of the (^{DH}DP), a series of acyl transfer assays involving recombinant didomains ACP-KS and DH-ACP from proteins GbnD4 and GbnD5 were performed by Dr. Matthew Jenner. To do so, an acetyl unit was loaded on the ACP of the ACP-KS and incubated with holo-DH-ACP. LC-MS analy-

sis revealed successful transfer of the acetyl unit onto the DH-ACP. The same assay was performed using a mutant version of ACP-KS to which the ^{DH}DP was removed (ACP-KS(Δ^{DH}DP)). Following incubation of acetyl-ACP-KS(Δ^{DH}DP) with DH-ACP no transfer of the acetyl unit was detected confirming the fundamental role of ^{DH}DP for the interaction between GbnD4 and GbnD5 across the KS/DH interface (see Fig. 5.6A). To confirm the function of ^{DH}DP a chemical cross-linking methodology was employed. *Trans*-b-chloroamide was utilised to trap the interaction between the ACP of DH-ACP and KS. The highly electrophilic reagent is synthesised as an acyl pantetheine and loaded onto the DH-ACP following a previously described procedure[196,197]. The resulting crypto-DH-ACP was incubated with ACP-KS and the reaction products were monitored by SDS-PAGE. A clear band at $\sim$ 130 kDa was observed corresponding to the cross-linked product (ACP-KS–DH-ACP). The same cross-linking experiment was performed incubating ACP-KS(Δ^{DH}DP) with crypto-DH-ACP to probe the role of ^{DH}DP. The absence of a band corresponding to the cross-linked product confirmed the fundamental role of ^{DH}DP in the association between the two proteins(see Fig. 5.6B).

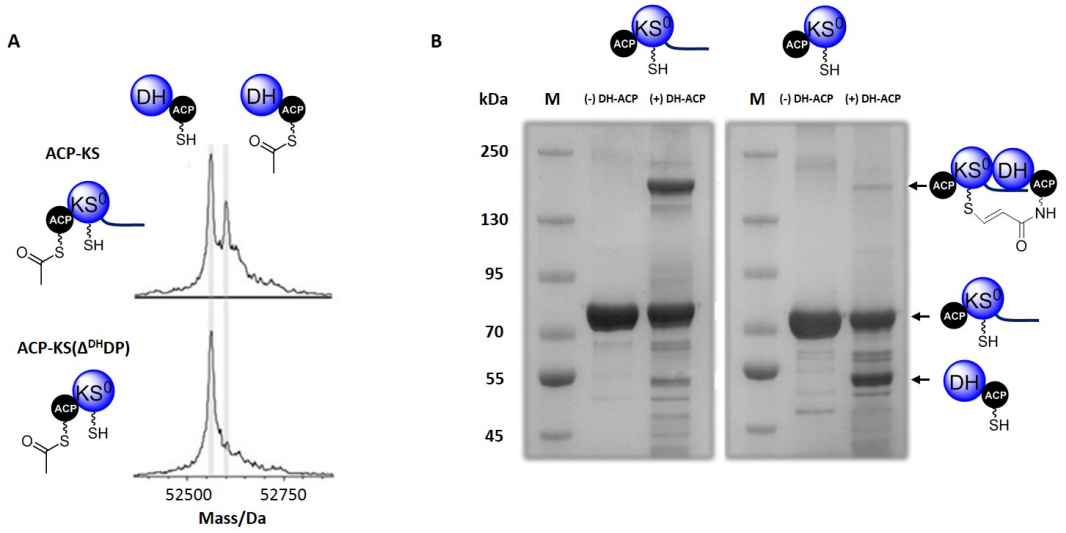


Figure 5.6: A) Deconvoluted mass spectra of GbnD5 holo-DH-ACP following incubation with GbnD4 Acetyl-ACP-KS14 (top) and Acetyl-ACP-KS14(Δ^{DH}DP) (bottom). Transfer of the acetyl moiety from Ac-ACP-KS14 to holo-DH-ACP can be monitored by additional signals corresponding to a +42 Da, mass shift on holo-DH-ACP. The results presented suggest that the acetyl group can be effectively transferred across the KS/DH protein boundary. Removal of the ^{DH}DP abolishes acetyl transfer, implying a crucial role for this region in association of the KS and DH.
B) SDS-PAGE (6%) analysis of cross-linking reactions between GbnD4 ACP-KS14 (left)/ACP-KS14(Δ^{DH}DP) (right) and GbnD5 crypto-DH-ACP. Presence of a cross-linked complex (130 kDa) is indicative of a protein-protein interaction, and can be observed by a clear additional band with ACP-KS14 (left). Removal of the ^{DH}DP prohibits complex formation, with only a faint band visible at the corresponding region with ACP-KS14(Δ^{DH}DP) (right).

5.2.3 Mapping the KS//DH split module junction interface

5.2.3.1 KS//DH peptide level analysis

After the identification of ^{DH}DP and the characterisation of its fundamental role in the interaction between GbnD4 and GbnD5, carbene footprinting experiments were performed to identify the binding patches involved in the interactions and gain insight into the structural arrangement of this particular split module junction. DH-ACP in the presence and absence of one equivalent of ^{DH}DP and isolated ^{DH}DP were labelled using AD and the reaction products were visualised by SDS-PAGE. The bands corresponding to the proteins were excised and digested using the procedure described in Chapter 2. Peptides were separated using nanoHPLC and analysed by MS in DDA mode on the five most intense ions in order to identify unlabelled and labelled ver-

sions of the digestion products. The data obtained was searched against a custom-made database including the sequences of DH-ACP and ^{DH}DP. The peptides identified accounted for 64% of the protein. The area spanning from R265 to R330 produced poor peptide coverage due to high basic residue content, which after tryptic digestion potentially generated small hydrophilic peptides that were not retained on the HPLC trapping column and therefore were not detected. Although the coverage obtained was not complete, it was considered sufficient for the purpose of the experiment given that peptides sampling areas all around the protein surface were detected and data obtained with complementary method were available. All identified peptides were also detected in their singly-labelled versions.

The fraction modified of each peptide derived from the digestion of isolated DH-ACP, DH-ACP bound to ^{DH}DP and isolated ^{DH}DP was calculated as described in Chapter 2. Briefly, chromatograms of each peptide in both their versions, labelled and unlabelled, were extracted, integrated and the ratio between the values of the peak areas were used to measure the fraction modified of each peptide. The results were compared between the isolated state of the proteins and their bound state in order to identify areas that were masked by the presence of the binding partner.

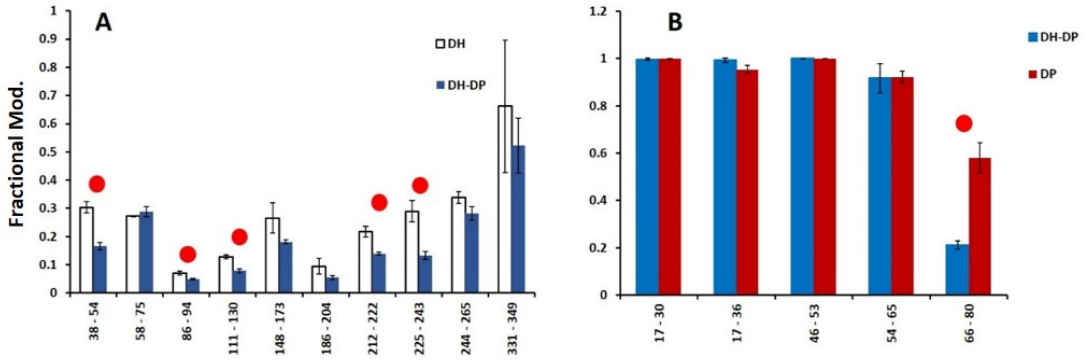


Figure 5.7: Fraction modified with AD reagent of DH-ACP peptides derived from digestion of isolated DH-ACP (figure A in white) and DH-ACP in the presence of ^{DH}DP (figure A in blue) and ^{DH}DP peptides derived from DH-ACP in the presence of ^{DH}DP (figure B in blue) and isolated ^{DH}DP (figure B in red). Six peptides showed significant and reproducible masking (Error bars are $\pm$ SD (n = 3), student's test p-value $\leq$0.05, indicated with red dots)

As shown in Figure 5.7, five peptides from DH-ACP and one peptide from ^{DH}DP showed significant masking (student's test p-value $\leq$0.05) in the presence of their respective binding partner. It is worth noting that the extent of labelling on peptides derived from ^{DH}DP is higher than the one observed for any folded protein on which the technique has been applied within this work. Four of the five ^{DH}DP peptides detected showed a fractional modification close to 1 against the average of $\sim$0.3 observed in DH-ACP. This phenomenon can be attributed to the lack of secondary structure elements in ^{DH}DP which implies a larger accessibility to the side chains of the docking peptide to the labelling reagent. The differences were mapped on a homology model of DH and the results are shown in Figure 5.8. Given its high level of disorder, a structure of ^{DH}DP is not available.

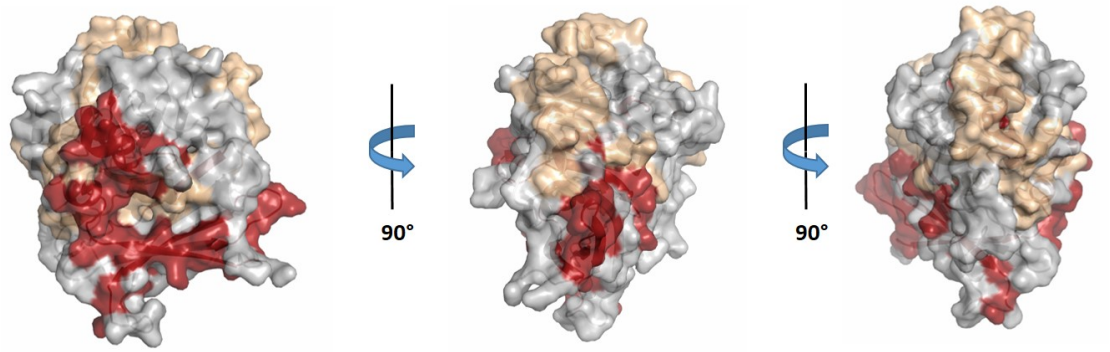


Figure 5.8: Homology model of DH showing the location of the five peptides (red) masked from labelling by ^{DH}DP binding. Areas coloured in wheat and grey correspond respectively to portions of the surface that did not show any masking and peptides that were not detected.

The area covered by the masked peptides resulted fairly extensive and spanned over two "faces" of the protein. Given that the ^{DH}DP is composed only by 90 amino acids, it seemed unlikely that the whole area that showed masking was involved in the binding. Furthermore only two small segments of ^{DH}DP were predicted to be involved in the interaction using ANCHOR (see section 5.2.1) suggesting that the contact surface does not extend over a large portion of DH. Upon a closer analysis of the spatial arrangements of the masked peptides within the protein it became evident that some of them span across two sides of the protein. Given that such mode of binding could not sensibly agree with the sizes of the proteins involved in the interactions, it was con-

cluded that the peptide level information obtained did not present sufficient resolution to pinpoint the binding site of ^{DH}DP to DH and more detailed information was needed.

5.2.3.2 KS//DH residue level analysis

5.2.3.2.1 DH binding surface In order to narrow down the interacting surface on DH, five peptides that showed significant decrease in labelling were subjected to CID fragmentation to pinpoint which residues were modified and calculate the extent of the modification at the residue level. Peptides 38-54 and 211-222 were found to be labelled on several residues, mostly of hydrophobic nature, yielding near-residue resolution information. Ten individual residues or small groups of residues were found to be significantly masked in the presence of ^{DH}DP (student's test p-value $\leq$0.05). The results are shown respectively in Fig. 5.9 A and Fig. 5.9 C. Per-residue analysis on peptide 86-94 showed labelling occurring on most residues, again mostly hydrophobic, but no significant masking was detected (shown in Fig. 5.9 B). Although 86-94 showed small but significant masking at the peptide level, such small difference resulted "diluted" over several residues leading to an overall absence of significant masking.

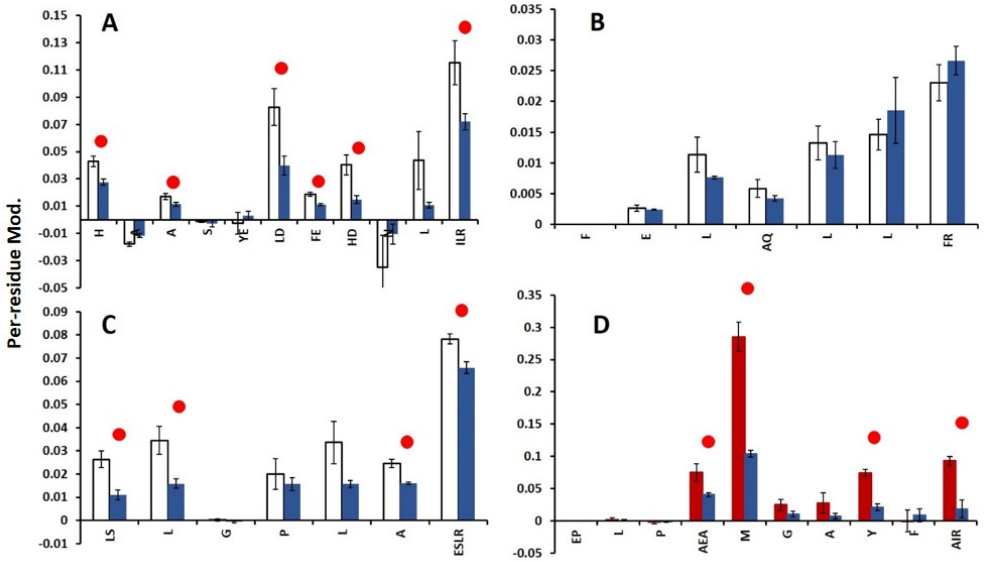


Figure 5.9: Per-residue modification calculated from MS/MS experiments of peptides showing significant masking at the peptide level. A,B and C respectively represent peptides 38-54, 86-94 and 211-222 from isolated DH-ACP (white) and DH-ACP in the presence of ^{DH}DP (blue). D represents peptide 66-80 from isolated ^{DH}DP (red) and ^{DH}DP in the presence of DH-ACP (blue). Several residues or small group of residues showed significant masking (Error bars are $\pm$ SD (n = 3), student's test p-value $\leq$0.05, indicated with red dots). Low abundance of peptides 111-130 and 225-243 (both showing peptide level significant masking) lead to poor MS/MS data quality from which it was impossible to obtain reliable per-residue information.

Unfortunately, isolation and fragmentation of labelled peptides 111-130 and 225-243, both showing significant masking at the peptide level, resulted in poor quality MS/MS data, probably because of the low abundance of such peptides. The low yield of CID fragmentation led to unreliable quantification of the per-residue modification therefore the low resolution structural information obtained from the peptide level analysis was kept in consideration for the identification of the binding patch.

The residues showing significant decrease in labelling (student's t-test $\leq$0.05) were visualised on a homology model of DH (Fig.5.10).

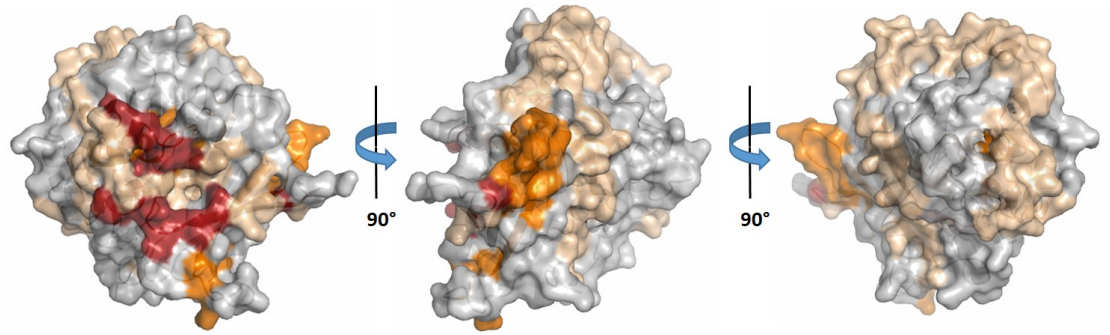


Figure 5.10: Homology model of DH showing the location of the residues (red) masked from labelling by ^{DH}DP binding. Areas coloured in wheat and grey correspond respectively to portions of the surface that did not show any masking and peptides that were not detected. Areas coloured in orange correspond to peptides showing significant masking at the peptide level but unable to locate and quantify the modification with amino acid resolution.

