University of Nottingham
UK | CHINA | MALAYSIA

Developing Carbene Footprinting Mass Spectrometry to Interrogate Protein-Ligand Interactions

by

James R. Lloyd, MSci (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

July 2023

Abstract

The interaction of proteins with small molecules or other proteins is essential to almost every cellular process. Because of this importance and the potential implication in disease pathogenesis, proteins are attractive drug targets. Improved knowledge of these relationships greatly benefits our understanding of key biological processes and aids in drug discovery efforts. Techniques that provide structural information on a short time scale and employ small amounts of sample are highly desired. Carbene footprinting is a recently developed mass spectrometry-based chemical labelling technique that probes protein interactions and conformation. The work reported in this thesis aims to further develop the application of carbene footprinting to the study of protein structure and interactions by mapping the contact interfaces of several clinically relevant protein systems.

Using the methodology, the contact surface between the human eukaryotic initiation factor 4A (eIF4A) and small molecular inhibitor hippuristanol was accurately mapped to the protein's C-terminal domain (CTD). The technique was successfully employed to study interactions between members of the inflammasome human Gasdermin D (hGSDMD) and hCaspase-1 (C285A) which revealed direct occupancy of the hCaspase-1 active-site by hGSDMD for the first time. Carbene footprinting was also applied to the hGSDMD pore-forming N-terminal domain (NTD) assembled in liposomes and then compared to the soluble monomer which showed masking effects consistent with oligomeric assembly and insertion into the lipid bilayer. Interactions between Caspase-1 (C285A) and the specific inhibitor O-desethyl-belnacasan (VRT-043198) were studied by carbene footprinting which revealed that the small molecule non-covalently occupied the active-site of a C285A mutant. Carbene footprinting was also applied to two alpha helical membrane proteins. The archaeal multidrug and toxic compound extrusion (MATE) transporter PfMATE was employed to optimise the labelling workflow to insoluble protein systems which notably revealed covalent modification of detergent molecules. Carbene labelling was also applied to the G protein-coupled receptor (GPCR) beta-1 adrenergic receptor (β_1AR). Activated and inactivated ternary β_1AR-nanobody (Nb) complexes highlighted binding of small molecules to the extracellular binding cavity as well as differential labelling changes on intracellular transmembrane helices (TMs) that indicated varying activation states of the receptor.

Acknowledgements

I would like to thank Professor Neil Oldham and Dr Jon Hopper for providing me with this opportunity and for being fantastic supervisors. I would like to thank everyone in the Oldham group (past and present, with a special thanks to Dr Jeddidiah Bellamy-Carter), everyone at OMass Therapeutics, my friends (particularly Christopher Merrett) and family for being so supportive and loving over the last four years.

"The satisfaction comes when you close PepFoot." - Cameron Baines

Publications

Mapping the interaction between eukaryotic initiation factor 4A and the inhibitor hippuristanol using carbene footprinting and mass spectrometry
James R. Lloyd, Amy Hogan, Vasileios Paschalis, Jeddidiah Bellamy-Carter, Andrew Bottley, Graham B. Seymour, Christopher J. Hayes, Neil J. Oldham.
Proteomics, 21, 2000288 (2021)

Mapping the binding interactions between human gasdermin D and human caspase-1 using carbene footprinting
James R. Lloyd, Antonio Biasutto, Katharina Dürr, Ali Jazayeri, Jonathan Hopper, Neil J. Oldham.
JACS AU, https://doi.org/10.1021/jacsau.3c00236 (2023)

Mapping binding interactions and structural changes of the beta-1 adrenergic receptor using carbene footprinting
James R. Lloyd, Parth Kapoor, Katharina Dürr, Ali Jazayeri, Jonathan Hopper, Neil J. Oldham.
Manuscript in preparation

Contents

List of Figures

List of Tables

List of Abbreviations

β1AR Beta-1 Adrenergic Receptor

β2AR Beta-2 Adrenergic Receptor

βME Beta-Mercapthoethanol

AC Alternating Current

ATP Adenosine Triphosphate

Br-NRF Bromine derivative of Norfloxacin

cAMP Cyclic Adenosine Monophosphate

CEM Chain-Ejection Model

CID Collision-Induced Dissociation

CL Cardiolipin

CMC Critical Micelle Concentration

CPK Corey–Pauling–Koltun

CRM Charged Residue Model

Cryo-EM Cryogenic Electron Microscopy

CSD Charge-State Distribution

CTD C-Terminal Domain

CTLA-4 Cytotoxic T-Lymphocyte-Associated Protein 4

DC Direct Current

DDA Data-Dependent Acquisition

DDM *N*-Dodecyl-β-Maltoside

DEER Double Electron-Electron Resonance

DH Dehydratase

DHD Dehydratase Docking

DIA Data-Independent Acquisition

DMSO Dimethyl Sulfoxide

DNA Deoxyribonucleic Acid

DSS Sodium Trimethylsilylpropanesulfonate

DTT Dithiothreitol

ECD Electron-Capture Dissociation

ECL Extracellular Loop

EGFR Epidermal Growth Factor Receptor

EIC Extracted Ion Chromatogram

eIF Eukaryotic Initiation Factor

ERRα Estrogen-Related Receptor Alpha

ESI Electrospray Ionisation

ETD Electron-Transfer Dissociation

FID Free Induction Decay

FPOP Fast Photochemical Oxidation Of Proteins

FRET Förster Resonance Energy Transfer

FTICR Fourier-Transform Ion Cyclotron Resonance

GAP GTPase-Activating Protein

GDP Guanosine Diphosphate

GFP Green Fluorescent Protein

GPCR G Protein-Coupled Receptor

GPI Glycosylphosphatidylinositol

GSDMD Gasdermin D

GTP Guanosine Triphosphate

HDX Hydrogen-Deuterium Exchange

HEWL Hen Egg-White Lysozyme

HGP Heteromeric G Protein

HRPF Hydroxyl Radical Protein Footprinting

ICL Intracellular Loop

ICR Ion Cyclotron Resonance

IEM Ion Evaporation Model

IMAC Immobilised Metal Affinity Chromatography

IMP Intergral Membrane Protein

IPTG Isopropyl β-D-1-thiogalactopyranoside

iTRAQ Isobaric Tags For Relative And Absolute Quantitation

LB Lysogeny Broth

LC Liquid Chromatography

LIT Linear Ion Trap

LPS Lipopolysaccharide

MATE Multidrug And Toxic Compound Extrusion

MICA Major Histocompatibility Complex Class I Chain-Related Molecule A

mRNA Messenger Ribonucleic Acid

MS Mass Spectrometry

MS/MS Tandem MS

mz Mass-To-Charge Ratio

NAG_5 Penta-*N*-Acetylchitopentaose

NaTDB Sodium 4-(3-Trifluoromethyl)-3H-Diazirin-3-yl)Benzoic Acid

Nb Nanobody

Nd:YAG Neodymium-Doped Yttrium Aluminum Garnet

NMR Nuclear Magnetic Resonance

NTD N-Terminal Domain

OD_{600} Optical Density (600 nm)

OG *N*-Octyl-β-D-Glucoside

OMPF Outer Membrane Protein F

PBS Phosphate-Buffered Saline (10 mM, pH 7.4)

PCNA Profilerating Cell Nuclear Antigen

PKA Protein Kinase A

PMP Peripheral Membrane Protein

POPC Phosphatidylcholine

PPARγ Peroxisome Proliferator-Activated Receptorγ

PTM Post-Translation Modification

RF Radio Frequency

RMSD Root Mean Square Deviation

RNA Ribonucleic Acid

RP-HPLC Reverse-Phase High-Performance Liquid Chromatography

SASA Solvent Accessible Surface Area

SDS-PAGE Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

SEC Size Exclusion Chromatography

SMALP Styrene Maleic-Acid Co-polymer Lipid Particle

SOD1 Superoxide Dismutase 1

SRM Selected Reaction Monitoring

SWIFT Stored Waveform Inverse Fourier Transform

T$_m$ Melting Temperature

TCEP Tris(2-carboxyethyl)phosphine

TDBA 4-(3-Trifluoromethyl)-3H-Diazirin-3-yl)Benzoic Acid

TEV Tobacco Etch Virus

TIC Total Ion Chromatogram

TM Transmembrane Helix

TMS Tetramethylsilane

tRNA Transfer Ribonucleic Acid

USP5 Ubiquitin Specific Protease 5

UTR Untranslated Region

UV Ultraviolet

UVPD Ultraviolet Photodissociation

VRT-043198 *O*-Desethylbelnacasan

VX-765 Belnacasan

1 Introduction

1.1 Proteins

Proteins are macromolecular workhorses involved in almost every living process. Proteins maintain cellular homeostasis through roles in cell signalling, defence, replication and metabolism. They exert their function through molecular recognition (the association of biological macromolecules with each other or small molecules through noncovalent interactions).[1] The shape of a protein, as well as its biochemical composition, determines what it can interact with and consequently, its overall function. There are an estimated ~20,000 proteins (and up to several million proteoforms[2]) encoded for in the human genome.[3] The enormous diversity of protein structures and potential binding partners correspond to an even larger number of functional outputs (there are an estimated 600,000 protein-protein interactions in the human interactome[4]) that finely govern the biochemistry of the cell.

1.1.1 Protein structure

Proteins are made up of one or more chains of amino acids. The three-dimensional arrangement (fold) of these biological building blocks is central to protein function.[1] There are four levels to protein structure (Figure 1.1):

- *Primary structure* is the exact sequence of amino acids in a polypeptide chain and is encoded for in the genetic information (genes). The primary sequence

of amino acids contains the necessary foundations for proteins to assume their native conformation.

- *Secondary structure* refers to regular repeating structural units between different amino acids, stabilised by intramolecular hydrogen bonds. Common secondary structure include alpha helices and beta strands.

- *Tertiary structure* is the overall three-dimensional structure of an individual polypeptide, stabilised by intramolecular interactions, including the hydrophobic effect, non-covalent interactions, disulfide bonds and salt bridges.

- *Quaternary structure* refers to the association of multiple (separate) polypeptide chains (often referred to as subunits in this context) that function as a single protein, stabilised by intermolecular interactions. These are referred to as dimers, trimers, etc.

1.1.2 Protein dynamics and interactions

Proteins are highly dynamic molecules. This can vary over a protein: some regions may be rigid whereas others highly flexible. The structural flexibility of a protein is inherent to its amino acid composition. The free-energy landscape theory provides a convenient theoretical framework for the kinetic and thermodynamic characterisation of conformational dispositions formed during protein folding.[5] According to this, protein folding is driven by a decrease in the Gibb's free energy of a protein-solvent system. The influence of the hydrophobic effect, salt bridges, intramolecular hydrogen bonds and van der Vaal interactions drives the protein structure down a series of funnel-like trajectories via intermediate states towards a native conformation at free-energy minima. These are rugged, filled with wells and traps, allowing a protein to exist as an ensemble of native conformers separated by low energy barriers that determine the transition rates between native conformers.[6]

Protein dynamics are also central to molecular recognition. The conformational changes associated with protein-ligand complexes can vary from small, local changes in side chain rotameric states to large global domain movements.[7] Three models have

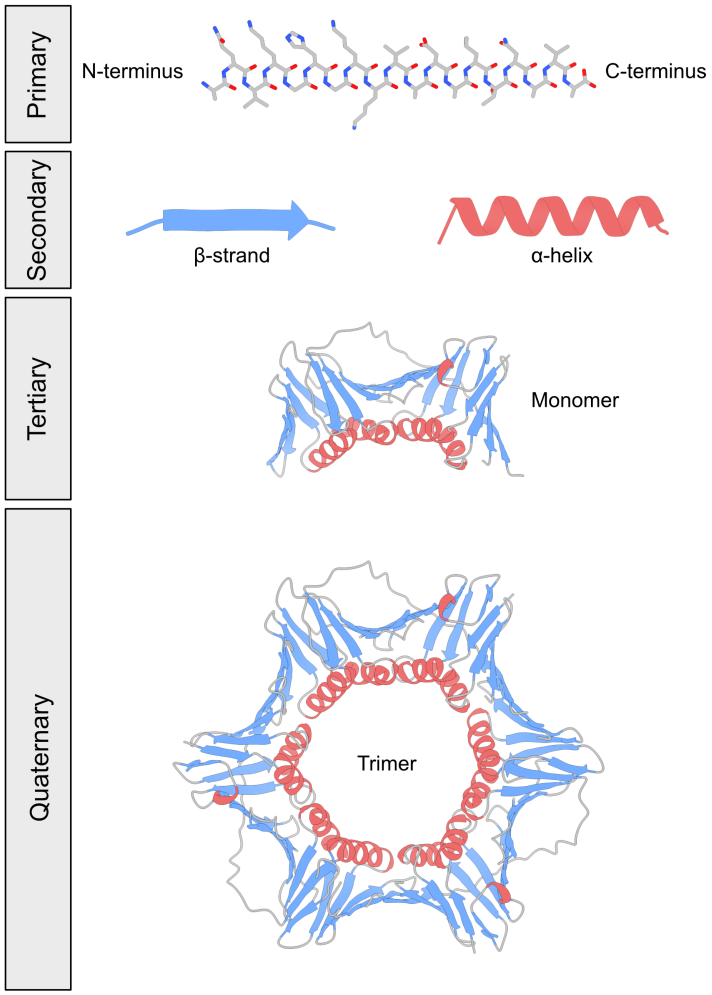


Figure 1.1 | Levels of protein structure. *Primary structure* — the sequence of amino acids in a polypeptide chain. *Secondary structure* — regular repeating structural units stabilised by hydrogen bonds on the polypeptide backbone. *Tertiary structure* — the overall three-dimensional structure of a polypeptide caused by different amino acid side chain interactions (the profilerating cell nuclear antigen (PCNA) monomer is shown). *Quaternary structure* — the association of multiple polypeptide chains (the PCNA trimer is shown (PDB 1AXC)).

been proposed to explain protein-ligand interactions, with each employing varying levels of structural change to accommodate the binding partner. All three models have been observed experimentally and it is likely that they exist simultaneously or sequentially and cover a broad spectrum of binding events (Figure 1.2).[1]

- The prototypical *lock-and-key hypothesis* necessitates perfectly complementary binding interfaces between the protein and ligand, omitting the need for any structural compensations.

- The *induced fit model* suggests that a defined binding site is flexible and can undergo structural shifts to accommodate to less complementary-shaped ligands.

- The *conformational selection model* suggests that a ligand can bind to the most suitable conformational state/substate of a protein.

Molecular recognition, like protein folding, is driven by a negative Gibbs free energy of a protein-solvent system.[1] This is the manifestation of the separate enthalpic and entropic contributions of the system. For a binding process, enthalpy changes represent noncovalent bond formation between protein-ligand and solvent-solvent molecules as well as disruption between protein-solvent and ligand-solvent molecules. Entropic contributions of a protein-solvent system include changes in the entropy of solvation (release of solvent upon surface burial), entropy of conformation (changes in conformational freedom) and entropy of rotation/translation (changes in rotational/translational degrees of freedom).[1] However, protein-ligand binding interactions are reversible events, and the strength of a protein-ligand interaction must be considered in light of the mean-lifetime of a complex. In this way, the amount of protein and ligand present, how much complex forms between them and how quickly these complexes fall apart, are all important aspects of protein-ligand interactions.[8] The association and dissociation between a protein and ligand can be described in a time-dependent manner where k_{on} and k_{off} represent the association and dissociation rates between protein and ligand, respectively. At equilibrium, higher concentration of free protein and ligand will favour the likelihood of each molecule coming into contact with the other. Collisions between protein and ligand are increased with larger size (including overall surface area of the protein but also of the binding site); however, not

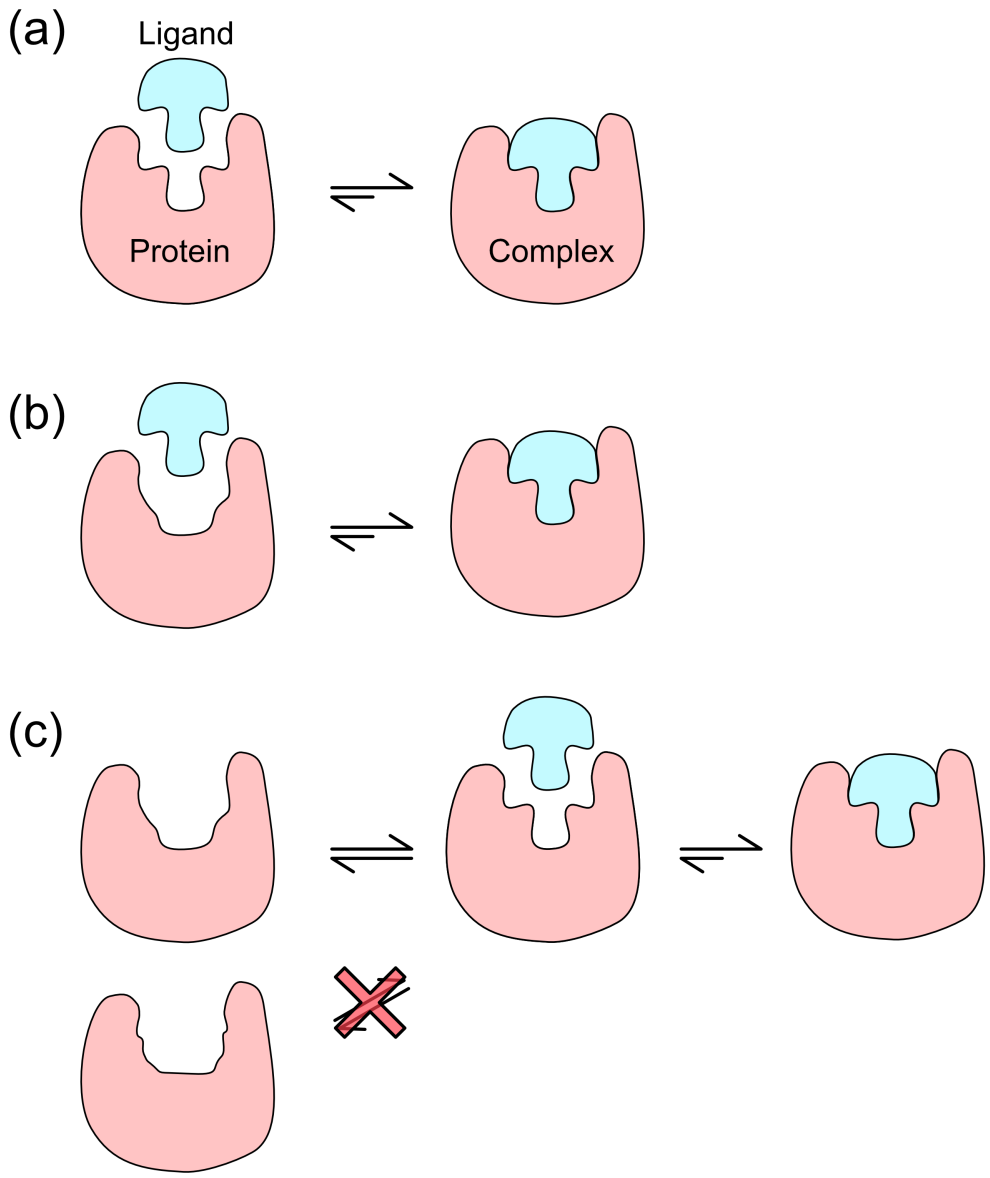


Figure 1.2 | Models of protein-ligand binding. (a) Lock-and-key model (b) Induced fit model (c) Conformational selection model.

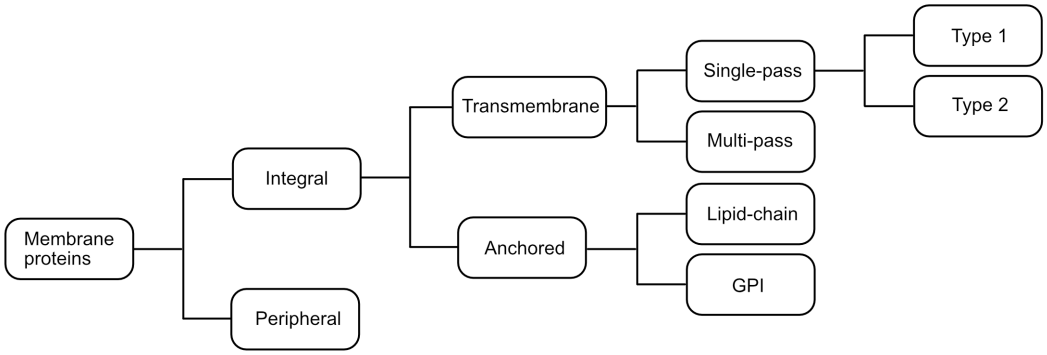
all collisions are productive.[1] In contrast, electrostatic interactions and surface chemical groups may guide a ligand towards the binding site. The dissociation of a complex is random. Binding interfaces made up of many weak interactions are considerably weaker (and consequently display a larger κ_{off}) than those consisting of few but stronger interactions. The dissociation constant (K_D) is often used to conveniently describe the dissociation of a complex into constituent molecules and is determined by the ratio of κ_{off}:κ_{on}.

1.1.3 Membranes and membrane proteins

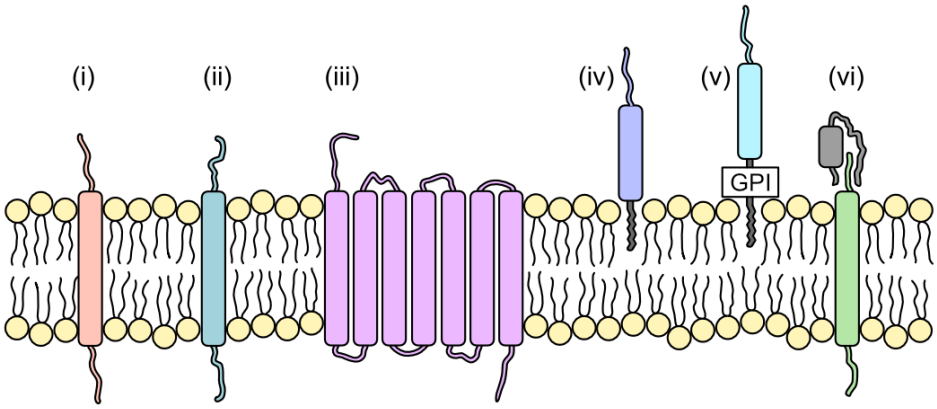
Cells are surrounded by a bilayer consisting of two layers of polar lipids, oriented with their hydrophobic tails pointing inwards. These thin, flexible sheets define and compartmentalise the cell into distinct aqueous environments, segregating chemical reactions and ultimately, aiding its overall biochemical efficiency.[9] However, membrane lipids are far from inert barrier structures, and contribute tremendously to cell functionality, for example, they potentiate budding, fusion, fission and tubulation events which in turn drive cellular division, reproduction and intracellular trafficking.

Membrane proteins make up an important part of biological membranes (~50 % by mass and ~30 % by area[10]). These 'gatekeepers of the cell' function as checkpoints that mediate the transmission of molecules and information across the lipid bilayer and have been implicated in molecular transport, signal transduction and catalysis. Membrane proteins can be classified depending on their localisation within the lipid bilayer (Figure 1.3a).[11] Integral membrane proteins (IMPs) are at least partially embedded in the bilayer whilst peripheral membrane proteins (PMPs) do not insert into the central hydrophobic portion of the lipid bilayer and instead associate with either integral membrane proteins or outer hydrophilic head groups of membrane lipids. IMPs can be further categorised into transmembrane and anchored membrane proteins. Transmembrane proteins span the entirety of the lipid bilayer whereas anchored (monotopic) membrane proteins are associated with (and inserted into) one side of the membrane. Transmembrane proteins can be grouped according to the number of times they span the lipid bilayer. Single-pass (bitopic) membrane proteins span the bilayer once, whilst multi-pass (polytopic) membrane proteins span it several times. Single-pass membrane proteins can also be classed according to the positioning of

N-terminal domains (NTD) and C-terminal domains (CTD) on different sides of the bilayer (Figure 1.3b). Type I transmembrane proteins have a cytoplasmic CTD and extracellular NTD whilst type II transmembrane proteins display a cytoplasmic NTD and an extracellular CTD. Anchored membrane proteins can be grouped according to their linkage to the lipid bilayer. They consist of lipid chain-anchored membrane proteins and GPI-anchored membrane proteins.



(a) Multilabel classification of membrane proteins



(b) Schematic of the six classes of membrane protein. (i) type-I (ii) type-II (iii) multi-pass (iv) lipid-chain anchored (v) GPI-achored (vi) peripheral (grey). Adapted from.[11]

Figure 1.3 | Membrane protein classification.

1.2 Methods for studying protein structure

Since proteins govern biological function, it follows that their dysfunction drives progression towards diseased states. Mutations in protein-encoding genes (either inherited or acquired) may promote disease-causing protein conformations through disrupted protein-deoxyribonucleic acid (DNA) interactions, protein misfolding

and new and undesired interactions with other proteins.[12] Pathogen-host protein interactions may also promote diseased states. Proteins are important targets of therapeutic intervention and subsequent drug discovery.[13] Membrane proteins are of particular pharmaceutical interest. Their roles in cell signalling and molecular transport implicate them in over 50% of drug targets.[14]

In vitro biophysical analysis of protein-ligand interactions — and in particular, structural interrogation of complexes — is an important stage of early drug discovery and more generally, a prerequisite to increasing our understanding of biology at the molecular level.[15] There are a range of tools available for the study of protein-ligand interactions; however, these first necessitate a high quality, homogeneous and functional (if appropriate) protein sample.

1.2.1 Protein production and purification

Obtaining recombinant proteins for *in vitro* analysis encapsulates two main processes: protein production and purification (Figure 1.4).

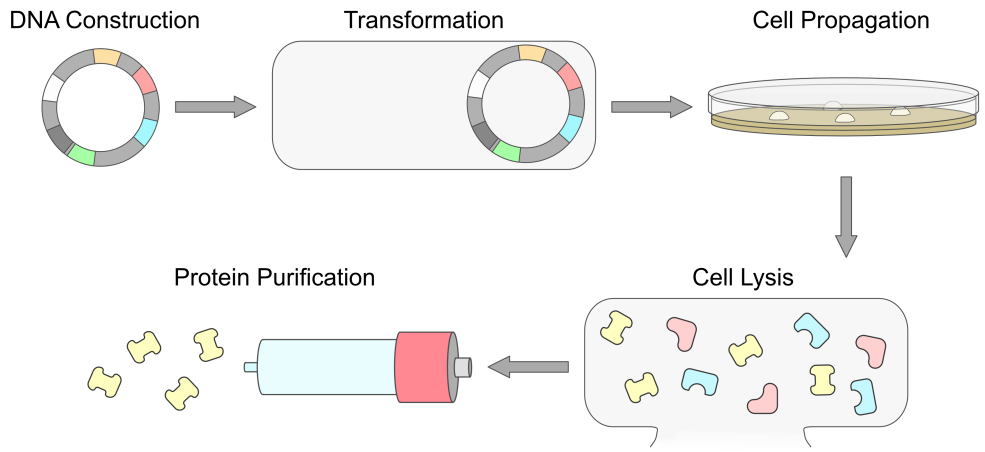


Figure 1.4 | Principles of protein expression and purification. Exogenous genetic material encoding a protein of interest is typically introduced into a chosen host. Cells are grown, such that the host expresses a large amount of the protein. Cells are collected and lysed before purification techniques are employed to isolate the protein of interest from the cell lysate.

Protein production typically involves manipulation of a host organism's genetic information such that it expresses (transcribes, translates and properly folds) a recombinant gene encoding a protein of interest.[16] There are a range of cell-based systems for the expression of recombinant proteins with varying advantages and

disadvantages. Bacteria are amongst the most commonly used expression hosts but the most appropriate host system often depends on the protein that is to be expressed. Bacteria are a convenient and cost-effective expression system due to their rapid growth rates (on cheap, minimal media) and ease of genetic manipulation. Indeed, plasmids (small circular pieces of DNA) facilitate cloning of genetic information into host cells and the use of strong promoters (e.g. T7) enable production of large amounts of protein. Promoters under control of the lac operon (whereby presence of the chemical inducer isopropyl β-D-1-thiogalactopyranoside (IPTG) binds to and release repressor proteins from the DNA) also enable toggling of gene expression. This is especially important for proteins which may be toxic to the cell. Unfortunately, bacteria lack the necessary molecular machinary to carry out post-translational modifications and eukaroytic, multi-domain proteins require alternative host expression systems.[17] Frequently used animal systems include insect, yeast and mammalian cells.

Protein purification involves isolation of a protein of interest from a host cell system.[16] Cells are lysed before the protein of interest is isolated from the cell lysate (usually by exploiting the protein's affinity, physico-chemical properties or size). Sonication or pressure homogenisation are popular methods to lyse cells but other techniques are available (freeze-thawing, proteases etc). There is no one best method, and the most appropriate process depends on the stability of the protein. The cell lysate may then be separated into soluble and insoluble fractions by way of centrifugation. Isolation of membrane proteins which are contained in the insoluble fraction follows several centrifugation steps to clarify cell membranes from cellular debris. Chromatographic techniques are used to obtain a protein isolate. Immobilised metal ion affinity chromatography (IMAC) is amongst the most commonly used chromatographic technique to isolate a protein of interest. A prerequisite of this is that the protein construct contains an affinity tag (typically a polyhistidine tag at either NTD or CTD). The fractionated lysate is combined with metal-charged affinity resins, causing electron donor groups on the histidine to form coordination bonds with the metal (often iron, nickel or cobalt). Imidazole solutions can then be used to elute the immobilised protein from the resin.[18] Further separation techniques can be employed to remove contaminants that co-purify with a recombinant protein. Size exclusion chromatography (SEC) involves passing the IMAC-purified proteins through

a porous matrix. Smaller molecules negotiate a longer route than larger molecules, causing size-based separation. SEC also permits buffer exchange which is particularly beneficial for removing high imidazole concentrations.

1.2.1.1 Difficulties in expressing and studying membrane proteins

Historically, the characterisation of membrane proteins has lagged behind soluble systems owing to difficulty by which proteins are obtained.[19] Membrane proteins are often found at low concentrations in the bilayer and their overexpression via heterologous systems can lead to aggregation in the cell. Many mammalian membrane proteins also require post-translational modifications (PTMs) which may be unavailable in bacterial host systems; however, bacteriorodopsin, a light-activated proton pump, is the only protein found in *Halobacterium salinarum* cell membranes and purification generally gives rise to high yields of the transporter.[19] The study of membrane proteins often also requires exchange into synthetic membrane mimetics; however, membrane proteins (and particularly transmembrane proteins) are normally deeply embedded within the lipid bilayer and removal of this can disrupt protein integrity. Moreover, the lipid bilayer does not solely exist to accommodate membrane proteins and the importance of the lipid milieu in maintaining native membrane protein structure and function is becoming increasingly recognised. For example, membrane lipids are known to exert an influence over membrane proteins by acting as substrates, stabilising oligomeric states, fine-tuning structures and allosterically modulating interactions.[20]

Typically, detergent systems are used to extract membrane proteins from the lipid bilayer and satisfy their high proportion of hydrophobic membrane-spanning residues. Detergents, like lipids, are amphipathic molecules consisting of a polar head group and hydrophobic tail. Above a critical micelle concentration (CMC), detergents form structural assemblies called micelles which feature topologically separate hydrophilic and hydrophobic sections. In a broad sense, detergents and lipids are both surfactants but are distinguished by the concentrations required for self-association and the shapes that these assemblies assume. The architecture of detergent micelles promotes interaction with, and solubilisation of, lipid bilayers (and associated proteins) into mixed micelle complexes. There are a great number of detergents available for membrane protein solubilisation. These can be classified

into four main categories: ionic detergents, bile acid salts, non-ionic detergents and zwitterionic detergents (Figure 1.5).[19] The physical properties of detergents vary greatly and are generally determined by the character and stereochemistry of the head group and tail. In addition to detergents, other solubilising agents exist which aim to preserve the native lipid environment, including bicelles, amphipols, nanodiscs and styrene maleic acid co-polymer lipid particles (SMALPs) (Figure 1.6).

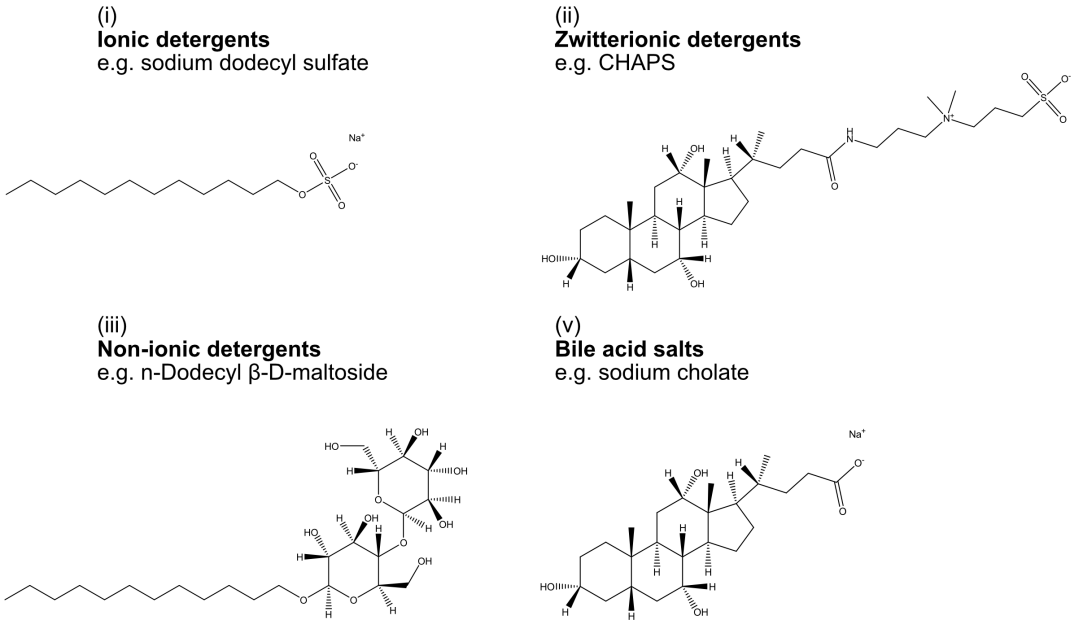


Figure 1.5 | Detergent classification. Detergents can be classified according to their electrical charge. Charged detergents are more denaturing than zwitterionic or non-ionic detergents.

1.2.2 X-ray crystallography

X-ray crystallography has long been a cornerstone of structural biology and can provide atomic-level information about protein structure.[21]

As the name suggests, X-ray crystallography necessitates the formation of protein crystals. The process of crystallisation is not well understood but is thought to include nucleation and growth stages.[22] In practice, a highly pure and concentrated protein solution (the concentration of which is initially below the solubility limit) has the solubility of the protein lowered (usually through addition of additives), rendering it in a supersaturated state. Proteins in a supersaturated state are thought

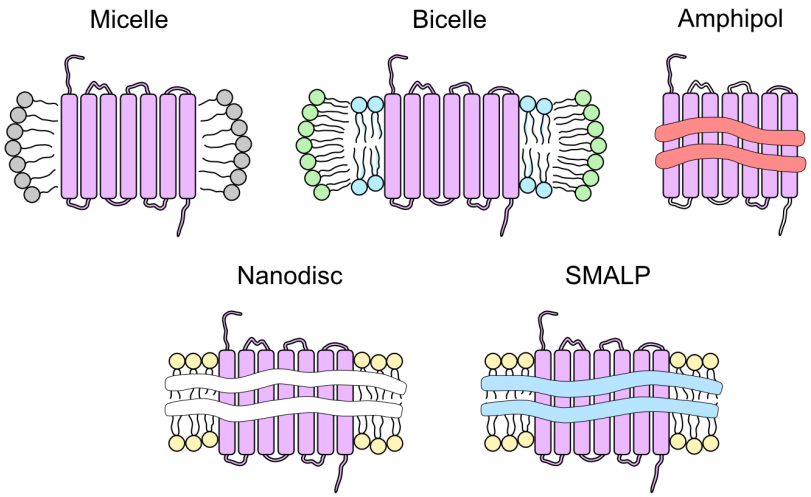


Figure 1.6 | **Lipid bilayer mimetics**. Frequently used membrane mimetics for membrane protein solubilisation include micelles, bicelles, amphipols, nanodiscs and SMALPs.

to form critical nuclei which promote crystal growth. The growth of protein crystals relies on trial-and-error and involves vigorous screening for suitable crystallisation conditions.[22] Once obtained, crystals are mounted and exposed to an X-ray beam and the resultant diffraction pattern recorded using charged coupled device technology (Figure 1.7). The diffraction data is processed. This is mathematically complex and involves: determination of the crystal system and unit cell dimensions, orientation of the beam, spot intensity, application of scale factors and amplitude of the reflection and phase angle between diffracted waves.[22] A fast fourier transform method is used to calculate structure factors and generate an electron density map which is used for model building.

The ability to obtain three-dimensional molecular structures has defined the field of structural biology and hugely accelerated structure-based drug discovery. Nevertheless, the technique has its drawbacks and is not necessarily suitable for all protein systems. X-ray crystallography is time-consuming, requires relatively large amounts of sample and furthermore, some proteins may not be amenable to the preparation (screening) stages. Crystals, by definition, also consist of repeated packing of identical units that give rise to a single rigid structure which may not reflect the true conformational range of a protein.

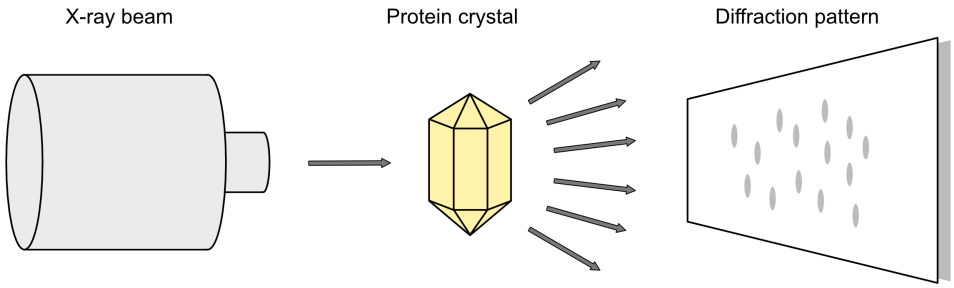


Figure 1.7 | **Principles of X-ray crystallography**. An X-ray beam is used to illuminate a protein crystal lattice. The resultant diffraction pattern is used to generate an electron density map from which a three dimensional atomic structure can be discerned.

1.2.3 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is another technique used to obtain three-dimensional information about the structure and dynamics of biological macromolecules. NMR spectroscopy monitors the quantum-mechanical properties of atomic nuclei.[23]

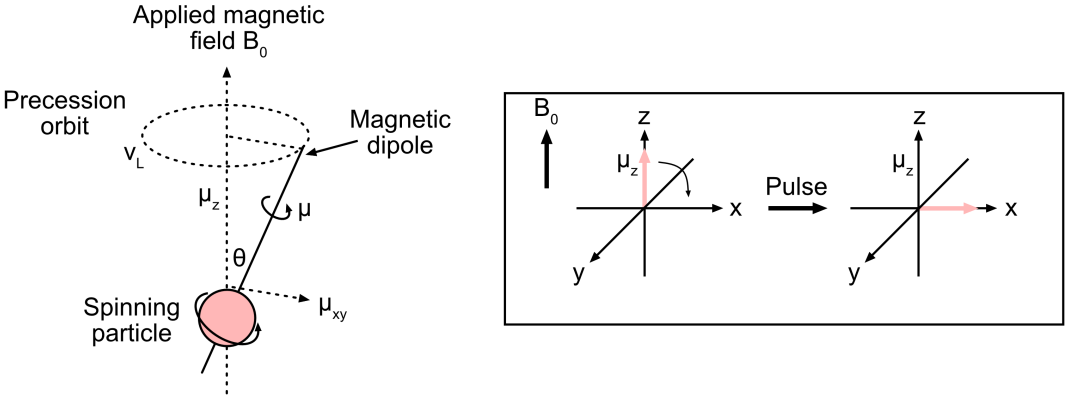


Figure 1.8 | **Principles of NMR spectroscopy**. Nuclei with odd mass numbers in the presence of an external magnetic field precess around the axis of the external field vector with a characteristic Larmor precession. In NMR spectroscopy, radio frequency pulses are applied to these nuclei such that the magnetic field component lies on the x-axis, rotating the bulk magnetisation onto the xy-plane. Relaxation of this magnetisation to the direction of the external magnetic field is monitored as a change in amplitude over time.

Atomic nuclei with odd mass numbers have the property of spin and will rotate around a given axis, described by spin angular momentum.[24] Nuclei with

spin also exhibit a magnetic field, characterised by the nuclear magnetic moment vector. When placed in an external magnetic field, nuclei precess around the axis of the external field vector, known as Larmor precession. To make this precession detectable, radio frequency pulses resonant with each unique Larmor precession and perpendicular to the external magnetic field are applied to samples, causing nuclei to rotate perpendicularly to the plane of the magnetic field (Figure 1.8). After the magnetic pulse, the magnetisation undergoes a relaxation process to the direction of the external magnetic field. A receiver coil detects the electric current induced by the rotating magnetic field. The declining amplitude is measured as a function of time, known as free induction decay (FID). A Fourier analysis is used to transform the FID to a function of frequency.[24] An NMR spectrum consists of a series of peaks of varying intensity, known as the chemical shift. These peaks are derived from the FID and reflect different atomic nuclei and their varying chemical environments. The chemical shift is relative to a reference compound such as sodium trimethylsilylpropanesulfonate (DSS) or tetramethylsilane (TMS). Proteins have a large number of NMR-active nuclei and NMR-spectra can be complex. Furthermore, relaxation rates are faster in larger molecules, leading to peak broadening and a reduced resolution, making spectral assignment difficult. Multidimensional NMR experiments (including correlated spectroscopy, total correlated spectroscopy and Nuclear Overhauser Effect spectroscopy) are often employed in protein NMR to address these limitations by introducing additional dimensions to the spectra which decrease the chance of overlaps and permit spectral resolution that would otherwise not be achieved in 1D NMR.[25]

NMR spectroscopy can provide valuable information relating to protein dynamics and interactions.[25] Advantages of the technique include its automation, non-destructive nature and ability to report on an atomic level. However, protein NMR experiments can be time-consuming and require large amounts of isotopically labelled sample.

1.2.4 Cryogenic electron microscopy

Cryogenic electron microscopy (cryo-EM) is another technique that has contributed enormously to the field of structural biology. Cryo-EM is a form of electron microscopy, most commonly transmission electron microscopy (TEM). In TEM, electrons from an electron gun are deflected and focused onto a specimen. Electrons penetrate

the specimen and project a 3D electric potential distribution function onto a two-dimensional plane before a computer 3D reconstruction is used to generate a magnified projection of the sample.[26] Vitrification of samples using liquid ethane overcomes several classical limitations of TEM, namely the incompatibility of hydrated samples with a high vacuum and damage of biological samples from high-energy electrons. Advancements in high-resolution image capturing devices and development of software for high-throughput, automatic data collection, processing and analysis have made entry into the field of cryo-EM and acquisition of detailed images of biological macromolecules easier.[26] Consequently, the number of submissions to the electron microscopy data bank have increased exponentially in recent years.

Moreover, cryo-EM maintains several advantages over other structural biology techniques. For one, it uses a low sample volume (0.1–1 µM, 3–5 µL). Vitrification also preserves samples under physiological conditions which is more likely to reflect their native state/structure compared to other techniques where additional additives are introduced to the sample.[27] Cryo-EM can be performed on large biological macromolecules, up to 500 kDa in size. The technique can also be performed on inhomogeneous samples, where advancements in algorithms enable statistical analyses that can classify molecules into separate conformational classes.

1.2.5 Difficulties in studying membrane proteins

Membrane proteins are typically more challenging to study by conventional structural biology approaches (including X-ray crystallography, NMR and cryo-EM) compared to soluble systems. For example, in X-ray crystallography, detergents and membrane mimetics encapsulate most of the protein, minimising the available surface area for formation of crystal contacts during crystallisation stages.[28] Moreover, membrane protein crystals are typically lower in resolution (>5 Å) than soluble protein crystals, and they are often also associated with anisotropy or twinning defects. Nevertheless, developments to crystallisation, including vapor diffusion crystallisation, *in meso* crystallisation and crystallisation chaperones have largely overcome these difficulties, and the technique is responsible for ~80 % of solved membrane protein structures. In contrast, drastic improvements to cryo-EM resolution through technological advancements have reignited interest in the technique to study membrane protein

structure.[29] Indeed, cryo-EM requires a small volume of sample and the technique is also tolerant to the presence of detergents and other membrane mimetics. Recently for example, single particle cryo-EM was used to determine the first sub nm structure of a SMA-solubilised membrane protein. Nevertheless, cryo-EM still suffers from size detection thresholds which limits study to membrane proteins greater than 50 kDa. NMR spectroscopy can also be used to study membrane proteins.[30] These include in solution and solid-state, where the latter permits study of membrane proteins embedded in lipid bilayers.

Despite the ongoing successes of x-ray crystallography, NMR spectroscopy and cryo-EM in the study of membrane proteins, a great deal of research is still being devoted to developing alternative techniques that facilitate study of this challenging family of proteins.

1.3 Mass spectrometry

Over the last 30 years, mass spectrometry (MS) has emerged as a powerful complementary technique to interrogate the noncovalent interactions of biological macromolecules. MS can provide valuable information such as binding site locations, dissociation constants, binding stoichiometry and conformational changes upon interaction.[31] Furthermore, its rapid and highly sensitive nature (requiring small amounts of sample) is extremely suitable for high-throughput, pharmaceutical-type study.

The field of MS encompasses a multitude of techniques that, broadly speaking, measure the response of ionised molecules to user-defined electric or magnetic fields and ultimately provide information on the mass-to-charge ratio (m/z) of ions.[32]

1.3.1 Ionisation techniques

The first step in a mass spectrometer is the production of gas phase ions. This is performed by an ionisation source. Ionisation sources can be classified as soft or hard depending on the amount of initial energy that they impart to the sample and the degree of fragmentation that occurs.

1.3.1.1 Electrospray ionisation

Electrospray ionisation (ESI) is a soft ionisation source. It can transfer ions to the gas phase without altering their integrity which is highly advantageous for structural mass spectrometry.[33]

In ESI, an analyte solution containing a nonvolatile solute dissolved in a volatile solvent is passed through a capillary emitter to an atmospheric pressure sampling plate/inlet capillary under high voltage.[34] As a result of the equilibrium between the surface tension of the solution and the electrostatic attraction to the counter electrode, and the electrostatic repulsion within the droplet, an elliptical droplet forms at the end of the capillary. At a given voltage, the Coulombic attraction overcomes the surface tension of the solution and the elliptical tip forms a Taylor Cone from which a fine spray of droplets is emitted (Figure 1.9).[33] Solvents used in ESI are often acidic to aid protonation of sample molecules in solution; however, additional redox reactions (anion discharge, anodic corrosion of capillary tip and solvent-oxidation) further aid the movement of excess ions (most commonly protons) into droplets.[35] In positive ion mode, these initial droplets carry a net positive charge. The droplets produced by the Taylor Cone undergo solvent evaporation which is often assisted through heating. The charge density of droplets gradually increases until the surface tension is balanced by Coulombic repulsion. This is known as the Rayleigh limit.[33] Droplets then undergo fission and explode into tens of smaller daughter droplets, with a larger droplet remaining behind. This process repeats until daughter droplets are only a few nanometers in diameter (Figure 1.9). The charge residue model (CRM), ion evaporation model (IEM) and chain-ejection model (CEM) describe the subsequent production of gas-phase ionised proteins from these Rayleigh droplets.[36] The CRM describes the production of ions from high-molecular weight species such as native folded proteins. Rayleigh-charged droplets undergo extensive desolvation until only a single analyte ion remains and any prior solvent charge is transferred to this. Low molecular weight species are thought to be transferred to the gas-phase by the IEM where excess charge emanating from the Rayleigh droplet causes repulsion and eventual expulsion of the analyte ion. The CEM describes the production of gas-phase ions from disordered,

unfolded proteins in which the polypeptide initially migrates to the surface of the Rayleigh droplet before one terminus exits the droplet and the rest follows.[36]

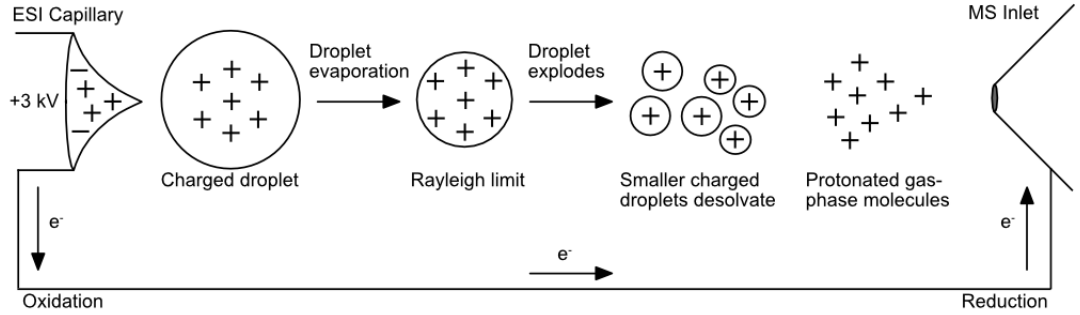


Figure 1.9 | Principles of ESI. A high voltage is applied to an analyte solution, causing dispersal of a fine mist of droplets. Nebulising gases and high temperatures are used to aid solvent evaporation before charged particles are ejected into the gas phase. Adapted from.[37]

Multiple charging of larger ionised molecules (i.e. peptides or proteins) is a feature of ESI and under appropriate experimental conditions, $[M + zH]^{Z+}$ ions are formed. Charge is predominantly due to proton adduction.[38] The presence of solvent additives and contaminants can lead to signal reduction. This may occur through ion suppression where contaminants reduce the total analyte signal as well as peak splitting due to adduct formation with other charge carriers. For example, the presence of Na^+ leads to sodium adduction forming $[M + (Z - i)H + iNa]^{Z+}$ ions, and mass heterogeneity caused by a variety of spectral peak adducts leads to a reduced signal-to-noise ratio.[39] Dialysis and desalting steps can be used to minimise the effect of contaminants on MS. Alternatively, ammonium salts such as ammonium acetate and ammonium bicarbonate are compatible with ESI.[40] In solution, ammonium cations pair with acidic functional groups whilst acetate and bicarbonate anions pair with basic functional groups. During the ion sampling stage of MS, sample ions undergo collisional activation which causes displacement of volatile ammonia and acetic/carbonic acid by proton transfer generating 'clean' protein ions. Several factors are known to influence the charge-state distribution (CSD) of proteins where physical dimensions of the biomolecule (primarily the surface area) are a major determinant.[34] Charge multiplicity is useful for analysing large proteins with mass analysers of a limited *m/z* range. Another advantage of ESI is that it may be integrated with liquid chromatography (LC). By separating complex mixtures into their relative constituents and introducing them

into the mass spectrometer, definitive identification and quantitative determination of desired compounds can be carried out.[33]

1.3.2 Mass analysers

Gaseous ions are next passed through the mass spectrometer to the mass analyser. The mass analyser uses static or dynamic electric or magnetic fields to separate charged particles according to their m/z values[32, 33].

1.3.2.1 Fourier-transform ion cyclotron resonance

The Fourier-transform ion cyclotron resonance (FTICR) mass analyser employs a superconducting magnet to generate a strong magnetic field. Ions moving in the presence of this magnetic field, with a component of their velocity perpendicular to its axis, will experience the Lorentz Force.[41] This causes charged particles to assume circular trajectories around the centre of the magnetic field axis. Ions of a given m/z will have the same trajectory. This so called cyclotron motion can be defined as:

$$\omega = \frac{qB}{m} \tag{1.1}$$

$$f = \frac{qB}{2\pi m} \tag{1.2}$$

where B is the magnetic field strength and ω (rad/s) and f (hertz, Hz) are expressions of cyclotron frequency.[42] Two additional ion motions are induced by the presence of trapping potentials that hold ions axially within the ion cyclotron resonance (ICR) cell: trapping oscillations between the two axial electrodes and a magnetron frequency that revolves around the centre axis of the magnetic field.[43] Fortunately the magnetron and trapping frequencies do not effect the cyclotron motion.[44]

During a measurement, ions must first be excited to detectable orbital radii. This is done through a 'chirp' or 'stored waveform inverse Fourier transform (SWIFT)' excitation.[33] Radio frequency (RF) potentials resonant with cyclotron frequencies of all ions are applied to excitation plates located along the y-axis of the cell (Figure 1.10). Importantly, all ions are excited to the same maximum radii (i.e. independent of m/z). Ions of the same m/z also rotate as coherent packets with their own characteristic

cyclotron frequencies; however, trajectories eventually dampen due to collisions with particles in the ICR cell. Detector plates located along the x-axis of the instrument detect packets of ions.[42] As the ions move past the plates, charge within the detection circuit moves to balance their presence and potential differences between plates are measured as a function of time.[41] Since the raw data represents an amalgamation of signals from all ion packets within the cell, it is essential to decipher data for each individual ion group. This is done through a type of mathematical manipulation called a Fourier Transform in which the time domain spectrum is converted to a frequency spectrum. The use of a calibration then permits conversion from frequency to m/z (in the form of a mass spectrum).[45]

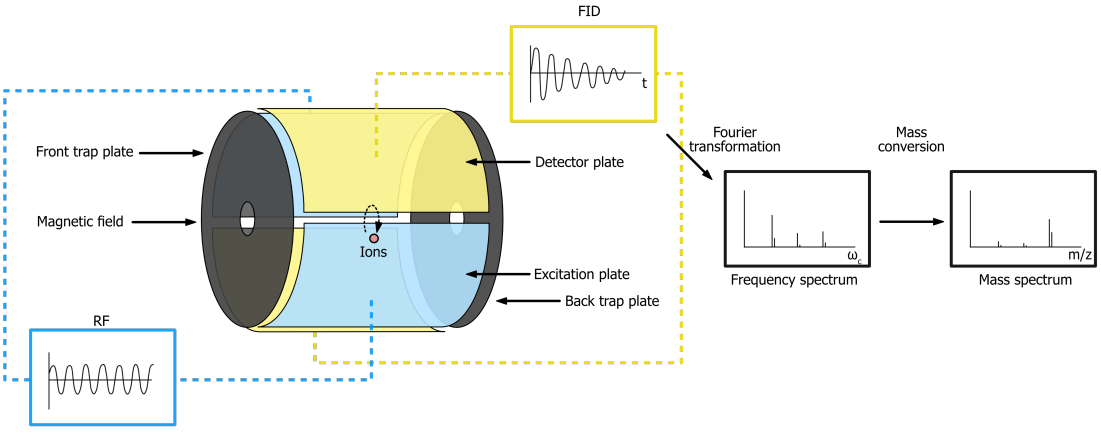


Figure 1.10 | Principles of FTICR. Ions trapped within a Penning Trap are excited to their resonant frequencies by an electric field. Ions with the same m/z orbit the cell with unique cyclotron frequencies. These frequencies are detected by detector plates as the ions pass nearby them. This free induction decay (FID) signal is converted to frequency spectrum via a Fourier Transform. A mass conversion is used to generate a mass spectrum from the frequency spectrum. Adapted from.[37]

FTICR provides ultra high resolution and mass accuracy allowing for unambiguous mass assignment and subsequent resolution of a given species which may be unachievable with other instruments. These properties are due to the measure of frequency which can be ascertained very accurately.[46] Moreover resolution is proportional to the acquisition time and magnetic field strength whilst inversely

proportional to m/z. In this way, FTICR instruments are kept at very low pressures (10^{-10} mbar) to reduce collisions with gas molecules and prevent the dampening of ion trajectories too quickly, permitting long acquisition times.[41]

1.3.2.2 Linear ion trap

The linear ion trap (LIT) confines ions in the radial dimension through the application of an alternating current (AC) voltage to four rods (usually hyperbolic) further divided into three sections (Figure 1.11).[37] The application of a direct current (DC) voltage to the end sections of the rods confines ions axially within the center of the trap. A positively charged ion will be attracted to a negative rod within the multipole and gain kinetic energy as it moves towards it; however, the alternating field spins these potentials, and that particular ion will see a decrease in its potential energy. Consequently, ions are sequentially brought back to the centre of the rods. Very large ions can have too much inertia and may be lost on the rods as neutrals. The degree of motion that these ions can undergo before hitting the rods is described by the Mathieu equation where the stability of an ion is given by two unitless parameters a and q and can be depicted within stability diagrams.[37] a refers to an ion's motion under a quadrupolar DC voltage and the q refers to an ion's motion under AC voltage. Since no quadrupolar DC voltage is present in an ion guide, q is the pertinent value. Stable q values are within 0-0.908 and an ion with a q value greater or equal to 0.908 can be ejected or lost on the rods as neutrals. Smaller ions have greater q values than larger ones and a smaller region of stability. Trapping ions within a LIT requires consideration of the focusing potential of heavy ions as well as the stable trajectory of light ions to ensure that a wide enough mass range is maintained.[37]

Particular ions can be isolated within the LIT through the ejection of unwanted species. This is done by applying AC waves corresponding to secular frequencies of unwanted ions to the trap in a simultaneous waveform — this is synonymous to excitation in the FTICR.[32] Ions then oscillate resonantly causing their destabilisation and ejection, leaving the selected ions behind in the trap. These can then be detected in subsequent mass analysers.

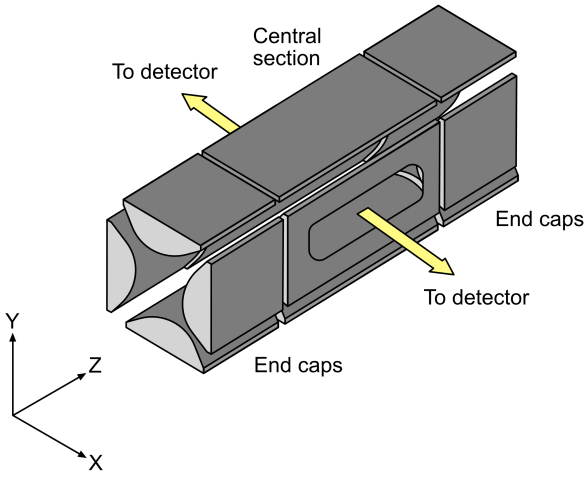


Figure 1.11 | Principles of the LIT. Ions are confined radially by quadrupolar rods and axially by static potentials applied to end electrodes.

1.3.3 Tandem mass spectrometry

MS-based proteomic workflows often include isolation and fragmentation steps for improved sample analysis.[33] This is typically carried out by tandem MS (MS/MS). In MS/MS, two or more mass analysers are coupled together with a collision cell. Typically, the first mass analyser identifies and isolates precursor ions before they are activated into fragments ions in the collision cell and the fragment ions detected in the second mass analyser.

MS/MS involves targeted and untargeted approaches. Untargeted MS/MS methods include data-dependent analysis (DDA) and data-independent analysis (DIA).[37] During DDA the most abundant ions (typically up to five species) from each precursor scan are transmitted to a second mass analyser where they are activated. Analysis is performed by database searching where experimental MS/MS spectra are compared to theoretical fragments of relevant peptides from a database of protein sequences. In contrast, DIA involves activating all ions across a precursor scan's mass range at delineated m/z windows. Data analysis involves a de-multiplexing search where MS/MS are combined into pseudo spectra by relating precursor and fragment ion elution profiles before comparing to known spectra using database searches. DDA is less computationally demanding than DIA and search times are far quicker. DDA datasets may contain gaps caused by variation in ion selection, effecting reproducibility (especially for less abundant ions). On the other hand, DIA introduces less bias as

all ions are activated; moreover, the approach does not necessitate knowledge of a protein sequence prior to analysis, and *de novo* approaches can be employed. However fragment ions may not be unambiguously recovered to precursor ions and algorithms may generate false positives where MS/MS spectra result from several precursors. False positives can also be generated in DDA where identical MS/MS spectra emerge from several different proteins. DDA and DIA can both be employed for protein/peptide quantification. Isobaric tags for relative and absolute quantitation (iTRAQ) is one method where proteins/peptides are chemically labelled at NTDs and side chain amines.[47] Activation of labelled ions generates low molecular mass reporters from which relative protein/peptide quantity can be inferred. However quantitative DIA exhibits lowers sensitivity than DDA due to reduced acquisition period per data point. Targeted MS/MS approaches include selected reaction monitoring (SRM).[37] During SRM a specific precursor ion is designated for activation. Targeted MS/MS approaches facilitate greater specificity and sensitivity for quantification than untargeted methods.

1.3.3.1 Collision-induced dissociation

Collision-induced dissociation (CID) is a widely used ion activation technique. It involves collision of inert gas molecules with selected ions at either high or low-energy. Heavier gases (xenon, argon, etc) can also be employed for more energetic collisions.[48] CID is considered an ergodic activation method where the rate of dissociation is slower than the rate of vibrational energy redistribution.[49] The result is that cleavage occurs at the kinetically weakest bonds.[33] For protonated peptides this is the amide bond; however, instead of cleavage via direct accumulation of vibrational energy at the amide bond, bond cleavage occurs via low-energy rearrangements as described by the mobile proton model.[50] This posits that ionising protons are initially located on basic sites, including arginine and lysine residues, and upon activation, these species are transferred to less basic regions, such as the amide bond.[51] Not only does this weaken the amide bond but it is also rendered more susceptible to nucleophilic attack and subsequent heterolytic cleavage, driving fragmentation. As a result, b_n and y_n ions are primarily formed in CID (Figure 1.12). In addition, non-specific cleavage as well as preferential cleavage at proline and acidic residues may also occur.[52]

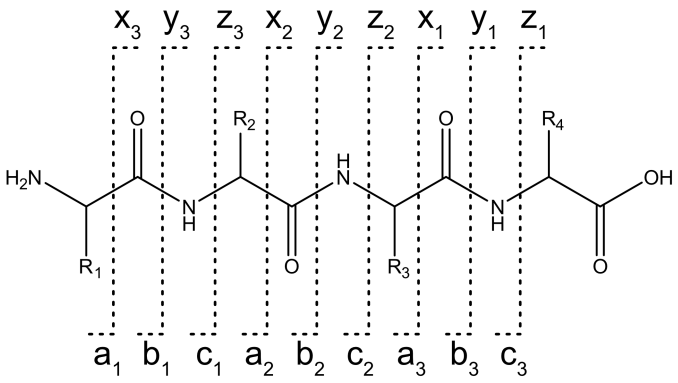


Figure 1.12 | Peptide fragmentation in mass spectrometry. Depending on the fragmentation technique employed, peptides may fragment at three different positions along the side chains (not shown here) and backbone, including the alkyl carbonyl bond (producing a and x ions), the peptide amide bond (producing b and y ions) and the amino alkyl bond (producing c and z ions).

1.3.3.2 Electron dissociation

Electron-based activation methods include electron-capture dissociation (ECD) and electron-transfer dissociation (ETD). In ECD, selected multiply charged positive ions are irradiated with a beam of low-energy electrons.[53] Capture of an electron takes place at the site of an ionising proton which generates an odd-electron radical cation with a reduced charge state, liberating energy.[33] In ETD, ions react with radical anions and are thought to generate similar species to ECD. Both ECD and ETD are non-ergodic activation methods where bond dissociation occurs more rapidly than the rate of energy redistribution.[54] This leads to cleavage of the N-Cα bond and production of c_n and z_n ions (Figure 1.12).[55] Electron-based activation methods provide complementary information to thermal activation methods. They are also especially useful for the study of post-translational modifications which often contain weak bonds.[56]

1.4 Bottom-up proteomics

MS-based bottom-up proteomics identifies and quantifies peptides (as well specific peptide PTMs) across a protein sequence. A typical bottom-up proteomic workflow consists of enzymatic digestion of a protein sample (either in-gel, in-solution or on-bead) into a complex mixture of peptides which are typically analysed by reverse-phase high-performance LC (RP-HPLC) coupled to ESI-MS, often aided by database-searching

algorithms (Figure 1.13).[57] Peptides are well-suited to this type of analysis: they are easily separated by RP-HPLC, ionise efficiently and exhibit predictable fragmentation patterns. This robust workflow is also suitable for high-throughput study and is often employed in systematic and global analysis of proteins. For example, bottom-up proteomics has contributed to the identification of the *Saccharomyces cerevisiae* proteome as well as initial drafts of the human proteome.[58]

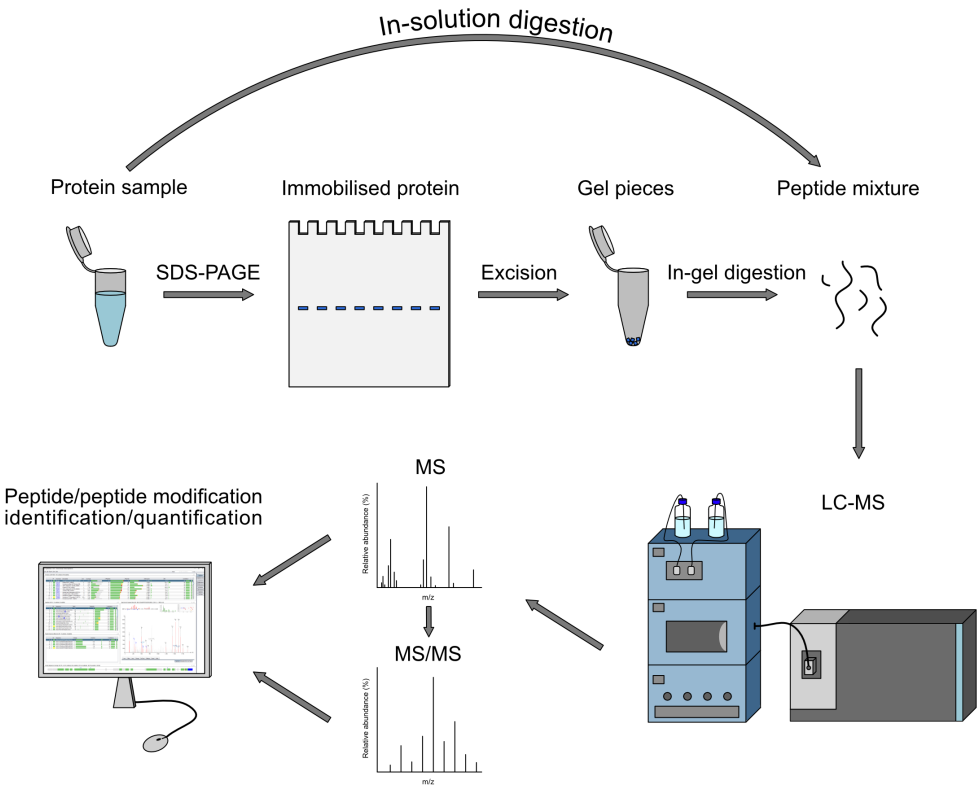


Figure 1.13 | A bottom-up proteomic workflow. A protein of interest is be digested by in-gel or in-solution methods. The peptide mixture is commonly separated using RP-HPLC before being analysed with MS. Peptides may be identified using database analysis of MS/MS fragments. In contrast, peptide modifications can be quantified using additional third-party software.

A common feature across all bottom-up proteomic experiments is protein digestion; however, despite advancements to sample preparation, peptide separation, mass spectrometry and database-searching algorithms, digestion is still usually performed with trypsin.[59] Commercial-grade trypsin displays high specificity and its proteolytic activity can be increased by engineering resistance to autolysis. Protein sequences unquestionably vary in their number of tryptic cleavage sites which potentially limits the number of resolvable peptides and overall attainable sequence coverage. Overly long or short peptides (produced by minimal or excessive proteolysis, respectively) are

unlikely to be resolved by RP-HPLC and MS. As such, much research has been devoted to enhancing the digestion stage of bottom-up proteomics and improving the overall sequence coverage. This has included changes to digestion buffers, reducing/alkylating agents and the introduction of solubilising agents that aid accessibility of proteases to polypeptide cleavage sites.[60,61] An obvious point of change; however, is the protease.[62]

1.4.0.1 Proteases

Several alternative proteases which vary in their amino acid specificity are available for protein digestion. The use of these enzymes (separately and in combination) can improve peptide sequence coverage. Enzyme optimisation is an important feature of bottom-up proteomic experiments. Commercially available proteases include members of the serine (including trypsin, chymotrypsin and elastase) and aspartic acid (including pepsin) classes. These endopeptidases function by directly cleaving the peptide bond. The Schecter and Berger nomenclature has been widely adopted to describe the activity of proteases (Figure 1.14).[63] By this convention protease subsites are termed S whilst corresponding amino acids (which bind subsites) on the polypeptide chain are named P. Both S and P sites are numbered according to their position relative to the scissile peptide bond with the absence and presence of an apostrophe denoting N- and C-terminal sides, respectively.

