Development of HDX-MS methods to further investigate complex protein samples

Luke Smith

St Cross College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2024

Abstract

Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) has become an essential technique in biological research for investigating the dynamic structural changes in proteins in response to various perturbations. This thesis explores the structural alterations that occur when proteins interact with different natural binding partners, aiming to identify and understand these differences. A significant contribution of this work is the development of a novel analytical script that streamlines the HDX-MS process. By integrating statistical testing, this tool rapidly identifies significant changes and presents the results clearly and concisely, enhanced by visual aids that facilitate the interpretation of structural variations across different experimental states. The thesis also includes several case studies of protein complexes analysed using HDX-MS, which had not previously been investigated with this technique. These findings were corroborated with complementary methods and aligned with expectations based on prior research.

Additionally, this thesis addresses the challenges posed by glycosylation, a post-translational modification that often complicates data interpretation and analysis, leading to gaps in sequence coverage. To overcome this, a method is presented that effectively manages glycan moieties, significantly improving the accuracy and completeness of glycoprotein analysis. This approach includes the use of a glycosidase that is active under typical quench conditions, this enables glycan cleavage while preserving the integrity of the native data throughout the labelling process. As a result, sequence coverage is enhanced, allowing for a more comprehensive understanding of the entire protein structure. Overall, this work offers advancements in both the analytical and practical aspects of HDX-MS, contributing to a deeper understanding of protein dynamics and the impact of glycosylation on protein behaviour

Development of HDX-MS methods to further investigate complex protein samples



Luke Smith

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2024

In loving memory of two remarkable individuals, Howard Smith and Mickey Lewis, whose absence is deeply felt but whose legacies resonate profoundly in my heart and in the fabric of this thesis.

To my beloved Grandad, Howard Smith, whose departure in late 2022 left a void that cannot be filled. A towering figure in my life, he embodied resilience, wisdom, and kindness. Grandad, you were more than a family patriarch; you were my role model. Your accomplishments, the strength of your character, and the love you showered upon our family have been guiding lights throughout my academic journey. Your enduring impact on our family's narrative and my own trajectory is immeasurable. I am eternally grateful for the privilege of being your grandchild.

In honor of Mickey Lewis, a titan within OUAFC, whose passing in early 2021 left an indelible mark on the Oxford football community and on my own life. Mickey was more than a coach; he was a force of passion and energy, infusing every training session and match with his contagious spirit. His twenty years of coaching the Blues were not just about the wins and losses but about fostering a sense of camaraderie and instilling a love for the beautiful game. Mickey's influence was the catalyst that reignited my passion for football upon arriving at Oxford, propelling me into the role of Blues captain and introducing me to a network of lifelong friends. Without OUAFC and Mickey, my Oxford experience would have been markedly different.

To Howard and Mickey, whose impact transcends the boundaries of personal and academic spheres, I dedicate this thesis. Your legacies continue to inspire and shape the person I am today. I am forever grateful for the privilege of having known and been influenced by two extraordinary souls. May your memories endure, guiding me through the chapters of my life that lie ahead.

Acknowledgements

Personal

Reflecting on my journey through this degree at Oxford, I am compelled to express my deepest gratitude to the exceptional individuals who have made this academic odyssey both possible and memorable.

First and foremost, my heartfelt thanks go to my supervisor, Prof Justin Benesch. His unwavering support, sage advice, and personal mentorship have been instrumental in transforming the dream of obtaining a PhD into a tangible reality. Justin, I am profoundly grateful for your guidance and generosity throughout my academic and personal endeavors.

I extend sincere appreciation to my esteemed peers and colleagues, Dr Joshua Bishop, Dr Dominik Saman, Dr Wiktoria Sadowska, and soon-to-be Dr Elena Holden, along with the entire Benesch group. Your camaraderie, valuable insights, and willingness to assist have enriched my academic experience. A special nod of gratitude to Dr Josh Bishop, whose friendship and counsel have been invaluable, extending beyond the academic realm into the collaborative development of the SAUSC analytical code.

In the realm of camaraderie and shared passion, I owe a debt of thanks to my friends and teammates within OUAFC. Representing the university and captaining the team during the 2021/22 season stands as a pinnacle achievement, and I want to express my gratitude to all my teammates and friends who made this possible. Matthew Naylor, Alfic Cicale, Ben Putland, Finlay Ryan-Philips, and many others – your presence has made my time here extraordinary.

Lastly, I want to acknowledge the unwavering support of my entire family, especially my parents, Sharon and Richard, and my sister, Colette. Their steadfast encouragement, love, and understanding have been my rock through the highs and lows of pursuing a PhD. I am forever grateful for their unwavering presence in my life.

To all those mentioned and countless others who have contributed to my academic and personal growth during these years, thank you. Your friendship, support, and love have left an indelible mark on this chapter of my life, and I am profoundly thankful for each one of you.

Abstract

Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) has become an essential technique in biological research for investigating the dynamic structural changes in proteins in response to various perturbations. This thesis explores the structural alterations that occur when proteins interact with different natural binding partners, aiming to identify and understand these differences. A significant contribution of this work is the development of a novel analytical script that streamlines the HDX-MS process. By integrating statistical testing, this tool rapidly identifies significant changes and presents the results clearly and concisely, enhanced by visual aids that facilitate the interpretation of structural variations across different experimental states. The thesis also includes several case studies of protein complexes analysed using HDX-MS, which had not previously been investigated with this technique. These findings were corroborated with complementary methods and aligned with expectations based on prior research.

Additionally, this thesis addresses the challenges posed by glycosylation, a post-translational modification that often complicates data interpretation and analysis, leading to gaps in sequence coverage. To overcome this, a method is presented that effectively manages glycan moieties, significantly improving the accuracy and completeness of glycoprotein analysis. This approach includes the use of a glycosidase that is active under typical quench conditions, this enables glycan cleavage while preserving the integrity of the native data throughout the labelling process. As a result, sequence coverage is enhanced, allowing for a more comprehensive understanding of the entire protein structure. Overall, this work offers advancements in both the analytical and practical aspects of HDX-MS, contributing to a deeper understanding of protein dynamics and the impact of glycosylation on protein behaviour

Contents

List of Figures **xiii**

List of Abbreviations **xv**

1 Introduction **1**
1.1 The evolving use of mass spectrometry within biological research . . 1
1.2 Native MS . . . 3
1.2.1 Sample Preparation . . . 4
1.2.2 Instrumentation . . . 4
1.2.2.1 Ion source - ESI . . . 5
1.2.2.2 Ion formation . . . 6
1.2.2.3 Mass Analysers . . . 7
1.2.2.4 Mass Detectors . . . 12
1.2.3 Example Experiments . . . 13
1.3 HDX-MS . . . 15
1.3.1 Experimental Workflow . . . 15
1.3.2 Labelling . . . 17
1.3.2.1 Continuous vs Pulsed labelling . . . 20
1.3.3 Quenching . . . 20
1.3.4 LC-MS Separation . . . 23
1.3.5 Data Analysis . . . 24
1.3.5.1 Back exchange . . . 25
1.3.6 Challenges Still Prevalent with the HDX Technique . . . 25
1.3.6.1 Understanding Differences of Deuterium Uptake . . 27

2 Development of our own analytical aid **29**
2.1 Abstract . . . 29
2.2 Introduction . . . 30
2.2.1 Different Analysis tools . . . 31
2.2.1.1 Deuteros 2.0 . . . 32
2.2.1.2 HDXanalyser . . . 33
2.2.1.3 HX-Express v2 . . . 34

2.2.1.4 HDX-Viewer . . . 34
2.2.1.5 MEMHDX . . . 35
2.2.2 Statistical Analysis . . . 35
2.2.2.1 Types of error . . . 36
2.2.3 Challenges in presenting/Analysing Data . . . 37
2.3 SAUSC . . . 38
2.3.1 T-test . . . 38
2.3.2 Global Threshold . . . 39
2.3.2.1 Determining Global Threshold value . . . 39
2.3.3 Hybrid Test . . . 39
2.3.4 Confidence levels . . . 40
2.3.5 Visual Aid - Structural Relevance . . . 41
2.3.5.1 PDB Colouring . . . 42
2.3.5.2 Uptake Difference . . . 42
2.3.5.3 Relative Difference . . . 43
2.3.5.4 Gradient Based Colour Scheme . . . 44
2.3.5.5 global vs time point dependent . . . 45
2.4 Results . . . 45
2.4.1 Figure Plotting . . . 46
2.4.1.1 Volcano Plots . . . 47
2.4.1.2 Woods Plots . . . 50
2.4.1.3 Cumulative Plots . . . 51
2.5 Discussion . . . 55

3 Investigating the affects of metal binding to Iso-Penicillin-N-synthase 57
3.1 Abstract . . . 57
3.2 Isopenicillin N Synthase . . . 58
3.2.1 Use of Zn/Cd(II) to Mimic Iron Binding . . . 59
3.2.2 ACV Tripeptide Binding Pocket . . . 60
3.2.3 Experimental plan . . . 61
3.2.4 Cross State Analysis . . . 62
3.2.5 Differences - Structural Insight Gained . . . 63
3.3 Methods . . . 65
3.3.1 HDX-MS . . . 65
3.4 Results . . . 66
3.4.1 Structural Relevance . . . 70
3.5 Discussion . . . 73
3.5.1 Octahedral Coordination . . . 73

4 How does the presence of the C2 domain affect the structure and function of SHIP1 and SHIP2 75
4.1 Abstract . . . 75
4.2 Introduction to SHIP1 and SHIP2 systems . . . 76
4.2.1 Role of proteins in disease . . . 77
4.2.1.1 Cancer . . . 77
4.2.1.2 Alzheimer's . . . 78
4.3 SHIP 1 and 2 Structural make-up . . . 78
4.4 Methods . . . 81
4.4.1 HDX . . . 81
4.5 Results . . . 83
4.5.1 HDX . . . 83
4.5.2 Activity assay . . . 88
4.5.3 Molecular Dynamics . . . 90
4.6 Discussion . . . 93
4.6.1 C2 domain modulates flexibility . . . 94

5 Overcoming Glycan Heterogeneity Using an Adapted HDX Workflow 97
5.1 Abstract . . . 98
5.2 Introduction . . . 98
5.2.1 Inherent Heterogeneity . . . 99
5.3 pH Problem . . . 101
5.3.1 PNGase F . . . 101
5.3.2 Enzyme Inactivity . . . 103
5.3.3 Acidophiles - Potential Answers . . . 104
5.4 Development of PNGase Rc Column . . . 105
5.4.1 Addition to the Typical Workflow . . . 105
5.4.2 Activity and Reproducibility . . . 106
5.4.3 Methods . . . 108
5.5 Results . . . 110
5.5.1 Choice of Protein . . . 110
5.5.2 HDX data post the addition of the PNGase Column . . . 111
5.5.2.1 Discerning between N and O-linked Glycans . . . 112
5.5.3 Column Performance . . . 113
5.5.4 Individual Protein optimisation . . . 114
5.5.5 Back exchange . . . 115
5.6 Discussion . . . 115

6 Overview **117**
6.1 Overview/Summary . . . 117
6.1.1 SAUSC . . . 117
6.1.2 Ligand binding HDX . . . 118
6.1.3 Dynamics of SHIP constructs . . . 119
6.1.4 Dealing with glycoforms . . . 120
6.1.5 Conclusion . . . 121

Appendices

A Investigating the affects of metal binding to Iso-Penicillin-N-synthase 125
A.1 Raw Uptake Plots . . . 125
A.1.1 IPNS vs IPNS + Zn(II) . . . 126
A.1.2 IPNS vs IPNS + Zn(II) and ACV . . . 128

B How does the presence of the C2 domain affect the structure and function in SHIP1 and SHIP2 **131**
B.1 Sequence Identification of SHIP1 and SHIP2 . . . 132
B.1.1 Sequence Alignment of SHIP1 and SHIP2 . . . 132
B.2 Raw Uptake Plots . . . 134
B.2.1 SHIP1 . . . 134
B.2.2 SHIP2 . . . 137
B.2.3 Full SAUSC output for SHIP2 . . . 139

C Overcoming Glycan Heterogeneity Using an Adapted HDX Workflow **141**
C.1 Glycan nomenclature . . . 142
C.2 Additional coverage maps of tested proteins . . . 143
C.2.1 Hp 1-1 . . . 143
C.2.2 VISTA . . . 144

References **147**

List of Figures

1.1 ESI MS process 6
1.2 Quadrupole Mass Analyser 8
1.3 Schematic of Synapt G2-Si 9
1.4 Orbitrap Mass analyser 11
1.5 Mass shift of Transferrin in the acquired native data 14
1.6 HDX workflow 17
1.7 EX1 vs EX2 examples 19
1.8 How the rate of exchange decreases with increased pH 22

2.1 Example Volcano plot provided by SAUSC 48
2.2 Uptake difference and Relative Uptake Difference vs -$\log_{10}$(P-Value) 49
2.3 Example woods plot provided by SAUSC 50
2.4 Cumulative Volcano plot with PDB visual aid example 52
2.5 Comparison of individual and cumulative plots 54

3.1 Binding of ACV in the active site of IPNS 58
3.2 Mechanism of the formation of IPN 59
3.3 Example of uptake differences occuring through a binding experiment 63
3.4 Common peptide coverage map for IPNS 67
3.5 Heat maps of raw HDX data for apo IPNS protein and comparative states. 68
3.6 Differential heat maps with coloured PDB file. 69
3.7 Cumulative plots for the state comparison of Cd addition 72

4.1 The breakdown of SHIP1 and 2 domains 79
4.2 The structure of the phosphatase and C2 domains of SHIP1 80
4.3 Coverage map of common peptides found across both SHIP1 constructs 84
4.4 Coverage map of common peptides found across both SHIP2 constructs 85
4.5 cumulative volcano with PDB for SHIP2 85
4.6 HDX heat maps of both SHIP1 and 2 and their respective structural aids 86
4.7 Activity assays of both SHIP1 and 2 89
4.8 MD simulation data of SHIP1 and 2 92

5.1 Typical Proteomic approach for cleaving glycans 102
5.2 Time series of PNGase F cleaving apo Transferrin at pH 8 102
5.3 PNGase F activity comparison at differing pH 104
5.4 Adapted workflow of the typical HDX-MS set up 106
5.5 Column Performance Evaluation 107
5.6 Activity of PNGase column in presence of typical quench additives . 108
5.7 Transferrin sequence coverage 111
5.8 Coloured PDB file of Transferrin in line with coverage map 112
5.9 Osteopontin sequence coverage with and without PNGase column . 113

A.1 peptides 1-36 uptake plots IPNS vs IPNS + Zn(II) 126
A.2 peptides 37-72 uptake plots IPNS vs IPNS + Zn(II) 127
A.3 peptides 1-36 uptake plots IPNS vs IPNS + Zn(II) and ACV 128
A.4 peptides 37-68 uptake plots IPNS vs IPNS + Zn(II) and ACV . . . 129

B.1 Aligned sequences of SHIP1 and SHIP2 133
B.2 peptides 1-48 uptake plots SHIP1 135
B.3 peptides 49-83 uptake plots SHIP1 136
B.4 peptides 1-36 uptake plots SHIP2 137
B.5 peptides 37-66 uptake plots SHIP2 138
B.6 Full SAUSC output for SHIP2 139

C.1 nomenclature for monosaccharides 142
C.2 Hp 1-1 coverage map 144
C.3 VISTA coverage map 144

List of Abbreviations

MS . . . . . Mass Spectrometry

HDX . . . . . Hydrogen-Deuterium Exchange

ESI . . . . . Electrospray ionisation

MALDI . . . . Matrix-assisted laser desorption/ionisation

nESI . . . . . nanoElectrospray ionisation

IEM . . . . . Ion Evaporation Model

CRM . . . . . Charge Residue Model

CEM . . . . . Chain Ejection Model

RF . . . . . Radiofrequency

DC . . . . . Direct Current

TOF . . . . . Time-of-Flight

LOD . . . . . Limit of detection

IR-MS . . . . . isotope ratio mass spectrometry

MS/MS . . . . Tandem MS

CID . . . . . Collision induced

ECD . . . . . Electron capture dissociation

D_2O . . . . . Deuterium Oxide

UHPLC . . . . Ultra High Performance Liquid Chromatography

PLGS . . . . . ProteinLynx Global Server

GdmCl . . . . Guanidinum Chloride

LC . . . . . Liquid Chromatography

SAUSC . . . . Sequence averaged uptake state comparison

PDB . . . . . Protein data bank

SEM . . . . . standard error mean

CI . . . . . Confidence Interval

DU . . . . . Deuterium Uptake

RD . . . . . Relative difference

1-D, 2-D, 3-D One-, two- or three-dimensional, referring in this thesis to spatial dimensions in an image.

IPNS . . . . . Iso Penicillin N Synthase

MD . . . . . Molecular Dynamics

ACV . . . . . δ-(L-α-aminoadipoyl)-L-cysteinyl-D-valine

2OG 2-oxoglutarate

IPN . . . . . Iso Penicillin N

PH . . . . . Pleckstrin Homology

Ptase . . . . . Phosphatase

L1, L2, L3...LN Loop 1, Loop 2, Loop 3...Loop N

AML . . . . . Acute Myeloid Leukemias

GWAS . . . . . Genome-wide Association Studies

SNPs . . . . . Single nucleotide polymorphisms

LOAD . . . . . Late-onset Alzheimers Disease

MDCC . . . . N-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin- 3-carboxamide

RMSFs . . . . Root-mean-square Fluctuations

PTM . . . . . Post-trnslational Modification

E:B . . . . . Enzyme-to-bead

SNFG . . . . . Symbol nomenclature for glycans

Neque porro quisquam est qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit...

There is no one who loves pain itself, who seeks after it and wants to have it, simply because it is pain...