Following residue level analysis, the surface that showed masking resulted to be less extended than the overly large patch identified through the peptide level analysis. Residues showing masking were pinpointed on a smaller area localised only on one face of the protein. Peptide 225-243, one of the two peptides yielding poor fragmentation data (orange in Fig.5.10), still raises some uncertainties about the location of the binding interface. Although spanning across a large area with part of it located away from the supposed binding site, the peptide includes a small portion located adjacent to the interacting surface identified by CID experiments. It is therefore possible to hypothesise that the area responsible for the decrease in fractional modification observed at the peptide level could be limited to such area, a possibility that would only lead to a minor extension of the identified binding surface.

5.2.3.2.2 ^{DH}DP binding surface Peptide 66-80 of ^{DH}DP showed a reduction in fractional modification from 0.581 when labelled in the absence of any binding partner to 0.211 when modified in the presence of DH-ACP. In order to narrow down the binding site the same CID fragmentation-based experiment was performed on 66-80 derived from ^{DH}DP labelled in the presence and absence of DH-ACP. Most residues resulted labelled with M73 being the amino acid exhibiting the highest extent of modification. Four among the single residues and small groups of residues showed a sig-

nificant and reproducible decrease in labelling when DH-ACP was present in solution (shown in Fig.5.11 A). The masked residues could not be visualised in three dimensions as the other examples described in this chapter because of the extreme flexibility that ^{DH}DP exhibits and the subsequent absence of a well defined structure available but the data was validated against NMR-based structural studies. BEST-TROSY-^{15}N-HSQC of ^{15}N labelled ^{DH}DP incubated with increasing concentrations of DH-ACP were acquired in order to identify residues involved in the interaction between the two domains (see Fig. 5.11 B). The NMR-based experiments were performed by Dr. Daniel Griffiths (University of Warwick)

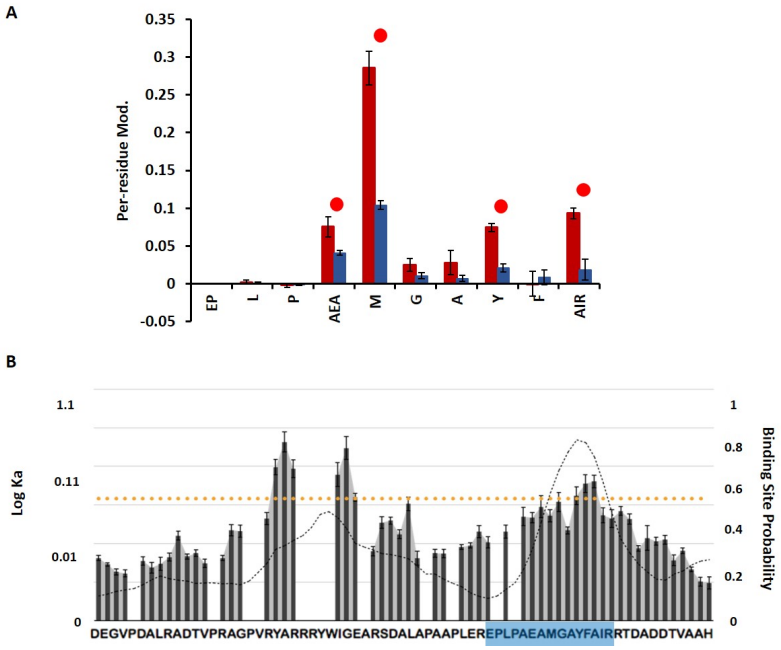


Figure 5.11: A) Per-residue modification calculated from MS/MS experiments of ^{DH}DP peptide 66-80 in the presence (blue) and absence (red) of DH-ACP. Several residues or small group of residues showed significant masking (Error bars are $\pm$ SD (n = 3), student's test p-value $\leq$0.05, indicated with red dots). B) Signal decay rates of assigned residues (black bars and grey area) in NMR titrations of ^{15}N labelled ^{DH}DP with DH-ACP. Decay rates above the threshold (decay rate mean + standard deviation; dotted orange line) indicate regions involved in binding. Dashed line indicate binding sites predicted using the ANCHOR server. The blue box indicates peptide 66-80. Residues involved in the interface identified through carbene footprinting and NMR chemical shift perturbation show excellent agreement.

Pleasingly, strongly decreasing signals were observed for the residues spanning from A75 to R80, the same region previously identified as involved in the binding of DH-ACP by carbene footprinting and predicted through the ANCHOR server. A second

arginine-rich region was also identified through chemical shift perturbation NMR experiments but such binding region was not confirmed using carbene footprinting because the presence of several trypsin cleavage sites lead to the generation of undetectable small hydrophilic peptides.

5.2.4 Mapping the DH/ACP interaction

In order to gain further insight into the spatial arrangement of the domains involved in the GbnD4/GbnD5 split module junction, the attention was turned to the identification of the binding patch responsible for the interaction between the DH domain and its donwstream ACP. Labelling experiments were performed on isolated DH, isolated ACP and on the DH-ACP didomain followed by SDS-PAGE analysis, in-gel digestion and HPLC-MS analysis (see Fig. 5.12).

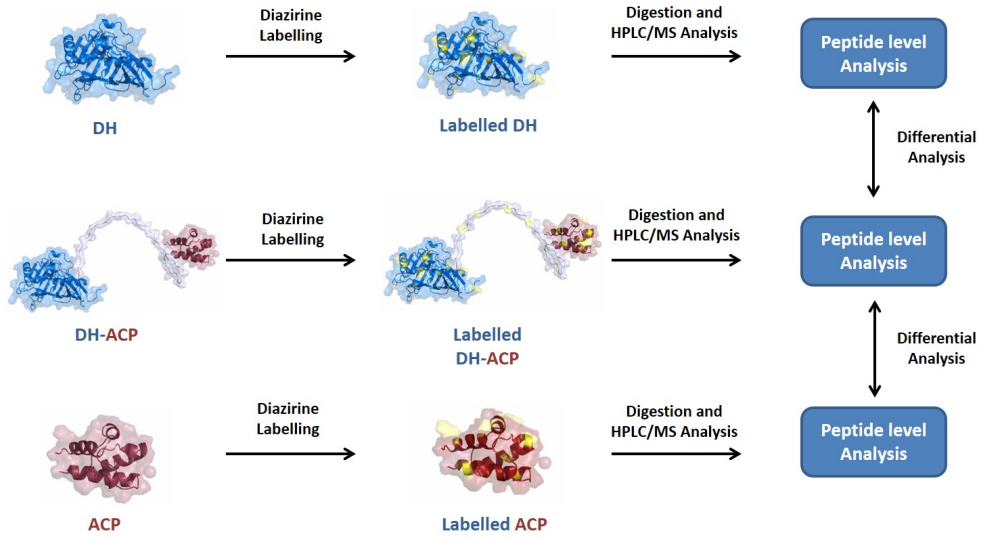


Figure 5.12: Workflow of DH-ACP binding interface footprinting. Isolated DH, DH-ACP and isolated ACP were labelled using AD, digested and analysed by HPLC/MS. Peptide level analysis results were compared across the different conditions to identify the binding surface.

5.2.4.1 DH/ACP peptide level analysis

The peak areas of extracted ion chromatograms of unlabelled and singly-labelled peptides derived from the digestion of the isolated proteins and the didomain were used

to calculate the fraction modified of each peptide. The results were compared in order to identify peptides showing significant differences to infer the binding interface involved in the DH/ACP interaction. The differential analysis of the peptide fractional modification revealed three peptides derived from DH and one peptide derived from the ACP showing a significant and reproducible decrease in fractional modification (student's t-test $\leq$0.05 annotated with red dots in Fig. 5.13)

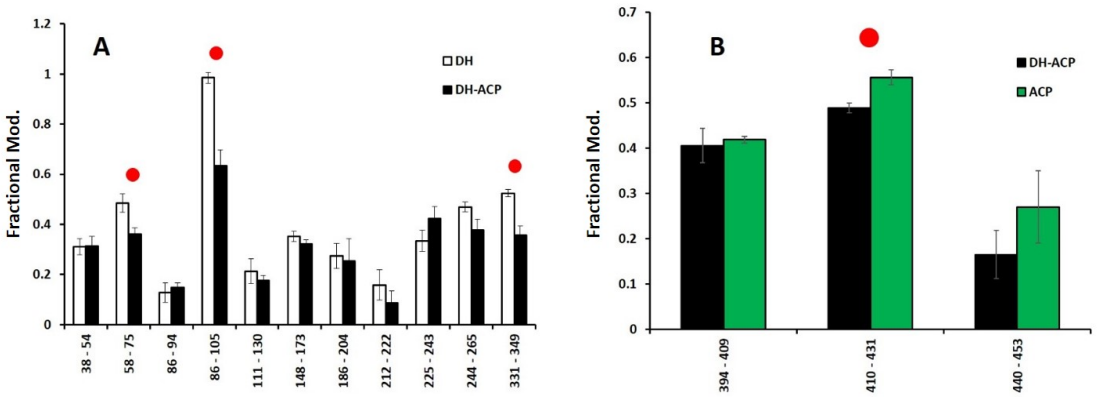


Figure 5.13: Fraction modified with AD reagent of DH-ACP peptides derived from digestion of isolated DH (figure A in white) and DH-ACP (figure A in black) and ACP peptides derived from DH-ACP (figure B in black) and isolated ACP (figure B in green). Four peptides showed significant and reproducible masking (Error bars are $\pm$ SD (n = 3), student's test p-value $\leq$0.05, indicated with red dots)

Peptide 86-105 showed the highest difference in fractional modification decreasing from 0.98 to 0.63 whereas the decrease exhibited by the other three peptides resulted somehow lower (-0.12% on peptide 58-75, -0.16% on peptide 331-349, -0.07% on peptide 410-431) but significant and reproducible. The low decrease in labelling associated with the binding event can be attributed to the nature of the interaction between the ACP and the DH domain; the dissociation constant of this interaction has not been evaluated but it is safe to assume that the complex cannot be extremely stable given that the ACP must interact with several other domains in order to deliver the substrate. The transient nature of the complex may lead to partial accessibility of the label to the binding site resulting in the observed low masking extent. Following visualisation of the masked areas on homology models of DH and ACP, a large patch was identified but the masking resulted extended over other two different areas of the protein surface

(Fig. 5.8 A). The same phenomenon was observed on the ACP with a large portion of one of the helices resulting masked in addition to a small helix turn located next to it (Fig. 5.8 B). The level of resolution obtained through the use of the peptide level analysis resulted once again not sufficient to confidently identify the contact surface of the two proteins; the presence of a large masked patch on the DH is incompatible with the small size and compact fold of the ACP and the two smaller masked areas located away from the patch raised further uncertainties. For these reasons it was concluded that CID fragmentation experiments were needed in order to increase the resolution of the information and narrow down the interacting surface of the two proteins.

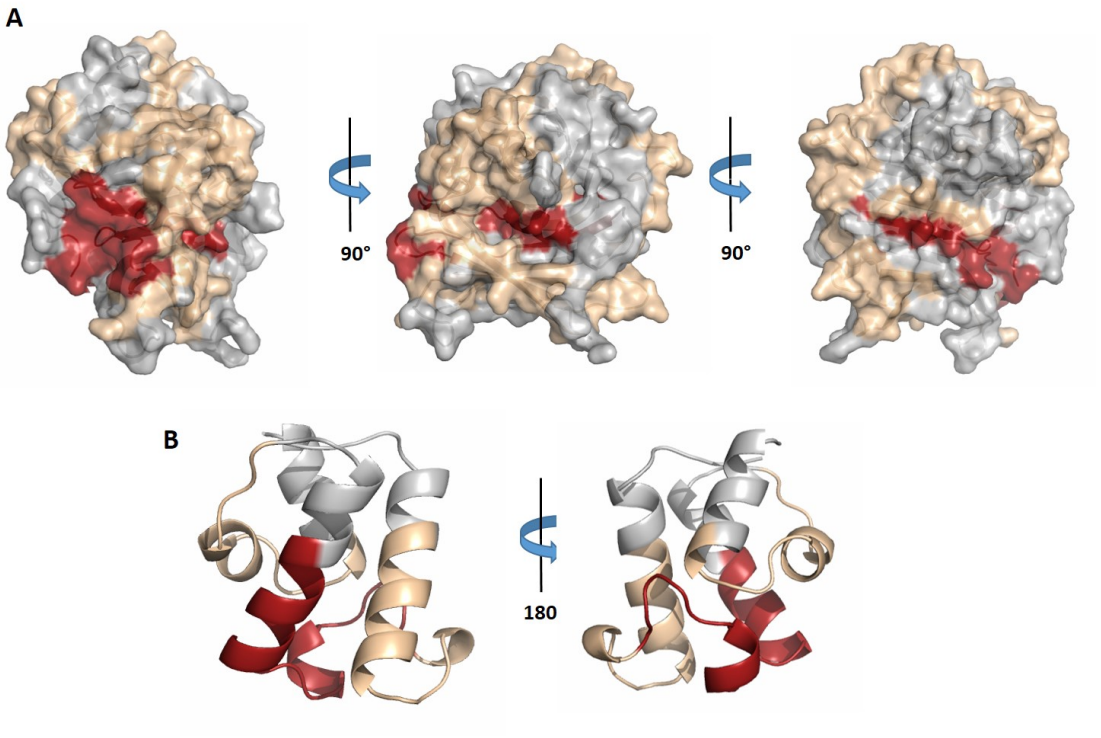


Figure 5.14: Homology models of DH (A) and ACP (B) showing the location of the four peptides (red) masked from labelling by the presence of the binding partner. Areas coloured in wheat and grey correspond respectively to portions of the surface that did not show any masking and peptides that were not detected.

5.2.4.2 DH//ACP residue level analysis

5.2.4.2.1 DH binding surface Singly labelled peptides 58-75, 86-105 and 331-449 derived from the digestion of isolated labelled DH and labelled DH-ACP were sub-

jected to CID fragmentation in order to locate and quantify the modification within the peptides and narrow down the area masked by the presence of the ACP. Most of the amino acids by which peptide 58-75 is composed of resulted labelled but interestingly only Leu64, a residue located in the major binding patch identified at the peptide level, showed reproducible and significant masking in the presence of ACP (Fig. 5.15 A). The rest of the amino acids C-terminal to Leu64 exhibited the same labelling extent in both DH and DH-ACP leading to the exclusion from the binding interface of one of the areas located away from the larger patch identified at the peptide level.

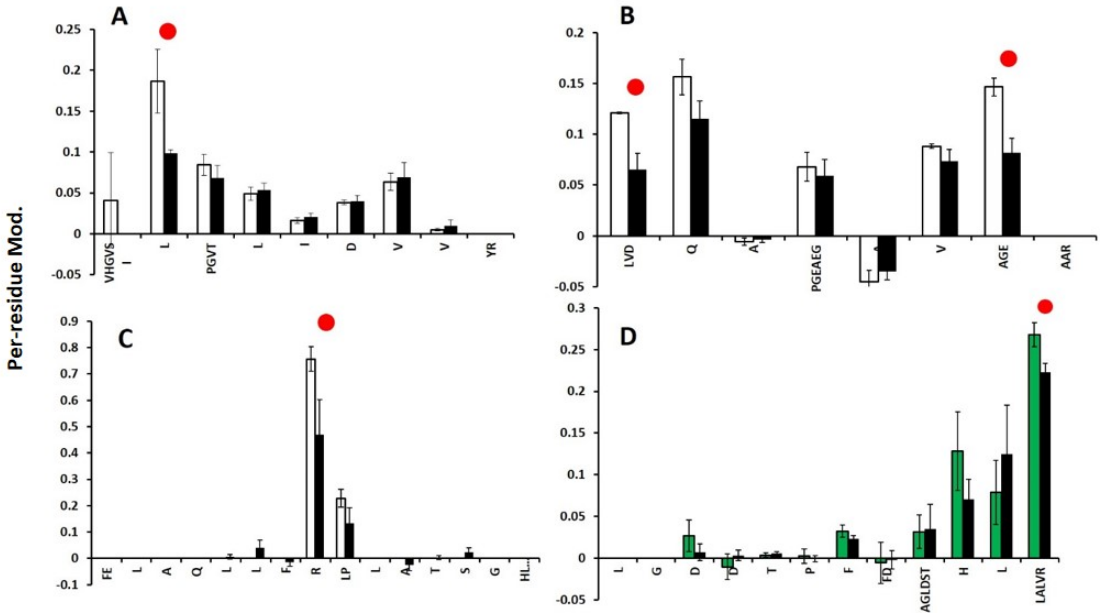


Figure 5.15: Per-residue modification calculated from MS/MS experiments of peptides showing significant masking at the peptide level. A,B and C respectively represent peptides 58-75, 331-349 and 86-105 from isolated DH (white) and DH-ACP (black). D represent peptide 410-431 from DH-ACP (black) and isolated ACP (green). Several residues or small group of residues showed significant masking (Error bars are $\pm$ SD (n = 3), student's test p-value $\leq$0.05, indicated with red dots).