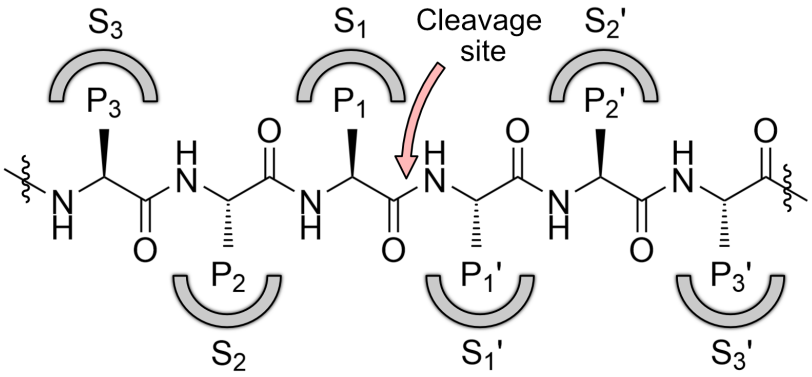


Figure 1.14 | Schechter-Berger nomenclature for protease activity. S and P sites are numbered according to their position relative to the scissile peptide bond with the absence and presence of an apostrophe denoting N- and C- terminal sides.

Serine and aspartic acid proteases achieve catalysis in different ways. Serine proteases consist of a catalytic triad (His57, Ser195 and Asp102).[64] Each residue in this triad plays an important role in catalysis. The serine serves as a nucleophile, attacking the carbonyl carbon of the peptide bond; the histidine accepts a proton from the hydroxyl group of serine and the aspartate hydrogen bonds to histidine to enhance its electronegativity. In contrast, aspartic acid proteases employ the β-carboxyl groups of conserved aspartic acids to activate a water molecule which hydrolyses the peptide bond.[65] Low pH (<2.5) is important for pepsin activity. This ensures protonation of carboxyl groups and disrupts net positive interactions to expose the enzyme's active site.[66] Protease cleavage specificity is primarily dictated by the S1 pocket but additional S pockets also contribute to polypeptide binding.[64] A comprehensive understanding of protease cleavage specificity requires an enormous amount of statistical data. The BRENDA database is one such tool that contains crucial information relating to enzyme specificity and function.[67] The specificity of serine proteases is dictated by their S1 pocket. Trypsin's S1 pocket is negatively charged which predisposes cleavage at arginine and lysine residues (Table 1.1).[68] Chymotrypsin exhibits a hydrophobic S1 pocket; therefore, nonpolar residues such as tyrosine, phenylalanine and tryptophan (and to a lesser extent methionine and leucine) are favoured.[68] Elastase contains a small S1 pocket. Consequently, cleavage at small aliphatic residues such as alanine, glycine and valine is preferred. Pepsin's specificity is dictated by a flexible loop.[69] Several residues on this region, including Tyr75, Gly76 and Thr77, hydrogen bond to the substrate, orienting it at the S1 subsite. Pepsin selectively cleaves at hydrophobic residues, including tyrosine, phenylalanine, tryptophan and leucine. Specificity for leucine and phenylalanine is greatest at pH 1 and decreases above pH 2.

Protease	Preferred P1 residues
Chymotrypsin	Tyrosine, Phenylalanine, Tryptophan, Leucine, Methionine
Elastase	Alanine, Valine, Leucine, Isoleucine, Serine, Glycine
Glu-C	Glutamate
Trypsin	Lysine, Arginine
Pepsin	Tyrosine, Phenylalanine, Tryptophan, Leucine

Table 1.1 | Preferred P1 sites of common proteases.

1.5 Mass spectrometry-based chemical labelling techniques

MS-based chemical labelling techniques are used to study protein-ligand and protein-protein interactions. Experiments follow a bottom-up proteomic design. The interaction of chemical reagents with accessible residues on a protein is usually carried out in the presence and absence of a binding partner. A bound ligand will shield a region of a protein from chemical modification whilst unbound sites remain accessible for labelling (Figure 1.15). Subsequent digestion and LC-MS stages enable identification of mass accumulations on peptides. The extent of chemical modification is compared at a peptide-level between ligand-treated and control samples. Differences in peptide labelling between samples may be caused direct ligand binding or conformational change induced by the interaction. Different MS-based chemical labelling techniques exist depending on the chemical label employed. These include hydrogen-deuterium exchange (HDX), hydroxyl radical protein footprinting (HRPF) and carbene footprinting.[70]

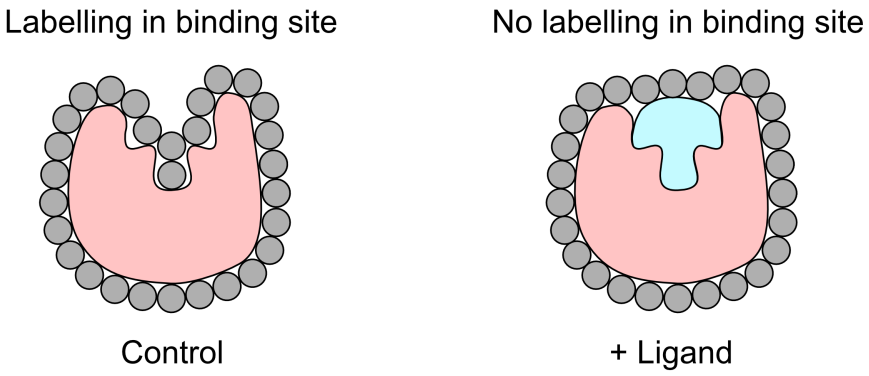


Figure 1.15 | Principles of chemical labelling to study protein interactions. Chemical labelling labelling is carried out on control and ligand-treated protein samples. Binding of the ligand causing masking of the binding site, prevent chemical labelling from occurring whilst in the control sample, this region is accessible for modification. A bottom-up proteomic approach can then be employed to identify peptide-level labelling and comparisons made between the two treatments.

Two distinct categories of chemical labelling approaches have been developed, based on the type of interaction the chemical probe forms with the protein. These include covalent and noncovalent chemical labelling techniques.

1.5.1 Noncovalent labelling

Noncovalent MS-based chemical labelling techniques employ chemical reagents that form noncovalent interactions with proteins. A characteristic feature of noncovalent labelling techniques is that chemical modification is reversible. This is best represented by HDX-MS. *The literature describes HDX-MS as a noncovalent labelling technique despite the covalent (albeit reversible) nature of the N-H bond.*

1.5.1.1 Hydrogen-deuterium exchange

HDX monitors the exchange of hydrogen for deuterium at amide linkages.[71] Other hydrogen atoms, including side chain hydrogens bound to heteroatoms and those covalently bound to carbon display impractical exchange that is not amenable to study (i.e. too fast and too slow, respectively).[72] Deuterium exchange at amide hydrogens is determined by temperature and pH and structural characteristics, namely solvent accessibility and the presence of hydrogen bonding. Slow deuterium uptake is indicative of reduced solvent accessibility and increased hydrogen bonding.[21]

Paterson and colleagues first used HDX-NMR to study protein-ligand interactions between horse cytochrome c and a monoclonal antibody.[73] HDX-NMR displayed poor spatial resolution and was soon replaced by MS which allowed analysis of larger proteins at lower concentrations. Katta and colleagues showed that HDX-MS could be used to probe the conformational changes of ubiquitin when sprayed in deuterated solvent.[74] However analysis of entire proteins limited structural resolution and prevented identification of regions undergoing exchange. Soon after, Zhang and Smith used proteolytic enzymes to digest proteins and subsequent LC-MS steps were employed to analyse deuterated peptides. This permitted increased resolution of deuterium exchange, setting the precedent for chemical labelling mass spectrometry.[75]

In a typical HDX experiment, deuterium exchange is carried out at neutral pH with excess D_2O. This is performed over a range of time points and quenched with low pH (<2.5) and temperature (0 °C).[76] Due to the labile nature of the N-H and N-D bonds, quenching conditions must be subsequently maintained otherwise undesired deuterium-hydrogen back-exchange may occur.[77] As such, downstream processing stages such as proteolysis and LC-separation are restricted. Historically,

this necessitated pepsin for proteolytic digestion of deuterated samples. However additional proteases that function at low pH, including ProAlanase from the fungus *Aspergillus niger* and Nepenthesin from the pitcher plant *Nepenthes*, have recently been discovered and isolated. These have increased the versatility of digestion conditions employed in HDX.[78,79] Tandem MS can be used to elucidate deuteration at the residue level; however, fragmentation techniques are limited to ECD, ETD and under certain conditions ultraviolet photodissociation (UVPD) since CID leads to scrambling of deuterium positions on the peptide.[80] Despite these pitfalls, use of HDX to provide structural assessment of proteins is extensive. For example, HDX has been used to identify the binding site of antibodies to ADAMTS13 which provided a greater understanding of acquired thrombotic thrombocytopenic purpura. In contrast, the technique has been employed in the analysis of the peroxisome proliferator-activated receptor γ (PPARγ). Different modulators were shown to impose varying degrees of dynamicity to the receptor which was linked to the pharmacological profiles of ligands.[81] HDX has also been applied to the study of membrane proteins, including G protein-coupled receptors (GPCRs). The beta-2 adrenergic receptor (β_2AR) was studied in complex with several ligands. Each complex afforded distinctive HDX fingerprints characterised by varying levels of stabilisation and conformational dynamics. In particular full agonists were shown to increase the mobility of transmembrane helix (TM) 8 at specific time points.[82,83] More recently, Yang and colleagues applied the technique to understand the structural relationship between the glucagon receptor and hormone glucagon. The authors showed that the peptide bound and stabilised an open conformation of the receptor. This allowed them to suggest a conformational selection model of ligand binding.[84] HDX-MS instrumentation is also highly advanced and deuteration and quenching steps can be fully automated. Quenching may be also be performed at as little as 50 ms time points.[85] Seetaloo and colleagues recently analysed hydrogen-exchange rates of wild type α-synuclein on the millisecond scale under extracellular, intracellular and lysosomal conditions.[86] The aggregation kinetics of α-synuclein fibrils were compared to structural differences in the monomer. This allowed the authors to implicate specific CTD residues in initial nucleation events that promoted aggregation of the protein, providing further insight into the pathogenesis of Parkinson's Disease.

1.6 Covalent labelling

Covalent labelling techniques employ chemical probes that covalently bond to a protein. The primary advantage of covalent approaches compared to noncovalent techniques is the irreversible nature of labelling which allows for robust downstream processing.[70] Common covalent labelling agents include hydroxyl radicals and carbenes but other irreversible probes exist, including trifluoromethyl and sulphate radicals, diethylpyrocarbonate (which selectively labels histidine) and glycine ethyl ester (which selectively labels acidic residues).

1.6.1 Hydroxyl radical protein footprinting

HRPF employs hydroxyl radicals to covalently modify protein structure. Mass shifts of +16 Da from the formation of a hydroxyl group and +14 Da from the formation of a carbonyl are among the most frequent adducts but further reactions involving particular amino acid side chains may occur.[70] Radical additions are rapid, occuring on a μs timescale.

Traditionally, hydroxyl radicals were generated through Fenton chemistry as well as the radiolysis of water.[70] The Fenton reaction relies on oxidation of Fe^{2+} and Fe^{3+} with H_2O_2 to generate hydroxyl radicals. Benefits of this methodology include broad accessibility to users without the need for expensive instrumentation; however, reactions are carried out at a low pH which is not necessarily suited to the study of proteins. Proteins may also be oxidised in the presence of H_2O_2. In contrast, water radiolysis and UV photolysis utilises Synchrotron-derived high energy photons to ionise water and produce hydroxyl radicals. The primary advantage of water radiolysis over Fenton chemistry is the use of water as a hydroxyl radical precursor. The administration of radiation is also rapid and highly controllable; however, Synchrotron sources are limited and radiolysis conditions may cause decarboxylation of Glu and Asp residues.

1.6.1.1 Fast photochemical oxidation of proteins

Fast photochemical oxidation of proteins (FPOP) is a recent development to HRPF made by Hambly and Gross.[87] FPOP utilises photolysis of hydrogen peroxide in the far-ultraviolet (UV) region to produce hydroxyl radicals Figure 1.16.[88] Compared to the prior methodology, FPOP allows for increased control over protein oxidation. A flow cell was employed to ensure that a single protein was irradiated once with a 248 nm excimer laser (17 ns laser pulse width and 50 mJ/s). Scavenging (preventing oxidation before irradiation) and quenching (preventing oxidation after irradiation) agents were introduced to prevent over-oxidation of the sample. The incorporation of unirradiated protein also allowed for background oxidation to be monitored.[89]

FPOP suffers from several drawbacks, namely discriminate reactivity towards amino acids (as seen in general with HRPF). Hydroxyl radicals display a preference toward sulfur-containing, heterocyclic and aromatic side-chains. The reactivity is Cys > Trp > Tyr > Met > Phe > His > Arg > Ile > Leu > Val > Pro > Gln > Thr > Lys > Ser > Glu > Ala > Asp > Asn > Gly (reactivities do not vary between FPOP and other HRPF-based methods). Rate constants vary some 1000-fold and modification of Ser, Glu, Ala, Asp, Asn and Gly is very low in FPOP studies.[90] Deconvolution of HRPF data is complex and analysis software is not as widely available compared to other labelling strategies.[89] FPOP is also limited to buffers that do not scavenge hydroxyl radicals. The use of irradiation wavelengths in the far-UV region *may* perturb protein structure; however, irradiation is conducted on a μs timescale so labelling should occur before any change to protein structure.

FPOP has been used to characterise various protein-ligand interactions including those between the human epidermal growth factor receptor (EGFR) and adnectin 1.[91] Decreased hydroxyl radical modification on domain 1 was suggested to indicate ligand binding whilst changes in labelling on distal residues were said to represent allosteric changes in residue side-chain orientation. Since hydroxyl radicals are alike molecular water, HRPF is also able to probe the solvent accessible surface area (SASA) of molecules. Sheng and colleagues used FPOP to report on the SASA of human copper-zinc superoxide dismutase 1 (SOD1), a β-barrel which is implicated in the early stages of neurodegenerative diseases.[92] They showed that hydroxyl radical modification

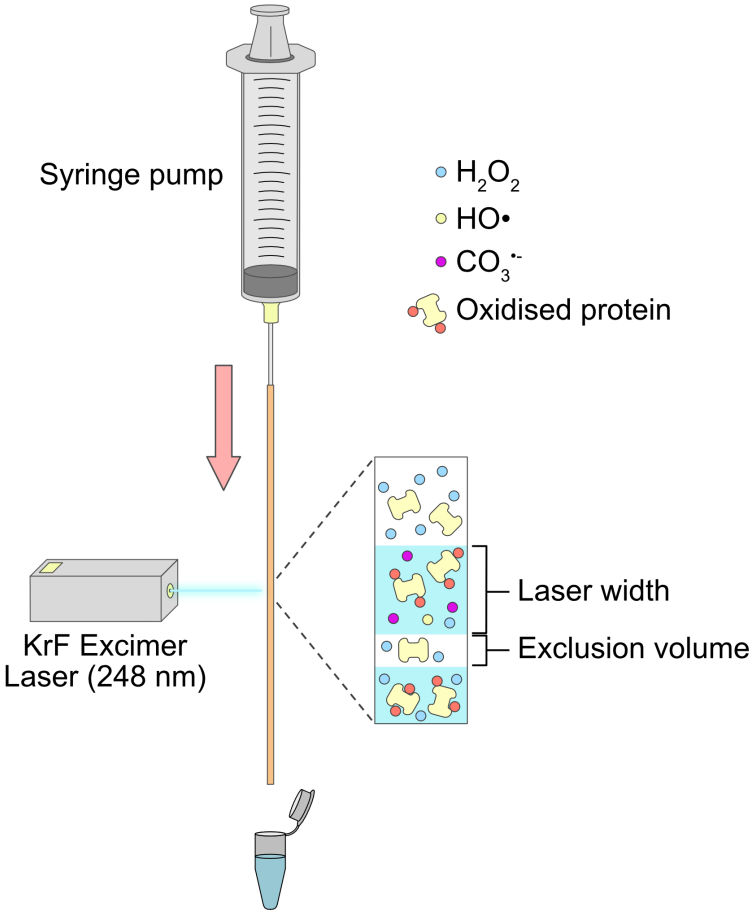


Figure 1.16 | Principles of FPOP. A solution containing a protein of interest and H_2O_2 is passed through 150–450 µm diameter silica tubing. A region of the tubing containing a transparent window is irradiated with an excimer laser, generating hydroxyl radicals that irreversibly label the protein. The sample flow and laser are adjusted to ensure that each protein is only irradiated once. Labelled samples are collected into a solution containing methione and catalase, to mitigate additional modification by remaining reactive species. Other additives are also added to samples prior to irradiation to prevent over oxidation of proteins.

correlated well with predicted SASA; however, the authors did observe modification of two residues with extremely low solvent exposure, located within the pore of the β-barrel. They suggested this was caused by either partial unfolding of the protein or long-distance tunneling from radical species at the surface of the pore. FPOP has also been applied to several GPCR systems. Du and colleagues used FPOP in combination with HDX to capture the conformational dynamics of β_2AR-G protein complexes. The authors highlighted that initial intermediate states acted as important selectivity filters that determined the endmost nucleotide-free β_2AR-G protein complex.[93] Unfortunately, due to sequence coverage issues the authors were not able to report on G protein contact sites of the receptor.

1.6.2 Carbene protein footprinting

Carbene protein footprinting is a covalent labelling technique. It exploits the inherent reactivity of carbenes to irreversibly label proteins.[94]

1.6.2.1 Carbenes

Carbenes are molecules that feature a neutral carbon with a valence of two with two unshared electrons. Carbenes can be further classified depending on their electronic configurations. Singlet carbenes contain a pair of nonbonding electrons within a shared orbital whilst triplet carbenes contain a single electron in two different orbitals and may be considered as diradicals (Figure 1.17).[95] The ground-state spin multiplicity of carbenes is dictated by steric and electronic effects of chemical substituents on the molecule where a large σ-p_π energy gap (>2 eV) favours a the singlet state and a small energy gap (<1.5 eV) drives the triplet state. Carbenes can insert into any X-H (where X is C, O, N, S), C=C and C=O bond on a nanosecond timescale. Singlet and triplet carbenes display divergent reactivity but share X-H insertion although mechanisms vary between the two.[95]

Carbenes can be formed in a number of ways but the photolysis of diazirines is commonly used in protein footprinting experiments. Diazirines are characterised by a constrained unsaturated three-membered ring consisting of two nitrogens and a carbon atom. Irradiation of diazirines at ~350 nm causes release of N_2 and formation

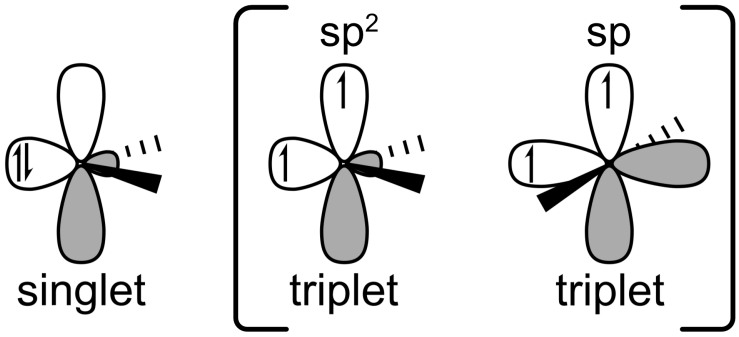


Figure 1.17 | Electronic configuration of carbenes. Singlet carbenes are spin-paired whilst triplet carbenes are paramagnetic and may be considered as diradicals.

of a carbene. This process may also generate linear diazoisomers. These display reduced reactivity compared to carbenes and may compromise labelling results through diffusion of the probe. Aryl-substituted diazirines have been reported to provide an increased carbene:diazo ratio.[96] Moreover, incorporation of a trifluromethyl group to the C1 diazirine carbon was shown to reduce reactivity of diazoisomers (Figure 1.18). Aryldiazirines also display a small energy difference between singlet and triplet states primarily react through the higher energy singlet state which is beneficial for footprinting experiments where carbene insertion is faster than the triplet state.

Carbene labelling of proteins was first reported in 2000 by Richards and colleagues.[97] Diazirine gas was continuously bubbled through a mixture of hen egg-white lysozyme (HEWL) and a ribonuclease-S peptide whilst irradiated with an Hg arc source. Very low carbene modification of the protein was observed due to the poor solubility of methylene in solution. Jumper and Schreimer reported a more efficient labelling precursor, photoleucine. The diazirine was employed to study the calmodulin-M13 complex. Irradiation was performed with an Nd:YAG pulsed laser (355 nm, 1000 Hz, 12.9 µJ/pulse). The authors showed reduced protein-level labelling on calmodulin bound to M13 compared to unbound protein caused by shielding effects at the binding interface.[98] However carbene labelling was still impeded by long irradiation times, high reagent concentrations, poor diffusion of the reagent prior to photolysis and quenching by water. The same group later revisited carbene footprinting of the calmodulin-M13 complex but instead employed two new diazirine precursors: 2-amino-4,4-azipentanoic acid and 4,4-azipentanoic acid.[94] Importantly, a flash-freezing step prior to irradiation with a Nd:YAG pulsed laser (355 nm, 10 Hz, 20 mJ/pulse) was

utilised to minimise diffusion of the carbene. Irradiated samples were further digested and fragmented with ETD and CID. The authors reported decreased peptide and residue level labelling in M13-treated calmodulin samples compared to apo protein. These masking effects corresponded to regions of the calmodulin-M13 binding site.

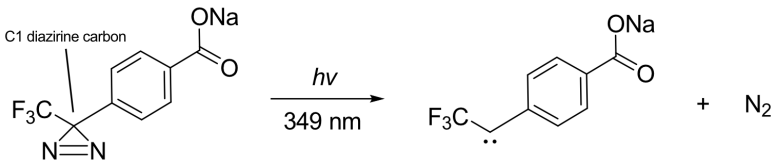


Figure 1.18 | Photolysis of NaTDB. Irradiation of NaTDB with a ~350 nm laser generates formation of a highly reactive carbene species.

Another development to the platform recently emerged when Manzi and colleagues reported a new aryldiazirine precursor, sodium 4-(3-Trifluoromethyl)-3H-diazirin-3-yl)benzoic acid (NaTDB) (Figure 1.18). The diazirine was designed for higher solubility in water and increased interaction with proteins prior to carbene formation. This permitted increased carbene modification at lower precursor concentration.[99] The authors went on to map several protein-ligand interactions using the probe, including between HEWL and penta-*N*-acetylchitopentaose (NAG_5) as well as between ubiquitin specific peptidase 5 (USP5) and ubiquitin. The same group later reported carbene labelling of the IMP OmpF. They demonstrated insertion of the probe into detergent micelles, highlighting the power of the technique to investigate insoluble protein systems.[100] However, carbene footprinting has not yet been applied to any α-helical membrane protein. These display much reduced chemical accessibility through helical packing and greater encapsulation from detergent molecules compared to β-barrels. More recently, carbene footprinting was employed to identify the interaction site of the gladiolin polyketide synthase subunits GbnD4 dehydratase docking (DHD) domain and GbnD5 dehydratase (DH) domain.[101] Lu and colleagues explored footprinting at the sub-residue level to gain insight into how the estrogen-related receptor alpha (ERRα) interacted with potential agonists.[77] Carbene footprinting has also been used to characterise six antibody binding epitopes to the major histocompatibility complex class I chain-related molecule A (MICA) and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) antibodies. Hogan and colleagues employed two separate diazirine probes, NaTDB and 3-azibutanol to achieve complementary labelling information

which facilitated fast and high-resolution epitope mapping at both the peptide and residue level.

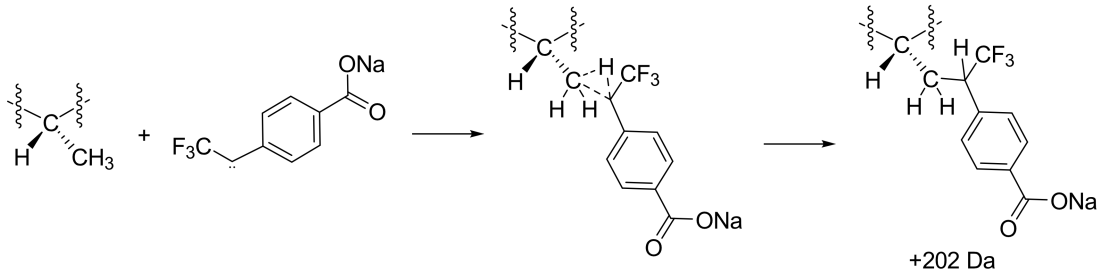


Figure 1.19 | NaTDB insertion into alanine. The singlet inserts into C-H bonds in a one-step process involving a three-carbon centre transition state. The insertion product generates a +202 Da mass shift.

Carbene footprinting maintains several advantages over HRPF. The indiscriminate nature of carbene insertion permits chemical labelling across all 20 amino acids (Figure 1.19); however, localisation of the diazirine to specific amino acid side-chains prior to carbene formation may incorporate selectivity into carbene insertion. For example, the aryldiazirine developed by Manzi and colleagues was thought to favour interactions with basic and hydrophobic residues due to the probe's aromatic ring and acidic functional group.[99] This finding was further corroborated by Ziemianowicz and colleagues.[102] Carbene insertion also displays easily recognisable mass shifts and insertion products compared to HRPF (where oxidised residues and reaction cascades can be difficult to link back to the primary labeling event). Carbenes are much less persistent than hydroxyl radicals owing to their shorter half life (nanoseconds compared to microseconds) and carbene activation in a frozen state also limits diffusion of probe minimising quenching effects. An important consideration of chemical labelling techniques is the impact on the sample, and carbene footprinting does not influence protein integrity and the structure of HEWL remained unperturbed in the presence of the diazirine.[99] Furthermore, irradiation is performed outside of the absorbance region of amino acids, rendering the process harmless to protein structure.[99] Additionally, the recent development of PepFoot, a semi-automated carbene footprinting data analysis software, has ensured rapid and accurate interpretation of labelling information.[103] The software distills out extracted ion chromatograms (EICs) for labelled (+202 Da) and unlabelled peptide (no mass shift) fragments from LC-MS analysis. Integration of

labelled and unlabelled chromatographic peak areas allows for the peptide fractional modification (F_{mod}) to be determined where:

$$F_{mod} = \frac{A_{labelled\ peptide}}{A_{unlabelled\ peptide} + A_{labelled\ peptide}} \tag{1.3}$$

Carbene modification at the residue level can also be determined by MS/MS. Fragmentation of labelled peptides permits assignment of labelled and unlabelled fragments and quantification of sub-peptide labelling information for more precise binding site assignments to be made.[99]

$$F_{mod(n_i)} = \frac{I(n_{i_{labelled\ peptide}})}{I(n_{i_{labelled\ peptide}}) + I(n_{i_{unlabelled\ peptide}})} \tag{1.4}$$

The primary drawback to carbene footprinting is the size of the diazirine probe. Larger labelling reagents may result in shielding effects over a greater area than the exact protein-binding partner contact interface leading to misleading results (Figure 1.20). Hogan and colleagues observed greater masking effects with NaTDB than 3-azibutanol during their antibody-epitope labelling research.[104] This was attributed to the smaller size and increased flexibility of 3-azibutanol compared to NaTDB. The reduced chemical accessibility of larger diazirine probes may also lead to reduced chemical modification of compact protein domains compared to smaller labelling reagents, preventing informative analysis of these regions. Initial labelling optimisation experiments are often conducted prior to differential study to assess global levels of modification. Indeed, peptides which display an f_{mod} of 0 will not be of use during a differential study (they may however inform on chemical accessibility of the protein though). Furthermore, whilst the nature of carbene insertion is relatively indiscriminate, diazirine probes - as mentioned above - exhibit varying amino acid specificity prior to carbene formation. This may lead to carbene insertion preferences that potentially reduce labelling on particular regions of the protein.[102] Jumper and colleagues identified a clear insertion bias towards glutamate residues when labelling calmodulin with 2-amino-4,4-azipentanoic acid. This was suggested to be caused by electrostatic interactions between carboxylic acids and reagent amino groups.[94] Lastly, carbene footprinting is a relatively recently developed technique and the carbene labelling infrastructure lags behind older, more refined systems like HDX and FPOP.

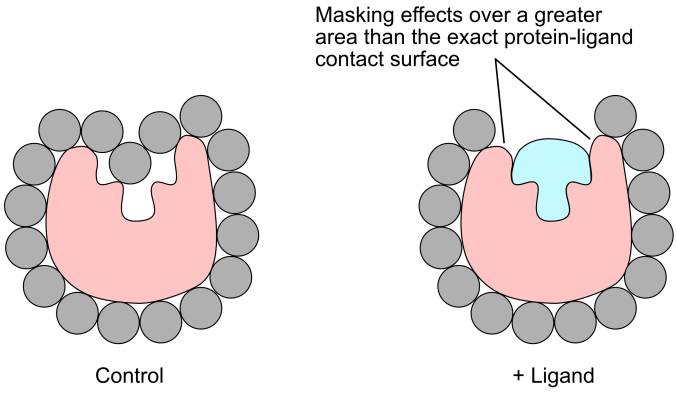


Figure 1.20 | Increased masking of a protein contact interface of large chemical labels. Large chemical labelling probes are more likely to show extended masking over a protein contact interface than small probes which are able to more accurately 'trace' the exact binding region.

2 Methodology

All aqueous solutions were prepared using Ultrapure water (18.2 MΩ) from a Millipore water purification system. All buffers were purchased from Thermo Fisher Scientific or Merck, unless otherwise specified.

2.1 Protein production

2.1.1 PfMATE

2.1.1.1 Expression

OverExpress C43 (DE3) Competent Cells (Lucigen) were thawed on ice. PfMATE plasmid (100 ng/µL, 0.5 µL) was added to cells (25 µL) and stirred briefly with a pipette tip. The mixture was transferred to a pre-chilled 15 mL Falcon tube (Thermo Fisher Scientific) and incubated on ice for 30 min. Cells were heat shocked in a 42 °C water bath for 45 s and incubated on ice for 2 min. SOC Recovery Media (475 µL, Lucigen) was added to the cells and the tube was placed in an Innova 44 Incubator Shaker Series (New Brunswick) at 200 rpm for 1 h and 37 °C. The transformed cells (150 µL) were plated on lysogeny broth (LB)-agar plates containing carbenicillin (50 µg/mL) and incubated at 200 rpm overnight at 37 °C in an Innova 44 Incubator Shaker Series (New Brunswick). A single colony was inoculated in LB (8 mL) containing carbenicillin (50 µg/mL) and incubated overnight at 37 °C. The overnight culture (8 mL) was inoculated in LB (1 L) containing carbenicillin (50 µg/mL) in a 2 L flask and incubated

at 37 °C and 230 rpm in an Innova 44 Incubator Shaker Series (New Brunswick) until the optical density at 600 nm (OD_{600}) measured between 0.6 and 0.8 using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix). IPTG was added to the culture (100 µg/mL) and incubated for a further 3 h at 37 °C.

2.1.1.2 Purification

Buffer	Composition
Lysis	300 mM NaCl, 5 mM 2-mercaptoethanol (βME), 20 mM Tris pH 7.4
Resuspension	100 mM NaCl, 20 % glycerol, 5 mM βME, 20 mM Tris, pH 7.4
A	200 mM NaCl, 10 % Glycerol, 20 mM Imidazole, 0.025 % *N*-dodecyl-β-maltoside (DDM), 50 mM Tris pH 7.4
B	100 mM NaCl, 10 % Glycerol, 500 mM Imidazole, 0.025 % DDM, 50 mM Tris pH 7.4
Dialysis	150 mM NaCl, 10 % Glycerol, 20 mM Imidazole, 0.025 % DDM, 50 mM Tris pH 7.4
SEC	130 mM NaCl, 10 % Glycerol, 0.025 % DDM, 50 mM Tris pH 7.4

Table 2.1 | PfMATE purification buffers

Cells were collected by centrifugation in an Avanti JXN-26 centrifuge (Beckman Coulter) at 5000 *g* using a JLA-8.1000 Rotor (Beckman Coulter) for 10 min at 4 °C. Cells were resuspended in phosphate-buffered saline (PBS) before being pelleted by centrifugation at 5000 *g* using a JA-25.50 Rotor (Beckman Coulter) for 10 min at 4 °C and stored at −80 °C. Cells were defrosted on ice before being resuspended in lysis buffer that had been supplemented with cOMPLETE Protease Inhibitor Cocktail (one tablet per 50 mL, Merck) (Table 2.1). Lysis buffer (20 mL) was added per litre of culture. The cell suspension was passed through a M-110 PS microfluidizer (Microfluidics) five times at 19,000 psi. The cell lysate was centrifuged at 2000 *g* using a JA-25.50 Rotor (Beckman Coulter) for 25 min at 4 °C. The supernatant was collected and centrifuged at 100 000 *g* in an Optima XPN-80 Ultracentrifuge (Beckman Coulter) using a Type 45 Ti (Beckman Coulter) Rotor for 2 h at 4 °C to pellet cell membranes. Cell membranes were resuspended in ice-cold resuspension buffer and homogenized using a borosilicate dounce homogeniser (Thermo Fischer Scientific). The concentration of the membrane resuspension was determined using a Protein 660 nm assay. Pierce 660 nm protein assay reagent (150 µL, Thermo Fischer Scientific) was added to 2x, 5x and 10x dilution of the membrane resuspension (10 µL) and a linear gradient of bovine serum albumin (BSA, 10 µL) standard in a 96-well plate (Starlab) and incubated for 5 min at room temperature. The concentration of the resuspended samples was estimated using

a Pherastar FSX Plate Reader (BMG Labtech) and the membrane resuspension was diluted (1 mg/mL) in resuspension buffer. This was supplemented with DDM (1 %, Anatrace) and incubated overnight at 4 °C with gentle agitation.

Solubilised membranes were clarified by centrifugation at 20 000 *g* using a JA-25.50 Rotor (Beckman Coulter) for 25 min at 4 °C before the supernatant was filtered using a Millex-GV syringe filter (0.22 µm, 33 mm, Merck). A 5 mL HisTrap (GE Healthcare) was installed on an AKTA Pure (Cytiva) that had been equilibrated in buffer A at a flow rate of 2 mL/min. Solubilised membranes were loaded onto the column at a flow rate of 0.5 mL/min and washed with 10 column volumes of buffer A. Immobilised membrane proteins were eluted with two column volumes of buffer B (Figure 2.1). Fractions containing eluted membrane protein were combined before beta-mercapthoethanol (βME, 5 mM) and His-tagged TEV Protease (1 mg/mL) provided by Dr Antonio Biasutto (OMass Therapeutics - Oxford, UK) were also added. Membrane protein samples were transferred to a Slide-A-Lyzer dialysis cassette (10 kDa MWCO, 3–12 mL, Thermo Fisher Scientific) and dialysed against dialysis buffer (5 L) overnight at 4 °C.

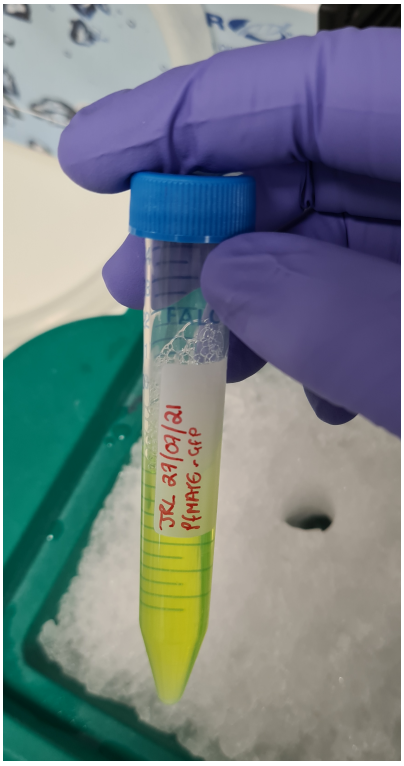


Figure 2.1 | **GFP-tagged PfMATE.** The GFP-tag on the PfMATE construct generated fluorescence in IMAC-purified protein.

The dialysed sample was recovered from the dialysis cassette and filtered using a Millex-GV syringe filter (0.22 µm, 33 mm, Merck). A 5 mL HisTrap (GE Healthcare) was installed on an AKTA Pure (Cytiva) that had been equilibrated in buffer B at a flow rate of 2 mL/min. TEV-cleaved membrane proteins were loaded onto the column at a flow rate of 0.5 mL/min and the flow-through was collected. The column was washed with 100 % buffer D for two column volumes and the eluent collected. Fractions from the flow through were concentrated to a final volume of approximately 2 mL using an Amicon Ultra-15 filter (50 kDa, MerckMillipore). A Superdex 200 Increase 10/300 GL (GE Healthcare) was installed on an AKTA Pure (Cytiva) and equilibrated in two column volumes of SEC buffer at a flow rate of 0.5 mL/min. Purified protein (500 µL) was injected onto the column before being washed with one column volume of SEC buffer. Fractions (1 mL) were collected in a Masterblock 96-well plate (Greiner Bio-One). Peaks were analysed by SDS-PAGE and relevant fractions were combined and concentrated using an Amicon Ultra-15 filter (50 kDa, MerckMillipore). Purified protein was transferred into 1.5 mL Eppendorf tubes, flash-frozen and stored at −80 °C.

2.1.2 Nanobodies

2.1.2.1 Expression

Genes encoding Nb80 and Nb60 were cloned into pET22b(+) expression vectors containing an N-terminal pelB leader sequence and C-terminal His-tag (GenScript). Plasmid (100 ng/µL, 1 µL) in nuclease-free water was mixed with OverExpress C43 (DE3) Competent Cells (50 µL, Lucigen) and incubated on ice for 30 min. Cells were heat shocked for 45 s at 42 °C in a water bath before incubating on ice for 2 min. SOC Recovery Medium (150 µL, Thermo Fisher Scientific) was added to cells and incubated at 220 rpm and 37 °C for 1 h. Cells (100 µL) were spread on LB agar plates supplemented with carbenicillin (50 µg/mL) and incubated at 37 °C for 18 h. A single, distinct colony was picked from a plate and inoculated in LB (50 mL) supplemented with carbenicillin (50 µg/mL) and incubated at 220 rpm and 37 °C for 18 h. Cultures were diluted in LB (OD_{600} 0.1, 1 L) supplemented with carbenicillin (50 µg/mL). Cultures were incubated at 220 rpm and 37 °C until an OD_{600} value of 0.6 was reached using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix). At this point cultures were removed

from the incubator for 30 min whilst the incubator temperature was lowered to 18 °C. Cultures were induced with IPTG (1 mM) and returned to the incubator for 18 h.

2.1.2.2 Purification

Buffer	Composition
TES	20 mM EDTA , 25 mM Tris pH 8, 20 % sucrose (w/v)
TES/4	5 mM EDTA, 5 mM Tris pH 8, 5 % sucrose (w/v)
Wash	20 mM HEPES pH 7.5, 200 mM NaCl, 5 mM Imidazole
Elution	20 mM HEPES pH 7.5, 200 mM NaCl, 150 mM Imidazole

Table 2.2 | Nanobody purification buffers

Cells were centrifuged at 6000 g in 1 L polypropylene bottles (Beckman Coulter) using an Avanti JXN-26 high-speed centrifuge (Beckman Coulter) containing a JLA-8.1000 rotor (Beckman Coulter) for 10 min at 4 °C. The supernatant was removed and the cells were resuspended in ice cold TES buffer supplemented with two cOMPLETE protease inhibitor tablets (Merck) (Table 2.2). These were incubated at 4 °C with stirring for 1 h. TES/4 buffer was added to the resuspended cells before stirring on ice for 1 h. The cell lysate was centrifuged at 10 000 g for 30 min at 4 °C. The supernatant was transferred to clean tubes and $MgCl_2$ (2 mM) was added to quench free EDTA. Cell lysate was incubated with Talon resin (1 mL) for 1 h at 4 °C. The supernatant was added to a 2.5 cm x 30 cm glass Econo-Column (BioRad) and the flow-through collected. Immobilised protein was washed with 20 CV wash buffer. Protein was eluted with elution buffer. Protein was buffer exchanged into wash buffer, concentrated and stored at −80 °C.

2.1.3 Beta-1 adrenergic receptor

2.1.3.1 Expression

Sf9 cells containing overexpressed beta-1 adrenergic receptor (β_1AR) with an N-terminal FLAG and strep tag and C-terminal His-tag were purchased from GenScript and provided by OMass Therapeutics; however, I performed subsequent protein purification steps.

2.1.3.2 Purification

Buffer	Composition
Lysis	1 mM EDTA, 20 mM Tris pH 8
Resuspension	1 mM EDTA, 20 mM Tris pH 8
Washing	0.2 mM EDTA, 20 mM Tris pH 8
Membrane resuspension	20 mM Tris pH 8
Solubilisation	700 mM NaCl, 6 mM Imidazole, 3 % DDM, 20 mM Tris pH 8
Dilution	350 mM NaCl, 3 mM Imidazole, 20 mM Tris pH 8
A	350 mM NaCl, 3 mM Imidazole, 0.05 % DDM, 20 mM Tris pH 8
B	350 mM NaCl, 250 mM Imidazole, 0.05 % DDM, 20 mM Tris pH 8

Table 2.3 | β_1AR purification buffers

The cell pellet was defrosted on ice and resuspended with lysis buffer (15 mL) and transferred to a 500 mL measuring beaker (Table 2.3). Cell lysis buffer was added to the pellet to bring the final volume to 300 mL. The resuspension was passed twice through a M-110 PS microfluidizer (Microfluidics) at 8,000 psi. This was collected and transferred to ultracentrifuge tubes (Beckman Coulter) and centrifuged at 9000 g for 20 min in an Optima XPN-80 Ultracentrifuge (Beckman Coulter) using a Type 45 Ti rotor (Beckman Coulter) for 20 min at 4 °C to pellet cell debris. The supernatant was collected and transferred to 65 mL ultracentrifuge tubes (Beckman Coulter). Samples were centrifuged at 170 000 g in an Optima XPN-80 Ultracentrifuge (Beckman Coulter) using a Type 45 Ti (Beckman Coulter) Rotor for 2 h at 4 °C to pellet cell membranes. Membranes were transferred to a borosilicate dounce homogeniser (Thermo Fischer Scientific) and manually homogenised in resuspension buffer (40 mL). The resuspension was made up to a final volume of 300 mL with resuspension buffer and transferred to 65 mL ultracentrifuge tubes (Beckman Coulter) before centrifuging at 170 000 g in an Optima XPN-80 Ultracentrifuge (Beckman Coulter) using a Type 45 Ti (Beckman Coulter) rotor for 2 h at 4 °C. This process was repeated twice using wash buffer and then membrane resuspension buffer with the latter being resuspended to a final volume of 30 mL in 15 mL Falcon tubes (Thermo Fisher Scientific). These were flash frozen in liquid nitrogen and stored at −80 °C overnight.

Membranes were defrosted on ice and incubated at an equal volume of solubilisation buffer before vortexing for a few s. Solubilised membranes were centrifuged at 120 000 g in an Optima XPN-80 Ultracentrifuge (Beckman Coulter) using a Type 45 Ti (Beckman Coulter) Rotor for 2 h at 4 °C to pellet any remaining

unsolubilised cell membranes. The supernatant was collected and diluted 1.5 fold with dilution buffer. The solution was passed through a Millex-GV syringe filter (0.45 µm, 33 mm, Merck). A 5 mL HiTRAP TALON column (Cytiva) was installed on an AKTA Pure (Cytiva) that had been equilibrated in buffer A at a flow rate of 2 mL/min. The filtered supernatant was loaded onto the column at a flow rate of 0.5 mL/min and the flow-through collected. The column was washed with 100 % buffer B for 10 column volumes and the eluent was also collected. Fractions were run on a 12 % SDS-PAGE gel and eluent corresponding to β_1AR was concentrated to a final volume of 7 mL using a protein concentrator (100 kDa, Thermo Fisher Scientific).

A HiTRAP 5 mL Sephadex G-25 desalting column (Cytiva) was installed on an AKTA Pure (Cytiva) that had been equilibrated in buffer A at a flow rate of 2 mL/min. Concentrated β_1AR was injected onto a 2 mL loop and eluted in buffer A at a flow rate of 1 mL/min in a 96-well plate. Fractions corresponding to β_1AR were collected, concentrated (0.75 mg/mL) and stored at −80 °C.

2.1.4 Beta-1 adrenergic receptor-nanobody binding assays

2.1.4.1 Size-exclusion chromatography co-elution of the beta-1 adrenergic receptor-nanobody complex

A Superdex 200 Increase 10/300 GL (GE Healthcare) was installed on an AKTA Pure (Cytiva) and equilibrated in two column volumes of buffer A (Table 2.3) at a flow rate of 0.3 mL/min. β_1AR-Nb ternary complex (200 µL) was injected onto the column before being washed with one column volume of buffer A at 0.3 mL/min. Fractions (1 mL) were collected in a Masterblock 96-well plate (Greiner Bio-One).

2.1.4.2 Sodium dodecyl sulfate–polyacrylamide gel electrophoresis and western blot analysis

Fractions corresponding to the three major SEC peaks from isoprenaline-Nb80 and carazolol-Nb60-treated samples were analysed by SDS-PAGE (4–12 %, Invitrogen) at 200 V, constant voltage, for 40 min. For Nb80-treated samples, two SDS-PAGE gels were run. One was analysed by Coomassie staining and the other by western blotting. The unstained gel was transferred to a nitrocellulose membrane using a Trans-Blot

Turbo Transfer System (BioRad). The membrane was incubated with SEA BLOCK Blocking buffer (Thermo Fisher Scientific) for 1 h at room temperature. The solution was discarded and replaced with primary anti penta histidine (BioRad) in washing buffer (5 µg antibody, 1x PBS, 0.2 % v/v Tween 20) and incubated with agitation overnight at 4 °C. The membrane was washed three times for 5 min with washing buffer (1x PBS, 0.2 % v/v Tween 20, 5 mL). The membrane was incubated with horseradish peroxidase secondary antibody conjugate (BioRad) for 1 h at 4 °C with agitation. This solution was discarded and the membrane washed three times for 5 min with washing buffer (1x PBS, 0.2 % v/v Tween 20, 5 mL). The membrane was incubated with Clarity Western ECL substrate (1:1 peroxide reagent and luminol, 7 mL, BioRad) and immediately imaged using the CCD-based mode on a ChemiDoc MP System (BioRad).

2.2 Production of photochemical probe

The free acid of TDBA was purchased from Merck. The sodium salt of TDBA (NaTDB) (100 µM solution) was made by adding a substoichiometric volume of sodium hydroxide (100 mM, 977 µL) to TDBA (25 mg) with stirring for 1 h at room temperature. The solution was filtered through a cotton wool plug to remove excess TDBA free acid. The pH of the resulting clear solution was checked with universal indicator paper to be ~7.5. The solution was stored in a clean vial at 4 °C.

2.3 Photochemical labelling

2.3.1 Eukaryotic initiation factor 4A

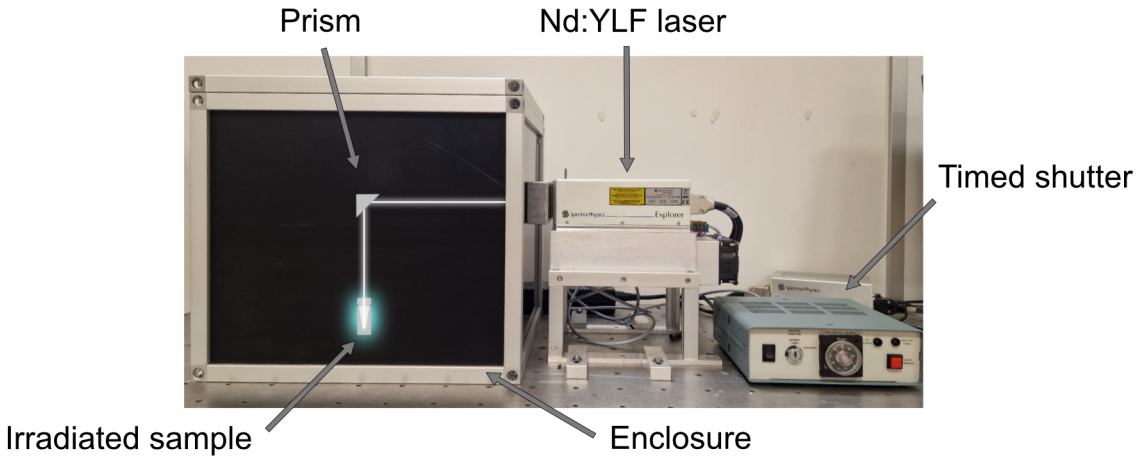


Figure 2.2 | Photochemical label set up. A Spectra Physics Explorer 349 laser operating at 349 nm, with a repetition frequency of 1000 Hz, and a pulse energy of 125 µJ is vertically refracted into flash-frozen samples by a 45° prism at timed intervals.

Hippuristanol was provided by Dr Jerry Pelletier (McGill University - Montreal, Canada) whilst eIF4A was provided by Dr Vasileios Paschalis (University of Nottingham - Nottingham, UK). Hippuristanol in 10 % methanol (280 µM, 1 µL) was added to a buffered (25 mM HEPES pH 8.0, 100 mM NaCl, 10 % glycerol) solution of eIF4A (14 µM, 9 µL). Methanol (10 %, 1 µL) was added to a buffered solution of eIF4A (14 µM, 9 µL). Samples were incubated for 10 min. An aqueous solution of NaTDB (40 mM, 10 µL) was combined with ligand-treated and control samples and incubated for a further 10 min. Aliquots (5 µL) were transferred to tapered autosampler vials (four replicates) and flash-frozen with liquid nitrogen. Samples were irradiated for 15 s using a Explorer One 349 laser (actively Q-switched Nd:YLF laser 349 nm wavelength, 1000 Hz repetition frequency, 125 µJ pulsed energy, Spectra Physics) that was vertically refracted into the vials by a 45° mirror (Figure 2.2).

2.3.2 *n*-Dodecyl-β-maltoside/PfMATE

2.3.2.1 *n*-Dodecyl-β-maltoside

An aqueous solution of NaTDB (40 mM, 11 µL) was combined with aqueous DDM (0.5 % w/v, Merck) and incubated for 10 min. An aliquot was transferred to a

tapered autosampler vial, snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

2.3.2.2 PfMATE

An aqueous solution of NaTDB (80 mM, 11 µL) was combined with an equal volume of buffered (130 mM NaCl, 10 % glycerol, 0.025 % DDM, 50 mM Tris pH 7.4) PfMATE (60 µM) and incubated for 10 min. Aliquots were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

2.3.3 Gasdermin D/Caspase-1

hGSDMD (in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP)), hCaspase-1 (C285A) (in 50 mM Tris-HCl pH 8.0, 5 mM DTT, 20 % sucrose), hGSDMD-NT pores in liposomes (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1,3-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol (CL) (1:1, Anatrace)) and VRT-043198 (in DMSO) were provided by Dr Antonio Biasutti (OMass Therapeutics - Oxford, UK).

2.3.3.1 Gasdermin D

Buffered (50 mM Tris-HCl pH 8.0, 5 mM DTT, 20 % sucrose) hCaspase-1 (151 µM, 5.6 µL) was added to buffered (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine hGSDMD (40 µM, 11.2 µL). Caspase-1 buffer (5.6 µL) was added to another buffered solution of hGSDMD (40 µM, 11.2 µL). Both samples were incubated on ice for 2 h. An aqueous solution of NaTDB (100 mM, 5 µL) was combined with hCaspase-1-treated and control hGSDMD samples and incubated for a further 10 min. Aliquots were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

2.3.3.2 Caspase-1

Buffered hGSDMD (151 µM, 5.6 µL) was added to buffered hCaspase-1 (40 µM, 11.2 µL). GSDMD buffer (5.6 µL) was added to another buffered solution of hCaspase-1 (40 µM,

11.2 µL). Both samples were incubated on ice for 2 h. An aqueous solution of NaTDB (100 mM, 5 µL) was combined with hGSDMD-treated and control hCaspase-1 samples and incubated for a further 10 min. Aliquots were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

VRT-043198 (1 mM, 2 µL) in DMSO (10 %) stored under N_2 was added to hCaspase-1 (C285A) (40 µM, 18 µL). DMSO (10 %, 2 µL) was added to another buffered solution of hCaspase-1 (40 µM, 18 µL). Both samples were incubated on ice for 30 min. An aqueous solution of NaTDB (100 mM, 5 µL) was combined ligand-treated and control samples and incubated for a further 10 min. Aliquots (5 µL) were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

2.3.3.3 Gasdermin D N-terminal pore liposomes

GSDMD-NT pores in POPC:CL (1:1) were resuspended with aqueous NaTDB (100 mM, 50 µL) using gentle agitation and incubated for 10 min. Aliquots were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

2.3.4 Beta-1 adrenergic receptor/nanobodies

2.3.4.1 Beta-1 adrenergic receptor

Carazolol and isoprenaline were provided by Dr Parth Kapoor (OMass Therapeutics - Oxford, UK)

Aqueous isoprenaline (1 mM, 2 µL) was added to two buffered (350 mM NaCl, 3 mM Imidazole, 0.05 % DDM, 20 mM Tris pH 8) solutions of β_1AR (18 µM, 15 µL). MilliQ water (2 µL) was added to another buffered solution of β_1AR (18 µM, 15 µL). All three samples were incubated on ice for 5 min. Buffered (20 mM HEPES pH 7.5, 200 mM NaCl, 5 mM Imidazole) Nb80 (173 µM, 2 µL) was added to an isoprenaline-treated β_1AR solution. Nb80 buffer was added to the remaining two β_1AR solutions. Samples were incubated for a further 35 min. An aqueous solution of the NaTDB (100 mM, 5 µL) was combined with ligand-treated and control β_1AR samples and incubated for

10 min. Aliquots (5 µL) were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

Carazolol in DMSO (1 mM, 2 µL) was added to two buffered solutions of β_1AR (18 µM, 15 µL). DMSO (10 %, 2 µL) was added to another buffered solution of β_1AR (18 µM, 15 µL). All three samples were incubated on ice for 5 min. Nb60 (186 µM, 2 µL) was added to a carazolol-treated β_1AR solution. Nb60 buffer was added to the remaining two β_1AR solutions. Samples were incubated for a further 35 min. An aqueous solution of the NaTDB (100 mM, 5 µL) was combined with ligand-treated and control β_1AR samples and incubated for 10 min. Aliquots (5 µL) were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

2.3.4.2 Nanobodies

β_1AR (18 µM, 50 µL) was added to a protein concentrator (100k MWCO, 0.1–0.5 mL, Thermo Fisher Scientific) that had been equilibrated with β_1AR buffer A (Table 2.3). This was centrifuged at 3000 *g* until the sample volume had at least halved.

Aqueous isoprenaline (1 mM, 2 µL) was added to a buffered solution of β_1AR (36 µM, 16 µL). MilliQ water (2 µL) was added to a separate buffered solution of β_1AR (36 µM, 16 µL). β_1AR samples were incubated on ice for 5 min. Both β_1AR samples were separately added to Nb80 (186 µM, 2 µL). MilliQ water (2 µL) and β_1AR buffer (16 µL) were also added to a separate Nb80 sample (186 µM, 2 µL). All three samples were incubated for a further 35 min. An aqueous solution of the NaTDB (100 mM, 5 µL) was combined with ligand-treated and control Nb80 samples and incubated for 10 min. Aliquots (5 µL) were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

Carazolol in DMSO (1 mM, 2 µL) was added to a buffered solution of β_1AR (36 µM, 16 µL). DMSO (10 %, 2 µL) was added to a separate buffered solution of β_1AR (36 µM, 16 µL). β_1AR samples were incubated on ice for 5 min. Both β_1AR samples were separately added to Nb60 (173 µM, 2 µL). DMSO (10 %, 2 µL) and β_1AR buffer (16 µL) were also added to a separate Nb60 sample (173 µM, 2 µL). All three samples were incubated for a further 35 min. An aqueous solution of the NaTDB (100 mM, 5 µL) was combined with ligand-treated and control Nb60 samples and incubated for 10 min.

Aliquots (5 µL) were transferred to tapered autosampler vials (four replicates), snap-frozen and irradiated at 349 nm as described for eIF4A (Section 2.3.1).

2.4 Sample preparation

2.4.1 Native polyacrylamide gel electrophoresis

Protein samples were combined with an equal volume of 4x native-PAGE buffer (Thermo Fisher Scientific) and incubated at room temperature for 10 min. These were loaded onto a 12 % TGX SDS-PAGE gel (BioRad). SDS-PAGE was conducted at 160 V for 50 min (PowerPac Basic, BioRad). The gel was washed with MilliQ water and heated by microwaving on 'high' for 2 min. The water was discarded and the process repeated twice more. The gel was then microwaved on medium with SimplyBlue SafeStain (Thermo Fisher Scientific) for 90 s and then further incubated at room temperature on a rocker for 10 min. The stain was removed and the gel was washed with MilliQ water and heated by microwaving on 'medium' for 2 min. The water was discarded and the process repeated twice more.

2.4.2 Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

Protein samples were combined with 6X SDS-PAGE reducing buffer (375 mM Tris-HCl pH 8.0, 9 % (w/v) SDS, 50 % (v/v) glycerol, 0.03 % (w/v) bromophenol blue, 9 % (v/v) βME, 1 µL) and incubated at room temperature for 10 min. These were loaded onto a 12 % TGX SDS-PAGE gel (BioRad). SDS-PAGE was conducted at 160 V for 50 min (PowerPac Basic, BioRad). The gel was stained as described in Section 2.4.1.

2.4.3 In-gel proteolytic digestion

Protein bands were excised using a scalpel, cut into 1 mm^2 pieces and destained with acetonitrile solution (MeCN, 50 %, 50 µL) for 10 min at room temperature. Gel pieces were dehydrated with MeCN (450 µL) with agitation for 3 min before the MeCN was removed. Gel pieces were treated with DTT solution to reduce protein disulfide bonds (10 mM, Merck, ammonium bicarbonate (AmBic) 100 mM, 50 µL) at 55 °C for 30 min

before being dehydrated with MeCN (450 µL). Gel pieces were then treated with iodoacetamide solution (55 mM, Merck, AmBic 100 mM, 50 µL) and incubated in the dark for 30 min to alkylate cysteine thiol groups before again being dehydrated with MeCN (450 µL). Gel pieces were finally incubated with protease solution (10 ng/µL, AmBic 50 mM, 50 µL) at the relevant temperature for 18 hour (Table 2.4). Formic acid (10 µL) was added to protein digests and kept on ice. Supernatant was removed from the gel pieces and centrifuged at 5000 *g* for 3 min. This was transferred to plastic autosampler vials for nano-LCMS analysis.

Protease	Temperature (°C)
Chymotrypsin	25
Glu-C	37
Pepsin	37
Trypsin	37

Table 2.4 | Protease incubation temperatures

2.4.4 Liquid chromatography-mass spectrometry

Digests were analysed with a Dionex U3000 nano-LC coupled to a ThermoFisher LTQ FT Ultra Mass Spectrometer (Thermo Fisher Scientific) containing a nano-ESI source. An injection volume of 3 µL was loaded onto a C18 Pepmap300 loading column (10 mm, 300 Å, 5 µm particle size, Thermo Fisher Scientific). Sample separation was performed using a C18 Pepmap300 column (150 mm × 75 µm, 300 Å, 5 µm particle size, Thermo Fisher Scientific) with a gradient of two mobile phases: mobile phase A (5 % MeCN, 0.1 % formic acid) and mobile phase B (95 % MeCN, 0.1 % formic acid).

2.4.5 Mass spectrometry

The mass spectrometer was operated in positive ion mode with a standard coated SilicaTip emitter (New Objective) nanospray source for analysis of eIF4A. However due to supply issues with coated nESI tips, uncoated tips (MSWil) were used in combination with an external voltage supply for all subsequent analyses. The inlet capillary of the mass spectrometer was held at 275 °C with a tube lens value of 145 V. For peptide identification in protease optimisation experiments, a DDA scan mode was employed in which the three most intense ions from each survey scan were fragmented in an

LTQ XL LIT and subsequently analysed. Ions were selected within a mass window of 2 Th. 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 s and an exclusion duration of 360 s. For labelled peptides, a full scan mode was employed. For sub-peptide analysis of carbene modified peptides, CID MS/MS was conducted on selected labelled peptides and detergent. This was carried out with a nominal energy of 35.0 and a mass window of 8 Th. The activation time was set at 30 ms and an activation-Q value of 0.250. A wideband activation was also applied to fragmentation of labelled detergent.

2.5 Data Analysis

2.5.1 Peptide identification

Peptide identity was confirmed by CID MS/MS of precursor peptide ions and database searching. RAW files were uploaded to SearchGUI version 4.1.11[105] and searched against a custom protein sequence database including the sequence of the protein of interest using the X!Tandem algorithm[106] and results visualised in PeptideShaker version 2.2.8.[107] Search settings included the relevant protease, a peptide length of 4-40 residues, three missed cleavages, a precursor charge of between +1-5, fragment b and y ions, fixed carbamidomethylation of cysteine residues, a precursor error of 0.05 Da and a fragment ion error of 0.1 Da.

2.5.2 Peptide-level label quantification

Quantification of carbene labelling at the peptide-level was carried out using PepFoot version 1.2.1.[103] RAW files were uploaded onto the software and converted to mz5 filetype. Spectral peaks from unlabelled and labelled EICs (from the same peptide) were manually inspected to ensure sampling of the correct ion, primarily using an increase in retention time of the labelled peptide. A mass shift of +202 Da was searched for, which corresponded to the insertion of one NaTDB molecule. Peptides displaying high modification with a single NaTDB molecule were also examined for multiple carbene insertion events (eg. +404 Da). This was carried out with a mass tolerance of

0.02 mmu, three missed cleavages, a peptide length of 4-40 residues and the relevant protease. Peak areas were integrated and fractional modifications generated from the ratio of labelled peak area to the sum of the labelled and unlabelled peak areas (Equation 2.1). Fractional modifications for each peptide were compared between treatments. Significant peptide-level labelling differences were discerned by way of a 2-tailed student T-test (P<0.05 and P<0.01).

$$F_{mod} = \frac{A_{labelled\ peptide}}{A_{unlabelled\ peptide} + A_{labelled\ peptide}} \tag{2.1}$$

2.5.3 Residue-level label quantification

Quantification of carbene labelling at the residue-level was carried out using MS/MS files. MS/MS spectra were combined over retention times pertaining to a particular precursor ion's EICs. This was again carried out with a precursor mass tolerance of 0.02 mmu. MS/MS spectra were inspected for b and y fragment ions before the ratio of labelled fragment intensities to the sum of labelled and unlabelled fragment intensities of one series were used to calculate residue-level fractional modifications (Equation 2.2). Absolute modifications were determined by calculating the difference in fractional modification between two consecutive fragments (Equation 2.3). In the event that fragment ions could not be detected, the residue was grouped together with the previous residue(s). Per-residue modification were averaged across identified fragments and compared between treatments. In either case, the total sub-peptide labelling was checked to ensure that this equalled the peptide-level modification. Significant sub-peptide level labelling differences were discerned by way of a 2-tailed student T-test (P<0.05 and P<0.01).

$$F_{mod(n_i)} = \frac{I(n_{i_{labelled}})}{I(n_{i_{labelled}}) + I(n_{i_{unlabelled}})} \tag{2.2}$$

$$abs.mod. = P[f.mod(n_i) - f.mod(n_{i-1})] \tag{2.3}$$

2.6 Computational modelling

2.6.1 Homology modelling

Homology models of eIF4A, β_1AR, GSDMD and PfMATE were generated for visualisation of carbene footprinting data.

2.6.1.1 Eukaryotic initiation factor 4A

The closed model of eIF4A was generated by I-TASSER with no specified template.[108,109] The top returned models all matched to closed forms of eIF4AI. The best model (displaying the lowest C-score) was selected and templated onto a mix of PDB structures, including predominantly 2HYI (eIF4AIII) but also 2J0S, 5IVL, 4D25, 4C9B.

2.6.1.2 Gasdermin D

The full length hGSDMD structure was generated by I-TASSER using PDB 6VIE as a template.[108,109] This was aligned to PDB 6KN0 in SwissModel.[110]

2.6.1.3 PfMATE

The outwards-facing model of PfMATE was generated by I-TASSER using PDB 3VVN as a template.[108,109]

2.6.1.4 Beta-1 adrenergic receptor

The fully-activated β_1AR structure was generated by I-TASSER using PDB 6H7J as a template.[108,109] Isoprenaline and Nb80 were aligned to this model with the ChimeraX matchmaker tool using the same PDB template.[111] The fully-inactivated β_1AR structure was generated by I-TASSER using PDB was generated using PDB 5JQH as a template. Carazolol and Nb60 were aligned to this model with the ChimeraX matchmaker tool using the same PDB template.

2.6.1.5 Molecular docking

Molecular docking was carried out using AutoDock Vina version 1.2.0.[112] PDBQT files for hippuristanol and eIF4A were generated using AutoDock Tools version 1.5.6. The grid box was applied over the protein's CTD. Docking was performed with an exhaustiveness of 9, an energy range of 9 and model output of 5.

2.6.2 Membrane and micelle modelling

The hGSDMD-NT pore was modelled in POPC:CL (1:1) membranes using the CharmmGUI Bilayer Builder.[113,114]

A DDM micelle containing 132 detergent monomers was modelled onto the PfMATE structure using the CharmmGUI Micelle Builder.[115]

2.6.3 Molecular visualisation

Protein structures were visualised using ChimeraX version 1.3.[111]

2.6.4 *In-silico* digestions

In-silico digestions were performed using the Peptide Cutter server,[116] utilising three missed cleavages and a peptide length of between 4-40 residues.