— Cicero's *de Finibus Bonorum et Malorum*

1 Introduction

Contents

1.1 The evolving use of mass spectrometry within biological research 1
1.2 Native MS 3
1.2.1 Sample Preparation 4
1.2.2 Instrumentation 4
1.2.3 Example Experiments 13
1.3 HDX-MS 15
1.3.1 Experimental Workflow 15
1.3.2 Labelling 17
1.3.3 Quenching 20
1.3.4 LC-MS Separation 23
1.3.5 Data Analysis 24
1.3.6 Challenges Still Prevalent with the HDX Technique . . 25

1.1 The evolving use of mass spectrometry within biological research

In the ever-evolving landscape of scientific research, we are now equipped with an array of techniques to tackle intricate biological inquiries. This chapter aims to shed light on several of these techniques used, with a particular focus on those applied within the context of this thesis.

Within the field of structural biology, a multitude of unresolved questions persist, many of which hold immense significance and relevance. These queries serve as catalysts for continually driving advancements of the instrumentation and, or the refinement of methodologies used. That aim is to help us unraveling phenomena and continuously pushing the boundaries of biological research. Through the application of diverse methodologies, we strive to unlock answers to these pressing questions, thereby propelling progress and innovation within the realm of structural biology and drug development.

As techniques continue to evolve in tandem with advancements in instrumentation and technology, researchers are increasingly empowered to tackle ever-more complex inquiries that were once deemed insurmountable. This continuous evolution opens up new avenues for exploration and discovery, pushing the boundaries of scientific understanding, and paving the way for transformative breakthroughs.

Within this thesis, two prominent techniques are native mass spectrometry (native MS) and Hydrogen-Deuterium exchange-MS (HDX-MS), both of which utilise the capabilities of a mass spectrometer. Despite sharing this common instrumentation, these techniques diverge significantly in their applications and functionalities.

Native MS is used to analyse intact biomolecules in their native state, without the need for prior chemical modification or denaturation. This technique offers insights into the composition, stoichiometry, and dynamics of macromolecular complexes, providing invaluable information on their structure and interactions[1]. The obstruction of this method when it was first developed came through the transfer of the intact complex in question into the gas phase, the development of the method in the early 1990s[2], means that it is close to four decades since we have been able to get intact protein complexes into the gas phase and analyse them by means of MS[3–6]. This has led to an increased contribution over recent years to the related fields of genomics and proteomics[7]. In this thesis, we use MS techniques to probe the function and structure of complex proteins. The development of MS methods over recent years means that we are able to investigate

intact complexes in the gas phase in their native state, which is crucial for biological studies when we recognise that the majority of protein complexes do not exist as single entities[8]. The development of softwares to help deconvolute complex mass spectra was an important step toward fully automated MS analysis of proteins and the way they assemble[9, 10].

However, HDX-MS is used to probe the conformational dynamics and structural changes of proteins. By measuring the exchange of labile hydrogen atoms with deuterium in solvent molecules, HDX-MS provides a unique glimpse into the folding, unfolding, and interactions of proteins under physiological conditions. This technique is particularly adept at elucidating protein folding pathways, ligand binding, and protein-protein interactions[11].

Although both native MS and HDX-MS employ the use of MS, they serve different purposes in structural biology research. Native MS offers insights into the assembly and composition of macromolecular complexes, whereas HDX-MS provides detailed information on protein dynamics and conformational changes. Together, these techniques complement each other, enabling a comprehensive understanding of complex biological systems.

1.2 Native MS

In the realm of native MS, samples are delicately introduced into the instrument via an electrospray ionisation (ESI) source. The hallmark of this method lies in its ability to seamlessly transfer proteins from the solution to the gas phase while preserving their structural integrity throughout the duration of the analysis.

The emergence of native MS as a stand alone technique marks a transformative moment in structural biochemistry, granting us unparalleled insights into the intricacies of protein structure and function [12]. By maintaining proteins in their native states, this technique has become a powerful tool to unravel the complexities of biomolecular interactions and dynamics, which finds extensive application in characterising the composition, stoichiometry, and architecture of protein assemblies [13].

The utilisation of native MS underscores its role in advancing our understanding of biological systems, offering a clear view into the molecular world with unprecedented clarity and precision.

1.2.1 Sample Preparation

Before entering the mass spectrometer, samples undergo a preparatory step: buffer exchange into a highly volatile buffer solution. Ammonuim acetate is the most commonly used buffer, but others can also be used.

This preparatory measure serves a dual purpose: enhancing the desolvation and ionisation efficiency of the analyte in question while minimising the formation of adducts, which bind salt molecules to the analyte of interest, which can obscure the signal of interest.

The rationale behind the buffer exchange process is to transition the sample from a non-volatile, potentially contaminant-rich buffer to one with high volatility. This transition facilitates more effective desolvation during the ionisation process, improving the transfer of analyte molecules into the gas phase and minimising the interference of solvent molecules [12]. This optimisation enhances the accuracy and precision of the following mass measurements.

Moreover, the avoidance of heavily salted buffers during ionisation is critical in preventing adduct formation. Adducts complicate data interpretation by masking the signal of the analyte and introducing ambiguity into the results. Using a volatile buffer, the risk of adduct formation is minimised, ensuring that the resulting mass spectra provide a clear and unambiguous depiction of the characteristics of the analyte of interest [12].

In essence, the buffer exchange step in native MS plays a vital role in enabling us to acquire clear and precise data for subsequent analysis and interpretation.

1.2.2 Instrumentation

A mass spectrometer fundamentally consists of three main parts; an ion source, a mass analyser and a mass detector. The make up of commercial mass spectrometers

can differ by having numerous of these with in one instrument. This will allow for different capabilities of the instrument and allow greater separation of ions and therefore greater sensitivity and resolution.

1.2.2.1 Ion source - ESI

The most widely used ionisation method used for MS studies of protein complexes is ESI[14–17]. This is due to the simplicity of the ESI method, without alteration of the native complex. Compared with matrix-assisted laser desorption/ionisation (MALDI), which tends to require the evaporation of solvent from an aqueous analyte solution which is made up of a 10-10,000-fold excess of organic acid. These conditions cause the protein to be altered from its native state[7].

A vital component within the nanoESI (nESI) - name chosen to reflect the low flow rates involved and amount of sample needed[18] - source is the gold-coated needle, which assumes a multifaceted role in the ionisation process. Firstly, the gold coating serves to facilitate the addition of charge to the proteins, a fundamental step for the subsequent ionisation. In addition, the gold-coated needle plays a crucial role in dissipating the volatile buffer in which the sample is prepared. As said above, this buffer is typically composed of ammonium acetate - a volatile buffer. This dissipation occurs through a fascinating process of Coulombic repulsion, in which the charge within a droplet exceeds the Rayleigh limit[19]. This limit represents the threshold for the number of charges within a charged droplet before a fission process occurs, leading to its explosion, leaving behind an intact protein molecule [20].

In typical HDX-MS systems, we tend to use a regular ESI source. This increases the sample volume per run and, therefore, the amount of sample required. However, it does allow for a large throughput compared to conventional native mass spectrometry methods that would tend to employ the nESI source.

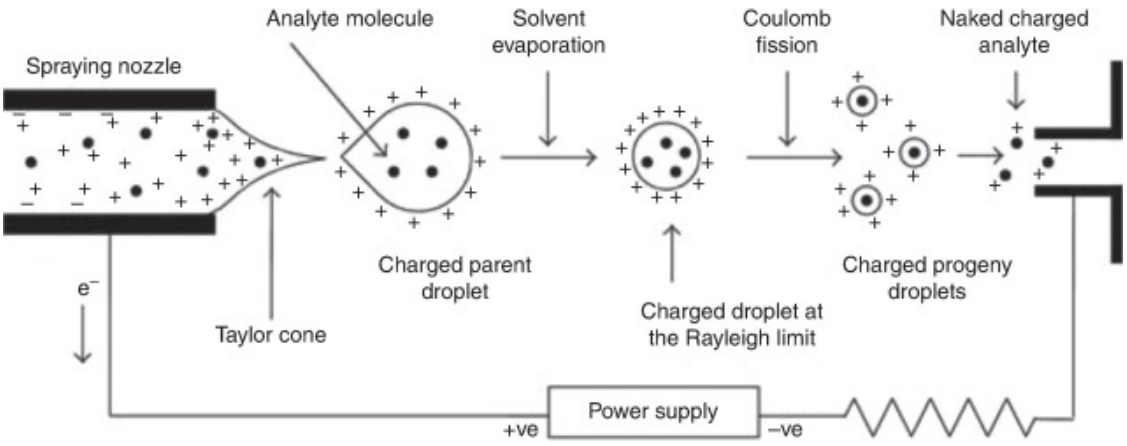


Figure 1.1: Schematic taken from the 2012 paper - A Technique to access the information beyond the molecular weight of the analyte [21]. The schematic shows the process involved within an ESI source to desolvate the sample. through

The schematic above shows the entire process that a sample undergoes through the ESI process, when entering the gas phase to be analysed on the mass spectrometer. The sample is subjected within the gold-coated needle to a high potential, which gives the sample charge.

1.2.2.2 Ion formation

The consensus in the literature is that no single model can explain all the experimental observations. Instead, different models or a combination of them may be needed to describe ion formation, depending on the specific ion being studied[22].

Two models commonly used to elucidate the formation of single ions are the ion evaporation model (IEM) and the Charge Residue Model (CRM). The IEM, initially proposed by Iribarne and Thomson, aims to elucidate the generation of atomic ions from randomly charged droplets created by a spray atomiser [20, 23]. According to this model, droplets undergo gradual shrinkage through evaporation until the electric field strength on their surface reaches a critical threshold, allowing the expulsion of solvated ions [23]. The IEM is well supported for small precharged species such as Na^+. Although IEM finds robust support for small precharged species such as Na^+, its applicability to proteins exhibiting IEM behaviour remains uncertain[24].

However, CRM posits that the electrospray process yields droplets, each containing a single analytical ion, which is liberated by solvent evaporation[20]. Even with

a low concentration of 1 pmol/µl, a droplet with a diameter of 200 nm typically houses fewer than one analyte molecule on average. A notable characteristic of this model is its assertion that the ionisation rate remains largely unaffected by the specific ion that is analysed[20].

Crucially, in the CRM, the availability of charge to the molecule hinges on the Rayleigh stability limit because the final droplet originates from a spraying process initiated by a Taylor cone. This understanding sheds light on the intricate mechanisms underlying the generation of single ions in ESI.

An additional model is used to explain the protein behaviour under non-native conditions. This model is called the chain ejection model (CEM)[25]. This model suggests that unfolded proteins, extended, are forced to the droplet surface by hydrophobic forces, and after that, the intermediates are ejected where the droplets carry extended protein tails of increased charge[25].

1.2.2.3 Mass Analysers

In MS analysis, a wide range of mass analysers have been developed. Only a small number have been applied to the analysis of large protein complexes[7]. Larger mass species have been known to be detected above 20,000 m/z[26–28].For these species, only certain mass analysers would suffice. whereas ions from individual peptides and denatured proteins typically lie below 3,000 m/z [7]. Some of the mass analysers commonly used in a typical experimental set up are as follows;

Quadrupole Mass Analyser

Quadrupole mass analysers, as their name implies, consist of four parallel cylindrical metal rods i.e poles, with hyperbolic interior surfaces. These rods are evenly spaced around a central axis. Ions traverse through the set of four rods functioning as a selective mass filter. This filtering mechanism operates on the principle of the ions' mass-to-charge ratio, which dictates their trajectories within the analyser. The fundamental operation of a quadrupole relies on the application of radio frequency (RF) and direct current (DC) voltages to non-adjacent rods in the set of four. This controlled application of electrical fields enables the quadrupole

to selectively guide ions based on their mass-to-charge ratios, ultimately allowing for precise separation and analysis of ions according to their mass properties[29].

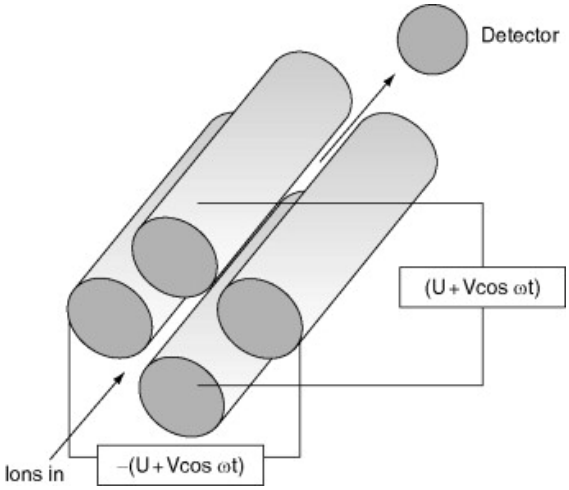


Figure 1.2: Schematic taken from Chapter 10 - MS in Contempory Practice in Clinical Chemistry[29]. showing the application and use of a Quadrupole mass analyser and how the DC voltage and RF voltage is applied.

Oscillating electrical fields are used to selectively stabilise or destabilise ions as they pass through an RF quadrupole field. Each pair of opposing rods is electrically connected, and an RF alternating current voltage is applied between the rods. A DC voltage is then superimposed on the RF voltage, which causes the ions to adopt an irregular, oscillatory trajectory, as they travel through the 3-D channel between the rods in a corkscrew path. Only ions with a selected m/z value are able to achieve a stable trajectory, which allows them to reach the detector[29].

The maximum m/z that can be resolved ($[\mathrm{m/z}]_{max}$) is given by

$$[\mathrm{m/z}]_{\mathrm{max}} = \frac{7 \times 10^6 V_m}{\nu^2 r_0^2} \tag{1.1}$$

V_m denotes the RF amplitude, ν is the RF frequency, and r_0 represents the inner radius between the quadrupole rods in metres. In mass-resolving mode, lowering the RF frequency (ν) increases the $[\mathrm{m/z}]_{max}$ that can be resolved by the quadrupole. This allows for the detection of higher m/z ions by adjusting the frequency accordingly[15, 30].

Quadrupoles can be used in many different experiments. They can help focus target ions toward the mass analyser or act as a selective mass filter to reduce the noise of unwanted analytes. In the majority of our HDX experiments, the quadrupole with in the synapt G2-Si is employed to filter masses in a range of around 200 - 2,000 m/z.

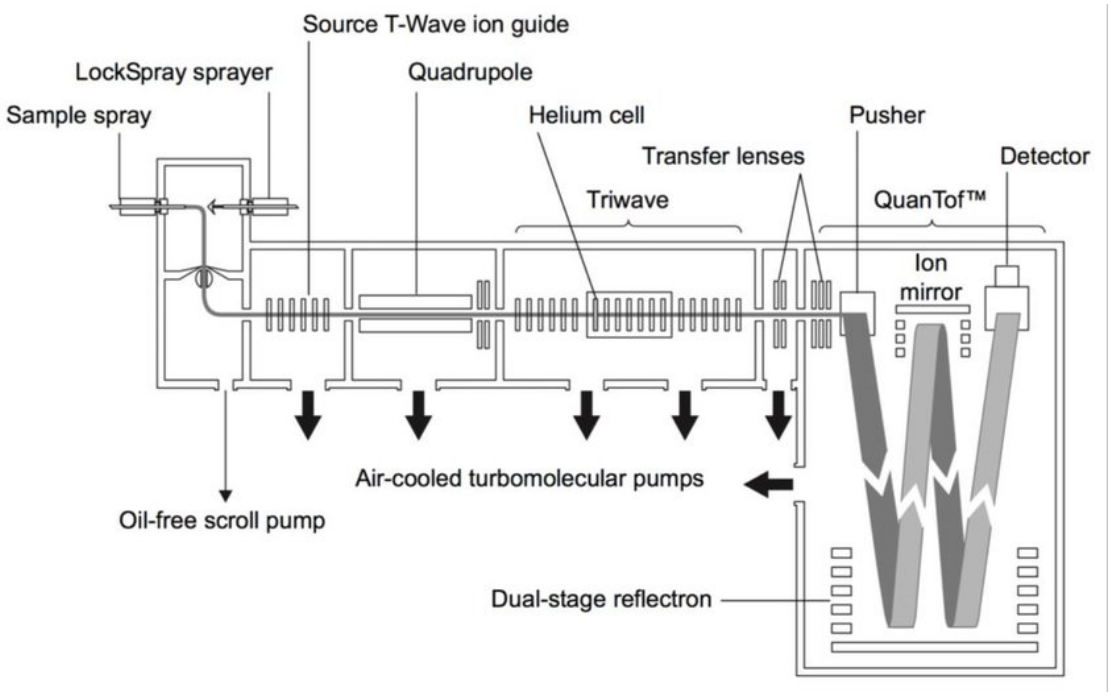


Figure 1.3: Schematic of Synapt G2-Si mass spectrometer altered for use from the waters website. Clearly showing different mass analysers and detectors used/available to use with in an experimental procedure.

In figure 1.3 above we can see the mass analysers used in a synapt G2-Si. As stated, the quadrupole is used as a filter in our work to decrease the level of unwanted noise. The helium cell can be used to increase signal separation through the use of ion mobility; however, it was not used throughout this thesis due to the fact the instrument was not set up or calibrated for such experiments. For future work, this is something that could be done to improve the quality of data collected, especially with the current advancements of instrumentation that is now available, namely the cyclic instrument discussed later.

Time-of-Flight (TOF) Mass Analyser

TOF analysers work by measuring the time it takes for ions of different m/z to travel a fixed distance in a vacuum under the influence of an electric field.

When a force is applied to a group of ions of differing m/z, separation of the ions is determined directly by their m / z as the ion velocity (v) is inversely related to the square root of m / z[29]. Here, (Q) is the interger charge of complex (Z) multiplied by the elementary charge (e).

$$v = \sqrt{(2V)(Q/m)} \tag{1.2}$$

Because of this, the greater the length of the drift tube can allow for greater separtion of the ion signal. This led to the use of reflectrons to increase the distance travelled and more recently circular drift tubes, which allowed for multi-pass experiments (multiple rounds through a circular drift tube to increase distance travelled, and therefore separation of analytes) on a cyclic instrument.This has been a very recent innovation of the instrumentation used in mass spectrometry experiments and has led to great improvements of the data acquired, particularly those in HDX-MS experiments.