Arg95 resulted to be the only residue exhibiting significant masking in peptide 86-105 with the remaining amino acids being unlabelled or equally labelled in both bound and unbound DH (see Fig. 5.15 C). The absence of difference in the labelling pattern lead to ruling out another large portion of the area identified at the peptide level. Two other small groups of residues were identified as masked in peptide 331-349 (see Fig.5.15 B) and after compiling the whole residue level information obtained together and visu-

alising it onto the DH homology model it was possible to narrow down the binding surface to a small patch located on the opposite "face" of the protein in respect with the previously identified ^{DH}DP contact surface (Fig. 5.16).

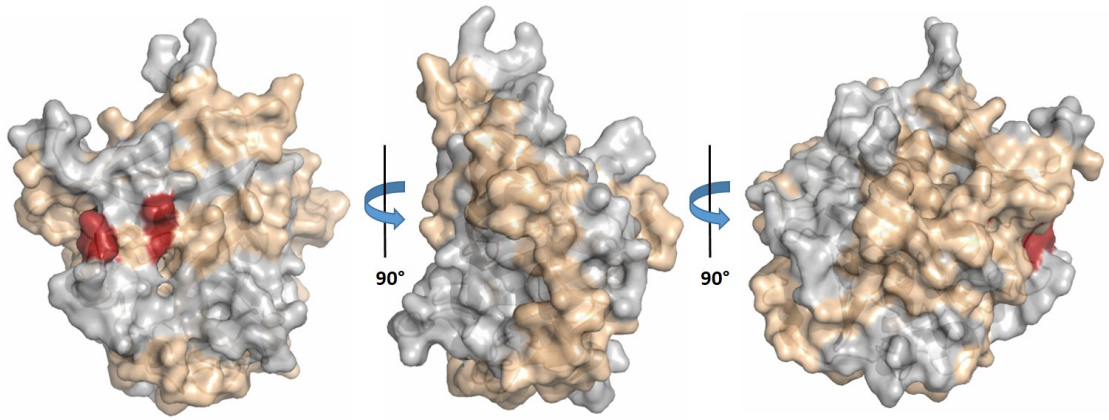


Figure 5.16: Homology model of DH showing the location of the residues (red) masked from labelling by ACP binding. Areas coloured in wheat and grey correspond respectively to portions of the surface that did not show any masking and peptides that were not detected.

5.2.4.2.2 ACP binding surface Peptide 410-431 exhibited a small but significant reduction (-0.07) in fractional modification when the ACP was bound to the DH domain. Fragmentation experiments revealed that most of the labelling was located towards the C-terminus of the peptide with the highest value observed on a small group of unresolved amino acids spanning from Leu428 to Arg431 (Fig. 5.15 D). The same small group of residues corresponded to the area found to be masked by the presence of DH with the remaining portion of the peptide unaffected by the binding event (Fig.5.17).

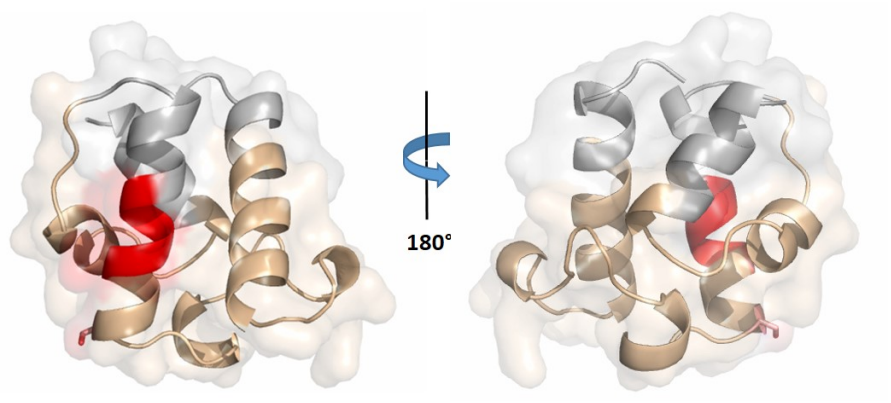


Figure 5.17: Homology model of ACP showing the location of the residues (red) masked from labelling by DH binding. Areas coloured in wheat and grey correspond respectively to portions of the surface that did not show any masking and peptides that were not detected.

5.2.4.2.3 DH-ACP modelling Once the area involved in the interaction between the two proteins was mapped, insight on the relative spatial arrangement of the two polypeptides was sought after. In order to do so, docking prediction using the ZDOCK server were performed employing the residue level information obtained through carbene footprinting as constraints. The ZDOCK server is a Fast Fourier Transform based protein docking program that searches all possible binding modes in the translational and rotational space between the two proteins and evaluates each pose, using a scoring function based on a combination of shape complementarity, electrostatics and statistical potential terms. This program allows the user to specify residues involved in the binding site or block amino acids from being part of the interacting surface and filters the results such that only structures in agreement with the information provided are returned. Arg94, Leu331, Val332 and Asp333 were specified as DH residues involved in binding whereas the small segment between Leu427 and Arg 431 were included in the ACP binding site. The software returned a small number of potential docked structures that were analysed in order to assess their agreement with the footprinting data and one of the complexes resulted highly compatible with the information gathered about the contact surface. The selected structure was then energy minimised using the ModRefiner algorithm and the final result revealed extensive contacts between the masked areas and the correspondent binding partner (shown in Fig. 5.18) confirming

the effectiveness of carbene footprinting for the mapping of binding interactions. Although involved in the binding interface, the identified protected portions of the proteins are not completely aligned. The small protected area on the ACP is involved in an interaction with a portion of the DH rich in basic residues which unfortunately was not possible to detect, probably because the tryptic digestion yielded small peptides not retained by the HPLC column used for the analysis. The residues that exhibited masking on DH instead seem to be interacting with a flexible loop of the ACP which was covered by the analysis but it did not show any significant difference in labelling at the peptide level. One of the potential reasons for the lack of detectable masking could be attributed to the nature of the interacting residues coinciding with tryptic cleavage sites which in case of labelling may be resistant to the action of trypsin.

Interestingly the contact surface presents a crevice at the bottom of which it is possible to identify the two catalytic residues of the DH domain, Asp233 and His56. The residues were identified through the catalytic motifs homology with the DH domain of erythromycin PKS.

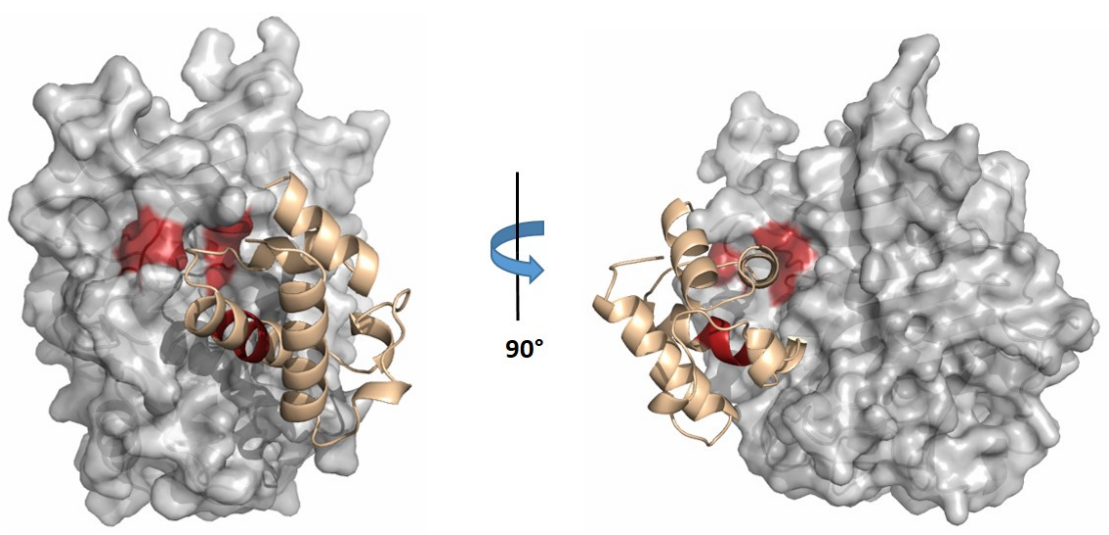


Figure 5.18: Model of the complex between DH (grey) and ACP (wheat) generated by ZDOCK using footprinting data as constraints (areas coloured in red). The structure was refined using ModRefiner. The model revealed extensive contacts DH and ACP through the areas identified using carbene footprinting.

The dehydration of the β-hydroxyl chain of the growing polyketide by the DH domain proceeds through the donation of a proton from an aspartic acid residue in the

HPALLD motif followed by the extraction of the α-proton by the catalytic histidine residue generating the final $\alpha\beta$-unsaturated product (see Fig.5.19 A). In the structure proposed the crevice is aligned to Ser423 of the ACP, the residue responsible for carrying the growing product through a tethered phosphopantetheinyl, chain and the distance between Ser423 and Asp233 was measured at 1.8 nm, approximately the length of a phosphopantetheinyl chain (shown in Fig.5.19 B). The alignment of the acyl-carrier serine residue with the crevice containing the catalytic residues provides further evidence towards the goodness of the proposed model.

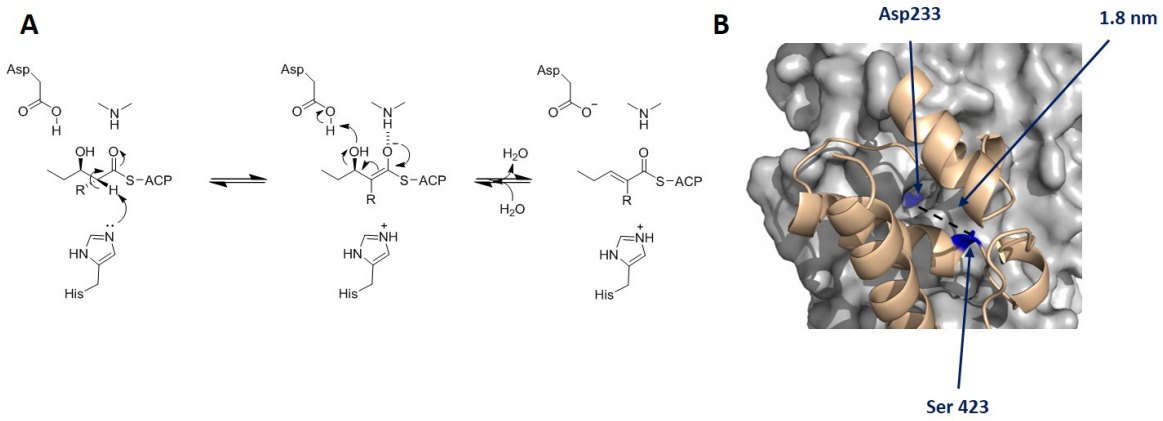


Figure 5.19: A) Proposed catalytic mechanism of DH domain. The abstraction of the C-H proton by a histidine residue and proton donation from an aspartic acid residue is followed by the release of water generating the unsaturated product. B) Detail of the DH-ACP binding interface. The crevice containing the catalytic residues aligns with the acyl-carrier Ser423. The distance between Ser423 and Asp223 was measured in Pymol at 1.8 nm.

5.3 Conclusions

The work in this chapter has successfully demonstrated that the interaction across the KS/DH disconnection in the Gladiolin trans-AT PKSs is directed by the presence of an additional polypeptide region located at the C-terminus of GbnD4, downstream to the KS domain. This docking region revealed to be intrinsically disordered by nature with few localised elements of secondary structure. Acetyl transfer assays and crosslinking experiments proved that the role of ^{DH}DP is essential for the association between

GbnD4 and GbnD5, and the absence of it inhibits the transfer of the growing product from a module to the downstream ACP across the boundary.

The N-terminal DH domain of GbnD5 was footprinted in the absence and presence of ^{DH}DP in order to identify the binding surface between the two polypeptides. Peptide level analysis revealed five significantly masked peptides but following mapping of the corresponding areas on a homology model of DH, the resolution of the information obtained resulted insufficient to confidently identify the binding patch. To improve the resolution of the information CID fragmentation experiments were performed on the protected peptides allowing the binding interface to be narrowed down and confidently mapped at near amino acid level. The same differential footprinting experiment on isolated ^{DH}DP allowed the identification of a small group of amino acids exhibiting lower labelling in the presence of DH. The same residues were confirmed to be involved in the protein-protein interaction through NMR titrations confirming the effectiveness of carbene-based footprinting.

In addition to the interaction between DH and ^{DH}DP, the contact surface between the GbnD5 DH domain and its downstream ACP was examined. Differential footprinting experiments allowed the identification of a small number of residues involved in the interaction. In order to gain insight into the tridimensional arrangement of the two proteins, the ACP was docked onto the surface of GbnD5 DH using ZDOCK with residue constraints derived from footprinting experiments. Following structure refinement using ModRefiner, the model obtained presented the masked areas on both DH and ACP involved in the interaction and the side chain of ACP acyl carrier residue Ser423 aligned with the active site of the DH domain (model shown in Fig.5.20).

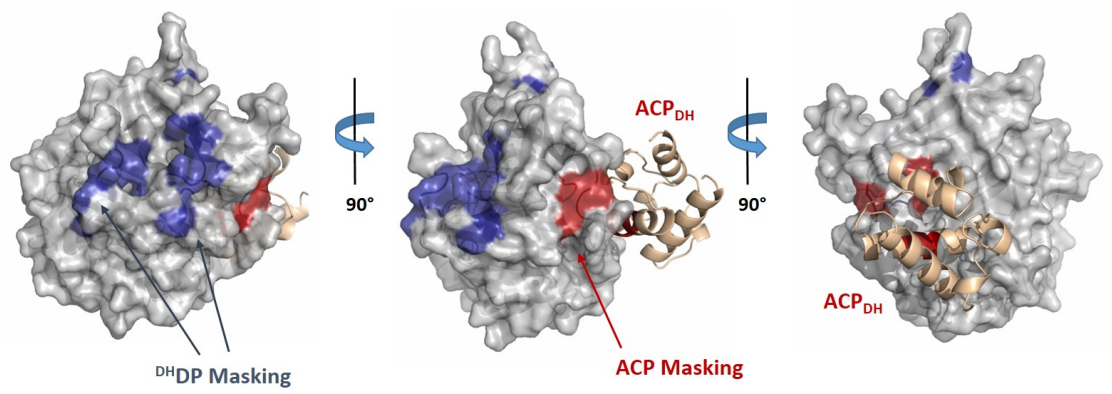


Figure 5.20: Model of the complex between DH (grey) and ACP (wheat) generated by ZDOCK using constraints derived from carbene footprinting, refined using ModRefiner. Two distinct binding patches were identified for ^{DH}DP (blue) and ACP (red).

The model highlighted distinct binding patches for the ^{DH}DP and ACP, allowing both GbnD4 KS and GbnD5 ACP to interact with GbnD5 DH simultaneously, allowing close proximity for effective acyl transfer across the protein junction. The results obtained confirm the potential of AD-based footprinting for the identification of protein-protein contact surfaces and the structural insight gained on the interaction may help future synthetic biology efforts to employ the chemistry of PKSs to synthesise bespoke bioactive compounds.