3

Mapping the interaction between the eukaryotic initiation factor 4A and the inhibitor hippuristanol using carbene footprinting and mass spectrometry

3.1 Introduction

3.1.1 Eukaryotic initiation factor 4A

The human eukaryotic initiation factor 4A (eIF4A) is a member of the DEAD-box family of helicases involved in displacing bound proteins and unwinding the 5'UTR of mRNA during translation.[117] The eIF4A family is comprised of three related proteins: eIF4AI (DDX2A), eIF4AII (DDX2B) and eIF4AIII (DDX48).[118] eIF4AI and eIF4AII display 90 % sequence similarity; however, only eIF4AI is required for cell viability. eIF4AI is also more abundant in cells and the most well characterised form of the protein.[119] The

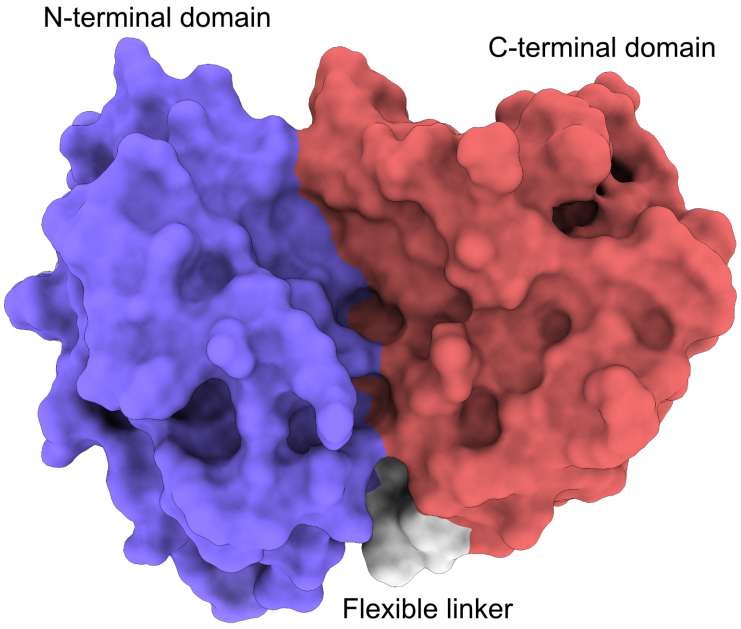
third paralog, eIF4AIII, is involved in exon junction complexes and only shares 67 % sequence similarity to eIF4AI.[120]

eIF4A is a dumbell-shaped protein which features two RecA-like domains joined by an intermediate, flexible linker (Figure 3.1a).[121,122] Conserved motifs line these domains and contribute to mRNA and ATP binding (Figure 3.1b).[123] The protein is conformationally dynamic: in the absence of mRNA and ATP, eIF4A occupies an open structure with no inter-domain contact; however, their binding prompts closure of the protein which promotes contact with conserved residues.[124] The closed state is transiently occupied and ATP hydrolysis is linked to displacement of mRNA-bound proteins and removal of secondary structures in mRNA, as well as release of mRNA and restoration of the protein's open unbound structure.[124] This conformational cycling between open and closed states is key to eIF4A activity.[125]

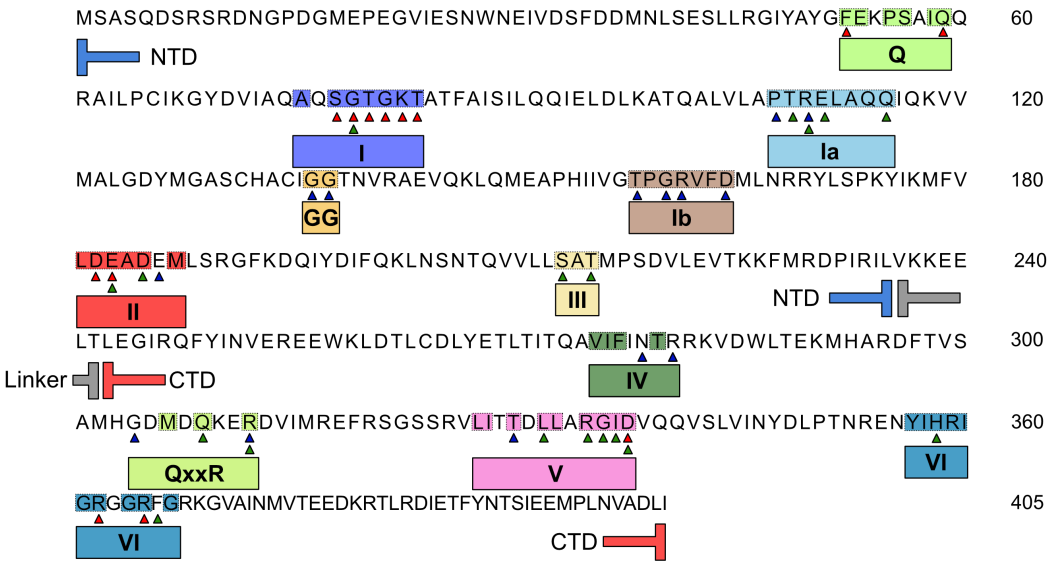
eIF4A makes up the eIF4F complex with two other proteins, eIF4E and eIF4G. eIF4E is a 5′-mRNA cap-binding protein and eIF4G is a scaffolding protein which associates eIF4A and eIF4E subunits.[126] After interaction of the mRNA 5′ cap with eIF4E in eIF4F, eIF4A serves to remove secondary structures from mRNA, further stimulated by eIF4B and eIF4H activity.[127] The ribosome is thought to display some mRNA unwinding activity; however, stable mRNA structures require the assistance of eIF4A.[128] The 43S pre-initiation complex — formed of the 40S ribosomal subunit, eIF3 and a ternary complex containing tRNA methionine-initiator, eIF2 and GTP — associates with the eIF4F complex and scans the 5′ UTR towards the start codon.[129] The 60S ribosomal subunit is then recruited at the start codon and the 80S ribosome is primed for translational elongation.[129]

3.1.2 Hippuristanol

Hippuristanol is a polyoxygenated steroid sourced from the coral *Isis hippuris* and a promising inhibitor of the early stages of protein synthesis.[130] NMR experiments on the CTD of eIF4A have suggested that hippuristanol binds directly to several conserved residues within this region of the protein.[131,132] Mutagenesis studies have further reinforced the importance of these residues in binding.[132] Furthermore, research using single-molecule Förster resonance energy transfer (FRET) experiments, in which



(a) eIF4A domains (Blue = NTD, grey = flexible linker, red = CTD). The closed protein structure is shown.



(b) eIF4A sequence with annotated motifs: Q (red), I (orange), Ia (yellow), GG (light green), Ib (dark green), II (cyan), III (light blue), IV (blue), QxxR (purple) and V (orchid). Arrows indicate interaction with ATP (red), RNA (blue) and inter-domains (green). NTD, linker and CTD are shown. N-Met is also included.

Figure 3.1 | Structure of eIF4A.

Cys residues of three eIF4A mutants were labelled with donor and acceptor dyes Cy3 and Cy5 maleimide derivatives, respectively, suggested that interaction with hippuristanol maintains full-length eIF4A in a closed conformation, consequently preventing mRNA, but not ATP, from binding the protein.[133] The use of hippuristanol to inhibit eIF4A selectively and consequently stall translation initiation is an attractive strategy for targeting human cancers (Figure 3.2) where cells with a malignant phenotype depend on not just a increase in protein synthesis but also an altered translational landscape where certain oncogenic mRNAs are up-regulated. Indeed, hippuristanol's anti-neoplastic activity is well-documented and demonstrates the potential for non-selectively targeting eIF4A/eIF4F in tumour cells.[134–136]

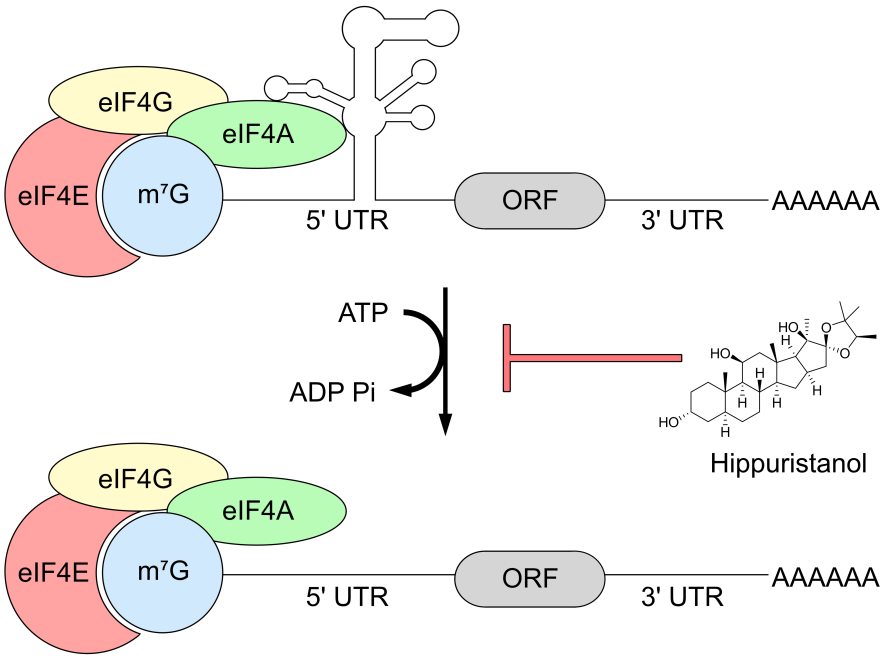


Figure 3.2 | eIF4A-mediated unwinding of 5′ UTR higher-order structures as part of the eIF4F complex and inhibition of this process by hippuristanol.

Here, I apply carbene footprinting of full-length eIF4A to map the interaction site of hippuristanol and examine the effect of binding upon the protein's structure.

3.2 Results and discussion

3.2.1 Carbene labelling of the eukaryotic initiation factor 4A-hippuristanol complex

eIF4A was kindly provided to me by Dr Vasileios Paschalis (University of Nottingham, United Kingdom) and hippuristanol by Dr Jerry Pelletier (McGill University, Canada). Initial optimisation of enzymatic digestion and labelling conditions were previously performed by Amy Hogan from the Oldham Group (University of Nottingham, United Kingdom). From this work, trypsin was identified as the most suitable protease which maximised peptide sequence coverage whilst 20 mM NaTDB provided optimal carbene modification of the helicase.

Differential footprinting of eIF4A was carried out by photo-chemical activation of the NaTDB in the presence and absence of hippuristanol, as described in the Materials and Methods section (Section 2.3.1). Following SDS-PAGE separation (Figure S1) and proteolysis with trypsin, the extent of labelling on peptides was analysed by LC-MS. A total of 32 tryptic peptides corresponding to 67.2 % coverage of eIF4A were identified. Theoretically, trypsin was able to generate 100 % sequence coverage (with peptides displaying a suitable length between 4-40 residues) and the inability to detect these was likely due to them remaining in the gel pieces or on the HPLC, eluting in the void volume, or being of too great an m/z to observe them by MS. Of these 32 peptides, 12 displayed significant differences (Student's t-test, $p < 0.05$) in labelling between control and ligand-treated samples, with six of these peptides showing even greater significance (Student's t-test, $p < 0.01$).

Pleasingly, peptides predicted to contain residues involved in direct hippuristanol binding by NMR[131] displayed significant reductions in F_{mod} by the carbene when hippuristanol was present (Figure 3.3). Peptide 335–353, whilst displaying very low labelling, showed a small but distinct masking event. Interestingly, when mapped onto the closed structure, the peptide occupied much of the cleft between domains (Figure 3.4b). Indeed, this peptide contained part of the V motif, involved in RNA binding, inter-domain contact and ATP binding, as well as direct interaction with hippuristanol (Figure 3.1b).[123] The inherently more sequestered and consequently,

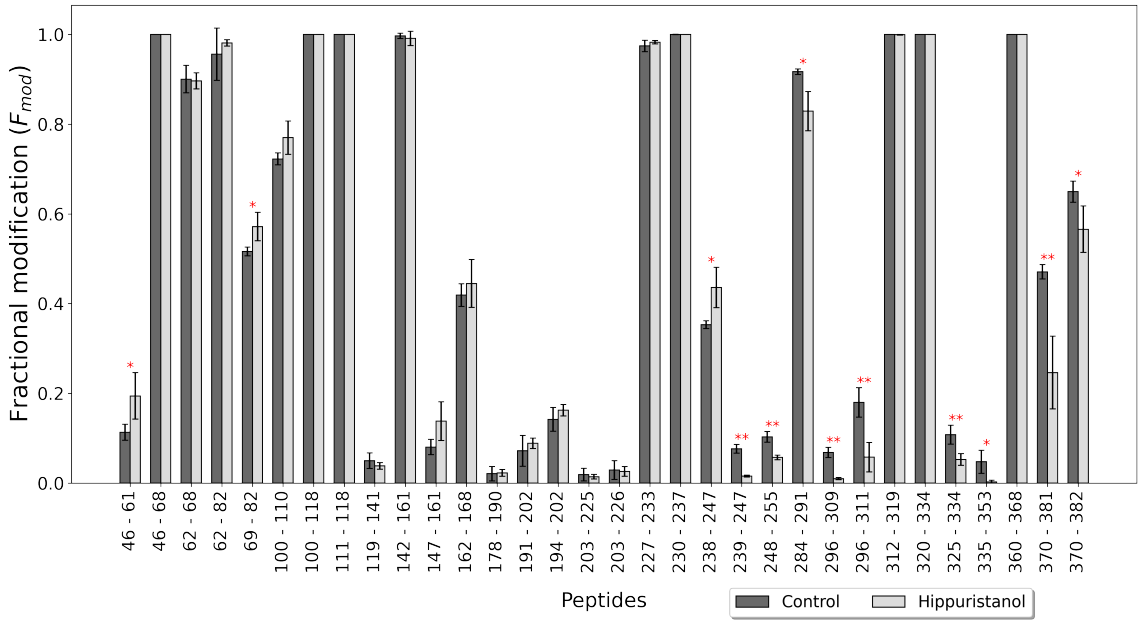
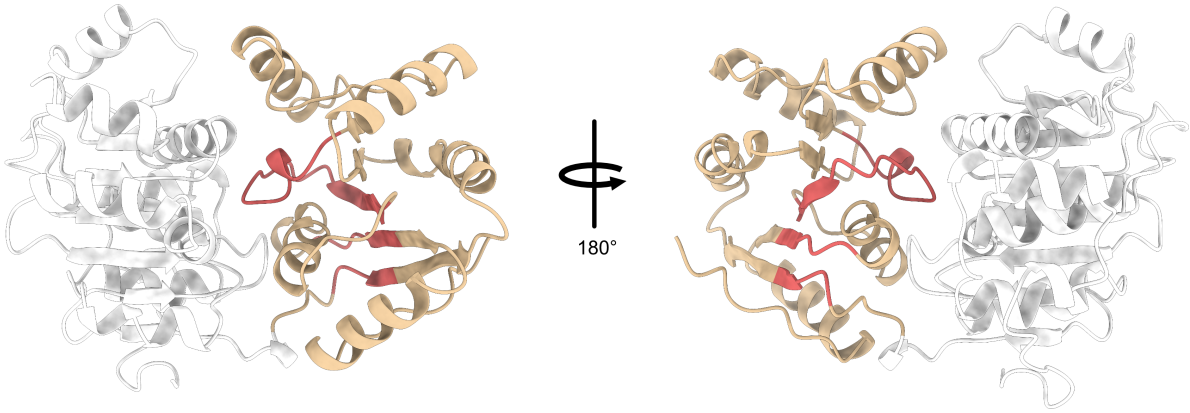


Figure 3.3 | Carbene footprinting of eIF4A with and without hippuristanol using a trypsin digestion. The carbene footprinting histogram is displayed. Fractional modification of each tryptic eIF4A peptide is shown with (light grey) and without (dark grey) 2:1 molar equivalents of hippuristanol. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = $P < 0.01$; * = $P < 0.05$).

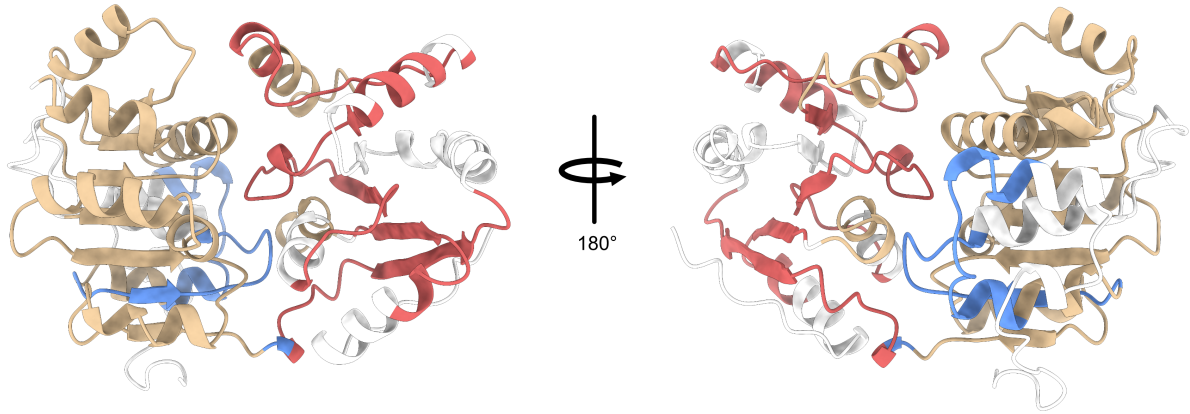
shielded nature of this peptide may explain low levels of labelling (consequently reducing its chemical accessibility) in both samples with hippuristanol possibly contributing to further masking through direct steric inhibition or inducing closure of the protein. Nonetheless, the extent of eIF4A closure due to hippuristanol activity is unknown and whether residues involved in native inter-domain contacts reform these interactions remains undetermined. It does appear logical that disruption of residues involved in RNA interaction and proper interdomain contact would impede nucleic acid binding and helicase functionality. Masking events were also observed at peptide 370–381 and the missed cleavage peptide 370–382, with the former showing higher significance than the latter. These were indicative of hippuristanol binding, and consistent with previous NMR data (Figure 3.4a, Figure 3.4d).[131] Peptides predicted to be within 5 Å of the hippuristanol binding site[131] also displayed changes in fractional modification in ligand treated samples. Interestingly, peptide 238–247 displayed an unmasking event whilst the cleavage variant 239–247 exhibited a masking event. Both peptides formed part of the flexible linker region (Figure 3.1b). This difference in labelling suggested that Lys238 was relatively highly labelled in the ligand-bound state. As such, it seems that binding of hippuristanol induced closure of eIF4A about

the flexible linker leading to an increase in chemical accessibility at Lys238 which ultimately led to an increase in the amount of carbene labelling. These observations highlighted that carbene footprinting could also be used to inform on protein dynamics, similarly to HDX. The N-terminus of eIF4A was not present in the construct used for the NMR study and so it was not possible to compare these effects with previous data.[131] Masking events on peptides 239–247, 248–255, and 325–334 were consistent with NMR data of the C-terminus, which predicted that residues on these peptides were located within 5 Å of hippuristanol. It should be borne in mind that the relatively large size of the carbene probe will result in masking events occurring over a larger area than the direct contact surface of the ligand (Figure 1.20). Interestingly, peptide 325–334 contains part of the V motif, including Thr329 which is implicated in RNA binding. Allosteric disruption of these conserved residues would likely hinder RNA interaction and eIF4A activity. Indeed, Lindqvist and colleagues postulated that hippuristanol interfered with Thr329 alignment, preventing interaction with RNA.[131]

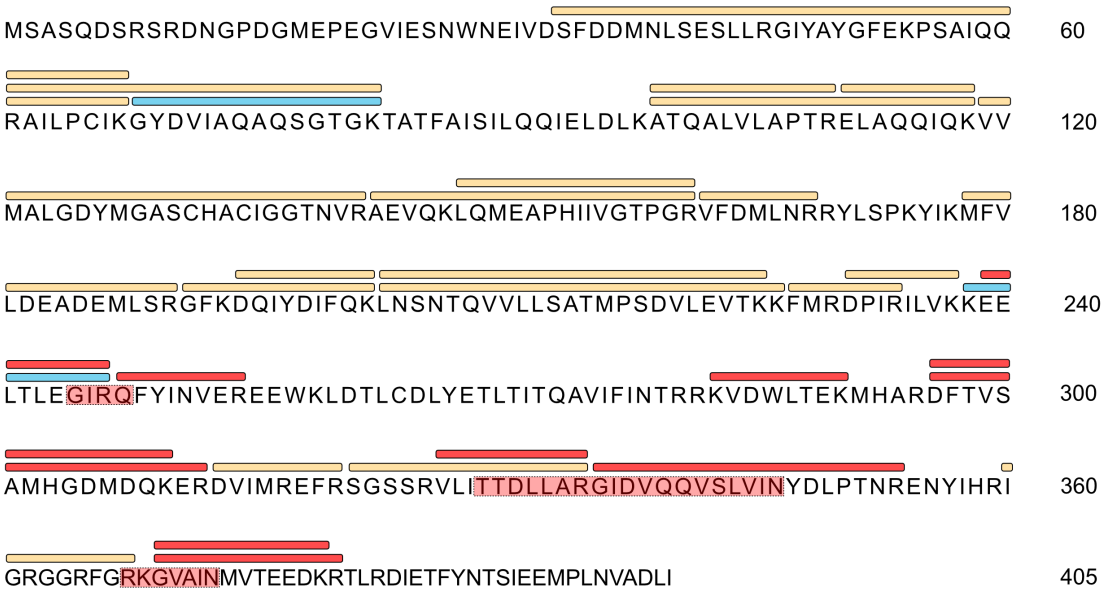
Masking events could not be observed for the previous peptide 320–335, suggesting that residues 320–324 were very highly labelled in the presence and absence of the ligand thereby concealing the masking event in peptide 320–335. Several differences in labelling — both masking and unmasking events — were observed in peptides not seen in NMR data, including peptides 46–61, 69–82, 284–291, 296–309, and 296-311. Unmasking events were observed for both N-terminal peptides 46–61 and 69–82. However, the missed cleavage peptide 46–68 did not display similar labelling events despite sharing residues 46–61. This suggested that the subsequent residues 62–68 were highly labelled and influenced overall fractional modification, consistent with our observations. This was further reinforced by peptide 62–82 displaying 100 % modification, whilst peptide 69–82 displayed an overall reduction in labelling. Interestingly, both 46–61 and 69–82 contained conserved residues involved in ATP binding. Since hippuristanol is not thought to impede ATP binding,[132] masking of implicated residues suggested that allosteric change led to higher probe insertion near or close to these residues. No NMR data were available for comparison here, as these residues were located in the NTD.[131] Peptides 296–309 and 296–311 were also of particular interest, containing residues from the QxxR motif that are thought to be involved in inter-domain contacts and RNA binding.[123] The observation of masking



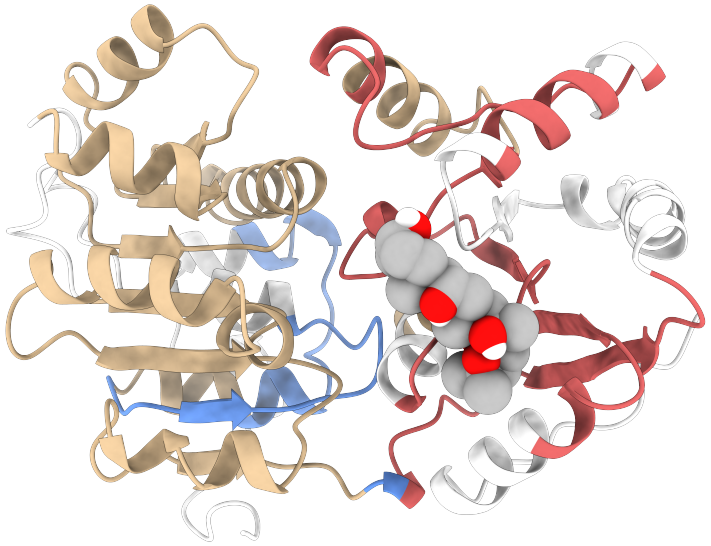
(a) Chemical shift perturbations identified from NMR analysis of eIF4A-CTD in the presence of hippuristanol mapped onto the closed eIF4A structure.[131] The authors suggested that these effects represented hippuristanol binding contacts. Colour scheme is as follows: red = NOES/residues predicted to be within 5 Å of hippuristanol, tan = no involvement in hippuristanol binding, white = no coverage.



(b) Carbene footprinting data mapped onto the closed eIF4A structure. Colour scheme is as follows: red = masking effect, blue = unmasking effect, tan = no change, grey = no labelling, white = no peptide coverage. The regions lacking sequence coverage were likely due to peptides being retained by gel pieces or the RP-HPLC column, or being outside of the scanned MS mass range.



(c) Carbene footprinting data mapped onto the eIF4A sequence. Bars above the sequence represent peptides whilst residues highlighted in red indicate predicted interaction regions/residues located within 5 Å of the inhibitor. Colouring is the same as above.



(d) eIF4A-hippuristanol complex. Labelling data was used to inform docking of hippuristanol to the eIF4A closed eIF4A structure using AutoDock Vina. Colouring is the same as above.

Figure 3.4 | Carbene labelling analysis of eIF4A-hippuristanol.

events at these peptides in ligand-treated samples supported the idea of hippuristanol triggering transition to some form of a closed structure in which RNA binding would be obstructed. Careful review of the NMR data from this region revealed a small but seemingly significant chemical shift perturbation around Asp305,[131] which sits within peptides 296–309 and 296–311.

3.3 Conclusions

In summary, I have shown that differential carbene footprinting of the complete eIF4A construct with and without hippuristanol identifies the binding site to be within the protein's CTD, which is consistent with previous NMR work.[131] Further conformational changes associated with ligand interaction were also revealed for the first time. These were principally located around the flexible linker between the NTD and CTD. Unfortunately, peptide-level labelling results were limited by their spatial resolution and future studies would benefit from targeted MS/MS analysis to determine exact contact interfaces and regions undergoing increased dynamicity. Nevertheless, the results demonstrate the feasibility of using carbene footprinting to understand and characterise protein-ligand interactions.

4

Mapping the binding interactions between gasdermin D and caspase-1 using carbene footprinting

4.1 Introduction

4.1.1 Gasdermin D

Gasdermin D (GSDMD) is a pore-forming protein and key initiator of pyroptosis, an inflammatory form of lytic cell death that occurs in response to diverse pathogenic and sterile insults.[137–139] GSDMD is a 484-residue (53 kDa) protein, encoded for by the GSDMD gene and a member of the gasdermin family of proteins. GSDMD is formed of two conserved domains, an NTD with an extended β-sheet core structure and a CTD that is further characterised by a linker region (linking N and C termini), a helix repeat-I bundle, a helix repeat-II bundle and an intermediated β-strand insertion.[140–142] In the resting state, the CTD interacts with the NTD and stabilises it, autoinhibiting the protein's pore-forming activity. However, upon appropriate host defense signals formation of the inflammasome leads to activation of inflammatory Caspases (namely Caspase-1) that cleave GSDMD's flexible linker. The NTD of GSDMD oligomerises at

the plasma membrane to form a 33-subunit transmembrane pore, approximately 320 Å in diameter, rupturing the membrane and inducing cell death (Figure 4.1).[143,144]

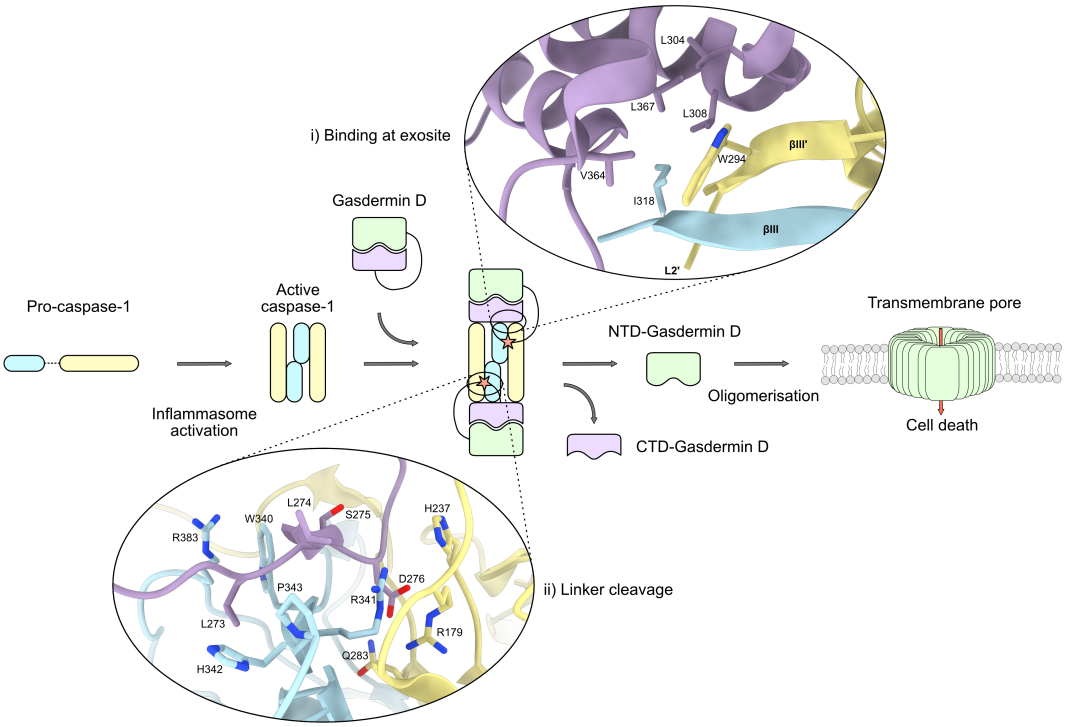


Figure 4.1 | Activation of NTD-Gasdermin D by inflammatory Caspase-1 and subsequent oligomerisation and transmembrane-pore formation in the lipid bilayer (human GSDMD/human Caspase-1 exosite-mediated binding adapted from PDB 6KN0, mouse GSDMD/human Caspase-1 linker binding to catalytic domain adapted from PDB 6VIE). GSDMD colored purple, Caspase-1 p20 subunit colored yellow, Caspase-1 p10 subunit colored light blue. Residues participating in these interfaces are shown.

4.1.2 Caspase-1

Caspase-1 is a cysteine protease formed from a heterodimer of p10 and p20 subunits.[145] The active enzyme is a dimer of heterodimers where each of the two catalytic domains span p10/p20 interfaces. These are comprised of the triad His237, Gly238 (this serves to anchor the P1 site to the catalytic site through backbone hydrogen bonding, forming the oxyanion hole) and nucleophilic Cys285. Caspase-1 recognises the tetrapeptide motif XXXD and induces cleavage after the aspartate residue.[146] The mechanism behind Caspase substrate specificity is poorly understood although it is recognised that Caspase-1 prefers hydrophobic/aromatic residues at position P4 (according to the Schecter-Berger nomenclature, which describes the enzyme and substrate sites

for proteases[63]) and small aliphatic residues at position P1'.[147] For mGSDMD, this region includes 272-SLLSDGIDE-280, where P4 is Leu273 and P1' is Gly277.[147] Whilst the XXXD motif defines a portion of hGSDMD that must interact with Caspase-1, it does not reveal details of other regions of the protein important for enzyme-substrate interaction.

Pyroptosis is recognised as a contributor to many human diseases, including cancer and inflammatory disorders. Inhibition of hGSDMD activation is therefore an attractive therapeutic strategy.[148] Belnacasan (VX-765) is a pro-drug and Caspase inhibitor,[149] which forms the active drug VRT-043198 (O-desethyl-belnacasan) upon esterase cleavage of VX-765 (Figure 4.2). A potent electrophile, VRT-043198, can modify the catalytic Cys285 thiol, thus impeding Caspase activity and preventing GSDMD activation. Administration of VX-765 to mice showed decreased lipopolysaccharide (LPS)-induced cytokine secretion and reduced inflammatory disease severity.[149] Phase IIa clinical trials of VX-765 were discontinued due to liver toxicity; however, development of structurally similar Caspase-targeting drugs continues.[150]

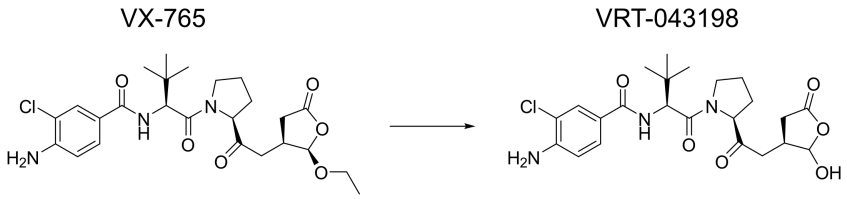


Figure 4.2 | Activation of VX-765. Esterase cleavage of VX-765 yields the electrophile VRT-043198.

Structural elucidation of the full-length human GSDMD-human Caspase-1 complex (hGSDMD/hCaspase-1) has proved difficult owing to the structural heterogeneity and flexibility of GSDMD. Recently, Wang and colleagues determined a structure of the complex between the CTD of hGSDMD and hCaspase-1.[151] This showed 2:2 binding stoichiometry of GSDMD:Caspase-1. It also highlighted the importance of the Caspase-1 βIII/βIII' sheet in mediating complex formation through its insertion into a hydrophobic groove on the CTD-GSDMD. Leu304, Leu308, Val364 and Val367 on hGSDMD were shown to make hydrophobic contacts with Trp294 and Ile318 on Caspase-1 and importantly, GSDMD-Caspase-1 binding was shown to be independent of the tetrapeptide cleavage site.[151] Unfortunately, the flexible linker

bearing this motif was not seen in the structure. Liu and colleagues showed that mouse GSDMD/human Caspase-1 (mGSDMD/hCaspase-1) interaction was mediated through the same hydrophobic exosite contacts.[147] This structure did include a truncated form of mGSDMD's linker loop and revealed Asp276, of the tetrapeptide, LLSD, buried into the Caspase active site. The authors also reported the P4 site residue $_{Gas}$Leu273 contacting $_{Casp}$Arg383, Trp340 and His342. His342 was also hydrogen bonded to the P5 site residue $_{Gas}$Ser272. At the P1' site, a main-chain hydrogen bond between $_{Gas}$Gly277 and $_{Casp}$His237 further anchored the linker to the catalytic groove whilst providing the conformational flexibility for the loop to exit the domain.[147] Despite these two structures providing complementary information into the molecular mechanisms behind GSDMD binding and activation, mouse and human GSDMD constructs only share 60 % sequence identity. As such, it is not known whether hGSDMD/hCaspase-1 make the same linker contacts to the catalytic domain of Caspase-1. Therefore, given the difficulty in generating crystal structures of full-length hGSDMD, I sought to further characterise the binding interactions between it and hCaspase-1.

Here I report the use of carbene footprinting to study three key aspects of the hGSDMD system. First, I provide an accurate map of the interactions between full-length hGSDMD and hCaspase-1 (C285A), including those at the active site as well as the exosite of hCaspase-1 (C285A). Second, I show changes in carbene labelling within the N-terminus of hGSDMD upon cleavage by hCaspase-1 (C285A) and associated pore formation in liposomes. Third, I detect and map non-covalent binding of the Caspase-1 inhibitor VRT-043198 to the active site of hCaspase-1 (C285A), providing evidence for the potential of these compounds as non-covalent inhibitors of hCaspase-1.

4.2 Results and discussion

4.2.1 Optimisation of sequence coverage and labelling

Full-length hGSDMD and hCaspase-1 were kindly provided to me by Dr Antonio Biasutti (OMass Therapeutics, United Kingdom). First I sought to optimise GSDMD and Caspase-1 (C285A) digestion and carbene labelling conditions. The catalytically inactive mutant was employed throughout this study to prevent enzymatic turnover of

GSDMD. Sequence coverage optimisation was performed to maximise the number of detectable peptides by LCMS. Optimisation of labelling conditions ensured peptide-level modification was at an appropriate level to report on differential binding partner masking effects.

4.2.1.1 Sequence coverage

In-silico digestion was performed on the sequences of both proteins. From these findings, trypsin and Glu-C were predicted to produce the best coverage for GSDMD (Figure 4.3), whilst trypsin, Glu-C and chymotrypsin generated the most useful peptides for Caspase-1 (Figure 4.4).

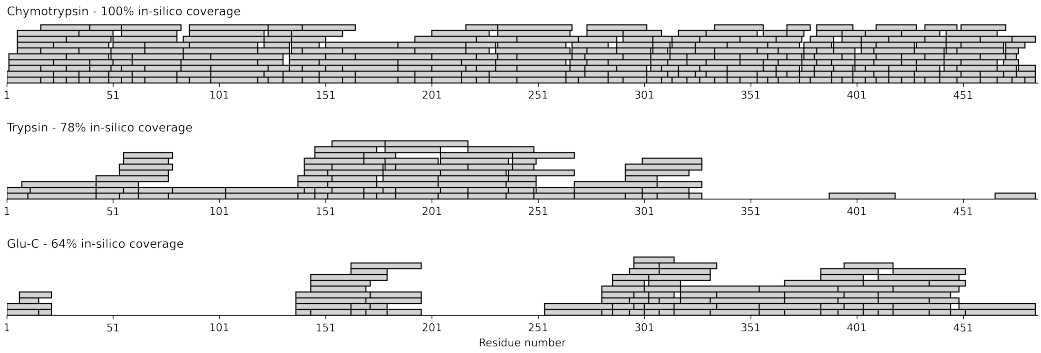
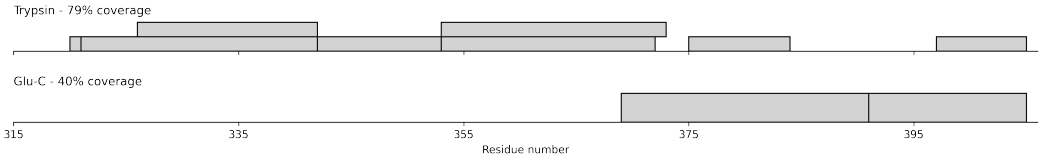
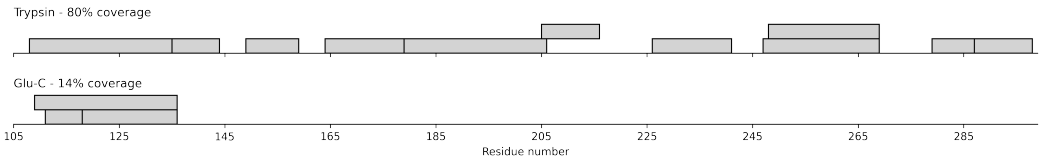


Figure 4.3 | *In-silico* digest of GSDMD with chymotrypsin, trypsin and Glu-C. Peptides were filtered between 4-40 residues with a maximum of three missed cleavages. Grey bars represent predicted peptides and overlapping bars indicate missed cleavages.



(a) *In-silico* digest of Caspase-1 p10 subunit with trypsin and Glu-C.



(b) *In-silico* digest of Caspase-1 p20 subunit with trypsin and Glu-C.

Figure 4.4 | *In-silico* digest of Caspase-1 with trypsin and Glu-C. Peptides were filtered between 4-40 residues with a maximum of three missed cleavages. Grey bars represent predicted peptides and overlapping bars indicate missed cleavages.

Tryptic digestion of GSDMD generated 26 peptides corresponding to 61 % sequence coverage, with much of the C-terminus remaining unrepresented (Figure 4.5). This region was recognised to contain a high Glu content and showed good theoretical coverage by Glu-C in *in silico* digestion. When digested with Glu-C, GSDMD revealed 29 peptides and a further 25 % gain in sequence coverage over the C-terminus (Figure 4.5). This multi-protease approach (where peptides from separate digests were pooled) was highly beneficial to maximising sequence coverage but, as with many techniques employed in protein structural study, it was not possible to achieve complete mapping of the protein sequence, which did place a limit on the information obtained. Tryptic digestion of the Caspase-1 p20 subunit generated 8 peptides corresponding to 79 % sequence coverage whilst proteolysis of the p10 subunit gave 80 % peptide coverage (Figure 4.6b). Digestion of Caspase-1 with Glu-C and chymotrypsin yielded no improvement in sequence coverage and so trypsin digestion only was employed for Caspase-1 (Figure 4.6a).

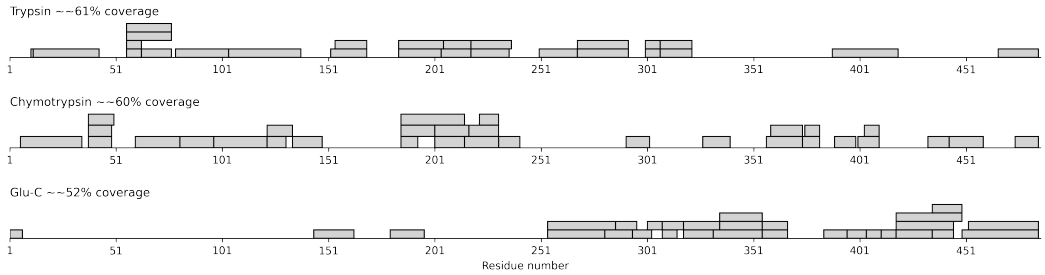
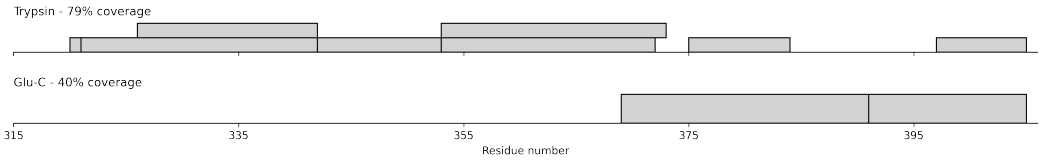


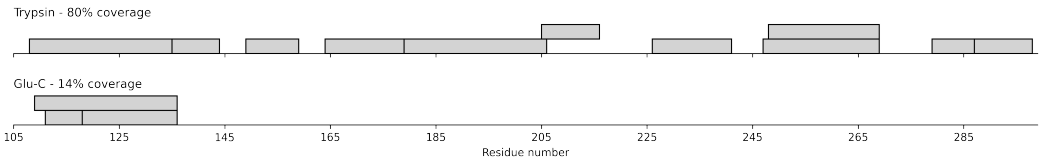
Figure 4.5 | Sequence coverage achieved by DDA LC-MS/MS analysis of hGSDMD following overnight digestion with trypsin, chymotrypsin and Glu-C. Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

4.2.1.2 Carbene labelling

Carbene labelling using NaTDB is typically carried out at low millimolar concentrations (10–50 mM) with most soluble proteins requiring 10–20 mM of the diazirine. Both GSDMD and Caspase-1 showed satisfactory levels of carbene modification at 20 mM NaTDB in which fewest peptides displayed no labelling (Figure 4.7, Figure 4.8). Caspase-1 displayed a distinctive labelling footprint. Regions of no labelling were largely mapped to the dimer interface, which may suggest a dimer. Previous work has shown that isolated Caspase-1 is essentially monomeric at low concentration,[152] and so it would be surprising if the dimer was present in the absence of GSDMD. Native



(a) Proteolytic digest of Caspase-1 p10 subunit with trypsin and Glu-C.



(b) Proteolytic digest of Caspase-1 p20 subunit with trypsin and Glu-C.

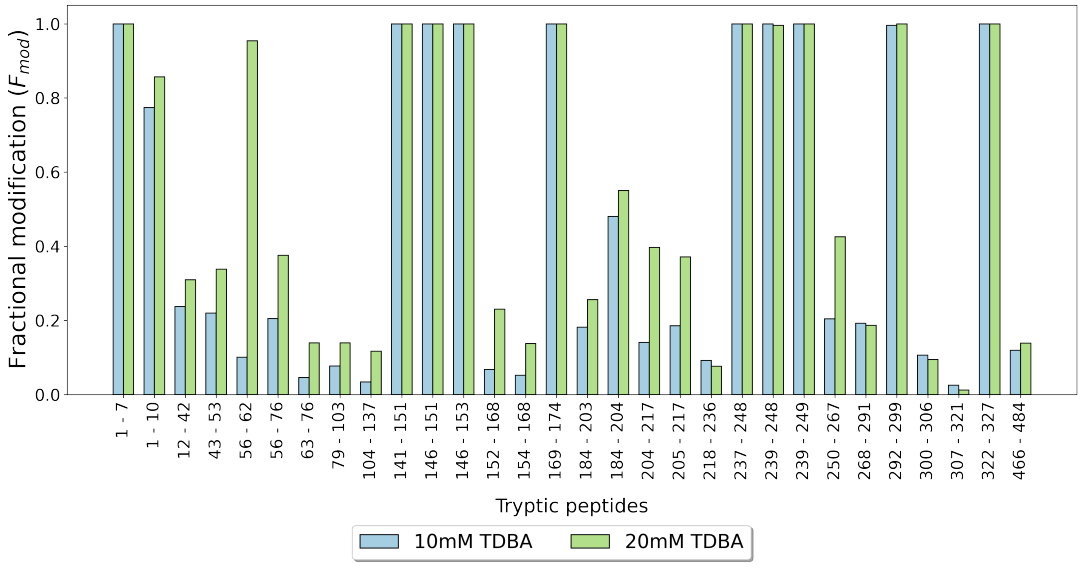
Figure 4.6 | Sequence coverage achieved by DDA LC-MS/MS analysis of Caspase-1 p10 and p20 subunits following overnight digestion with trypsin and Glu-C. Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

gel electrophoresis of Caspase-1 gave a single band of low mobility (Figure 4.9), but this may be due to the relatively high pI of hCaspase-1, rather than a multimeric form.

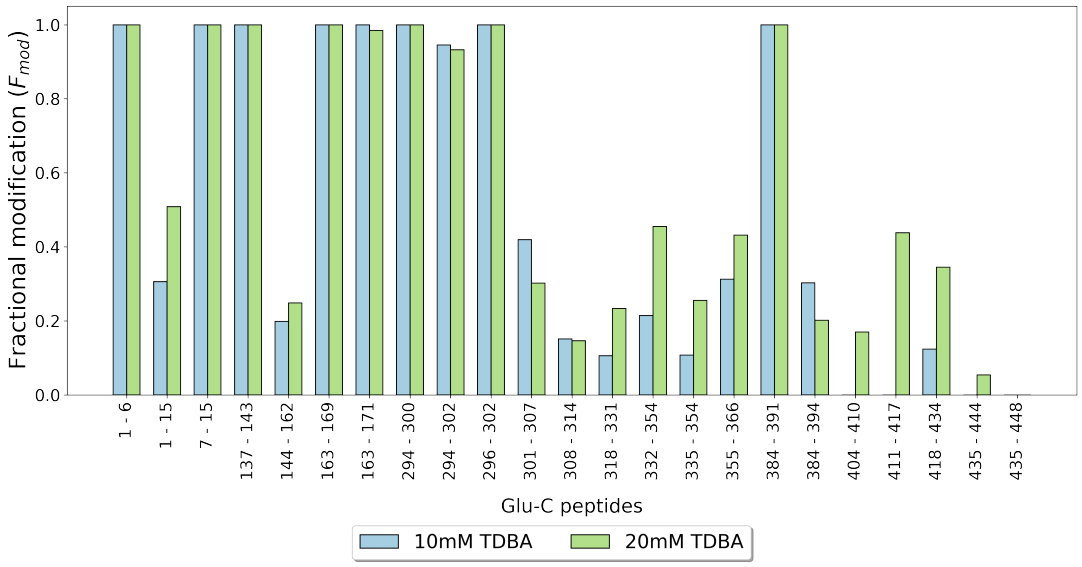
4.2.2 Carbene labelling of the gasdermin D-caspase-1 complex

Following carbene footprinting of GSDMD and Caspase-1 in isolation, labelling was carried out in the presence and absence of two-fold excess binding partner, as described in the Methods section (Section 2.3.3). Native gel electrophoresis revealed that a two-fold excess was required to ensure full complex formation in each case (Figure 4.9). Additionally, the presence of 20 mM diazirine showed no disruption to the complex, consistent with previous findings by us that NaTDB does not perturb protein-protein or protein-ligand interactions.[99]

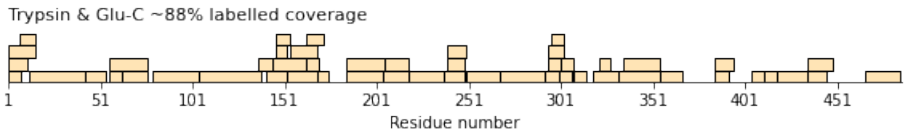
Differentially footprinted GSDMD (Figure S2) (±Caspase-1, two-fold excess) revealed a labelling reduction on tryptic peptide 300-306, only, in the presence of Caspase-1 (Figure 4.10a). A similar comparison of Glu-C-derived GSDMD peptides showed that four exhibited a significant decrease in labelling due to Caspase-1 masking: namely peptides 301-307, 332-354, 335-354 and 355-366 (Figure 4.11a). The fact that peptides 300-306 and 301-307, derived from trypsin and Glu-C digestion, respectively, both showed significant reductions in labelling in the presence of Caspase-1 provides mutually supportive data for masking of this region of GSDMD due to Caspase-1



(a) Carbene labelling optimisation of hGSDMD with a trypsin digestion. The fractional modification of each tryptic hGSDMD peptide is shown at 10 and 20 mM NaTDB.

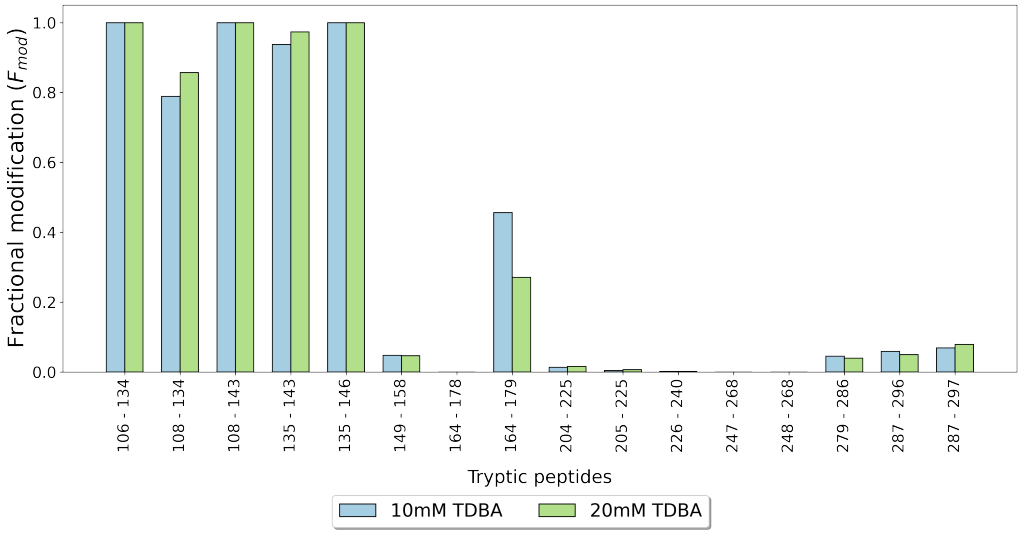


(b) Carbene labelling optimisation of hGSDMD with a Glu-C digestion. The fractional modification of each Glu-C hGSDMD peptide is shown at 10 and 20 mM NaTDB.

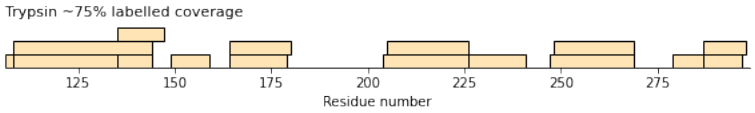


(c) Sequence coverage plot of carbene labelled hGSDMD tryptic and Glu-C peptides. Oranges bars represent observed peptides and overlapping bars represent missed cleavages.

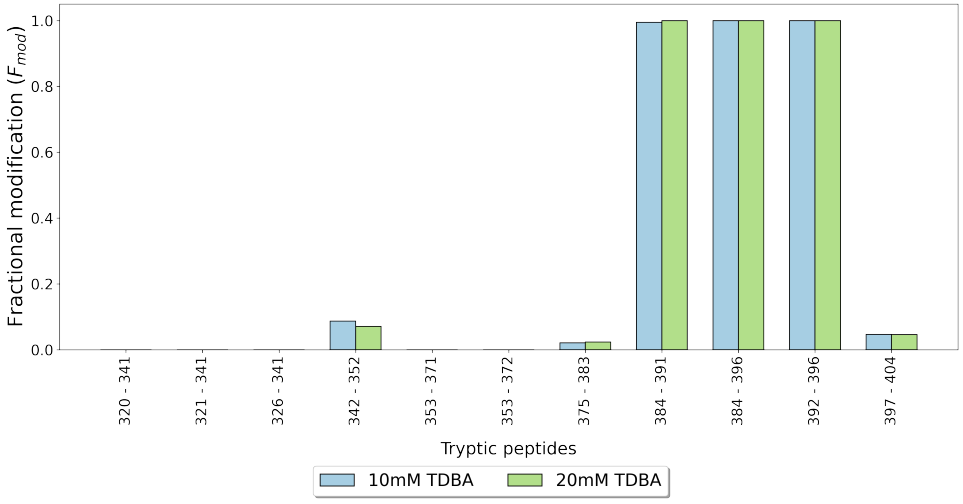
Figure 4.7 | Carbene labelling optimisation of hGSDMD.



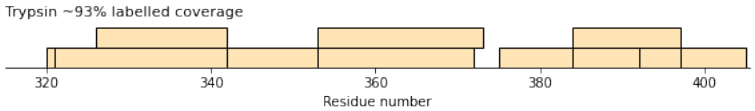
(a) Carbene labelling optimisation of hCaspase-1 p20 subunit with a trypsin digestion. The fractional modification of each tryptic p20 peptide is shown at 10 and 20 mM NaTDB.



(b) Sequence coverage plot of labelled hCaspase-1 p20 tryptic peptides. Oranges bars represent observed peptides and overlapping bars represent missed cleavages.



(c) Carbene labelling optimisation of hCaspase-1 p10 subunit with a trypsin digestion. The fractional modification of each tryptic p10 peptide is shown at 10 and 20 mM NaTDB.



(d) Sequence coverage plot of labelled hCaspase-1 p10 tryptic peptides. Oranges bars represent observed peptides and overlapping bars represent missed cleavages.

Figure 4.8 | Carbene labelling optimisation of hCaspase-1.

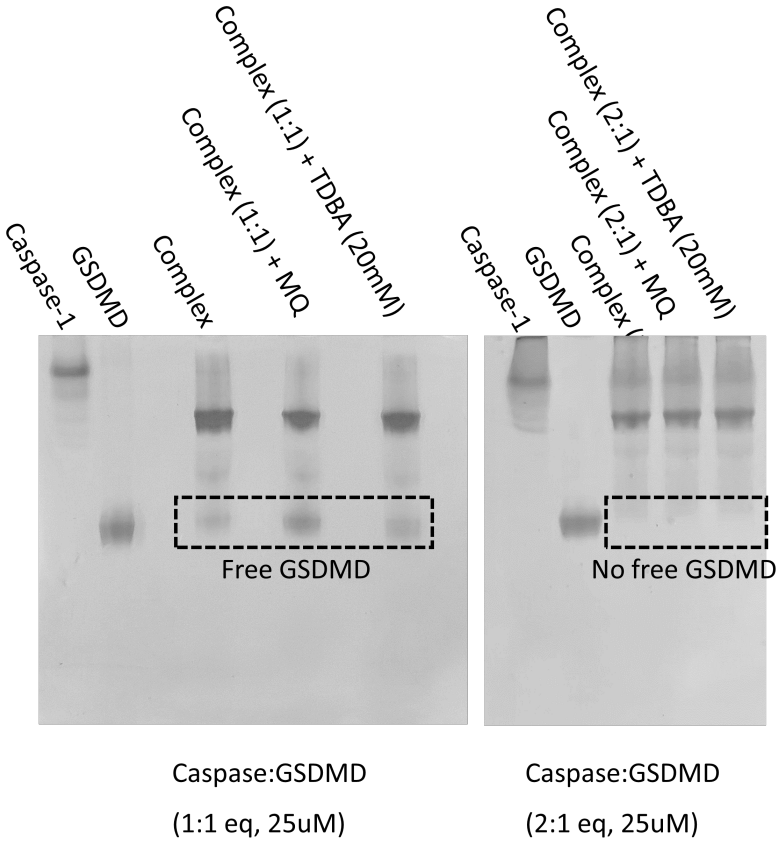
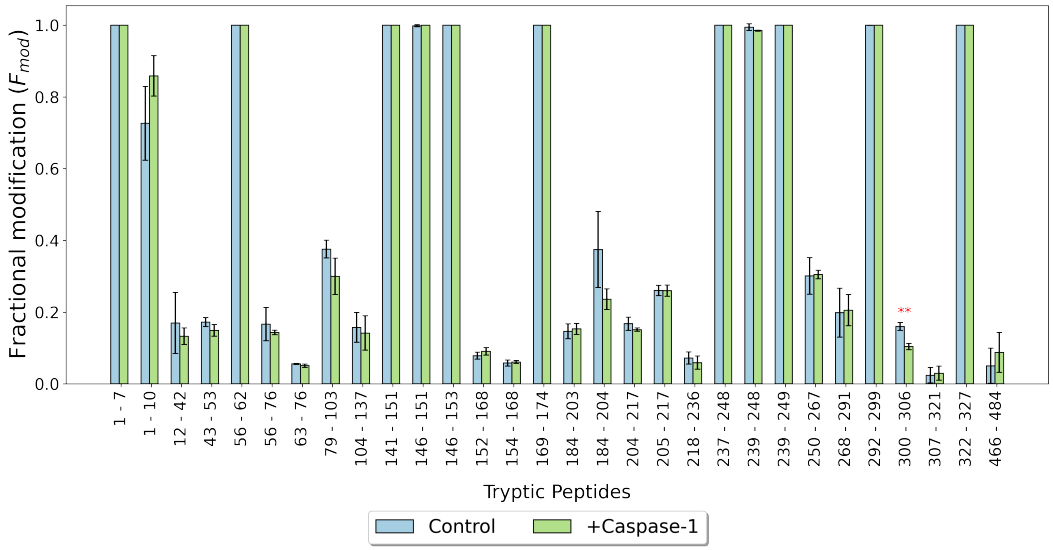


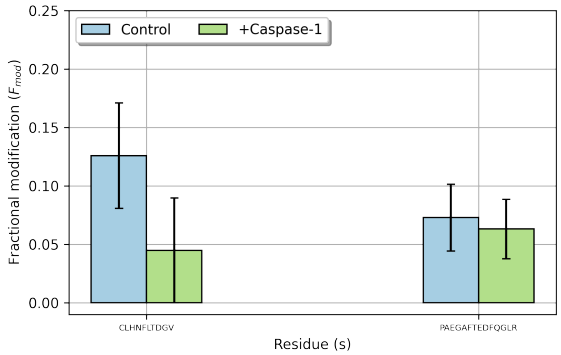
Figure 4.9 | Native-PAGE of the hCaspase-1(C285A)-hGSDMD complex. This was conducted at 1:1 and 2:1 equimolar ratios and revealed that no free hGSDMD is present with two equivalents of Caspase-1 added.

binding. Moreover, labelling reduction in contiguous Glu-C peptides 332-354 and 355-366 shows that the binding interaction also extends over these regions of GSDMD.

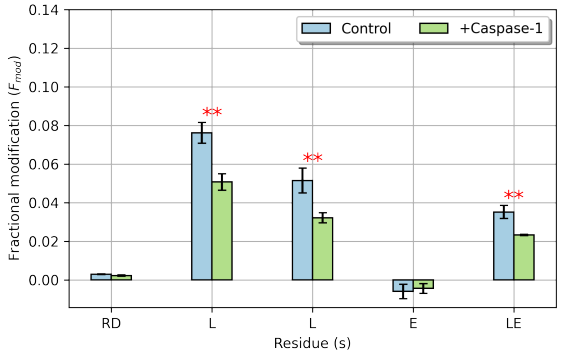
GSDMD residues Leu304, Leu308, Val364 and Leu367 are documented to form hydrophobic contacts with Caspase-1 βIII/βIII' (Figure 4.1), confirming that the observed labelling reduction at tryptic GSDMD peptide 300-306 was due to masking effects associated with the GSDMD-Caspase-1 exosite interaction.[151] CID MS/MS of the labelled GSDMD peptide ELELLDR revealed distinct masking effects at Glu300-Leu301, Leu303 and Leu304, reinforcing the suspected binding interactions around this region with amino acid residue-level resolution (Figure 4.10c). Congruently, the Glu-C GSDMD peptide 301-307 also displayed a masking event, corroborating our findings from the tryptic digest, further suggesting that Leu304 was masked by the interaction with Caspase-1 at the exosite. Indeed, MS/MS analysis showed a single specific masking event at this residue (Figure 4.11b). No labelling differences were observed on GSDMD peptide 308-321, however interrogation of the crystal structure showed that this peptide (excluding Leu308) was located away from the exosite, suggesting that any labelling differences would be largely hidden by neighboring residues with high chemical accessibility and therefore difficult to detect.[151] Further indication of exosite interaction was demonstrated by masked Glu-C GSDMD peptides. Peptide 355-366 contained Val364, which was known to make contacts with Caspase-1, and it appeared logical that the labelling reduction observed on this peptide reflected binding interactions involving the residue. MS/MS of labelled CLVLSSGMLVPE revealed that residues 355-361 (CLVLSSG) were labelled and remaining 361-366 (MLVPE) were unlabelled, hence locating masking effects to the N-terminal half of the peptide (Figure 4.11d). Interrogation of the CTD-GSDMD/Caspase-1 crystal structure revealed that GSDMD peptide 355-361 was extremely proximal to Caspase-1 βIII/βIII' and despite these residues not directly contacting the heterodimer, masking was presumably due to blocking by bound Caspase-1.[151] GSDMD peptides 332-354 or 335-354 did not contain any residues that were known to form direct contacts with Caspase-1. MS/MS showed that residues 332-342 were unlabelled whilst carbene modification was found on residues 343-354, forming a contiguous region of masking with the following peptide, 355-361 (Figure 4.11c). It should be borne in mind that the size of NaTDB (approximately 8.5 Å in length) may result in masking effects over a greater



(a) Carbene footprinting histogram of hGSDMD alone (control), and with hCaspase-1(C285A) with a trypsin digestion. The fractional modification of each tryptic hGSDMD peptide is shown with (green) and without (blue) 2:1 molar equivalents of hCaspase-1(C285A). Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = $P < 0.01$; * = $P < 0.05$).

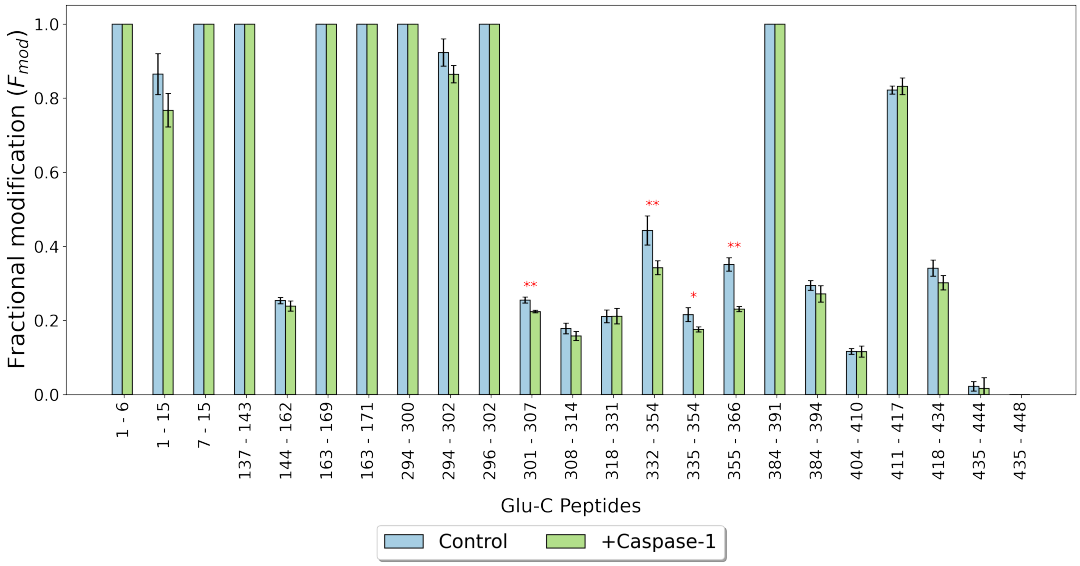


(b) Sub-peptide analysis of CLHNFLTDGVPAEGAFTEDFQGLR (peptide 268-291).

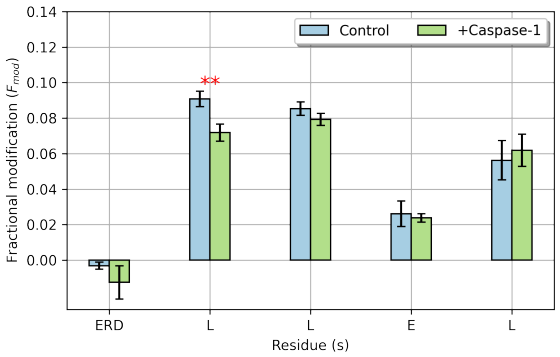


(c) Sub-peptide analysis of ELELLDR (peptide 300-306).

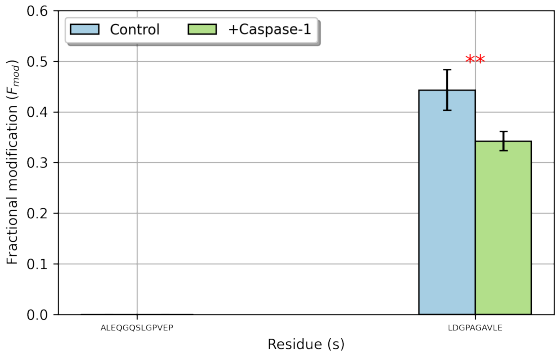
Figure 4.10 | Carbene footprinting of hGSDMD with and without hCaspase-1 (C285A) using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and hCaspase-1 (C285A)-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.



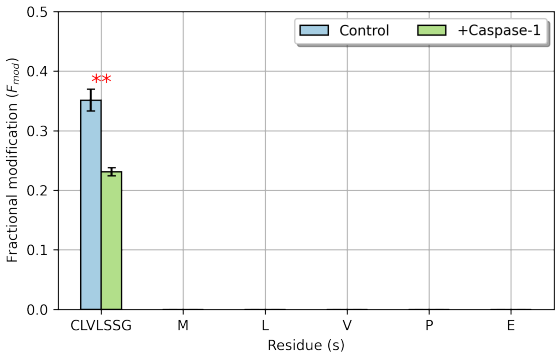
(a) Carbene footprinting histogram of hGSDMD alone (control), and with hCaspase-1(C285A) with a Glu-C digestion. The fractional modification of each Glu-C hGSDMD peptide is shown with (green) and without (blue) 2:1 molar equivalents of hCaspase-1(C285A). Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



(b) Sub-peptide analysis of LELLDRE (peptide 301-307).



(c) Sub-peptide analysis of ALEQGQSLGPVEPLDGPAGAVLE (peptide 332-354).



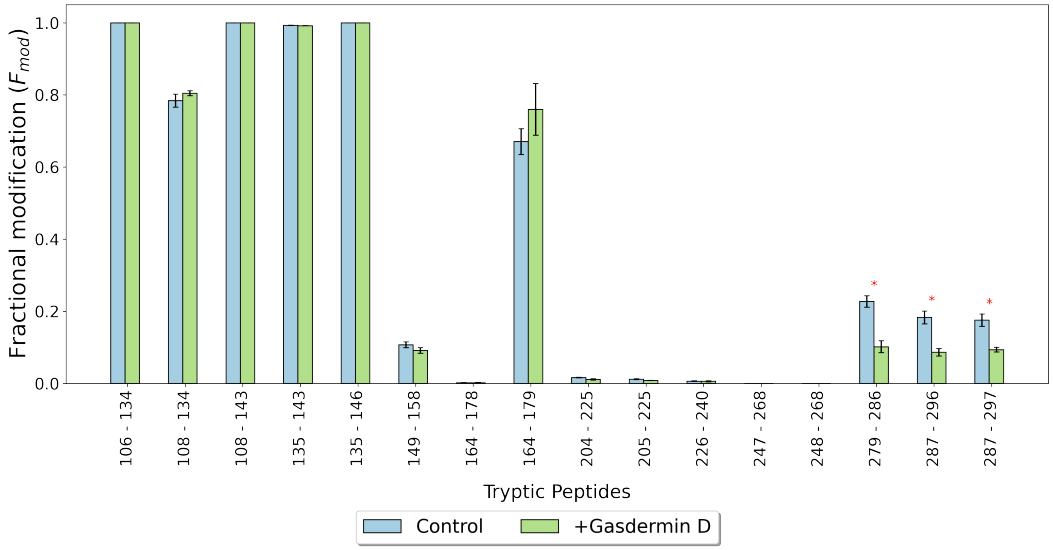
(d) Sub-peptide analysis of CLVLSSGMLVPE (peptide 355-366).

Figure 4.11 | Carbene footprinting of hGSDMD with and without hCaspase-1 (C285A) using a Glu-C digestion. Full scan data highlighted Glu-C peptide-level differences in carbene modification between control and hCaspase-1 (C285A)-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

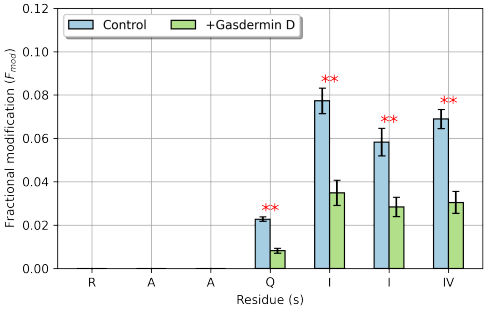
area than the exact contact surface (e.g. Figure 1.20), which may reflect the observed results. However, in keeping with our observations at peptide 355-361, it appeared the masking effects on 343-354 represented similar steric protection caused by Caspase-1 βIII' proximity, again reiterating exosite interaction between GSDMD and the cysteine protease (Figure 4.1).

On the Caspase-1 p20 subunit, masking events were observed at peptides 279-286, 287-296 and 287-297 (Figure 4.12a, Figure S3). Two significant labelling reductions were also seen on the p10 subunit of hCaspase-1, at peptides 342-352 and 375-383 (Figure 4.13a). Masking on the p20 subunit mapped to βIII and the L2 loop, whilst labelling reduction on p20 peptides 287-296 and 287-297 provided further evidence for GSDMD/Caspase-1 binding at the exosite region, in accordance with GSDMD footprinting data (Figure 4.10, Figure 4.11, Figure 4.14).