Orbitrap Mass Analyser

Orbitrap analysers confine ions within an electrostatic field established between an outer barrel-shaped electrode and a central spindle electrode. Within this setup, ions undergo oscillatory motion around the central electrode, producing a frequency spectrum that correlates with their mass-to-charge (m/z) ratios. Renowned for their exceptional resolving power, mass accuracy, and sensitivity, Orbitrap analysers are well suited to analyse complex mixtures and facilitate high-throughput analyses.

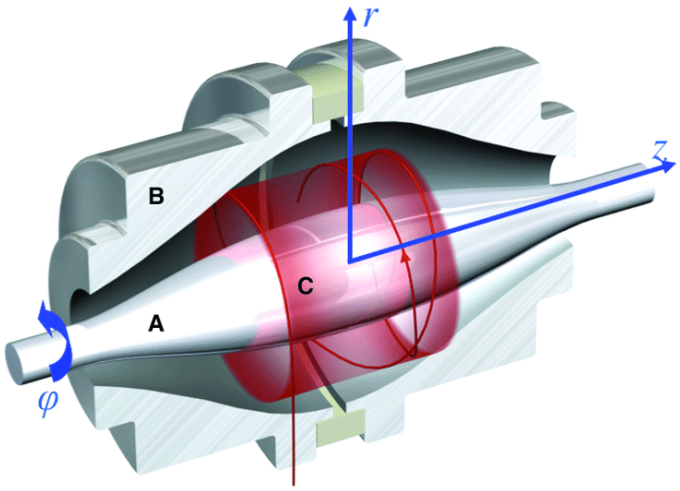


Figure 1.4: Schematic of an Orbitrap mass analyser showing: **a** spindle-like-shaped central electrode, **b** outer electrode, and **c** In red is an image current of ions moving between the two halves of the outer Orbitrap electrode, separated by the ceramic ring that goes around the orbitrap. Taken from the orbitra mass analyser - Overview and applications in proteomics[31].

Stable ion trajectories in an Orbitrap analyser are characterised by a combination of rotational motion around the central axis and axial oscillations. The parameter ω_z, which represents the frequency of these axial oscillations, depends solely on the ratio of the ionic charge q to the ionic mass m_i, as well as the strength of the electric field or curvature[32]. Remarkably, ω_z remains unaffected by the tangential velocity and spatial spread of the ions. This fundamental property underscores the independence of ion stability from factors such as velocity and spatial distribution, emphasising the robustness and reliability of Orbitrap analysers in accurately measuring the mass-to-charge ratio of ions.

$$\omega_z = \sqrt{(z/m_i) \cdot k} \tag{1.3}$$

.

m/z ration is simply related to the frequency of oscillation along the z-axis[33] as the frequency of such oscillation is inversely proportional to the square root of the mass-to-charge of the ions[32].

1.2.2.4 Mass Detectors

The term "limit of detection" (LOD), also known as the detection limit, may seem self-explanatory, yet it is often confused with sensitivity. The LOD denotes the minimum concentration or quantity of analyte necessary to produce a discernible signal above background noise. In any test / measurement, a sufficient concentration of analytes must be present to produce a signal that can reliably be distinguished from 'analytical noise', the signal produced in the absence of analytes (blank)[34]. This limit tends to be the amount of analyte required to produce the mean signal value of a blank sample +2 standard deviations[34].

While sensitivity plays a crucial role in achieving low detection limits, it is essential to recognise the distinction between sensitivity and the detection limit. Although sensitivity influences the LOD, the detection limit represents a separate measure. In trace analysis, the detection limit can be expressed in relation to the analytical context.

Several types of mass detector are employed in mass spectrometers, each offering unique capabilities and applications.

Photo Multiplier

In a photomultiplier conversion dynode detector, the detection process begins when ions collide with a dynode[35]. This collision induces the emission of electrons from the dynode surface. Subsequently, these emitted electrons impinge upon a phosphor screen, leading to the emission of photons.

The photons released from the phosphor screen then proceed through a multiplier section of the detector. Within this multiplier section, amplification occurs in a cascade-like fashion, akin to the mechanism observed in an electron multiplier. This amplification process involves the multiplication of electrons through a series of dynodes, resulting in a significant increase in the signal strength.

Overall, the photomultiplier conversion dynode detector utilises a multi-step process to convert ion signals into measurable electrical signals. Beginning with ion-induced electron emission from the dynode, followed by photon emission from the phosphor screen, and culminating in electron multiplication within the multiplier

section, this detector offers sensitive detection capabilities suitable for various analytical applications.

1.2.3 Example Experiments

As mentioned above, MS stands as a versatile technique with profound utility in structural chemical biology[36]. It empowers us as researchers to explore protein samples comprehensively, enabling the characterisation of various aspects such as composition and stoichiometries through intact mass analysis.

MS facilitates a diverse array of experiments, allowing us to investigate intricate molecular structures and dynamics in detail.

MS also supports the analysis of protein-protein and protein-ligand interactions, contributing to our understanding of complex biological systems. Techniques such as cross-linking MS and HDX-MS can reveal information about protein conformational changes, binding sites, and dynamic interactions. These insights are instrumental in fields like drug discovery, where understanding the interaction between proteins and potential therapeutic compounds is key to developing effective treatments.

In addition, MS can be used in quantitative proteomics to measure protein abundance changes in response to various stimuli or conditions. Techniques such as isotope labelling and label-free quantitation enable an accurate comparison of protein levels between different samples, providing insights into cellular responses and disease mechanisms.

In summary, MS is a powerful and versatile tool in structural chemical biology. It enables a comprehensive protein characterisation, a detailed analysis of molecular structures, and molecular dynamics, and it provides critical insights into protein function, regulation, and interactions. This makes MS an indispensable technique for advancing our understanding of complex biological systems and driving innovations in biotechnology, medicine, and related fields.

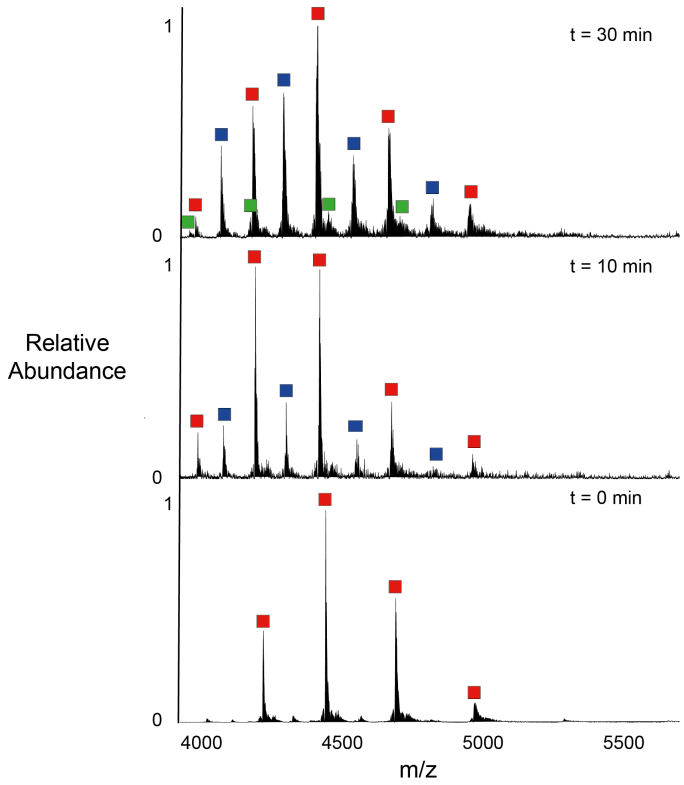


Figure 1.5: Using The treatment of Transferrin with PNGase F as an example for MS. Here we can see the gradual mass shifts and the subsequent alteration of the raw MS data acquired of Transferrin across a 30 min period of treatment with PNGase F

Here in figure 1.5 we observe the changes in transferrin mass (Tf) after treatment with PNGase F. Treatment of Tf with a PNGase results in the cleavage of one or two glycans bound to the transferrin amino acid backbone. This enzymatic cleavage leads to the appearance of two new masses over time. These new masses correspond to transferrin molecules that have lost one or two glycans, respectively. As the reaction progresses, the abundance of these new masses increases in relation to the native, apo glycosylated form of transferrin. This increase in abundance indicates that more transferrin molecules are undergoing deglycosylation as the treatment with PNGase F continues, providing a clear temporal profile of glycan removal. The masses are clearly indicated in figure 1.5 and labelled using a red square for the fully glycosylated molecules and blue and green squares for deglycosylated molecules N-1 and N-2.

1.3 HDX-MS

HDX-MS is a sophisticated analytical technique used to investigate the structural features and dynamic properties of proteins[37]. In HDX-MS, proteins are immersed in a solution containing deuterium oxide (D_2O), where deuterium atoms readily exchange with the labile amide hydrogens present in the protein backbone[38]. This exchange occurs rapidly under physiological conditions because of the kinetic isotope effect, where the heavier deuterium isotope replaces the lighter hydrogen isotope.

By exposing the protein to a highly deuterated environment, we effectively label the exchangeable hydrogen atoms in the protein with deuterons. After a defined labelling period, typically ranging from seconds to hours depending on the protein's dynamics, the exchange reaction is quenched by lowering the pH and temperature to halt further exchange. The protein is then enzymatically digested into smaller peptides while still in the deuterated state.

The resulting mixture of deuterated peptides is subjected to MS analysis, where the masses of the peptides and the extent of deuterium incorporation are measured. This allows us to pinpoint regions of the protein that undergo significant conformational changes or interactions with other molecules, as these areas will exhibit altered deuterium uptake compared to a reference state.

Through comprehensive data analysis and comparison between different experimental conditions or protein states, HDX-MS provides valuable insights into the dynamics and structural heterogeneity of proteins. This information is crucial for understanding various biological processes, protein-ligand interactions, and the mechanisms underlying diseases, making HDX-MS a powerful tool in structural biology and drug discovery.

1.3.1 Experimental Workflow

In a typical experimental workflow of HDX-MS, several key steps are involved, each crucial to obtaining meaningful structural insights into proteins.

The process begins with labelling, where the protein of interest is immersed in deuterium oxide (D_2O). This labelling initiates the exchange of labile amide

hydrogens with deuterons. Multiple labelling times are often used to capture the dynamics of protein structure over time.

Following labelling, a quenching step is essential to stop further exchange of deuterium atoms. This is achieved by diluting the labelled protein solution in an acidic buffer, effectively locking it in the deuterium that has already exchanged. The quench buffer may include additives to help unfold the protein, facilitating better digestion rates.

Digestion is then initiated, typically using pepsin as it remains active in the acidic pH range required for HDX-MS. Digestion can be carried out either online, using a protease pepsin column, or in solution using pepsin beads. Recently, alternative proteases for digestion have been explored, offering flexibility in experimental design.

The protein is broken down into peptide fragments during digestion. These peptides are subsequently separated using an Ultra-High Performance Liquid Chromatography (UHPLC) setup before being injected into the mass spectrometer for analysis.

The analysis of the data obtained involves several software tools. Peptide fragments are initially identified using software such as ProteinLynx Global Server (PLGS). Subsequently, the uptake values are calculated by comparing the levels of deuterium incorporation at different time points to the undeuterated (t_0) state.

The final step involves data interpretation and manipulation. Various analytical techniques can be employed to extract meaningful information, allowing us to elucidate structural changes and dynamics within the protein of interest. This comprehensive workflow enables us to gain valuable insight into protein conformational dynamics and interactions, contributing to our understanding of biological processes and facilitating drug discovery efforts.

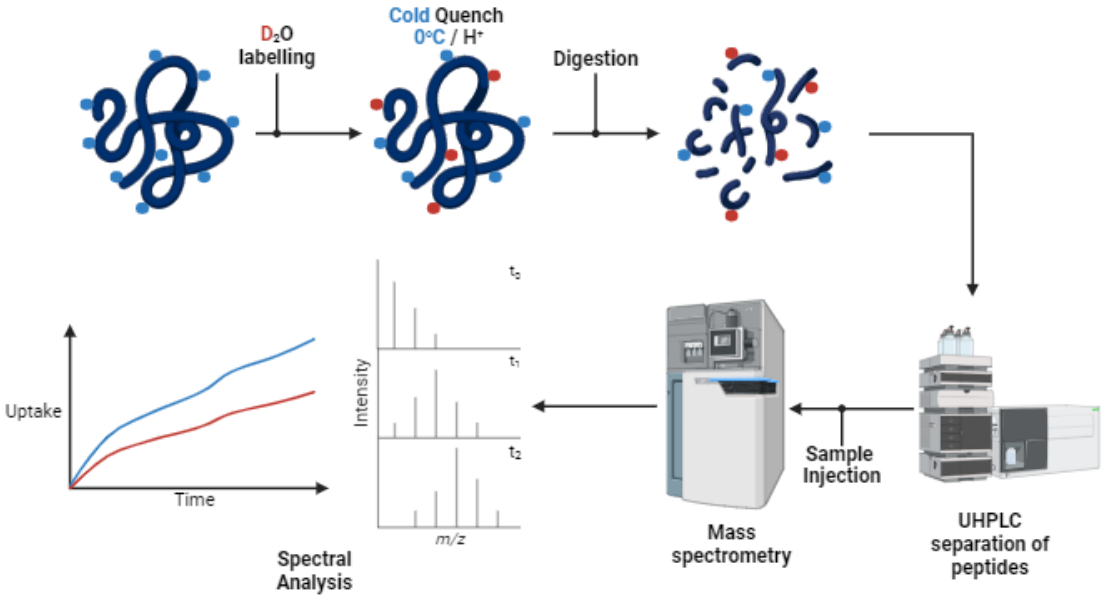


Figure 1.6: Typical HDX workflow Here is the typical workflow for an experimental HDX set up.Showing steps of labelling initiation, quenching, digestion, peptide separation, sample injection and data analysis.

Figure 1.6 shows a typical complete experimental workflow from start to finish. In the subsequent pages of this thesis we will look at each step shown in further detail. Some experimentalists have been known to adapt some of these steps for differing research goals, as we have done in the chapter C. For the purpose of this introductory chapter, we will focus on the most basic setup.

1.3.2 Labelling

To begin labelling, the protein sample is introduced into a solution rich in deuterium oxide (D_2O), meticulously designed to mirror the composition of the equilibration buffer. Within this deuterated environment, the amide hydrogens of the protein amide backbone undergo an exchange with deuterium atoms. This exchange sets the stage for subsequent analysis, enabling the identification of regions within the protein that display differential uptake when comparing distinct states. By precisely controlling the labelling conditions, we can capture the intricate dynamics of protein structure and discern subtle changes in the incorporation of deuterium, which serve as indicators of conformational alterations or molecular interactions. This strategic

approach in HDX-MS lays the foundation for comprehensive investigations into the structural dynamics of proteins, ultimately advancing our understanding of biological processes and informing drug discovery efforts.

Calculating exchange rates between hydrogen and deuterium reveals two different kinetic regimes, as outlined in a study by Vinciauskaite et al. (2023) [39]. These regimes are predicated on the prevailing forward reaction of hydrogen to deuterium in a highly deuterated environment, where the concentration of deuterium significantly exceeds that of hydrogen.

The rate of exchange (k_{ex}) can be expressed using the equation:

$$k_{ex} = \frac{(k_{op})(k_{ch})}{(k_{cl} + k_{ch})} \tag{1.4}$$

Here, k_{op} represents the rate of opening, k_{cl} denotes the rate of closing, and k_{ch} stands for the chemical rate constant of individual backbone amides.

In the first scenario (EX1), where k_{ch} greatly surpasses k_{cl}:

$$k_{ch} >> k_{cl} \tag{1.5}$$

The exchange rate approximates the rate of opening:

$$k_{ex} \approx k_{op} \tag{1.6}$$

Consequently, the mass envelopes in the deuterium uptake display a bimodal distribution in the raw data, indicative of two distinct populations[39].

In contrast, in the second scenario (EX2), where k_{ch} is significantly smaller than k_{cl}:

$$k_{ch} << k_{cl} \tag{1.7}$$

The exchange rate can be approximated as the product of kch and kop:

$$k_{ex} \approx k_{ch}k_{op} \tag{1.8}$$

This leads to a gradual shift of one mass population towards higher masses in the deuterium uptake, reflecting a more continuous distribution of exchange rates[39].

Understanding these distinct kinetic regimes and their implications on mass envelopes is essential to accurately interpret HDX-MS data and unraveling the dynamics of protein structures.

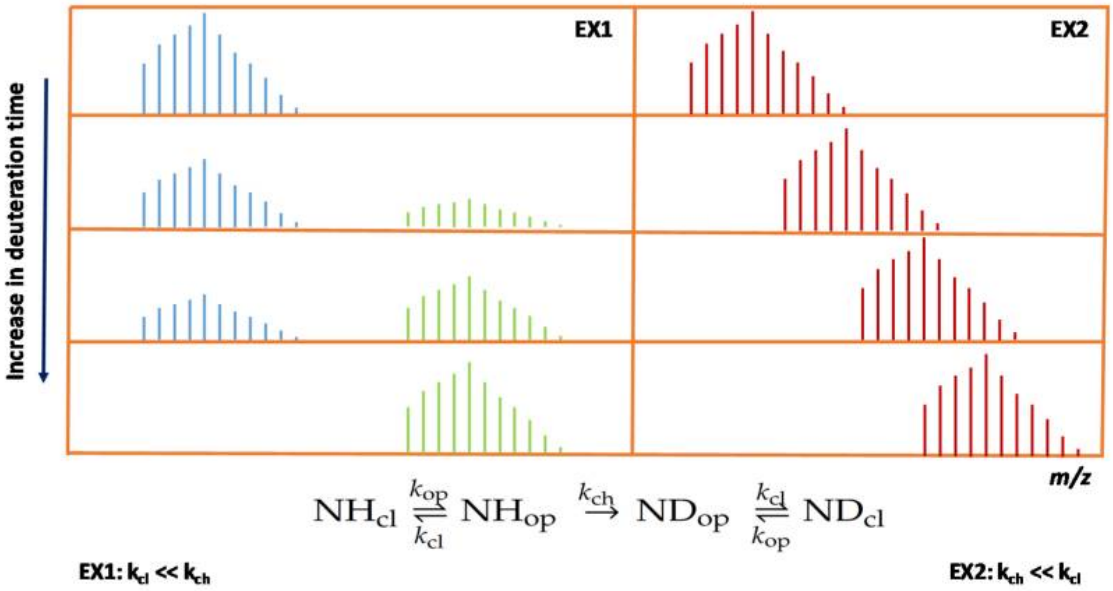


Figure 1.7: Schematic example of how raw EX1 data looks compared to raw EX2 data. showing clearly over time the appearance and disappearance of the bimodal nature of EX1 behaviour vs the steady shift toward higher m/z for EX2

Here is a visual representation taken from HDX MS: A Novel Structural Biology Approach to the Structure, Dynamics, and Interactions of Proteins and Their Complexes[37], illustrating the concepts described earlier. In EX1 scenarios, distinct bimodal peaks emerge during labelling experiments, indicating the presence of two separate populations within the protein sample. This phenomenon occurs because some regions of the protein exchange hydrogen for deuterium quickly (fast exchange), while others exchange at a slower rate (slow exchange), resulting in two distinct m/z populations in the MS data.