CHAPTER 6

Carbene footprinting reveals binding interfaces of a multimeric membrane spanning protein

6.1 Introduction

The role of membrane proteins is undoubtedly vital for the homeostasis of the cell. Given their pivotal role in a vast number of cellular functions including respiration[198] transport, and signal transduction[199], an increasing number of proteins belonging to this class are drug targets[200]. Membrane proteins are usually classified into two main categories: peripheral and integral. Peripheral membrane proteins are involved in temporary interactions with the lipid bilayer or with integral membrane proteins (IMPs) whereas biomolecules categorised in the latter group are permanently attached to the membrane. Proteins spanning across the two sides of the bilayer present an amphiphilic nature, with the area in contact with the hydrophobic core of the membrane containing mostly non-polar amino acids and the areas extending into the aqueous environment possessing hydrophilic character. Obtaining structural information on integral membrane protein is notoriously more difficult than with water-soluble proteins because

their hydrophobic nature requires additives to be employed in order to solubilise them and preserve their native conformation in water. Various detergents have been employed for the extraction of these proteins from the membrane environment and their encapsulation within micelles[201]. Although effective in the stabilisation of the protein in solution, micelles are not the best representation of the natural membrane environment of these proteins. Their curvature restriction and the potential excess of detergent in solution has been shown to potentially perturb the interactions of these proteins with their substrate[202]. For this reason several methodologies have been successfully employed to constitute an environment more similar to the native one for IMPs, including liposomes, nanodiscs and monolayers. Nevertheless, detergents have been widely used, under carefully optimised conditions, for the study of membrane protein structure and interactions via NMR, X-ray crystallography and mass spectrometry[201].

6.1.1 MS-based techniques for the study of MPs

The use of mass spectrometry for the investigation of membrane protein interactions has been pioneered by the Robinson group[203]. Their work demonstrated that large multi-subunit protein complexes, such as ATP synthase, can be electrosprayed when solubilised in micelles and detected following the removal of bound detergent molecules using collisional activation. Such approach reveals the stoichiometry of the assembly and the presence of small molecules bound to it but does not provide structural information about the binding interface.

Various footprinting techniques have been used to investigate membrane protein structure and the conformational changes associated with various triggering events. In a few cases HDX has proven successful in reporting the accessibility of solvent-exposed backbone hydrogen atoms of micelle-solubilised integral membrane proteins. Examples include the investigation of the conformational dynamics of Glycerol Facilitator (GF)[204], the study on folding kinetics of bacteriorhodopsin[205], and the conformational analysis of γ-glutamyl carboxylase in nanodiscs[206]. The main drawback associated

with the application of this technique to IMPs is related to the difficult digestion of these highly hydrophobic proteins under typical acidic quenching conditions.

Oxidatively labelling has also been used for MS-based structural studies. The first FPOP oxidative labelling on a membrane protein has been carried out by Konermann et al. on bacteriorhodopsin[207]. The study revealed that solvent exposed readily oxidizable side chains are highly more reactive towards hydroxyl radicals compared to amino acids of transmembrane regions. Fenton chemistry has also been used to produce hydroxyl radicals and oxidize the porin OmpF *in vivo* revealing its voltage gating[208]. An interesting application of oxidative labelling was proposed by Chance et al. in 2009[209]. Using radiolytic hydroxyl radical labelling on rhodopsin the authors observed oxidation on both solvent-accessible and solvent-inaccessible regions, in contrast with previous labelling experiments showing the absence of oxidation on transmembrane regions. Using a combination of radiolytic labelling and $H_2{}^{18}O$ exchange they found that the oxidation in buried regions derived from tightly bound waters. The oxidation of such residues was found to be dependent on the activation state of rhodopsin suggesting a possible mechanism for water-dependent communication in G protein-coupled receptors.

More recently FPOP labelling has been employed to probe the conformation and topology of light harvesting complex 2 (LH2) in nanodiscs[87] and the differences in solvent accessibility of OmpT solubilised in micelles and in amphipols[71]. These studies revealed that both amphipols and nanodisc provide better protection to membrane proteins than micelles do rationalising the observed higher stability in both systems compared to detergent-based media.

6.1.2 OmpF

OmpF is a porin found in bacteria responsible for the non-specific transport of small hydrophilic molecules (600-700 Da) through the outer membrane and its high-resolution structure was solved in 1992[210]. Antibiotic resistance has been linked to the loss of the

porin and mutation of key residues has been shown to result in an altered response towards β-lactam antibiotics[211]. These studies suggest that a decrease in cell permeability leads to a reduced accumulation of antibiotic resulting in the loss of efficiency of the drug. This resistance mechanism has been proved to work in association with other mechanisms such as the expression of efflux pumps and enzymatic drug inactivation. The crystal structure of OmpF reveals a homo-trimer assembly of 16-stranded β-barrel subunits (see figure 6.1). Each subunit contains a pore with a loop restricting the channel whose permeability is regulated by charge distribution.

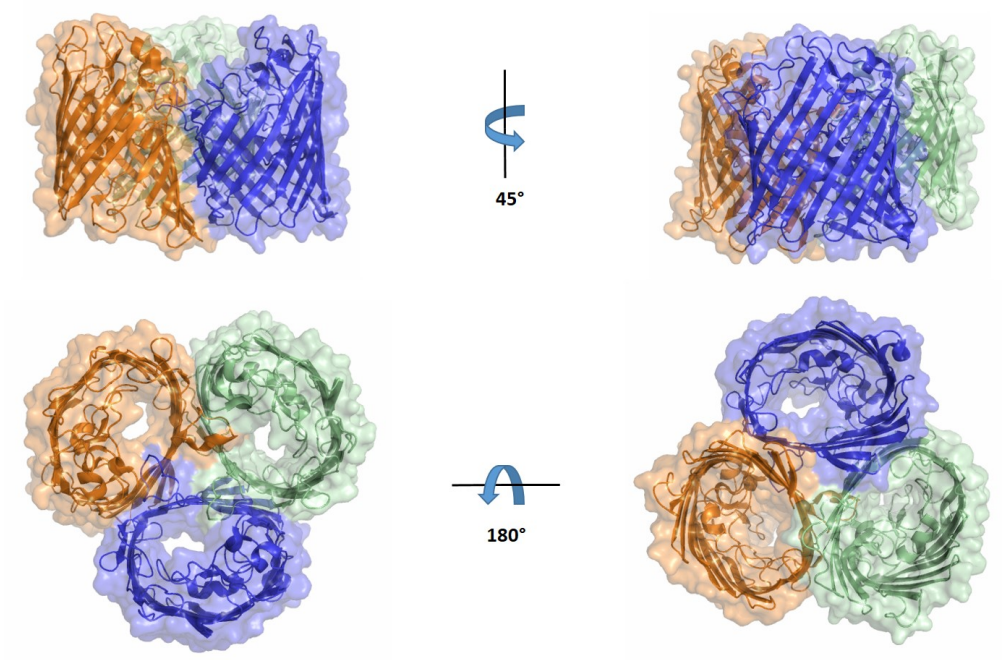


Figure 6.1: Cartoon and surface representation of OmpF homotrimer. Rendered in pymol from pdb 3K1B.

The use of AD-based footprinting on OmpF solubilised in octyl-glucoside (OG) micelles is described in this chapter. The photo-reagent showed preferential labelling of residues located in transmembrane regions. The absence of labelling at the interface between the subunits and on solvent-exposed areas suggested that AD tends to partition in the micelle before the photolysis and labels areas of the protein in proximity of detergent molecules. The data shows how the behaviour of AD in the presence of micelle-solubilised membrane proteins provides a novel MS-based methodology to

identify the contact areas between the subunits of MP assemblies and gain insight into their quaternary structure.

6.2 Results and discussion

6.2.1 Footprinting of OmpF in OG micelles

6.2.1.1 Peptide level analysis

Given the amphipatic nature of AD, it was postulated that the label would preferentially associate with molecules of the detergent, orderly arranged in a micelle, prior to photolysis. It was hypothesised that the reagent, partitioned in such fashion, upon photo-activation in conditions of slow diffusion (-78 °C) would react with the highly hydrophobic portions of membrane proteins spanning through the core of the micelle. To test this idea labelling experiments were performed on OmpF solubilised in 0.1% octyl glucoside (*n*-octyl-β-D-glucoside). OG is a non-ionic surfactant commonly used for membrane protein solubilisation and its amphiphilic nature is provided by the presence of a polar portion, a glucose moiety, linked through a ether bond to a hydrophobic hydrocarbon chain (shown in Fig. 6.2). The critical micellar concentration (CMC) reported for the surfactant is 0.025 M, therefore the labelling experiments were performed in conditions favouring the formation of micelles.

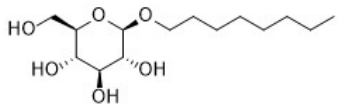


Figure 6.2: Structure of octyl glucoside

Laser irradiation of OG-solubilised OmpF in the presence of increasing concentrations of AD was performed in order to define optimal labelling conditions followed by SDS-PAGE separation, in-gel tryptic digestion and calculation of the per-peptide fractional modification. Twenty two peptides were detected, accounting for 86% coverage of the

protein sequence, of which 12 showed significant AD labelling (fractional modification >5%, shown in Fig. 6.3). As reported in chapter 3, 4 and 5, high levels of labelling are achievable on water-soluble proteins when irradiated in the presence of 5-10 mM AD. OmpF solubilised in OG micelles instead exhibited sub-optimal levels of labelling at such concentration and resulted significantly labelled only when irradiation was performed in the presence of 50 mM of the probe.

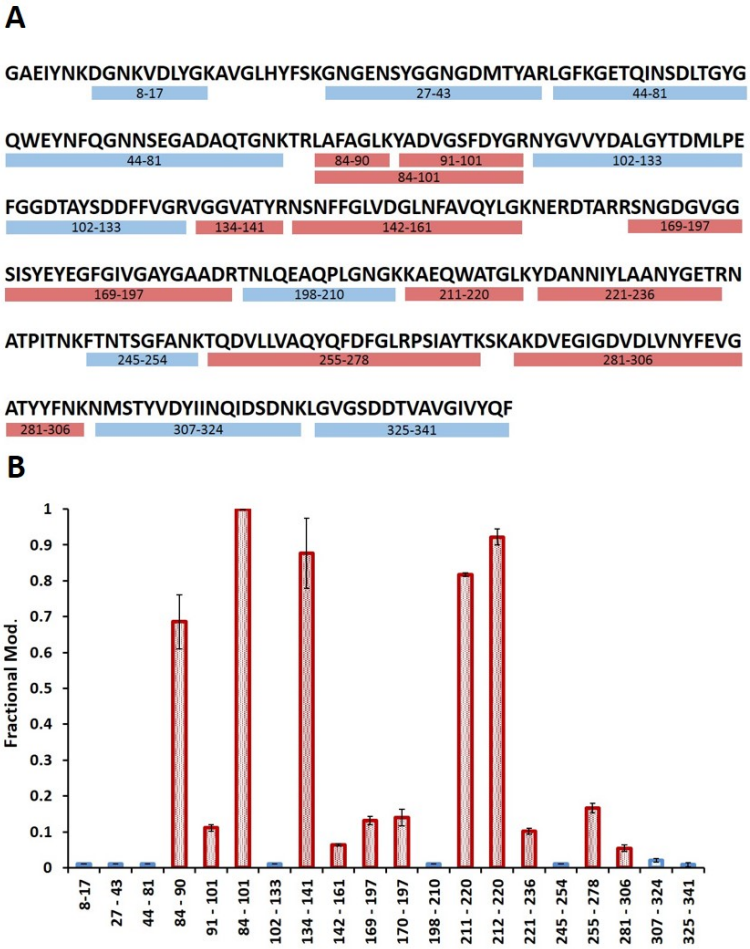


Figure 6.3: A) Sequence coverage of OmpF. Unlabelled and significantly AD labelled peptides (fractional modification > 0.05) indicated respectively by blue and red bars. B) Fractional modification of AD labelled peptides. Blue and red bars correspond respectively to peptides found to be unlabelled or labelled.

The peptides showing significant labelling (fractional modification > 0.05) were mapped onto the structure of OmpF revealing absence of carbene modification localised around the portion of the β-barrel involved in the interaction between the subunits of the trimer

(shown in Fig.6.4 A). The protein resulted labelled almost exclusively on one face of the monomer, a portion of the assembly protected from the aqueous environment by close contact with the cellular membrane (shown in Fig. 6.4 B) as reported in previous structural studies[210]. Rosenbusch and co-workers identified two series of hydrophobic residues flanking the OmpF β-barrel and postulated that they correspond to the boundaries of the area buried by the membrane in physiological conditions (shown in Fig. 6.4 C).

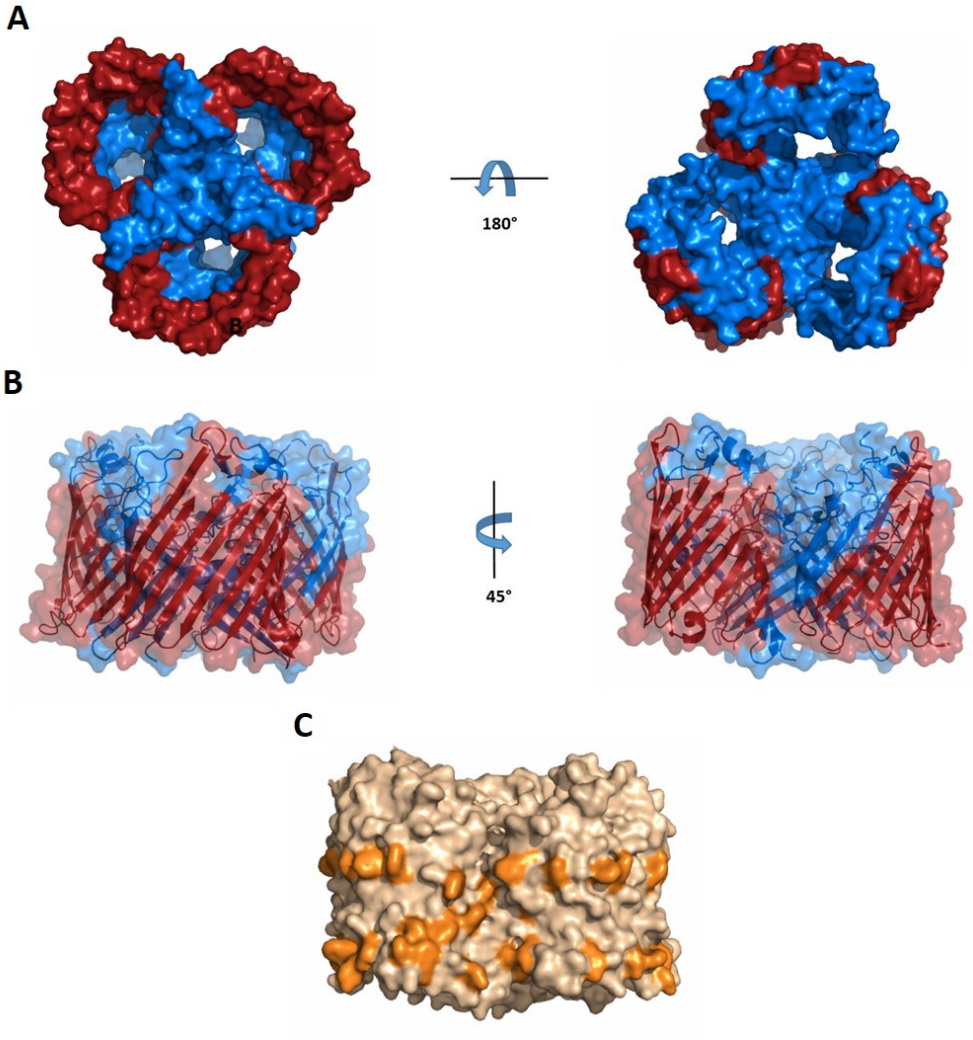


Figure 6.4: Top-down (A) and side view (B) of surface and cartoon representation of the OmpF homotrimer. Red and blue areas represent respectively labelled and unlabelled areas. Residues coloured in orange in C represent hydrophobic residues expected to delimit the membrane-spanning area.