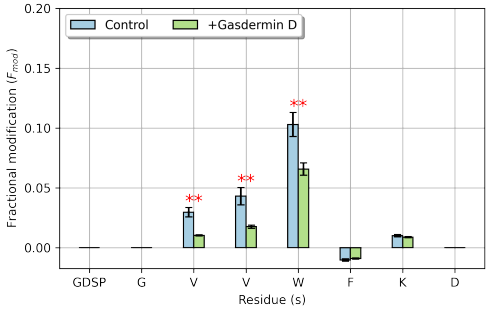
MS/MS analysis of p20 peptide 287-297 revealed residue-level labelling, with specific reductions at Trp294, Val293 and Val292 (Figure 4.12c). Given the role that Trp294 plays in binding to the hydrophobic groove of GSDMD,[151] labelling reductions were attributed to formation of these contacts, highlighting the power that carbene footprinting and specifically, MS/MS-based approaches, have in identifying high resolution interaction sites (Figure 4.14). Due to low sequence coverage and lack of labelling on the NTD of the Caspase-1 p10 subunit, I was not able to detect the binding of βIII' to the hydrophobic groove of GSDMD.



(a) Carbene footprinting histogram of the hCaspase-1(C285A) p20 subunit alone (control), and with hGSDMD with a trypsin digestion. The fractional modification of each tryptic hCaspase-1 p20 subunit peptide is shown with (green) and without (blue) 2:1 molar equivalents of hGSDMD. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).

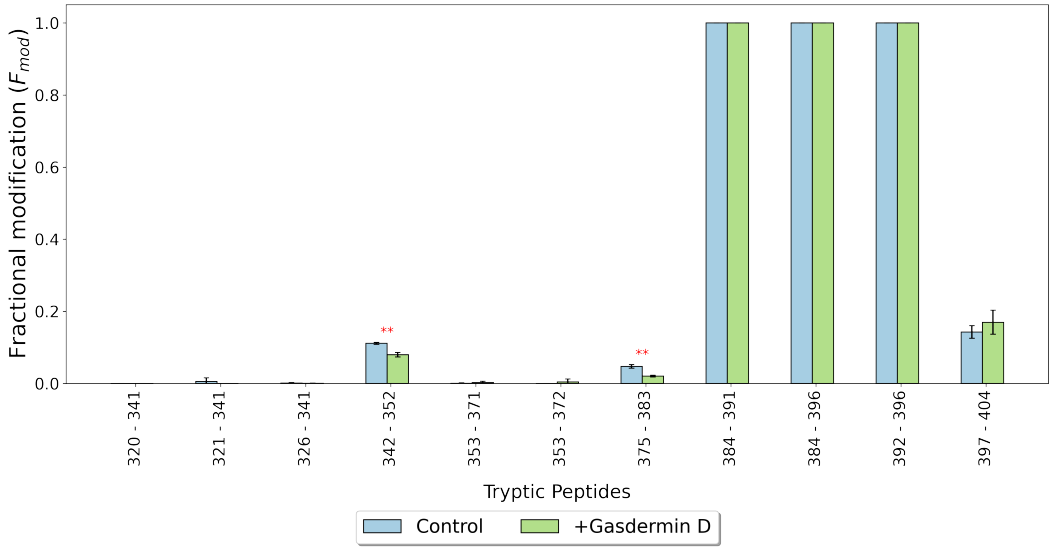


(b) Sub-peptide analysis of VIIIQAAR (peptide 279-286).

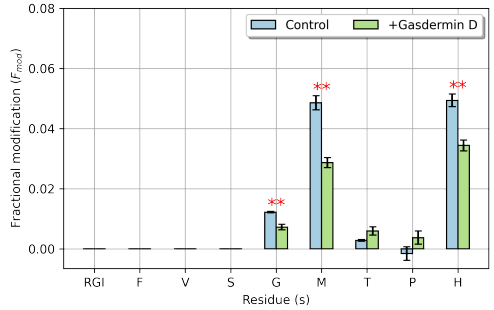


(c) Sub-peptide analysis of GDSPGVVWFKD (pepide 287-297).

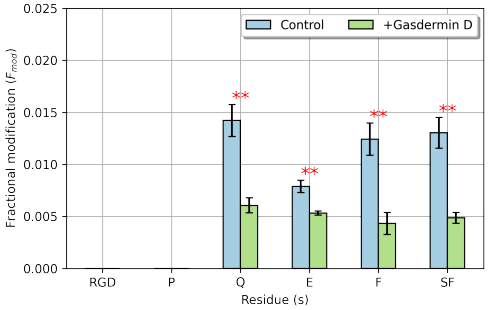
Figure 4.12 | Carbene footprinting of the hCaspase-1 (C285A) p20 subunit with and without hGSDMD using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and hGSDMD-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.



(a) Carbene footprinting histogram of the hCaspase-1(C285A) p10 subunit alone (control), and with hGSDMD with a trypsin digestion. The fractional modification of each tryptic hCaspase-1 p10 subunit peptide is shown with (green) and without (blue) 2:1 molar equivalents of hGSDMD. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



(b) Sub-peptide analysis of HPTMGSVFIGR (peptide 342-352).



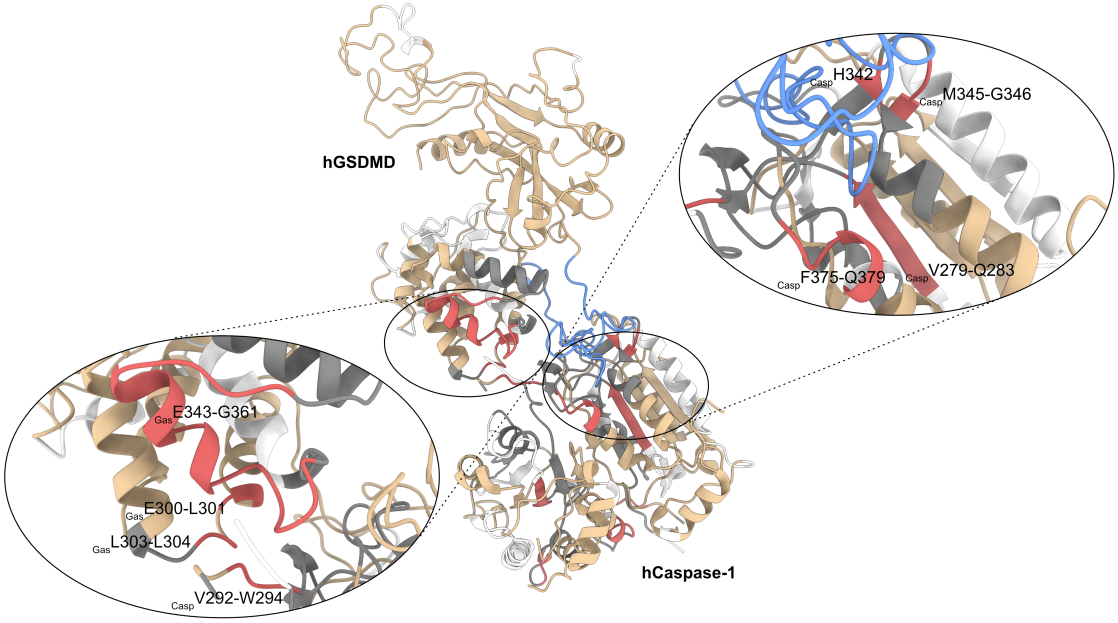
(c) Sub-peptide analysis of FSFEQPDGR (peptide 375-383).

Figure 4.13 | Carbene footprinting of the hCaspase-1 (C285A) p10 subunit with and without hGSDMD using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and hGSDMD-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

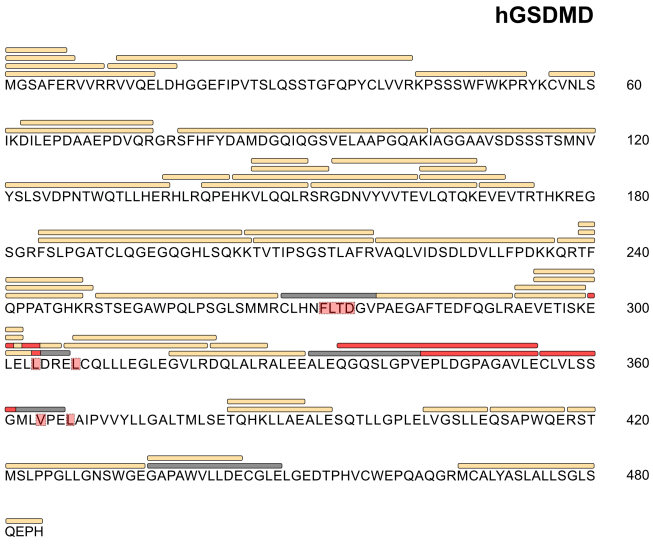
Attention was next turned to the catalytic domain of Caspase-1 and, specifically, whether its interaction with the GSDMD linker region (containing the cleavage site) could be detected by carbene footprinting. Whilst these interactions remain unknown, it was anticipated that hGSDMD/hCaspase-1 would make analogous binding contacts to those of the mGSDMD/hCaspase-1 structure.[147] Unfortunately, masking effects were not observed on the GSDMD linker peptide 268-291 (Figure 4.10a). This peptide contained the tetrapeptide FLTD and so was anticipated to contact the Caspase-1 active site. MS/MS of the labelled GSDMD peptide 268-291 showed that carbene modification was located between regions 268-277 and 278-291 (Figure 4.10b). The former region contained the (FLT)D cleavage site, and exhibited a three-fold reduction in labelling in the presence of Caspase-1, but this difference was not statistically significant perhaps due to structural flexibility of the loop. Returning to Caspase-1 p20, peptide 279-286 constituted much of the Caspase-1 catalytic site and the observed masking event at this peptide evidenced its interaction with the linker region. Indeed, further MS/MS analysis revealed carbene modification to residues 279-283 but not residues 284-286 (Figure 4.12c), meaning the masking observed was confined to the five N-terminal residues. Since the C285A Caspase-1 construct was employed in this study to prevent catalytic turnover of GSDMD, interaction of $_{Casp}$Ala285 with the linker peptide was not anticipated. However, $_{Casp}$Gln283 is known to form side-chain hydrogen bonds to $_{Gas}$Asp276 and the observed masking events on Caspase-1 peptide VIIIQ supported the notion of GSDMD linker binding to the catalytic domain of Caspase-1. Therefore, despite being unable to detect masking on the hGSDMD linker directly, due to a lack of labelling, differential study of both proteins allowed characterisation of active site-based interactions on Caspase-1 (Figure 4.14).

For the Caspase-1 p10 subunit, significant reductions in labelling were seen at peptides 342-352 and 375-383. Much of the region around peptide 342-352 is known to contact the mGSDMD linker.[147] For example, $_{Casp}$Arg341, Trp340 and Pro343, are in proximity of $_{Gas}$Leu274 and Ser275 whilst $_{Gas}$Leu273 and Ser272 are also reported to make Van der Waal and hydrophobic contacts with $_{Casp}$Arg383, Trp340 and His342.[147] The masking effects observed on Caspase-1 peptide 342-352 support linker binding on and around this region of Caspase-1 (Figure 4.13a). Indeed, MS/MS showed residue-level masking events at His342, Met345 and Gly346, with very low/no labelling

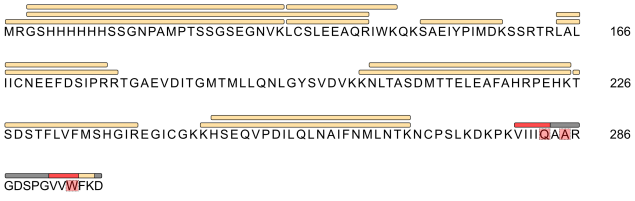
observed at Pro343 (Figure 4.13b). Therefore, given the known role that this region, and indeed, $_{Casp}$His342 play in mGSDMD linker binding, I was able to again report binding to the hCaspase-1 (C285A) catalytic domain (Figure 4.14). This was also reinforced by peptide-level masking on Caspase-1 peptide 375-383, given Arg383's role in contacting mGSDMD. MS/MS showed carbene modification to residues 375-379, with labelling differences occurring on all labelled residues (Figure 4.13c). Whilst labelling did not occur on Arg383, masking events on nearby residues again reflected proximity to the hGSDMD linker.



(a) Combined carbene footprinting data from separate differential studies mapped onto the full-length hGSDMD-hCaspase-1 (C285A) structure. Colour scheme is as follows: red = masking effect, tan = no change, grey = no labelling, white = no peptide coverage. The flexible hGSDMD linker loop is coloured blue for clarity.



(b) Carbene footprinting data mapped onto the hGSDMD sequence. Bars above the sequence represent peptides whilst residues highlighted in red indicate predicted interaction regions. Colouring is the same as above.



(c) Carbene footprinting data mapped onto the hCaspase-1 p20 subunit sequence. Colouring is the same as above.

(d) Carbene footprinting data mapped onto the hCaspase-1 p10 subunit sequence. Colouring is the same as above.

Figure 4.14 | Carbene labelling analysis of hGSDMD-hCaspase-1 (C285A).

4.2.3 Carbene labelling of gasdermin-D N-terminal pore liposomes

hGSDMD N-terminal pores in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1,3-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol (CL) liposomes were kindly provided to me by Dr Antonio Biasutti (OMass Therapeutics, United Kingdom). These were labelled to identify changes associated with protein oligomerisation and lipid-binding (Figure S4). The relative change in fractional modification was compared between full-length monomer hGSDMD and the oligomerised NTD pore version (Figure 4.15). Most tryptic peptides displayed a reduction in carbene labelling compared to full-length GSDMD; however, several regions showed no change in fractional modification between the two species. Labelling reductions were mapped on the GSDMD-NTD pore subunits which revealed extensive masking effects on membrane-spanning β-sheets and oligomerisation contact surfaces (Figure 4.15, Figure 4.16). Several relative masking events were observed on seemingly exposed regions however these constituted long tryptic peptides which also partly contacted interfacial sites or the lipid bilayer. This masking likely reflected the combined steric effects of pore formation and lipid-insertion, compared to labelling of the more accessible soluble monomer. Further to this, I was unable to separate the effects that pore formation and lipid-insertion had on relative change to carbene modification since many peptides were in contact with both the bilayer and neighbouring GSDMD-NTD subunits, and long tryptic peptides were more likely to be simultaneously involved in several of these processes. Indeed, interpretation of the observed labelling effects were only made possible by the deposited X-ray pore structure (PDB 6VFE) and it appears that future labelling studies into such systems would benefit from alternative proteases that generate shorter peptides and/or MS/MS approaches to aid resolution. Nonetheless, peptides on the solvent-accessible lip portion of GSDMD-NT primarily displayed no change in carbene modification compared to the full-length monomer, suggesting no

change in chemical accessibility of these regions. Taken together, the observed changes were consistent with monomer self-assembly and pore formation and highlight the versatility of using carbene footprinting to identify gross structural changes of lipid-embedded protein assemblies, which are challenging targets for structural study.

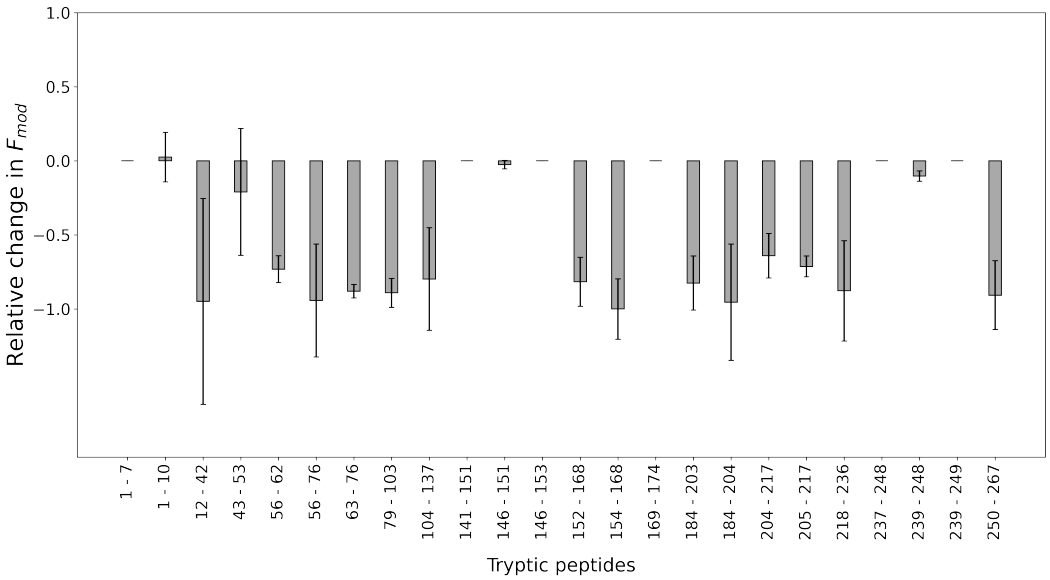
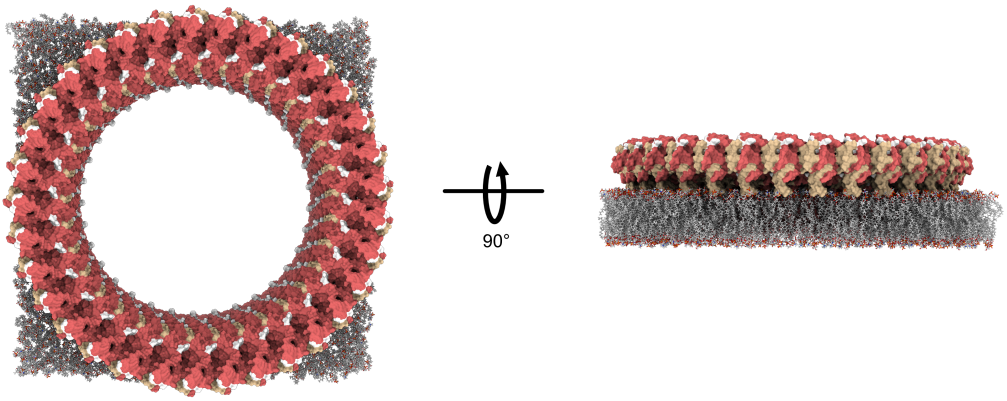


Figure 4.15 | Relative change in fractional modification on tryptic peptides between the hGSDMD-NT pore and full-length monomer.

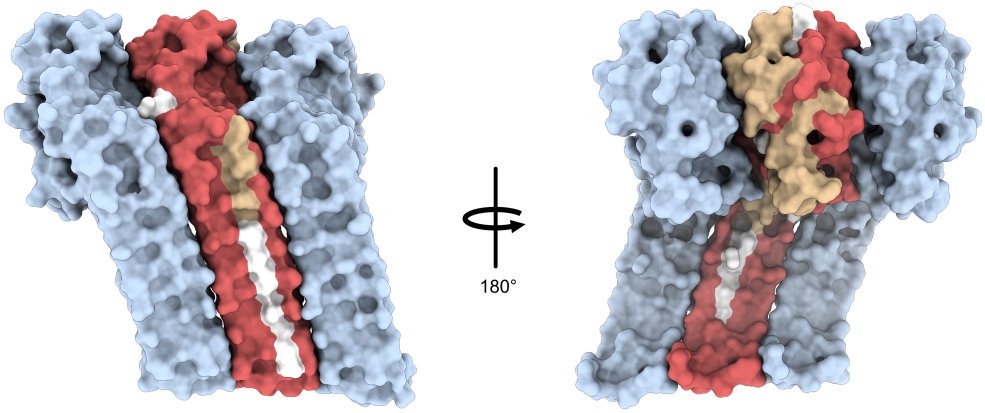
4.2.4 Carbene labelling of the caspase-1-VRT-043198 complex

Carbene labelling was next applied to Caspase-1 in the presence and absence of VRT-043198 (the active metabolite of VX-765), as described in the Materials and Methods section (Section 2.3.3.2, Figure S5). This compound normally reacts with the active site Cys285 thiol group of Caspase-1.[149] The C285A mutant employed in this study was therefore unable to undergo covalent modification. Shape complementarity has been suggested to mediate important binding contacts between VRT-043198 and Caspase-1 which increase the selectivity and overall efficacy of interaction. As such, I sought to determine whether VRT-043198 would bind non-covalently to the protease in the absence of Cys285.

The addition of DMSO (1 % v/v) to Caspase-1 solution was necessary to solubilise VRT-043198. For the differential study, an equivalent amount of DMSO was added to the control and VRT-043198 treated protein. The addition of DMSO slightly reduced overall levels of labelling compared to without it. The addition of VRT-043198 induced

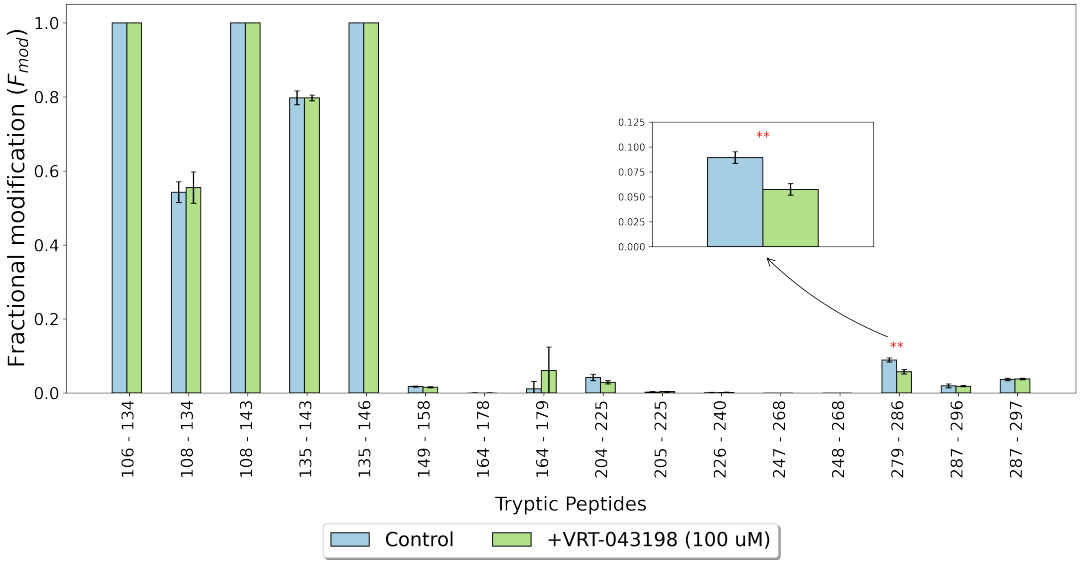


(a) Carbene footprinting data mapped onto the hGSDMD-NT (33mer) pore (PDB 6VFE) liposomes. Colour scheme is as follows: red = masking effect, tan = no change, grey = no labelling, white = no peptide coverage. Colouring is according to relative change in fractional modification compared to hGSDMD full-length labelling. The pore liposome structure was generated using the CHARMM-GUI.

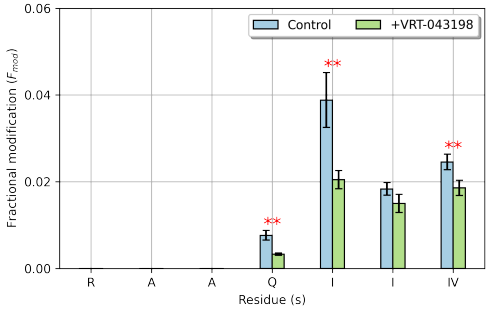


(b) Carbene footprinting data mapped onto the hGSDMD-NT trimer substructure (derived from PDB 6VFE). Colour scheme is the same as above.

Figure 4.16 | Carbene labelling analysis of the hGSDMD-NT pore.

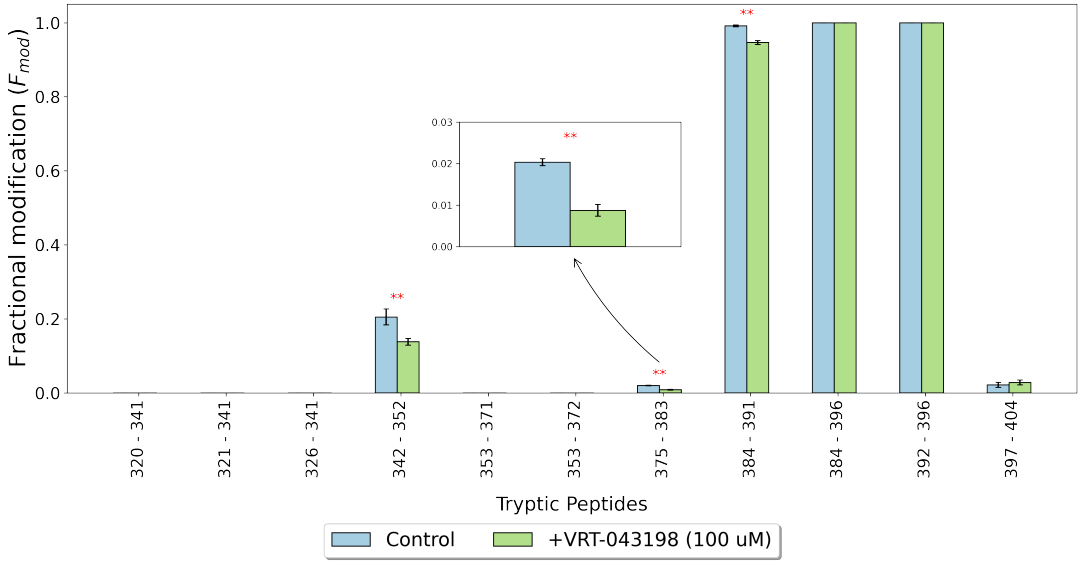


(a) Carbene footprinting histogram of the hCaspase-1(C285A) p20 subunit alone (control), and with VRT-043198 with a trypsin digestion. The fractional modification of each tryptic hCaspase-1 p20 subunit peptide is shown with (green) and without (blue) 100 µM of VRT-043198. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).

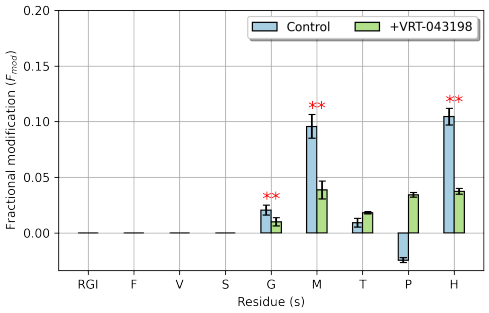


(b) Sub-peptide analysis of VIIIQAAR (peptide 279-286).

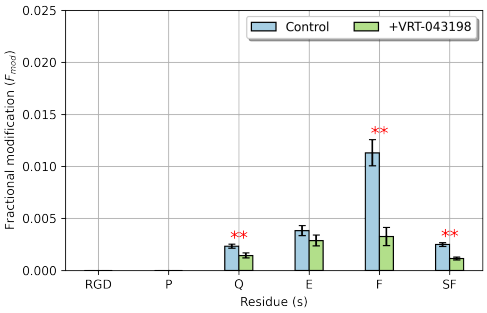
Figure 4.17 | Carbene footprinting of the hCaspase-1 (C285A) p20 subunit with and without VRT-043198 using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and VRT-043198-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.



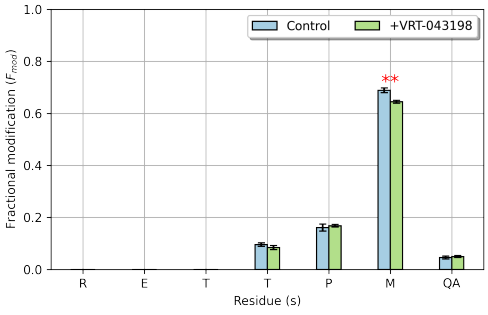
(a) Carbene footprinting histogram of the hCaspase-1(C285A) p10 subunit alone (control), and with VRT-043198 with a trypsin digestion. The fractional modification of each tryptic hCaspase-1 p10 subunit peptide is shown with (green) and without (blue) 100 µM of VRT-043198. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



(b) Sub-peptide analysis of HPTMGSVFIGR (peptide 342-352).



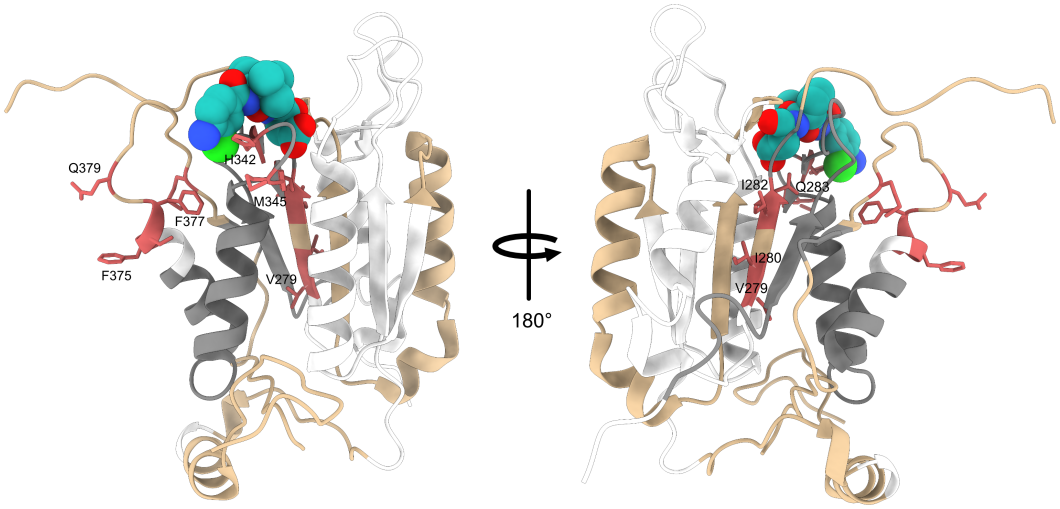
(c) Sub-peptide analysis of FSFEQPDGR (peptide 375-383).



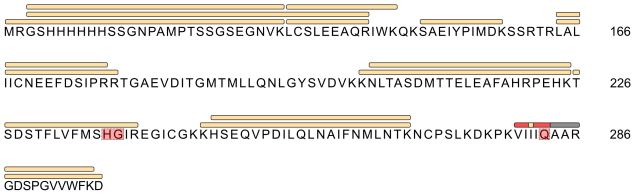
(d) Sub-peptide analysis of AQMPTTER (peptide 384-391).

Figure 4.18 | Carbene footprinting of the hCaspase-1 (C285A) p10 subunit with and without VRT-043198 using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and VRT-043198-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

a single labelling reduction on the p20 subunit at peptide 279-286 (Figure 4.17a) and three significant labelling reductions on the p10 subunit, including at peptides 342-352, 375-383 and 384-391 (Figure 4.18a). Pleasingly, when mapped to the Caspase-1/VRT-043198 crystal structure (PDB 6PZP), these masking events localised to the ligand binding site, suggesting that despite the absence of the catalytic residue, the metabolite was binding to Caspase-1 (C285A) in a similar region to that of wildtype hCaspase-1.[153] MS/MS was again utilised to provide higher resolution labelling information on masked peptides. CID fragmentation showed consistent fragmentation patterns, and similar residue-level labelling, to those observed in the GSDMD/Caspase-1 labelling study (Section 4.2.2). The active site peptide 279-286 on p20 displayed significant residue-level labelling effects at residues Val279-Ile280, Ile282 and Gln283 (Figure 4.17b). Labelling reductions at Gln283 were attributed to hydrogen bond formation with the carboxylic acid of VRT-043198 whilst differences on neighboring residues were likely caused by proximity of the ligand and consequential steric hindrance to the carbene. On the p10 subunit, Val338 and Trp340 were predicted to form pi-alkyl interactions with the pyrrolidine ring whilst Val348 and Pro343 were suggested to interact with the tert-butyl moiety. Pi-pi interactions were also predicted between His342 and the aromatic ring of the chloroaniline group, as well as hydrogen bond formation between Arg341 and the benzamide moiety and Ser339 and a secondary amide moiety. Ser347 was also anticipated to form van der Waals interactions with the metabolite. The reduction in carbene labelling on peptide 342-352 in the presence of VRT-043198 likely reflected these interactions, with MS/MS revealing significant labelling events on His342, Met345 and Gly346 (Figure 4.18a, Figure 4.18b). Peptide-level labelling reductions at 375-383 and 384-391 (Figure 4.18a) presumably reflected Arg383's pi-cation interaction with VRT-043198's chloroaniline moiety but also by the overall proximity of these peptides to the compound, causing masking from the photoactive probe. MS/MS displayed significant labelling events on Phe375-Ser383, Phe376 and Glu378 and Met386 (Figure 4.18c, Figure 4.18d). These findings demonstrate a specific non-covalent interaction between VRT-043198 and hCaspase-1 (C285A) and highlight the ability of carbene footprinting methodology to rapidly discern the binding of non-covalent inhibitors in the hCaspase-1 active site (Figure 4.19).



(a) Carbene footprinting data mapped onto the hCaspase-1 (C285A)-VRT-043198 structure. Colour scheme is as follows: red = masking effect, tan = no change, grey = no labelling, white = no peptide coverage.



(b) Carbene footprinting data mapped onto the hCaspase-1 p20 subunit sequence. Bars above the sequence represent peptides whilst highlighted residues indicate predicted interaction regions. Colouring is the same as above.

(c) Carbene footprinting data mapped onto the hCaspase-1 p10 subunit sequence. Colouring is the same as above.

Figure 4.19 | Carbene labelling analysis of hCaspase-1-VRT-043198.

4.3 Conclusions

In summary, using a dual-protease approach, I have shown that carbene footprinting mass spectrometry accurately maps the exosite and catalytic domain interaction between full-length human GSDMD and human Caspase-1. Limited peptide-level labelling resolution was overcome by MS/MS which was used to highlight residue-level masking events at interaction sites, consistent with CTD-hGSDMD and full-length-mGSDMD X-ray crystallography structures, but with the benefit of using full length hGSDMD and reporting data on the linker peptide occupancy of the hCaspase-1 active site. Poor sequence coverage and carbene modification restricted full analysis of interaction, particularly on the hCaspase-1 (C285A) p10 subunit; nevertheless, these hindrances were largely overcome through labelling and protease optimisation, as well as differential study of both hGSDMD and hCaspase-1 (C285A). Application of carbene footprinting to the pore-forming NTD of GSDMD showed how changes in labelling compared to the full-length protein were consistent with oligomerisation. The technique was also used to show non-covalent interaction of VRT-043198 with the protease (C285A), in a similar structural arrangement to that of the wildtype. These results demonstrate the feasibility of using carbene footprinting to understand and characterise protein-ligand interactions.

5 Optimisation of labelling a helical integral membrane protein in detergent

5.1 Introduction

5.1.1 PfMATE

The movement of metabolites and xenobiotic compounds across the cell membrane is essential to cell functionality. This is mediated by cell membrane-embedded transporters. The multidrug and toxic compound extrusion (MATE) proteins are a relatively recently identified family of multidrug transporters widely distributed in all domains of life.[154] These are secondary active transporters that employ electrochemical gradients of cations to drive a range of chemically and structurally-diverse compounds from the cell.[155] They support a myriad of cellular activities from iron homeostasis in plants to the excretion of organic cation/anions from the kidneys in humans.[156] Nevertheless the overlapping substrate specificity profiles of MATEs has implicated these promiscuous transporters in multidrug resistance through removal of antimicrobial and chemotherapeutic agents from the cell.[157]

The MATE family can be classified into three subfamilies based on sequence similarity, including NorM, DinF and eukaryotic subfamilies.[157] These share ~40 %

sequence similarity.[158] The NorM and DinF subfamilies both consist of eubacterial and archael members and utilise electrochemical gradients of Na^+ and H^+ for transport. The eukaryotic subfamily only comprises of eukaryotic members which employ electrochemical gradients of H^+. MATE proteins typically consist of 400-500 amino acids.[159] These form a conserved fold of 12 TMs (mammalian MATEs have an extra TM) arranged into two pseudo-symmetrical lobes - the N-lobe and the C-lobe that produce a central V-shaped cavity situated halfway into the lipid bilayer.[160] This cavity is further classified into the N-lobe and C-lobe cavity. Crystallographic studies of MATE proteins - supported by biophysical analysis - have highlighted an alternating access mechanism of transport where conformational changes in the transporter are coupled to movement of the substrate across the membrane. This is often an antiport mechanism where MATEs occupy an outward-facing structure in their resting-state and are oriented towards either the extracellular or periplasmic side of the cell (Figure 5.1).[156] H^+ or Na^+ binding, and subsequent translocation down an electrochemical gradient drive conformational rearrangements into an inward-facing structure. In this orientation the transporter is able to bind to its substrate before undergoing another conformational change to an outward-facing conformation, followed by release of the substrate. Doubly occluded conformational states are thought to be occupied between inward- and outward-facing rearrangements where periplasmic and cytoplasmic gates restrict access to substrate binding sites.[156] Additionally, other monovalent cations including K^+, Rb^+ and Li^+ have been reported to further influence substrate efflux in some MATEs.[156]

5.1.2 PfMATE transport mechanism

PfMATE is a member of the DinF subfamily of MATE transporters found in the thermophile *Pyrococcus furiosus*. This transporter is thought to couple H^+ ions to substrate transport. PfMATE exhibits an unusually hydrophobic cleft whilst other binding sites seen in other MATES are typically polar. Crystal structures of PfMATE in an outward-facing conformation have shed insight into the mechanism of substrate binding and ejection. Research by Tanaka and colleagues showed that Br-NRF bound to the N-lobe cavity of its central cleft.[161] The authors suggested that the molecule was primarily recognised through shape complementarity, further aided by hydrogen bonding between the ligand's 4-oxo group and Gln34, the amine group of the 7-

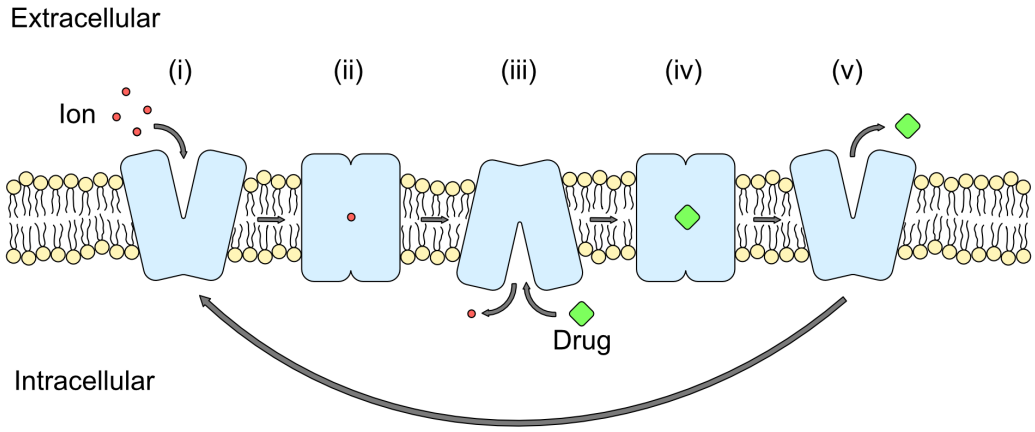


Figure 5.1 | Antiport mechanism of transport in MATE proteins. (i) In the resting state, an outward-facing conformation is occupied allowing binding of a monovalent cation, inducing a conformational change. (ii) Transient doubly occluded structure. (iii) Formation of an inward-facing conformation, allowing release of the ion, binding of a substrate and another conformational change. (iv) Transient doubly occluded structure. (v) Formation of the outward-facing structure and release of the substrate in the extracellular space.

piperazine moiety and Asn157, and the 3-carboxylate group and Asn180. PfMATE crystals collected at differing pH highlighted changes to the conformation of TM1.[161] At pH 7, the helix was shown to occupy a straight conformation whilst at pH 6 it displayed a bent structure in which the helix was kinked at Gly30 and Pro26, causing it to occupy the Norfloxacin derivative binding site. Further comparisons between these conformers showed that in the bent conformation, protonation of Asp41 had induced drastic rearrangements in a hydrogen bonding network involving hydrophilic side chains of TM1, TM4, TM5 and TM6. The authors proposed a model where, in the outward-facing, substrate-bound state, protonation of Asp 41 induced substrate ejection through bending and protrusion of TM1 into the substrate binding pocket.[161] Nevertheless calculated pKas of bent and straight crystal structures were <4, and as such, even in the low pH crystal structure PfMATE would be expected to be deprotonated.[160] The related DinF family member DinF-Bh from *Bacillus halodurans* does not exhibit bending of TM1, despite a more appropriate (experimentally determined) pKa of 7.[162] However, research on VcmN from *Vibrio cholerae* has supported TM1 bending being an important aspect of the MATE transport cycle.[156] Recently, reinterpretation of these crystal structures has lead to the identification of a Na^+ binding site within the NTD including Asn180, Thr202 and Ala198 and stabilised by a proton bound to Asp184 that mediates carboxylate-carboxylate interactions with Asp 41.[163] MD simulations further

showed that this cation binding pocket was strongly selective of Na^+ ions over K^+ ions but only marginally discriminate over H^+ ions.[163]

Recent inward-facing structures have provided further insight into PfMATE's transport mechanism.[164] TM1 and TM7 were shown to be loosely bound to their respective lobes. The increased flexibility was suggested to induce rigid body movements in these lobes, driving a switch between conformations - similar to the rocker-switch mechanism of transport observed in the major facilitator superfamily (MFS) of transporters (indeed, prior to their reclassification in 1998, MATE transporters were cateogrised as MFS transporters[159]). Further support of this rocker-switch mechanism came from observations that the extracellular-facing side of the transporter had closed and formed a 10 Å thick barrier, mediated by a network of Van der Waal and hydrophobic interactions. Interestingly, the intracellular gate observed in the outward-facing conformation was 14 Å thick but involved both ionic and hydrophobic interactions. This greater stability may perhaps account for the majority of outward-facing MATE structures observed; however, Zakwreskra and colleagues only observed inward-facing PfMATE structures when the transporter was treated with native archaeal lipids.[164] This hinted that PfMATE could be involved in lipid transport; however, according to the antiport model, substrate specificity should dictate the outward-facing conformation rather than the inward-facing structure.[156] Indeed, MD simulations have shown that stable lipid binding is only observed in an outward-facing conformation.[165] Recently, DEER spectroscopy has been used to exhaustively investigate conformational rearrangements of PfMATE and the energetic changes that underpin these transitions.[166] In this work, low pH was used to experimentally mimic protonation. Conformational changes were not observed in DDM-solubilised PfMATE at low pH. In contrast, changes in distance between spin pairs were observed at low pH in the presence of native and even nonnative lipids. Nevertheless, this was not observed under physiological pH with the addition of lipid, suggesting that protonation was crucial for changes in PfMATE conformation. Following on from this, the authors sought to identify residues involved in this protonation event.[166] Mutagenesis was used to implicate Glu163 (found on the cytoplasmic 4-5 loop) and Tyr224 (found on the cytoplasmic 6-7 loop) in a conformational switch to an inward-facing conformation. The authors proposed that in the outward-facing conformation, Glu163 and Tyr224 interact

through a pi-charge interaction. However, upon protonation of Glu163, this belt-like structure is disrupted, releasing tension in the cytoplasmic 6-7 loop and allowing TM3 to tilt, and induce isomerisation to the inward-facing conformation.[156]

As of writing, OmpF is the only carbene labelled membrane protein that has been reported.[100] This β-barrel displayed high sequence coverage following tryptic digestion as well as moderate carbene modification of transmembrane spanning β-sheets. In contrast, PfMATE, as well as the majority of clinically relevant membrane protein receptors display highly compact alpha-helical arrangements and do not share the heightened chemical accessibility and wealth of lysine and arginine residues that β-barrels do (making tryptic digestion difficult). As such, optimisation of carbene labelling and digestion conditions of helical IMPs is necessary to report on these clinically relevant systems.

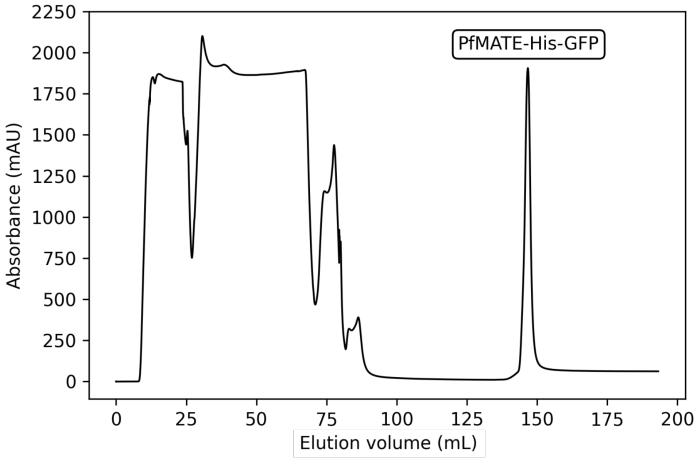
Here, for the first time, I sought to apply and optimise varying aspects of carbene footprinting to the IMP, PfMATE solubilised in DDM. This work represents development of the covalent labelling workflow, and more generally, gel-based proteomics, to facilitate improved bottom-up analysis of membrane protein systems.

5.2 Results and discussion

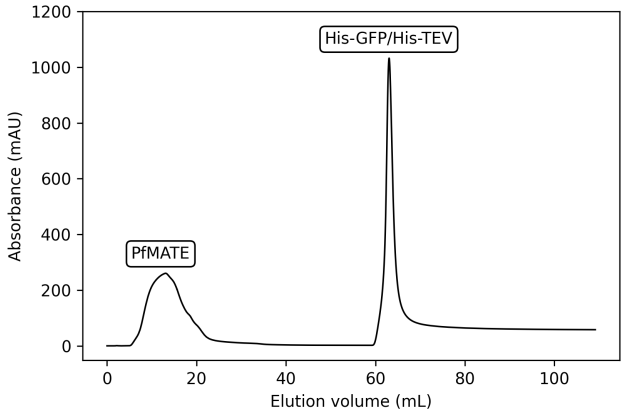
5.2.1 PfMATE production

I performed PfMATE expression and purification of PfMATE to produce protein samples for carbene labelling experiments. Expression was performed in BL21 cells and GFP-tagged PfMATE was successfully solubilised in DDM and purified with IMAC (Figure 5.2a), as described in the methodology (Section 2.1.1).

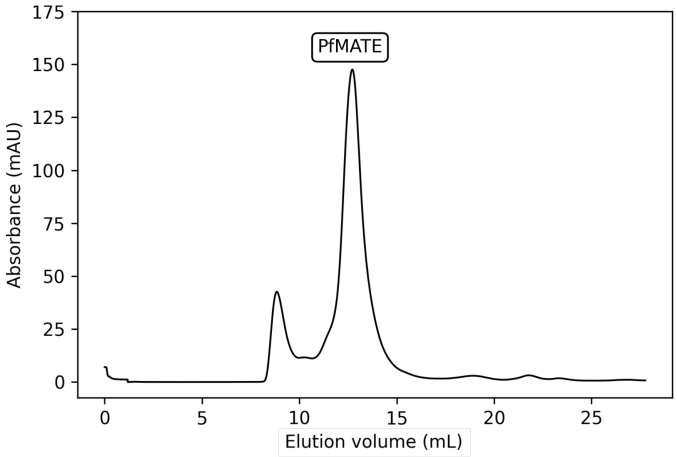
TEV Protease successfully cleaved off the His-tag and GFP-tag and reverse IMAC gave the processed PfMATE from the flow-through (Figure 5.2b). PfMATE samples were further purified and buffer exchanged using SEC (Figure 5.2c). SDS-PAGE showed that PfMATE samples were pure (Figure 5.2d).



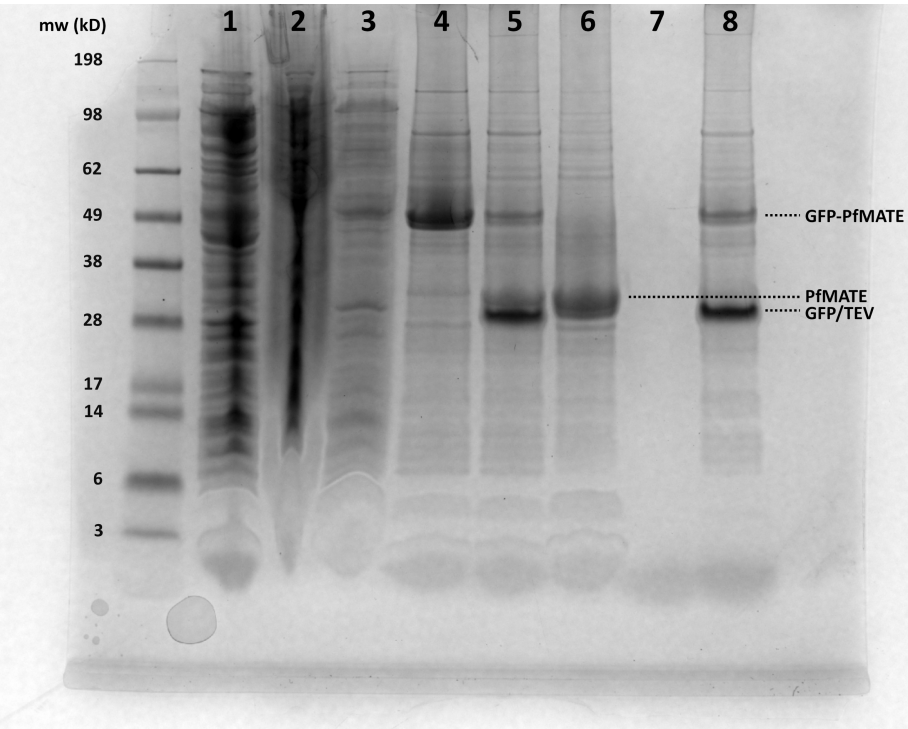
(a) PfMATE-His-GFP IMAC Trace. The PfMATE-His-GFP peak is indicated, eluting at ~145 minute. The high UV observed between ~10–90 min was attributed to the elution of other (non-His-tagged) membrane proteins.



(b) PfMATE Reverse IMAC Trace. The PfMATE peak is indicated, eluting at ~5 minute.



(c) PfMATE SEC Trace. The PfMATE peak is indicated, eluting at ~13 minute.



(d) PfMATE purification gel. Lane 1 - Cell lysate, Lane 2 - Resuspended membranes, Lane 3 - Solubilised membranes, Lane 4 - Immobilised metal affinity chromatography, Lane 5 - TEV cleavage, Lane 6 - SEC purification, Lane 7 - Concentrated flow through, Lane 8 - Reverse immobilised metal affinity chromatography eluent.

Figure 5.2 | PfMATE purification.

5.2.2 Carbene modification of n-dodecyl-beta-maltoside micelles

Since carbene labelling experiments were to be performed on PfMATE and other membrane proteins containing DDM, it was expected that NaTDB would react with the detergent in accordance with results reported by Manzi and colleagues when labelling OG-solubilised OmpF.[100] To investigate this, 0.02 % (w/v) DDM in water was labelled with 20 mM NaTDB. Irradiated samples were diluted and analysed by LC-MS/MS/MS.

The labelled DDM TIC showed a number of intense peaks eluting between 45–50 min (Figure 5.3). Only one of these peaks corresponded to unlabelled DDM dimer (1021 *m/z* - 2x 510 *m/z*) whilst all subsequent peaks corresponded to the labelled monomer 713 *m/z* (510 *m/z* (DDM) + 202 *m/z* (carbene) + proton (1 *m/z*)) suggesting that the carbene was inserting into differing positions on the detergent molecule, influencing chromatographic retention time.

To gain further insight into the localisation of the carbene on the detergent, the labelled monomer was selected for fragmentation with CID (Figure 5.4). This resulted in the production of two major species, including 551 *m/z* and 389 *m/z* that were generated

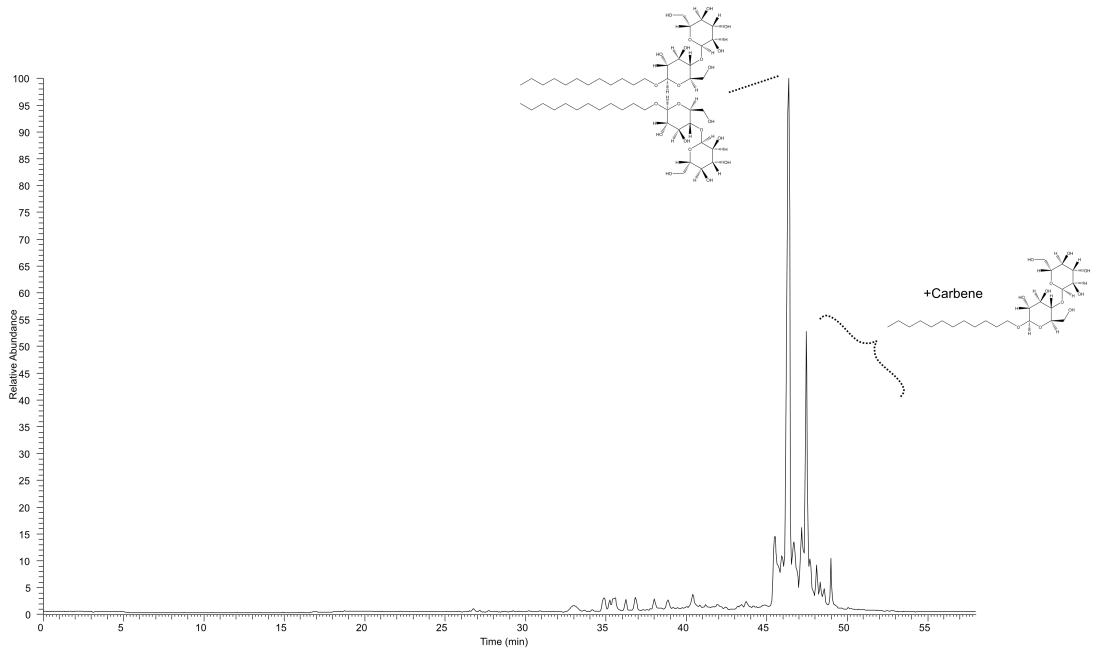


Figure 5.3 | **TIC of labelled DDM.** Several several large peaks were observed. A single peak corresponded to the unlabelled DDM dimer whilst the remainder represented carbene modified DDM monomers.

through successive fragmentation of the maltose head group, characterised by loss of one (162 *m/z*) or two (324 *m/z*) glucose units. The presence of +202 *m/z* on either species suggested that the label was modifying the detergent alkyl chain and at no point in this analysis was a modified maltose or glucose molecule observed.

A wideband activation (using the FTICR) was then applied to CID fragmentation of the labelled monomer which again resulted in loss of the maltose head group but now also showed extensive fragmentation of the detergent tail, highlighting carbene modification of successive CH_2 groups and reiterating carbene modification at different positions along the tail (Figure 5.5). Given the architecture of detergent micelles in which hydrophobic tail groups are internalised and sugar head groups are exposed to solvent, carbene labelling of the alkyl chain suggested that the diazirine was inserting itself into the micelle and intercalating with detergent micelles prior to irradiation. These results suggested that carbene labelling of PfMATE in DDM were feasible.

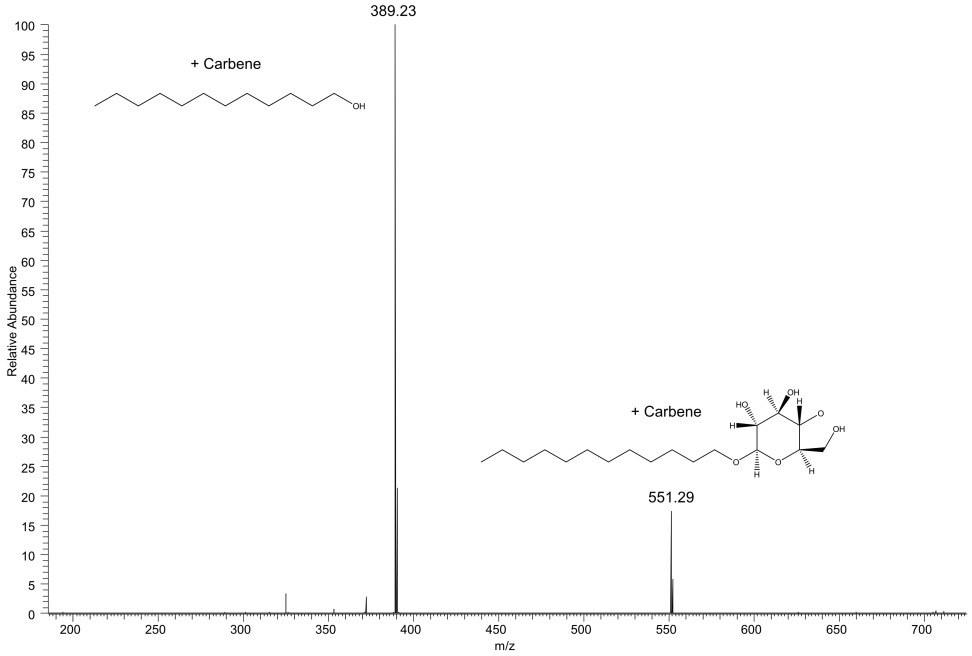


Figure 5.4 | Combined MS^2 spectrum of labelled DDM monomer. CID fragmentation of the 713 *m/z* parent ion generated two major fragments, including 389 *m/z* and 551 *m/z*.

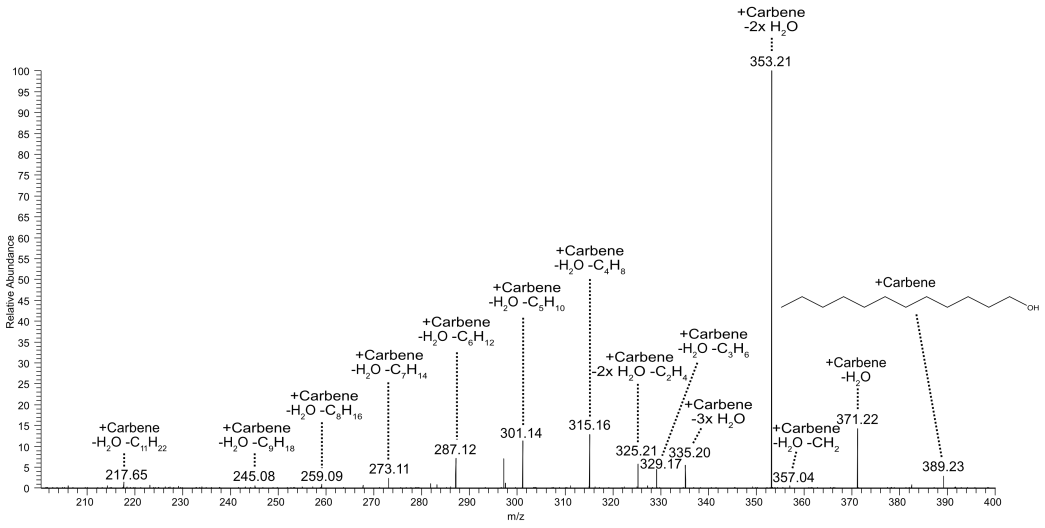


Figure 5.5 | Combined MS^3 (with an additional wideband activation) spectrum of labelled DDM monomer. CID fragmentation showed successive loss of CH_2 groups, highlighting carbene labelling at multiple positions along the detergent tail.

5.2.3 Carbene footprinting of PfMATE

5.2.3.1 Optimisation of digestion conditions

Proteolytic digestion is a prerequisite to bottom-up proteomic analysis and a key determinant in the peptides observed by LC-MS. Indeed, bottom-up proteolytic analysis suffers from two main challenges: forming and detecting peptides.[167] Membrane proteins are typically hydrophobic and often lack tryptic cleavage sites which may impede the generation and extraction of peptides. Therefore, the choice of proteolytic conditions used to digest membrane proteins was an important consideration in carbene labelling experiments.

Initial *in-silico* digestions were performed on the PfMATE sequence using the PeptideCutter server to identify appropriate proteases for experimental sequence coverage analysis. PfMATE was theoretically digested with chymotrypsin, trypsin and pepsin. Peptides were filtered to include a length of between 4 and 40 residues and a maximum of three missed cleavages. Lengths beyond these ranges were unlikely to be resolved by RP-HPLC or MS.

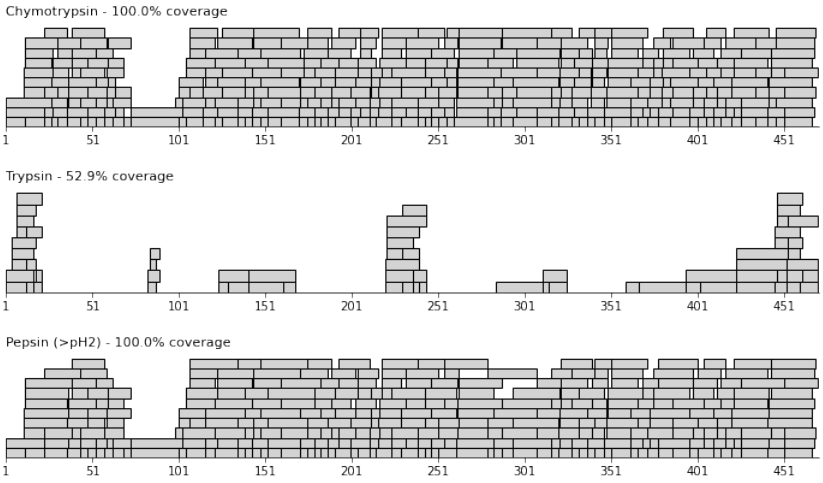


Figure 5.6 | *In-silico* digest of PfMATE with chymotrypsin, trypsin and pepsin. Peptides were filtered between 4-40 residues with a maximum of three missed cleavages. Grey bars represent predicted peptides and overlapping bars indicate missed cleavages.

The *in-silico* digestions, whilst displaying an unrealistic level of sequence coverage, proved useful in highlighting the inaccessible regions of proteolytic cleavage (Figure 5.6). This was certainly the case for trypsin where cleavage sites were lacking and multiple regions of the sequence remained uncovered. In contrast, chymotrypsin and pepsin (pH < 2) generated a huge number of theoretical peptides with both proteases achieving 100 % sequence coverage. This reflected the highly hydrophobic

nature of the membrane protein and abundance of cleavage sites. The lack of cleavage sites for chymotrypsin and pepsin between residues 75-100 resulted in only two hypothetical peptides of the appropriate length covering this region. Conversely, several short tryptic peptides were predicted to be generated around residues 75-80. Therefore, trypsin could be used in parallel with chymotrypsin or pepsin to provide complementary data in this region.

In-gel digestions were conducted at 1 h and 18 h intervals (in line with literature and manufacturer's recommendations for digestion time ranges), with and without ProteaseMax - a mass spectrometry-compatible surfactant that has been shown to increase peptide coverage during proteolytic digestion by solubilising hydrophobic regions of proteins.[168] A DDA method was used to identify digested peptides where the top four most abundant ions in each precursor scan were fragmented with a collision gas prior to species identification using a database search. In-solution digestions were not performed on samples since DDM (a non-ionic detergent) was not compatible with LC-MS.

The chymotryptic digestion showed that an overnight incubation provided higher sequence coverage compared to a 1 h digestion (Figure 5.7). Of the overnight digestions, the incorporation of ProteaseMAX only seemed to slightly benefit sequence coverage. In contrast, the inclusion of ProteaseMAX had a much larger benefit on 1 h digestions. Despite *in-silico* digests suggesting complete coverage of residues 405-469, nearly all chymotryptic digests lacked sequence coverage in this region. The lack of peptides observed in this region may be explained by them being retained in gel pieces or on the RP-HPLC column, being outside of the *m/z* range scanned during MS analysis or being too low in signal intensity to be selected for DDA analysis.

Sequence coverage with trypsin was sparse across different digestion conditions and reflected *in-silico* predictions (Figure 5.8). Moreover, the 1 hour digest treated with ProteaseMAX generated higher sequence coverage than the overnight digest with the surfactant, inline with the manufacturer guidelines that suggested that longer digestion periods with the surfactant may impede peptide recovery rates. No tryptic peptides were observed between residues 75-100 suggesting that trypsin may not be able to provide complementary structural data to chymotrypsin and pepsin in this region.

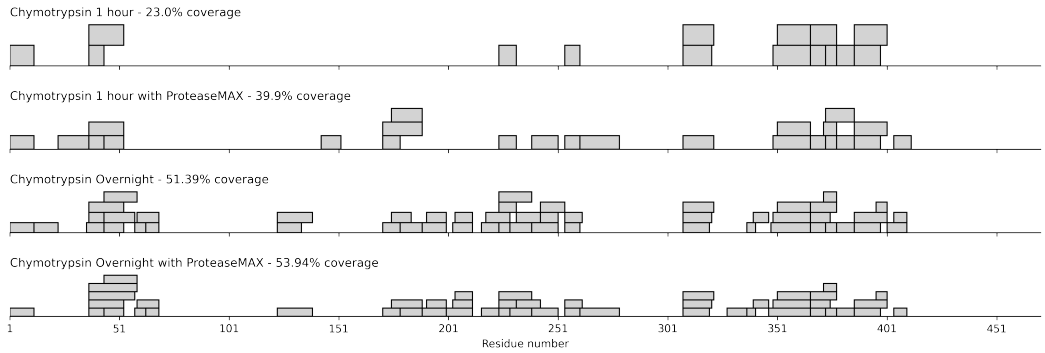


Figure 5.7 | Sequence coverage achieved by DDA LC-MS/MS analysis of PfMATE following digestion with chymotrypsin for 1 h and overnight time points, with and without ProteaseMAX Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

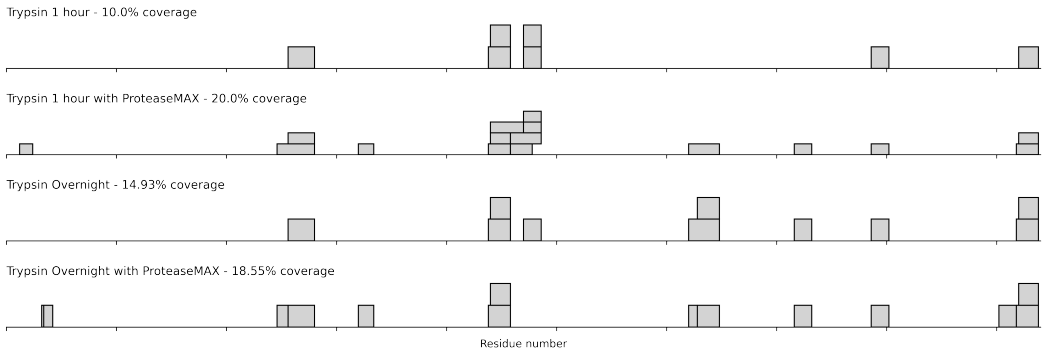


Figure 5.8 | Sequence coverage achieved by DDA LC-MS/MS analysis of PfMATE following digestion with trypsin for 1 h and overnight time points, with and without ProteaseMAX Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

Sequence coverages obtained from pepsin digests were consistent across different experimental conditions (Figure 5.9). The overnight digest with ProteaseMAX led to the highest sequence coverage; however, given that the surfactant was acid-labile and all pepsin digestions were carried out at pH 2, it was surprising that an effect on sequence coverage was observed. Further to this, the addition of ProteaseMAX to the 1 h digestion showed no increase in sequence coverage. Despite the *in-silico* digest showing 100 % sequence coverage, several regions lacked sequence coverage. This again included residues 75-100 and the C-terminal region (as observed with chymotrypsin) but also several other locations including peptides 25-50, 220-240 and 295-340.

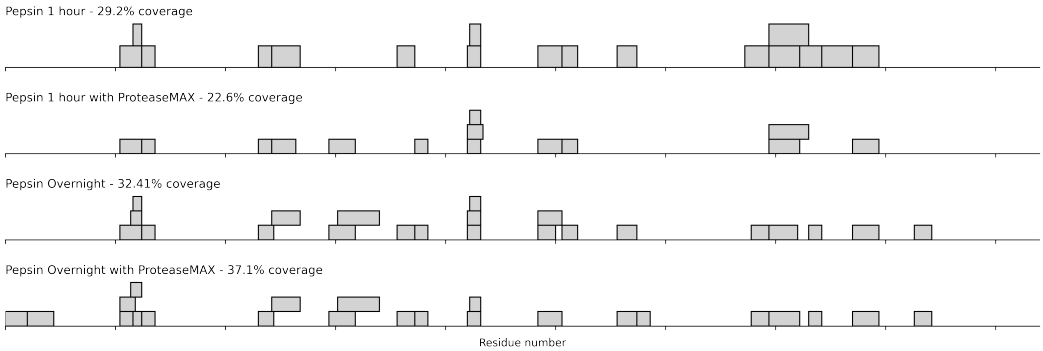


Figure 5.9 | Sequence coverage achieved by DDA LC-MS/MS analysis of PfMATE following digestion with pepsin for 1 h and overnight time points, with and without ProteaseMAX Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

5.2.3.2 Optimisation of carbene labelling conditions

Initial carbene footprinting experiments were performed on singlet PfMATE samples (N=1) at 10, 20 and 40 mM NaTDB followed by in-gel proteolysis to determine which diazirine concentration was optimal for labelling of the membrane protein. It was envisioned that a higher NaTDB concentration would be needed compared to typical concentrations used in soluble protein labelling (around 10–20 mM) in order to overcome label quenching by the detergent micelle.