In contrast, in EX2 scenarios, as labelling time increases, there is a noticeable gradual shift toward higher values of m / z. This gradual increase reflects the progressive incorporation of deuterium atoms into the protein structure over time. Here, the exchange rates across the protein sample vary more continuously, resulting in a single population of peptides shifting to higher m/z values in the mass spectra.

These visual patterns in the MS data are indicative of the differing kinetic behaviours observed in HDX-MS experiments. They provide valuable insights into the dynamics of protein structure and the accessibility of amide hydrogens to solvent exchange. Understanding these patterns is crucial for interpreting the biological implications of protein dynamics and stability under varying experimental conditions. In this thesis, we did not see an example of EX1 style kinetics in our HDX-MS data across the systems that we investigated.

1.3.2.1 Continuous vs Pulsed labelling

Most HDX-MS experiments involve continuous labelling due to the fact that they are much easier to perform technically[40–43]. However, this does not mean that it is the only type of HDX-MS experiment available[44].

In a continuously labelled experiment, the protein is exposed continuously to D_2O levels and the relative intensities of two envelopes of isotopic peaks give an integrated account of populations participating in the folding/unfolding process[45]. In contrast, in a pulsed experiment the protein is only briefly exposed to D_2 O and then the two envelopes give an instantaneous snapshot of the folded/unfolded populations[45].

Pulsed labelling experiments have been known to be used down to the ms time scale. allowing for potential of previously unseen kinetics to be examined[46].

1.3.3 Quenching

Within the HDX-MS workflow, the quenching step plays a critical role in ensuring the acquisition of precise and biologically relevant data. Its primary objective is to stop further HDX reactions within the protein sample, thereby preserving the deuterium-labelled states for subsequent analysis. This preservation is crucial to accurately capture the structural dynamics of the protein under study.

The quenching process involves lowering the pH of the solution to approximately 2.5. At this acidic pH, the exchange of labile hydrogens with deuterons is significantly suppressed. The low pH environment stabilises the deuterium-labelled states by minimising the mobility of amide hydrogens, effectively "locking" the exchanged

deuterons in place along the protein backbone. This immobilisation ensures that the structural information captured during the labelling phase is retained and accurately reflects the protein's native conformation at the time of quenching.

Additionally, the acidic quench solution often contains additives such as urea or guanidinium chloride. These additives assist in unfolding the protein structure, which enhances the accessibility of amide hydrogens for labelling during the initial phase of the HDX experiment. By optimising the quenching conditions, we are able to maximise the fidelity of our HDX-MS data, enabling precise mapping of protein dynamics and interactions under physiological or experimental conditions.

In summary, the quenching step in HDX-MS is pivotal for stabilising deuterium-labelled protein states and minimising further exchange, thus ensuring the reliability and accuracy of subsequent data analysis and interpretation.

The graph presented below is an adaptation of the graph shown in "Bottom-up HDX MS: data analysis and interpretation" by Brown et al. [47]. It illustrates a significant observation regarding the rate of HDX (k, in s^{-1}) as a function of pH. In particular, the exchange rate exhibits a decreasing trend, reaching a minimum around pH 2.5. This pH range corresponds to a point where both acidic and basic catalysed exchanges of hydrogens are minimised.

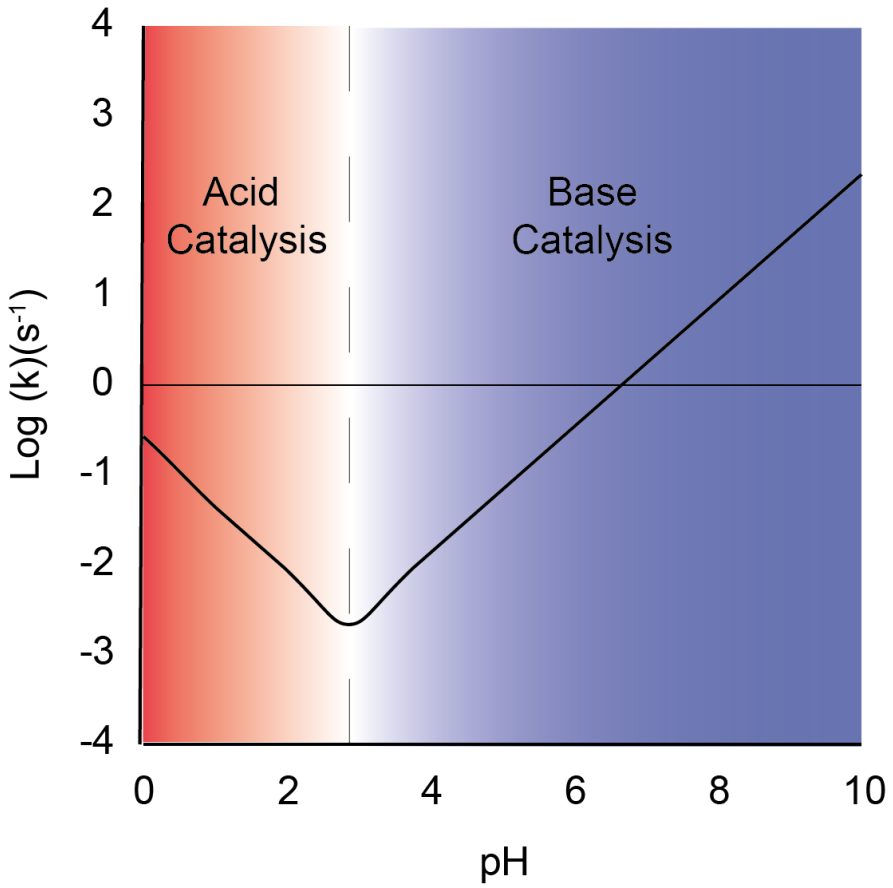


Figure 1.8: Visual representation of how the rate of exchange of labile hydrogens changes with an increase and decrease of pH centred around a minima of ≈2.5. Explaining the reason that we reduce the pH during the quench process.

At lower pH values, such as pH 3, the environment is sufficiently acidic to suppress the rate of hydrogen exchange. Acidic conditions favour protonation of basic residues and reduce the availability of hydrogens for exchange. Similarly, at higher pH values, basic catalysed exchange processes become more prominent because of the deprotonation of acidic residues, facilitating faster hydrogen exchange rates.

Understanding the pH dependence of amide hydrogen exchange rates is crucial within HDX-MS experiments. It informs the experimental design by influencing the choice of buffer conditions to control exchange rates and maximise the sensitivity of detecting structural changes in proteins. Moreover, this knowledge aids in interpreting experimental results, particularly in elucidating the solvent accessibility and dynamics of protein structures under varying pH conditions.

Moreover, the quench solution often incorporates additives such as urea or guanidinium chloride (GdmCl). These agents serve a dual purpose: firstly, they facilitate the unfolding of protein structures, and enhance the digestion efficiency in the subsequent step. This meticulous quenching process is instrumental in preserving the structural fidelity of proteins and facilitating an accurate interpretation of the HDX-MS results.

1.3.4 LC-MS Separation

Liquid chromatography coupled with MS (LC-MS) is a powerful analytical technique that integrates the separation capabilities of liquid chromatography with the mass analysis capabilities of MS [48]. At its core, LC-MS operates on the principle of exploiting differences in the interaction between analytes and a stationary phase within a chromatographic column.

In LC-MS, a liquid mobile phase is directed, often under high pressure, through a chromatographic column packed with a stationary phase. This stationary phase can consist of various materials, such as silica or polymer-based resins, depending on the specific separation requirements. As the sample passes through the column, different analytes interact differently with the stationary phase, leading to their separation on the basis of factors such as polarity, size, and affinity.

Following separation, elution of the analytes is achieved through the application of a gradient. This gradient typically involves the use of pH-adjusted water or organic solvent mixtures, such as water-acetonitrile or water-methanol blends. By systematically altering the composition of the mobile phase, linear or non-linear, analytes are selectively eluted from the column based on their respective affinities with the stationary phase[48].

Once separated, the eluted analytes entered the mass spectrometer for analysis. Here, they undergo ionisation and subsequent mass analysis, enabling the determination of their molecular masses and structural characteristics. The combination of

liquid chromatography and MS thus provides a robust platform for the comprehensive analysis of complex mixtures, making it invaluable in a wide range of scientific disciplines, including proteomics, metabolomics, and pharmaceutical analysis.

1.3.5 Data Analysis

Analysing HDX-MS data is a laborious task, which over recent years has prompted the development of numerous software solutions over recent years to streamline and expedite the process. As discussed extensively in Chapter 2, these software tools offer invaluable assistance in unraveling the intricacies of protein dynamics captured through HDX.

At its core, the objective of HDX-MS data analysis is to discern variations in deuterium uptake across different regions of the protein under varying experimental conditions. This entails comparing the extent of deuterium incorporation at specific amino acid residues or peptide segments between different states of the protein. However, the crux of the analysis lies not only in identifying these differences but also in deciphering the underlying reasons behind them.

To achieve this, sophisticated algorithms and statistical methods are employed to correlate the observed changes in deuterium uptake with structural or functional alterations within the protein. Factors such as protein-protein interactions, ligand binding, solvent accessibility, and conformational dynamics all contribute to the differential deuterium labelling patterns observed in HDX-MS experiments. By meticulously dissecting these intricate relationships, we can unravel the molecular mechanisms driving protein behaviour and gain deeper insights into biological processes.

Therefore, while HDX-MS data analysis may indeed be time-consuming, the use of specialised software tools allows us to extract meaningful information from complex datasets efficiently.

1.3.5.1 Back exchange

To ensure that your LC-MS system is of a level of suitability that will allow for reproducible data, some would recommend doing a back exchange analysis to ensure deuterium recovery is consistent. Although with some experiments it may not be needed it is considered good experimental practice to do so as it ensures that a suitable level of deuterium is retained across the experiment[44].

The back exchange of a system can be calculated using the following equation[40, 42, 44, 49]

$$Backexchange = (1 - \frac{m_{100\%} - m_{0\%}}{N \times D_{frac}}) \times 100\% \tag{1.9}$$

where $m_{0\%}$ is the mass of the non-deuterated peptide, $m_{100\%}$ is the maximum labelled peptide mass, N is the number of hydrogens of theoretical exchangeable backbone amide in the peptide, and D_{frac} is the fraction of Deuterium in the labelling buffer[44].

1.3.6 Challenges Still Prevalent with the HDX Technique

Analyis and presentation; In the HDX community, there remains an ongoing quest to establish the most effective and "correct" approach to conveying and interpreting findings. Despite the availability of numerous software tools, many of which have been developed in recent years to facilitate the analysis of large datasets (as discussed in Chapter 2, a consensus on the optimal method to present results has not yet been reached.

Our research goals/aims tend to influence the ways in which we analyse and present our data. leading to there not being a single particularly agreed upon way in which to do so.

The key considerations to determine the most appropriate presentation format include the complexity of the data, the intended audience, and the overarching research goals. We may opt for visually engaging graphical representations to highlight key findings and trends, while others may favour detailed tables or statistical analyses to provide a comprehensive overview of the data.

Ultimately, the choice of presentation format should aim to effectively communicate the significance of the experimental findings and facilitate the audience's understanding and interpretation. By tailoring the presentation approach to suit the specific requirements of the research inquiry, we can maximise the impact and utility of their findings within the scientific community.

Complex protein samples; A growing concern within the scientific community revolves around enhancing the ability to address complex research questions by optimising the analysis of intricate protein solutions. Recent attention has been directed towards refining the HDX workflow to better accommodate the study of challenging protein types, particularly membrane proteins and heavily glycosylated proteins.

In Chapter C, we dive deeper into the advances made in this area, specifically focussing on strategies to improve the analysis of highly glycosylated proteins. These proteins present unique challenges as a result of the presence of complex carbohydrate structures attached to their amino acid residues. The intricacies associated with glycan structures necessitate innovative approaches to ensure the accurate characterisation and interpretation of HDX data.

As scientists, we have been actively exploring novel methodologies and analytical techniques to address these challenges. This includes the development of specialised protocols for the enzymatic cleavage of glycans prior to HDX analysis, as well as the utilisation of advanced MS techniques capable of distinguishing between glycosylated and nonglycosylated peptides.

By enhancing our ability to analyse complex protein solutions, particularly those containing membrane proteins and heavily glycosylated proteins, we aim to unlock new insights into fundamental biological processes and disease mechanisms. These advances have the potential to significantly impact various fields, including structural biology, drug discovery, and biomarker identification.

1.3.6.1 Understanding Differences of Deuterium Uptake

Effective data analysis in HDX-MS experiments is pivotal for elucidating significant findings. This process relies on identifying and interpreting differences observed under various experimental conditions, providing insights into structural dynamics and interactions within proteins.

The HDX rate within specific peptides is intricately linked to the protection factors of individual amide hydrogens within the quaternary structure. These protection factors are extensively discussed in Chapter 3 and reflect the stability and accessibility of amide hydrogens to the solvent environment.

During experimental perturbation, such as ligand binding or other stimuli, variations in deuterium uptake become apparent. These differences signify structural alterations within the protein in specific regions influenced by external factors introduced in differential states[50].

"The only true wisdom is in knowing you know nothing."

— Socrates

2

Development of our own analytical aid

Contents

2.1 Abstract 29
2.2 Introduction 30
2.2.1 Different Analysis tools 31
2.2.2 Statistical Analysis 35
2.2.3 Challenges in presenting/Analysing Data 37
2.3 SAUSC 38
2.3.1 T-test 38
2.3.2 Global Threshold 39
2.3.3 Hybrid Test 39
2.3.4 Confidence levels 40
2.3.5 Visual Aid - Structural Relevance 41
2.4 Results 45
2.4.1 Figure Plotting 46
2.5 Discussion 55

2.1 Abstract

This chapter introduces SAUSC, an analytical framework for decoding peptide dynamics across diverse experimental conditions and time points. By combining advanced statistical methods with clear visualisations, SAUSC allows researchers to identify significant patterns in peptide behaviour. Central to the framework are Woods plots, which map spatial peptide patterns along protein sequences, and

Volcano plots, which highlight statistical significance and the magnitude of uptake differences; both plots can be visualised from individual labelling times as well as using the additional cumulative plots. All plots use consistent colour gradients to improve data interpretation. SAUSC's systematic and iterative approach enables deeper insights into peptide behaviour, providing a robust tool for understanding protein function in biological systems.

2.2 Introduction

In the HDX community, there is an ongoing and nuanced discourse surrounding the optimal methodologies for handling and presenting collected data. Given the widespread adoption of this technique within the scientific community to address various problems, the analysis, interpretation, and presentation of data pose considerable challenges in achieving both completeness and clarity.

Due to the subjectivity in determining a universally "correct" or "complete" approach to data handling, we directed our efforts toward developing an analytical tool that we could use to gain structural information of the protein at question in a quick and simple manner that is easily presented through the use of visual aids. We hoped that this would empower users through the ability to dynamically adjust and fine-tune various parameters and thereby facilitate a comprehensive representation of experimentally acquired information in a clear and concise way.

The tool in which we developed was inspired by a number of analytical softwares already available, which we thought that could be expanded further. Across comparative experimental states, we perform a full cross-state analysis using data available from the raw exported csv data from Dynamx. Each sequence (peptide) identified from the (non-deuterated) peptide mapping run will undergo a state comparison, and differences will be deemed significant/insignificant through a standard statistical test.

A notable feature of our tool is the user's ability to specify a desired level of confidence. This parameter significantly influences the selection and application of agreed-upon statistical tests, as well as the desired filtering of significant and

insignificant findings. By affording users the flexibility to tailor the tool to their specific needs, we enhance the depth and granularity of the data portrayal.

Within our analysis tool, the adaptability of the PDB colouring scheme is further enhanced through a normalisation process that incorporates both global (across exposures) and time-point dependencies.

The global and time-point dependencies embedded in the normalisation process enable users to maintain a coherent and consistent colour scale across diverse datasets. By establishing a baseline that considers the entirety of the dataset globally (across all exposures), our tool ensures a comprehensive understanding of the peptide dynamics. Simultaneously, the time point dependence enables users to discern temporal variations and trends within the experimental data.

2.2.1 Different Analysis tools

What should we represent in our findings? Is there a right or wrong way to represent our findings? What do we see as being complete?

Currently, a large number of HDX analytical tools are available for HDX users. All of which give the user differing outputs and capabilities. Again, we highlight the issue of what a total and complete analysis consists of. To list but a few of the tools that are commonly used;

- Deuteros 2.0
- HDXanalyzer
- HX-Express v2
- HDX-Viewer
- MEMHDX

2.2.1.1 Deuteros 2.0

Deuteros 2.0 was developed to offer a quick and informative comparison of HDX-MS datasets. This was to fill the void of tools available that offer both statistical analysis and advanced visualisation [51]. This served as inspiration for our own analysis tool in SAUSC.