Assuming a certain degree of overlapping between the areas of OmpF in close contact with OG micelles and biological membranes, the labelling pattern observed points to the confirmation of the hypothesis that AD preferentially labels areas of the protein buried by the micelle. Figure 6.4 A shows how the intracellular portion of OmpF (left) results labelled whereas the extracellular domain does not show any significant modification (right). This observation is in agreement with the model proposed by Rosenbusch which depicts the intracellular side of the channel almost delimited by one of the two series of hydrophobic residues delimiting the membrane spanning region. The extracellular domain instead shows a larger area extending from the other series of hydrophobic residues delimiting the membrane. The absence of labelling on such portion (shown in Fig.6.4 A-left) confirms that labelling occurs exclusively in proximity of the core of the micelle. Further evidence corroborating the theory is provided by the absence of labelling observed on the highly solvent-exposed loop located in the barrel, whose role is to provide gating function to the channel. The areas included within the two series of hydrophobic residues resulted significantly labelled, excluding the areas located at the interface of the trimer, which cannot be in contact with OG molecules in order to interact and assembly in its native conformation and therefore should not be exposed to AD prior and during its activation.

Analysis of the labelling pattern showed a clear correspondence between AD-modified hydrophobic areas and their location in proximity to the expected non-polar portion of the micelle in which the protein is solubilised. Thus, the absence of labelling can be used to infer whether certain portions of the protein surface are in close contact with ordered detergent molecules and, in principle, as a reporter for the interfacial areas of multi-subunit membrane proteins.

6.2.1.2 Residue level analysis

Analysis of the labelling pattern at the level of resolution achievable using tryptic digestion was found to be sufficient to map the membrane-spanning surface of the OmpF

homo-trimeric assembly in OG. However, in order to increase the resolution of the technique and narrow down the location of labelled amino acids, MS/MS experiments were performed on several labelled peptides. Figure 6.5 shows the average numbers of labels per residue, or small groups of residues when CID fragmentation was found to be sub-optimal, observed on peptides 211-220 and 221-236. Results revealed that AD-labelling occurred mainly on amino acids located towards the center of the membrane-spanning region (Fig. 6.5) and towards the intracellular side of the channel.

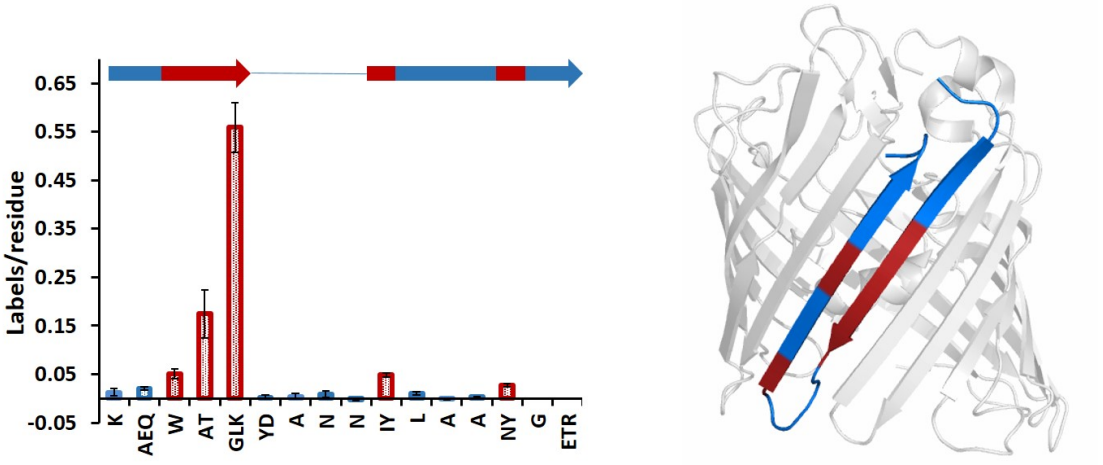


Figure 6.5: Per-residue modification observed on peptides 211-220 and 221-236. Red and blue bars represent respectively labelled and unlabelled residues (left). Cartoon representation of OmpF highlighting labelled (red) and unlabelled (blue) residues. Rendered in pymol from pdb 3K1B (right).

The per-residue modification observed on eight labelled peptides including 211-220 and 221-236 was then mapped on a model of Ompf including the spatial arrangements of the membrane protein with respect to the hydrocarbon core of the lipid bilayer[212,213]. The analysis confirmed that labelled areas were located within or in close proximity of the bilayer (shown in Fig.6.6).

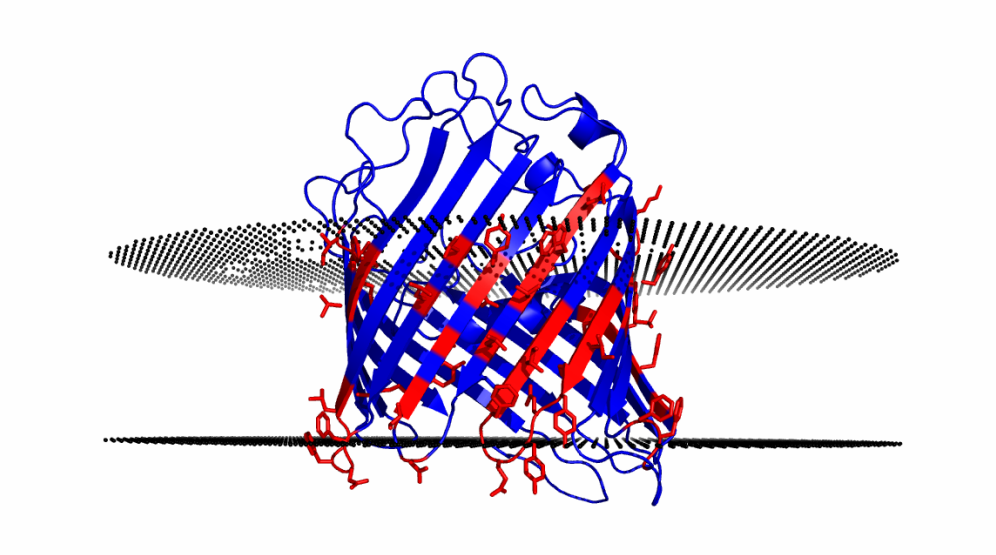


Figure 6.6: Cartoon representation of OmpF highlighting labelled (red) and unlabelled (blue) residues. The model includes the boundaries of a lipid bilayer in which the protein is embedded in physiological conditions. Rendered in Pymol from pdb: 3pox.

These observations strengthened the hypothesis according to which the aryl diazirine probe preferentially partitions in OG micelles and, following photo-activation in conditions favoring the absence of diffusion, labels amino acids located in the hydrophobic core of the micelle. Several studies have shown that the quaternary structure of multi-subunit membrane protein assemblies is retained in OG micelles, making AD an ideal reagent to label the membrane-spanning regions of MPs and gain insight into the contact surface of membrane protein complexes.

6.2.1.3 Carbene footprinting of OmpF in OG micelles using Photoleucine

As described in chapters 1 and 3, Photoleucine has been shown to be able to label soluble proteins but to be significantly less effective compared to AD. Its ability to label membrane proteins in micelles or other systems though has not been previously described. OmpF in OG micelles was again used as a model system to test the efficiency and the selectivity of PLeu by irradiating the protein in the presence of 100 mM of the reagent. The standard protocol of in-gel digestion followed by nano-HPLC-MS analysis was employed to calculate the fractional modification obtained on each identified

peptide. The extent of labelling resulted to be poor with only two peptides showing significant modification (fractional modification > 0.05, Fig. 6.7). Interestingly one of the two labelled peptides, 102-133, did not show any modification when the protein was labelled with AD. When mapped on the structure of OmpF, the peptide resulted to correspond to the solvent-exposed loop located inside the channel which suggests that PLeu can potentially label solvent exposed portions of membrane proteins. The region is protected from contact with detergent molecules of micelles reinforcing the theory of selective AD labelling regions of the protein in proximity to the core of the micelle. Nevertheless Photoleucine proved to be unsuitable for membrane protein foot-printing because of the low labelling efficiency shown and the extremely poor labelling coverage achieved.

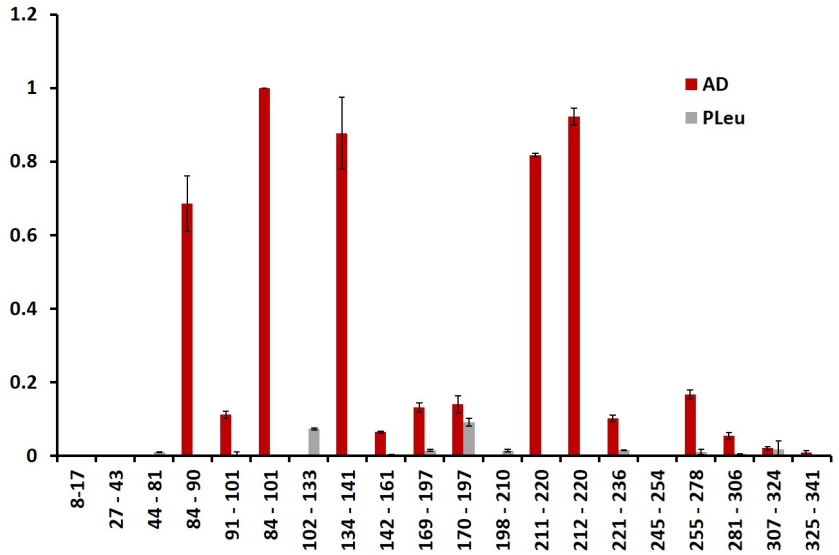


Figure 6.7: Fractional modification observed labelling OmpF in 0.1% OG with 50 mM AD (red) and 100 mM Photoleucine (grey). Only two peptides resulted significantly labelled by PLeu (fractional modification > 0.05).

6.2.2 Characterisation of AD reactivity in OG

6.2.2.1 AD distribution in OG micelles

In order to further investigate the behavior of AD in the presence of OG micelles, the probe was photo-activated at various concentrations in the presence of the detergent

at a concentration over its CMC and ubiquitin, a model soluble protein. As shown in chapter 3, AD labelling of ubiquitin in the presence of 10 mM concentration of the probe results in the insertion of up to 3 molecules of reagent. It was postulated that if the label is sequestered by the detergent micelles then the labelling of ubiquitin should be negatively affected given that the protein is soluble in water and should not be captured in OG micelles. When labelled in the standard conditions used for soluble proteins (77 K, 10 mM AD) in the presence of OG, ubiquitin showed an extent of modification value of 0.16 against the 1.14 obtained in absence of OG (see Fig. 6.8). An increase in concentration of the probe resulted in higher labelling (extent of modification respectively of 0.29 and 0.43 in the presence of 25 mM and 50 mM AD) but still significantly lower than the modification obtained in absence of OG. Thus, it appeared that diazirine, and thus carbene, availability was reduced by the presence of the OG detergent which suggested its capture by micelles.

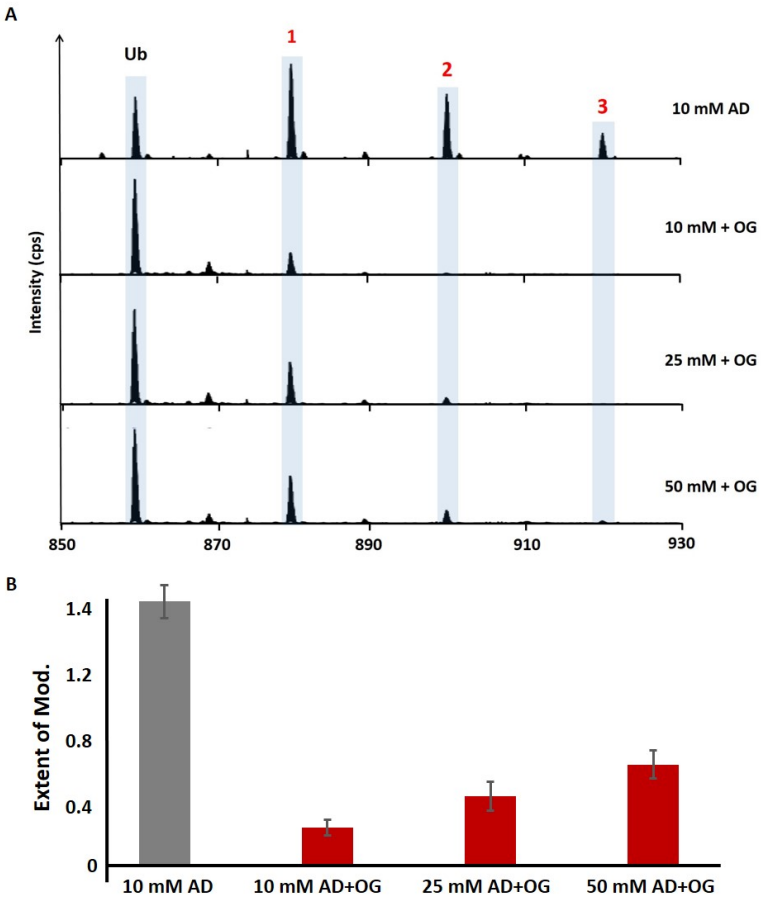


Figure 6.8: (A) ESI mass spectra of the 10+ charge state of ubiquitin photolabelled with 10 mM AD in absence of OG and with increasing concentration of AD in the presence of OG. (B) AD labelling of ubiquitin in standard conditions resulted in an extent of modification of 1.14. Labelling was reduced by the presence of the detergent to 0.16, 0.29 and 0.43 in the presence of OG using respectively 10 mM, 25 mM and 50 mM concentrations of AD.

6.2.2.2 Reaction products of AD in OG solutions

To examine the AD-OG interactions further, the reagent was photo-activated in the presence and absence of the detergent. the resulting mixture was separated by HPLC and the products were identified by MS and UV spectroscopy. As shown in chapter 3, when irradiated in absence of OG, the resulting carbene generates several products originated from the reaction with solvent molecules and with a second unactivated molecule of AD, the latter being the analyte generating the most intense response (shown in Fig. 6.9 A). The product profile significantly changed when the reaction was performed in the presence of OG. The dimer signal was significantly reduced in

the presence of OG and a new set of signal appeared at a retention time centered on five minutes. All the new peaks showed an *m/z* in agreement with the addition of one molecule of AD to OG and the multiplicity suggested different positional isomers of the labelled product (shown in Fig. 6.9 B).

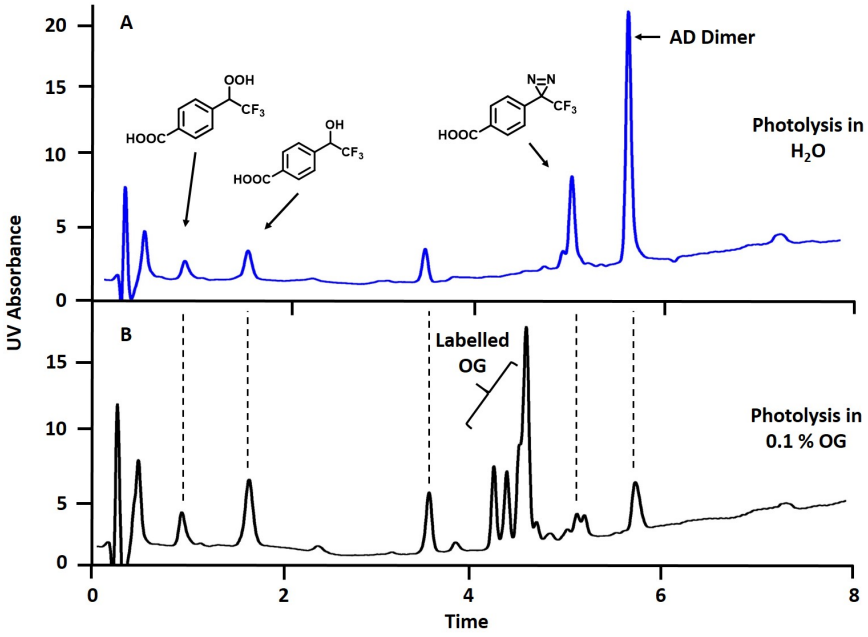


Figure 6.9: HPLC-UV chromatograms of the post-photolysis reaction mixture obtained from laser irradiation of AD in the presence (B) and absence of OG (A). The presence of the detergent lead to a significant decrease of the AD dimer and the appearance of a set of peaks identified as AD labelled OG

The decrease in the dimer abundance and the appearance of intense labelled OG signals suggested that AD preferentially associates with the detergent prior to photolysis rather than other molecules of the probe, which are favored in absence of OG.

In order to gain insight into the location of the label within the micelle, MS/MS experiments on all the different isomers of labelled OG were performed. The modified detergent, detected as a non-covalent proton-bound dimer, was isolated and fragmented by collision induced dissociation resulting in a spectrum dominated by the signal of the protonated AD OG monomer (Fig. 6.10).