The tryptic digest of labelled PfMATE corresponded to 19 % peptide sequence coverage. The digest showed an upwards trend in f_{mod} from 10–40 mM NaTDB (Figure 5.10a). An exception to this was peptide 236-243 (EILKDILR) which showed the highest labelling with 20 mM NaTDB. Peptides 221-235 (DTYVDITLRDFSPSR),

230-235 (DFSPSR) and 230-239 (DFSPSREILK) showed 100 % carbene modification at all three NaTDB concentrations. Despite promising levels of carbene labelling an obvious drawback to the tryptic digestion was the poor sequence coverage.

The chymotryptic digest of labelled PfMATE corresponded to 70 % peptide sequence coverage, 15 % higher than the coverage observed during the DDA-based protease sequence coverage optimisation (full-scan events were used when labelling). However, this may be explained by lower intensity peptides which were not picked for activation. The chymotryptic digest of labelled PfMATE also showed upwards trend in f_{mod} from 10–40 mM NaTDB (Figure 5.10b). Two peptides, including 321-328 (YAIKIAFM) and 398-405 (TIFRTLVM) displayed 100 % carbene modification at 10, 20 and 40 mM diazirine concentration whilst the majority of remaining peptides showed 0 % modification at all probe concentrations. The dual-digest corresponded to 71 % labelled sequence coverage (Figure 5.10c). The labelling optimisation results suggested that 40 mM NaTDB in combination with tryptic and chymotryptic digests provided the greatest carbene labelling with complementary sequence coverage.

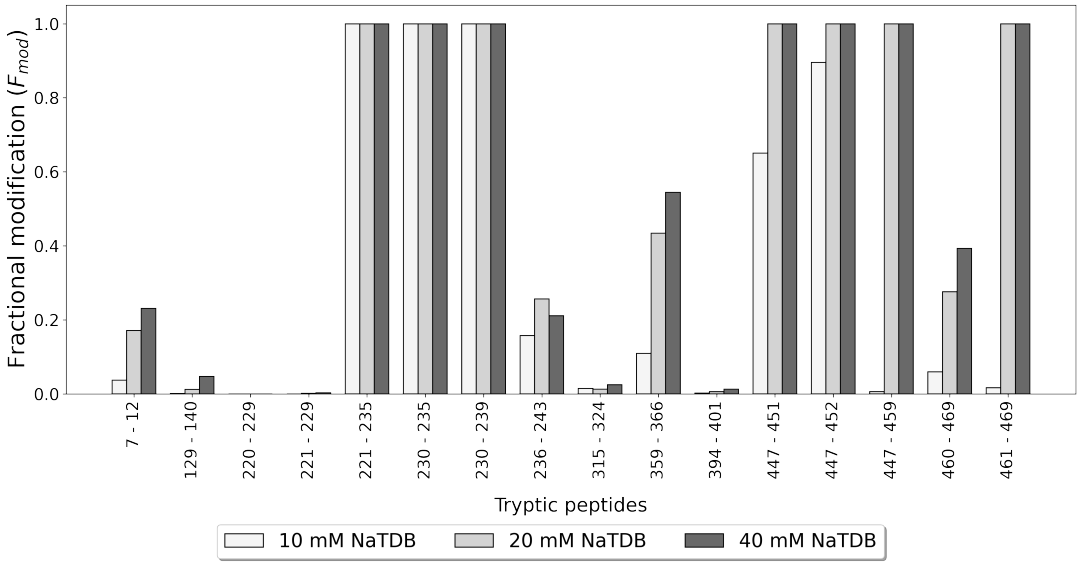
Protein	Optimum digestion conditions	Optimum labelling conditions
PfMATE	Chymotrypsin Overnight with ProteaseMAX	40 mM NaTDB

Table 5.1 | Summary of optimal digestion and carbene labelling conditions for PfMATE.

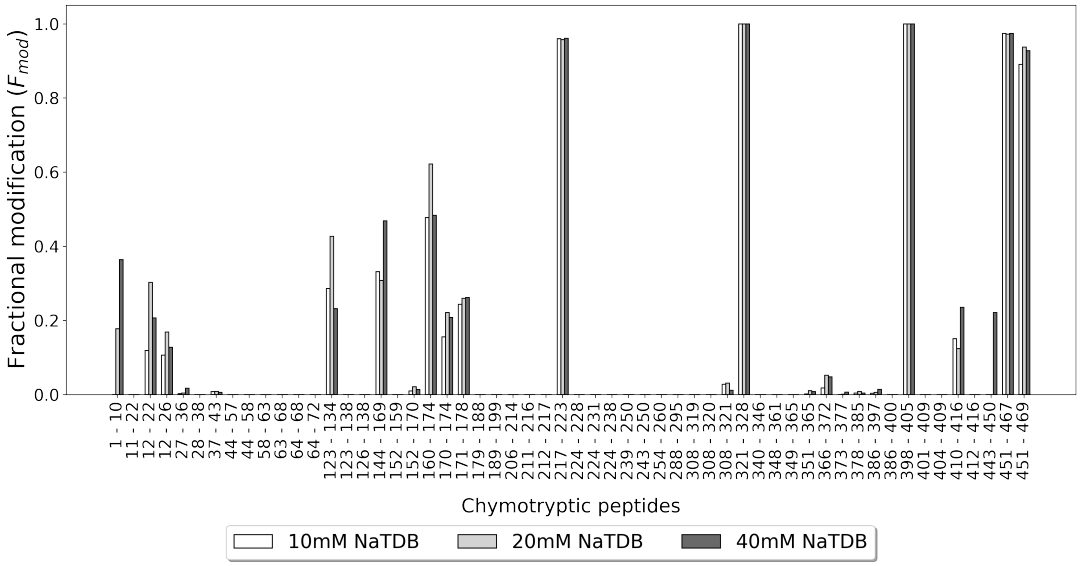
5.2.3.3 Optimised carbene labelling of PfMATE

Following successful optimisation of digestion and labelling conditions (Table 5.1), modification of replicate PfMATE samples was performed at 40 mM NaTDB before trypsin and chymotrypsin were employed for proteolytic digestion (Figure S6).

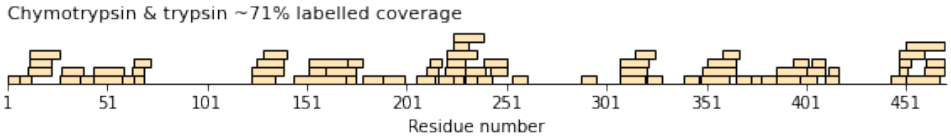
The tryptic digest of labelled PfMATE displayed high chemical modification (>0.8 f_{mod}) on peptides 221-235 (DTYVDITLRDFSPSR), 230-235 (DFSPSR) and 230-239 (DFSPSREILK), and peptides 447-451 (ISALK), 471-452 (ISALKK) and 447-459 (ISALKKTSATGGK) (Figure 5.11a). In contrast, peptides 220-229 (RDTYVDITLR) and 221-229 (DTYVDITLR) displayed almost no carbene labelling despite both peptides sharing "DTYVDITLR" with peptide 221-235 that reported an f_{mod} of 1. It should be borne in mind that peptide-level labelling is contained over the length of an entire peptide, and the inclusion of an additional amino acid may drastically influence the f_{mod}.



(a) Carbene labelling optimisation of PfMATE with a trypsin digestion. The fractional modification of each tryptic PfMATE peptide is shown at 10, 20 and 40 mM NaTDB.



(b) Carbene labelling optimisation of PfMATE with a trypsin digestion. The fractional modification of each chymotryptic PfMATE peptide is shown at 10, 20 and 40 mM NaTDB.



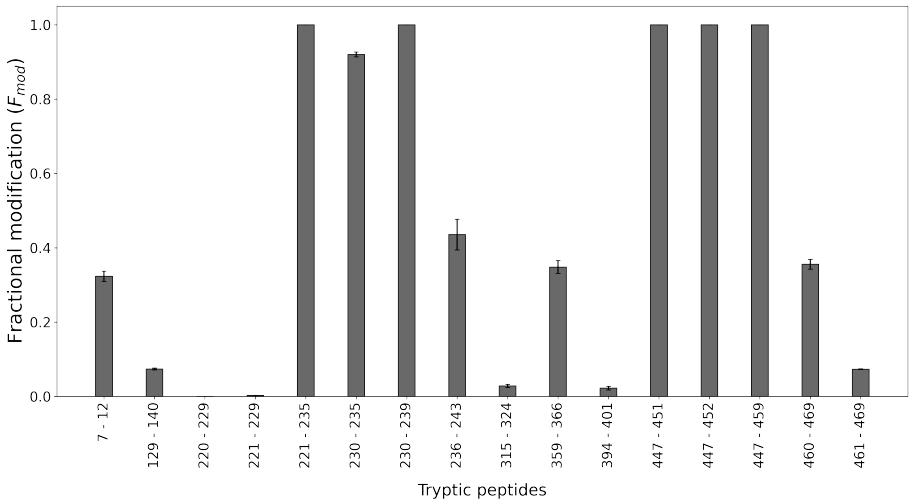
(c) Sequence coverage plot of concatenated carbene labelled PfMATE chymotryptic and tryptic peptides showing overall labelling coverage. Orange bars represent observed peptides and overlapping bars represent missed cleavages.

Figure 5.10 | Carbene labelling optimisation of PfMATE.

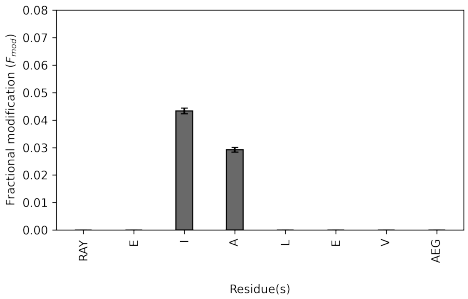
This suggested that modification of residues on "DFSPSR" were responsible for high carbene labelling, and was in agreement with the high level of modification observed on both of the following two peptides 230-235 and 230-239 (which contained the peptide "DFSPSR"). Interestingly, when mapped to the PfMATE structure, "DTYVDITLR" and "DFSPSR" both corresponded to unstructured loop regions at solvent-exposed portions of the protein, suggesting that disparities in labelling were instead due to residue-specific labelling preferences (Figure 5.11i). Peptides 315-324 (LETAYLYAIK) and 394-401 (SLILTIFR) also displayed low carbene modification (<0.1 f_{mod}). These peptides were located at transmembrane-spanning regions of the protein and were appropriately surrounded by detergent molecules (Figure 5.11j). Therefore it appeared that whilst NaTDB molecules were able to penetrate micelles, the presence of detergent molecules reduced chemical labelling at transmembrane-spanning regions.

To gain higher resolution insight into carbene modification, CID fragmentation was conducted on several labelled tryptic peptides. These included peptides 129-140 (GEAVELAIEYAR), 230-235 (DFSPSR), 236-243 (EILKDILR), 359-366 (GDLISALR), 394-401 (SLILTIFR), 447-459 (ISALKKTSATGGK) and 461-469 (ASGENLYFQ) (Figure 5.11). MS/MS analysis of tryptic peptides (apart from terminal peptide 461-469) showed prominent y ion series, owing to the presence of C-terminal basic residues. In most cases, labelling data was resolved to the residue level however carbene modification of peptides 230-235 and 447-459 could only be discerned to the sub-peptide level due to poor intensity of labelled y ions.

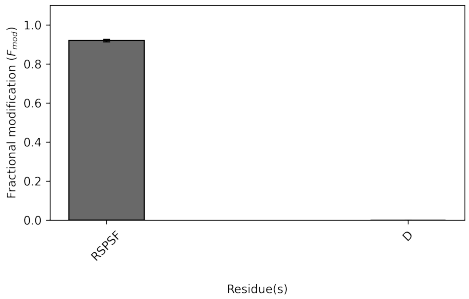
The chymotryptic digest of labelled PfMATE displayed an average peptide f_{mod} of 0.16 compared to 0.47 observed in the tryptic digest (Figure 5.12a). Of the chymotryptic peptides, 64 % showed a f_{mod} of 0.1 or less, compared to only 25 % of tryptic peptides. These differences likely represented the varying cleavage sites between trypsin and chymotrypsin, where tryptic cleavage sites were inherent to regions of greater solvent exposure (and greater chemical accessibility to the diazirine) and therefore increased likelihood of carbene labelling. Chymotrypsin's specificity towards hydrophobic residues enabled digestion of transmembrane-spanning residues, which highlighted reduced labelling of the these peptides. This was likely due to reduced chemical accessibility caused by helical packing and enclosure by detergent molecules. High chemical modification (>0.8 f_{mod}) was observed on



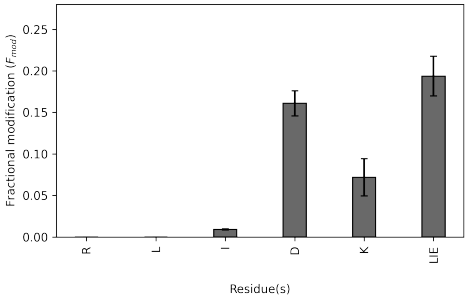
(a) Carbene footprinting histogram of PfMATE with a trypsin digestion. The fractional modification of each tryptic PfMATE peptide is shown. Error bars are ± standard deviation (n = 4).



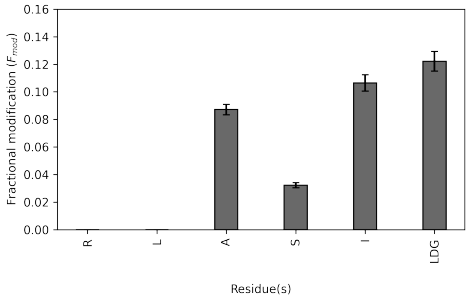
(b) Sub-peptide analysis of GEAVELAIEYAR (peptide 129-140).



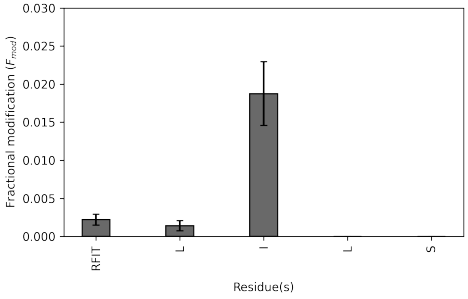
(c) Sub-peptide analysis of DFSPSR (peptide 230-235).



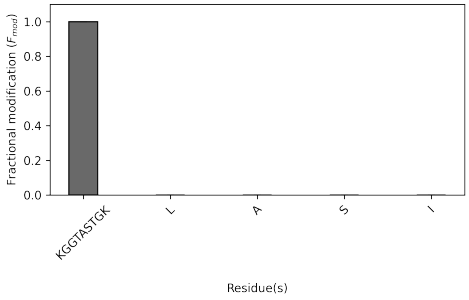
(d) Sub-peptide analysis of EILKDILR (peptide 236-243).



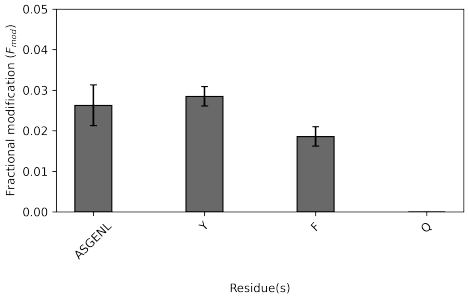
(e) Sub-peptide analysis of GDLISALR (peptide 359-366).



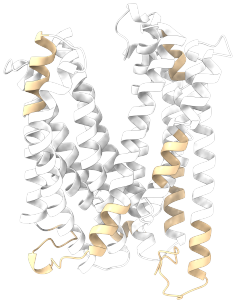
(f) Sub-peptide analysis of SLILTIFR (peptide 394-401).



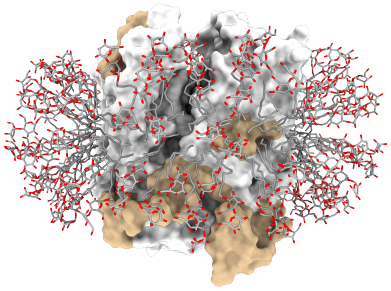
(g) Sub-peptide analysis of ISALKKTSATGGK (peptide 447-459).



(h) Sub-peptide analysis of ASGENLYFQ (peptide 461-469).



(i) Tryptic carbene footprinting data mapped onto the PfMATE structure. Colour scheme is as follows: tan = labelling, grey = no labelling, white = no peptide coverage.

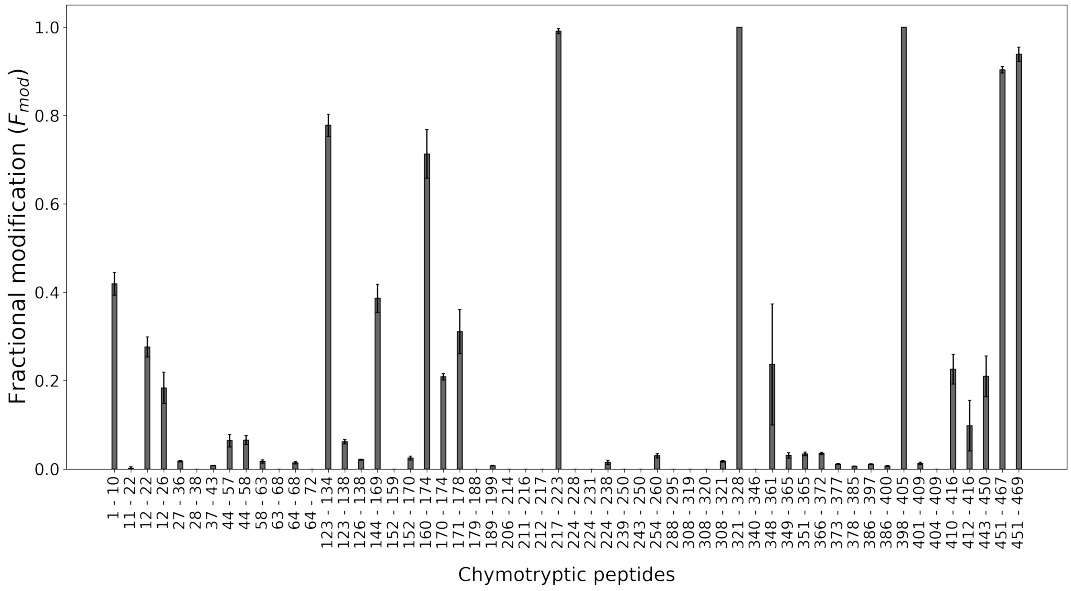


(j) Tryptic carbene footprinting data mapped onto the PfMATE DDM-solubilised structure, generated by the CHARMM-GUI. Colour scheme is the same as above.

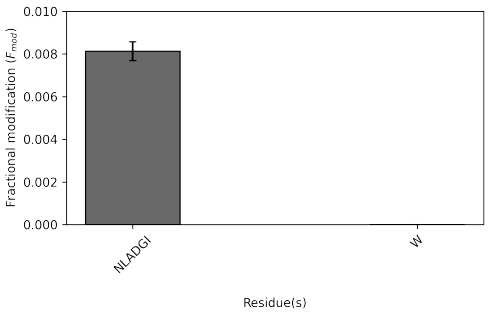
Figure 5.11 | **Carbene footprinting of PfMATE using a trypsin digestion.** MS/MS of labelled peptides revealed higher resolution labelling information.

chymotryptic peptides 123-134 (RSMGAKGEAVEL), 217-223 (FVKRDTY), 321-328 (YAIKIAFM), 398-405 (TIFRTLVM), 451-467 (KKTSATGGKRASGENLY) and 451-469 (KKTSATGGKRASGENLYFQ). Despite these peptides sharing overlapping coverage with the tryptic peptides, very little quantitative similarity was observed, again suggesting that specific residues were responsible for differing degrees of peptide-level modification. Further to this, highly promiscuous proteases (i.e. those which cleave at multiple residues) that would likely generate multiple missed cleavages could be used to improve the resolution of labelling data by comparing labelling between overlapping fragments and attributing changes in modification to those residues which are or are no longer present. This would omit the need for additional MS/MS steps which would be of particular use for peptides that did not fragment in a desired manner. The development of such a framework (i.e. using 'overcleavage' to generate overlapping modified peptides) to aid labelling analysis should be explored in future work. Returning to the labelled chymotryptic digest, peptides 321-328 and 398-40. These peptides were located on central transmembrane-spanning helices yet displayed high carbene modification and again suggested that despite being surrounded by detergent molecules and neighbouring helices, the diazirine probe showed either high chemical accessibility and/or specificity to residues on these peptides.

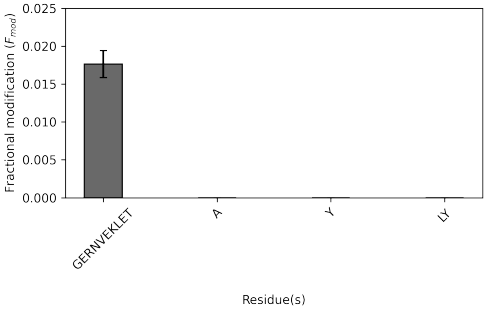
CID fragmentation was also conducted on labelled chymotryptic peptides. These included peptides NLADGIW (37-43), GERNVEKLETAYLY (308-321), SESAQVIKGDLISAL (351-365), RTLPVFL (366-372), VLTPF (373-377) and GMMTSAMF (378-385) (Figure 5.12). In contrast to tryptic MS/MS analysis, fragmentation of labelled chymotryptic peptides primarily showed b ion series with reduced unlabelled fragment ion intensities compared to tryptic y ion intensities making comprehensive residue-level analysis difficult. Nevertheless, in most cases, fragmentation enabled determination of labelling to the sub-peptide label. These results highlighted carbene insertion at specific residues and sub-peptides, and indicated that forthcoming footprinting experiments on other membrane protein systems followed by MS/MS analysis would enable high-resolution analysis of labelling differences.



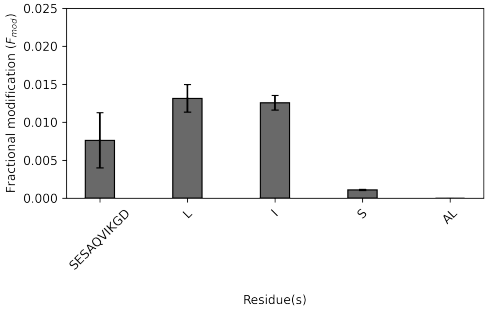
(a) Carbene footprinting histogram of PfMATE with a chymotrypsin digestion. The fractional modification of each chymotryptic PfMATE peptide is shown. Error bars are ± standard deviation (n = 4).



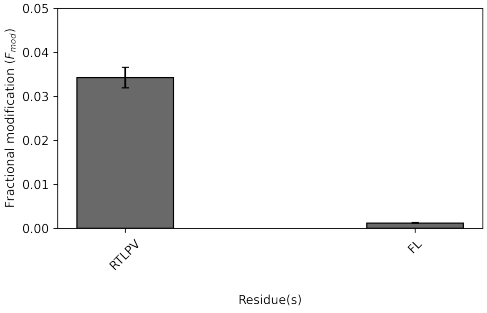
(b) Sub-peptide analysis of NLADGIW (peptide 37-43).



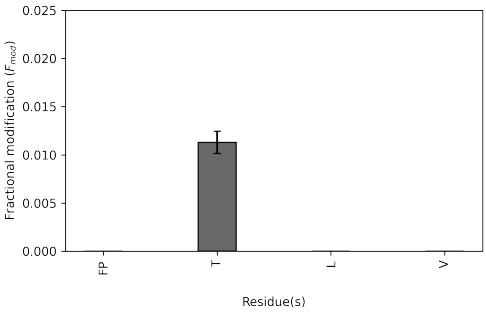
(c) Sub-peptide analysis of GERNVEKLETAYLY (peptide 308-321).



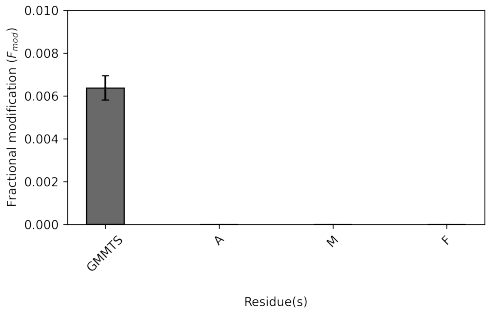
(d) Sub-peptide analysis of SESAQVIKGDLISAL (peptide 351-365).



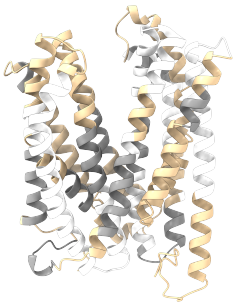
(e) Sub-peptide analysis of RTLPVFL (peptide 366-372).



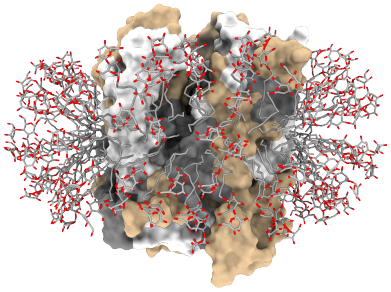
(f) Sub-peptide analysis of VLTPF (peptide 373-377).



(g) Sub-peptide analysis of GMMTSAMF (peptide 378-385).



(h) Chymotryptic carbene footprinting data mapped onto the PfMATE structure. Colour scheme is as follows: tan = labelling, grey = no labelling, white = no peptide coverage.



(i) Chymotryptic carbene footprinting data mapped onto the PfMATE DDM-solubilised structure, generated by the CHARMM-GUI. Colour scheme is the same as above.

Figure 5.12 | Carbene footprinting of PfMATE using a chymotrypsin digestion. MS/MS of labelled peptides revealed higher resolution labelling information.

5.3 Conclusions

In summary, aspects of the carbene footprinting technique have been optimised to facilitate labelling of the membrane protein PfMATE. The use of chymotrypsin with ProteaseMAX showed improved peptide-level sequence coverage compared to chymotrypsin alone. Moreover, pooling of tryptic and chymotryptic peptides enabled higher sequence coverage than when a single protease was used alone. Labelling at 40 mM NaTDB showed increased carbene modification compared to those (lower) concentrations typically employed with soluble systems. However, the overall sequence coverage and fractional modification was lower than typically observed with soluble proteins. Carbene modification of transmembrane-spanning helices was also seen, in agreement with previous research that the diazirine precursor can insert into detergent micelles.[100] Indeed, labelling of detergent micelles showed modification of the detergent tail, with MS/MS analysis highlighting insertion at varying positions along its length. MS/MS was also conducted on selected modified tryptic and chymotryptic peptides, enabling sub-peptide labelling analysis. These results represent developments to carbene footprinting mass spectrometry and highlight its potential in studying membrane protein-ligand interactions.

6

Mapping binding interactions and structural changes of the beta-1 adrenergic receptor using carbene footprinting

6.1 Introduction

6.1.1 G protein-coupled receptors

GPCRs are the largest family of transmembrane proteins.[169] GPCRs play vital roles in signal transduction by responding to a diverse set of endogenous and exogenous ligands and have been implicated in the perception of smell, taste and light.[170] Due to this role in vertebrate physiology, GPCRs are frequently implicated in disease. They are also a rich source of therapeutic targets in the pharmaceutical industry. Out of all human-encoded GPCRs, 10 % are current drug targets[171] (this constitutes 35 % of all clinically-approved therapeutic agents[172]); however, their inherent instability in detergents has historically hampered their structural interrogation. It has only been until relatively recently — owing to developments in protein engineering and breakthroughs to structural techniques — that the field of GPCR structural biology has emerged.

6.1.1.1 G protein-coupled receptor structure

GPCRs consist of an extracellular N-terminus accompanied by seven transmembrane helices (TMs) that are linked by three extracellular loops (ECLs) and three intracellular loops (ICLs) as well as a final intracellular C-terminus. GPCRs can be classified according to their structural and functional similarity. Historically a class A-F system was used for this; however, the GRAFS classification system was more recently proposed based on phylogenetic analysis of human GPCRs.[171] This consists of the Glutamate (G), Rhodopsin (R), Adhesion (A), Frizzled/Taste2 (F), Secretin (S) families. TMs assume a pore-like structure in the lipid bilayer that serves to transduce extracellular stimuli through the cell membrane to the intracellular environment. GPCR activity is associated with ligand binding to the extracellular side of the receptor. This induces structural rearrangements in TMs, facilitating interaction of effector proteins with the receptor's intracellular region. Bound effector proteins can then produce an intracellular signalling response. The conformational molecular switches which accompany receptor activation are predominantly observed in TMs which shift, kink and tilt relative to one another. However as key mediators in signalling responses, GPCRs do not exist as simple binary switches but rather as highly dynamic proteins that assume an ensemble of functionally distinct populations where subtle conformational changes can shift the inactive-active GPCR equilibrium and bias the overall signalling response. As such ligands may be classified by the extent to which they modulate basal GPCR activity.[173] Ligands that elicit a physiological response are known as agonists whilst those that dampen them are termed inverse agonists. Antagonists obstruct agonist binding and impede downstream signalling events. The magnitude of the signalling response elicited by agonists, inverse agonists and antagonists can be described by preceding terms full, partial or neutral. In addition, other small molecules can bind to allosteric sites, distinct from the defined orthosteric ligand binding site to influence GPCR signalling. There are also bitopic modulators which bind to both orthosteric and allosteric sites. Additionally, GPCRs can dimerise (as well as form higher-order oligomers). Furthermore GPCR dimers often maintain functionality; moreover, research is emerging that receptor-receptor interactions may actually underpin activation through allosteric modulation.[174]

GPCR TMs typically exhibit high structural similarity. N and C-terminii, as well as the ICL that traverses TMs 5 and 6, display the most divergence, reflecting the structurally diverse ligands and effector proteins that the receptors interact with.[175]

6.1.1.2 G protein-coupled receptor signalling

Canonical GPCR signalling occurs through heteromeric G proteins (HGPs) (Figure 6.1). HGPs consist of three subunits: Gα, Gβ and Gγ. There are four classes of HGP Gα subunit: $G\alpha_s$, $G\alpha_i/G\alpha_o$, $G\alpha_q/G\alpha_{11}$ and $G\alpha_{12}/G\alpha_{13}$.[176] G protein-based receptor activation involves changes to the nucleotide-bound state of Gα. In the GPCR's inactivated state, Gα is bound to GDP between its α-helical domain and Ras-like GTPase domain and can assemble with the Gβγ dimer to form an inactive heterotrimer. Upon receptor activation, GPCRs mediate the exchange of GDP for GTP on Gα. GDP release is thought to occur through domain separation. A nucleotide-free state is thought to be short-lived owing to the high concentration of GTP within cells. Binding of GTP to Gα causes the heteromeric G protein to dissociate into Gα and Gβγ. These then modulate downstream signalling events.[169] Gα is classically known to target adenylyl cyclase, resulting in changes to the level of cAMP. This regulates protein kinase A (PKA) activity, leading to changes in physiological function. The G protein signalling cycle is completed when GTP bound to Gα is hydrolysed, allowing the subunit to reassociate with Gβγ. Hydrolysis may occur through inherent GTPase activity of Gα. Uncoupling of GTP-bound Gα may be slowed by effector proteins. GAPs can also accelerate the GTPase activity of Gα.

6.1.2 Beta-adrenergic receptors

The βAR-adrenergic receptors (βAR) are members of the Rhodopsin family (R) of GPCRs. They consist of three subtypes: β_1-adrenergic receptor (β_1AR), β_2-adrenergic receptor (β_2AR) and the β_3-adrenergic receptor (β_3AR). β_1AR is predominantly expressed in cardiac tissue where it plays a role in heart function (Figure 6.2). The receptor is endogenously targeted by the catecholamines adrenaline and noradrenaline. This invokes the HGP signalling cycle mediated by $G\alpha_s$ which results in phosphorylation of calcium-channel regulatory proteins by protein kinase A (PKA)

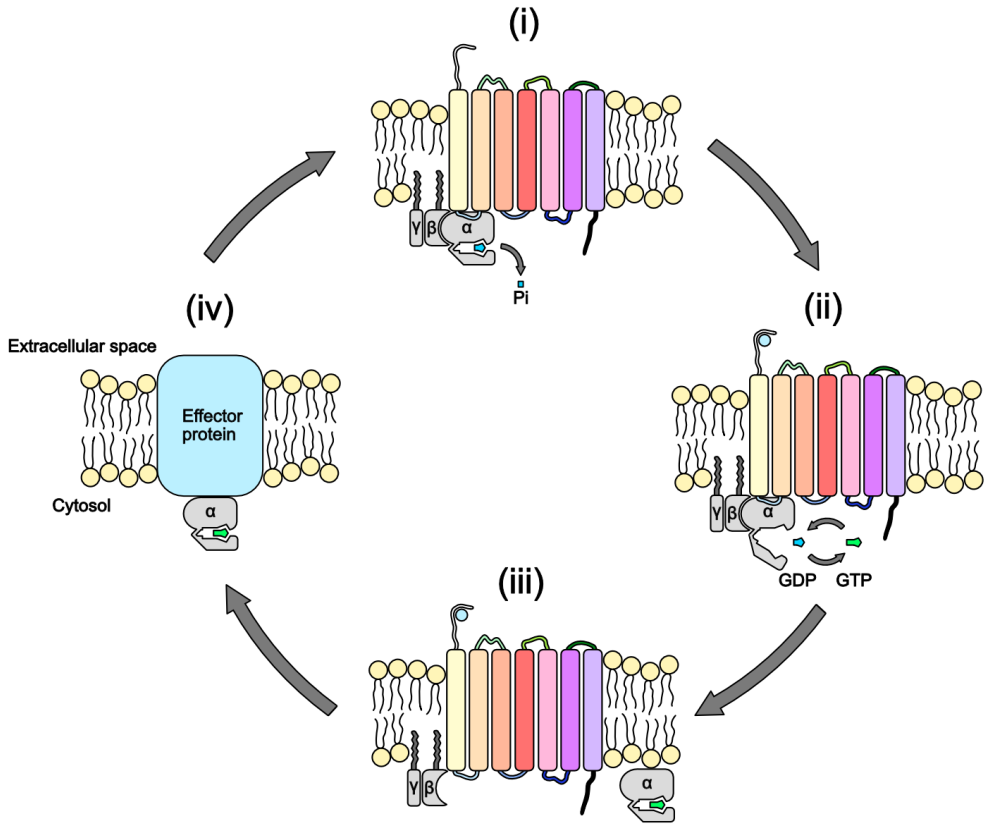


Figure 6.1 | **Heterotrimeric G-protein signalling cycle.** (i) In the basal state, the HGP (formed of Gα and G$\beta\gamma$) associates with the GPCR. (ii) GPCR activation leads to binding of GTP to Gα. (iii) Gα dissociates from G$\beta\gamma$. (iv) Gα and G$\beta\gamma$ (not shown) can interact with downstream effector proteins to mediate physiological changes.

and an influx of calcium into the cytosol from the sarcoplasmic reticulum.[169] Increased levels of intracellular calcium cause myocardial contraction. Constitutive $G\alpha_s$ signalling has been associated with heart failure. β_1AR antagonists and inverse agonists have been developed, including the β-blockers which are used to treat myocardial infarction and increased hypertension. Furthermore, previous research has showed β_1AR/arrestin-mediated activation of the EGFR signalling cascade in the context of catecholamine-based cardioprotection, highlighting the potential of targeting non-classical signalling pathways initiated by β_1AR in the treatment of heart disease.[177]

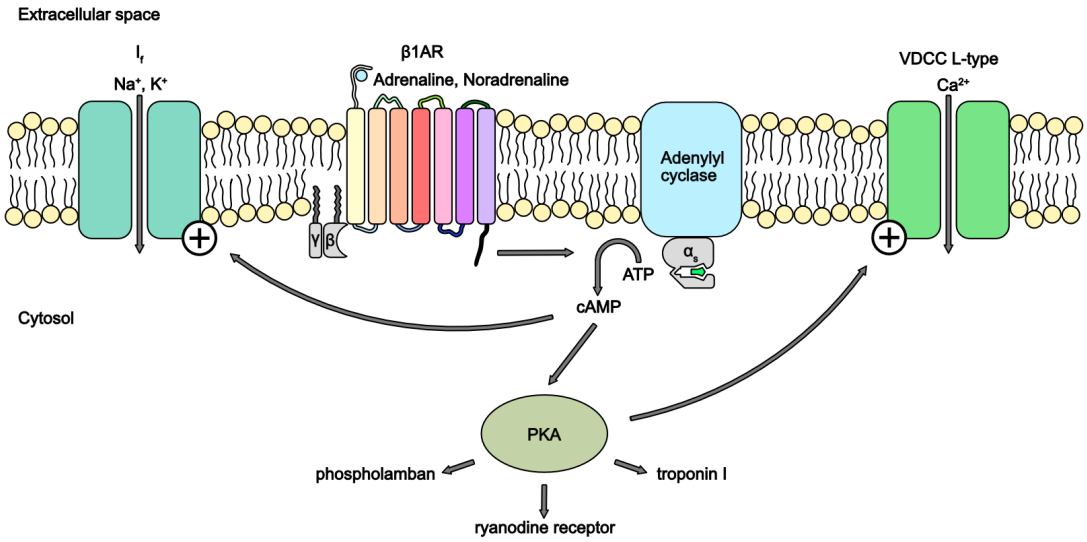


Figure 6.2 | β_1AR signalling pathway in cardiac cells. Pacemaker channel (I_f), voltage-dependent calcium channel (VDCC), PKA.

6.1.3 Insight from beta-1 adrenergic receptor crystal structures

The study of human β_1AR has been difficult owing to low conformational thermostability in detergent. In contrast, turkey β_1AR exhibits a 10 °C higher melting temperature (T_m) than its human counterpart. Turkey β_1AR also shares 82 % sequence similarity with human β_1AR and 67 % sequence similarity with human β_2AR. The turkey variant is therefore an attractive alternative for the structural study of β_1AR. Alanine-scanning has been used to further enhance the thermostability of turkey β_1AR. Single mutations that individually increased the T_m of the receptor were randomly combined to identify the most thermostable turkey β_1AR.[178] β_1AR-m23 displayed a

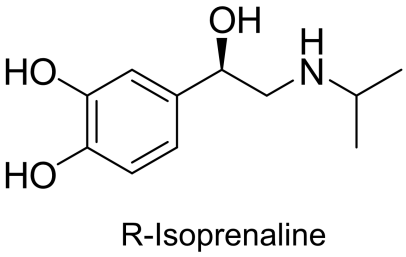
21 °C higher T_m than the wildtype turkey receptor. Subsequent incorporation of these mutations into human β_1AR led to an increase of the T_m by 17 °C compared to native human β_1AR.[179] Warne and colleagues collected the first crystal structure of turkey β_1AR using the m23 variant.[180] They showed that transmembrane domain topology between β_1AR and β_2AR was highly similar with superimposed C_α atoms displaying an RMSD of 0.25 Å.[180] The ECLs were also comparable but ECL2 of β_1AR was shown to have partial helical structure and stabilised by Na^+.[180,181] Structural similarity between β_1AR and β_2AR was also observed on ICL1; however, ICL2 on β_1AR exhibited a slight helical structure but this was suggested to have arisen as an artifact due to exogenous stabilising agents.[180]

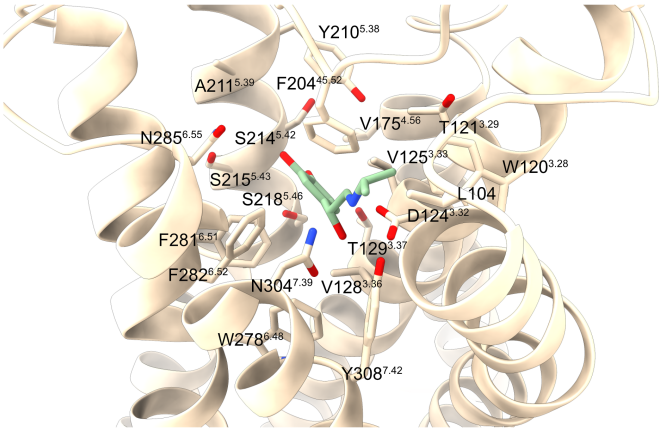
Several structures of β_1AR bound to agonists, inverse agonists and antagonists have since been released. Some of these have included ternary complexes bound to G protein mimics, including nanobodies (Nbs) and more recently, G proteins and β-arrestins. Structural comparison between these states has provided valuable insight into the conformational rearrangements associated with β_1AR activation. Nbs have been particularly important in highlighting the conformational range of GPCRs. These antibody fragments, comprising a monomeric variable antibody domain, are being increasingly employed as chaperones to preserve transient protein states. A range of Nbs have been raised (mainly from immunisation of Camelids against a desired antigen, and subsequent isolation of mRNA that encodes the variable domain of the heavy chain) which stabilise various activation states of βARs.[182]

Structures of β_1AR bound to the full agonists, isoprenaline and carmoterol, and partial agonists, salbutamol and dobutamine showed almost identical molecular networks in the catecholamine binding pocket despite having distinctly different pharmacological profiles (Figure 6.3).[183] *In this chapter β_1AR residue numbering is based on the employed construct sequence; however, the Ballesteros-Weinstein numbering scheme is also indicated.*[184] *This consists of two numbers where the first designates the helix and the second, the residue position relative to the most conserved amino acid which is defined as 50. For example,* $^{3.41}$ *indicates a residue on TM3, nine residues before the most conserved residue.* These structures revealed that the core interactions included hydrogen bond formation between Asp124$^{3.32}$ and a secondary amine, Asn304$^{7.39}$ and a β-hydroxyl group and Ser214$^{5.42}$ and Asn285$^{6.55}$ to the catechol moiety. Several subtle structural differences

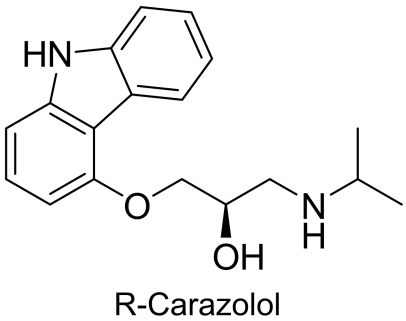
were observed between agonists and partial agonists. Most notably, agonists were seen to induce a rotamer change in Ser215$^{5.43}$ allowing the residue to hydrogen bond to Asn285$^{6.55}$. This rotamer change also enabled hydrogen bond formation to Ser218$^{5.46}$. The authors observed a ~1 Å contraction of the binding cavity between TM5 and TM7 in agonist-bound structures and suggested that the overall strengthening of these contacts was responsible for increasing the efficacy of a ligand. Warne and colleagues recently crystallised several ternary β_1AR-Nb complexes.[185] Nb80 which displays G protein like characteristics and stabilises the activated β_1AR state was shown to bind to the intracellular G protein binding cavity.[185] The CDR3 loop of Nb80 adopted a hairpin conformation enabling it (and to a lesser extent CDR1 and CDR2 loops) to bind through hydrophobic interactions. Active β_1AR-Nb80 structures revealed that agonists induced a greater reduction to the internal volume of the extracellular ligand binding cavity compared to antagonists. This was again correlated to an increase in the number of hydrogen bond interactions and strength of contacts.[185] The ternary complexes also highlighted the prototypical molecular switch of GPCR activation in which TM6 undergoes a 14 Å movement away from the helical bundle to accommodate an intracellular binding partner.[185] Warne and colleagues had previously proposed that the concomitant strengthening of the TM5-TM6 interface and weakening of the TM4-TM5 interface by agonist-induced rotameric changes was crucial for transition to an activated state.[180] However crystallisation of activated βAR in the presence of an orthosteric agonist alone has not yet been possible.[182] β_1AR-Nb80 complexes also highlighted several other molecular switches, including the ionic lock and water lock. The ionic lock serves to maintain the receptor in an inactivated state and is mediated by a salt bridge between Arg139$^{3.50}$ of the conserved motif 'DRY' and Glu260$^{6.30}$. Upon activation, Warne and colleagues observed that the 'ionic-lock' had broken, facilitating movement of TM6.[185] Activated structures were also observed to contain the water-lock; where a hydrogen bond formed between Tyr318$^{7.53}$ on the conserved motif NPxxY and Tyr230$^{5.58}$ coordinated by a water molecule, served to stabilise the receptor's active structure.[185] Su and colleagues recently resolved the Gs bound β_1AR structure using cryo-EM and observed that changes in β_1AR conformation were consistent with previous active ternary β_1AR-Nb complexes.[186]

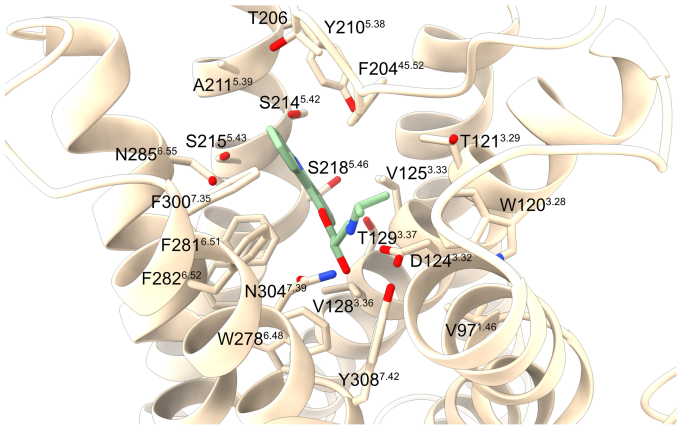
Nb60 is another recently discovered antibody fragment which stabilises the inactive β_2AR conformation.[187] Unliganded β_2AR has been shown to occupy two inactive states termed S1 and S2.[187] S1 represents an ionic-lock bound state and S2 represents an ionic-lock broken state. Binding of the inverse agonist carazolol (Figure 6.3) to β_2AR has been shown to stabilise the S1 state whereas the ternary β_2AR-Nb60 complex has been shown to populate the S2 state.[187] Interaction between these proteins occurs through disruption of the ionic lock and formation of a salt bridge between Arg142$^{3.50}$ of the DRY motif and residues on the Nb60 CDR3 loop.[187] Other than these differences, structures of β_2AR bound to carazolol, and carazolol and Nb60 are highly similar and display an RSMD of 0.3 Å.





(a) Structure of R-isoprenaline and binding of the full agonist (light green) to β_1AR (tan).





(b) Structure of R-carazolol and binding of the partial inverse agonist (light green) to β_1AR (tan).

Figure 6.3 | Structures of R-isoprenaline and R-carazolol and their binding contacts with β_1AR.

Here, I report the use of carbene footprinting to accurately map the interactions with activated and inactivated β_1AR-Nb ternary complexes for the first time. The use of MS/MS allowed direct protein-ligand and protein-protein interactions as well as binding partner-induced structural protein changes to be mapped to resolution approaching the amino acid level. The agreement between these data and those afforded by previous studies is remarkable and demonstrates that carbene footprinting can provide high quality structural information for this important and challenging class of protein.

6.2 Results and discussion

6.2.1 Beta-1 adrenergic receptor purification

I purified thermostabilised β_1AR (Figure 6.4) to produce protein samples for carbene labelling experiments. His-tagged β_1AR was successfully solubilised in DDM and purified using IMAC (Figure 6.5b). Purified β_1AR was observed to run at an apparent molecular mass of ~30 kDa on the SDS-PAGE gel despite being a ~40 kDa protein. This was anticipated as membrane proteins are well documented to migrate to lower molecular weights due to incomplete denaturation and reduced SDS interaction caused by associated detergents.[188]

6.2.2 Nanobody purification

I successfully expressed and purified Nbs to generate protein samples for carbene labelling experiments on active and inactive ternary β_1AR-Nb complexes (Figure 6.6).

6.2.3 Nanobody binding assays

A SEC-based co-elution assay was used to determine binding between β_1AR and Nbs. β_1AR was incubated with either isoprenaline and Nb80, or carazolol and Nb60 before samples were analysed by SEC (Figure 6.7a, Figure 6.8a). Eluent spanning the two largest SEC peaks were analysed by SDS-PAGE (Figure 6.7b, Figure 6.8b). In both cases co-elution of β_1AR with Nb80 and Nb60 was observed, indicating binding between

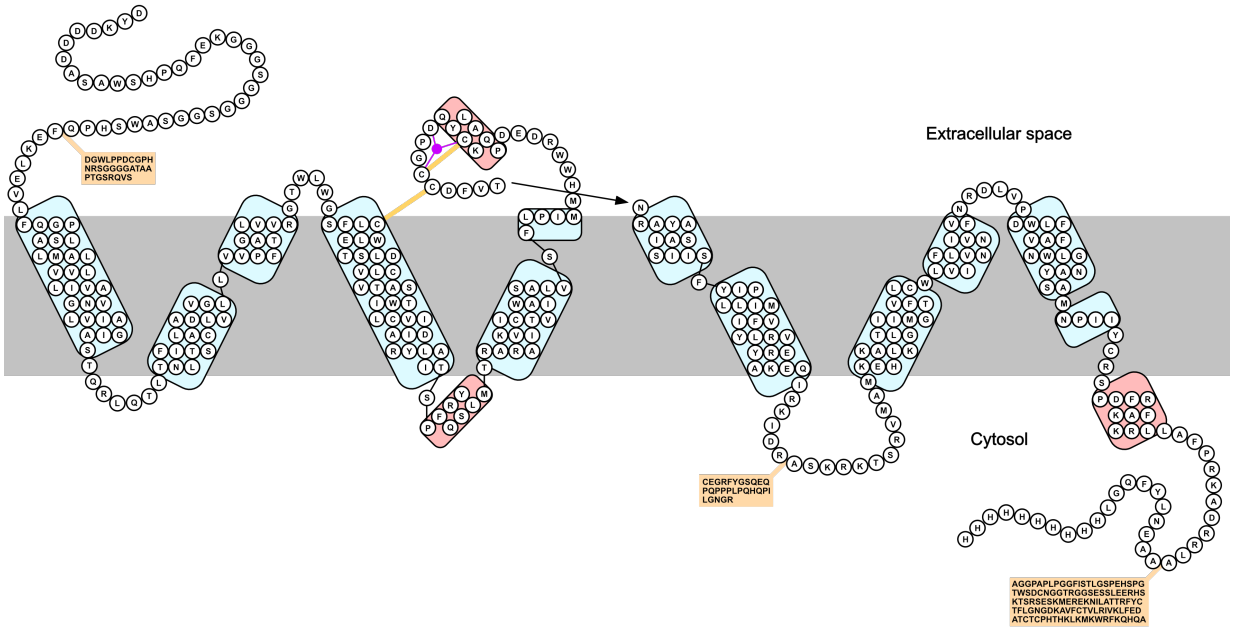
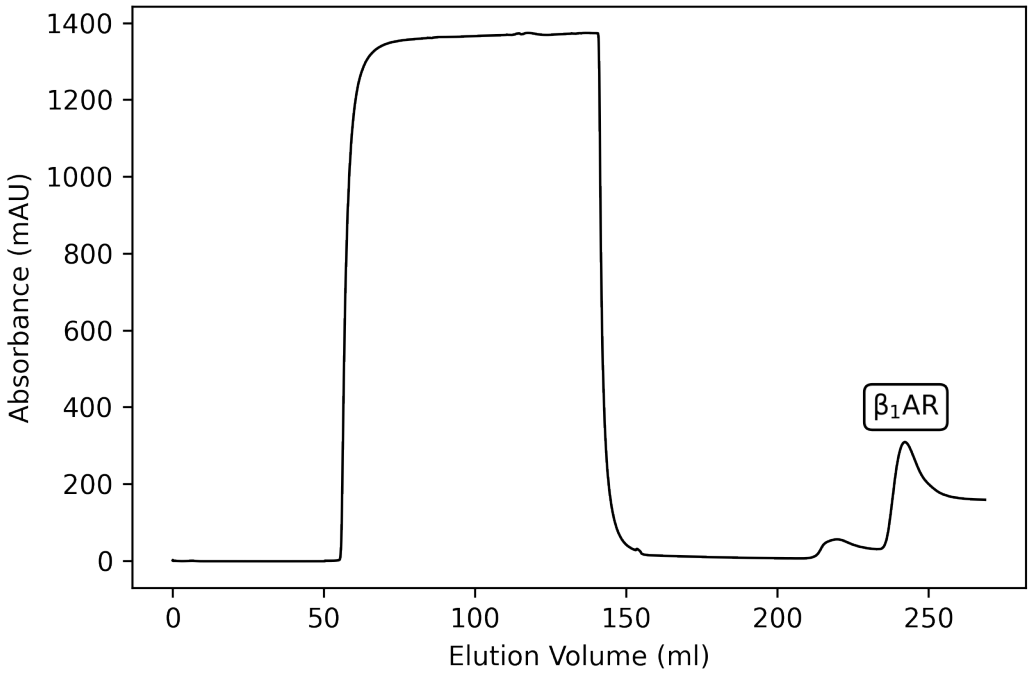
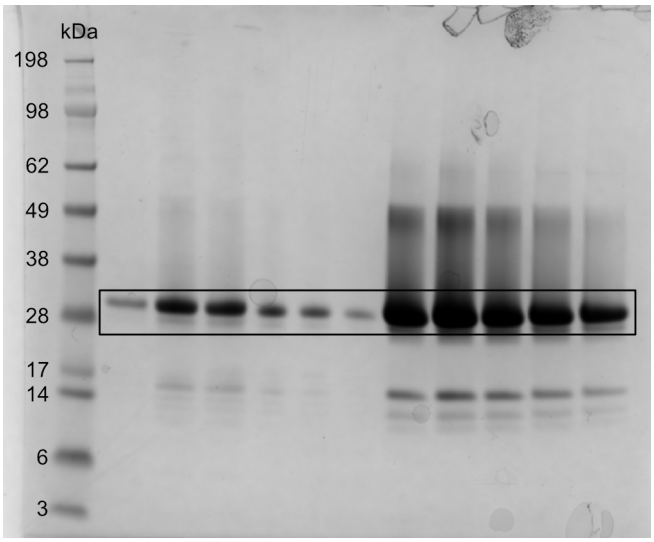


Figure 6.4 | β_1AR construct map. The β_1AR construct employed in this study contained several well documented thermostabilising mutations and deletions. The construct was generated by OMass Therapeutics. Transmembrane-spanning helices are indicated by blue rectangles whilst pink rectangles represent extracellular and intracellular helices. Deletions are indicated by yellow boxes.

the receptor and Nbs. The quantitative difference in band intensities between β_1AR and Nbs was attributed to the molecular weights of either protein (40 and 14 kDa, respectively) and the amount of available surface area for staining with coomassie. Free β_1AR and Nb was observed at either end of the SEC trace. Western blotting was also performed on fractions from the β_1AR-Nb80 co-elution fractions against an anti penta histidine antibody (Figure 6.7c). This further validated the presence of both proteins and highlighted a more accurate stoichiometric representation of β_1AR-Nb80 binding.

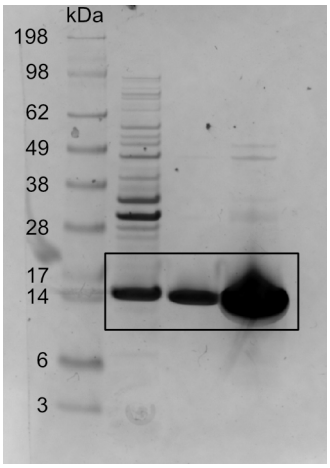


(a) β_1AR IMAC chromatogram. Membranes containing β_1AR were solubilised in DDM and isolated using IMAC. The peak corresponding to elution of β_1AR is indicated. The high UV observed between ~55–150 min was attributed to the elution of other (non-His-tagged) membrane proteins.

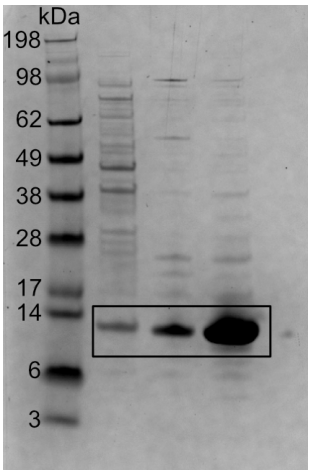


(b) β_1AR purification gel. Fractions from the IMAC purification of DDM-solubilised β_1AR were analysed by SDS-PAGE. Bands corresponding to β_1AR are indicated.

Figure 6.5 | β_1AR purification.

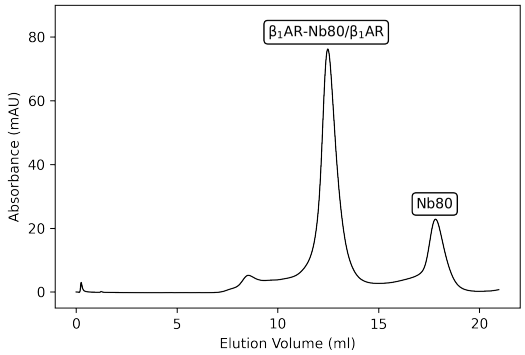


(a) Nb80 purification.

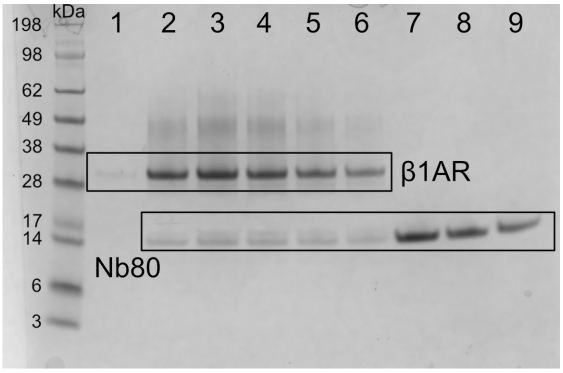


(b) Nb60 purification.

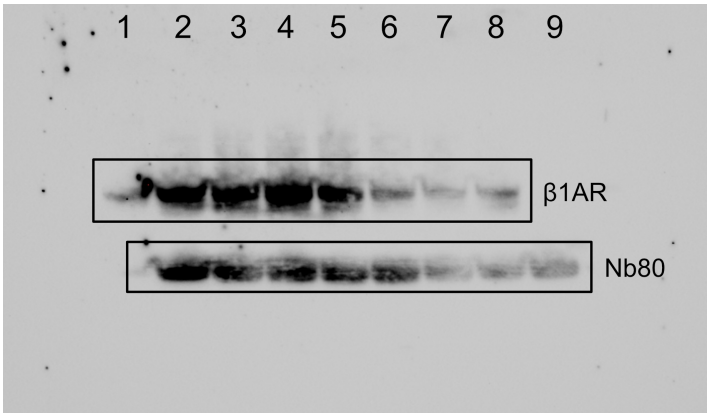
Figure 6.6 | Nanobody purification. SDS-PAGE of Nb80 and Nb60 purifications. Bands corresponding to either protein are indicated. In each case, lanes correspond to the molecular weight ladder, flow-through, wash and eluent, respectively.



(a) β_1AR-Nb80 co-elution SEC chromatogram. β_1AR was incubated with isoprenaline and Nb80 on ice for 1 h before analysing with SEC. Peaks corresponding to the β_1AR-Nb80 complex or either protein are indicated.

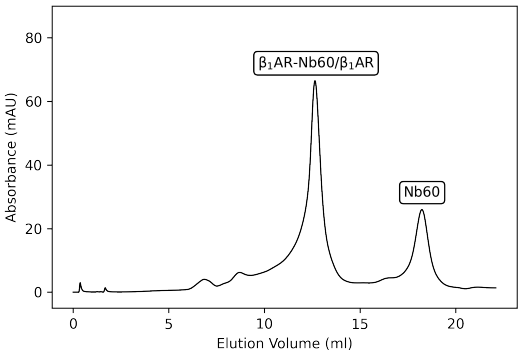


(b) β_1AR-Nb80 co-elution SDS-PAGE. Fractions spanning two indicated peaks were analysed by SDS-PAGE (lanes 1-9). Co-elution of β_1AR with Nb80 was observed, indicating interaction of the proteins.

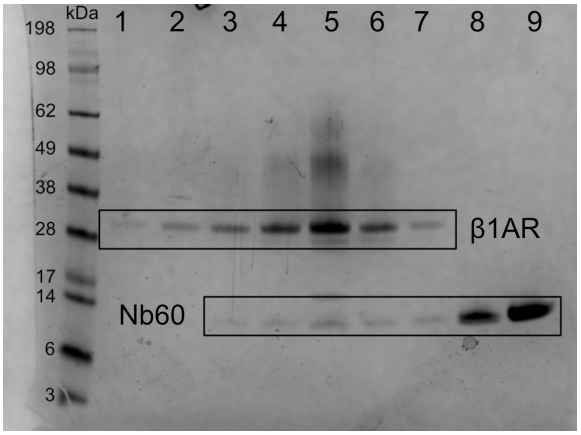


(c) β_1AR-Nb80 co-elution western blot Fractions spanning two indicated peaks (lanes 1-9) were analysed by western blotting against anti-His antibodies. Co-elution of β_1AR with Nb80 was again observed, reinforcing interaction of the proteins.

Figure 6.7 | SEC co-elution of β_1AR with isoprenaline and Nb80.



(a) β_1AR-Nb60 co-elution SEC chromatogram. β_1AR was incubated with carazolol and Nb60 on ice for 1 h before analysing with SEC. Peaks corresponding to the β_1AR-Nb60 complex or either protein are indicated.



(b) β_1AR-Nb60 co-elution SDS-PAGE. Fractions spanning two indicated peaks were analysed by SDS-PAGE (lanes 1-9). Co-elution of β_1AR with Nb60 was observed, indicating interaction of the proteins.

Figure 6.8 | SEC co-elution of β_1AR with carazolol and Nb60.

6.2.4 Optimisation of footprinting conditions

6.2.4.1 Optimisation of digestion conditions

Proteolytic digestion is fundamental to bottom-up proteomic analysis and the primary determinant in the peptides observed by LC-MS. The choice of proteolytic conditions was therefore an important consideration in carbene footprinting of β_1AR. *In-silico* digestions were performed on the β_1AR, Nb80 and Nb60 sequences using the PeptideCutter server to determine the most effective protease(s) that maximised peptide sequence coverage.[116] Chymotrypsin, trypsin and pepsin were used to theoretically digest β_1AR whilst chymotrypsin and trypsin were used to cleave Nb80 and Nb60. Peptides were filtered to include a length of between 4 and 40 residues and a maximum of three missed cleavages.

The β_1AR *in-silico* digest displayed 100 % sequence coverage with chymotrypsin, trypsin and pepsin (Figure 6.9); however, the lack of tryptic cleavage sites between residues 39-142 resulted in single peptide coverage of this region of β_1AR (i.e. no missed cleavages).

In-silico digestion of Nb80 revealed 100 % sequence coverage with chymotrypsin and trypsin (Figure 6.10). The absence of tryptic cleavage sites led to only single peptides covering N- and C-terminal regions of the protein.

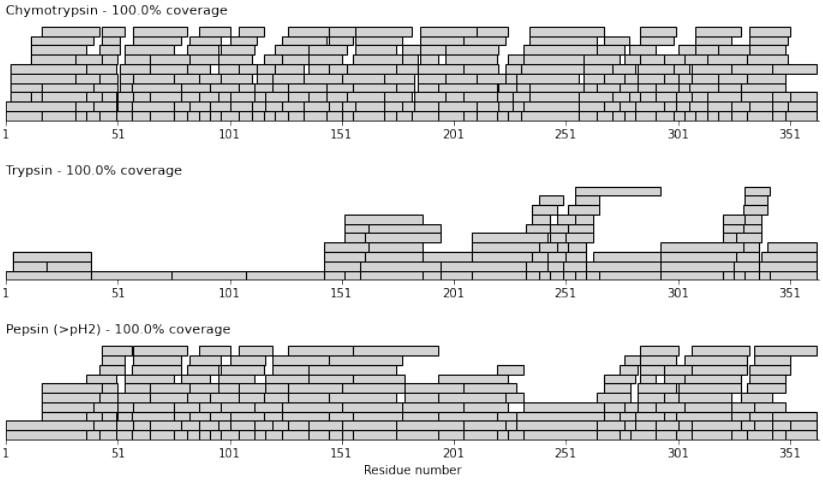


Figure 6.9 | ***In-silico* digest of β_1AR with chymotrypsin, trypsin and pepsin.** Peptides were filtered between 4-40 residues with a maximum of three missed cleavages. Grey bars represent predicted peptides and overlapping bars indicate missed cleavages.

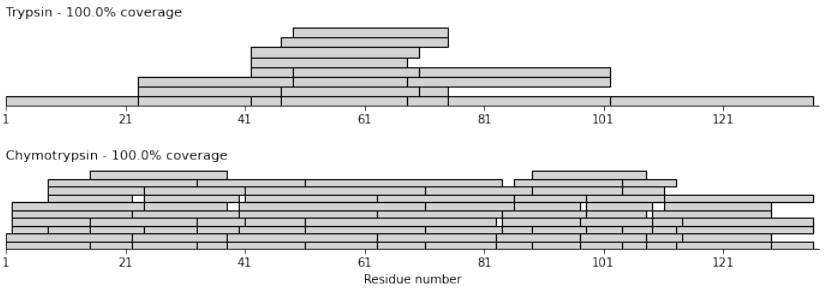


Figure 6.10 | ***In-silico* digest of Nb80 with chymotrypsin and trypsin.** Peptides were filtered between 4-40 residues with a maximum of three missed cleavages. Grey bars represent predicted peptides and overlapping bars indicate missed cleavages.

The Nb60 *in-silico* digest also displayed 100 % sequence coverage with chymotrypsin and trypsin (Figure 6.11). Similarly to the Nb80 *in-silico* digest, single peptides covered N- and C-terminal regions of the sequence, suggesting that for both proteins chymotrypsin may be the more appropriate protease to maximise sequence coverage.

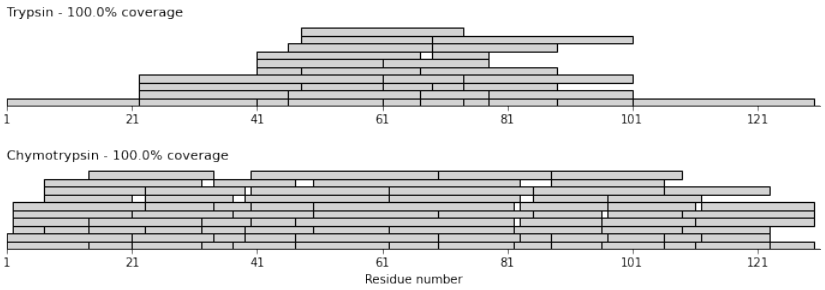


Figure 6.11 | *In-silico* digest of Nb60 with chymotrypsin and trypsin. Peptides were filtered between 4-40 residues with a maximum of three missed cleavages. Grey bars represent predicted peptides and overlapping bars indicate missed cleavages.

After performing *in-silico* digestions on β_1AR, Nb80 and Nb60, experimental sequence coverage was determined using LC/MS operated in DDA mode. β_1AR was digested separately with chymotrypsin, trypsin and pepsin for either 1 h or overnight as well as with and without and ProteaseMAX. This surfactant was previously identified to provide increased sequence coverage for PfMATE (Section 5.2.3.1). Nb80 and Nb60 were digested overnight with chymotrypsin and trypsin.

The overnight chymotrypsin digestion utilising ProteaseMAX generated the highest sequence coverage (~66 %) of all proteolytic conditions (Figure 6.12). The absence of the surfactant showed a ~22 % decrease in sequence coverage compared to the overnight chymotrypsin digest with the surfactant, in agreement with previous work that the surfactant may aid bottom-up proteolytic analysis of membrane proteins. The 1 h digestion with ProteaseMAX similarly led to a ~15 % increase in peptide coverage compared to the 1 h digestion without the surfactant.

Tryptic digests of β_1AR revealed that a maximum sequence coverage of ~32 % was obtained when digested overnight without ProteaseMAX. The large discrepancy between theoretical and experimental sequence coverage was presumably due to the inability to extract peptides out of the gel pieces and/or LC column. 1 h digests with

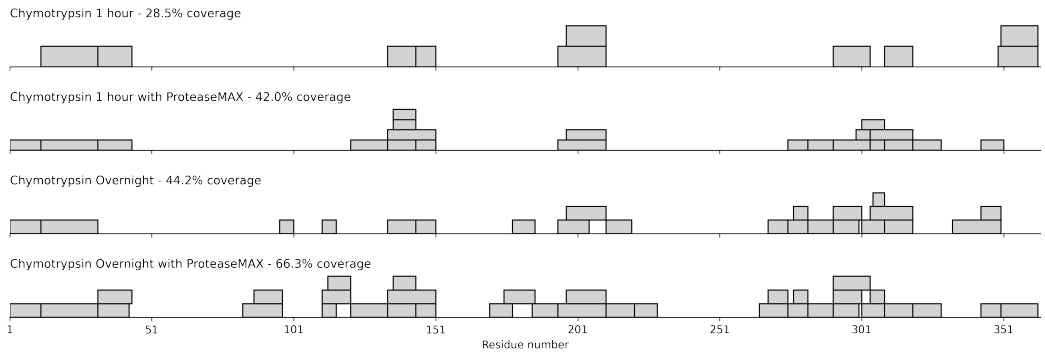


Figure 6.12 | Sequence coverage achieved by DDA LC-MS/MS analysis of β_1AR following digestion with chymotrypsin for 1 h and overnight time points, with and without ProteaseMAX. Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

and without ProteaseMAX displayed a near identical set of peptides (Figure 6.13) where no increase in sequence coverage was observed with the introduction of ProteaseMAX on the 1 h digestion.