Deuteros 2.0 showcases a comprehensive suite of capabilities designed to elevate the analytical process in HDX-MS. This latest version introduces a spectrum of advanced functionalities, including robust mechanisms for back-exchange correction, sophisticated data summarisation techniques, in-depth statistical analyses at the peptide level, and a suite of cutting-edge features for intricate data visualisation.

One of the distinguishing features of Deuteros 2.0 lies in its ability to address back-exchange correction, a critical aspect of HDX-MS data analysis. The tool employs meticulous algorithms to rectify and enhance the accuracy of results by accounting for the influence of back-exchanged hydrogen atoms, ensuring the fidelity of the acquired data and accounting for any instrumental/experimental set up issues that could lead to an increased level of back exchange.

Furthermore, Deuteros 2.0 introduces a refined approach to data summarisation, streamlining the assimilation of complex datasets. Its intelligent algorithms distill voluminous information into concise and meaningful summaries, providing researchers with a comprehensive, yet easily interpretable, overview of their experimental findings. This is done by interpreting the data collected across each charge state of each peptide and comparing those across each experimental state to gain the uptake values of each peptide which are then deemed to be significant or not significant through the statistical test the user deems fit to use.

The peptide-level analysis capabilities of Deuteros 2.0 set it apart as a powerful tool for discerning subtle variations and trends within the data. Researchers can delve into granular insights at the peptide level, enabling a more nuanced understanding of the experimental outcomes.

In addition to its analytical prowess, Deuteros 2.0 offers a variety of advanced data plotting features. These functionalities go beyond conventional visualisation

methods, allowing users to create intricate plots that facilitate a deeper exploration of the underlying patterns and relationships in their data; this is something that we found very attractive with in our analysis and something we tried to build on with the SAUSC script.

2.2.1.2 HDXanalyser

HDXanalyser emerged in response to a notable gap in the landscape of HDX software, specifically the absence of robust and clearly defined statistical treatment methods prevalent in existing tools at the time of its publication. HDXanalyser, serves as a sophisticated software solution meticulously crafted using a combination of R, Python, and the RPY2 interface[52]. There are three main modules in the HDXanalyser package;

- Data Processing

  Implemented in Python, this software generates centroids represented by m/z values and computes the centroid differences between comparative states. The purpose is to discern significant changes in specific regions, providing a quantitative measure to evaluate variations under different experimental conditions. Through this approach, the software serves as a powerful analytical tool, providing a better understanding of molecular differences and facilitating the identification of regions undergoing significant alterations in mass spectrometry data [52].

- Statistical analysis

  Developed using the R programming language, this tool employs two distinct statistical analyses: the paired Student's t-test and the linear combination of the intercept for both multiple regression and ANCOVA models. This dual analytical approach improves the software's ability to effectively explore and interpret data, providing researchers with a robust toolkit to conduct in-depth statistical assessments within the R environment [52].

- The WxPython user interface is used to enhance the user interface, enabling seamless integration for data visualisation in graphs and the presentation of statistical analysis outputs. This integration optimises the user experience, providing a cohesive platform for effective data interpretation and display of the results of the analysis[52].

2.2.1.3 HX-Express v2

HX-Express v2 represents a significant advancement in providing a more comprehensive analysis. Although numerous tools excel in analysing peptides with clear distinctions and no overlap, HX-Express v2 distinguishes itself by addressing the challenge of overlapping peptides. Developers have implemented an alternative, and by their own characterisation, a simpler approach to extract deuterium levels from overlap peptides, thus expanding the scope of analysis and enhancing the sequence coverage[53]. The strategy combines and streamlines the binomial distribution fitting described by Chik et al.[54] and asymmetric least squares regression, as commonly used in baseline correction algorithms[55]. This highly efficient method also functions as a rapid approach to deconvoluting bimodal spectra, such as EX1 kinetic signatures. This holds even for data sets characterised by low signal-to-noise ratios or the presence of overlapped species[53].

2.2.1.4 HDX-Viewer

HDX-Viewer stands out as a user-friendly tool that eliminates the need for prior programming expertise. It distinguishes itself by enabling the visualisation of individual state data, which separates it from other visual aid tools such as MEMHDX. Unlike MEMHDX, which is limited to differential analysis and restricted to peptides passing statistical tests, HDX-Viewer provides a more comprehensive platform for single-state data visualisation [56]. This is designed to enhance users' protein analysis by allowing for improved biological interpretation through enhanced visualisation capabilities.

2.2.1.5 MEMHDX

This interactive tool facilitates seamless and expeditious validation, visualisation, and comparative analysis of relative deuterium incorporation within the amino acid sequence and 3D structural framework. By providing nuanced insights into both spatial and temporal dynamics, it serves as an invaluable resource for in-depth protein analysis[57].

The estimation of fixed effects vectors β and the parameter u_r involved the utilisation of the 'nmle' R package. Employing the restricted maximum-likelihood method, a linear mixed-effects model was fitted to discern patterns in deuterium incorporation rates across conditions. To assess the statistical significance of both the condition and its impact on the interaction on protein deuterium uptake, they conducted dual Wald tests, which yielded two distinct P-values per peptide. The condition-associated P-value scrutinises the null hypothesis of no disparity in deuterium uptake between conditions 1 and 2, while the interaction-associated P-value assesses the null hypothesis regarding the constancy of the deuterium uptake behaviour between conditions 1 and 2, accounting for the temporal dynamics inherent in the HDX reaction[57]. This method is designed to effectively regulate the proportion of hypotheses rejected erroneously, ensuring a robust control mechanism to maintain statistical integrity and minimising Type I errors in hypothesis testing[58].

2.2.2 Statistical Analysis

In the realm of HDX-MS workflows, a multitude of potential sources converge, each capable of introducing error into the analytical process. These sources span from the intrinsic characteristics of the sample itself to the intricacies of instrumental calibration and data manipulation, all of which require attention and require an understanding for their effective mitigation.

In particular, the unique composition and dynamic behavior of the sample wield significant influence, with factors such as protein stability, solvent interactions, and structural heterogeneity playing pivotal roles. As samples progress through pre-injection handling, meticulous care is essential to maintain the sample integrity,

regulate environmental conditions, and minimise any extraneous sources of variability that may compromise the quality of data that are to be collected.

Moreover, the precision and accuracy of mass measurement techniques constitute critical considerations, necessitating vigilant oversight of instrumental parameters such as ionisation efficiency and resolution, alongside robust calibration protocols to ensure reliable performance over the duration of the experiment. Currently, rigorous calibration routines are imperative to counteract instrumental drift and maintain optimal analytical performance.

Finally, effective handling and interpretation of the large amounts of data generated require sophisticated computational approaches, which include data pre-processing, noise reduction, and advanced statistical analyses to distill meaningful insights from complex datasets [59].

When these issues are coupled, they lead to the need for stringent statistical measures to be implemented to minimise the false-positive rate at which we define the peptide uptake behaviour to be inherently different across differing experimental states.

2.2.2.1 Types of error

When handling data, there are a number of different errors that can arise when applying statistical testing. This can become a problem when analysing HDX-MS data as you are usually handling huge data sets with numerous differential states, time points, and you are wanting to make inferences about the properties of a population within that data set. For example, does adding a binding partner create a significant difference in Deuterium uptake at any particular point within a protein.

The use of random samples offers significant advantages because of the impracticality of measuring the entire population in most cases. However, there are trade-offs associated with sample usage. These samples typically represent only a tiny fraction of the total population. Consequently, they may occasionally lead to substantial misrepresentations of the population, increasing the likelihood of Type I and Type II errors in hypothesis tests.

A Type I error occurs in research when the null hypothesis is rejected, leading to the mistaken conclusion that significant differences exist, when in reality there are none. Essentially, this means falsely claiming that the groups or variables differ, resulting in a false positive[60]. So in relation to HDX-MS data, if we did not eradicate errors of type I fashion, we may wrongly conclude that there were differences present across the experimental states when, in fact, there were no differences.

On the other hand, a type II error occurs when we declare that there were no differences between the study groups, when, in fact, there were[60]. This would drastically minimise the power of our results. Power is the ability to correctly reject a null hypothesis that is in fact false[61].

Null hypothesis is...	True	False
Rejected	Type I error - False positive	Correct Decision
Not rejected	Correct Decision	Type II error - False negative

Table 2.1: Hypothesis testing and different types of error that can occur through statistical testing.

2.2.3 Challenges in presenting/Analysing Data

As evidenced by the vast array of diverse analytical tools available and showcased above, each offering unique capabilities for data analysis, it becomes apparent that within the HDX community, no single tool encompasses the full spectrum of analytical capacity. Rather, existing tools tend to be specialised, with some emphasising visual aids for structural representation, while others prioritising the extraction of kinetic insights from raw data at the peptide level.

The ensuing challenge revolves around determining the required output to glean comprehensive insights from the data garnered through experimental work. Often, the nature of the inquiry within one's work dictates the necessary approach. In structural studies, for example, there is often a greater demand for rich visual aids while retaining the integrity of the raw data information.

2.3 SAUSC

In order to tackle the above challenges, this project focused on developing SAUSC, a standalone script designed to analyse complex variations in higher-order protein structures. The script is compatible with PyMOL and requires only a .csv file exported from DynamX.

SAUSC enhances flexibility by allowing users to apply various colouring schemes to visualise multiple parameters on output figures and structural models. It also includes normalisation options to handle differing maxima, aiding in the interpretation of structural dynamics over time. Users can preset statistical tests and adjust confidence levels, providing control over the depth and precision of the analysis.

In the following section, we will guide you through the entire process, including the calculations used, the adjustable parameters available, and the way we generate visual output.

2.3.1 T-test

The T-test used in SAUSC is the Welch's T-test. This is a variation of the T-test that does not require equal standard deviations across the two samples. Suitable for instances where the standard deviation value may differ significantly, which can occur across differential experimental states.

$$t = \frac{\Delta\overline{X}}{s_{\Delta\overline{X}}} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{s^2_{\overline{X_1}} + s^2_{\overline{X_2}}}} \tag{2.1}$$

This is the test in which we gain a calculated p-value which we use throughout our portrayal of the findings in the figures then created. The p-value is a significant number in statistical testing, as it signifies the likelihood of the observations found by chance. i.e., the lower the p-value, the more statistically assured you can be that the result is true, therefore being able to reject or accept the null hypothesis. The t score calculated above gives us a corresponding p-value.

2.3.2 Global Threshold

The Global threshold test was a test developed by Hageman and Weis [59]. This test threshold is determined by the quality of data across the entire data set collected. It is determined using a few simple statistical measurements to obtain a propagated standard error of the mean (SEM)[59].

2.3.2.1 Determining Global Threshold value

- Calculate a pooled sample standard deviation (s_p) for all N experimental standard deviations (s_j) and corresponding sample sizes (n_j)

$$s_p = \sqrt{\frac{\Sigma_{j=1}^{N}(s_j(n_j-1))}{\Sigma_{j=1}^{N}(n_j-1)}} \tag{2.2}$$

- calculate propagated standard error of the mean (SEM) for a $\Delta\overline{HX}$ value.

$$SEM_{\Delta\overline{HX}} = \sqrt{\frac{s_p^2}{n_a} + \frac{s_p^2}{n_b}} \tag{2.3}$$

- Using now calculated $SEM_{\Delta\overline{HX}}$ and the corresponding t-value for desired significance and degrees of freedom calculate confidence interval (CI)

$$\pm CI_{\Delta\overline{HX}} = t_{\frac{a}{2},df} \times SEM_{\Delta\overline{HX}} \tag{2.4}$$

- $|CI_{\Delta\overline{HX}}|$ calculated is the threshold used then to determine the significance of confidence 1-α[59]

This value tends to be much smaller than that which would be required using a Bonferroni corrected significance for the same α values.

2.3.3 Hybrid Test

A 'Hybrid Test' incorporates both the tests mentioned/discussed above i.e the T-test and the calculated Global Threshold. This uses strengths of both tests to yet again determine if a peptide exhibits significant differences across the measured data sets while maintaining power and minimising the risk of acceptance of a negative result.

	Welch's test		Bonferroni correction		Global $\overline{HX}$		Hybrid test	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
null 1	24	4	0	0	5	0	1	0
null 2	40	9	0	0	17	0	9	0
null 3	6	0	0	0	1	0	0	0
null 4	10	0	0	0	2	0	1	0
null 5	49	12	1	0	22	0	7	0
combined	129 (5.6%)	25 (1.1%)	1 (0.04%)	0 (0.0%)	47 (2.0%)	0 (0.0%)	18 (0.8%)	0 (0.0%)

Table 2.2: Table showing the percentage of type I errors from comparison of various statistical approaches taken from the Hageman and Weis paper of reliably identifying significant differences [59]

Using the hybrid test of α values < 0.5 we can assume from table 2.2 that less than 1% of the results potentially carry errors of type 1. Using the hybrid test increases the power of the statistical testing by maintaining significantly different values. However, the increase in accepting false positive results compared to the Bonferroni corrected value for each individual test involved is clear. This balance is one that we can accept when we use the global threshold in conjunction with the T-test as the likelihood of a type I error drops significantly when the two tests are used in conjunction. Inline with errors accepted by the the Bonferroni-correction method, whilst only requiring a much smaller differential value for uptake, therefore minimising type II error also.

2.3.4 Confidence levels

The determination of confidence levels in statistical analysis encompasses various factors. These levels are typically defined using a specific α value, which serves as a threshold for significance. The confidence in the results is then quantified as 1-α, where a higher value indicates a greater level of confidence. Consequently, the more stringent we are in refining our results, the higher the level of confidence we can achieve, thereby minimising the risk of both Type I errors. A balance, however, must be found, as by minimising type I error, can come at times hand in hand with an increase in type II error.

In statistical hypothesis testing, α values are commonly set at no greater than 0.05. This threshold is chosen to strike a balance between maintaining statistical power, the ability to detect true differences, and mitigating the risk of

false positives. By adhering to α values within this range, we aim to maximise statistical power in identifying genuine differences while simultaneously minimising the occurrence of false positives.

2.3.5 Visual Aid - Structural Relevance

Within the realm of structural biology, it is of utmost importance to ensure the reliability and validity of the conclusions drawn from the collected data. This assurance comes from not only meticulous data analysis but also from the manner in which the findings are communicated. The overarching objective is to present the data and discoveries in a manner that facilitates a comprehensive understanding of the function and dynamics of the biological molecule under investigation.

Central to this goal is meticulous analysis of the data, where rigorous methods are employed to ensure accuracy and robustness. Moreover, equal emphasis is placed on the presentation of findings, as clarity in communication is pivotal in being able to elucidate the complexities of molecular behaviour.

Ultimately, the goal is to furnish a framework through which the intricate interplay of molecular mechanisms can be comprehensively comprehended, thereby advancing our understanding of biological systems and excelling our future research endeavours.

Numerous analysis tools within the HDX community, as discussed earlier in this chapter, adopt unique approaches to data interpretation. In the development of SAUSC, our primary focus was to empower users with the flexibility to customise the presentation of findings according to their specific requirements. This flexibility allows users to customise the analytical process to yield diverse outputs, from which the question at hand can drive our focus in the desired output.

Central to the SAUSC platform is the incorporation of customisable features that enable users to manipulate data output effectively. By offering various colour schemes and gradients inspired by PDB colouring, SAUSC provides users with a powerful means to enhance the structural relevance and significance of their findings.

This emphasis on customisation and flexibility not only enriches the user experience but also facilitates a deeper understanding of the structural implications of the data. By empowering users to tailor their data presentation. SAUSC fosters a more nuanced understanding of the underlying molecular dynamics and highlights the true significance of the findings.

2.3.5.1 PDB Colouring

In our pursuit of structural relevance, our focus extends beyond merely identifying differences across experimental states. It is imperative to discern the structural implications underlying these variations, as they serve as the crux of our investigation.

Although heat maps serve as effective tools for visualising disparities, for example, if it is found that there is observed protection at residues x to y upon the addition of a binding partner, this only scratches the surface. To truly grasp the significance of these findings, we must delve deeper, bridging the gap from primary amino acid sequences to higher-order structural changes.

The key question arises: How do these observed variations manifest in the context of the overall protein structure, and what insights can be gleaned regarding structural alterations upon any state comparison? The overlay of our findings on PDB structures provides invaluable visual aids, offering a comprehensive framework to interpret the implications of these variations.

By integrating experimental data with structural models, we can unravel the intricate interplay between sequence-level (primary) changes and their implications on the three-dimensional architecture of the protein. This holistic approach not only enhances our understanding of structural dynamics but also provides crucial insights into the functional implications of molecular interactions.

2.3.5.2 Uptake Difference

At the heart of the analysis lies the assessment of differences in peptide-level uptake. This involves examining changes in deuterium uptake values derived from the collected data.