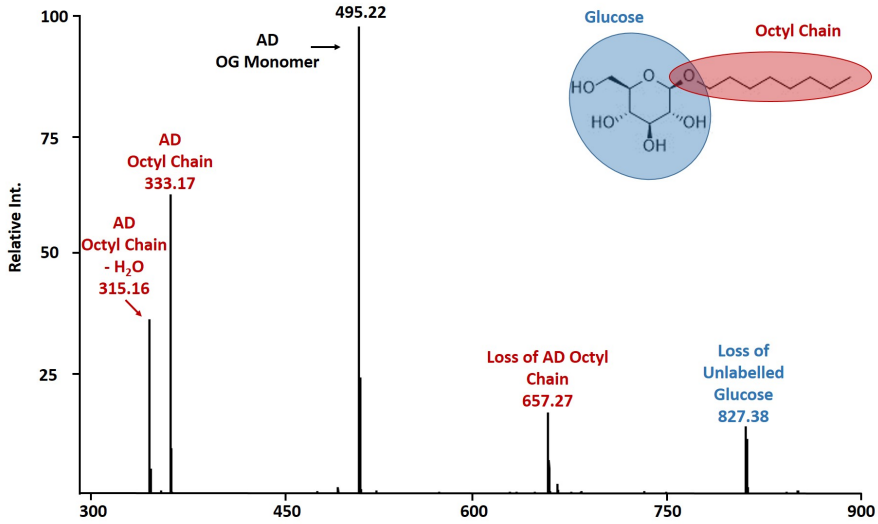


Figure 6.10: CID MS/MS spectrum of *m/z* 989.4 (dimeric labelled OG). The fragmentation pattern suggested reaction of the carbene exclusively with the octyl chain.

CID generated fragments corresponding to the AD labelled octyl chain and its dehydrated version, to the loss of labelled alkyl chain from the dimeric precursor and to the loss of unlabelled glucose. All of the above showed a discrepancy between their observed and theoretical *m/z* $\leq$1.00 ppm. The absence of a signal corresponding to the labelled sugar or loss of labelled sugar from the precursor, in combination with the detection of various labelled versions of the hydrophobic portion of the detergent suggest that the labelling event is occurring exclusively on the core of the micelle. The same pattern was observed for all the AD OG isomers but the absence of different pathways of fragmentation did not allow the location of the different modified positions over the octyl chain.

Formula	Measured m/z	Theoretical m/z	ppm error
$C_{17}H_{22}F_3O_2$	315.1566	315.1566	0.00
$C_{17}H_{24}F_3O_3$	333.1671	333.1672	0.30
$C_{23}H_{34}F_3O_8$	495.2195	495.2200	1.00
$C_{29}H_{44}F_3O_{13}$	657.2723	657.2728	0.76
$C_{40}H_{57}F_6O_2$	827.3791	827.3799	0.96

Table 6.1: Observed and theoretical m/z of CID MS/MS fragments shown in Fig. 6.10.

6.2.3 Conclusions

The work highlighted in this chapter has demonstrated that AD-based carbene footprinting has the potential to reveal the binding interfaces of multi-subunit membrane spanning proteins.

AD labelling of the OG solubilised homotrimeric OmpF porin followed by proteolysis and nanoHPLC-MS analysis revealed modification exclusively located on the membrane spanning region of the protein suggesting partitioning of the reagent in the hydrophobic core of the OG micelles. The absence of labelling on the surface between the subunits of the trimer was attributed to the absence of detergent molecules located at the interface and it was shown to be potentially used to identify such regions.

Labelling experiments on ubiquitin, a model water-soluble protein, in the presence and absence of OG revealed a significant decrease in the extent of modification when the detergent was part of the mixture. This phenomenon suggested reduced availability of the reagent in the presence of OG and confirmed that the probe is sequestered by the micelles before activation.

Further evidence of AD capturing by OG micelles was provided by the identification of the photolysis products of the probe irradiated in solutions of OG detergent and in standard labelling conditions. Several isomers resulting from the addition of AD to OG were detected when the reaction was conducted in OG-containing solution and MS/MS experiments revealed that the modification occurs exclusively on the hydrophobic portion of the molecule.

Taken together, these results strongly suggest that the amphipathic diazirine probe efficiently inserts into the OG micelles. This favourable property explains why the membrane-spanning regions of OmpF, which are expected to be in close contact with the OG micelle are labelled so effectively and the absence of modification on regions located at the interface between subunits. The results clearly show the potential of the use of AD labelling on micelle solubilised-membrane protein complexes in combination

with mass spectrometry detection to gain insight into their quaternary structure. The possibility of labelling IMPs within the membrane makes the technique complementary to other covalent labelling methodologies, such as hydroxyl radical footprinting, and together they present the potential to be used as a more comprehensive structural tool to study the structure of integral membrane proteins.

References

[1] H. Saibil. Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews Molecular Cell Biology*, 14(10):630–642, 2013.

[2] M. Jucker and L. C. Walker. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*, 501(7465):45–51, 2013.

[3] S. B. Prusiner. A unifying role for prions in neurodegenerative diseases. *Prion*, 7: 7–8, 2013.

[4] T. P. Silverstein. The Real Reason Why Oil and Water Don't Mix. *Journal of Chemical Education*, 75(1):116, jan 1998.

[5] A. Wlodawer, W. Minor, Z. Dauter, M. Jaskolski, and B. Physics. HHS Public Access. *FEBS Journal*, 275(1):1–21, 2015.

[6] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature (London, United Kingdom)*, 181:662–666, 1958.

[7] E. Chen. X-ray crystallography - http://260h.pbworks.com/w/page/30814223/X Ray Crystallography.

[8] M. Hong, Y. Zhang, and F. Hu. Membrane protein structure and dynamics from NMR spectroscopy. *Annu. Rev. Phys. Chem.*, 63:1–24, 2012.

[9] S. O. Smith. Structure and activation of the visual pigment rhodopsin. *Annual Review of Biophysics*, 39:309–328, 2010.

[10] D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science (Washington, D. C.)*, 280(5360): 69–77, 1998.

[11] Y. Zhou, J. H. Morals-Cabral, A. Kaufman, and R. MacKinnon. Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 Å resolution. *Nature (London, United Kingdom)*, 414(6859):43–48, 2001.

[12] S. B. Long, X. Tao, E. B. Campbell, and R. MacKinnon. Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. *Nature (London, United Kingdom)*, 450(7168):376–382, 2007.

[13] A. Arora, F. Abildgaard, J. H. Bushweller, and L. K. Tamm. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nature Structural Biology*, 8(4):334–338, 2001.

[14] B. Liang and L. K. Tamm. Structure of outer membrane protein G by solution NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 104(41):16140–16145, 2007.

[15] P. M. Hwang, W.-Y. Choy, E. I. Lo, L. Chen, J. D. Forman-Kay, C. R. H. Raetz, G. G. Prive, R. E. Bishop, and L. E. Kay. Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 99(21):13560–13565, 2002.

[16] T. Wyttenbach and M. T. Bowers. Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment. *J. Phys. Chem. B*, 115(42): 12266–12275, 2011.

[17] O. S. Skinner, A. D. Catherman, B. P. Early, P. M. Thomas, P. D. Compton, and N. L. Kelleher. Fragmentation of Integral Membrane Proteins in the Gas Phase. *Analytical chemistry*, 2015.

[18] L. Konermann, S. Vahidi, and M. A. Sowole. Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules. *Analytical Chemistry*, 86(1):213–232, 2014.

[19] P. W. A. Devine, H. C. Fisher, A. N. Calabrese, F. Whelan, D. R. Higazi, J. R. Potts, D. C. Lowe, S. E. Radford, and A. E. Ashcroft. Investigating the Structural Compaction of Biomolecules Upon Transition to the Gas-Phase Using ESI-TWIMS-MS. *Journal of the American Society for Mass Spectrometry*, page Ahead of Print, 2017.

[20] B. C. Bohrer, S. I. Merenbloom, S. L. Koeniger, A. E. Hilderbrand, and D. E. Clemmer. Biomolecule analysis by ion mobility spectrometry. *Annu. Rev. Anal. Chem.*, 1:293–327, 2008.

[21] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem. (Washington, DC, U. S.)*, 82(22):9557–9565, 2010.

[22] M. F. Bush, I. D. G. Campuzano, and C. V. Robinson. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Anal. Chem. (Washington, DC, U. S.)*, 84(16):7124–7130, 2012.

[23] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, and C. V. Robinson. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.*, 3(7):1139–1152, 2008.

[24] T. Wyttenbach, M. Grabenauer, K. Thalassinos, J. H. Scrivens, and M. T. Bowers. The Effect of Calcium Ions and Peptide Ligands on the Relative Stabilities of the Calmodulin Dumbbell and Compact Structures. *Journal of Physical Chemistry B*, 114(1):437–447, 2010.

[25] M. Jenner, J. Ellis, W. C. Huang, E. Lloyd Raven, G. C. K. Roberts, and N. J. Oldham. Detection of a protein conformational equilibrium by electrospray

ionisation-ion mobility-mass spectrometry. *Angewandte Chemie - International Edition*, 50(36):8291–8294, 2011.

[26] T. W. Knapman, V. L. Morton, N. J. Stonehouse, P. G. Stockley, and A. E. Ashcroft. Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. *Rapid Communications in Mass Spectrometry*, 24 (20):3033–3042, 2010.

[27] J. N. Rabuck, S.-J. Hyung, K. S. Ko, C. C. Fox, M. B. Soellner, and B. T. Ruotolo. Activation State-Selective Kinase Inhibitor Assay Based on Ion Mobility-Mass Spectrometry. *Anal. Chem. (Washington, DC, U. S.)*, 85(15):6995–7002, 2013.

[28] Y. Zhong, S.-J. Hyung, and B. T. Ruotolo. Ion mobility-mass spectrometry for structural proteomics. *Expert Rev. Proteomics*, 9(1):47–58, 2012.

[29] A. Politis, A. Y. Park, S.-J. Hyung, D. Barsky, B. T. Ruotolo, and C. V. Robinson. Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes. *PLoS One*, 5(8):No pp. given, 2010.

[30] S. K. Chowdhury, V. Katta, and B. T. Chait. Probing Conformational-Changes in Proteins by Mass-Spectrometry. *Journal of the American Chemical Society*, 112(24): 9012–9013, 1990.

[31] I. A. Kaltashov, C. E. Bobst, and R. R. Abzalimov. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Science*, 22(5):530–544, 2013.

[32] R. Beveridge, Q. Chappuis, C. MacPhee, and P. Barran. Mass spectrometry methods for intrinsically disordered proteins. *Analyst (Cambridge, U. K.)*, 138(1):32–42, 2013.

[33] L. Testa, S. Brocca, and R. Grandori. Charge-Surface Correlation in Electrospray Ionization of Folded and Unfolded Proteins. *Anal. Chem. (Washington, DC, U. S.)*, 83(17):6459–6463, 2011.

[34] D. A. Shepherd, K. Holmes, D. J. Rowlands, N. J. Stonehouse, and A. E. Ashcroft. Using Ion Mobility Spectrometry-Mass Spectrometry to Decipher the Conformational and Assembly Characteristics of the Hepatitis B Capsid Protein. *Biophysical Journal*, 105(5):1258–1267, 2013.

[35] L. Konermann, S. Vahidi, and M. A. Sowole. Mass spectrometry methods for studying structure and dynamics of biological macromolecules. *Anal Chem*, 86 (1):213–232, 2014.

[36] P. Kebarle and U. H. Verkerk. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.*, 28(6):898–917, 2009.

[37] M. Dole, L. L. Mack, and R. L. Hines. Molecular beams of macroions. *Journal of Chemical Physics*, 49(5):2240–&, 1968.

[38] J. Fernandez de la Mora. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta*, 406 (1):93–104, 2000.

[39] L. Konermann, E. Ahadi, A. D. Rodriguez, and S. Vahidi. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem. (Washington, DC, U. S.)*, 85(1):2–9, 2013.

[40] K. D. Rand, F. W. Lund, S. Amon, and T. J. D. Jørgensen. Investigation of amide hydrogen back-exchange in Asp and His repeats measured by hydrogen (1H/2H) exchange mass spectrometry. *International Journal of Mass Spectrometry*, 302(1-3):110–115, 2011.

[41] J. Ahn, M. C. Jung, K. Wyndham, Y. Q. Yu, and J. R. Engen. Pepsin Immobilized on High-Strength Hybrid Particles for Continuous Flow Online Digestion at 10 000 psi. *Anal. Chem. (Washington, DC, U. S.)*, 84(16):7256–7262, 2012.

[42] L. M. Jones, H. Zhang, I. Vidavsky, and M. L. Gross. Online, High-Pressure Digestion System for Protein Characterization by Hydrogen/Deuterium Exchange

and Mass Spectrometry. *Anal. Chem. (Washington, DC, U. S.)*, 82(4):1171–1174, 2010.

[43] P. G. Fajer, G. M. Bou-Assaf, and A. G. Marshall. Improved Sequence Resolution by Global Analysis of Overlapped Peptides in Hydrogen/Deuterium Exchange Mass Spectrometry. *J. Am. Soc. Mass Spectrom.*, 23(7):1202–1208, 2012.

[44] W. Hu, B. T. Walters, Z.-Y. Kan, L. Mayne, L. E. Rosen, S. Marqusee, and S. W. Englander. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.*, 110(19):7684–7689, S7684/1–S7684/23, 2013.

[45] E. Althaus, S. Canzar, C. Ehrler, M. R. Emmett, A. Karrenbauer, A. G. Marshall, A. Meyer-Base, J. D. Tipton, and H.-M. Zhang. Computing H/D-exchange rates of single residues from data of proteolytic fragments. *BMC Bioinformatics*, 11:424, 2010.

[46] Z.-Y. Kan, B. T. Walters, L. Mayne, and S. W. Englander. Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. *Proc. Natl. Acad. Sci. U. S. A.*, 110(41):16438–16443, 2013.

[47] M. Modzel, P. Stefanowicz, and Z. Szewczuk. Hydrogen scrambling in non-covalent complexes of peptides. *Rapid Communications in Mass Spectrometry*, 26 (23):2739–2744, 2012.

[48] L. Konermann, J. Pan, and Y.-H. Liu. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society reviews*, 40(3):1224–1234, 2011.

[49] A. J. Percy, M. Rey, K. M. Burns, and D. C. Schriemer. Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-A review. *Anal. Chim. Acta*, 721:7–21, 2012.

[50] M. J. Chalmers, S. A. Busby, B. D. Pascal, G. M. West, and P. R. Griffin. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert Rev. Proteomics*, 8(1):43–59, 2011.

[51] G. Stjepanovic, K. Kapp, G. Bange, C. Graf, R. Parlitz, K. Wild, M. P. Mayer, and I. Sinning. Lipids Trigger a Conformational Switch That Regulates Signal Recognition Particle (SRP)-mediated Protein Targeting. *J. Biol. Chem.*, 286(26): 23489–23497, 2011.

[52] T. R. Keppel, B. A. Howard, and D. D. Weis. Mapping Unstructured Regions and Synergistic Folding in Intrinsically Disordered Proteins with Amide H/D Exchange Mass Spectrometry [Erratum to document cited in CA155:477933]. *Biochemistry*, 51(39):7812, 2012.

[53] D. Balasubramaniam and E. A. Komives. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim. Biophys. Acta, Proteins Proteomics*, 1834(6):1202–1209, 2013.

[54] M. Rey, E. Forest, and L. Pelosi. Exploring the Conformational Dynamics of the Bovine ADP/ATP Carrier in Mitochondria. *Biochemistry*, 51(48):9727–9735, 2012.

[55] S. Mehmood, C. Domene, E. Forest, and J.-M. Jault. Dynamics of a bacterial multidrug ABC transporter in the inward- and outward-facing conformations. *Proc. Natl. Acad. Sci. U. S. A.*, 109(27):10832–10836, S10832/1–S10832/11, 2012.

[56] V. L. Mendoza and R. W. Vachet. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews*, 28(5):785–815, 2009.

[57] T. Orban, S. Gupta, K. Palczewski, and M. R. Chance. Visualizing water molecules in transmembrane proteins using radiolytic labeling methods. *Biochemistry*, 49(5):827–834, 2010.