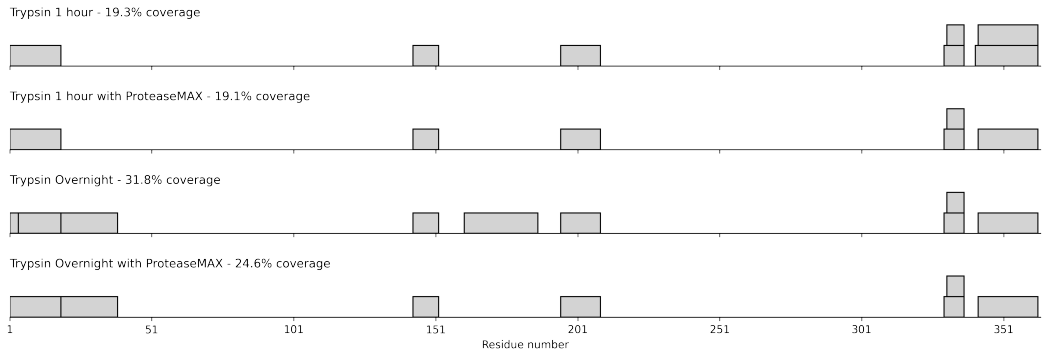


Figure 6.13 | Sequence coverage achieved by DDA LC-MS/MS analysis of β_1AR following digestion with trypsin for 1 h and overnight time points, with and without ProteaseMAX. Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

The overnight β_1AR peptic digest without ProteaseMAX led to the highest level of sequence coverage (Figure 6.14). The 1 h digestion benefited from the introduction of ProteaseMAX which showed a two-fold increase in sequence coverage compared to the 1 h digestion without the surfactant. This was unexpected given the surfactant's acid-labile nature.

The overnight Nb80 tryptic digest led to a higher sequence coverage compared to the overnight chymotryptic digest (Figure 6.15).

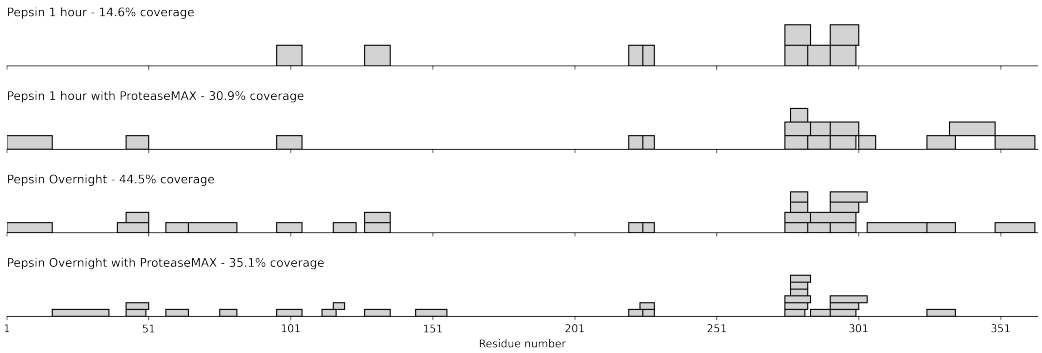


Figure 6.14 | Sequence coverage achieved by DDA LC-MS/MS analysis of β_1AR following digestion with pepsin for 1 h and overnight time points, with and without ProteaseMAX. Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

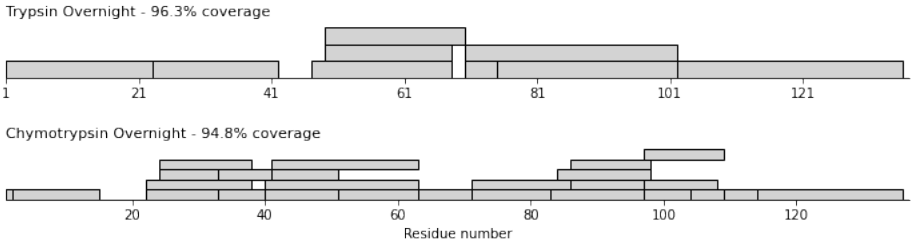


Figure 6.15 | Sequence coverage achieved by DDA LC-MS/MS analysis of Nb80 following digestion with trypsin and chymotrypsin for overnight. Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

In contrast to the Nb80 digests, the overnight Nb60 chymotryptic digest led to a higher peptide sequence coverage compared to the overnight tryptic digest (Figure 6.16).

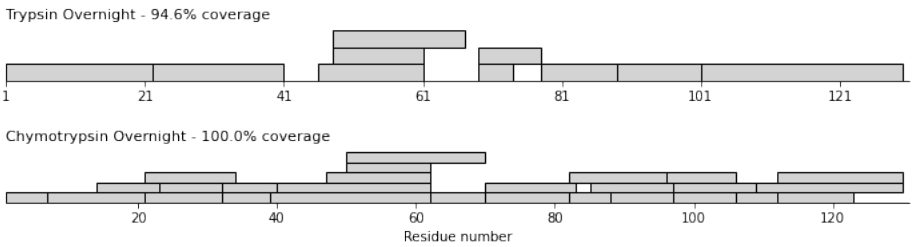


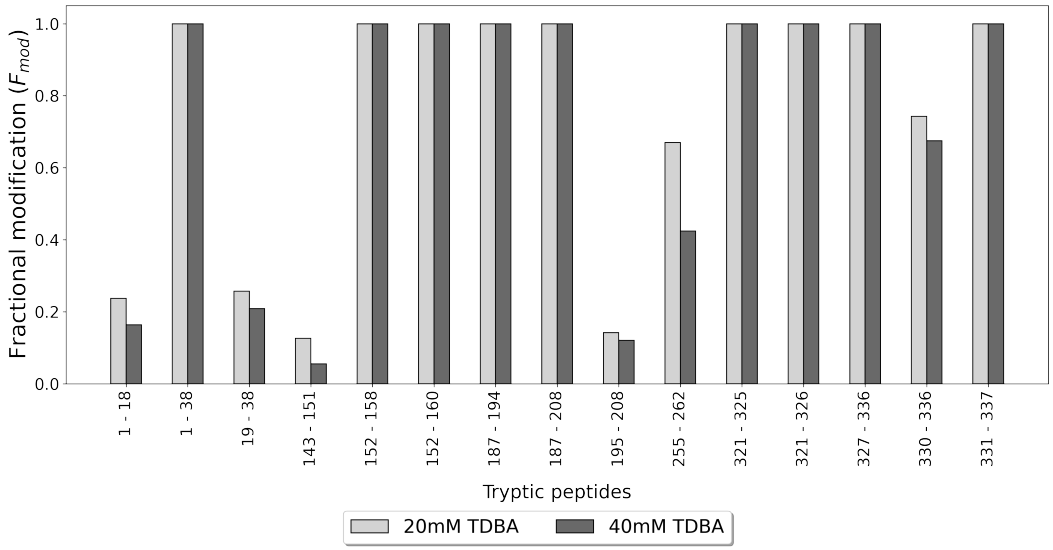
Figure 6.16 | Sequence coverage achieved by DDA LC-MS/MS analysis of Nb60 following digestion with trypsin and chymotrypsin for overnight. Grey bars represent DDA-identified peptides and overlapping bars indicate missed cleavages.

6.2.4.2 Optimisation of carbene labelling conditions

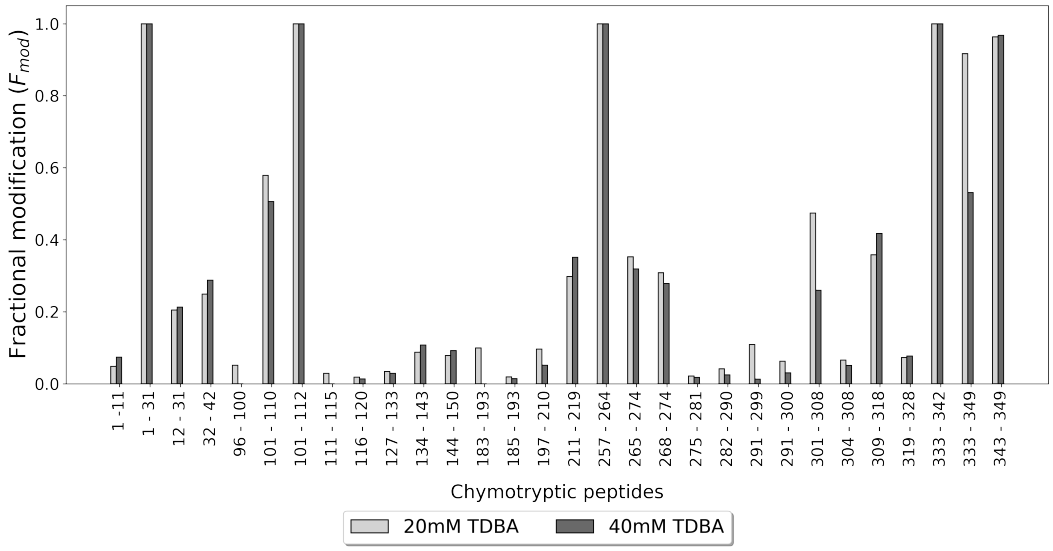
Carbene footprinting experiments were performed on singlet (N=1) β_1AR, Nb80 and Nb60 samples to determine which diazirine concentration was optimal for differential labelling studies. Labelling of β_1AR was performed at 20 and 40 mM NaTDB whilst the Nbs were modified at 10 and 20 mM NaTDB. It was envisioned that a higher NaTDB concentration would be needed than typical concentrations used in soluble protein labelling (~10–20 mM) to overcome label quenching by detergent micelles and the more compact protein structure as observed with chemical modification of PfMATE (Section 5.2.3.1).

The tryptic digest of labelled β_1AR corresponded to 28.5 % modified sequence coverage. Labelled peptides showed a reduction in F_{mod} from 20 and 40 mM NaTDB (Figure 6.17a). Interestingly, PfMATE showed higher labelling with increased NaTDB concentration. Nevertheless, the cause of the labelling reduction with higher NaTDB concentration remained unknown. Several peptides, including 1-38 (DYKDDDDASAWSHPQFEKGGGSGGGSGGSAWSHPQFEK), 152-158 (YQSLMTR), 152-160 (YQSLMTRAR), 187-194 (DEDPQALK), 187-208 (DEDPQALKCYQDPGCCDFVTNR), 321-325 (SPDFR), 321-326 (SPDFRK), 327-336 (AFKRLLAFPR) and 331-337 (LLAFPRK) displayed 100 % modification at either probe concentration. Despite promising levels of carbene labelling, the trypsin digestion produced poor sequence coverage. The chymotryptic digest of labelled

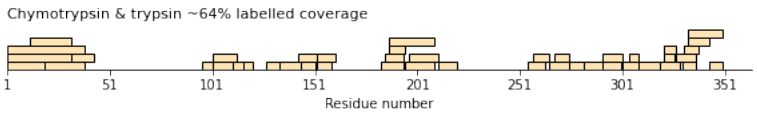
β_1AR corresponded to 59 % sequence coverage. The digest of labelled β_1AR revealed mixed labelling results from 20 and 40 mM NaTDB with 20 mM NaTDB generally showing increased peptide-level labelling compared to 40 mM in line with the tryptic digest results (Figure 6.17b). Four peptides, including 1-31 (DYKDDDDASAWSHPQFEKGGGSGGGSGGSAW), 101-112 (GATLVVRGTWLW), 257-264 (AMKEHKAL) and 333-342 (AFPRKADRRL), displayed 100 % carbene modification at 20 and 40 mM diazirine. The pooled peptides from separate tryptic and chymotryptic digests corresponded to 64 % labelled sequence coverage (Figure 6.17c). This peptide coverage was greater than what West and colleagues observed in their HDX analysis of β_2AR.[82] The labelling optimisation results suggested that 20 mM NaTDB in combination with tryptic and chymotryptic digests provided the greatest carbene modification with complementary sequence coverage.



(a) Carbene labelling optimisation of β_1AR with a trypsin digestion. The fractional modification of each tryptic β_1AR peptide is shown at 20 and 40 mM NaTDB.



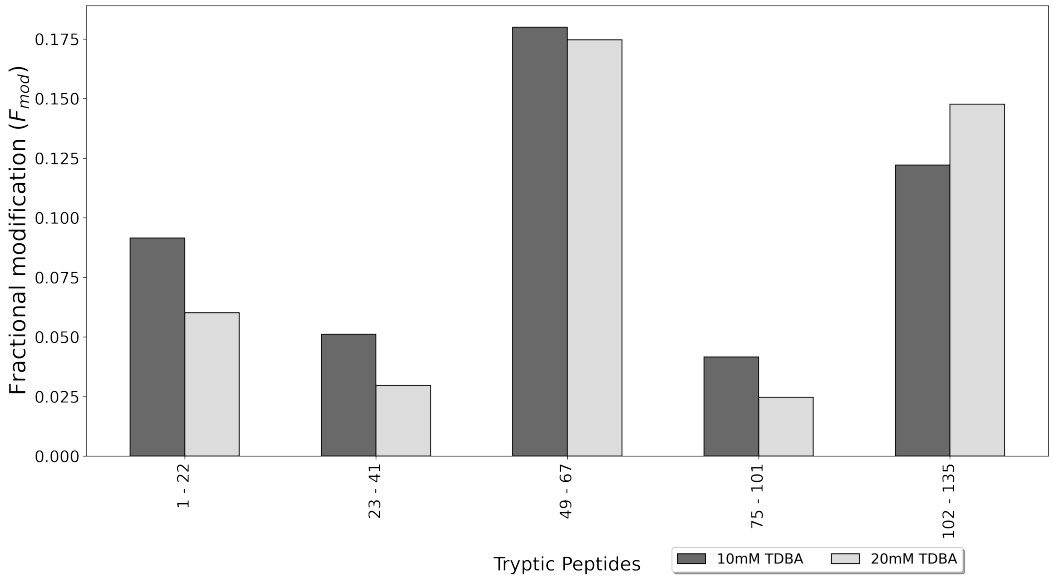
(b) Carbene labelling optimisation of β_1AR with a chymotrypsin digestion. The fractional modification of each chymotryptic β_1AR peptide is shown at 20 and 40 mM NaTDB.



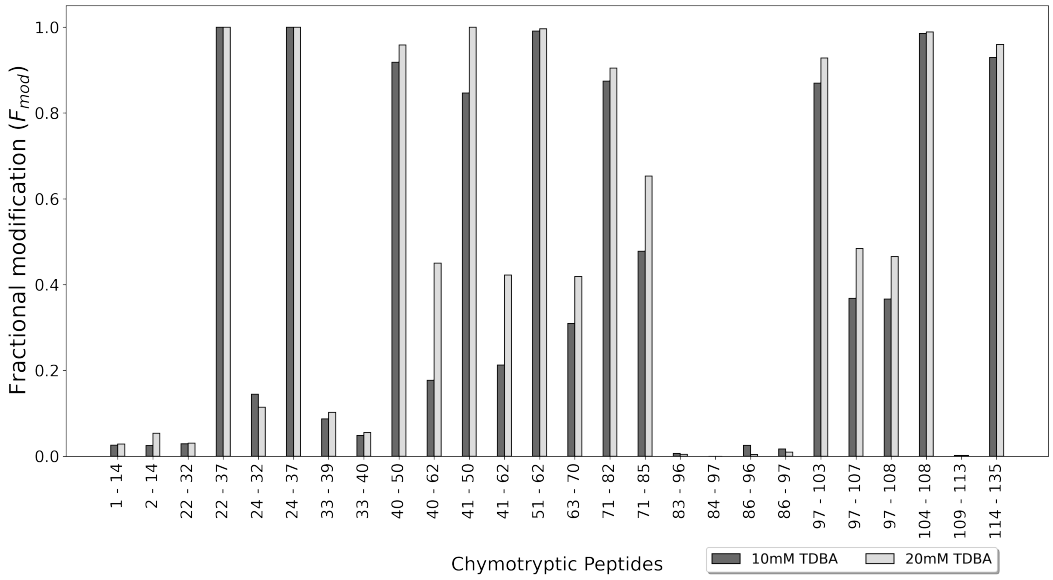
(c) Sequence coverage plot of concatenated carbene labelled β_1AR chymotryptic and tryptic peptides showing overall labelling coverage. Orange bars represent observed peptides and overlapping bars represent missed cleavages.

Figure 6.17 | Carbene labelling optimisation of β_1AR.

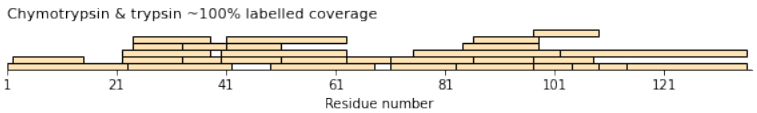
The tryptic digest of labelled Nb80 exhibited a total of 90 % sequence coverage. Peptides displayed no clear labelling trend from 10 and 20 mM NaTDB with certain peptides displaying higher modification at either 10 mM or 20 mM diazirine (Figure 6.18a). The labelled chymotryptic digest exhibited 95 % sequence coverage. Labelled chymotryptic Nb80 peptides showed a pronounced increase in f_{mod} from 10 and 20 mM NaTDB compared to the tryptic digest (Figure 6.18b). No labelling was observed on peptide 84-97 (QMNSLKPEDTAVYY) whilst peptides 22-37 (RLSCAASGSIFSINTM) and 24-37 (SCAASGSIFSINTM) displayed 100 % labelling at either diazirine concentration. The dual-digest (wherein peptides from the two separate digests were combined) corresponded to 100 % labelled sequence coverage (Figure 6.18).



(a) Carbene labelling optimisation of Nb80 with a trypsin digestion. The fractional modification of each tryptic Nb80 peptide is shown at 10 and 20 mM NaTDB.



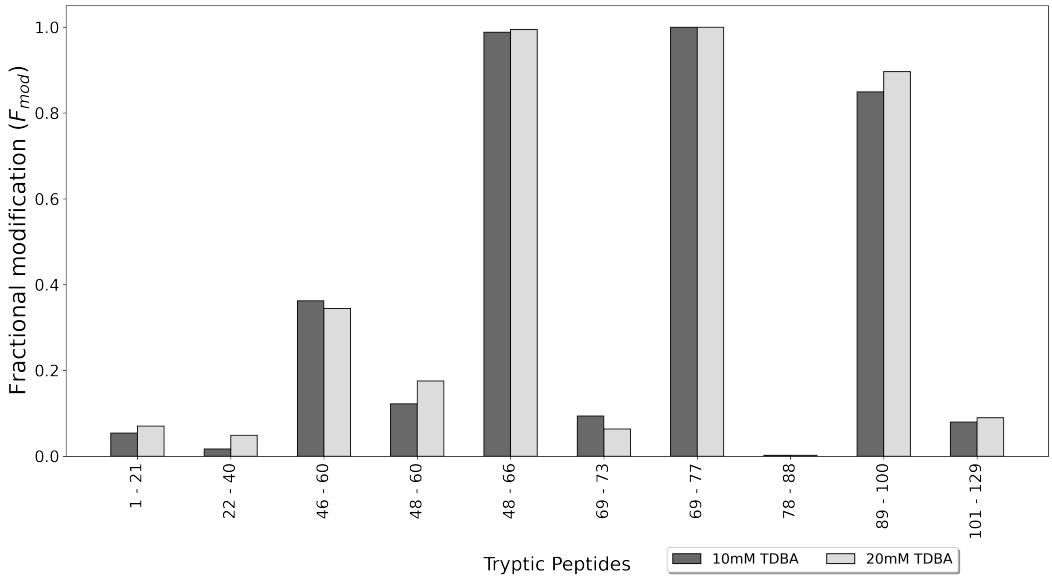
(b) Carbene labelling optimisation of Nb80 with a chymotrypsin digestion. The fractional modification of each chymotryptic Nb80 peptide is shown at 10 and 20 mM NaTDB.



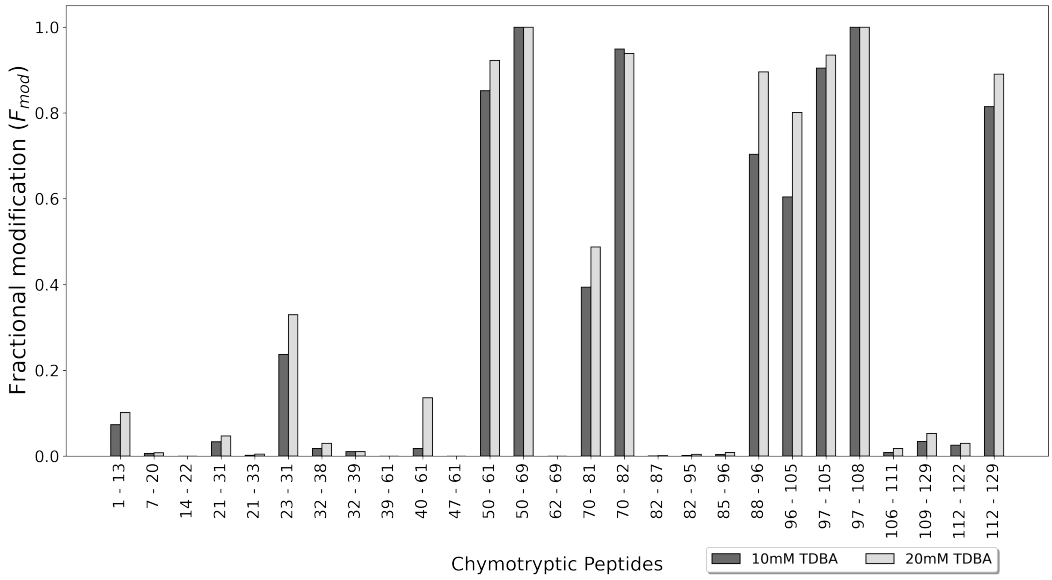
(c) Sequence coverage plot of concatenated carbene labelled Nb80 chymotryptic and tryptic peptides showing overall labelling coverage. Orange bars represent observed peptides and overlapping bars represent missed cleavages.

Figure 6.18 | Carbene labelling optimisation of Nb80.

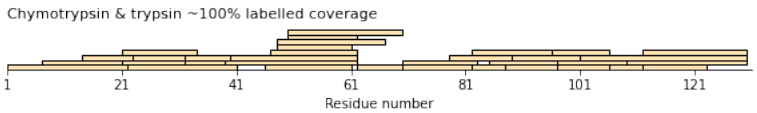
The tryptic digest of labelled Nb60 displayed 88 % sequence coverage. The digest showed a slight trend from 10 and 20 mM NaTDB (Figure 6.19a). Peptide 46-60 (LRELVAAITSGGSTK) displayed higher modification at 10 mM whilst the remaining peptides showed increased modification at 20 mM diazirine. The labelled chymotryptic digest exhibited 100 % sequence coverage. These peptides also showed a clear increase in carbene modification from 10 and 20 mM (Figure 6.19b). No modification was observed on peptides 14-22 (VQAGGSLRL), 39-61 (YRQAPGKLRELVAAITSGGSTKY), 47-61 (RELVAAITSGGSTKY) or 62-69 (ADSVKGRF) at either NaTDB concentration. In contrast, peptides 50-69 (VAAITSGGSTKYADSVKGRF) and 97-108 (CNAKVAGTFSIY) exhibited 100 % labelling at either diazirine concentration. The dual-digest (wherein peptides from the two separate digests were combined) corresponded to 100 % labelled sequence coverage (Figure 6.19).



(a) Carbene labelling optimisation of Nb60 with a trypsin digestion. The fractional modification of each tryptic Nb80 peptide is shown at 10 and 20 mM NaTDB.



(b) Carbene labelling optimisation of Nb60 with a chymotrypsin digestion. The fractional modification of each chymotryptic Nb80 peptide is shown at 10 and 20 mM NaTDB.



(c) Sequence coverage plot of concatenated carbene labelled Nb60 chymotryptic and tryptic peptides showing overall labelling coverage. Orange bars represent observed peptides and overlapping bars represent missed cleavages.

Figure 6.19 | Carbene labelling optimisation of Nb80.

Protein	Optimum digestion conditions	Optimum labelling conditions
β_1AR	Chymotrypsin Overnight with ProteaseMAX	20 mM NaTDB
Nb80	Chymotrypsin Overnight	20 mM NaTDB
Nb60	Chymotrypsin Overnight	20 mM NaTDB

Table 6.1 | Summary of optimal digestion and carbene labelling conditions for β_1AR, Nb80 and Nb60.

6.2.5 Differential study of beta-1 adrenergic receptor

After optimising β_1AR sequence coverage and labelling conditions (Table 6.1), I sought to capture interaction and conformational changes of β_1AR in the presence of orthosteric ligands and intracellular binding partners using carbene footprinting.

6.2.5.1 Activation of beta-1 adrenergic receptor

Binding of isoprenaline and Nb80 to β_1AR are known to be required for full receptor activation (Figure 6.20). In addition to capturing interactions at extracellular and intracellular binding sites, I was interested to determine whether carbene footprinting could detect the conformational switches that accompanied receptor activation, including motion of TM6.[189] Binding of isoprenaline to β_1AR was not predicted to cause any large conformational changes in the receptor and masking was only anticipated at the orthosteric binding site.[190]

Enzymatic digestion of β_1AR was carried out separately with chymotrypsin and trypsin to maximise peptide coverage and insight into receptor interactions and dynamics. Carbene labelling was compared between unliganded and binding partner-treated protein. MS/MS was also performed on selected peptides which displayed carbene labelling changes compared to control samples; however, due to sample conservation (MS/MS experiments were run separately to peptide-level labelling experiments), intensity of certain labelled peptides and co-elution of ions of greater intensity, MS/MS was not performed on all peptides displaying labelling differences.

Reductions in carbene modification were observed on several β_1AR peptides in the isoprenaline-treated samples compared to control samples (Figure 6.21, Figure S7). Pleasingly when mapped to the β_1AR structure, masking effects localised to the extracellular binding cavity, displaying remarkable agreement with crystal

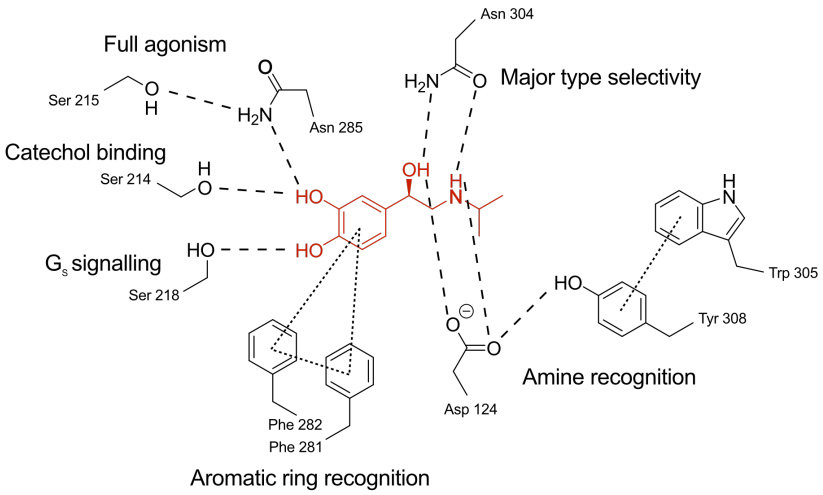
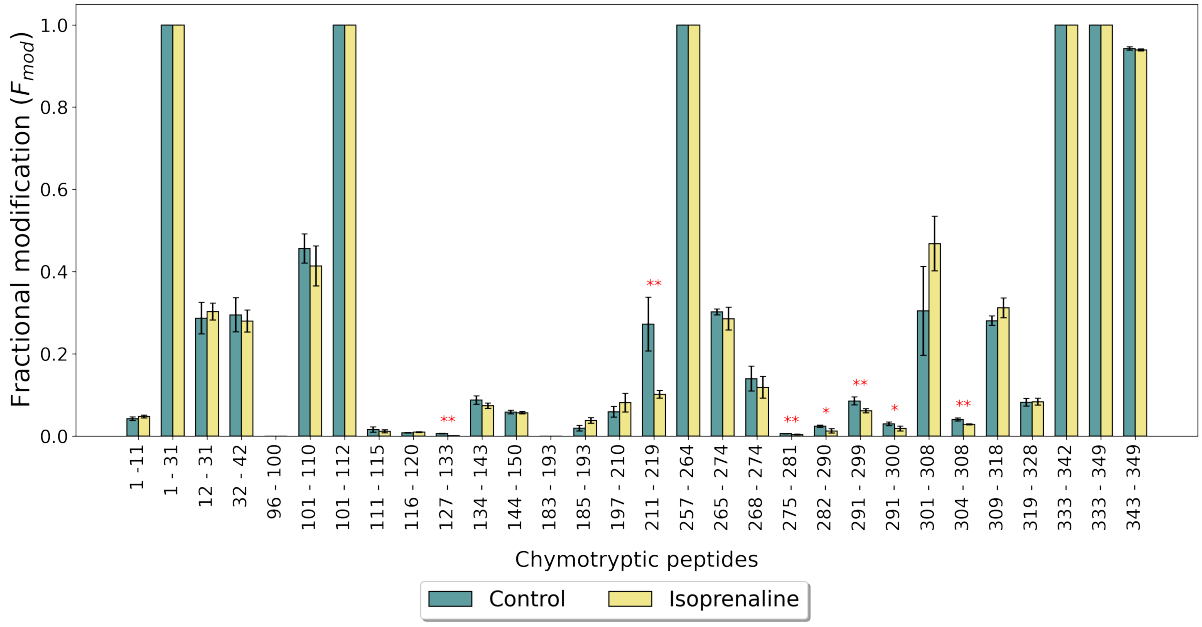


Figure 6.20 | Isoprenaline-β_1AR binding interactions.

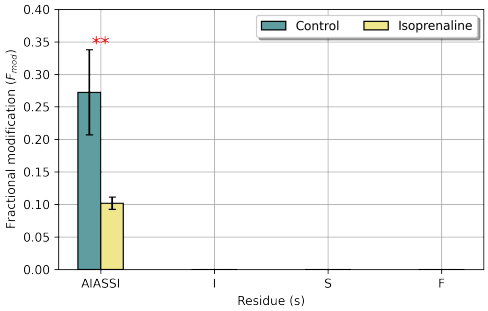
structures. Masked peptides included: $127^{3.35}$-$133^{3.41}$ (CVTASIW), $211^{5.39}$-$219^{5.47}$ (AIASSIISF), $275^{6.45}$-$281^{6.51}$ (TLCWLPF), $282^{6.52}$-$290^{6.60}$ (FLVNIVNVF), $291^{6.61}$-$299^{7.34}$ (NRDLVPDWL), $291^{6.61}$-$300^{7.35}$ (NRDLVPDWLF) and $304^{7.39}$-$308^{7.43}$ (NWLGY). Peptide 211-219 showed the greatest reduction in carbene labelling in the presence of isoprenaline compared to the control. This peptide contained Ser$214^{5.42}$, Ser$215^{5.43}$ and Ser$218^{5.46}$. Interaction of isoprenaline with Ser$214^{5.42}$ is known to induce rotamer changes in Ser$215^{5.43}$ facilitating several new interactions within the binding cavity, including hydrogen bond formation between itself and Asn$285^{6.55}$ but also between the catechol and Ser$218^{5.46}$. MS/MS revealed that carbene labelling was located on residues $211^{5.39}$-$216^{5.44}$ (AIASSI) whilst residues $217^{5.45}$-$219^{5.49}$ (ISF) showed no covalent modification (Figure 6.21b), further supporting the interaction of these residues with isoprenaline. Masking effects on peptides $275^{6.45}$-$281^{6.51}$ and $282^{6.52}$-$290^{6.60}$ likely represented pi interactions of Phe$281^{6.51}$ and Phe$282^{6.52}$ with the aromatic ring of the catechol group and interaction of Asn$285^{6.55}$ with Ser$218^{5.46}$. The masking effect on peptide $304^{7.39}$-$308^{7.43}$ presumably reflected interaction of Asn$304^{7.39}$ with the ethanolamine moiety of isoprenaline. MS/MS revealed specific masking effects on peptide Asn$304^{7.39}$-Trp$305^{7.40}$ and Leu$306^{7.41}$ (Figure 6.21f), further highlighting the power that MS/MS-based approaches have in identifying high resolution protein-ligand interactions. The reductions in carbene modification on peptides $291^{6.61}$-$299^{7.34}$ and $291^{6.61}$-$300^{7.35}$ also reinforced binding of isoprenaline to the binding cavity despite these peptides lacking residues involved in contacting the ligand. Isoprenaline has

been shown to induce a significant contraction of the β_1AR binding cavity, reducing its internal volume by 42 %.[185] A contraction to this extent would likely restrict chemical accessibility of the labelling probe and limit carbene modification at the catecholamine binding site which is what I observed. Rößler and colleagues applied NMR spectroscopy to the study of β_1AR with several small molecules and intracellular binding partners.[190] β_1AR NMR spectra were largely unaffected by the presence of isoprenaline and the authors were unable to discern ligand binding but suggested that isoprenaline slightly shifted the β_1AR equilibrium state towards the active state despite the receptor continuing to be stabilised by the inactive conformation. My data therefore highlight binding of a small molecule to the β_1AR orthosteric site for the first time (as of writing), using non-crystallographic techniques.

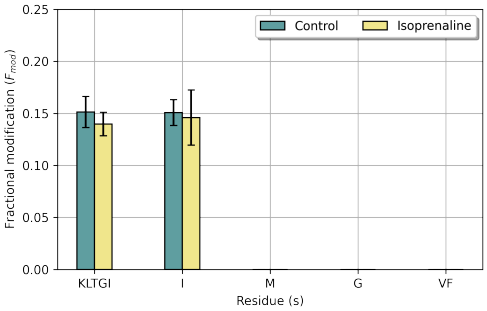
Carbene footprinting was next performed on β_1AR in the presence of isoprenaline and Nb80 (Figure 6.22). Pleasingly the same peptide and MS/MS-level masking differences were identified around the orthosteric site as observed with isoprenaline alone (Figure 6.21), reiterating binding of the agonist to the orthosteric binding cavity. Several additional masking effects were observed away from this region, including on peptides $134^{3.42}$-$143^{3.51}$ (TLCVIAIDRY), $144^{3.52}$-$150^{34.51}$ (LAITSPF), $265^{6.35}$-$274^{6.44}$ (KTLGIIMGVF) and $268^{6.38}$-$274^{6.44}$ (GIIMGVF). Peptides $265^{6.35}$-$274^{6.44}$ and $268^{6.38}$-$274^{6.44}$ were located on TM6. This transmembrane helix is documented to undergo a large conformational change upon receptor activation, shifting outwards by ~14 Å to accommodate the intracellular binding partner.[191] MS/MS analysis revealed specific masking effects on residues $265^{6.35}$-$269^{6.39}$ (KTLGI) and $Ile270^{6.40}$ (Figure 6.22d). Interrogation of the crystal structure (Figure 6.33a) revealed that these regions were proximal to Nb80 on the intracellular side of the receptor. This suggested that labelling reductions were caused by binding of Nb80 but masking effects due to conformational change in TM6 could not be eliminated. Indeed, it remained unknown how a conformational change on TM6 would effect chemical accessibility and carbene modification of the helix. It must be borne in mind that carbene footprinting captures the entire breadth of labelled protein states where the fractional modification represents the average amount of carbene insertion over the irradiation period. As such, loosely stabilised conformations whose populations are not fully shifted to a predominant equilibrium state, or even coexisting conformational



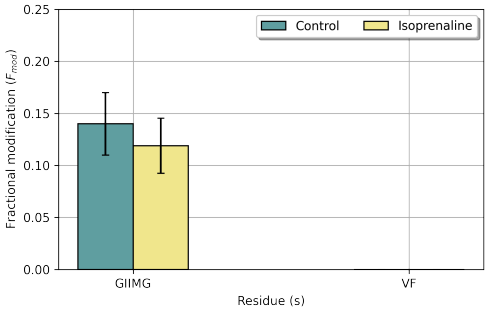
(a) Carbene footprinting histogram of β_1AR alone (control), and with isoprenaline with a chymotrypsin digestion. The fractional modification of each chymotryptic β_1AR peptide is shown with (yellow) and without (turquoise) 100 µM of isoprenaline. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



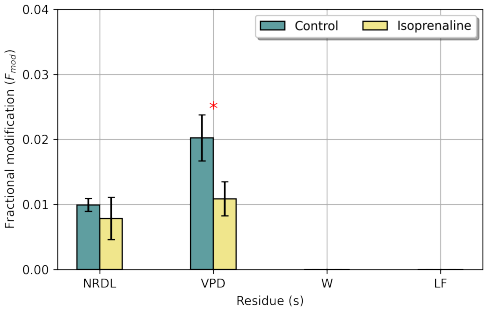
(b) Sub-peptide analysis of labelled AIASSIISF (peptide 211-219).



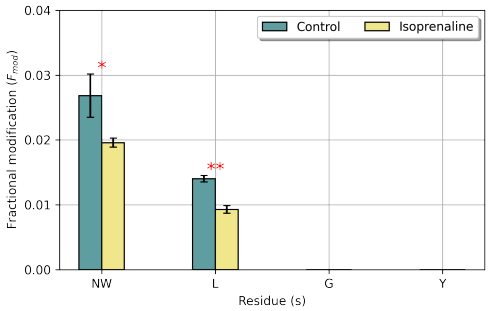
(c) Sub-peptide analysis of labelled KTLGIIMGVF (peptide 265-274).



(d) Sub-peptide analysis of labelled GIIMGVF (peptide 268-274).



(e) Sub-peptide analysis of labelled NRDLVPDWLF (peptide 291-300).



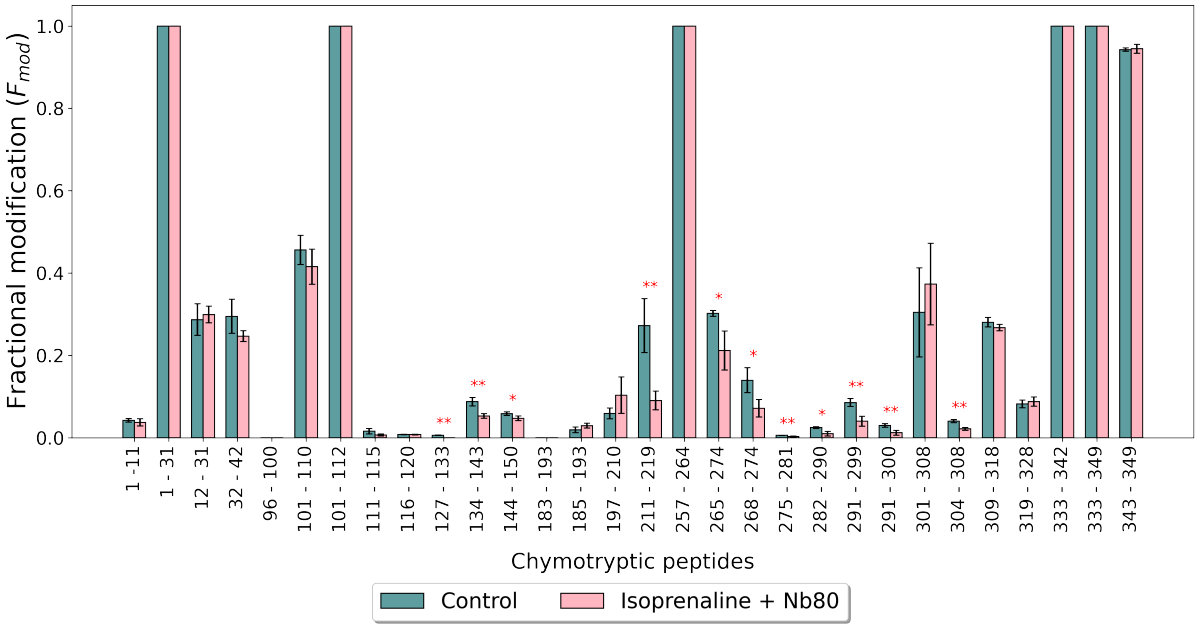
(f) Sub-peptide analysis of labelled NWLGY (peptide 304-408).

Figure 6.21 | Carbene footprinting of β_1AR with and without isoprenaline using a chymotrypsin digestion. Full scan data highlighted chymotryptic peptide-level differences in carbene modification between control and isoprenaline-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

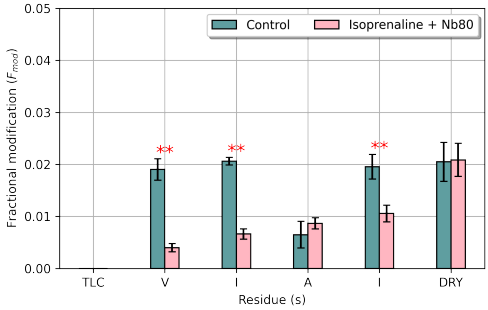
states may be concealed. Furthermore, subtle conformational transitions which do not lead to large changes in chemical accessibility and therefore lack an obvious differential masking effect may also remain unrecognised. Regardless, the masking effects away from the extracellular binding site were at least indicative of receptor activation, and highlighted for the first time that carbene footprinting could effectively capture molecular switches and/or interaction of an intracellular binding partner to a GPCR. Peptide $257^{6.27}$-$264^{6.34}$ (AMKEHKAL), located on the intracellular tip of TM6 exhibited 100 % labelling in control and ternary complex samples. This highlighted a caveat of many covalent labelling metrics where values are bound between 0-1 and a peptide displaying complete modification in the presence and absence of a binding partner may actually be more or less labelled in either case. Therefore, whilst $257^{6.27}$-$264^{6.34}$ was 100 % labelled in either treatment, I was unable to discern whether isoprenaline and Nb80 binding led to a quantitative difference on the peptide. One solution may involve repeating footprinting experiments with a lower concentration of NaTDB (<20 mM) which in turn may reduce the labelling from 100 % and potentially enable more informative analysis of this peptide. Indeed, peptide-level labelling titrations could be explored as important refinements of the labelling optimisation workflow. Peptides $134^{3.42}$-$143^{3.51}$ and $144^{3.52}$-$150^{34.51}$ were located on TM3. Peptide $134^{3.42}$-$143^{3.51}$ contained the conserved 'DRY' motif. This motif was located adjacent to peptide $144^{3.52}$-$150^{34.51}$. In the receptor's inactive conformation, $Arg142^{3.50}$ of the DRY motif forms a salt bridge with $Glu260^{6.30}$ locking the receptor in an idle state but upon

receptor activation, movement of TM6 breaks this interaction. This region and the majority of 144$^{3.52}$-150$^{34.51}$ are located proximally to the intracellular Nb binding cavity. Masking effects on these peptides in the presence of isoprenaline and Nb80 (whilst not observed in the isoprenaline-treated sample) again suggested at either β_1AR-Nb interactions and/or changes to intramolecular bonding within the receptor associated with activation. MS/MS of labelled peptide 134$^{3.42}$-143$^{3.51}$ highlighted distinct masking effects on residues Val137$^{3.45}$-Ile138$^{3.46}$ and Ile140$^{3.48}$ (Figure 6.22b). Examination of the crystal structure (Figure 6.33a) revealed that Ile138$^{3.46}$ was within 5 Å of Nb80 and it is plausible that the masking effect on Ile138$^{3.46}$ and peptide 144$^{3.52}$-150$^{34.51}$ represented interaction with the Nb. On the other hand, Ile140$^{3.48}$ was within 10 Å of Nb80 and it appeared more likely that this masking effect reflected some degree of conformational change such as the breaking of the Arg142$^{3.50}$ salt bridge with Glu285$^{6.30}$. No change in carbene modification was observed on peptide 309$^{7.44}$-318$^{7.53}$ in the activated complex, suggesting that formation of the water-lock did not lead to differences in chemical accessibility of the labelling probe. Rößler and colleagues identified changes to all labelled methionine signals in β_1AR NMR spectra in the presence of isoprenaline and Nb80 despite three out of 11 methionine residues occurring in the intracellular binding cavity.[190] They inferred that these changes reflected structural transition of the receptor to the activated state and interaction with the intracellular binding partner. My results corroborate previous knowledge of the β_1AR system, reiterating binding of isoprenaline to the β_1AR orthosteric site, interaction with Nb80 and structural changes in the receptor associated with these contacts.

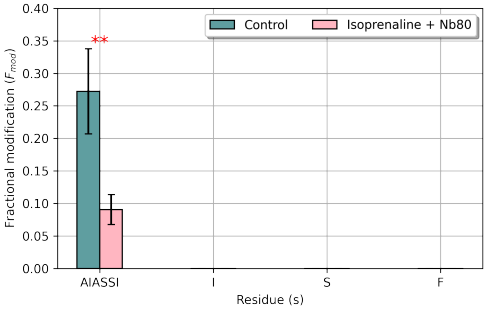
β_1AR footprinting experiments were repeated as above but trypsin was employed to digest samples. A single masked peptide was observed in β_1AR treated with isoprenaline, compared to control samples (Figure 6.23). Peptide 195-208$^{5.36}$ (CYQDPGCCDFVTNR) was located near to the extracellular binding cavity which suggested the labelling reduction was due to ligand masking. This effect was not observed on the chymotryptic peptide 197-210$^{5.38}$ (QDPGCCDFVTNRAY) despite sharing the majority of residues with the tryptic peptide. This indicated that either Cys195 and Tyr196 were masked in the presence of the ligand or that labelling differences were located this shared region (i.e. QDPGCCDFVTNR) and increased labelling of Ala209$^{5.37}$ and Tyr210$^{5.38}$ veiled these effects on the chymotryptic peptide.



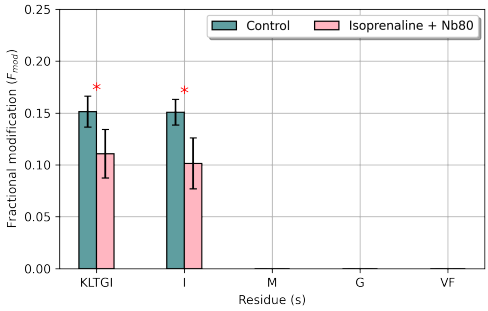
(a) Carbene footprinting histogram of β_1AR alone (control), and with isoprenaline and Nb80 with a chymotrypsin digestion. The fractional modification of each chymotryptic β_1AR peptide is shown with (pink) and without (turquoise) 100 µM of isoprenaline and 2:1 molar equivalents of Nb80. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



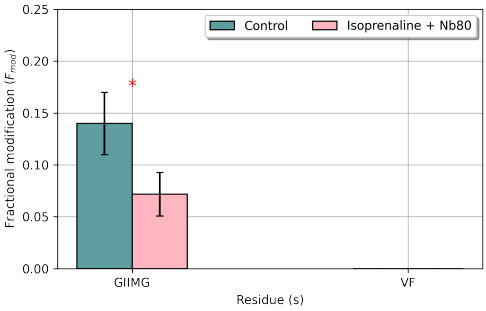
(b) Sub-peptide analysis of labelled TLCVIAIDRY (peptide 134-143).



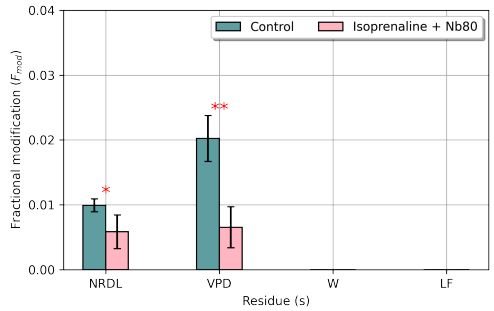
(c) Sub-peptide analysis of labelled AIASSIISF (peptide 211-219).



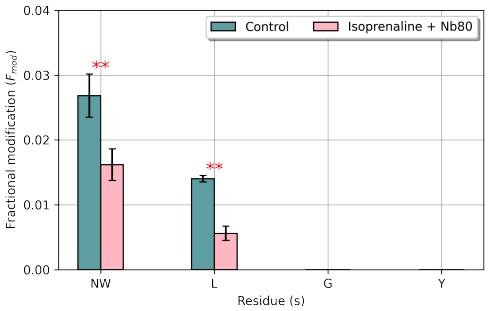
(d) Sub-peptide analysis of labelled KTLGIIMGVF (peptide 265-274).



(e) Sub-peptide analysis of labelled GIIMGVF (peptide 268-274).



(f) Sub-peptide analysis of labelled NRDLVPDWLF (peptide 291-300).

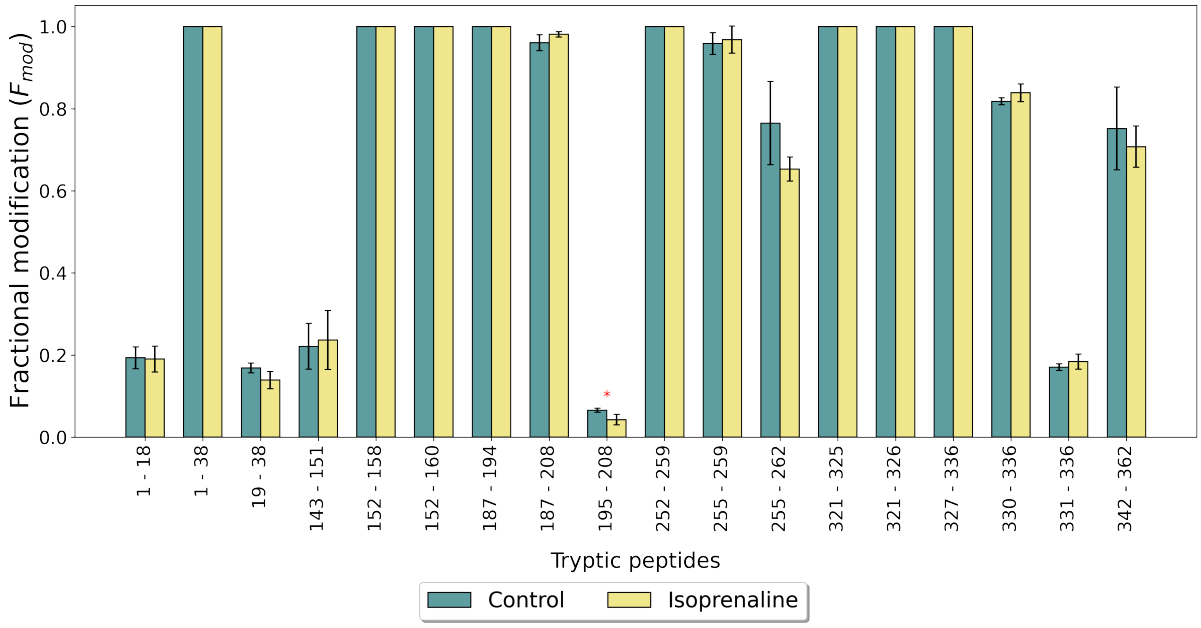


(g) Sub-peptide analysis of labelled NWLGY (peptide 304-308).

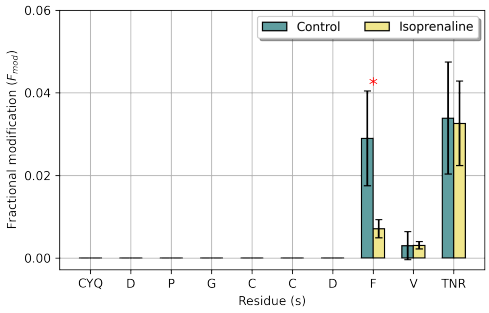
Figure 6.22 | Carbene footprinting of β_1AR with and without isoprenaline and Nb80 using a chymotrypsin digestion. Full scan data highlighted chymotryptic peptide-level differences in carbene modification between control and isoprenaline and Nb80-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

MS/MS revealed carbene modification was located on residues $204^{45.42}$-$208^{5.36}$ (FVTNR) with specific masking at Phe$204^{45.52}$ in the presence of isoprenaline (Figure 6.23b). Analysis of the crystal structure (Figure 6.33a) showed that this residue was within ~2 Å of isoprenaline but noncovalent interactions between the catecholamine and ECL2 were not predicted.[183] The masking effect at Phe$204^{45.52}$, on ECL2, presumably represented steric inhibition of the labelling probe by binding of isoprenaline. These results highlighted how a multi-protease approach could provide complementary labelling information.

The tryptic digest of β_1AR treated with isoprenaline and Nb80 displayed labelling reductions on two peptides compared to the β_1AR control, including on peptide 195-$208^{5.36}$ (CYQDPGCCDFVTNR) as previously seen with isoprenaline (Figure 6.23) but also on peptide $143^{3.51}$-$151^{34.52}$ (YLAITSPFR) (Figure 6.24). MS/MS analysis of 195-$208^{5.36}$ was consistent with previous observations which further supported interaction between β_1AR and isoprenaline. The labelling reduction on peptide $143^{3.51}$-$151^{34.52}$ was consistent with the masked chymotryptic peptide $144^{3.52}$-$150^{34.51}$ (LAITSPF), reinforcing this region's involvement in receptor activation and highlighting the accuracy and reliability with which carbene footprinting can identify protein interactions (Figure 6.22). MS/MS analysis revealed specific masking effects on residues $143^{3.51}$-$145^{3.52}$ (YL) and $148^{3.56}$-$151^{34.52}$ (SPFR) (Figure 6.24b). Interrogation of the crystal structure revealed direct proximity of these regions but particularly residues $148^{3.56}$-



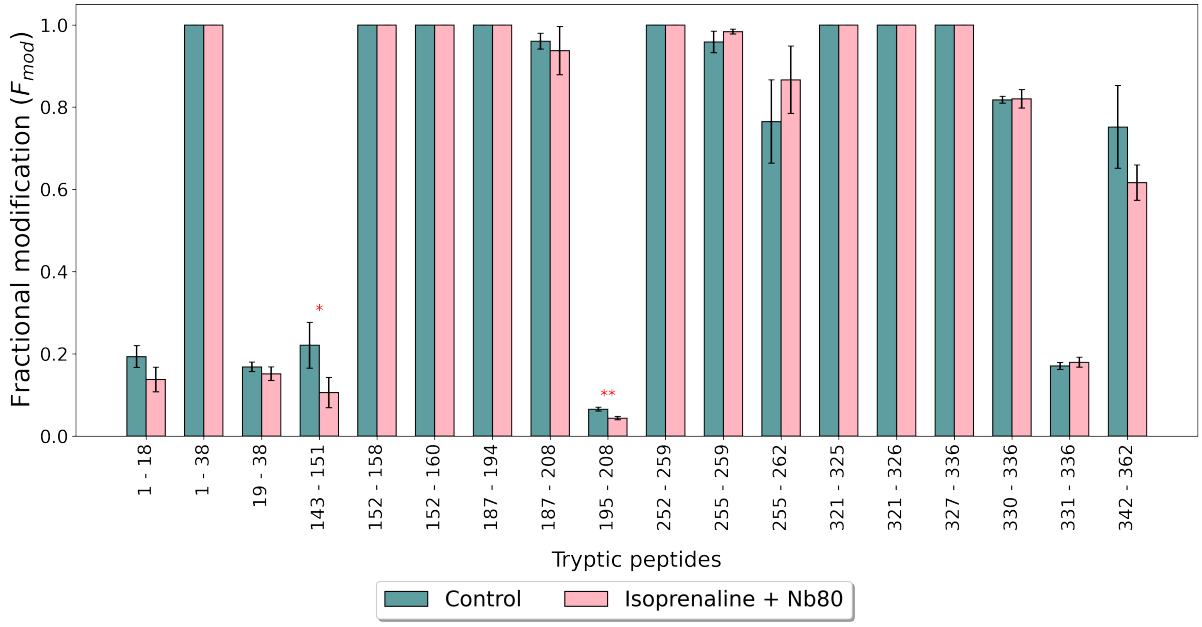
(a) Carbene footprinting histogram of β_1AR alone (control), and with isoprenaline with a trypsin digestion. The fractional modification of each tryptic β_1AR peptide is shown with (yellow) and without (turquoise) 100 µM of isoprenaline. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



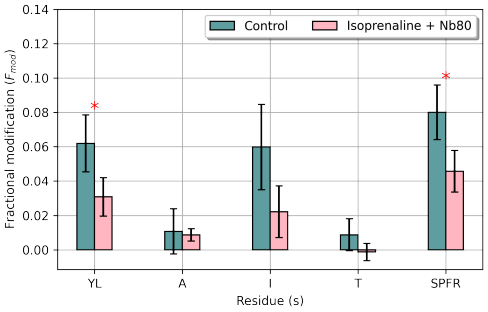
(b) Sub-peptide analysis of labelled CYQDPGCCDFVTNR (peptide 195-208).

Figure 6.23 | Carbene footprinting of β_1AR with and without isoprenaline using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and isoprenaline-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

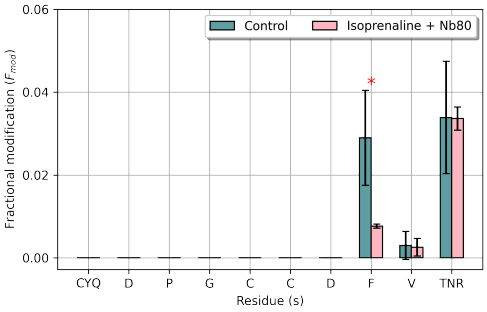
$151^{34.52}$ to Nb80 (Figure 6.33a) again suggesting at masking effects caused by protein-protein interactions and binding partner-induced structural change.



(a) Carbene footprinting histogram of β_1AR alone (control), and with isoprenaline and Nb80 with a trypsin digestion. The fractional modification of each tryptic β_1AR peptide is shown with (pink) and without (turquoise) 100 µM of isoprenaline and 2:1 molar equivalents of Nb80. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



(b) Sub-peptide analysis of labelled YLAITSPFR (peptide 143-151).



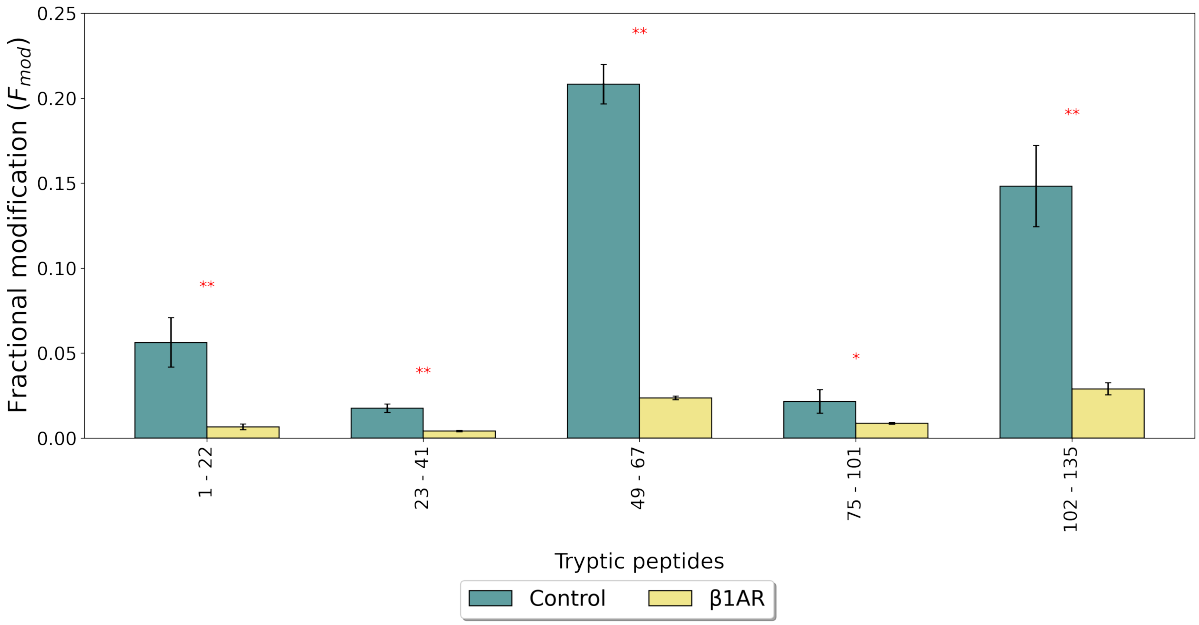
(c) Sub-peptide analysis of labelled CYQDPGCCDFVTNR (peptide 195-208).

Figure 6.24 | Carbene footprinting of β_1AR with and without isoprenaline and Nb80 using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and isoprenaline and Nb80-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

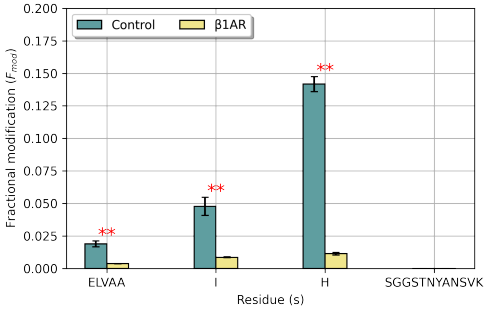
Lastly, carbene footprinting was applied to Nb80 in the presence of β_1AR, and β_1AR and isoprenaline (Figure S8). Binding of Nbs to βAR receptors occurs in the absence of small molecules and β_2AR bound to Nb80 has been shown to increase the

binding affinity to isoprenaline by 95-fold.[192] As such, masking effects on Nb80 were expected with β_1AR in both the presence and absence of isoprenaline.

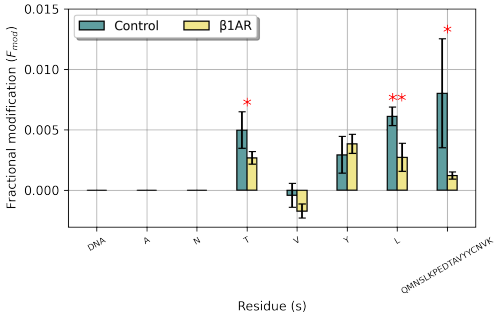
Nb80 treated with β_1AR displayed extensive reductions in carbene modification across every tryptic peptide compared with control samples (Figure 6.25). This was explained by the largely antiparallel β-sheet nature of Nb80 and the long tryptic peptides which were generated, where each peptide spanned the length of the protein and presumably contained residues involved in binding to β_1AR or that were at least proximal to it (Figure 6.33a). Indeed, analysis of the crystal structure revealed that every tryptic Nb80 peptide contained at least one residue that was within 6 Å of β_1AR. MS/MS allowed sub-peptide labelling analysis of two peptides. Fragmentation of labelled peptide 49-67 (ELVAAIHSGGSTNYANSVK) revealed that all carbene modification (and masking effects) were located on residues 49-53 (ELVAA), Ile54 and His55 (Figure 6.25b). Examination of the crystal structure (Figure 6.33a) showed that these residues were located within the intracellular binding cavity of β_1AR which again indicated binding of Nb80 to β_1AR. Fragmentation of labelled peptide 75-101 (DNAANTVYLQMNSLKPEDTAVYYCNVK) showed that masking effects were located on Thr80, Leu83 and residues 84-101 (QMNSLKPEDTAVYYCNVK) (Figure 6.25c). Puzzlingly, Thr80 and Leu83 were located distally from β_1AR. The C-terminal end of peptide 84-101 was located proximally to the interaction site and it appeared plausible that the masking effect on the peptide reflected this contact interface. These results showed that carbene footprinting of Nb80 could be used to highlight its interaction with β_1AR.



(a) Carbene footprinting histogram of Nb80 alone (control), and with β_1AR with a trypsin digestion. The fractional modification of each tryptic Nb80 peptide is shown with (yellow) and without (turquoise) 2:1 molar equivalents of β_1AR. Error bars are $\pm$ standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



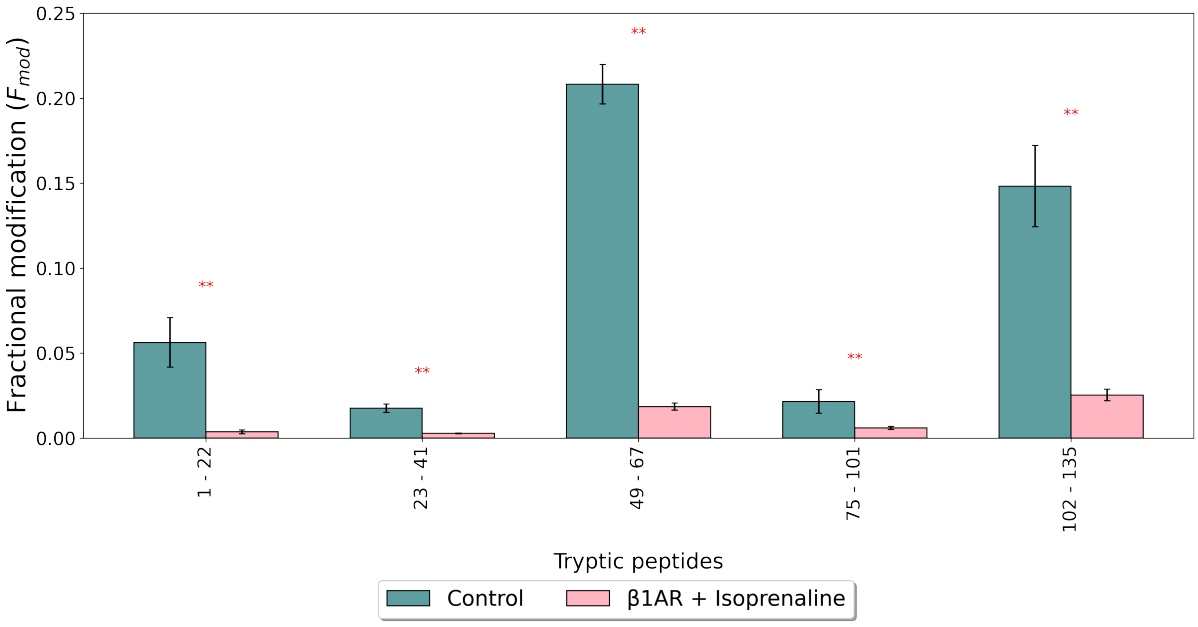
(b) Sub-peptide analysis of labelled ELVAAIHSGGSTNYANSVK (peptide 49-67).



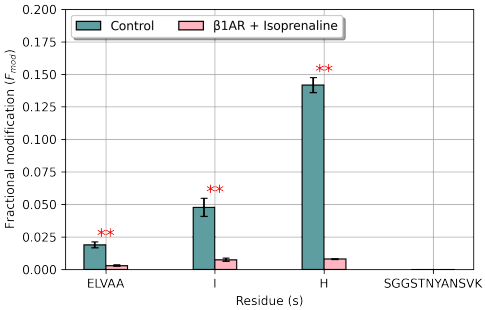
(c) Sub-peptide analysis of labelled DNAANTVYLQMNSLKPEDTAVYYCNVK (peptide 75-101).

Figure 6.25 | Carbene footprinting of Nb80 with and without β_1AR using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and β_1AR-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

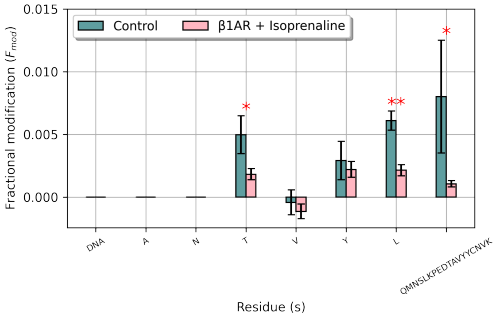
Nb80 treated with β_1AR and isoprenaline also displayed reductions in carbene modification across all tryptic peptides compared with control samples. This was in agreement with previous observations (Figure 6.26) and in accordance with the literature, and again highlighted binding of the Nb to β_1AR irrespective of agonist binding.[192]



(a) Carbene footprinting histogram of Nb80 alone (control), and with β_1AR and isoprenaline with a trypsin digestion. The fractional modification of each tryptic Nb80 peptide is shown with (pink) and without (turquoise) 2:1 molar equivalents of β_1AR and 100 µM isoprenaline. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



(b) Sub-peptide analysis of labelled ELVAAIHSGGSTNYANSVK (peptide 49-67).