The uptake value (DU) can be determined for any peptide that is identified in both experimental states, typically spanning a specific range of residues (e.g., from residue x to y). This calculation is performed at various individual time points (t) and is compared to the uptake value in the unlabelled initial state (0).

$$\mathrm{DU}_t^{x\to y} = \mathrm{m}_t^{x\to y} - \mathrm{m}_0^{x\to y} \tag{2.5}$$

When making comparisons between different experimental states, these uptake values serve as crucial metrics for calculating ΔDU. This represents the disparity in deuterium uptake of a specific peptide at a given time point between two comparative states, typically delineated as a control state and comparison state. Calculating Δ DU is at the forefront of any valuable insight gained about the dynamic changes that occur within peptides under varying experimental conditions, helping to clarify structural and functional differences. This is due to the fact that we are in essence measuring deuterium uptake spatially across a structure, and therefore the difference (Δ DU) between the experimental states shown in Equation 2.6 is the difference in deuterium uptake between those two different states that are calculated in Equation 2.5.

$$\Delta\mathrm{DU}_t^{x\to y} = (Control)\mathrm{DU}_t^{x\to y} - (Comparison)\mathrm{DU}_t^{x\to y} \tag{2.6}$$

2.3.5.3 Relative Difference

The relative uptake difference (RD) is derived from the absolute uptake difference, which we previously calculated. This absolute difference is divided by the total number of exchangeable amide hydrogens present within the peptide under consideration. Using this approach, we ensure that the resultant values are inherently relative. This adjustment is crucial because, without it, longer peptides would exhibit seemingly greater differences, solely because of their length, even if the actual level of protection or deprotection remain consistent. Therefore, this normalisation process allows for a more accurate comparison between peptides of varying lengths, eliminating biases introduced by the peptide size.

$$\Delta RD = \pm \frac{\Delta \mathrm{DU}_t^{x \to y}}{\#ExchangableAmides} \times 100\% \tag{2.7}$$

By standardising the analysis through this normalisation procedure, we ensure that peptides of different lengths are placed on an equitable footing for comparative assessment. This method facilitates the identification of regions within the protein in which the most pronounced structural and dynamic alterations manifest across the states compared. By enabling a comprehensive examination of these changes without the confounding influence of peptide length, we gain deeper insights into the underlying mechanisms driving variations in protection and deprotection levels. Thus, this approach enhances our ability to discern and interpret subtle, yet significant, distinctions in peptide behaviour under experimental conditions.

2.3.5.4 Gradient Based Colour Scheme

Using the aforementioned computations, we can integrate a colour-based gradient system across the generated figures, along with the respective mapping onto the PDB structures. This strategic approach ensures a consistent colour scheme throughout our analysis, thereby enhancing the depth of information conveyed in our data presentation.

Colourisation is determined by the ΔDU or RD figures computed above and can be normalised in two distinct ways, offering users greater flexibility for information analysis. These normalisation methods include global normalisation and time-point-dependent normalisation.

Global normalisation provides a comprehensive overview of the data, allowing users to observe patterns and trends across the entire dataset. However, time-point-dependent normalisation offers a more granular perspective, enabling users to analyse fluctuations and variations at specific time points throughout the dataset.

By offering these normalisation options, users can tailor their analysis approach according to the specific insights they seek, thereby facilitating a more comprehensive understanding of the data and its underlying dynamics. This systematic approach

improves the interpretability and utility of our analysis results, contributing to more informed decision-making processes.

2.3.5.5 global vs time point dependent

In contrast to the global normalisation approach, which considers the maximum absolute value of |Δ RD| throughout the experimental data, regardless of time points. The time-point-dependent normalisation method offers users a more nuanced analysis. Here, each time point, designated as t_1, t_2, etc., receives its own individual ± min / max values for colour normalisation.

This tailored approach enables users to dive deeper into the dynamics of each time point, unveiling subtle variations and trends that may not be as apparent when viewing the data globally. By assigning unique ±min/max values to each time point, users gain a better understanding of the differences and developments that occur within and between peptides in the temporal spectrum.

For instance, this method facilitates the identification of variations specific to particular time points that might not be as clearly discernible when employing a global normalisation strategy. Early time points, for instance, might exhibit differences of varying magnitudes that could be overlooked in a broader global context. Conversely, later time points might reveal trends or discrepancies that are not immediately evident when considering the data as a whole.

By allowing for both global and time-point-dependent normalisation, our approach empowers users to explore and interpret their data from multiple perspectives, uncovering insights that might otherwise remain obscured. This flexibility ensures that the colour scheme remains meaningful and informative throughout the entire data analysis process, fostering a more comprehensive understanding of the underlying dynamics and facilitating informed decision-making.

2.4 Results

Our aim was to provide a comprehensive output using the points discussed above. We focused on expanding the already available figures from other softwares to

maintain the greatest level of depth possible. Allowing for a greater ability to alter these figures at will whilst maintaining a consistent colour scheme throughout to allow users to interpret and identify trends across complex data sets.

2.4.1 Figure Plotting

In our data analysis approach, we opted to use the figures generated using Deuteros Lau et al.[51]. Both Woods Plots and Volcano Plots offer distinct insights to users, enriching their understanding of the statistical analysis performed and elucidating the magnitude of change observed, as well as the specific regions along the primary sequence where these alterations occur. Woods plots offer a holistic view of the extent and location of changes across individual residue numbers and peptide lengths, providing valuable insights into the magnitude and spatial distribution of alterations. In contrast, Volcano Plots focus on presenting statistical analysis outcomes, focussing on highlighting significant differences to facilitate the identification of key findings.

Both types of plot consistently employ the previously mentioned colour scheme throughout the analysis. This scheme utilises varying shades of blue, white, and red to emphasise the magnitude of differences for specific significant peptides in both 2D and 3D representations. This uniform application of colour enhances the interpretability and continuity of the visualisations, enabling researchers to discern patterns and trends with clarity across different analytical contexts.

By incorporating 3D representations alongside these graphs, we enhance the depth of analysis, enabling users to draw more robust conclusions. These figures not only facilitate the interpretation of complex data, but also serve as valuable tools for communicating research findings effectively. Thus, through the combined use of our gradient colour scheme consistently employed within the Woods Plots, Volcano Plots, and 3D representations, our data analysis framework provides a comprehensive approach to understanding and interpreting the results obtained.

2.4.1.1 Volcano Plots

Volcano plots provide a comprehensive visual representation of peptides shared between different time points and compared states. In these plots, peptides are depicted as circular dots, each coloured according to the gradient determined previously. The intensity of the colour, ranging from deep blue to red, reflects the magnitude of the relative difference (RD) for each peptide. Peptides that do not exceed the predetermined significance boundaries are coloured grey, indicating their insignificance.

Significantly different peptides are highlighted, exceeding the calculated ΔDU threshold and passing the t-test at the desired level of confidence, according to the chosen normalisation method, either global or time-point-dependent gradient. This selective colorisation ensures that peptides that exhibit substantial differences are easily identifiable with respect to the complexity of the data set.

In a volcano plot, the X-axis represents the absolute difference in the uptake of an individual peptide (Δ DU) between the comparative states at individual time points. Significance is determined by whether this difference exceeds the predetermined $\pm$global threshold ($\pm$ $CI_{\Delta\overline{HX}}$), as described by Equation 2.4. Currently, significance also depends on passing the t test, typically indicated by the Y axis, represented as -$\log_{10}$(p-value). The p-value here is determined by your desired confidence level, where p represents the 1-α value. This setup ensures that both the magnitude of difference and the statistical significance of the observed changes are effectively visualised and interpreted within the volcano plot. Due to the inverse relation of the p-value and the -$\log_{10}$ (p-value), we see that the most significant peptides appear in the upper portion of volcano plots, evident in figures 2.1 and 2.2.

Both constraints are visually represented by dashed lines within the volcano plot, delineating clear boundaries that demarcate areas of significance. This graphical representation enhances the interpretability of the plot by visually highlighting regions where the absolute difference in peptide uptake exceeds the predetermined global threshold and where the corresponding statistical significance, determined

by the t-test, is met. Thus, the dashed lines serve as visual aids, facilitating the identification and interpretation of significant findings within the plot.

Using this approach, volcano plots offer a clear and concise overview of peptide dynamics, facilitating the identification of meaningful trends and patterns across different experimental conditions and time points. Visual differentiation between significant and non-significant peptides enhances the interpretability of the plots, enabling researchers to focus on the most relevant findings for further analysis and exploration.

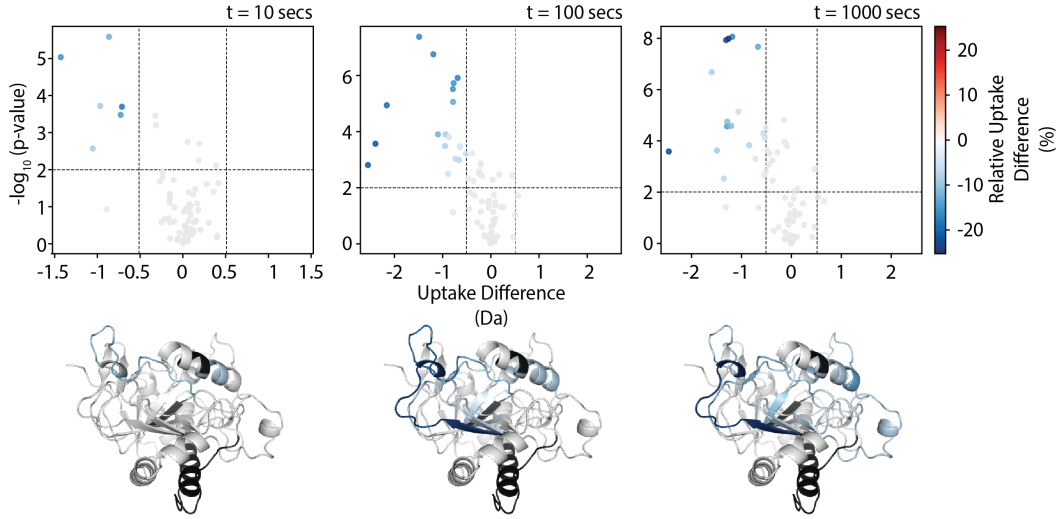


Figure 2.1: Volcano plot of three different time points produced by SAUSC in the analysis of IPNS and IPNS-Cd comparative states. The time points correspond to labeling times of 10, 100, and 1000 seconds. Colouring is consistent across all time points due to the global colour scheme implemented. The deeper the colour red/blue, the greater the respective deprotection/protection, as indicated by the gradient colour scale on the right-hand side of the figure. The 1-D information gained from the significant peptides has then been mapped onto the 3-D structures at each of these time points to allow for a greater visual representation. This mapping enhances the understanding of the spatial distribution of protection and deprotection across the protein's structure over time.

The atypical volcano plot we created is a basic scatter plot with 3 dimensions of data - Uptake Difference (x-axis), -$\log_{10}$(P-value) (y-axis) and the relative uptake difference as the colour gradient. We have enabled the user to be able to change these and create a scatter plot that would allow for any combination of the measured components (Uptake difference, Relative uptake difference, P-

value, and -$\log_{10}$(P-value). In figure 2.2 there are differing combinations of what the user may decide to do.

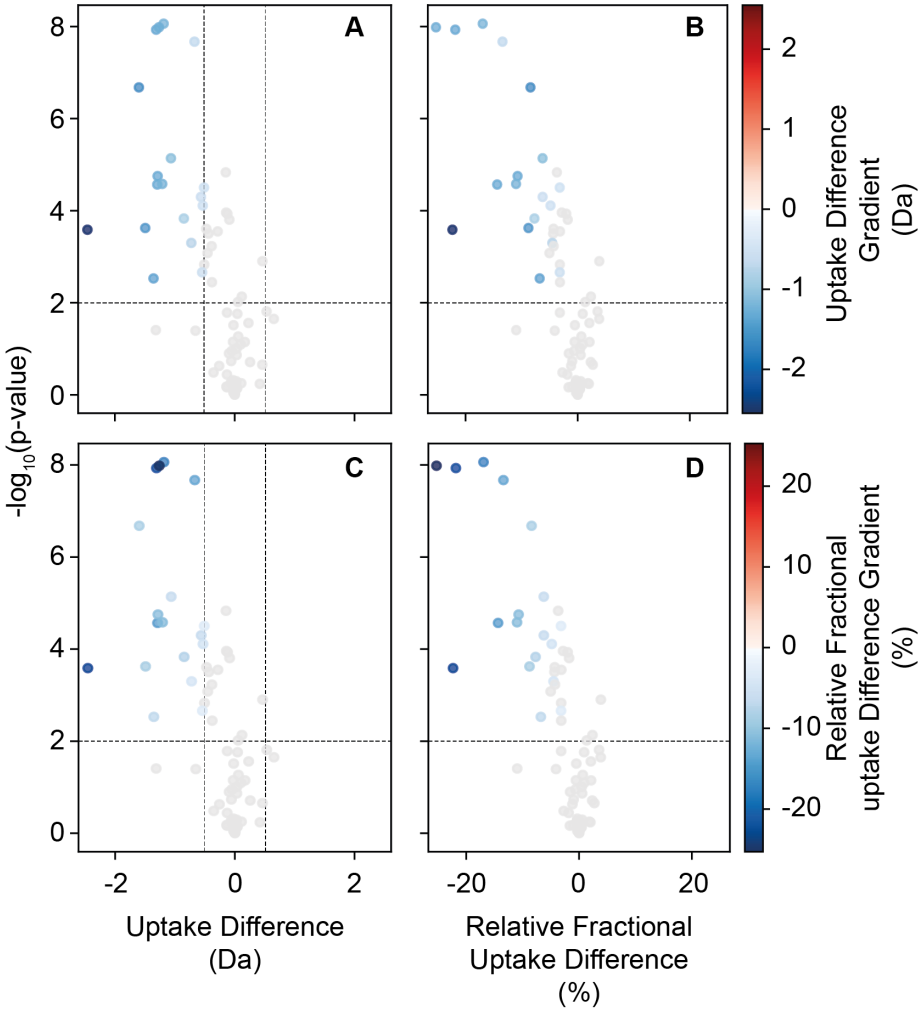


Figure 2.2: Here we have plotted four different Scatter plots. All from the same 1000 sec labelled data of IPNS compared to the Cd bound state.**A** and **B**shows the Uptake difference and relative uptake difference respectively vs -$\log_{10}$(P-Value) with the coloured gradient being the Uptake difference. **C** and **D** is showing the uptake difference and relative uptake difference respectively vs -$\log_{10}$(P-Value) with the coloured gradient being the relative uptake difference. **A** and **D** we are losing a data component as the colour gradient matches the x-axis

As we can see from figure 2.2 plot **B** and **C** maximise the information gained from the collected data and are able to portray in similar, yet different ways three different components of the findings. **A** and **D** although looking similar to the other plots, do actually lose a component through the colour scheme matching that of the x-component in the plots. Therefore, in plots **A** and **D** the left most

point is represented by the deepest blue, as expected when the x-axis is also the coloured gradient. Contrastingly, **B** and **C** maintain the three components, and we can see that the colour is independent of the x position, giving us another element of depth to the figure.

2.4.1.2 Woods Plots

In a Woods plot, the protein's primary sequence is mapped along the x-axis, presenting residue numbers distinctly. This layout provides a visual depiction of the individual peptide initiation and termination points, each represented by a horizontal bar that spanning from the peptide's N-terminus to its C-terminus.

On the y-axis, the graph illustrates the absolute difference in uptake of an individual peptide (Δ DU) in various comparative states at different time points. Similar to the Volcano Plot, significance determination in the Woods Plot follows predefined criteria. Dashed lines are strategically placed along the y-axis to delineate the boundaries of the $\pm$global threshold ($\pm$ $CI_{\Delta\overline{HX}}$), aiding users in understanding the significance thresholds.

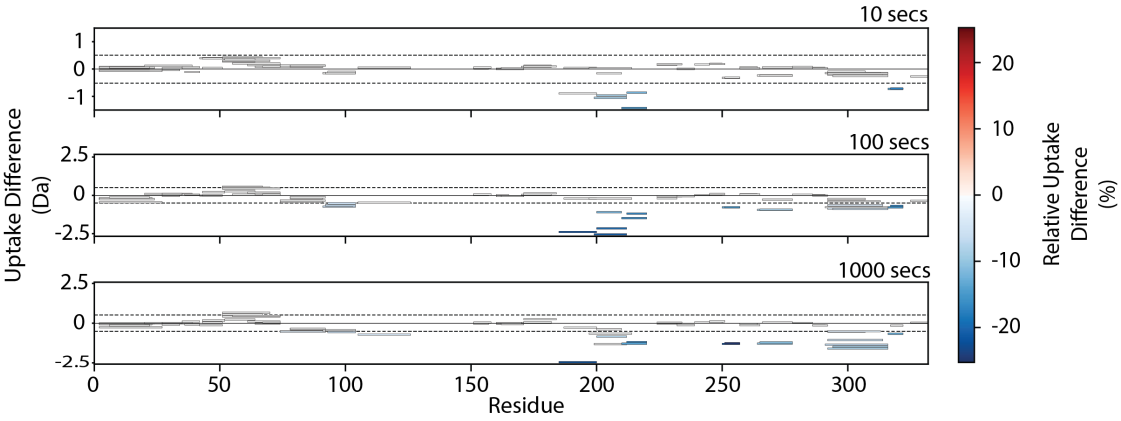


Figure 2.3: Woods plot of three different time points, 10,100 and 100 seconds produced by SAUSC in analysis of IPNS and IPNS-Cd comparative states. Colouring consistent across all time points due to a Global colouring scheme implemented. Deeper the colour Red/Blue greater the respective deprotection/protection as is made clear with the gradient bar to the right hand side of the figure.

Here in figure 2.3, we see the three time points of the data collected in IPNS in the presence and absence of cadmium (Cd^{2+}). The time points collected correspond to labelling times of 10, 100, and 1000 seconds. The bars in the plot represent the

individual peptides identified by sequencing - which is a non-labelled run to gain a t_0 that acts as a reference throughout the experiment - with colours indicating the results of significance testing. Specifically, the hybrid test was employed to determine significance, with grey bars representing non-significant peptides, and coloured bars indicating peptides that exhibited significantly different uptakes between the two compared states.

The colours of significant peptides are based on their relative uptake differences, with the gradient used being consistent between the different figures produced, including the one shown in Figure 2.1. This colour gradient allows for a clear visual representation of the degree of uptake difference, facilitating an easier interpretation of the data.

2.4.1.3 Cumulative Plots

Cumulative plots, in contrast to regular plots, showcase the cumulative difference of a specific peptide across states throughout the duration of the experiment. This distinction requires a variation in the statistical tests used. Although the hybrid approach remains intact, involving both Welch's t test and the global threshold, the criteria for passing these tests differ from those used for individual time points.