[58] O. T. Akinsiku, E. T. Yu, and D. Fabris. Mass spectrometric investigation of protein alkylation by the RNA footprinting probe kethoxal. *Journal of Mass Spectrometry*, 40(10):1372–1381, 2005.

[59] G. J. Carven and L. J. Stern. Probing the Ligand-Induced Conformational Change in HLA-DR1 by Selective Chemical Modification and Mass Spectrometric Mapping. *Biochemistry*, 44(42):13625–13637, 2005.

[60] C. Hager-Braun and K. B. Tomer. Characterization of the Tertiary Structure of Soluble CD4 Bound to Glycosylated Full-Length HIV gp120 by Chemical Modification of Arginine Residues and Mass Spectrometric Analysis. *Biochemistry*, 41 (6):1759–1766, 2002.

[61] R. E. Iacob, Z. Keck, O. Olson, S. K. H. Foung, and K. B. Tomer. Structural elucidation of critical residues involved in binding of human monoclonal antibodies to hepatitis C virus E2 envelope glycoprotein. *Biochimica et Biophysica Acta, Proteins and Proteomics*, 1784(3):530–542, 2008.

[62] S. Akashi, M. Shirouzu, T. Terada, Y. Ito, S. Yokoyama, and K. Takio. Characterization of the structural difference between active and inactive forms of the Ras protein by chemical modification followed by mass spectrometric peptide mapping. *Analytical Biochemistry*, 248(1):15–25, 1997.

[63] S. Izumi, H. Kaneko, T. Yamazaki, T. Hirata, and S. Kominami. Membrane Topology of Guinea Pig Cytochrome P450 17α Revealed by a Combination of Chemical Modifications and Mass Spectrometry. *Biochemistry*, 42(49):14663–14669, 2003.

[64] A. Gonzalez de Peredo, C. Saint-Pierre, A. Adrait, L. Jacquamet, J.-M. Latour, I. Michaud-Soret, and E. Forest. Identification of the Two Zinc-Bound Cysteines in the Ferric Uptake Regulation Protein from Escherichia coli: Chemical Modification and Mass Spectrometry Analysis. *Biochemistry*, 38(26):8582–8589, 1999.

[65] H. P. Happersberger, M. Przybylski, and M. O. Glocker. Selective bridging of biscysteinyl residues by arsonous acid derivatives as an approach to the characteri-

zation of protein tertiary structures and folding pathways by mass spectrometry. *Analytical Biochemistry*, 264(2):237–250, 1998.

[66] S. Barbirz, U. Jakob, and M. O. Glocker. Mass spectrometry unravels disulfide bond formation as the mechanism that activates a molecular chaperone. *Journal of Biological Chemistry*, 275(25):18759–18766, 2000.

[67] B. C. Gau. The Advancement of Mass Spectrometry-based Hydroxyl Radical Protein Footprinting: Application of Novel Analysis Methods to Model Proteins and Apolipoprotein E. *All Theses and Dissertations*, PhD(January), 2011.

[68] C. C. Jumper, R. Bomgarden, J. Rogers, C. Etienne, and D. C. Schriemer. High-resolution mapping of carbene-based protein footprints. *Analytical Chemistry*, 84 (10):4411–4418, 2012.

[69] D. M. Hambly and M. L. Gross. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *Journal of the American Society for Mass Spectrometry*, 16(12):2057–2063, 2005.

[70] B. C. Gau, J. S. Sharp, D. L. Rempel, and M. L. Gross. Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Analytical Chemistry*, 81 (16):6563–6571, 2009.

[71] T. G. Watkinson, A. N. Calabrese, J. R. Ault, S. E. Radford, and A. E. Ashcroft. FPOP-LC-MS/MS Suggests Differences in Interaction Sites of Amphipols and Detergents with Outer Membrane Proteins. *Journal of The American Society for Mass Spectrometry*, pages 35–40, 2016.

[72] S. D. Maleknia and K. M. Downard. Advances in radical probe mass spectrometry for protein footprinting in chemical biology applications. *Chemical Society reviews*, 43(10):3244–58, 2014.

[73] T. D. Tullius and B. A. Dombroski. Hydroxyl radical footprinting - High-resolution information about DNA protein contacts and application to lambda-

repressor and CRO protein. *Proceedings of the National Academy of Sciences of the United States of America*, 83(15):5469–5473, 1986.

[74] S. H. Shanblatt and A. Revzin. Interactions of the catabolite activator protein (CAP) at the galactose and lactose promoters of Escherichia coli probed by hydroxyl radical footprinting. The second CAP molecule which binds at gal and the one CAP at lac may act to stimulate transcriptio. *Journal of Biological Chemistry*, 262(24):11422–11427, 1987.

[75] L. Wang and M. R. Chance. Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Analytical chemistry*, 83(19):7234–41, 2011.

[76] B. Sclavi, S. Woodson, M. Sullivan, M. Chance, and M. Brenowitz. Following the folding of RNA with time-resolved synchrotron X-ray footprinting. In G. K. Ackers and M. L. Johnson, editors, *Energetics of Biological Macromolecules, Pt B*, volume 295, pages 379–402. 1998.

[77] K. D. Asmus. Pulse radiolysis methodology. *Methods in Enzymology*, 105(Oxygen Radicals Biol. Syst.):167–178, 1984.

[78] J. J. Hayes, L. Kam, and T. D. Tullius. FOOTPRINTING PROTEIN-DNA COMPLEXES WITH GAMMA-RAYS. *Methods in Enzymology*, 186:545–549, 1990.

[79] S. D. Maleknia, M. R. Chance, and K. M. Downard. Electrospray-assisted modification of proteins: a radical probe of protein structure. *Rapid Communications in Mass Spectrometry*, 13(23):2352–2358, 1999.

[80] C. Y. Ralston, B. Sclavi, M. Sullivan, M. L. Deras, S. A. Woodson, M. R. Chance, and M. Brenowitz. Time-resolved synchrotron X-ray footprinting and its application to RNA folding. *Rna-Ligand Interactions Pt A*, 317:353–368, 2000.

[81] K. Takamoto and M. R. Chance. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annual Review of Biophysics and Biomolecular Structure*, 35:251–276, 2006.

[82] G. H. Xu and M. R. Chance. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews*, 107(8):3514–3543, 2007.

[83] J. W. H. Wong, S. D. Maleknia, and K. M. Downard. Hydroxyl radical probe of the calmodulin-melittin complex interface by electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 16(2):225–233, 2005.

[84] B. C. Gau, J. S. Sharp, D. L. Rempel, and M. L. Gross. Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Analytical Chemistry*, 81 (16):6563–6571, 2009.

[85] A. N. Calabrese, J. R. Ault, S. E. Radford, and A. E. Ashcroft. Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. *Methods (Amsterdam, Netherlands)*, 89:38–44, 2015.

[86] K. S. Li, D. L. Rempel, and M. L. Gross. Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1-42 Aggregation. *Journal of the American Chemical Society*, 138(37):12090–12098, 2016.

[87] Y. Lu, H. Zhang, D. M. Niedzwiedzki, J. Jiang, R. E. Blankenship, and M. L. Gross. Fast Photochemical Oxidation of Proteins Maps the Topology of Intrinsic Membrane Proteins: Light-Harvesting Complex 2 in a Nanodisc. *Analytical Chemistry*, 88(17):8827–8834, 2016.

[88] J. A. Espino, V. S. Mali, and L. M. Jones. In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Analytical Chemistry*, 87(15):7971–7978, 2015.

[89] S. Vahidi and L. Konermann. Probing the Time Scale of FPOP (Fast Photochemical Oxidation of Proteins): Radical Reactions Extend Over Tens of Milliseconds. *Journal of the American Society for Mass Spectrometry*, 27(7):1156–1164, 2016.

[90] J. R. Pliego Jr. and W. B. De Almeida. A New Mechanism for the Reaction of Carbenes with OH Groups. *Journal of Physical Chemistry A*, 103(20):3904–3909, 1999.

[91] D. Robinette, N. Neamati, K. B. Tomer, and C. H. Borchers. Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics. *Expert Review of Proteomics*, 3(4):399–408, 2006.

[92] F. Kotzyba-Hibert, I. Kapfer, and M. Goeldner. Recent trends in photoaffinity labeling. *Angewandte Chemie, International Edition in English*, 34(12):1296–1312, 1995.

[93] M. Hashimoto and Y. Hatanaka. Recent progress in diazirine-based photoaffinity labeling. *European Journal of Organic Chemistry*, (15):2513–2523, 2008.

[94] J. Das. Aliphatic diazirines as photoaffinity probes for proteins: Recent developments. *Chemical Reviews*, 111(8):4405–4417, 2011.

[95] A. B. Kumar. *Design, Synthesis and Evaluation of Novel Diazirine Photolabels with Improved Ambient Light Stability and Fluorous-Based Enrichment Capacity.* PhD thesis, University of South Florida, 2012.

[96] A. B. Kumar, J. M. Anderson, and R. Manetsch. Design, synthesis and photoactivation studies of fluorous photolabels. *Organic & Biomolecular Chemistry*, 9(18): 6284–6292, 2011.

[97] J. Brunner, H. Senn, and F. M. Richards. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *Journal of Biological Chemistry*, 255(8):3313–3318, 1980.

[98] R. Bonneau and M. T. H. Liu. Quantum Yield of Formation of Diazo Compounds from the Photolysis of Diazirines. *Journal of the American Chemical Society*, 118(30): 7229–7230, 1996.

[99] T. Wakahara, Y. Niino, T. Kato, Y. Maeda, T. Akasaka, M. T. H. Liu, K. Kobayashi, and S. Nagase. A Nonspectroscopic Method To Determine the Photolytic Decomposition Pathways of 3-Chloro-3-alkyldiazirine: Carbene, Diazo and Rearrangement in Excited State. *Journal of the American Chemical Society*, 124(32):9465–9468, 2002.

[100] H. Bayley and J. R. Knowles. Photogenerated reagents for membrane labeling. 2. Phenylcarbene and adamantylidene formed within the lipid bilayer. *Biochemistry*, 17(12):2420–2423, 1978.

[101] R. A. G. Smith and J. R. Knowles. Aryldiazirines. Potential reagents for photolabeling of biological receptor sites. *Journal of the American Chemical Society*, 95(15): 5072–5073, 1973.

[102] R. A. G. Smith and J. R. Knowles. Preparation and photolysis of 3-aryl-3H-diazirines. *Journal of the Chemical Society-Perkin Transactions 2*, (7):686–694, 1975.

[103] R. R. Lucchese and H. F. Schaefer III. Extensive configuration interaction studies of the methylene singlet-triplet separation. *Journal of the American Chemical Society*, 99(20):6765–6766, 1977.

[104] C. A. Richards Jr., S.-J. Kim, Y. Yamaguchi, and H. F. Schaefer III. Dimethylcarbene: A Singlet Ground State?. *Journal of the American Chemical Society*, 117(40): 10104–10107, 1995.

[105] M. J. S. Dewar, R. C. Haddon, and P. K. Weiner. MINDO [modified intermediate neglect of differential overlap]/3 study of the electronic states of methylene. *Journal of the American Chemical Society*, 96(1):253–255, 1974.

[106] R. K. Lengel and R. N. Zare. Experimental determination of the singlet-triplet splitting in methylene. *Journal of the American Chemical Society*, 100(24):7495–7499, 1978.

[107] W. T. Borden, N. P. Gritsan, C. M. Hadad, W. L. Karney, C. R. Kemnitz, and M. S. Platz. The Interplay of Theory and Experiment in the Study of Phenylnitrene. *Accounts of Chemical Research*, 33(11):765–771, 2000.

[108] M. Nassal. 4'-(1-Azi-2,2,2-trifluoroethyl)phenylalanine, a photolabile carbene-generating analog of phenylalanine. *Journal of the American Chemical Society*, 106 (24):7540–7545, 1984.

[109] J. Brunner. Photochemical labeling of apolar phase of membranes. *Methods in Enzymology*, 172(Biomembranes, Pt. S):628–687, 1989.

[110] T. Weber and J. Brunner. 2-(Tributylstannyl)-4-[3-(trifluoromethyl)-3H-diazirin-3-yl]benzyl Alcohol: A Building Block for Photolabeling and Crosslinking Reagents of Very High Specific Radioactivity. *Journal of the American Chemical Society*, 117(11):3084–3095, 1995.

[111] J. Das. Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments. *Chemical Reviews (Washington, DC, United States)*, 111(8):4405–4417, 2011.

[112] F. M. Richards, R. Lamed, R. Wynn, D. Patel, and G. Olack. Methylene as a possible universal footprinting reagent that will include hydrophobic surface areas: Overview and feasibility: Properties of diazirine as a precursor. *Protein Science*, 9 (12):2506–2517, 2000.

[113] N. J. Turro, Y. Cha, and I. R. Gould. Reactivity and intersystem crossing of singlet methylene in solution. *Journal of the American Chemical Society*, 109(7):2101–2107, 1987.

[114] P. O. Craig, D. B. Ureta, and J. M. Delfino. Probing protein conformation with a minimal photochemical reagent. *Protein Sci*, 11(6):1353–1366, 2002.

[115] G. E. Gómez, A. Cauerhff, P. O. Craig, F. A. Goldbaum, and J. M. Delfino. Exploring protein interfaces with a general photochemical reagent. *Protein science : a publication of the Protein Society*, 15(4):744–52, 2006.

[116] D. B. Ureta, P. O. Craig, G. E. Gómez, and J. M. Delfino. Assessing native and non-native conformational states of a protein by methylene carbene labeling: The case of Bacillus licheniformis β-lactamase. *Biochemistry*, 46(50):14567–14577, 2007.

[117] G. E. Gómez, M. R. Mundo, P. O. Craig, and J. M. Delfino. Probing protein surface with a solvent mimetic carbene coupled to detection by mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 23(1):30–42, 2012.

[118] G. E. Gómez, J. L. E. Monti, M. R. Mundo, and J. M. Delfino. Solvent Mimicry with Methylene Carbene to Probe Protein Topography. *Analytical Chemistry*, 87 (19):10080–10087, 2015.

[119] C. E. McKenna, A. M. Simeonov, H. Eran, and M. Bravo-Leerabhandh. Reduction of cyclic and acyclic diazene derivates by Azotobacter vinelandii nitrogenase: diazirine and trans-dimethyldiazene. *Biochemistry*, 35(14):4502–4514, 1996.

[120] M. Suchanek, A. Radzikowska, and C. Thiele. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nature Methods*, 2(4):261–267, 2005.

[121] C. C. Jumper and D. C. Schriemer. Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Analytical Chemistry*, 83(8):2913–2920, 2011.

[122] Z. Zhang and R. W. Vachet. Kinetics of Protein Complex Dissociation Studied by Hydrogen/Deuterium Exchange and Mass Spectrometry. *Analytical Chemistry*, 87(23):11777–11783, 2015.

[123] J. J. Thomson. *Rays of Positive Electricity and their Application to Chemical Analyses*. Number Copyright (C) 2017 American Chemical Society (ACS). All Rights Reserved. Longmans, 1913.

[124] G. Squires. Francis Aston and the mass spectrograph. *J. Chem. Soc., Dalton Trans.*, (23):3893–3900, 1998.

[125] E. De Hoffmann Stroobant V. *Mass spectrometry: principles and applications*. Wiley, 3 edition, 2007.

[126] J. B. Fenn. Electrospray wings for molecular elephants (Nobel lecture). *Angew Chem Int Ed Engl*, 42(33):3871–3894, 2003.

[127] J. A. Loo. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews*, 16(1):1–23, 1997.

[128] J. Fenn. Electrospray ionisation for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, 1989.

[129] G. Taylor. Disintegration of Water Drops in an Electric Field. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 280(1382):383, 1964.

[130] J. H. Gross. *Mass spectrometry: a textbook*. Springer, 2 edition, 2004.

[131] A. Gomez and K. Tang. Charge and fission of droplets in electrostatic sprays. *Phys. Fluids*, 6(1):404–414, 1994.

[132] J. V. Iribarne and B. A. Thomson. Evaporation of small ions from charged droplets. *Journal of Chemical Physics*, 64(6):2287–2294, 1976.

[133] S. Nguyen and J. B. Fenn. Gas-phase ions of solute species from charged droplets of solutions. *Proc. Natl. Acad. Sci. U. S. A.*, 104(4):1111–1117, 2007.

[134] M. Gamero-Castano and J. Fernandez de la Mora. Direct measurement of ion evaporation kinetics from electrified liquid surfaces. *J. Chem. Phys.*, 113(2):815–832, 2000.