(c) Sub-peptide analysis of labelled DNAANTVYLQMNSLKPEDTAVYYCNVK (peptide 75-101).

Figure 6.26 | Carbene footprinting of Nb80 with and without β_1AR and isoprenaline using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and β_1AR and isoprenaline-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

6.2.5.2 Inactivation of beta-1 adrenergic receptor

After successfully footprinting the activated receptor complex, I was interested to determine if carbene labelling could be applied to the inactivated complex and how footprinting data differed between the apo and carazolol-bound states. I was also

interested to determine how the footprinting results compared to the activated state (Section 6.2.5.1).

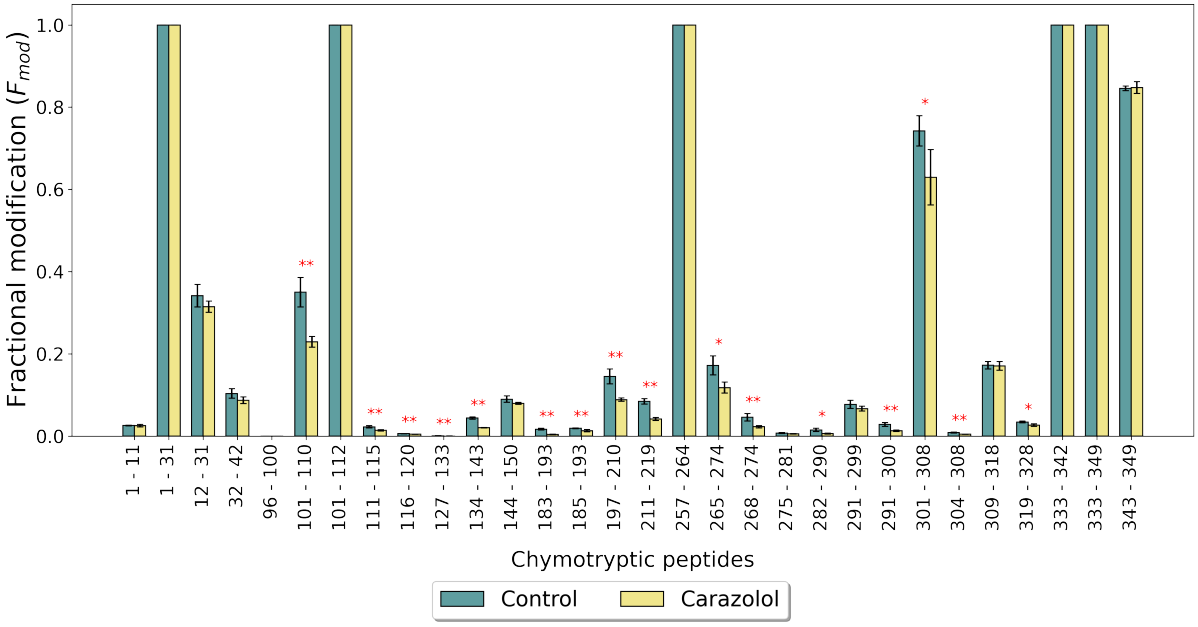
Binding of the inverse agonist carazolol and Nb60 to β_1AR were required to elicit full receptor inactivation. Enzymatic digestion of β_1AR was carried out separately with chymotrypsin and trypsin to maximise peptide coverage and insight into receptor interactions and dynamics. MS/MS was also performed on selected peptides that displayed labelling changes compared to control samples to improve labelling resolution. Again due to sample conservation, intensity of certain labelled peptides and co-elution of higher intensity ions, MS/MS was not performed on all labelled peptides that displayed masking effects.

Reductions in carbene modification were observed on several β_1AR chymotryptic peptides in carazolol-treated samples compared to control samples (Figure 6.27a, Figure S9). The majority of masking effects localised to the extracellular binding cavity, highlighting binding of the inverse agonist. Several of these labelling differences were shared with those seen in isoprenaline-treated samples but additional masking effects were also identified. These were located at the tip of TM2, TM3 and TM4 and included peptides $101^{2.61}$-$110^{23.50}$ (GATLVVRGTW), $111^{23.51}$-$115^{3.23}$ (LWGSF), $116^{3.24}$-$120^{3.28}$ (LCELW), $183^{4.64}$-193 (HWWRDEDPQAL), 185-195 (WRDEDPQAL), 197-$210^{5.38}$ (QDPGCCDFVTNRAY) and $301^{7.36}$-$308^{7.43}$ (VAFNWLGY). Masking effects were also observed on some membrane-spanning helices, including peptides $144^{3.52}$-$150^{34.51}$ (LAITSPF), $265^{6.35}$-$274^{6.44}$ (KTLGIIMGVF) and $268^{6.38}$-$274^{6.44}$ (GIIMGVF) and $319^{7.54}$-$328^{8.54}$ (CRSPDFRKAF). Peptide $275^{6.45}$-$281^{6.51}$ (TLCWLPF) was not observed to display masking effects whilst these were observed in both isoprenaline-treated chymotryptic digests suggesting differences in chemical accessibility at the orthosteric site between the agonist and inverse agonist. The regions of additional masking around the orthosteric cavity appeared to reflect the increased size of carazolol compared to isoprenaline. The reduction in fractional modification on peptide $101^{2.61}$-$110^{23.50}$, located on the extracellular tip of TM2 reinforced this notion where MS/MS revealed specific masking effects on residues $101^{2.61}$-$106^{2.66}$ and Arg$107^{2.67}$ (Figure 6.27b). The adjacent missed cleavage 101-112 (GATLVVRGTWLW) displayed 100 % labelling indicating that residues Leu$111^{23.51}$-Trp$112^{23.52}$ were highly modified and prevented masking effects from being identified. Residues $101^{2.61}$-$106^{2.66}$ were located within

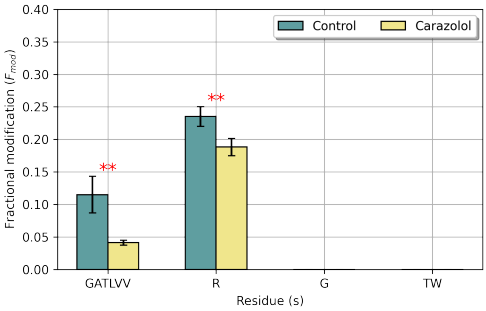
7 Å of the isopropylamino group. However Arg107$^{2.67}$, located on ECL1, was distal from carazolol (~15 Å) and the masking effect appeared to reflect a change in protein conformation on this unstructured region, presumably as a result of ligand binding. A similar effect was observed on the adjacent peptide 111$^{23.51}$-115$^{3.23}$, located on ECL1 and the extracellular portion of TM2. This peptide appeared too distal to the binding cavity to reflect ligand contacts. Similar observations were made on peptides 183$^{4.64}$-193 and 185-195, located on ECL2 which again reiterated shielding effects in the presence of carazolol. West and colleagues used HDX to probe binding of several catechols to β_2AR.[82] They observed stabilising effects (reductions in deuterium uptake) around ECL2 with clenbuterol (partial agonist), alprenolol (antagonist), timolol (partial inverse agonist) and carazolol (inverse agonist) but not isoprenaline. These observations were suggested to reflect interaction of the ligands with ECL2, facilitating the loop's stabilisation, and our carbene labelling results corroborate this notion.[82] Masking of peptide 197-210$^{5.38}$ appeared to highlight extended binding contacts of carazolol's carbazole group compared to the catechol moiety on isoprenaline. Moreover masking on this peptide was not observed in isoprenaline-treated samples. MS/MS displayed specific labelling reductions on Val205 and residues 208$^{5.36}$-210$^{5.38}$ (Figure 6.27d). Phe206$^{45.52}$ and Tyr210$^{5.38}$ are predicted to form strong hydrophobic contacts with carazolol's carbazole group. Masking was observed on the tryptic peptide 195-208$^{5.36}$ (CYQDPGCCDFVTNR) in both isoprenaline treatments but this presumably reflected proximity of the agonist to the peptide rather than any direct interactions. On the other hand, the observed masking at Val205 (adjacent to Phe206$^{45.52}$) and residues 208$^{5.36}$-210$^{5.38}$ in the presence of carazolol likely reflected direct contacts with the partial inverse agonist. Similarly, masking of peptide 301$^{7.36}$-308$^{7.43}$ was not observed in isoprenaline-treated samples. This peptide was a missed cleavage, containing the observed chymotryptic peptide 304$^{7.39}$-308$^{7.43}$ (NWLGY) which displayed sub-peptide level masking effects at Asn304$^{7.39}$-Trp305$^{7.40}$ and Leu306$^{7.41}$. However residues 304$^{7.39}$-308$^{7.43}$ exhibited far lower carbene modification than the missed cleavage 301$^{7.36}$-308$^{7.43}$. This suggested that residues 301$^{7.36}$-303$^{7.38}$ were responsible for the high level of carbene labelling. Interrogation of the β_1AR-carazolol crystal structure (Figure 6.34a) revealed that Val301$^{7.36}$ was within 7.8 Å of the inverse agonist, compared to 8.5 Å with isoprenaline. This small decrease in proximity may have been responsible for

reduced carbene modification on Val301$^{7.36}$ in carazolol-treated samples compared to isoprenaline-treated samples and contributed to the overall masking effect on the peptide. This would also explain why no masking effect was observed on peptide 301$^{7.36}$-308$^{7.43}$ in isoprenaline-treated samples where residues 301$^{7.36}$-303$^{7.38}$ were not masked but displayed a high degree of labelling, concealing those changes on residues 304$^{7.39}$-308$^{7.43}$. The remaining masking effects observed on transmembrane-spanning helices were slightly more unanticipated. RMSD between unliganded and carazolol-bound β_1AR was 0.737 Å, and 0.524 Å between isoprenaline and carazolol-bound β_1AR, suggesting at minimal structural change between unliganded, agonist and inverse agonist-bound states. However my results suggest pronounced stabilisation effects on helix three and six in the presence of carazolol. These findings are also supported by work conducted by West and colleagues who showed distinct shielding effects on these helices (and TM5) in the presence of carazolol.[82] Protection effects were also observed on peptide 319$^{7.49}$-328$^{7.53}$, located on the small intracellular helix VIII, in carazolol-treated samples compared to control samples. West and colleagues also observed minimal deuterium exchange in HDX experiments up to 60 s on this region of β_2AR.[82] Isoprenaline-treated samples were shown to induce slow rates of exchange after this time point and the authors suggested that this region was highly conformationally dynamic. These suggestions were reinforced by active and inactive β_2AR crystal structures but particularly at the NPxxY motif which displayed a more extended and therefore less protected conformation in the active, agonist-bound state, which corroborated our findings.[192] These results show that in addition to identifying ligand binding at the orthosteric binding site, carbene footprinting can also discern between subtle differences in binding contacts and potentially identify conformational effects induced by ligand binding not yet detected by NMR spectroscopy or X-ray crystallography. Of course, further repeat experiments could be employed to mitigate potential false positive results.

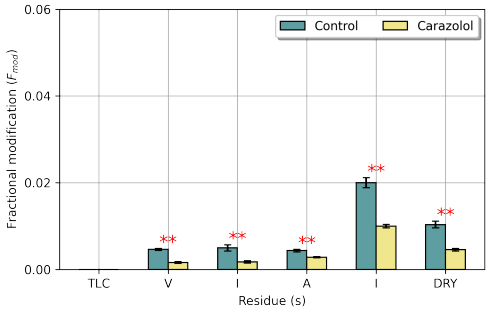
Changes in carbene labelling were observed on several β_1AR chymotryptic peptides for carazolol and Nb60-treated samples compared to control samples (Figure 6.28a). These changes were generally consistent between carazolol, and carazolol and Nb60-treated β_1AR but peptides 144$^{3.52}$-150$^{34.51}$ (LAITSPF) and 309$^{7.44}$-318$^{7.53}$ (ANSAMNPIIY) displayed unmasking effects in the ternary complex whilst



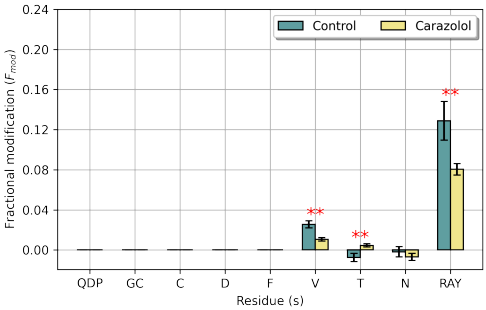
(a) Carbene footprinting histogram of β_1AR alone (control), and with carazolol with a chymotrypsin digestion. The fractional modification of each chymotryptic β_1AR peptide is shown with (yellow) and without (turquoise) 100 µM of carazolol. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



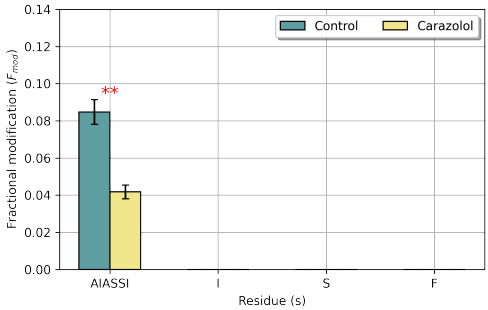
(b) Sub-peptide analysis of labelled GATLVVRGTW (peptide 101-110).



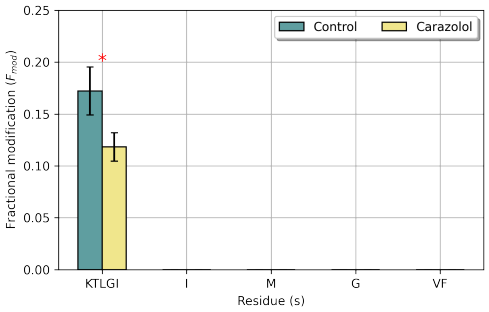
(c) Sub-peptide analysis of labelled TLCVIAIDRY (peptide 134-143).



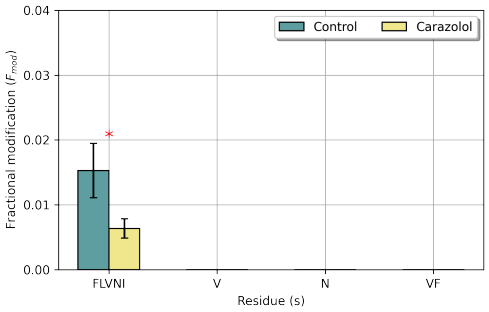
(d) Sub-peptide analysis of labelled QDPGCCDFVTNRAY (peptide 197-210).



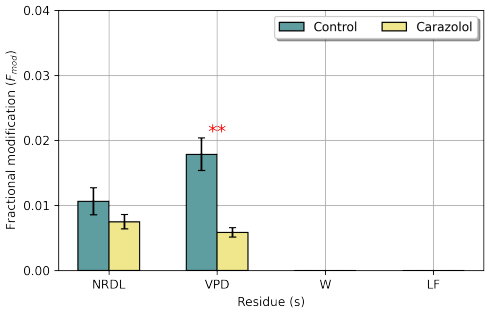
(e) Sub-peptide analysis of labelled AIASSIISF (peptide 211-219).



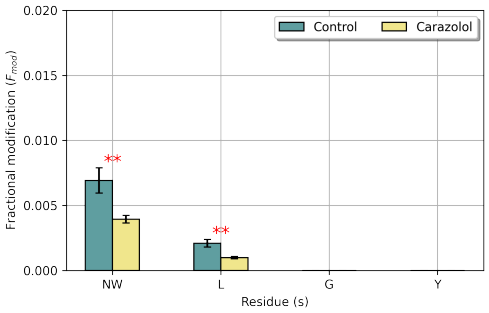
(f) Sub-peptide analysis of labelled KTLGIIMGVF (peptide 265-274).



(g) Sub-peptide analysis of labelled FLVNIVNVF (peptide 282-290).



(h) Sub-peptide analysis of labelled NRDLVPDWLF (peptide 291-300).



(i) Sub-peptide analysis of labelled NWLGY (peptide 304-308).

Figure 6.27 | Carbene footprinting of β_1AR with and without carazolol using a chymotrypsin digestion. Full scan data highlighted chymotryptic peptide-level differences in carbene modification between control and carazolol-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

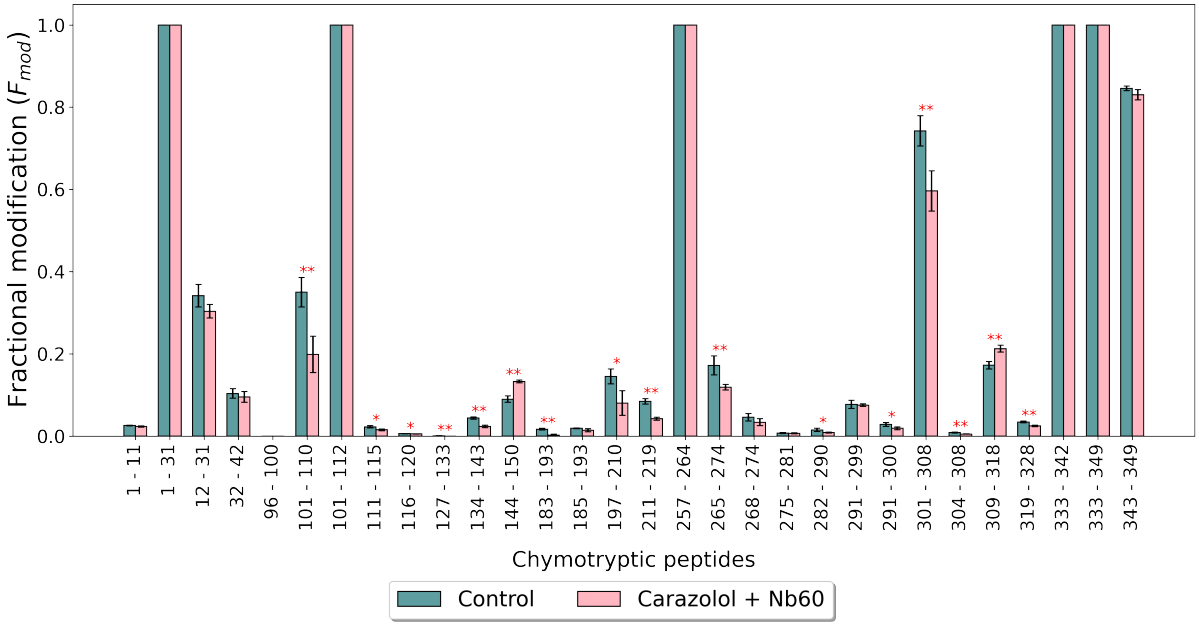
peptide $268^{6.38}$-$274^{6.44}$ (GIIMGVF) exhibited loss of significant masking effects (Figure 6.27a, Figure 6.28a). The unmasking effects observed on peptides $144^{3.52}$-$150^{34.51}$ and $309^{7.44}$-$318^{7.53}$ likely represented conformational change to TM3/ICL2 and TM7 as interaction with Nb60 would not be expected to cause a gain in carbene modification at/near to contact sites. Peptide $144^{3.52}$-$150^{34.51}$ showed specific unmasking events on residues $144^{3.52}$-$147^{3.55}$, $Ser148^{3.56}$ and $Phe150^{34.51}$ (Figure 6.28d) whilst peptide $309^{7.44}$-$318^{7.53}$ exhibited an increase in carbene modification on residues $309^{7.44}$-$312^{7.47}$ and $314^{7.49}$-$318^{7.53}$ (Figure 6.28k). The observed gain in carbene labelling on TM3 may be explained by the S2 (broken ionic-lock) conformation; yet, if this was the case, it raises the question of why similar unmasking effects were not identified in the activated ternary complex where this interaction would also be disengaged. Moreover, peptide $309^{7.44}$-$318^{7.53}$, located distally from the Nb60 binding site, was not predicted to undergo any conformational change in the inactivated state. Only in the ternary agonist effector protein complex does $Tyr319^{7.53}$ of the NPxxY motif form a water-mediated hydrogen bond with $Tyr227^{5.58}$ and even then, carbene labelling changes were not observed on this peptide in the presence isoprenaline and Nb80. Rößler and colleagues only observed slight changes to labelled methionine β_1AR NMR signals between apo and Nb60-bound β_1AR and proposed that their findings highlighted binding of the Nb with minimal perturbation to the overall receptor conformation.[190] In contrast, our findings indicate more substantial conformational change to β_1AR upon interaction with Nb60 that significantly increases the chemical accessibility of TM3/ICL2 and TM7. Grahl and colleagues highlighted the fluidity of β_1AR, showing that even when bound to antagonists, the receptor is in rapid equilibrium between inactive and preactive states with bias towards the inactivated state.[191] The authors showed that the presence of an intracellular binding partner is required to fully saturate the activated state. However this work did not include inactivatory Nbs, such as Nb60. Therefore, I suggest that carbene footprinting data reflect distinct conformational states of activated and inactivated β_1AR ternary complexes. Whilst these states are known to contain regions of local structural similarity (including the broken ionic-lock), it appears that global structural differences (and resultant changes to chemical accessibility) and complete saturation of these conformation states lead to well-defined carbene labelling profiles. The masking effects observed on chymotryptic β_1AR peptides $134^{3.42}$-$143^{3.51}$

and $265^{6.35}$-$274^{6.44}$ were consistent between carazolol, and carazolol and Nb60-treated samples. Reductions in carbene modification were also observed on these peptides in the β_1AR-Nb80 ternary complex. Since masking effects were not observed on these peptides in the presence of isoprenaline, differences were partly attributed to protein-protein interactions between the receptor and Nb80; however, this could not be the case for β_1AR treated with carazolol which suggested that binding of the inverse agonist was associated with changes in receptor structure, consistent with the above deductions. Surprisingly the Nb80 and Nb60-receptor ternary complexes displayed differing protein-protein contact surfaces (Figure 6.33b, Figure 6.33c, Figure 6.34b, Figure 6.34c). The Nb60-β_1AR contact surface was estimated (using X-ray structures) at 850 Å^2 compared with 1185 Å^2 for the activated ternary receptor complex.[190] 31 β_1AR residues were located within 5 Å of Nb60 with the majority of these residues located on TM1 and ICL2; however, 39 β_1AR residues were located within 5 Å of Nb80 with most of these residues situated on TM3, TM5 and TM6. These contrasting contact interfaces reflected conformational change to TM5 and TM6 in the activated β_1AR state where movement of the helices is essential for creating an intracellular G protein binding cavity. The comparison of contact interfaces reiterated that masking effects on $134^{3.42}$-$143^{3.51}$ and $265^{6.35}$-$274^{6.44}$ in the β_1AR-Nb80 complex were caused by Nb binding whereas in the carazolol treatment and β_1AR-Nb60 ternary complex, masking likely reflected conformational change/stabilisation effects on these regions. Minimal sequence coverage on TM1 and TM5 as well as 100 % fractional modification on ICL3/TM6 peptides meant that the majority of Nb interactions on β_1AR remained unrecognised. For example, residues Tyr152-Arg159 were predicted to be within 5 Å of Nb60 however this exact tryptic peptide displayed 100 % fractional modification in the presence or absence of the Nb, meaning that labelling differences remained unidentified. Therefore, improvement of β_1AR peptide sequence coverage should be investigated. Zhang and colleagues showed increased sequence coverage of β_2AR when proteolytic digestion was carried out in the presence of detergents.[83] Unfortunately, non-ionic surfactants like DDM are not compatible with nESI-MS and difficulties in its removal meant that this strategy was not employed in my experiments. Further work could also employ alternate proteases to provide increased sequence coverage, such as the α-lytic enzyme or ProAlanase. In addition, the S-trap could be used to potentially improve

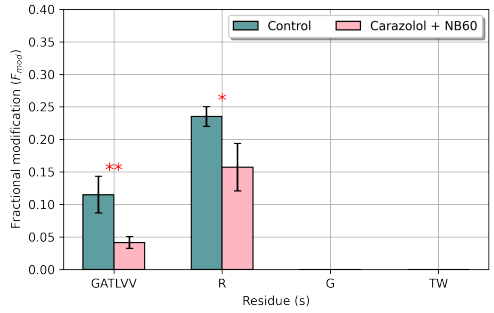
peptide sequence coverage. This technology employs SDS to solubilise proteins during a proteolytic digestion step and removal of the surfactant is carried out by centrifugation stages enabling MS analysis. Importantly, the methodology has been shown to increase peptide coverage for membrane proteins.[193] Lastly, carbene labelling experiments could be repeated at lower concentrations of NaTDB (>20 mM) which may enable more informative analysis of peptides on ICL3/TM6.

The tryptic digest of β_1AR treated with carazolol exhibited a single masking effect on peptide 195-208$^{5.36}$ (CYQDPGCCDFVTNR) compared to control β_1AR (Figure 6.29a). This tryptic peptide was also previously masked under both isoprenaline conditions. MS/MS showed specific masking effects at 195-196 (CY) and Pro198 which were again consistent with previous agonist treatments (Figure 6.29b). Interrogation of the crystal structure showed these residues were far too distal from the binding pocket to constitute direct binding contacts despite the slightly larger size of carazolol compared to isoprenaline. This again suggested that small molecule binding induced stabilisation effects on ECL2.

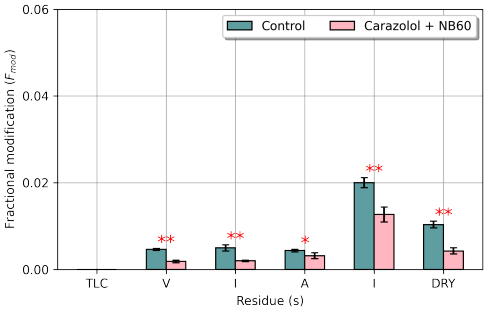
The tryptic digest of β_1AR treated with carazolol and Nb60 displayed an unmasking event on peptide 143$^{3.51}$-151$^{34.52}$ (YLAITSPFR) and masking events on peptides 195-208$^{5.36}$ (CYQDPGCCDFVTNR) and 330$^{8.56}$-336 (RLLAFPR) compared to control β_1AR (Figure 6.30a). Pleasingly, the unmasking event on tryptic peptide 143$^{3.51}$-151$^{34.52}$ further corroborated findings from the chymotryptic digest where peptides around TM3/ICL2 displayed a gain in carbene modification in the presence of carazolol and Nb60 (Figure 6.28a). This unmasking event reiterated distinct conformational changes to TM3 in the inactive ternary complex despite work by Rößler and colleages that suggested otherwise.[190] Nevertheless, the resolution of their work was largely limited by the number of available methionine residues which presumably prevented identification of these effects on TM3. The labelling reduction on peptide 195-208$^{5.36}$ was consistent with previous tryptic β_1AR digests (Figure 6.30c) and again suggested conformational changes in ECL2 caused by ligand binding. The masking event on peptide 330$^{8.56}$-336 was not observed on the subsequent peptide 331$^{8.57}$-336, suggesting that the drop in chemical modification was occurring on Arg330$^{8.56}$, located on intracellular helix VIII. The chymotryptic digest also displayed masking effects on this region of the receptor in the presence of both carazolol, and carazolol and Nb60;



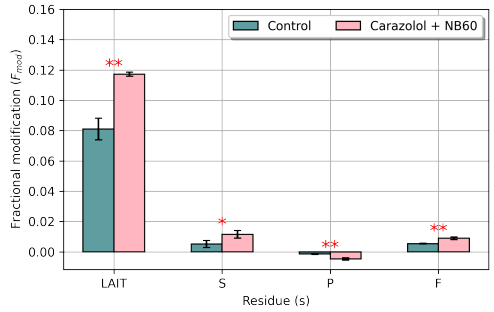
(a) Carbene footprinting histogram of β_1AR alone (control), and with carazolol and Nb60 with a chymotrypsin digestion. The fractional modification of each chymotryptic β_1AR peptide is shown with (pink) and without (turquoise) 100 µM of carazolol and 2:1 molar equivalents of Nb60. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



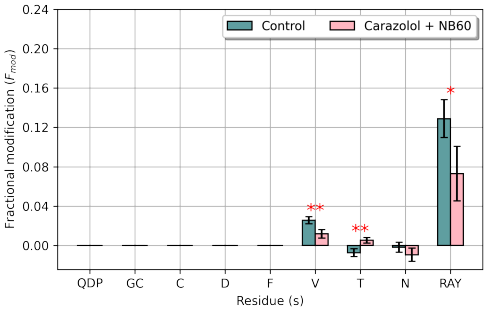
(b) Sub-peptide analysis of labelled GATLVVRGTW (peptide 101-110).



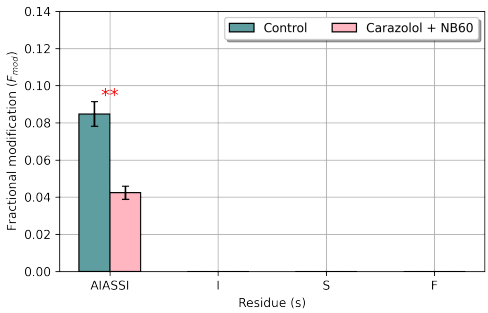
(c) Sub-peptide analysis of labelled TLCVIAIDRY (peptide 134-143).



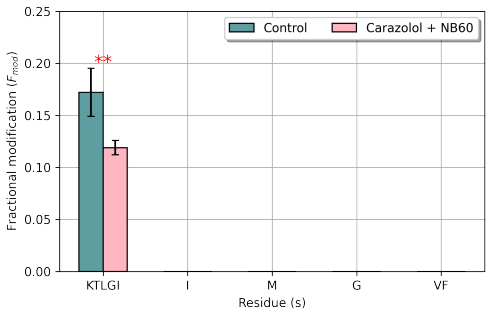
(d) Sub-peptide analysis of labelled LAITSPF (peptide 144-150).



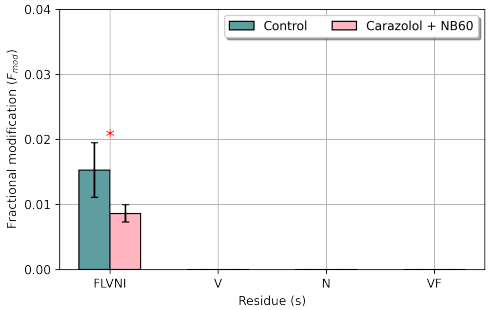
(e) Sub-peptide analysis of labelled QDPGCCDFVTNRAY (peptide 197-210).



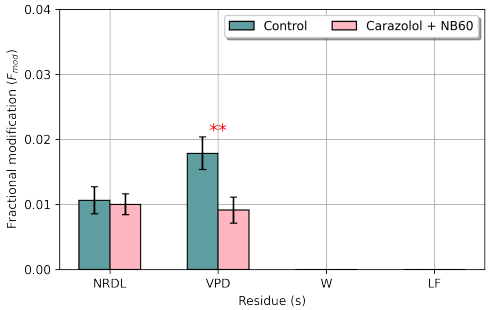
(f) Sub-peptide analysis of labelled AIASSIISF (peptide 211-219).



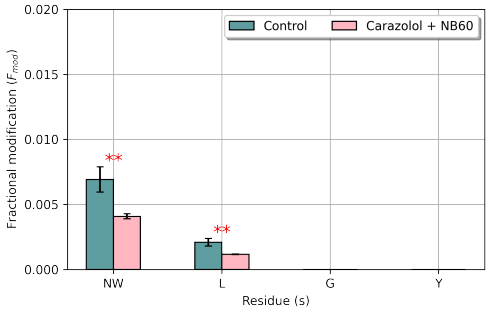
(g) Sub-peptide analysis of labelled KTLGIIMGVF (peptide 265-274).



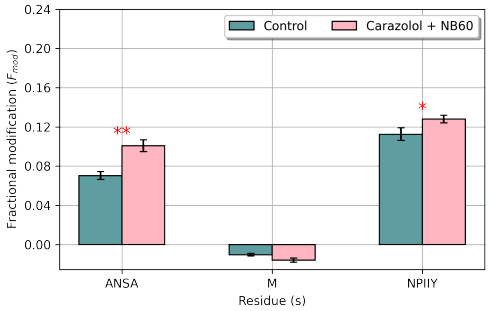
(h) Sub-peptide analysis of labelled FLVNIVNVF (peptide 282-290).



(i) Sub-peptide analysis of labelled NRDLVPDWLF (peptide 291-300).

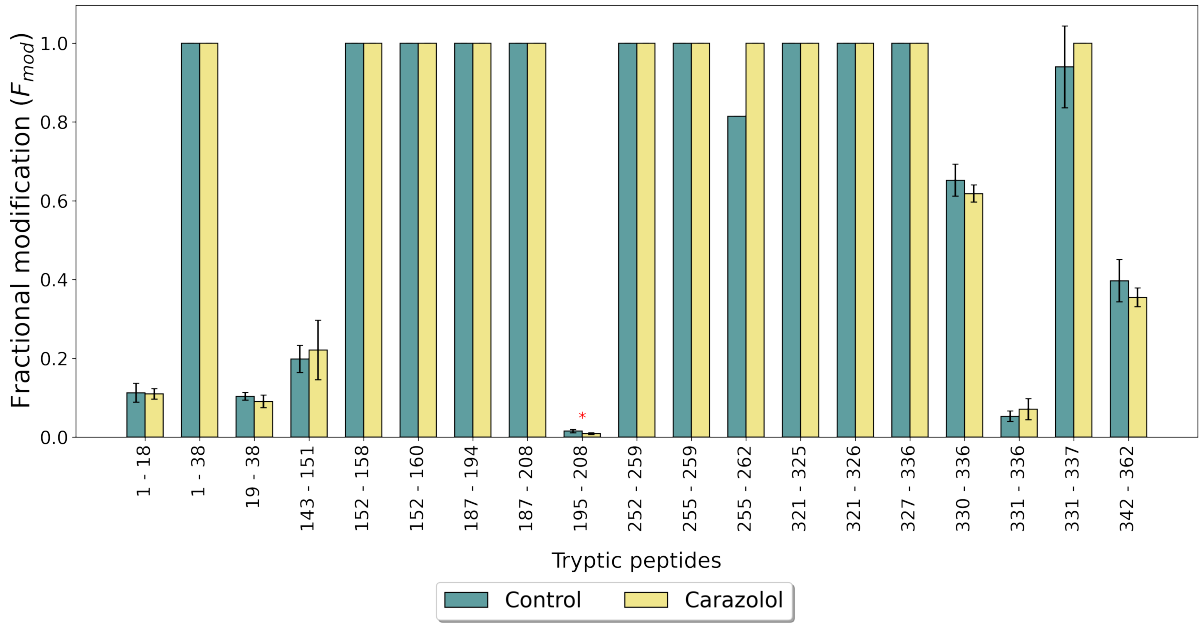


(j) Sub-peptide analysis of labelled NWLGY (peptide 304-308).

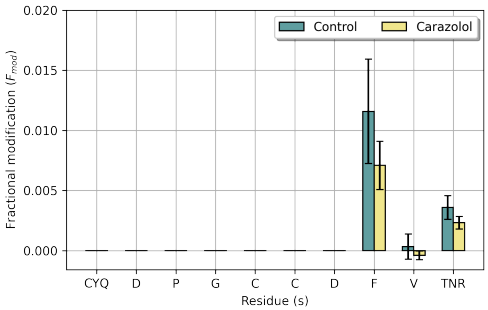


(k) Sub-peptide analysis of labelled ANSAMNPIIY (peptide 309-318).

Figure 6.28 | Carbene footprinting of β_1AR with and without carazolol and Nb60 using a chymotrypsin digestion. Full scan data highlighted chymotryptic peptide-level differences in carbene modification between control and carazolol and Nb60-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.



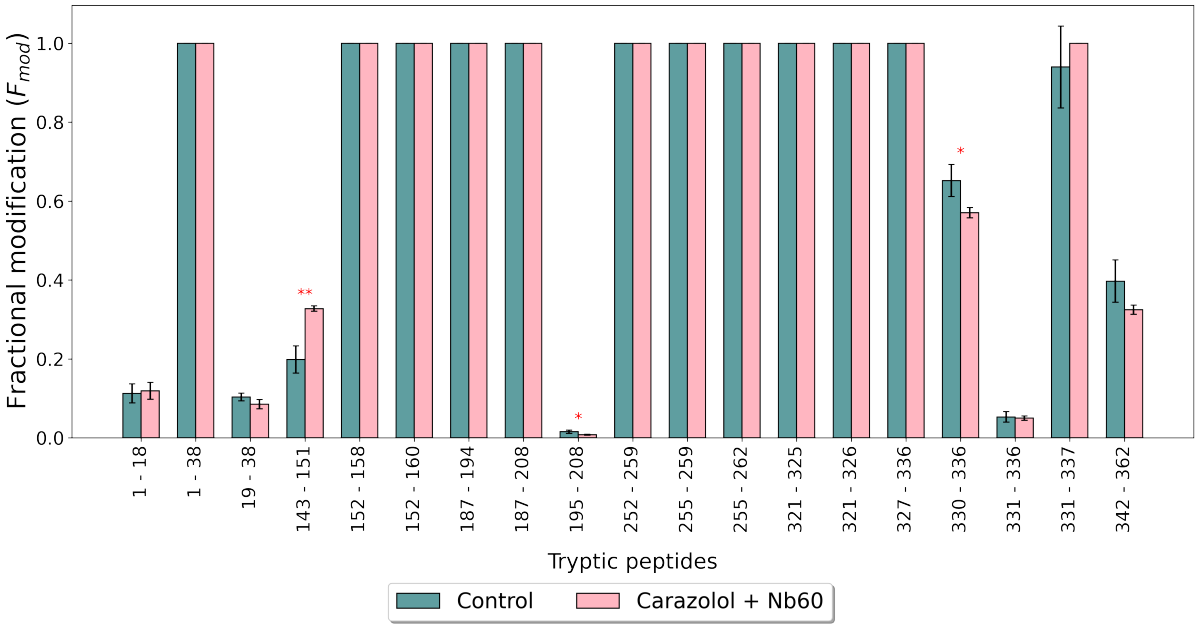
(a) Carbene footprinting histogram of β_1AR alone (control), and with carazolol with a trypsin digestion. The fractional modification of each tryptic β_1AR peptide is shown with (yellow) and without (turquoise) 100 µM of carazolol. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



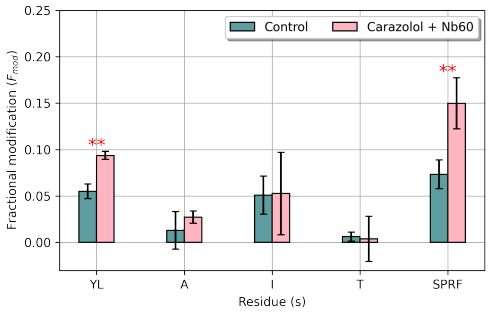
(b) Sub-peptide analysis of labelled CYQDPGCCDFVTNR (peptide 195-208).

Figure 6.29 | Carbene footprinting of β_1AR with and without carazolol using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and carazolol-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

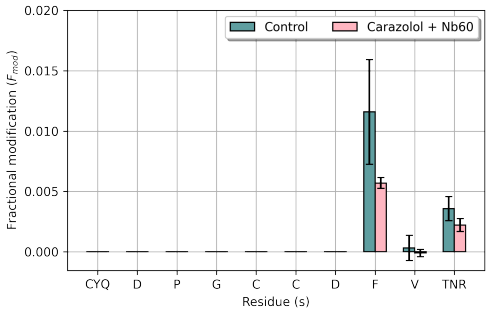
however, this masking event was not observed in the presence of carazolol alone, suggesting additional stabilisation on helix VIII in the inactive ternary complex.



(a) Carbene footprinting histogram of β_1AR alone (control), and with carazolol and Nb60 with a trypsin digestion. The fractional modification of each tryptic β_1AR peptide is shown with (pink) and without (turquoise) 100 µM of carazolol and 2:1 molar equivalents of Nb60. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).



(b) Sub-peptide analysis of labelled YLAITSPFR (peptide 143-151).



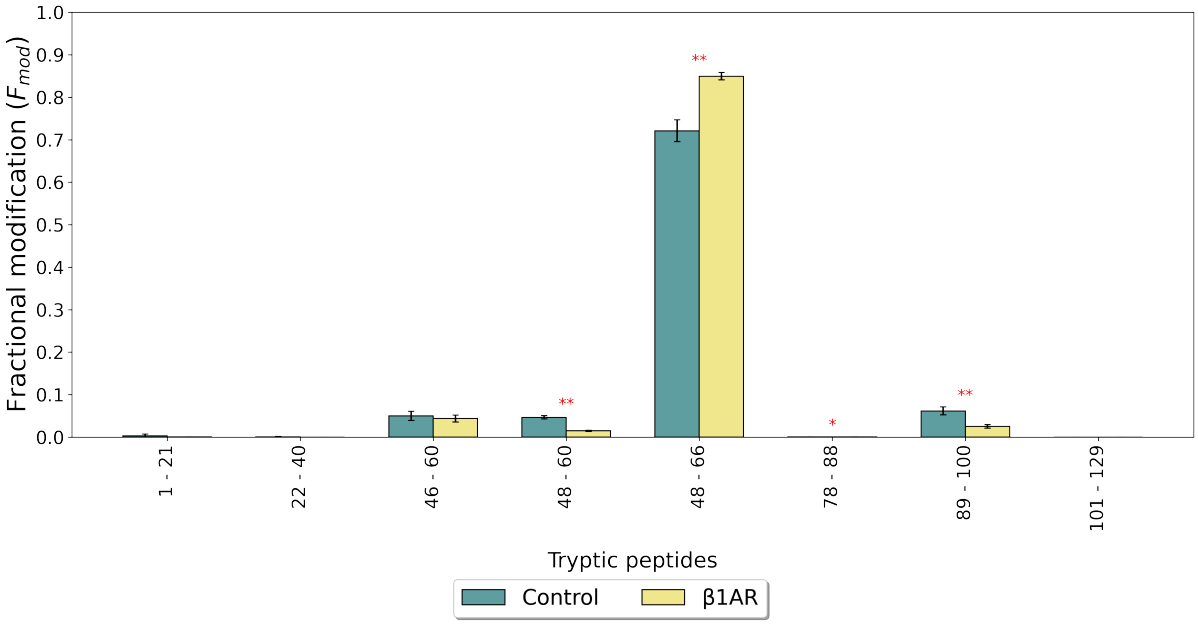
(c) Sub-peptide analysis of labelled CYQDPGCCDFVTNR (peptide 195-208).

Figure 6.30 | Carbene footprinting of β_1AR with and without carazolol and Nb60 using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and carazolol and Nb60-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.

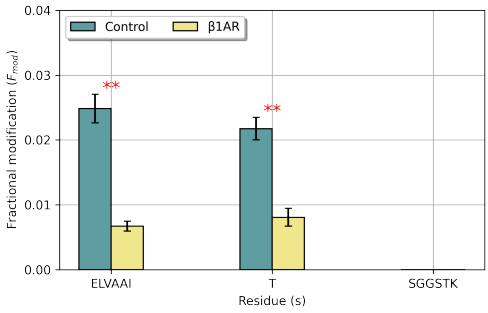
Carbene footprinting was next applied to Nb60 in the presence of either β_1AR or β_1AR and carazolol (Figure S10). Masking effects on Nb60 were expected with β_1AR both in the presence and absence of carazolol, similarly to footprinting of Nb80 (Figure 6.25, Figure 6.26).

Nb60 exhibited three masking effects and one unmasking event in the presence of β_1AR compared to without the receptor (Figure 6.31a). Reductions in labelling were observed on peptides 48-60 (ELVAAITSGGSTK), 78-88 (NTVYLQMNSLK) and 89-100 (AEDTAVYYCNAK) whilst the gain in fractional modification was seen on peptide 48-66 (ELVAAITSGGSTKYADSVK). Significant labelling reductions in the presence of the β_1AR were fewer compared those observed for Nb80. This presumably reflected the difference in binding surfaces between Nb60 and Nb80 to β_1AR. 19 Nb60 residues were located within 5 Å of β_1AR, compared to 27 residues for Nb80. Masking on peptide 48-60 reinforced Nb interaction with β_1AR. MS/MS revealed specific masking effects on residues 48-53 and Thr54 (Figure 6.31b) which were located on and around the CDR2 loop (Figure 6.34). This region is known to contact β_1AR and the observed masking effects likely reflected reduced chemical accessibility due to complex formation. An unmasking effect was located on the missed cleaved peptide 48-66. Given that sub-peptide labelling analysis was conducted on peptide 48-60, I was able to conclude that the unmasking event occurred on the subsequent residues 61-66 (YADSVK). Examination of the crystal structure revealed that these residues were located on a loop, distal from β_1AR, implying a gain in chemical accessibility and labelling due to a probable conformational change associated with β_1AR-Nb binding. Masking of peptide 89-100 also reiterated Nb60-β_1AR binding. This peptide was located on β-strand B, adjacent to the CDR2 loop. Pleasingly MS/MS analysis revealed that carbene modification was located towards the C-terminal side of the peptide, specifically on residues 96-100 (Figure 6.31c) which were proximal to the β_1AR-Nb60 contact interface. These results again highlighted interaction between Nb60 and β_1AR.

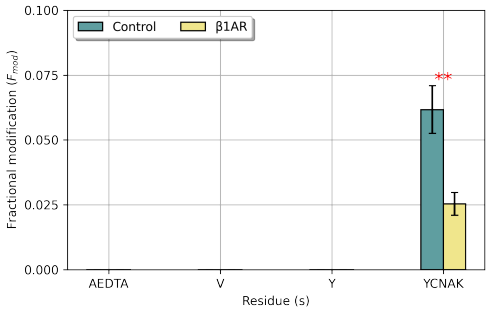
The tryptic digest of Nb60 in the presence of β_1AR and carazolol showed identical labelling changes to Nb60 treated with the receptor alone (Figure 6.31a, Figure 6.32a). This was anticipated since binding of Nbs to βARs is not dependent on small molecules. These results were also consistent with labelling of Nb80.[191]



(a) Carbene footprinting histogram of Nb60 alone (control), and with β_1AR with a trypsin digestion. The fractional modification of each tryptic Nb60 peptide is shown with (yellow) and without (turquoise) 2:1 molar equivalents of β_1AR. Error bars are $\pm$ standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P $<$ 0.01; * = P $<$ 0.05).

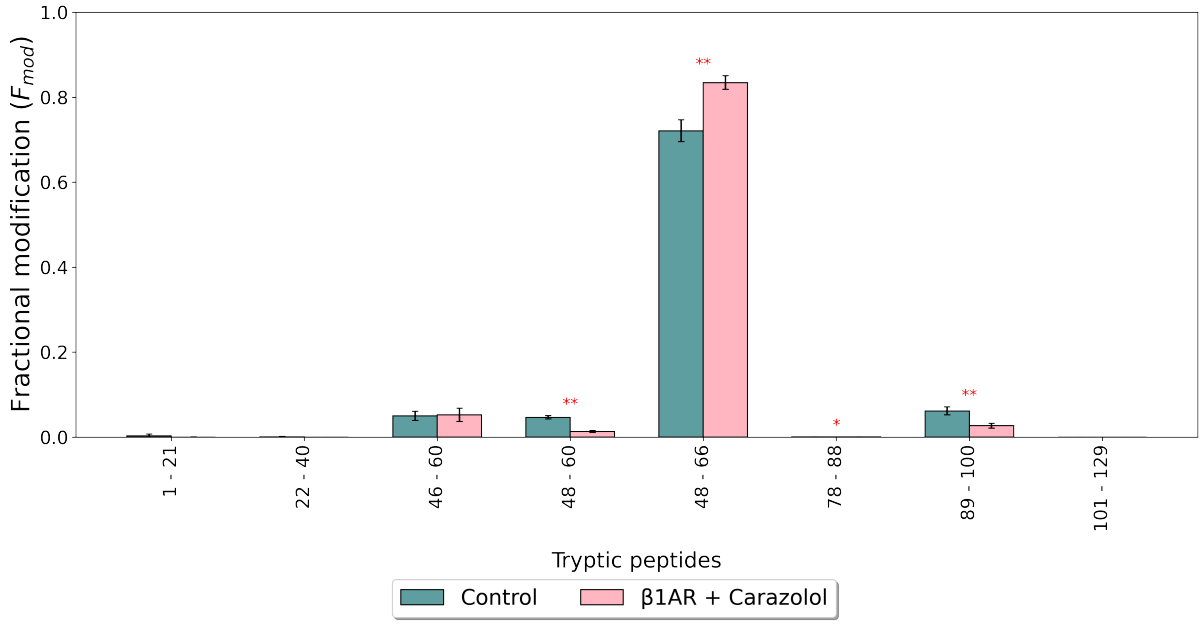


(b) Sub-peptide analysis of labelled ELVAAITSGGSTK (peptide 48-60).

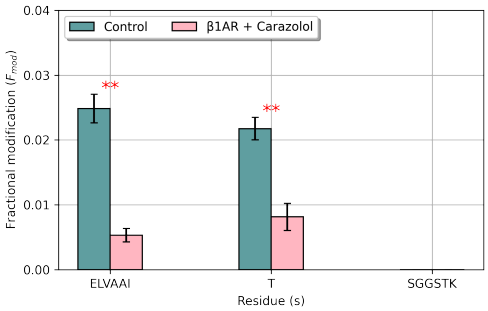


(c) Sub-peptide analysis of labelled AEDTAVYYCNAK (peptide 89-100).

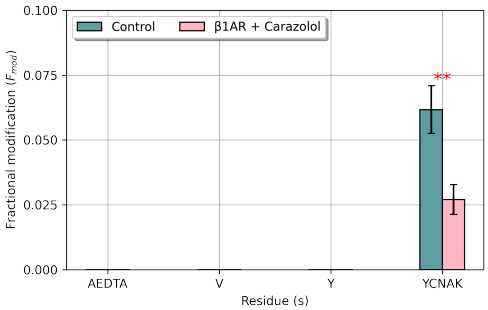
Figure 6.31 | Carbene footprinting of Nb60 with and without β_1AR using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and β_1AR-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.



(a) Carbene footprinting histogram of Nb60 alone (control), and with β_1AR and carazolol with a trypsin digestion. The fractional modification of each tryptic Nb60 peptide is shown with (pink) and without (turquoise) 2:1 molar equivalents of β_1AR and 100 µM of carazolol. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).

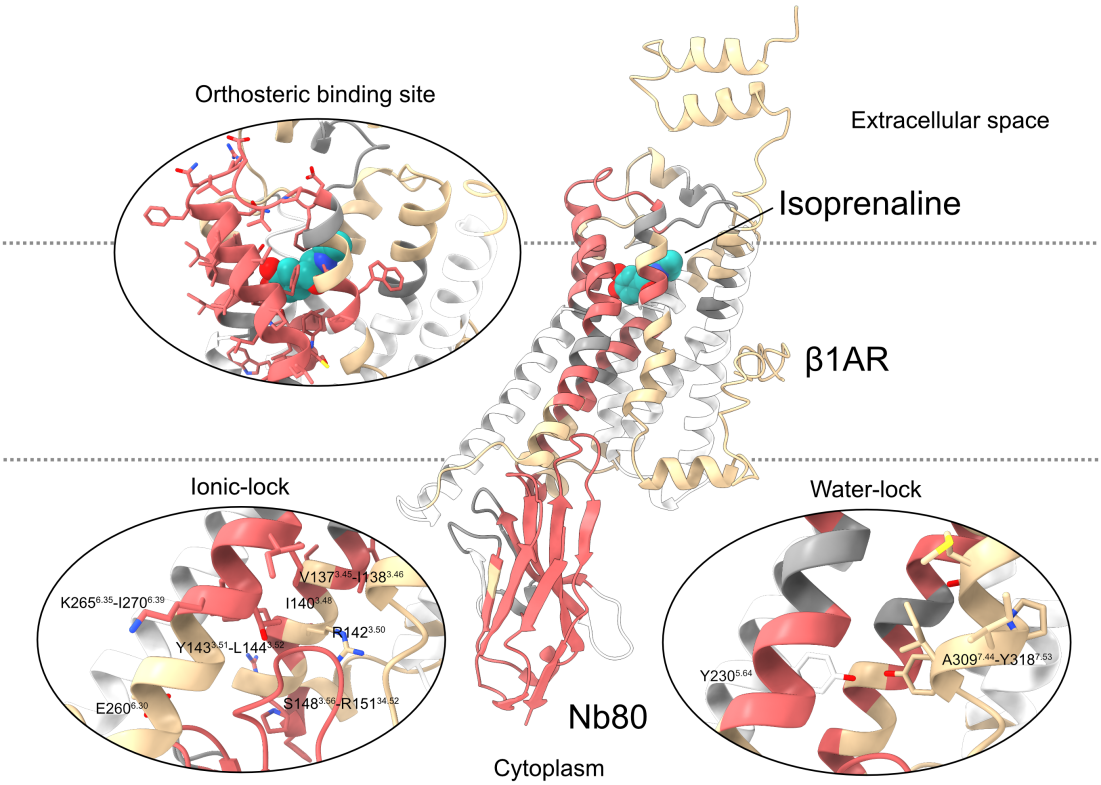


(b) Sub-peptide analysis of labelled ELVAAITSGGSTK (peptide 48-60.)

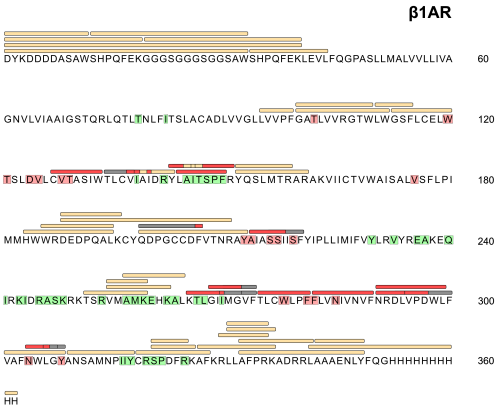


(c) Sub-peptide analysis of labelled AEDTAVYYCNAK (peptide 89-100).

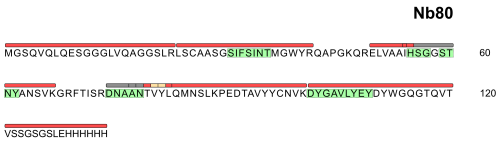
Figure 6.32 | Carbene footprinting of Nb60 with and without β_1AR and carazolol using a trypsin digestion. Full scan data highlighted tryptic peptide-level differences in carbene modification between control and β_1AR and carazolol-treated samples. MS/MS of these labelled peptides revealed higher resolution labelling information.



(a) Combined carbene footprinting data mapped onto the ternary β_1AR-Nb80 complex with isoprenaline bound. Colour scheme is as follows: red = masking effect, tan = no change, grey = no labelling, white = no peptide coverage.

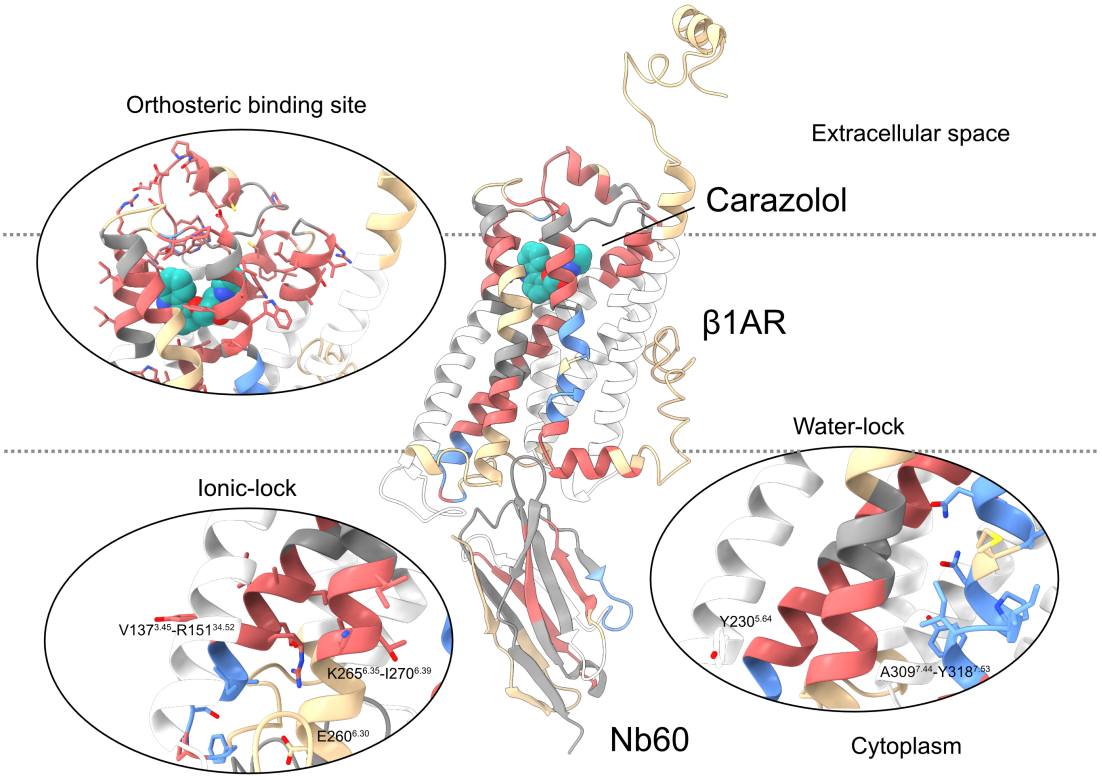


(b) Carbene footprinting data mapped onto the β_1AR sequence (Figure S13). Bars above the sequence represent peptides whilst residues highlighted in red indicate predicted interaction regions with isoprenaline, and residues highlighted in green represent regions within 5 Å of Nb80. Colour scheme is the same as above.

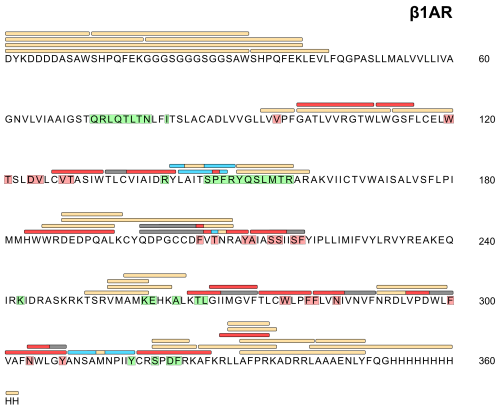


(c) Carbene footprinting data mapped onto the Nb80 sequence. Colour scheme is the same as above.

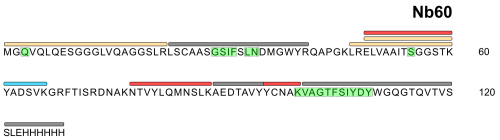
Figure 6.33 | Carbene labelling analysis of the β_1AR-Nb80-isoprenaline complex.



(a) Combined carbene footprinting data mapped onto the ternary β_1AR-Nb60 complex with carazolol bound. Colour scheme is as follows: red = masking effect, blue = unmasking effect, tan = no change, grey = no labelling, white = no peptide coverage.



(b) Carbene footprinting data mapped onto the β_1AR sequence (Figure S15). Bars above the sequence represent peptides whilst residues highlighted in red indicate predicted interaction regions with carazolol, and residues highlighted in green represent regions within 5 Å of Nb60. Colour scheme is the same as above.



(c) Carbene footprinting data mapped onto the Nb60 sequence. Colour scheme is the same as above.

Figure 6.34 | Carbene labelling analysis of the β_1AR-Nb60-carazolol complex.

6.3 Conclusions

In summary, I have applied carbene footprinting mass spectrometry to the study of a GPCR for the first time. The technique was performed on agonist and inverse agonist-bound β_1AR as well as fully active and inactive ternary receptor-Nb complexes. A dual-protease approach was employed to maximise peptide sequence coverage of the receptor and MS/MS was used to highlight sub-peptide level changes in carbene labelling. In each case, carbene footprinting accurately mapped the β_1AR orthosteric binding site. Extended masking was observed around the binding cavity in carazolol-treated samples compared to isoprenaline-treated samples which may have reflected the inverse agonist's bulkier size and expanded contact network. Changes in carbene labelling were also observed on the intracellular side of the receptor in ternary β_1AR-Nb complexes. These differed between isoprenaline and carazolol-treated ternary complexes, reflecting capture of Nb binding and highlighting the conformational range between active and inactive states. Carbene footprinting was also performed on Nbs in the presence of β_1AR and either ligand. Large changes in fractional modification were identified over the surface of either Nb, further reiterating binding of the heavy-chain antibody fragments to the receptor.

This work represents further development to carbene footprinting mass spectrometry and demonstrates the feasibility of using the technique to understand and characterise membrane protein (and in particular, GPCR)-ligand interactions and conformational changes.

7 Conclusions

The work presented in this thesis was aimed at developing carbene footprinting mass spectrometry to interrogate protein interactions. This was carried out by applying the methodology to several clinically relevant protein-ligand and protein-protein systems.

Carbene footprinting was performed on the full-length eIF4A helicase construct in the presence and absence of the natural product hippuristanol from the sea sponge *Isis hippuris*. This led to distinct masking effects within the helicase's CTD that were indicative of ligand binding, despite lacking nearly 33 % total sequence coverage, highlighting the dependence of the technique on peptide coverage, and the need in future studies to exhaustively optimise proteolytic digestion conditions. Nevertheless, despite such shortcomings, my results were in agreement with previous NMR research that showed interaction between the ligand and CTD construct where several masked peptides from the carbene footprinting study were shown to contain residues that had been identified to directly interact with the ligand, or that were at least 5 Å from it. This reinforced ligand binding at the CTD on a full-length construct for the first time. Unmasking events were also observed on linker loop and NTD peptides in the presence of hippuristanol. The presence of overlapping peptides caused by missed cleavages enabled higher resolution labelling data to be inferred without the need for targeted MS/MS approaches; however, the majority of tryptic peptides did not overlap which limited labelling resolution to the peptide level. Furthermore, observed tryptic peptides were often tens of residues long which impeded a detailed analysis of ligand interaction sites and highlighted the need in future studies to employ MS/MS to

improve labelling resolution. Nevertheless, unmasking events were rationalised as local regions of allosteric change with heightened chemical accessibility that had not been identified in the previous NMR study. Taken together, our results were representative of hippuristanol binding at the eIF4A CTD and subsequent conformational change to a closed state.

Carbene footprinting was applied to the full-length hGSDMD pore-forming protein and catalytically inactive form of the cysteine protease hCaspase-1 (C285A). Aspects of the carbene footprinting workflow were extensively optimised for hGSDMD and hCaspase-1 (C285A). Separate enzymatic digestions with Glu-C and trypsin were employed on hGSDMD to provide improved total sequence coverage over the protein compared to a single enzyme alone. The differential carbene footprinting study was conducted on both hGSDMD and hCaspase-1 (C285A). Carbene footprinting results highlighted exosite and catalytic site interactions. MS/MS was also employed to fragment peptides displaying labelling differences. This permitted sub-peptide labelling resolution, often at the residue level. These high resolution labelling differences displayed remarkable agreement with the available crystal structures. However, two hGSDMD peptides displayed masking effects at the exosite region beyond peptides expected to be involved in this region (as identified from hGSDMD-CTD/hCaspase-1 crystal structures). Whilst these results may have been explained by local regions of conformational stabilisation caused by protein-protein interactions, masking effects beyond the exact binding site, caused by the size of the probe could not be disregarded. Other mass spectrometry-based labelling techniques that employ smaller probes, like HRPF, HDX and residue-specific labelling agents (e.g. glycine ethyl ester), would unlikely suffer from this effect. With that said, the size of NaTDB and its chemical substituents are responsible for its easily discernible mass shifts and improved diazirine solubility and overall labelling efficiency, respectively. Certain regions of hCaspase-1 displayed minimal carbene modification, despite labelling optimisation experiments. Poor carbene labelling of the N-terminus of the hCaspase-1 (C285A) p10 subunit meant that interactions at the exosite remained unidentified; however, by performing the footprinting study on both proteins (i.e. hGSDMD $\pm$Caspase-1 and Caspase $\pm$hGSDMD) masking effects were still observed on hGSDMD at this interaction site. Circumventing pitfalls in this way may not

always be possible (i.e. in a differential protein-ligand system, as seen with hCaspase-1(C285A)-VRT-043198 in which a single protein is being studied) and it seems that a comprehensive labelling optimisation experiment which includes alternative diazirine probes, probe concentrations, irradiation times and protein buffers should be considered prior to a differential study. Nevertheless, optimisation of digestion and labelling conditions may only prove *so* useful, and a protein system that displays poor chemical accessibility and/or infrequent proteolytic cleavage sites will limit informative analysis of interaction sites. Fortunately in the case of hCaspase-1/VRT-043198, identification of binding interactions was still possible. These results highlighted VRT-043198 binding at the same location to that of the covalent-bound structure which reinforced the significance of shape complementarity in ligand binding and importantly, suggested that new non-covalent hCaspase-1 inhibitors could be developed. One principle not considered in this study was the effect of the buffer or ligand on the photochemical reagent and in particular, whether a selective modification to these molecules occurred. Indeed, local quenching of the probe by the ligand would be particularly problematic for a footprinting study if it led to additional masking events at a binding site. Slight reduction in carbene labelling was observed in the presence of DMSO during during the VRT-043198 study and a systematic analysis of the effects of different protein/ligand buffers on protein labelling would be beneficial for future labelling studies.

Carbene footprinting was applied to PfMATE. This system was used to inform labelling and digestion conditions for an α-helical membrane protein for the first time. Pertinently, carbene modification was observed on transmembrane spanning regions of the transporter albeit at low levels. This was in agreement with observations that detergent micelles were only modified at tail groups which was also what Manzi and colleagues observed with modification of OG.[100] Higher NaTDB concentrations than typically used for labelling of globular proteins were required for satisfactory modification of PfMATE, presumably due to quenching from detergent molecules. However, the poor overall labelling and sequence coverage meant that any further differential studies would be severely limited in the scope of their analysis. These results were in contrast to those by Manzi and colleagues who showed high sequence coverage and covalent modification of the β-barrel. This presumably reflected the increased

chemical accessibility and number of tryptic cleavage sites in β-barrels compared to α-helical membrane proteins. Indeed, tight packing of the 12 transmembrane-spanning helices was primarily thought to limit carbene modification, and as discussed in the above paragraph, such protein systems with low chemical accessibility and poor sequence coverage may not be suited for such study; nevertheless, this work proved valuable in highlighting such caveats but also in developing a workflow to improve analysis of this challenging family of proteins.

Carbene footprinting was applied to β_1AR in complex with small molecule ligands isoprenaline and carazolol, and their respective intracellular binding partners Nb80 or Nb60. Separate digestion with trypsin and chymotrypsin was employed to maximise sequence coverage of the receptor, similarly to optimised proteolysis of hGSDMD and PfMATE. The observed β_1AR sequence coverage was greater than that observed by West and colleagues in their HDX analysis of β_2AR.[82] The irreversible nature of covalent labelling techniques permitted access to a wide variety of proteases (compared to HDX where proteases are limited by digestion pH and temperature), potentiating high sequence coverage by pooling together peptides from separate digestions. Nevertheless the marginal improvements in sequence coverage were met with a significant increase in workload, and it appeared that prioritising sequence coverage through this multi-protease approach may not always be worth the concomitant increase in sample preparation, sample consumption and data analysis, especially if peptides are not located in 'informative' regions of the protein. Indeed, sample consumption was a pertinent issue in this study where proteolytic digestion of the β_1AR complexes with more than two proteases (in addition to trypsin and chymotrypsin) was not sustainable. Furthermore preliminary analysis of observed β_1AR peptides showed coverage in the extracellular and intracellular binding cavities, alleviating the need for further protease digestions. Covalent modification of β_1AR was higher than observed with the previously studied α-helical membrane protein PfMATE which was presumably due to increased overall chemical accessibility compared to the MATE transporter. However, carbene modification of the receptor dropped was lower at 40 mM NaTDB compared with 20 mM NaTDB. No rationale was provided for this observation and it appears that further research into the mechanism behind carbene modification is required. One possibility was that the improved modification of β_1AR compared to PfMATE led to a

lower labelling saturation concentration of the receptor where higher concentrations of NaTDB above this level led to multiple label insertions and reductions in the amount of singly labelled peptides. However, multiply-labelled peptides (i.e. +404 Da) were searched for during the analysis but none were observed, and more work into this phenomenon should be conducted. Pleasingly, distinct carbene footprinting profiles were observed for each of the different β_1AR complexes. Consistent masking effects were observed at extracellular binding cavity in ligand-treated samples, irrespective of whether an intracellular binding partner was present. This was indicative of small molecule binding; however, extended masking at the extracellular region was observed with carazolol compared to isoprenaline which was attributed to the slightly larger size of the inverse agonist compared to the agonist, as well as additional regions of stabilisation primarily induced on extracellular loops. Stabilisation effects in the presence of carazolol were also seen on TM3/ICL2 and TM6. These observations were also seen reported by West and colleagues,[82] suggesting that carbene footprinting could be used to detect subtle allosteric changes in the GPCR. This notion was reinforced by differential study of ternary active and inactive Nb complexes where masking and unmasking effects were attributed to Nb binding, well known molecular switches and local regions of masking/unmasking. Interestingly, the majority of overlapping peptides from chymotryptic and tryptic digests showed congruent labelling effects, further highlighting the accuracy with the technique can report on. However, as would be expected from the mediocre sequence coverage, analysis of Nb binding was partially limited. Fortunately, differential footprinting on the Nbs themselves showed masking effects that were indicative of receptor binding. This study represents the first use of carbene footprinting to report on GPCR interactions and dynamics.