The cumulative difference, measured across the entire experiment rather than at individual time points, prompts adjustments in the threshold calculations. Specifically, when calculating the confidence interval ($\pm CI_{\Delta\overline{HX}}$) for cumulative differences, the global threshold is multiplied by the number of time points of labelling t_n (denoted as n when $n \neq 0$). This modification ensures that for a peptide to pass the cumulative test, its average difference must consistently exceed or fall below the calculated ± CI value at all labelling times.

$$\text{Cumulative} \pm CI_{\Delta\overline{HX}} = t_{\frac{\alpha}{2},df} \times SEM_{\Delta\overline{HX}} \times \text{number of labelling time points} \quad (2.8)$$

In the Welch t test, while the assessment of the repeatability and likelihood of a peptide remains consistent with the typical t test, there is a variance in the standard

deviation calculation. Here, the standard deviations of both states are pooled across all labelling times to derive a composite standard deviation for the peptide.

By integrating these adjusted statistical tests, the cumulative plots offer a comprehensive assessment of peptide behaviour over time. This nuanced approach accounts for the cumulative nature of the data, ensuring robust statistical analysis and interpretation. Through meticulous consideration of cumulative differences and standardised testing procedures, researchers can gain deeper insights into the dynamics of peptide behaviour between experimental conditions and time points.

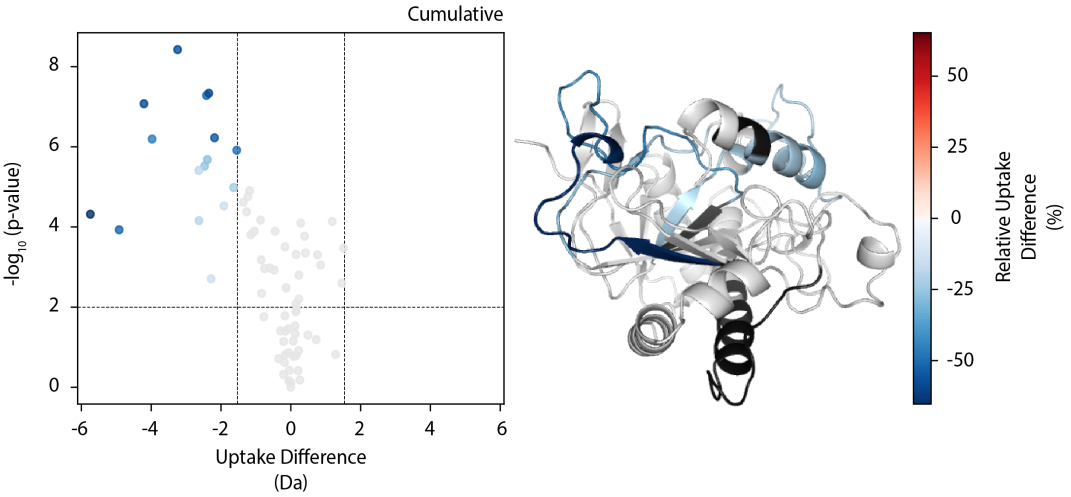


Figure 2.4: Cumulative Volcano plot with PDB colour mapping Here is an example of a cumulative volcano plot from data gathered on IPNS when compared with the addition of Cd(II) metal ion. Showing the differences clearly when compared to that of the individual time points found in 2.1.To aid visualisation these cumulative differences have then been mapped on to the PDB structure of IPNS (1BK0).

Comparing the example Cumulative Volcano Plot in Figure 2.4 to the SAUSC Volcano Plots in Figure 2.1, it is evident that the criteria for passing the t-test on the y-axis remain consistent across both types of plots. However, there's a notable difference in the bounds of the global threshold, as indicated by the dashed vertical line. This change is attributed to the cumulative nature of the differences being analysed in the cumulative plots, contrasting with the singular differences examined in the regular Volcano Plots.

Similarly, the same comparative analysis can be extended to Woods plots. Although the statistical testing procedures remain unchanged, the significance

bounds are subject to adjustment. This adjustment reflects the cumulative nature of the data being analysed in Woods plots, necessitating recalibration of the significance thresholds to account for the accumulation of differences over time.

One of the primary advantages of cumulative plots is their resilience to the impact of individual time points with suboptimal quality or unexpected variability. In contrast to analyses at the individual time-point level, where poor quality data or unusually large standard deviations at a single time point could lead to the overlooking of significant differences, cumulative plots offer a more robust assessment.

Consider a scenario in which data quality issues or increased variability are observed at a particular time point in an experiment. At the individual time point level, such anomalies might result in the dismissal of a peptide as not significantly different, potentially leading to its oversight. However, when comparing across cumulative plots, discrepancies and anomalies of this nature are more readily discernible.

Cumulative plots provide researchers with a broader perspective by aggregating data across multiple time points, enabling the identification of patterns and trends that might remain hidden when analysing individual time points separately. This expanded view improves sensitivity to data irregularities, ensuring a more comprehensive assessment of peptide dynamics. Consequently, researchers are better equipped to detect significant differences that might otherwise be overlooked due to isolated instances of poor data quality or variability. As a result, the cumulative approach adopted in SAUSC instils greater confidence in the reliability and robustness of the analyses performed.

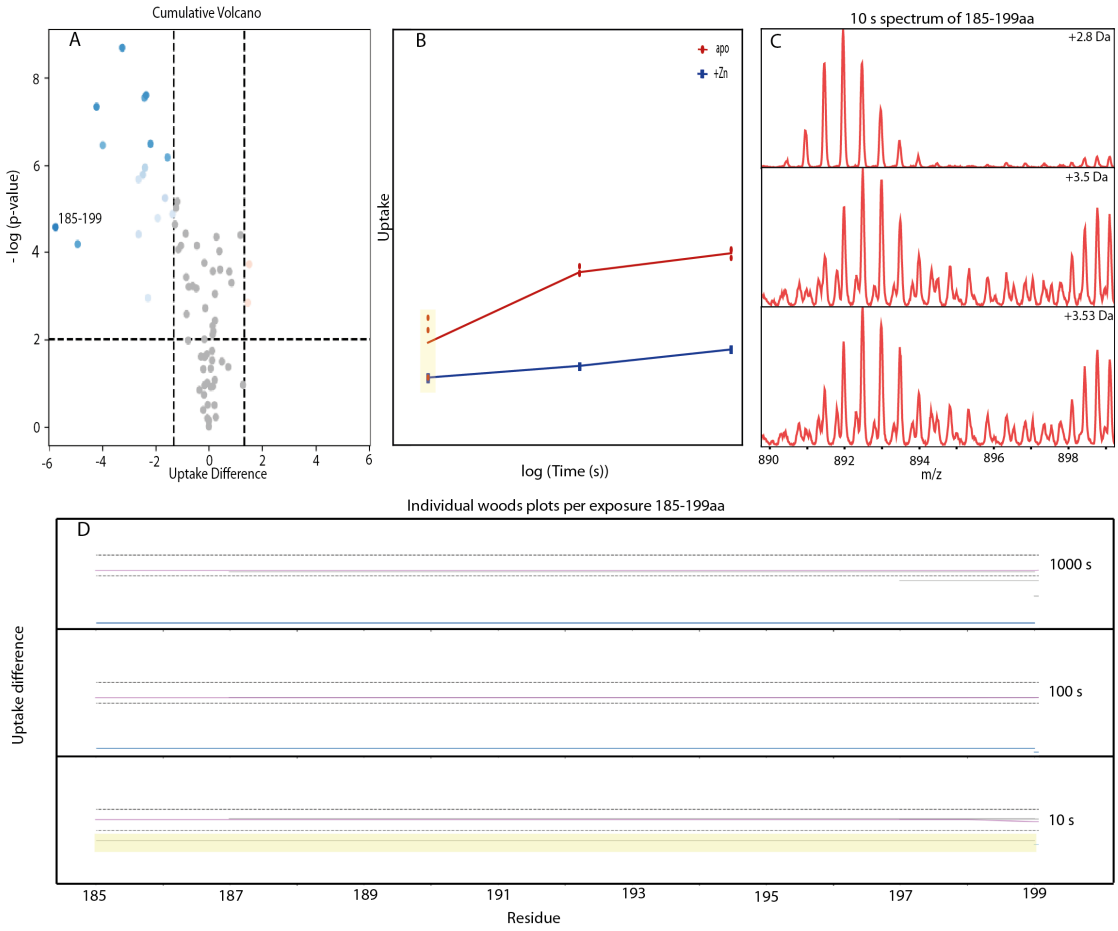


Figure 2.5: Comparison of individual and cumulative plots. Here we can see a comparison of cumulative and individual plots and start to gauge how the quality of data or lack thereof can affect the statistical testing performance at different levels.**A**is the cumultaive plot of the IPNS vs IPNS+Zn data.**B**is the raw Deuterium uptake curve of both states, IPNS in red and +Zn in blue.**C**Raw MS data of Peptide residues 185-199 at the 10s time point.**D**Woods plots zoomed in to show Peptide residues 185-199 at 10, 100 and 100s labelling time - 10s peptide highlighted in yellow as was deemed not significantly different.

Figure 2.5 clearly illustrates how the quality of the data significantly influences the outcome of statistical tests, particularly at the individual time point level compared to cumulative data analysis. The impact of the standard deviation on the t test for individual time points can affect the ability to detect differences accurately. Peptides that appear insignificant when analysed individually may, in fact, exhibit significant differences when examined cumulatively.

The complementary nature of cumulative and singular time-point analyses

enables researchers to conduct comprehensive investigations, facilitating the identification of genuine differences with greater ease. Using both approaches, researchers can mitigate the potential pitfalls associated with poor data quality and variability, thereby enhancing the robustness and reliability of their findings. This integrated approach ensures a greater understanding of peptide dynamics and fosters confidence in the validity of the statistical analyses conducted.

In essence, the transition from singular to cumulative analysis necessitates recalibration of significance bounds to accurately reflect the cumulative nature of the data. By adjusting the global threshold and other significance criteria accordingly, researchers ensure that the statistical analysis remains robust and also aligned with the nature of the data being analysed. Through consideration of these adjustments, SAUSC enables a comprehensive and accurate interpretation of peptide dynamics across varying experimental conditions and time points.

2.5 Discussion

The SAUSC framework is designed to decode the complex dynamics of peptide behaviour at various experimental conditions and time points. By integrating advanced statistical methods with cutting-edge visualisation techniques, researchers can uncover subtle but significant patterns in peptide datasets.

At its core, SAUSC is focused on comparative analysis of the uptake values of experimental states, allowing researchers to track peptide behaviour in diverse experimental settings. This involves aggregating uptake data from individual peptides to reveal broader trends and dynamics beyond isolated conditions and time points.

A cornerstone of SAUSC lies in the use of both Woods plots and Volcano plots, each of which offers unique insights into the dynamics of the peptides. Woods plots the primary sequence of the protein of interest in an intricate manner, mapping the initiation and termination points of each peptide. Through the depiction of horizontal bars spanning from the N-terminus to the C-terminus, Woods plots provide us with a vivid representation of spatial patterns along the protein sequence, facilitating insightful analyses.

In contrast, Volcano Plots focus on the statistical outcomes derived from the data, utilising -$\log_{10}$(p) values to denote significance and absolute differences in peptide uptake to quantify magnitude. These plots serve as a focused lens through which researchers can discern significant differences among the complexity of the data. The determination of significance in both the Woods and Volcano plots adheres to predefined criteria, ensuring the robustness and rigour of the analytical process.

Integral to SAUSC is the consistent application of colour-based gradients across both Woods and Volcano Plots. This uniform colour scheme, calibrated to reflect the magnitude of differences for significant peptides, enhances the interpretability and coherence of the visualisations. By conveying the intensity of changes in a two- and three-dimensional representation, this colour gradient helps researchers uncover nuanced patterns and trends within the data.

Furthermore, SAUSC embraces a sequential approach to data analysis, allowing for iterative exploration and refinement of our understanding of peptide dynamics. By scrutinising peptide behaviour across multiple time points and experimental conditions in a systematic manner, researchers can uncover subtle variations and trends that may elude detection in isolated analyses. This iterative process fosters a comprehensive understanding of the underlying dynamics, empowering researchers to draw robust conclusions and formulate informed hypotheses.

In summary, SAUSC represents a sophisticated analytical framework for unraveling the intricate dynamics of peptide behaviour. By seamlessly integrating advanced statistical methodologies with state-of-the-art visualisation techniques. SAUSC offers researchers a comprehensive toolkit for uncovering meaningful insights into protein function and dynamics. Through its systematic approach and iterative analysis process, SAUSC empowers researchers to explore, interpret, and ultimately comprehend the complexities of peptide behaviour within biological systems.

Alles Gescheite ist schon gedacht worden.
Man muss nur versuchen, es noch einmal zu denken.

All intelligent thoughts have already been thought;
what is necessary is only to try to think them again.

— Johann Wolfgang von Goethe

3

Investigating the affects of metal binding to Iso-Penicillin-N-synthase

Contents

3.1 Abstract 57
3.2 Isopenicillin N Synthase 58
3.2.1 Use of Zn/Cd(II) to Mimic Iron Binding 59
3.2.2 ACV Tripeptide Binding Pocket 60
3.2.3 Experimental plan 61
3.2.4 Cross State Analysis 62
3.2.5 Differences - Structural Insight Gained 63
3.3 Methods 65
3.3.1 HDX-MS 65
3.4 Results 66
3.4.1 Structural Relevance 70
3.5 Discussion 73
3.5.1 Octahedral Coordination 73

3.1 Abstract

This chapter explores the critical role of the iron(II) centre in the catalytic mechanism of isopenicillin N (IPN) biosynthesis. The iron centre adopts an octahedral coordination geometry, providing multiple binding sites that sequentially interact with important molecular entities, including the amide side chain of Gln330, water

molecules, the substrate's thiol group, d-valine, molecular oxygen, and reactive oxygen species. These coordinated interactions are essential for driving the catalytic process. Furthermore, our findings highlight the importance of key residues: His214, Asp216, and His270, whose reduced uptake in the presence of metal-binding partners indicates their direct involvement in maintaining the enzyme's catalytic activity. Disruption of these residues would likely alter their function. This detailed understanding of the dynamic coordination at the iron(II) centre offers valuable insights into the mechanistic steps of Isopenicillin-N-synthase (IPNS) and its role in penicillin biosynthesis.

3.2 Isopenicillin N Synthase

IPNS is a non-heme iron-dependent oxidase that belongs to the 2-oxoglutarate (2OG)-dependent dioxygenase family. It catalyses the formation of IPN from δ-(L-α-aminoadipoyl)-L-cysteinyl-D-valine (ACV), a key step in penicillin biosynthesis. IPNS has been extensively studied since the discovery of penicillin antibiotics in the 20^{th} century [62–64].

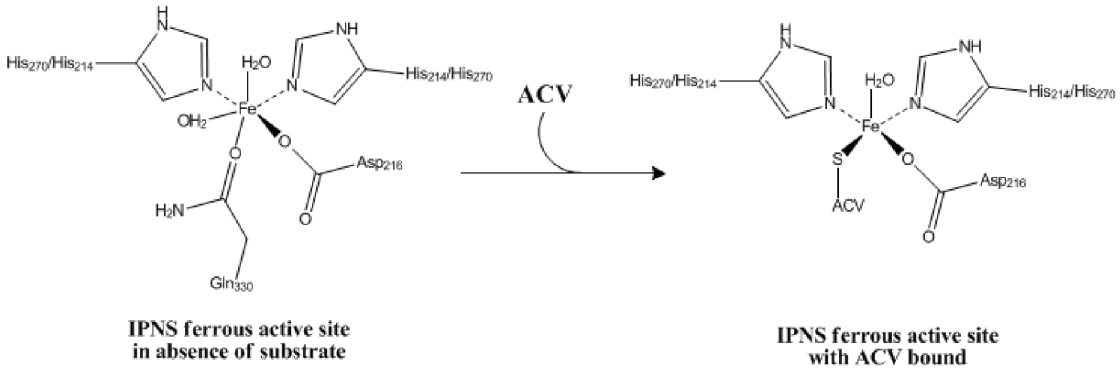


Figure 3.1: Molecular diagram showing the binding of tripeptide ACV with in the active site of IPNS and the displacement from the iron core. Taken from Roach et al [65].

In the active site of IPNS, a Fe(II) ion is coordinated by two histidine residues, one aspartate residue, one glutamine residue, and two water molecules in the absence of a bound substrate [65]. Upon binding to the substrate (ACV), the glutamine residue and one water molecule are displaced, leaving the two histidines (His214 and

His270) and aspartate (Asp216) conserved. These residues are key to the enzymatic activity of IPNS, particularly in the context of the HDX data collected [66].

IPNS catalyses the conversion of ACV and dioxygen to IPN, releasing two water molecules as by-products [65]. The catalytic mechanism begins with ACV's thiolate binding to the Fe(II) ion at the active site, triggering a reaction cascade that leads to the formation of a monocyclic intermediate, a crucial step in the enzyme's function [65, 67]. These studies have significantly deepened our understanding of the molecular mechanisms involved in IPNS's role in antibiotic biosynthesis.