[135] B. E. Winger, K. J. Light-Wahl, R. R. O. Loo, H. R. Udseth, and R. D. Smith. Observation and implications of high mass-to-charge ratio ions from electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.*, 4(7):536–545, 1993.

[136] E. D. Hoffmann and V. Stroobant. *Mass Spectrometry - Priniples and Applications.*, volume 29. Wiley, 2007.

[137] V. H. Wysocki, G. Tsaprailis, L. L. Smith, and L. A. Breci. Mobile and localized protons: A framework for understanding peptide dissociation. *Journal of Mass Spectrometry*, 35(12):1399–1406, 2000.

[138] R. S. Johnson, S. A. Martin, and K. Biemann. Collision-induced fragmentation of (M + H)+ ions of peptides. Side chain specific sequence ions. *Int. J. Mass Spectrom. Ion Processes*, 86:137–154, 1988.

[139] X. Tang and R. K. Boyd. An investigation of fragmentation mechanisms of doubly protonated tryptic peptides. *Rapid Commun. Mass Spectrom.*, 6(11):651–657, 1992.

[140] A. L. McCormack, A. Somogyi, A. R. Dongre, and V. H. Wysocki. Fragmentation of protonated peptides: surface-induced dissociation in conjunction with a quantum mechanical approach. *Anal. Chem.*, 65(20):2859–2872, 1993.

[141] A. Somogyi, V. H. Wysocki, and I. Mayer. The effect of protonation site on bond strengths in simple peptides: application of Ab initio and modified neglect of differential overlap bond orders and modified neglect of differential overlap energy partitioning. *J. Am. Soc. Mass Spectrom.*, 5(8):704–717, 1994.

[142] K. A. Cox, S. J. Gaskell, M. Morris, and A. Whiting. Role of the site of protonation in the low-energy decompositions of gas-phase peptide ions. *J. Am. Soc. Mass Spectrom.*, 7(6):522–531, 1996.

[143] B. Palzs and S. Suhal. Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews*, 24(4):508–548, 2005.

[144] A. R. Dongre, J. L. Jones, A. Somogyi, and V. H. Wysocki. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.*, 118(35): 8365–8374, 1996.

[145] T. Yalcin, C. Khouw, I. G. Csizmadia, M. R. Peterson, and A. G. Harrison. Why are B ions stable species in peptide spectra? *J. Am. Soc. Mass Spectrom.*, 6(12): 1165–1174, 1995.

[146] T. Yalcin, I. G. Csizmadia, M. R. Peterson, and A. G. Harrison. The structure and fragmentation of B n (nâĽě3) ions in peptide spectra. *J Am Soc Mass Spectrom*, 7 (3):233–242, 1996.

[147] M. J. Nold, C. Wesdemiotis, T. Yalcin, and A. G. Harrison. Amide bond dissociation in protonated peptides. Structures of the N-terminal ionic and neutral fragments. *Int. J. Mass Spectrom. Ion Processes*, 164(1,2):137–153, 1997.

[148] M. J. Polce, D. Ren, and C. Wesdemiotis. Dissociation of the peptide bond in protonated peptides. *J. Mass Spectrom.*, 35(12):1391–1398, 2000.

[149] B. Paizs, S. Suhai, B. Hargittai, V. J. Hruby, and A. Somogyi. Ab initio and MS/MS studies on protonated peptides containing basic and acidic amino acid residues I. Solvated proton vs. salt-bridged structures and the cleavage of the terminal amide bond of protonated RD-NH2. *Int. J. Mass Spectrom.*, 219(1):203–232, 2002.

[150] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, and M. Mann. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols*, 1(6):2856–2860, 2006.

[151] K. D. Rand, S. D. Pringle, M. Morris, J. R. Engen, and J. M. Brown. ETD in a traveling wave ion guide at tuned Z-spray ion source conditions allows for site-specific hydrogen/deuterium exchange measurements. *J. Am. Soc. Mass Spectrom.*, 22(10): 1784–1793, 2011.

[152] C. C. Jumper, R. Bomgarden, J. Rogers, C. Etienne, and D. C. Schriemer. High-Resolution Mapping of Carbene-Based Protein Footprints. *Analytical Chemistry*, 84(10):4411–4418, 2012.

[153] J. A. Espino, V. S. Mali, and L. M. Jones. In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Analytical Chemistry*, 87(15):7971–7978, 2015.

[154] Y. Yan, G. Chen, H. Wei, R. Y. C. Huang, J. Mo, D. L. Rempel, A. A. Tymiak, and M. L. Gross. Fast photochemical oxidation of proteins (FPOP) maps the epitope

of EGFR binding to adnectin. *Journal of the American Society for Mass Spectrometry*, 25(12):2084–2092, 2014.

[155] J. Brunner, H. Senn, and F. M. Richards. "3-trifluoromethyl-3-phenyldiazirine - A new carbene generating group for photolabelling reagents. *Journal of Biological Chemistry*, 255(8):3313–3318, 1980.

[156] A. Admasu, A. D. Gudmundsdottir, M. S. Platz, D. S. Watt, S. Kwiatkowski, and P. J. Crocker. A laser flash photolysis study of p-tolyl(trifluoromethyl)carbene. *Journal of the Chemical Society, Perkin Transactions 2: Physical Organic Chemistry*, (5):1093–1100, 1998.

[157] C. C. Blake, L. N. Johnson, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma. Crystallographic Studies of the Activity of Hen Egg-White Lysozyme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 167(1009): 378–388, 1967.

[158] R. B. Von Dreele. Binding of N-acetylglucosamine oligosaccharides to hen egg-white lysozyme: a powder diffraction study. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, D61(1):22–32, 2005.

[159] M. Schindler, Y. Assaf, N. Sharon, and D. M. Chipman. Mechanism of lysozyme catalysis: role of ground-state strain in subsite D in hen egg-white and human lysozymes. *Biochemistry*, 16(3):423–431, 1977.

[160] L. Manzi, A. S. Barrow, D. Scott, R. Layfield, T. G. Wright, J. E. Moses, and N. J. Oldham. Carbene footprinting accurately maps binding sites in protein-ligand and protein-protein interactions. *Nature Communications*, 7:13288, 2016.

[161] A. J. Cross and G. R. Fleming. Influence of inhibitor binding on the internal motions of lysozyme. *Biophysical journal*, 50(3):507–512, 1986.

[162] S. Elsasser and D. Finley. Delivery of ubiquitinated substrates to protein-unfolding machines. *Nature Cell Biology*, 7(8):742–749, 2005.

[163] O. Staub and D. Rotin. Role of ubiquitylation in cellular membrane transport. *Physiological reviews*, 86(2):669–707, 2006.

[164] G. L. Dianov, C. Meisenberg, and J. L. Parsons. Regulation of DNA repair by ubiquitylation. *Biochemistry (Moscow)*, 76(1):69–79, 2011.

[165] V. Chau, J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, and A. Varshavsky. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science (Washington, DC, United States)*, 243(4898):1576–1583, 1989.

[166] L. Sun and Z. J. Chen. The novel functions of ubiquitination in signaling. *Current Opinion in Cell Biology*, 16(2):119–126, 2004.

[167] L. Hicke and R. Dunn. Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annual Review of Cell and Developmental Biology*, 19:141–172, 4 plates, 2003.

[168] J. Spence, R. R. Gali, G. Dittmar, F. Sherman, M. Karin, and D. Finley. Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell (Cambridge, Massachusetts)*, 102(1):67–76, 2000.

[169] D. Komander. The emerging complexity of protein ubiquitination. *Biochemical Society Transactions*, 37(5):937–953, 2009.

[170] C. Heride, S. Urbe, and M. J. Clague. Ubiquitin code assembly and disassembly. *Current Biology*, 24(6):R215–R220, 2014.

[171] F. E. Reyes-Turcu, J. R. Shanks, D. Komander, and K. D. Wilkinson. Recognition of polyubiquitin isoforms by the multiple ubiquitin binding modules of isopeptidase T. *Journal of Biological Chemistry*, 283(28):19581–19592, 2008.

[172] S. Dayal, A. Sparks, J. Jacob, N. Allende-Vega, D. P. Lane, and M. K. Saville. Suppression of the Deubiquitinating Enzyme USP5 Causes the Accumulation of

Unanchored Polyubiquitin and the Activation of p53. *J. Biol. Chem.*, 284(8):5030–5041, 2009.

[173] Z.-P. Xia, L. Sun, X. Chen, G. Pineda, X. Jiang, A. Adhikari, W. Zeng, and Z. J. Chen. Direct activation of protein kinases by unanchored polyubiquitin chains. *Nature (London, U. K.)*, 461(7260):114–119, 2009.

[174] G. V. Avvakumov, J. R. Walker, S. Xue, A. Allali-Hassani, A. Asinas, U. B. Nair, X. Fang, X. Zuo, Y.-X. Wang, K. D. Wilkinson, and S. Dhe-Paganon. Two ZnF-UBP Domains in Isopeptidase T (USP5). *Biochemistry*, 51(6):1188–1198, 2012.

[175] F. E. Reyes-Turcu, J. R. Horton, J. E. Mullally, A. Heroux, X. D. Cheng, and K. D. Wilkinson. The ubiquitin binding domain ZnFUBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell*, 124(6):1197–1208, 2006.

[176] D. Scott, R. Layfield, and N. J. Oldham. Ion mobility-mass spectrometry reveals conformational flexibility in the deubiquitinating enzyme USP5. *Proteomics*, 15 (16):2835–41, 2015.

[177] D. J. Newman and G. M. Cragg. Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.*, 70(3):461–477, 2007.

[178] D. J. Newman, G. M. Cragg, and K. M. Snader. The influence of natural products upon drug discovery (Antiquity to late 1999). *Nat. Prod. Rep.*, 17(3):215–234, 2000.

[179] C. Hertweck. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem., Int. Ed.*, 48(26):4688–4716, 2009.

[180] C. Hertweck. Hidden biosynthetic treasures brought to light. *Nat. Chem. Biol.*, 5 (7):450–452, 2009.

[181] S. Smith and S.-C. Tsai. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.*, 24(5):1041–1072, 2007.

[182] J. N. Copp and B. A. Neilan. The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria. *Appl. Environ. Microbiol.*, 72(4):2298–2305, 2006.

[183] C. Khosla. Structures and mechanisms of polyketide synthases. *J. Org. Chem.*, 74 (17):6416–6420, 2009.

[184] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, and J. Piel. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.*, 26(2):225–233, 2008.

[185] J. Staunton and K. J. Weissman. Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep.*, 18(4):380–416, 2001.

[186] B. Shen. Polyketide biosynthesis beyond the type I, II, and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, 7:285–295, 2003.

[187] Y.-Q. Cheng, G.-L. Tang, and B. Shen. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.*, 100(6):3149–3154, 2003.

[188] E. J. N. Helfrich and J. Piel. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.*, 33(2):231–316, 2016.

[189] R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, and K. J. Weissman. The Structure of Docking Domains in Modular Polyketide Synthases. *Chem. Biol.*, 10(8):723–731, 2003.

[190] R. S. Gokhale and C. Khosla. Role of linkers in communication between protein modules. *Curr. Opin. Chem. Biol.*, 4(1):22–27, 2000.

[191] T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith, and D. H. Sherman. Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chem. Biol.*, 4(1):41–52, 2009.

[192] L. Katz. The DEBS paradigm for type I modular polyketide synthases and beyond. *Methods Enzymol.*, 459(Complex Enzymes in Microbial Natural Product Biosynthesis, Part B):113–142, 2009.

[193] T. A. M. Gulder, M. F. Freeman, and J. Piel. The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules. *Top Curr Chem*, 2011.

[194] J. Dorival, T. Annaval, F. Risser, S. Collin, P. Roblin, C. Jacob, A. Gruez, B. Chagot, and K. J. Weissman. Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.*, 138 (12):4155–4167, 2016.

[195] J. Zeng, D. T. Wagner, Z. Zhang, L. Moretto, J. D. Addison, and A. T. Keatinge-Clay. Portability and Structure of the Four-Helix Bundle Docking Domains of trans-Acyltransferase Modular Polyketide Synthases. *ACS Chem. Biol.*, 11(9): 2466–2474, 2016.

[196] S. Kapur, A. Worthington, Y. Tang, D. E. Cane, M. D. Burkart, and C. Khosla. Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase. *Bioorganic & Medicinal Chemistry Letters*, 18(10): 3034–3038, 2008.

[197] A. S. Worthington, H. Rivera, J. W. Torpey, M. D. Alexander, and M. D. Burkart. Mechanism-Based Protein Cross-Linking Probes To Investigate Carrier Protein-Mediated Biosynthesis. *ACS Chemical Biology*, 1(11):687–691, 2006.

[198] F. Krause. Detection and analysis of protein-protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (membrane) protein complexes and supercomplexes. *Electrophoresis*, 27(13):2759–2781, 2006.

[199] J. Pan, M. J. Misamore, Q. Wang, and W. J. Snell. Protein transport and signal transduction during fertilization in Chlamydomonas. *Traffic (Oxford, United Kingdom)*, 4(7):452–459, 2003.

[200] D. Grimm, J. Bauer, J. Pietsch, M. Infanger, J. Eucker, C. Eilles, and J. Schoenberger. Diagnostic and therapeutic use of membrane proteins in cancer cells. *Current Medicinal Chemistry*, 18(2):176–190, 2011.

[201] A. M. Seddon, P. Curnow, and P. J. Booth. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta, Biomembranes*, 1666 (1-2):105–117, 2004.

[202] A. J. Rice, F. J. D. Alvarez, A. L. Davidson, and H. W. Pinkett. Effects of lipid environment on the conformational changes of an ABC importer. *Channels (Austin, Tex.)*, 8(4):327–333, 2014.

[203] A. Laganowsky, E. Reading, J. T. S. Hopper, and C. V. Robinson. Mass spectrometry of intact membrane protein complexes. *Nature Protocols*, 8(4):639–651, 2013.

[204] Y. Pan, H. Piyadasa, J. D. O'Neil, and L. Konermann. Conformational Dynamics of a Membrane Transport Protein Probed by H/D Exchange and Covalent Labeling: The Glycerol Facilitator. *J. Mol. Biol.*, 416(3):400–413, 2012.

[205] A. Khanal, Y. Pan, L. S. Brown, and L. Konermann. Pulsed hydrogen/deuterium exchange mass spectrometry for time-resolved membrane protein folding studies. *Journal of Mass Spectrometry*, 47(12):1620–1626, 2012.

[206] C. R. Morgan, C. M. Hebling, K. D. Rand, D. W. Stafford, J. W. Jorgenson, and J. R. Engen. Conformational transitions in the membrane scaffold protein of phospholipid bilayer nanodiscs. *Molecular and Cellular Proteomics*, 10(9):M111.010876, 11 pp., 2011.

[207] Y. Pan. (1609) Structural Characterization Of An Integral Membrane Protein By Oxidative Methionine Labeling And Mass Spectrometry. *Analytical chemistry*, 118 (1):2009, 2009.

[208] Y. Zhu, T. Guo, J. E. Park, X. Li, W. Meng, A. Datta, M. Bern, S. K. Lim, and S. K. Sze. Elucidating in vivo structural dynamics in integral membrane protein by

hydroxyl radical footprinting. *Molecular and Cellular Proteomics*, 8(8):1999–2010, 2009.

[209] T. E. Angel, S. Gupta, B. Jastrzebska, K. Palczewski, and M. R. Chance. Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proceedings of the National Academy of Sciences*, 106(34):14367–14372, 2009.

[210] S. W. Cowan, T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius, and J. P. Rosenbusch. Crystal structures explain functional properties of two E. coli porins. *Nature*, 358(6389):727–733, 1992.

[211] B. K. Ziervogel and B. Roux. The binding of antibiotics in OmpF porin. *Structure*, 21(1):76–87, 2013.

[212] E. Pebay-Peyroula, R. M. Garavito, J. P. Rosenbusch, M. Zulauf, and P. A. Timmins. Detergent structure in tetragonal crystals of OmpF porin. *Structure (London)*, 3(10):1051–1059, 1995.

[213] A. H. O'Keeffe, J. M. East, and A. G. Lee. Selectivity in lipid binding to the bacterial outer membrane protein OmpF. *Biophysical Journal*, 79(4):2066–2074, 2000.