The findings presented in this body of work have significantly contributed to the development of carbene footprinting mass spectrometry as a method to interrogate protein interactions.

References

1. Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17*, DOI:10.3390/ijms17020144.
2. Aebersold, R. *et al.* How many human proteoforms are there? *Nat. Chem. Biol.* **2018**, *14*, 206–214.
3. Omenn, G. S. *et al.* Research on the Human Proteome Reaches a Major Milestone: >90% of Predicted Human Proteins Now Credibly Detected, According to the HUPO Human Proteome Project. *J. Proteome Res.* **2020**, *19*, 4735–4746.
4. Kuzmanov, U.; Emili, A. Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Med.* **2013**, *5*, 37.
5. Yang, L.-Q.; Sang, P.; Tao, Y.; Fu, Y.-X.; Zhang, K.-Q.; Xie, Y.-H.; Liu, S.-Q. Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn.* **2014**, *32*, 372–393.
6. Chen, K.; Kurgan, L. A.; Ruan, J. Prediction of flexible/rigid regions from protein sequences using k-spaced amino acid pairs. *BMC Struct. Biol.* **2007**, *7*, 25.
7. Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18908–18913.
8. Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 386–391.
9. Schmitt, C.; Lippert, A. H.; Bonakdar, N.; Sandoghdar, V.; Voll, L. M. Compartmentalization and Transport in Synthetic Vesicles. *Front. Bioeng. Biotechnol.* **2016**, *4*, DOI:10.3389/fbioe.2016.00019.
10. Opekarová, M.; Tanner, W. Specific lipid requirements of membrane proteins—a putative bottleneck in heterologous expression. *Biochim. Biophys. Acta Biomembr.* **2003**, *1610*, 11–22.
11. Chou, K.-C.; Elrod, D. W. Prediction of membrane protein types and subcellular locations. *Proteins* **1999**, *34*, 137–153.
12. Gonzalez, M. W.; Kann, M. G. Translational Bioinformatics (Chapter 4: Protein Interactions and Disease). *PLoS Comput. Biol.* **2012**, *8*, DOI:10.1371/journal.pcbi.1002819.
13. González-Ruiz, D.; Gohlke, H. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr. Med. Chem.* **2006**, *13*, 2607–2625.
14. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug. Discov.* **2006**, *5*, 993–996.
15. Iii, Y.; R, J. A century of mass spectrometry: from atoms to proteomes. *Nat. Methods* **2011**, *8*, 633–637.

16. Gräslund, S. *et al.* Protein production and purification. *Nat. Methods* **2008**, *5*, 135–146.

17. Edwards, A. M.; Arrowsmith, C. H.; Christendat, D.; Dharamsi, A.; Friesen, J. D.; Greenblatt, J. F.; Vedadi, M. Protein production: feeding the crystallographers and NMR spectroscopists. *Nat. Struct. Mol. Biol.* **2000**, *7*, 970–972.

18. Bornhorst, J. A.; Falke, J. J. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* **2000**, *326*, 245–254.

19. Seddon, A. M.; Curnow, P.; Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta Biomembr.* **2004**, *1666*, 105–117.

20. Bolla, J. R.; Agasid, M. T.; Mehmood, S.; Robinson, C. V. Membrane Protein–Lipid Interactions Probed Using Mass Spectrometry. *Annu. Rev. Biochem.* **2019**, *88*, 85–111.

21. Eschweiler, J. D.; Kerr, R.; Rabuck-Gibbons, J.; Ruotolo, B. T. Sizing Up Protein–Ligand Complexes: The Rise of Structural Mass Spectrometry Approaches in the Pharmaceutical Sciences. *Annu. Rev. Anal. Chem.* **2017**, *10*, 25–44.

22. McPherson, A.; Gavira, J. A. Introduction to protein crystallization. *Acta. Crystallogr. F. Struct. Biol. Commun.* **2013**, *70*, 2–20.

23. Cala, O.; Guillière, F.; Krimm, I. NMR-based analysis of protein–ligand interactions. *Anal. Bioanal. Chem.* **2014**, *406*, 943–956.

24. Bothwell, J. H. F.; Griffin, J. L. An introduction to biological nuclear magnetic resonance spectroscopy. *Biol. Rev.* **2011**, *86*, 493–510.

25. Marion, D. An Introduction to Biological NMR Spectroscopy. *Mol. Cell. Proteomics* **2013**, *12*, 3006–3025.

26. Ceska, T.; Chung, C.-W.; Cooke, R.; Phillips, C.; Williams, P. A. Cryo-EM in drug discovery. *Biochem. Soc. Trans.* **2019**, *47*, 281–293.

27. Milne, J. L. S.; Borgnia, M. J.; Bartesaghi, A.; Tran, E. E. H.; Earl, L. A.; Schauder, D. M.; Lengyel, J.; Pierson, J.; Patwardhan, A.; Subramaniam, S. Cryo-electron microscopy: A primer for the non-microscopist. *FEBS J.* **2013**, *280*, 28–45.

28. Birch, J.; Axford, D.; Foadi, J.; Meyer, A.; Eckhardt, A.; Thielmann, Y.; Moraes, I. The fine art of integral membrane protein crystallisation. *Methods* **2018**, *147*, 150–162.

29. Thonghin, N.; Kargas, V.; Clews, J.; Ford, R. C. Cryo-electron microscopy of membrane proteins. *Methods* **2018**, *147*, 176–186.

30. Opella, S. J.; Marassi, F. M. Applications of NMR to membrane proteins. *Arch. Biochem. Biophys.* **2017**, *628*, 92–101.

31. Pacholarz, K. J.; Garlish, R. A.; Taylor, R. J.; Barran, P. E. Mass spectrometry based tools to investigate protein–ligand interactions for drug discovery. *Chem. Soc. Rev.* **2012**, *41*, 4335–4355.

32. Savaryn, J. P.; Toby, T. K.; Kelleher, N. L. A researcher's guide to mass spectrometry-based proteomics. *Proteomics* **2016**, *16*, 2435–2443.

33. Stroobant, E. d. H. V. *Mass Spectrometry: Principles and Applications*, third edition ed.; John Wiley & Sons, Ltd: England, 2012.

34. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85*, 2–9.

35. Mora, J. F. d. l.; Van Berkel, G. J.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. Electrochemical processes in electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2000**, *35*, 939–952.

36. Banerjee, S.; Mazumdar, S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* **2012**, *2012*, 282574.

37. McCullagh, J.; Oldham, N. *Mass Spectrometry*, first edition ed.; Oxford University Press: Oxford, 2019.

38. Kaltashov, I. A.; Abzalimov, R. R. Do ionic charges in ESI MS provide useful information on macromolecular structure? *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1239–1246.

39. Metwally, H.; McAllister, R. G.; Konermann, L. Exploring the mechanism of salt-induced signal suppression in protein electrospray mass spectrometry using experiments and molecular dynamics simulations. *Anal. Chem.* **2015**, *87*, 2434–2442.

40. Konermann, L. Addressing a Common Misconception: Ammonium Acetate as Neutral pH "Buffer" for Native Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1827–1835.

41. Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. Principles of Fourier transform ion cyclotron resonance mass spectrometry and its application in structural biology. *Analyst* **2005**, *130*, 18–28.

42. Marshall, A. G.; Grosshans, P. B. Fourier transform ion cyclotron resonance mass spectrometry: the teenage years. *Anal. Chem.* **1991**, *63*, 215–229.

43. Jertz, R.; Friedrich, J.; Kriete, C.; Nikolaev, E. N.; Baykut, G. Tracking the Magnetron Motion in FT-ICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1349–1366.

44. Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A. Fourier Transform Mass Spectrometry. *Mol. Cell Proteomics* **2011**, *10*, M111.009431.

45. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.

46. Kooijman, P. C.; Nagornov, K. O.; Kozhinov, A. N.; Kilgour, D. P. A.; Tsybin, Y. O.; Heeren, R. M. A.; Ellis, S. R. Increased throughput and ultra-high mass resolution in DESI FT-ICR MS imaging through new-generation external data acquisition system and advanced data processing approaches. *Sci. Rep.* **2019**, *9*, 8.

47. Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007**, *7*, 340–350.

48. Wells, J. M.; McLuckey, S. A. Collision-induced dissociation (CID) of peptides and proteins. *Methods. Enzymol.* **2005**, *402*, 148–185.

49. Lifshitz, C. Intramolecular energy redistribution in polyatomic ions. *J. Phys. Chem.* **1983**, *87*, 2304–2313.

50. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399–1406.

51. Pramanik, B., Chen, G., Gross, M. L., Eds. *Protein and Peptide Mass Spectrometry in Drug Discovery*, 1st ed.; John Wiley & Sons, Ltd: Hoboken, N.J, 2011.

52. Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88*, 30–51.

53. Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* **2018**, *37*, 750–771.

54. Sleno, L.; Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **2004**, *39*, 1091–1112.

55. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.

56. Bakhtiar, R.; Guan, Z. Electron Capture Dissociation Mass Spectrometry in Characterization of Peptides and Proteins. *Biotechnol. Lett.* **2006**, *28*, 1047–1059.

57. Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **2020**, *8*, 14.

58. Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **2009**, *138*, 795–806.

59. Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond trypsin. *FEBS J* **2015**, *282*, 2612–2626.

60. Goodman, J. K.; Zampronio, C. G.; Jones, A. M. E.; Hernandez-Fernaud, J. R. Updates of the In-Gel Digestion Method for Protein Analysis by Mass Spectrometry. *Proteomics* **2018**, *18*, e1800236.

61. Müller, T.; Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Mol. Cell Proteomics.* **2017**, *16*, 1173–1187.

62. Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nat. Protoc.* **2016**, *11*, 993–1006.

63. Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.

64. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102*, 4501–4524.

65. Brik, A.; Wong, C.-H. HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* **2003**, *1*, 5–14.

66. Raufman, J.-P. In *Encyclopedia of Gastroenterology*; Johnson, L. R., Ed.; Elsevier: New York, 2004; pp 147–148.

67. Schomburg, I.; Chang, A.; Schomburg, D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* **2002**, *30*, 47–49.

68. Keil, B. *Specificity of Proteolysis*, first edition ed.; Springer Publishing: Berlin, 2012.

69. Okoniewska, M.; Tanaka, T.; Yada, R. The role of the flap residue, threonine 77, in the activation and catalytic activity of pepsin A. *Protein Eng. Des. Sel.* **1999**, *12*, 55–61.

70. Wang, L.; Chance, M. R. Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Mol. Cell Proteomics* **2017**, *16*, 706–716.

71. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert Rev. Proteomics* **2011**, *8*, 43–59.

72. Engen, J. R. Analysis of Protein Conformation and Dynamics by Hydrogen/Deuterium Exchange MS. *Anal. Chem.* **2009**, *81*, 7870–7875.

73. Paterson, Y.; Englander, S. W.; Roder, H. An Antibody Binding Site on Cytochrome C Defined by Hydrogen Exchange and Two-Dimensional NMR. *Science* **1990**, *249*, 755–759.

74. Katta, V.; Chait, B. T.; Carr, S. Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 214–217.

75. Zhang, Z.; Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. *Protein Sci.* **1993**, *2*, 522–531.

76. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **2019**, *16*, 595–602.

77. Lu, G.; Xu, X.; Li, G.; Sun, H.; Wang, N.; Zhu, Y.; Wan, N.; Shi, Y.; Wang, G.; Li, L.; Hao, H.; Ye, H. Sub-residue Resolution Footprinting of Ligand-Protein Interactions by Carbene Chemistry and Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 947–956.

78. Samodova, D.; Hosfield, C. M.; Cramer, C. N.; Giuli, M. V.; Cappellini, E.; Franciosa, G.; Rosenblatt, M. M.; Kelstrup, C. D.; Olsen, J. V. ProAlanase is an Effective Alternative to Trypsin for Proteomics Applications and Disulfide Bond Mapping. *Mol. Cell Proteomics* **2020**, *19*, 2139–2157.

79. Rey, M.; Yang, M.; Burns, K. M.; Yu, Y.; Lees-Miller, S. P.; Schriemer, D. C. Nepenthesin from Monkey Cups for Hydrogen/Deuterium Exchange Mass Spectrometry. *Mol. Cell Proteomics* **2013**, *12*, 464–472.

80. Ferguson, P. L.; Pan, J.; Wilson, D. J.; Dempsey, B.; Lajoie, G.; Shilton, B.; Konermann, L. Hydrogen/deuterium scrambling during quadrupole time-of-flight MS/MS analysis of a zinc-binding protein domain. *Anal. Chem.* **2007**, *79*, 153–160.

81. Hamuro, Y.; Coales, S. J.; Morrow, J. A.; Molnar, K. S.; Tuske, S. J.; Southern, M. R.; Griffin, P. R. Hydrogen/deuterium-exchange (H/D-Ex) of PPAR LBD in the presence of various modulators. *Protein Sci.* **2006**, *15*, 1883–1892.

82. West, G. M.; Chien, E. Y. T.; Katritch, V.; Gatchalian, J.; Chalmers, M. J.; Stevens, R. C.; Griffin, P. R. Ligand-dependent perturbation of the conformational ensemble for the GPCR beta2 adrenergic receptor revealed by HDX. *Structure* **2011**, *19*, 1424–1432.

83. Zhang, X.; Chien, E. Y.; Chalmers, M. J.; Pascal, B. D.; Gatchalian, J.; Stevens, R. C.; Griffin, P. R. Dynamics of the Beta2-adrenergic G-protein coupled receptor revealed by hydrogen-deuterium exchange. *Anal. Chem.* **2010**, *82*, 1100–1108.

84. Yang, L. *et al.* Conformational states of the full-length glucagon receptor. *Nat. Commun.* **2015**, *6*, 7859.

85. Kish, M.; Smith, V.; Lethbridge, N.; Cole, L.; Bond, N. J.; Phillips, J. J. Online Fully Automated System for Hydrogen/Deuterium-Exchange Mass Spectrometry with Millisecond Time Resolution. *Anal. Chem.* **2023**, *95*, 5000–5008.

86. Seetaloo, N.; Zacharopoulou, M.; Stephens, A. D.; Kaminski Schierle, G. S.; Phillips, J. J. Millisecond Hydrogen/Deuterium-Exchange Mass Spectrometry Approach to Correlate Local Structure and Aggregation in -Synuclein. *Anal. Chem.* **2022**, *94*, 16711–16719.

87. Hambly, D.; Gross, M. Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin. *Int. J. Mass Spectrom.* **2007**, *259*, 124–129.

88. Li, K. S.; Shi, L.; Gross, M. L. Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Acc. Chem. Res.* **2018**, *51*, 736–744.

89. Johnson, D. T.; Di Stefano, L. H.; Jones, L. M. Fast photochemical oxidation of proteins (FPOP): A powerful mass spectrometry-based structural proteomics tool. *J. Biol. Chem.* **2019**, *294*, 11969–11979.

90. Zhang, B.; Rempel, D. L.; Gross, M. L. Protein Footprinting by Carbenes on a Fast Photochemical Oxidation of Proteins (FPOP) Platform. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 552–555.

91. Yan, Y.; Chen, G.; Wei, H.; Huang, R. Y.-C.; Mo, J.; Rempel, D. L.; Tymiak, A. A.; Gross, M. L. Fast Photochemical Oxidation of Proteins (FPOP) Maps the Epitope of EGFR Binding to Adnectin. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 2084–2092.

92. Sheng, Y.; Capri, J.; Waring, A.; Valentine, J. S.; Whitelegge, J. Exposure of Solvent-Inaccessible Regions in the Amyloidogenic Protein Human SOD1 Determined by Hydroxyl Radical Footprinting. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 218–226.

93. Du, Y. *et al.* Assembly of a GPCR-G Protein Complex. *Cell* **2019**, *177*, 1232–1242.

94. Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. High-resolution mapping of carbene-based protein footprints. *Anal. Chem.* **2012**, *84*, 4411–4418.

95. Bourissou, D.; Guerret, O.; Gabbaï, F. P.; Bertrand, G. Stable Carbenes. *Chem. Rev.* **2000**, *100*, 39–92.

96. West, A. V.; Muncipinto, G.; Wu, H.-Y.; Huang, A. C.; Labenski, M. T.; Jones, L. H.; Woo, C. M. Labeling Preferences of Diazirines with Protein Biomolecules. *J. Am. Chem. Soc.* **2021**, *143*, 6691–6700.

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

98. Jumper, C. C.; Schriemer, D. C. Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Anal. Chem.* **2011**, *83*, 2913–2920.

99. Manzi, L.; Barrow, A. S.; Scott, D.; Layfield, R.; Wright, T. G.; Moses, J. E.; Oldham, N. J. Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. *Nat. Commun.* **2016**, *7*, 13288.

100. Manzi, L.; Barrow, A. S.; Hopper, J. T. S.; Kaminska, R.; Kleanthous, C.; Robinson, C. V.; Moses, J. E.; Oldham, N. J. Carbene Footprinting Reveals Binding Interfaces of a Multimeric Membrane-Spanning Protein. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 14873–14877.

101. Jenner, M.; Kosol, S.; Griffiths, D.; Prasongpholchai, P.; Manzi, L.; Barrow, A. S.; Moses, J. E.; Oldham, N. J.; Lewandowski, J. R.; Challis, G. L. Mechanism of Intersubunit Ketosynthase-Dehydratase Interaction in Polyketide Synthases. *Nat. Chem. Biol.* **2018**, *14*, 270–275.

102. Ziemianowicz, D. S.; Bomgarden, R.; Etienne, C.; Schriemer, D. C. Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2011–2021.

103. Bellamy-Carter, J.; Oldham, N. J. PepFoot: A Software Package for Semiautomated Processing of Protein Footprinting Data. *J. Proteome Res.* **2019**, *18*, 2925–2930.

104. Hogan, J. M.; Lee, P. S.; Wong, S. C.; West, S. M.; Morishige, W. H.; Bee, C.; Tapia, G. C.; Rajpal, A.; Strop, P.; Dollinger, G. Residue-Level Characterization of Antibody Binding Epitopes Using Carbene Chemical Footprinting. *Anal. Chem.* **2023**, *95*, 3922–3931.

105. Barsnes, H.; Vaudel, M. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *J. Proteome Res.* **2018**, *17*, 2552–2555.

106. Bjornson, R. D.; Carriero, N. J.; Colangelo, C.; Shifman, M.; Cheung, K.-H.; Miller, P. L.; Williams, K. X!!Tandem, an Improved Method for Running X!Tandem in Parallel on Collections of Commodity Computers. *J. Proteome Res.* **2008**, *7*, 293–299.

107. Vaudel, M.; Burkhart, J. M.; Zahedi, R. P.; Oveland, E.; Berven, F. S.; Sickmann, A.; Martens, L.; Barsnes, H. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* **2015**, *33*, 22–24.

108. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **2008**, *9*, 40.

109. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738.

110. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381–3385.

111. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82.

112. Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898.

113. Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009**, *97*, 50–58.

114. Lee, J. *et al.* CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **2019**, *15*, 775–786.

115. Cheng, X.; Jo, S.; Lee, H. S.; Klauda, J. B.; Im, W. CHARMM-GUI Micelle Builder for Pure/Mixed Micelle and Protein/Micelle Complex Systems. *J. Chem. Inf. Model.* **2013**, *53*, 2171–2180.

116. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Springer Protocols Handbooks; Humana Press: Totowa, NJ, 2005; pp 571–607.

117. Chang, J. H.; Cho, Y. H.; Sohn, S. Y.; Choi, J. M.; Kim, A.; Kim, Y. C.; Jang, S. K.; Cho, Y. Crystal structure of the eIF4A–PDCD4 complex. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3148–3153.

118. Li, Q.; Imataka, H.; Morino, S.; Rogers, G. W.; Richter-Cook, N. J.; Merrick, W. C.; Sonenberg, N. Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol. Cell. Biol.* **1999**, *19*, 7336–7346.

119. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419.

120. Chan, C. C.; Dostie, J.; Diem, M. D.; Feng, W.; Mann, M.; Rappsilber, J.; Dreyfuss, G. eIF4A3 is a novel component of the exon junction complex. *RNA* **2004**, *10*, 200–209.

121. Jiang, C. *et al.* Targeting the N Terminus of eIF4AI for Inhibition of Its Catalytic Recycling. *Cell Chem. Biol.* **2019**, *26*, 1417–1426.

122. Andersen, C. B. F.; Ballut, L.; Johansen, J. S.; Chamieh, H.; Nielsen, K. H.; Oliveira, C. L. P.; Pedersen, J. S.; Séraphin, B.; Le Hir, H.; Andersen, G. R. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **2006**, *313*, 1968–1972.

123. Sengoku, T.; Nureki, O.; Nakamura, A.; Kobayashi, S.; Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* **2006**, *125*, 287–300.

124. Theissen, B.; Karow, A. R.; Köhler, J.; Gubaev, A.; Klostermeier, D. Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 548–553.

125. Lorsch, J. R.; Herschlag, D. The DEAD Box Protein eIF4A. 2. A Cycle of Nucleotide and RNA-Dependent Conformational Changes. *Biochemistry* **1998**, *37*, 2194–2206.

126. Brito Querido, J.; Sokabe, M.; Kraatz, S.; Gordiyenko, Y.; Skehel, J. M.; Fraser, C. S.; Ramakrishnan, V. Structure of a human 48S translational initiation complex. *Science* **2020**, *369*, 1220–1227.

127. Rogers, G. W.; Richter, N. J.; Lima, W. F.; Merrick, W. C. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.* **2001**, *276*, 30914–30922.

128. Takyar, S.; Hickerson, R. P.; Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **2005**, *120*, 49–58.

129. Jackson, R. J.; Hellen, C. U. T.; Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 113–127.

130. Pestova, T. V.; Hellen, C. U. T. Small molecule derails translation initiation. *Nat. Chem. Biol.* **2006**, *2*, 176–177.

131. Lindqvist, L.; Oberer, M.; Reibarkh, M.; Cencic, R.; Bordeleau, M.-E.; Vogt, E.; Marintchev, A.; Tanaka, J.; Fagotto, F.; Altmann, M.; Wagner, G.; Pelletier, J. Selective Pharmacological Targeting of a DEAD Box RNA Helicase. *PloS ONE* **2008**, *3*, e1583.

132. Cencic, R.; Pelletier, J. Hippuristanol - A potent steroid inhibitor of eukaryotic initiation factor 4A. *Translation* **2016**, *4*, e1137381.

133. Sun, Y.; Atas, E.; Lindqvist, L. M.; Sonenberg, N.; Pelletier, J.; Meller, A. Single-molecule kinetics of the eukaryotic initiation factor 4AI upon RNA unwinding. *Structure* **2014**, *22*, 941–948.

134. Cramer, Z.; Sadek, J.; Vazquez, G. G.; Di Marco, S.; Pause, A.; Pelletier, J.; Gallouzi, I.-E. eIF4A inhibition prevents the onset of cytokine-induced muscle wasting by blocking the STAT3 and iNOS pathways. *Sci. Rep.* **2018**, *8*, 8414.

135. Cencic, R.; Robert, F.; Galicia-Vázquez, G.; Malina, A.; Ravindar, K.; Somaiah, R.; Pierre, P.; Tanaka, J.; Deslongchamps, P.; Pelletier, J. Modifying chemotherapy response by targeted inhibition of eukaryotic initiation factor 4A. *Blood Cancer J.* **2013**, *3*, e128.

136. Ishikawa, C.; Tanaka, J.; Katano, H.; Senba, M.; Mori, N. Hippuristanol reduces the viability of primary effusion lymphoma cells both in vitro and in vivo. *Mar. Drugs* **2013**, *11*, 3410–3424.

137. He, W.-t.; Wan, H.; Hu, L.; Chen, P.; Wang, X.; Huang, Z.; Yang, Z.-H.; Zhong, C.-Q.; Han, J. Gasdermin D is an executor of pyroptosis and required for interleukin-1Beta secretion. *Cell Res.* **2015**, *25*, 1285–1298.

138. Shi, J.; Zhao, Y.; Wang, K.; Shi, X.; Wang, Y.; Huang, H.; Zhuang, Y.; Cai, T.; Wang, F.; Shao, F. Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* **2015**, *526*, 660–665.

139. Kayagaki, N. *et al.* Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature* **2015**, *526*, 666–671.

140. Kuang, S.; Zheng, J.; Yang, H.; Li, S.; Duan, S.; Shen, Y.; Ji, C.; Gan, J.; Xu, X.-W.; Li, J. Structure insight of GSDMD reveals the basis of GSDMD autoinhibition in cell pyroptosis. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10642–10647.

141. Liu, Z.; Wang, C.; Rathkey, J. K.; Yang, J.; Dubyak, G. R.; Abbott, D. W.; Xiao, T. S. Structures of the Gasdermin D C-Terminal Domains Reveal Mechanisms of Autoinhibition. *Structure* **2018**, *26*, 778–784.

142. Liu, Z.; Wang, C.; Yang, J.; Zhou, B.; Yang, R.; Ramachandran, R.; Abbott, D. W.; Xiao, T. S. Crystal Structures of the Full-Length Murine and Human Gasdermin D Reveal Mechanisms of Autoinhibition, Lipid Binding, and Oligomerization. *Immunity* **2019**, *51*, 43–49.

143. Liu, X.; Zhang, Z.; Ruan, J.; Pan, Y.; Magupalli, V. G.; Wu, H.; Lieberman, J. Inflammasome-activated gasdermin D causes pyroptosis by forming membrane pores. *Nature* **2016**, *535*, 153–158.

144. Xia, S.; Zhang, Z.; Magupalli, V. G.; Pablo, J. L.; Dong, Y.; Vora, S. M.; Wang, L.; Fu, T.-M.; Jacobson, M. P.; Greka, A.; Lieberman, J.; Ruan, J.; Wu, H. Gasdermin D pore structure reveals preferential release of mature interleukin-1. *Nature* **2021**, *593*, 607–611.

145. Wei, Y.; Fox, T.; Chambers, S. P.; Sintchak, J.; Coll, J. T.; Golec, J. M.; Swenson, L.; Wilson, K. P.; Charifson, P. S. The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. *Chem. Biol.* **2000**, *7*, 423–432.

146. Franchi, L.; Eigenbrod, T.; Muñoz-Planillo, R.; Nuñez, G. The inflammasome: a caspase-1-activation platform that regulates immune responses and disease pathogenesis. *Nat. Immunol.* **2009**, *10*, 241–247.

147. Liu, Z.; Wang, C.; Yang, J.; Chen, Y.; Zhou, B.; Abbott, D. W.; Xiao, T. S. Caspase-1 Engages Full-Length Gasdermin D through Two Distinct Interfaces That Mediate Caspase Recruitment and Substrate Cleavage. *Immunity* **2020**, *53*, 106–114.

148. Shi, J.; Gao, W.; Shao, F. Pyroptosis: Gasdermin-Mediated Programmed Necrotic Cell Death. *Trends Biochem. Sci.* **2017**, *42*, 245–254.

149. Wannamaker, W.; Davies, R.; Namchuk, M.; Pollard, J.; Ford, P.; Ku, G.; Decker, C.; Charifson, P.; Weber, P.; Germann, U. A.; Kuida, K.; Randle, J. C. R. (S)-1-((S)-2-{[1-(4-amino-3-chloro-phenyl)-methanoyl]-amino}-3,3-dimethyl-butanoyl)-pyrrolidine-2-carboxylic acid ((2R,3S)-2-ethoxy-5-oxo-tetrahydro-furan-3-yl)-amide (VX-765), an orally available selective interleukin (IL)-converting enzyme/caspase-1 inhibitor, exhibits potent anti-inflammatory activities by inhibiting the release of IL-1beta and IL-18. *J. Pharmacol. Exp. Ther.* **2007**, *321*, 509–516.

150. Dhani, S.; Zhao, Y.; Zhivotovsky, B. A long way to go: caspase inhibitors in clinical use. *Cell Death Dis.* **2021**, *12*, 1–13.

151. Wang, K.; Sun, Q.; Zhong, X.; Zeng, M.; Zeng, H.; Shi, X.; Li, Z.; Wang, Y.; Zhao, Q.; Shao, F.; Ding, J. Structural Mechanism for GSDMD Targeting by Autoprocessed Caspases in Pyroptosis. *Cell* **2020**, *180*, 941–955.

152. Datta, D.; McClendon, C. L.; Jacobson, M. P.; Wells, J. A. Substrate and inhibitor-induced dimerization and cooperativity in caspase-1 but not caspase-3. *J. Biol. Chem.* **2013**, *288*, 9971–9981.

153. Li, H.; Guo, Z.; Chen, J.; Du, Z.; Lu, H.; Wang, Z.; Xi, J.; Bai, Y. Computational research of Belnacasan and new Caspase-1 inhibitor on cerebral ischemia reperfusion injury. *Aging* **2022**, *14*, 1848–1864.

154. Brown, M. H.; Paulsen, I. T.; Skurray, R. A. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Mol. Microbiol.* **1999**, *31*, 394–395.

155. Du, D.; Wang-Kan, X.; Neuberger, A.; van Veen, H. W.; Pos, K. M.; Piddock, L. J. V.; Luisi, B. F. Multidrug efflux pumps: structure, function and regulation. *Nat. Rev. Microbiol.* **2018**, *16*, 523–539.

156. Claxton, D. P.; Jagessar, K. L.; Mchaourab, H. S. Principles of Alternating Access in Multidrug and Toxin Extrusion (MATE) Transporters. *J. Mol. Biol.* **2021**, *6*, 166959.

157. Kusakizako, T.; Miyauchi, H.; Ishitani, R.; Nureki, O. Structural biology of the multidrug and toxic compound extrusion superfamily transporters. *Biochim. Biophys. Acta Biomembr.* **2020**, *1862*.

158. Moriyama, Y.; Hiasa, M.; Matsumoto, T.; Omote, H. Multidrug and toxic compound extrusion (MATE)-type proteins as anchor transporters for the excretion of metabolic waste products and xenobiotics. *Xenobiotica* **2008**, *38*, 1107–1118.

159. Omote, H.; Hiasa, M.; Matsumoto, T.; Otsuka, M.; Moriyama, Y. The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends Pharmacol. Sci.* **2006**, *27*, 587–593.

160. Radchenko, M.; Symersky, J.; Nie, R.; Lu, M. Structural basis for the blockade of MATE multidrug efflux pumps. *Nat. Commun.* **2015**, *6*, 7995.

161. Tanaka, Y.; Hipolito, C. J.; Maturana, A. D.; Ito, K.; Kuroda, T.; Higuchi, T.; Katoh, T.; Kato, H. E.; Hattori, M.; Kumazaki, K.; Tsukazaki, T.; Ishitani, R.; Suga, H.; Nureki, O. Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature* **2013**, *496*, 247–251.

162. Lu, M.; Radchenko, M.; Symersky, J.; Nie, R.; Guo, Y. Structural insights into H+-coupled multidrug extrusion by a MATE transporter. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1310–1317.

163. Ficici, E.; Zhou, W.; Castellano, S.; Faraldo-Gómez, J. D. Broadly conserved Na^+-binding site in the N-lobe of prokaryotic multidrug MATE transporters. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6172–6181.

164. Zakrzewska, S.; Mehdipour, A. R.; Malviya, V. N.; Nonaka, T.; Koepke, J.; Muenke, C.; Hausner, W.; Hummer, G.; Safarian, S.; Michel, H. Inward-facing conformation of a multidrug resistance MATE family transporter. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 12275–12284.

165. Verhalen, B.; Dastvan, R.; Thangapandian, S.; Peskova, Y.; Koteiche, H. A.; Nakamoto, R. K.; Tajkhorshid, E.; Mchaourab, H. S. Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature* **2017**, *543*, 738–741.

166. Jagessar, K. L.; Claxton, D. P.; Stein, R. A.; Mchaourab, H. S. Sequence and structural determinants of ligand-dependent alternating access of a MATE transporter. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 4732–4740.

167. Zhang, X. Less is More: Membrane Protein Digestion Beyond Urea–Trypsin Solution for Next-level Proteomics*. *Mol. Cell Proteomics* **2015**, *14*, 2441–2453.

168. Pirmoradian, M.; Budamgunta, H.; Chingin, K.; Zhang, B.; Astorga-Wells, J.; Zubarev, R. A. Rapid and Deep Human Proteome Analysis by Single-dimension Shotgun Proteomics. *Mol. Cell Proteomics* **2013**, *12*, 3330–3338.

169. Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **2009**, *459*, 356–363.

170. Heng, B. C.; Aubel, D.; Fussenegger, M. An overview of the diverse roles of G-protein coupled receptors (GPCRs) in the pathophysiology of various human diseases. *Biotechnol. Adv.* **2013**, *31*, 1676–1694.

171. Hu, G.-M.; Mai, T.-L.; Chen, C.-M. Visualizing the GPCR Network: Classification and Evolution. *Sci. Rep.* **2017**, *7*, 15495.

172. Sriram, K.; Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* **2018**, *93*, 251–258.

173. Tate, C. G. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends. Biochem. Sci.* **2012**, *37*, 343–352.

174. Lazim, R.; Suh, D.; Lee, J. W.; Vu, T. N. L.; Yoon, S.; Choi, S. Structural Characterization of Receptor-Receptor Interactions in the Allosteric Modulation of G Protein-Coupled Receptor (GPCR) Dimers. *Int. J. Mol. Sci.* **2021**, *22*, 3241.

175. Erlandson, S. C.; McMahon, C.; Kruse, A. C. Structural Basis for G Protein–Coupled Receptor Signaling. *Annu. Rev. Biophys.* **2018**, *47*, 1–18.

176. Tuteja, N. Signaling through G protein coupled receptors. *Plant Signal. Behav.* **2009**, *4*, 942–947.

177. Noma, T.; Lemaire, A.; Naga Prasad, S. V.; Barki-Harrington, L.; Tilley, D. G.; Chen, J.; Le Corvoisier, P.; Violin, J. D.; Wei, H.; Lefkowitz, R. J.; Rockman, H. A. Beta-Arrestin–mediated Beta1-adrenergic receptor transactivation of the EGFR confers cardioprotection. *J. Clin. Invest.* **2007**, *117*, 2445–2458.

178. Serrano-Vega, M. J.; Magnani, F.; Shibata, Y.; Tate, C. G. Conformational thermostabilization of the Beta1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 877–882.

179. Serrano-Vega, M. J.; Tate, C. G. Transferability of thermostabilizing mutations between beta-adrenergic receptors. *Mol. Membr. Biol.* **2009**, *26*, 385–396.

180. Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G.; Tate, C. G.; Schertler, G. F. Structure of a Beta1-adrenergic G protein-coupled receptor. *Nature* **2008**, *454*, 486–491.

181. Zarzycka, B.; Zaidi, S. A.; Roth, B. L.; Katritch, V. Harnessing Ion-Binding Sites for GPCR Pharmacology. *Pharmacol. Rev.* **2019**, *71*, 571–595.

182. Martin, C. *et al.* Rational Design of Nanobody80 Loop Peptidomimetics: Towards Biased Beta2 Adrenergic Receptor Ligands. *Chemistry* **2017**, *23*, 9632–9640.

183. Warne, T.; Moukhametzianov, R.; Baker, J. G.; Nehmé, R.; Edwards, P. C.; Leslie, A. G. W.; Schertler, G. F. X.; Tate, C. G. The structural basis for agonist and partial agonist action on a Beta(1)-adrenergic receptor. *Nature* **2011**, *469*, 241–244.

184. Ballesteros, J.; Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *J. Neurosci. Methods* **1995**, *25*, 366–428.

185. Warne, T.; Edwards, P. C.; Doré, A. S.; Leslie, A. G. W.; Tate, C. G. Molecular basis for high-affinity agonist binding in GPCRs. *Science* **2019**, *364*, 775–778.

186. Su, M. *et al.* Structural Basis of the Activation of Heterotrimeric Gs-Protein by Isoproterenol-Bound Beta1-Adrenergic Receptor. *Mol. Cell* **2020**, *80*, 59–71.

187. Staus, D. P. *et al.* Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation. *Nature* **2016**, *535*, 448–452.

188. Rath, A.; Glibowicka, M.; Nadeau, V. G.; Chen, G.; Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1760–1765.

189. Moukhametzianov, R.; Warne, T.; Edwards, P. C.; Serrano-Vega, M. J.; Leslie, A. G. W.; Tate, C. G.; Schertler, G. F. X. Two distinct conformations of helix 6 observed in antagonist-bound structures of a Beta1-adrenergic receptor. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8228–8232.

190. Rößler, P.; Mayer, D.; Tsai, C.-J.; Veprintsev, D. B.; Schertler, G. F. X.; Gossert, A. D. GPCR Activation States Induced by Nanobodies and Mini-G Proteins Compared by NMR Spectroscopy. *Molecules* **2020**, *25*, 5984.

191. Grahl, A.; Abiko, L. A.; Isogai, S.; Sharpe, T.; Grzesiek, S. A high-resolution description of Beta1-adrenergic receptor functional dynamics and allosteric coupling from backbone NMR. *Nat. Commun.* **2020**, *11*, 2216.

192. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the Beta2 adrenoceptor. *Nature* **2011**, *469*, 175–180.

193. Bang, G.; Lee, H.; Kim, H.; Han, E. H.; Park, Y. H.; Kim, J. Y. Comparison of protein characterization using In solution and S-Trap digestion methods for proteomics. *Biochem. Biophys. Res. Commun.* **2022**, *589*, 197–203.

Appendices

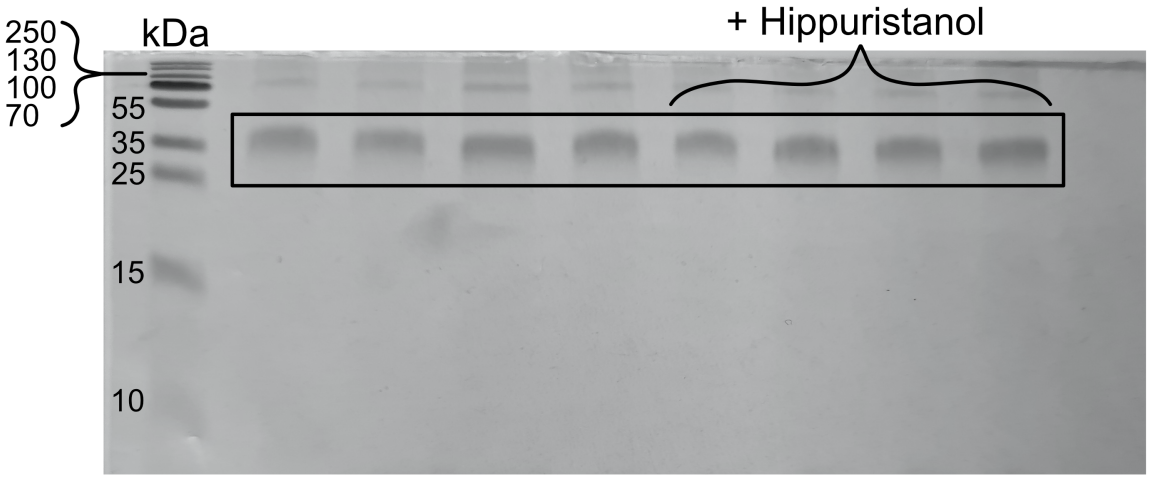


Figure S1 | SDS-PAGE of carbene labelled eIF4A with and without hippuristanol. eIF4A (~46 kDa) is indicated.

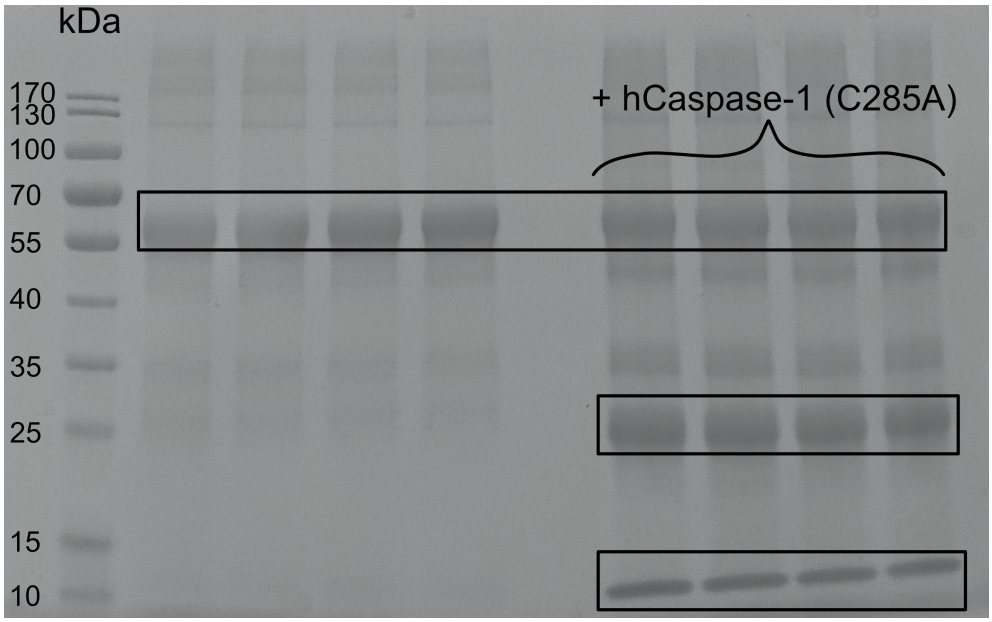


Figure S2 | SDS-PAGE of carbene labelled hGSDMD with and without hCaspase-1 (C285A). hGSDMD (~53 kDa), hCaspase-1 p20 (~20 kDa) and p10 (~10 kDa) subunits are indicated.

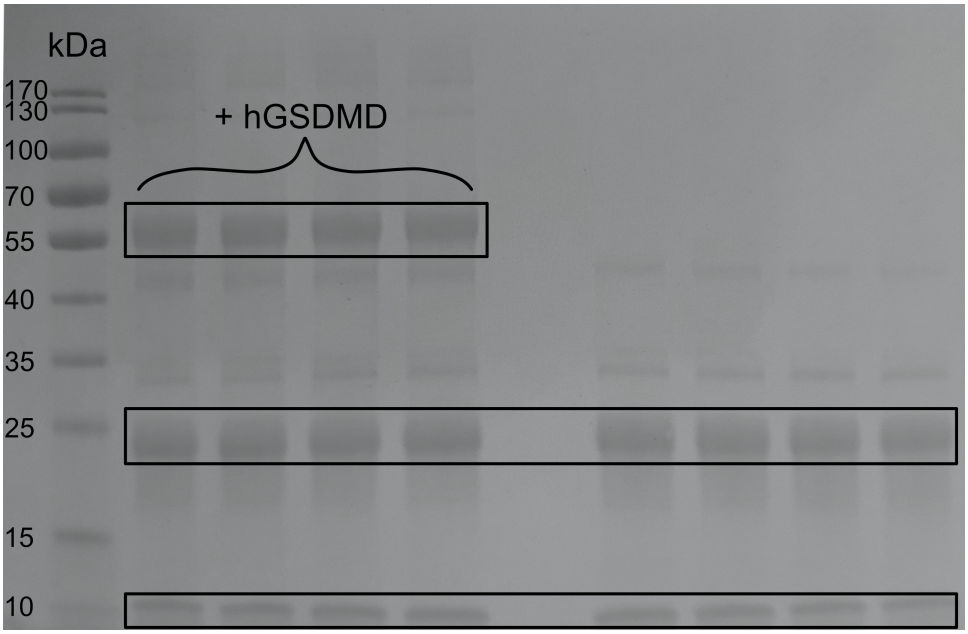


Figure S3 | SDS-PAGE of carbene labelled hCaspase-1 (C285A) with and without hGSDMD. hGSDMD (~53 kDa), hCaspase-1 p20 (~20 kDa) and p10 (~10 kDa) subunits are indicated.

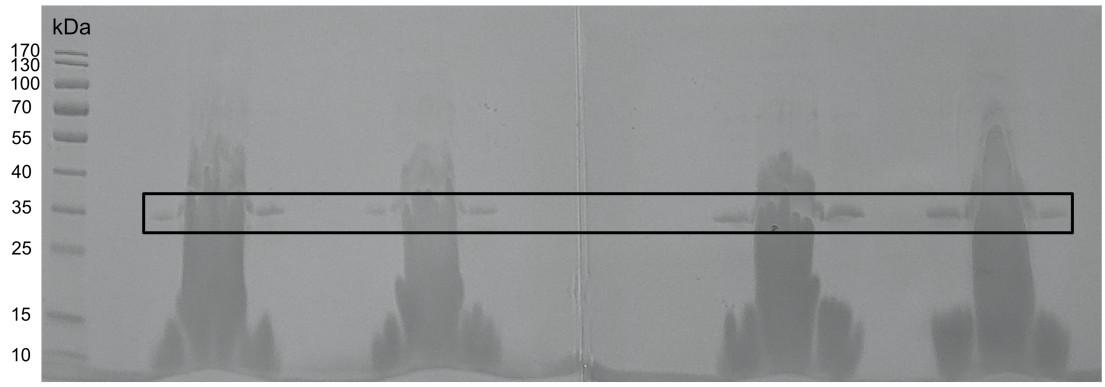


Figure S4 | SDS-PAGE of carbene labelled hGSDMD-NT pores. hGSDMD-NT (~31 kDa) is indicated.

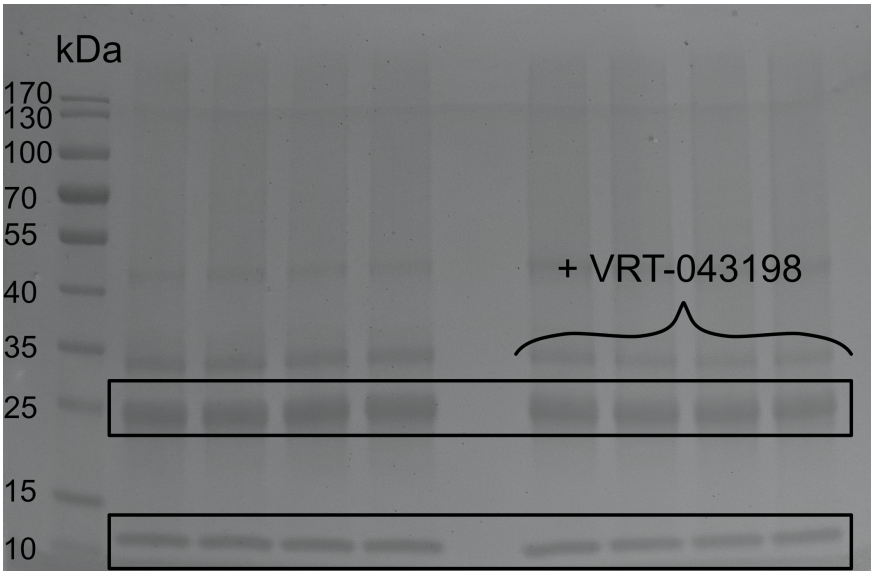


Figure S5 | SDS-PAGE of carbene labelled hCaspase-1 (C285A) with and without VRT-043198. hCaspase-1 p20 (~20 kDa) and p10 (~10 kDa) subunits are indicated.

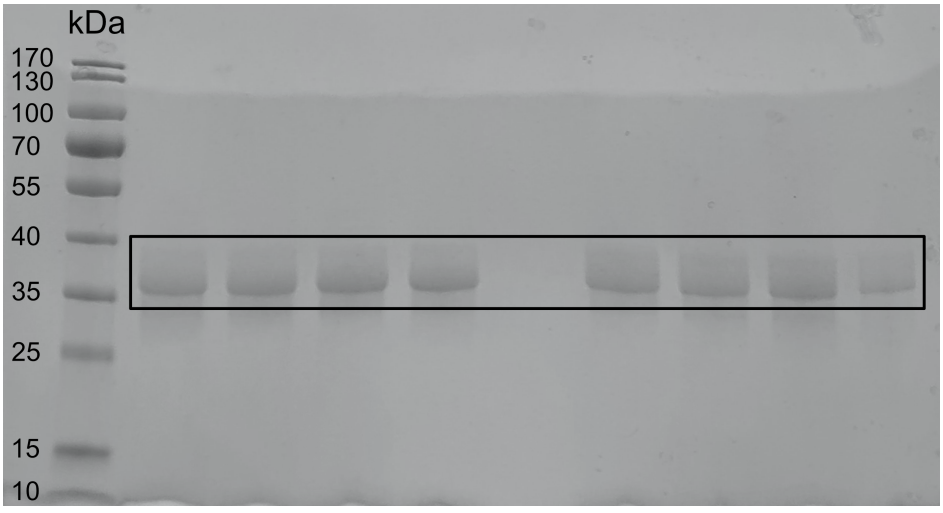


Figure S6 | SDS-PAGE of carbene labelled PfMATE. PfMATE (~47 kDa) is indicated.

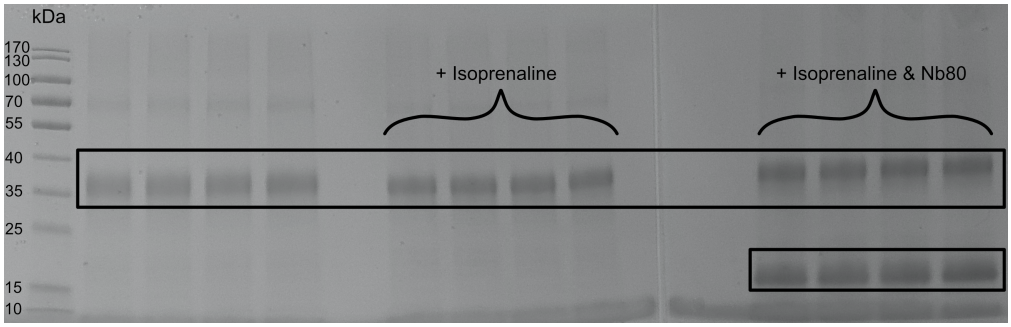


Figure S7 | SDS-PAGE of carbene labelled β_1AR with and without isoprenaline, and isoprenaline and Nb80. β_1AR (~40 kDa) and Nb80 (~14 kDa) are indicated.

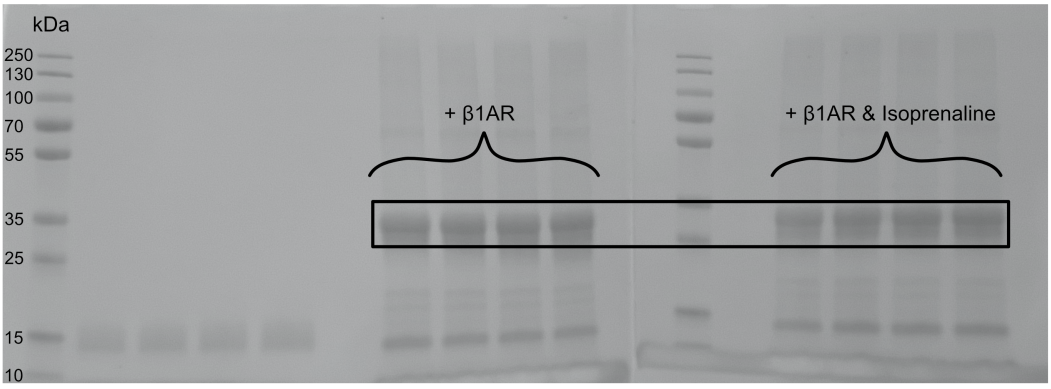


Figure S8 | SDS-PAGE of carbene labelled Nb80 with and without β_1AR, and β_1AR and isoprenaline. β_1AR (~40 kDa) and Nb80 (~14 kDa) are indicated.

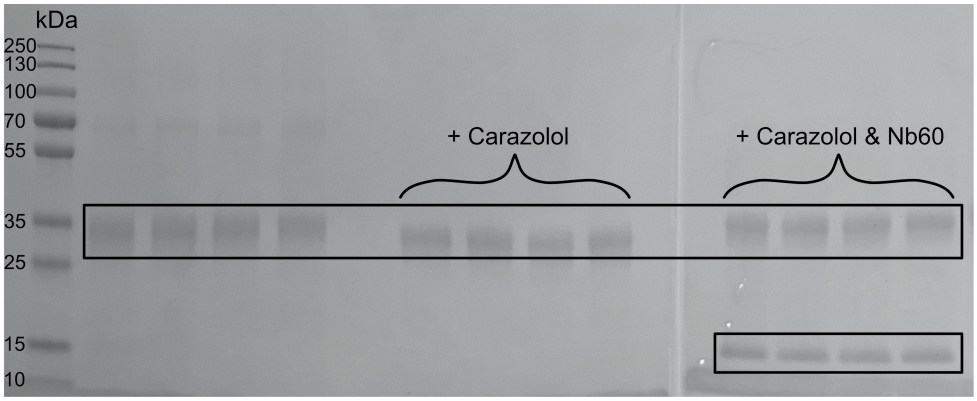


Figure S9 | SDS-PAGE of carbene labelled β_1AR with and without carazolol, and carazolol and Nb60. β_1AR (~40 kDa) and Nb60 (~13 kDa) are indicated.

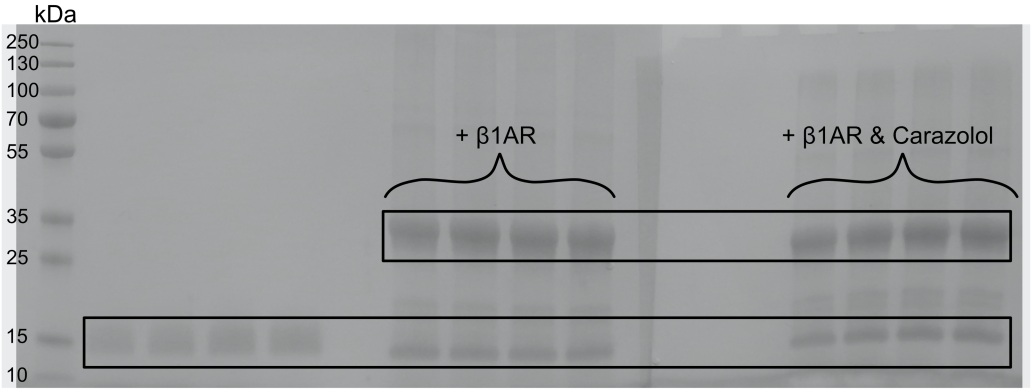


Figure S10 | SDS-PAGE of carbene labelled Nb60 with and without β_1AR, and β_1AR and carazolol. β_1AR (~40 kDa) and Nb60 (~13 kDa) are indicated.

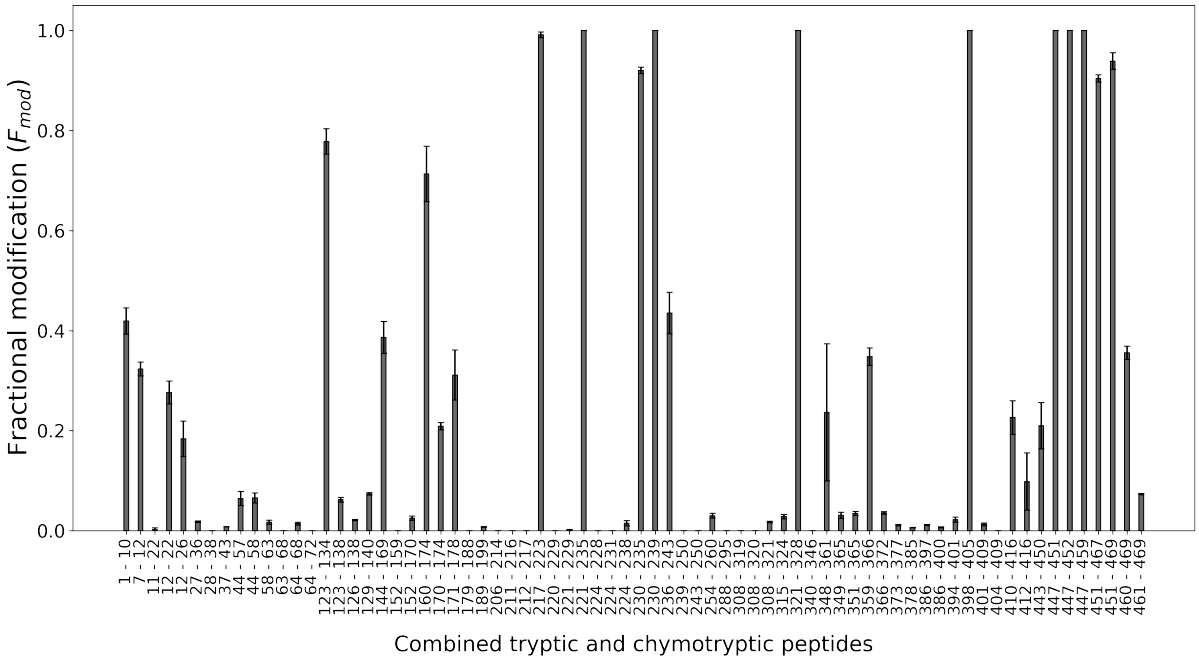


Figure S11 | Pooled carbene footprinting histogram of PfMATE with a trypsin and chymotrypsin digestion. The fractional modification of each tryptic and chymotryptic PfMATE peptide is shown. Error bars are ± standard deviation (n = 4).

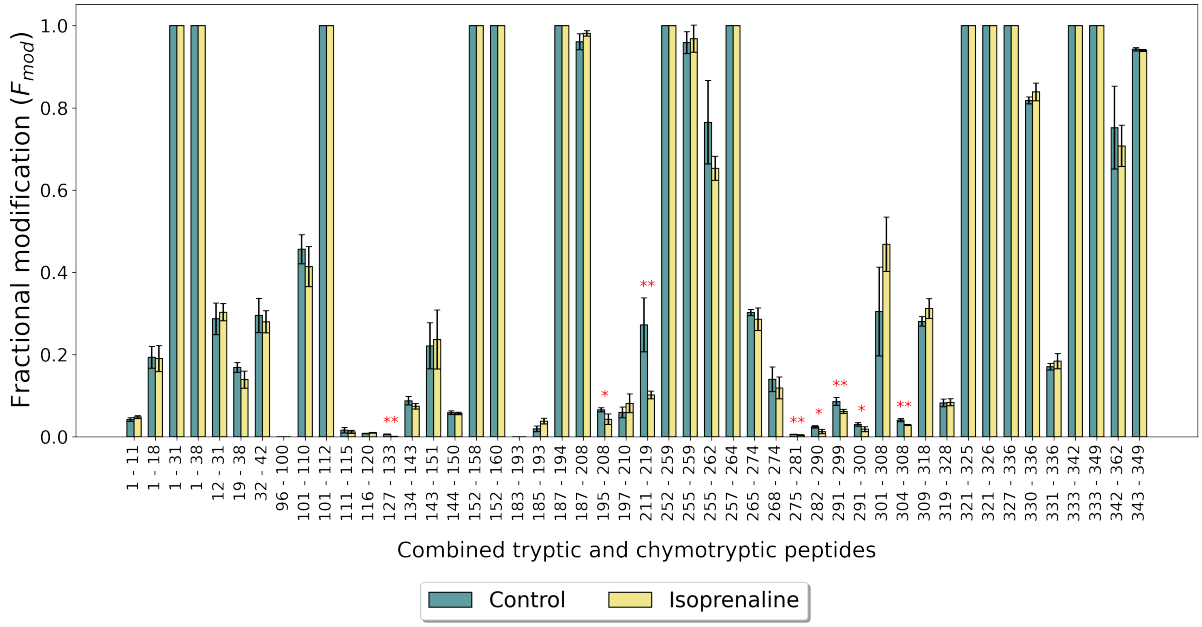


Figure S12 | Pooled carbene footprinting histogram of β_1AR alone (control), and with isoprenaline with a trypsin and chymotrypsin digestion. The fractional modification of each tryptic and chymotryptic β_1AR peptide is shown with (yellow) and without (turquoise) 100 µM of isoprenaline. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).

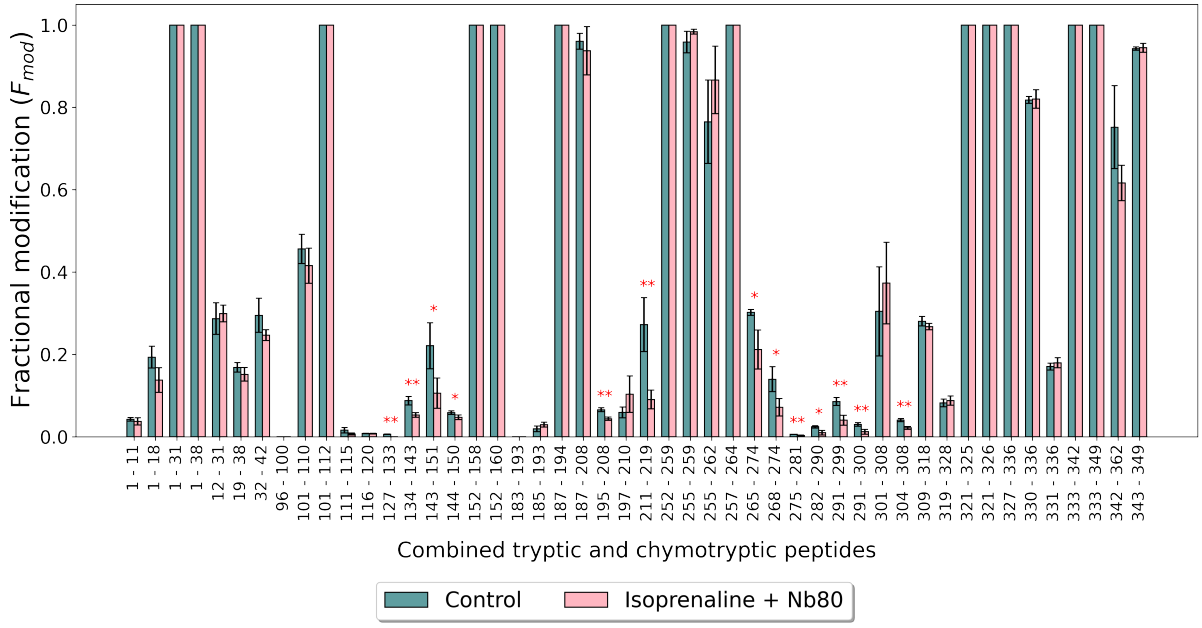


Figure S13 | Pooled carbene footprinting histogram of β_1AR alone (control), and with isoprenaline and Nb80 with a trypsin and chymotrypsin digestion. The fractional modification of each tryptic and chymotryptic β_1AR peptide is shown with (yellow) and without (pink) 100 µM of isoprenaline and 2:1 molar equivalents of Nb80. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).

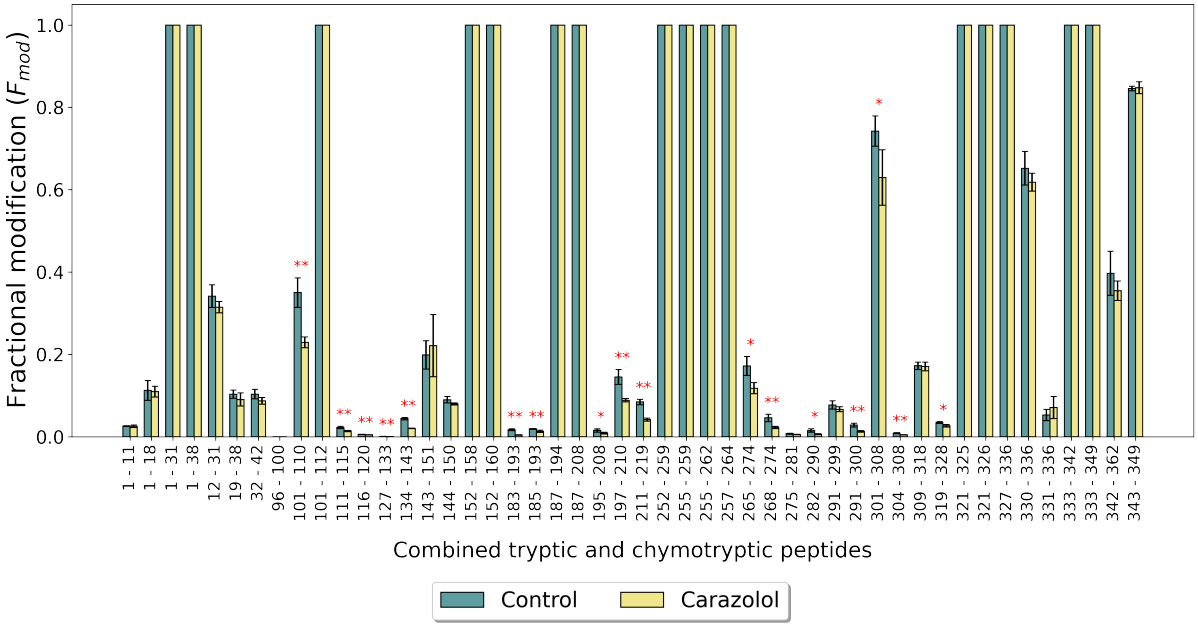


Figure S14 | Pooled carbene footprinting histogram of β_1AR alone (control), and with carazolol with a trypsin and chymotrypsin digestion. The fractional modification of each tryptic and chymotryptic β_1AR peptide is shown with (yellow) and without (turquoise) 100 µM of carazolol. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).

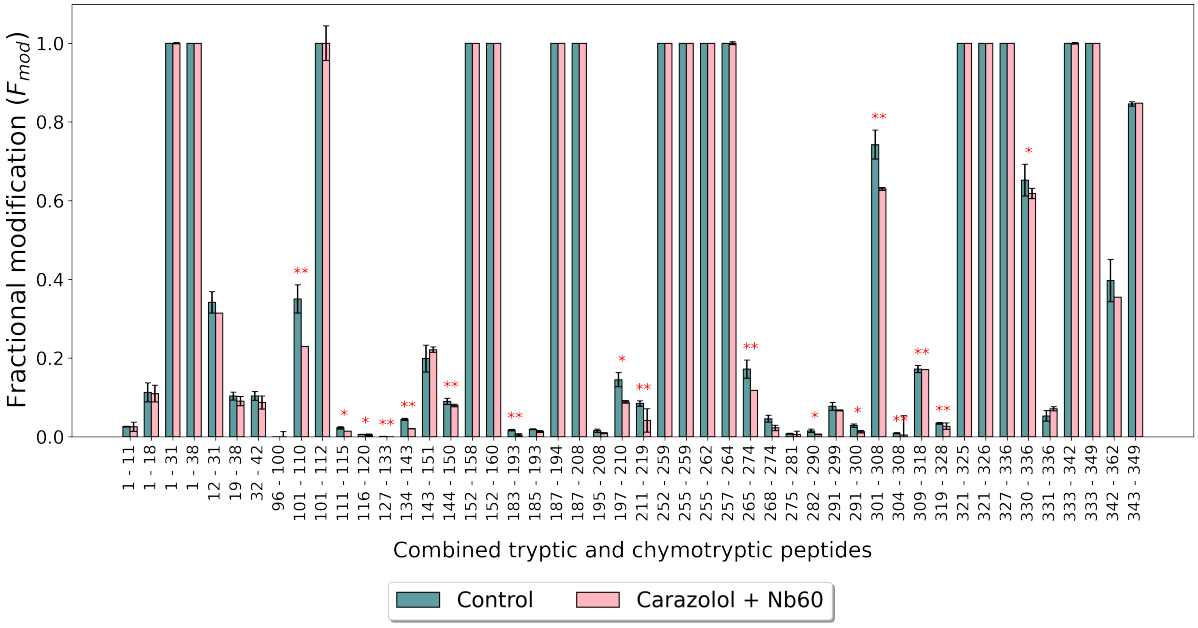


Figure S15 | Pooled carbene footprinting histogram of β_1AR alone (control), and with carazolol and Nb60 with a trypsin and chymotrypsin digestion. The fractional modification of each tryptic and chymotryptic β_1AR peptide is shown with (yellow) and without (pink) 100 µM of carazolol and 2:1 molar equivalents of Nb60. Error bars are ± standard deviation (n = 4). Asterisks denote significant difference between samples (Student t-test; ** = P < 0.01; * = P < 0.05).