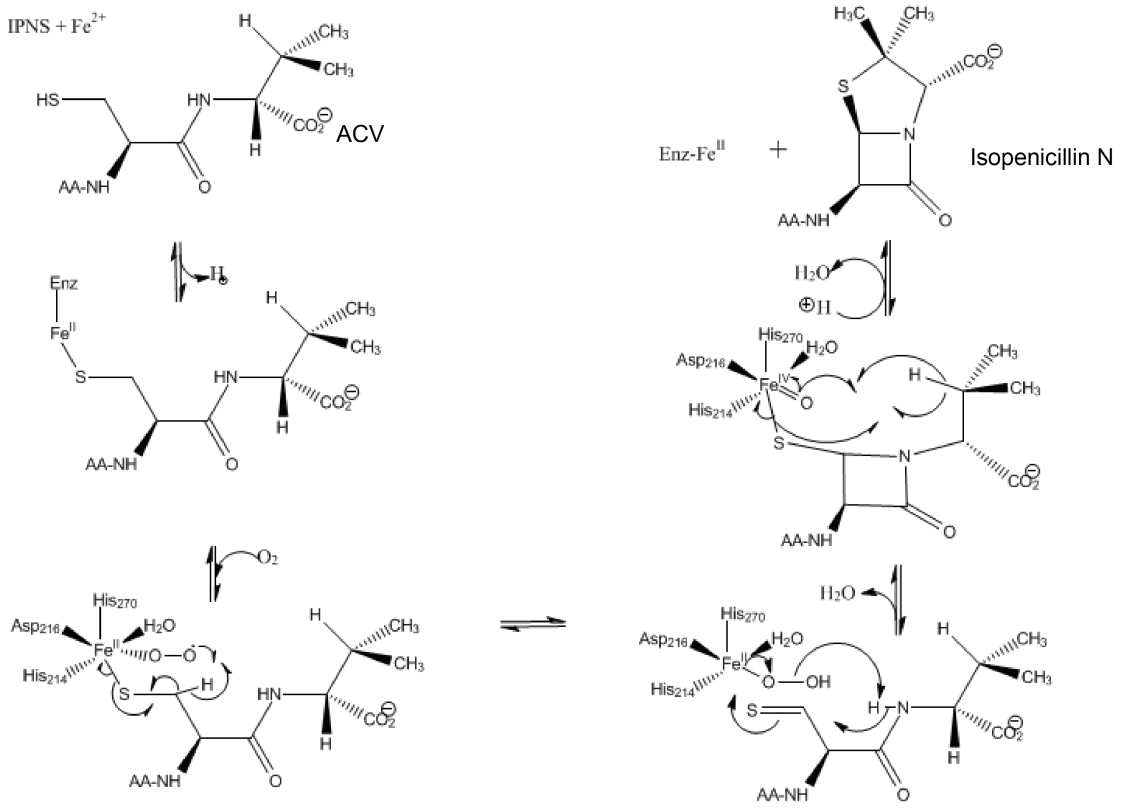


Figure 3.2: molecular diagram showing the reactive intermediates and the step wise formation of IPN from ACV along with two intermediate water molecules being formed. Taken from Roach et al [65].

3.2.1 Use of Zn/Cd(II) to Mimic Iron Binding

In our experiments, we employed Zinc (Zn(II)) and Cadmium (Cd(II)) ions as substitutes for the Iron (Fe(II)) ion to mimic its binding effect within the enzyme's active site. The objective was to induce metal binding within the active site while

avoiding the active turnover of products. Using zinc and cadmium ions, we aimed to investigate the impact of conformational changes resulting solely from metal binding, without the potential interference of product turnover. This approach allowed us to focus exclusively on elucidating the effects of metal coordination on the enzyme's structure and how that affected the proteins' conformational dynamics. The avoidance of product turnover helped preserve the integrity of the experimental conditions, ensuring that any observed changes could be attributed directly to metal binding-induced conformational alterations. Zn(II) and Cd(II) were selected for use due to their ability to 'mimic' binding[68] as well as their increased stability in association with protein complexes[69].

3.2.2 ACV Tripeptide Binding Pocket

According to the findings reported by Roach et al. [65], it is hypothesised as previously said that during the formation of penicillin, the ACV tripeptide undergoes ligation to the iron core. This process results in the creation of a vacant coordination site on the iron atom, which subsequently facilitates the binding of dioxygen, shown in Figure 3.1. This interaction is proposed to enable the removal of the necessary hydrogens from ACV without direct involvement or assistance from specific protein residues[65].

This mechanism suggests a pivotal role for the iron core in orchestrating the chemical transformation associated with penicillin biosynthesis. By creating a reactive environment at the iron centre, the biosynthetic pathway can proceed efficiently, highlighting the intricate interplay between small-molecule substrates and enzymatic cofactors in natural product biosynthesis.

Changes in the conformational dynamics of ligand-activated receptors play a crucial role in their functional regulation upon ligand binding. The use of HDX measurements combined with mass spectrometry has emerged as a powerful and efficient method for characterising alterations in the conformational dynamics of proteins after ligand binding. This technique not only offers high sensitivity in detecting ligand interactions and changes in receptor dynamics but also provides

valuable mechanistic insights into the regulation of ligands. Through HDX coupled with mass spectrometry, it is possible to gain a deeper understanding of the molecular mechanisms underlying ligand regulation, thus advancing our knowledge of receptor-ligand interactions and their functional consequences [70].

In this chapter, we will explore how the binding of the metal ion to the active core of IPNS alters the structural dynamics of the protein and what that means for the protein function. The works of Chapman and Rutledge[71] give us insight into how the mechanism of binding occurs, as we can see exact residues involved in binding through their crystallographic approach. They were able to clearly demonstrate that binding of the ligand occurs in very clear and concise areas of the protein, namely His214, Asp216, and His270. This provides clear focal points for our HDX application, allowing us to observe how the protein dynamics changes during the mechanistic binding of the ligand. HDX-MS was determined to be a technique of interest when looking at this complex because of its ability to detect real-time dynamics upon perturbation of the control.

This synergistic approach enables a comprehensive exploration of protein dynamics at both the structural and functional levels, facilitating a more holistic understanding of the biological processes at hand. The integration of HDX-MS with other techniques not only enhances the accuracy and resolution of the obtained data but also offers a multifaceted perspective on the dynamics of protein-ligand interactions. Consequently, this integrative approach holds significant promise in advancing our understanding of complex biological systems and unraveling the intricacies of the molecular mechanisms governing protein behaviour.

3.2.3 Experimental plan

In a typical experimental setup for binding ligands, there exists a reference state, often referred to as the *apo* state, alongside one or two comparison states representing the ligand-bound states. For instance, in the case of proteins that require a catalytic metal ion such as IPNS or in enzymatic systems that involve substrates like the tripeptide ACV, which will be explored in this chapter, these comparison states

are crucial. By examining and analysing the differences across these distinct states, we can uncover how protein dynamics and conformational changes are influenced by the addition of the substrate. This approach provides valuable insight into the mechanistic intricacies of the complex under investigation. Through systematic measurement and analysis, we are able to elucidate the dynamic interplay between proteins and their ligands, offering a deeper understanding of the underlying molecular mechanisms governing their interactions.

Recent technological advancements have revolutionised the efficiency and accuracy of experimental procedures, particularly in the realm of protein dynamics analysis. A key driver of this transformation is the widespread adoption of liquid handling robots, which have greatly enhanced the capabilities and accessibility of such experiments. These robots facilitate precise labelling times, improved incubation conditions, reduced back-exchange, and improved overall throughput. As a result, we are now able to optimise instrumental time more effectively, ensuring that we can extract the maximum value from our experiments, whilst maintaining the best standard of repeatability allowing for more concise and trusted results. This enhanced efficiency and precision pave the way for deeper insights into protein behaviour and dynamics, accelerating progress in the field of molecular biology and biochemistry.

3.2.4 Cross State Analysis

Analysing protein behaviour across different states offers valuable insight into their dynamic properties. Identifying and comprehending the differences between these compared states, we can then discern what constitutes a significant deviation, as elaborated in Chapter 2. This understanding serves as a foundation for drawing conclusions regarding the mechanistic implications of perturbations induced within our comparative states. Through meticulous examination and interpretation of these differences, we can begin unraveling the underlying molecular mechanisms that govern protein responses to various stimuli. This analytical approach enables

the formulation of informed hypotheses and facilitates a deeper understanding of the intricate dynamics underlying biological processes.

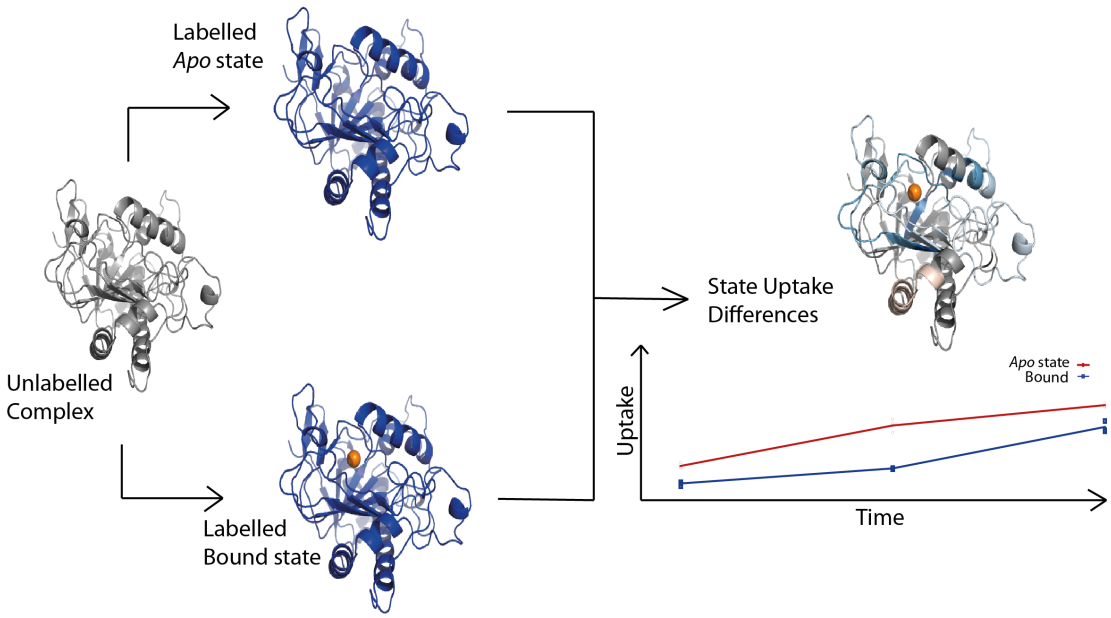


Figure 3.3: This figure shows in simplistic terms the differences that occur across states. looking at the uptake of state 1 (labelled, *apo* above) vs the uptake of state 2(labelled, labelled above) we are able to easily identify differences that occur in the labelling phase of our experiment.

We speak extensively in chapter 2 about differences that occur between state comparisons and how we can handle these differences to gain the greatest knowledge to allow insight into protein behaviour. It is important to note that in figure 3.3 the colours of the structures in the two comparative states are not indicative of any uptake data and serve purely as a depiction of what one might do in a comparative experiment.

3.2.5 Differences - Structural Insight Gained

There are several factors that can influence and contribute to changes in deuterium uptake:

- Protein Structure: Variations in HDX rates arise because of differences in the secondary structure of a protein complex. For example, α-helices and β-sheets

exhibit distinct exchange rates owing to variations in bonding patterns and solvent accessibility within these regions.

- Solvent Accessibility: Regions located within the core of a protein structure may display slower exchange rates as they are buried from the protein surface, thus less accessible to the labelling buffer. In contrast, solvent-exposed areas on the protein surface demonstrate faster exchange rates due to their accessibility to solvent molecules[72].

- Dynamics: Changes in protein dynamics, such as conformational fluctuations, can affect the accessibility of hydrogen atoms to solvent molecules. Regions undergoing dynamic conformational alterations may exhibit fluctuations in Deuterium uptake rates[73]. An example of this would be the hinge mechanism present between the S1 and S2 subunits of the sars-cov2-spike protein in the presence of the ACE2 receptor.

- Ligand Binding: The binding of ligands or small molecules can perturb the hydrogen-bonding network within the protein structure, leading to conformational changes. This interaction may result in alterations in the deuterium uptake rates in specific regions of the protein.

- Post-translational Modification: Modifications such as phosphorylation or glycosylation can induce conformational changes and alter protein dynamics. These modifications impact the accessibility of hydrogen atoms to solvent molecules, thereby influencing Deuterium uptake rates in the modified regions of the protein[74].

Each of these factors can play and contribute to a significant role in modulating the dynamics and conformational changes of proteins, ultimately influencing their biological functions.

Upon understanding the factors influencing changes in deuterium uptake, we gain insights into the molecular dynamics of proteins and the underlying mechanisms driving these alterations. Using this understanding, we can infer the occurrence

of molecular events within the protein structure and discern the contributing factors to the observed changes. Through the analysis of HDX data, we were able to delineate potential binding sites at which interactions with ligands or other molecules may occur. Additionally, this method enables us to unravel the mechanistic pathways of change, elucidating the sequential events that lead to conformational alterations. Furthermore, HDX data offer valuable information for visualising dynamic protein behaviour, providing a glimpse into the transient states adopted by the protein over a certain time period. Ultimately, the comprehensive analysis of HDX data empowers us in our quest to unravel the intricacies of protein dynamics and functionality, paving the way for deeper insights into biological processes and molecular interactions.

3.3 Methods

3.3.1 HDX-MS

The study began with the preparation of the apo-IPNS sample, sourced from the Schofield group. The sample was carefully diluted to a concentration of 16µM using an equilibration buffer consisting of PBS at pH 7.4.

For the labelling step, 5 µL of this protein sample was further diluted to a total volume of 50 µL. This was achieved by adding a labelling buffer composed of PBS in D_2O with a pD of 7.4 (measured pH 7.0), creating a precisely controlled 90% D_2O environment. The labelling process was conducted over a range of time intervals.

Binding runs were completed by adding (Zn / Cd) Cl_2 and the tripeptide ACV. This was done to maintain the same concentration of the IPNS enzyme in the injected solution.

After the specified times, the labelling reaction was rapidly quenched by adding an equal volume of ice cold quench buffer, consisting of 50 mM KH_2PO_4 and 4 M urea, adjusted to a pH of 2.

The following phase involved online protein digestion using a Waters Enzymate BEH pepsin column, maintained at 20 ° C. After digestion, the resulting peptides were temporarily trapped in a Waters BEH C18 VanGuard precolumn for 3 minutes

at a flow rate of 200 µL/min using buffer A, which contained 0.1% formic acid at approximately pH 2.5. The peptides were then transferred to a Waters BEH C18 analytical column.

The elution was carried out using a linear gradient with buffer B, composed of 0.1% formic acid in acetonitrile, also at a pH of approximately 2.5, with a flow rate of 40 µL/min. To ensure accuracy and minimise back exchange, all trapping and peptide separation steps were carried out at approximately 0 ° C.

Data acquisition was consistent with all conditions and was collected using the MSE workflow on a synapt G2-Si instrument. This MSE workflow alternates between high and low collision energies in the hope of fragmenting all precursor ions with the acquisition window. Each data point was recorded in triplicate to ensure reliability. Furthermore, the mass spectrometer was independently calibrated with NaI, and a lock mass correction using LeuEnk was applied to the collected MS data.

For data analysis, peptide identification was performed using the ProteinLynx Global Server (PLGS, Waters Corporation, UK), and the isotope uptake of the identified peptides was precisely measured using DynamX v3.0 software.

3.4 Results

The coverage of IPNS in all states was greater than 88%. The +Zn(II) state found 72 common peptides, and the additional state of +Zn(II) and ACV found only 68 peptides. This decrease in peptides was more than likely due to some overlap and interference of the signal of the ACV tripeptide. The raw uptake plots for the three comparative states can be found in figures A.1, A.2, A.3 and A.4.

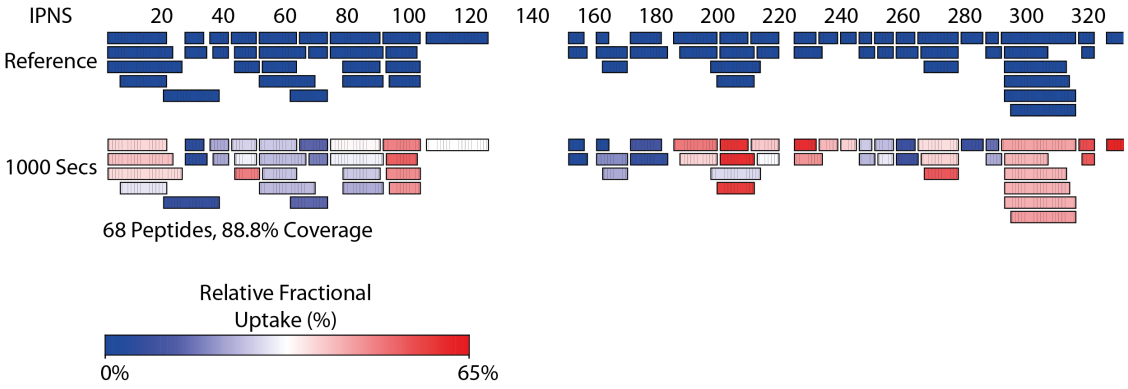


Figure 3.4: Here is the common peptide coverage map for apo IPNS. The reference state is shown above, with peptides identified presented as blue rectangle from stretching the length along the primary amino acid sequence. Below the reference time point is overlayed relative fractional uptake of each individual peptide, this is normalised to the highest value and coloured accordingly

The map of coverage of common peptides identified across the three comparative states is shown above throughout our experiment. The primary amino acid sequence is numbered from the N-terminus to the C-terminus in increments of 20 along the x-axis. The reference data (t = 0) is presented at the top of the figure, with each peptide represented by a rectangle that spans the length of that particular peptide.

In the lower portion of the map, common peptides that have been found and identified across all time points have been colour coded according to their fractional deuterium uptake at the final time point (1000 seconds) of the experiment for the apo state, referred to simply as IPNS when making comparisons across different states. This visualisation technique allows for an additional dimension of analysis, illustrating where along the primary sequence the protein exhibits greater or lesser deuterium uptake in the absence of any additional stimuli.

Upon examination, the heat map shown above exhibits regions of greater uptake (red in colour) at both the N and C terminal regions of IPNS. This is consistent with the loss of structure and the increased accessibility of solvents to protein termini[75, 76]. This work states that 80.3% of the N-terminal residues and 86.1% of the C-terminal residues are exposed compared to 32% for all residues. Therefore, we can generally expect an increase in uptake at the terminal ends of proteins, as seen here in Figure 3.4.

We can discern other regions of the protein that are more dynamic or flexible, as indicated by higher incorporation of deuterium, versus regions that are more stable and structured, showing less uptake. This detailed mapping is invaluable for understanding the structural and functional dynamics of the protein under study.

There are three other distinct areas that exhibit similar properties of higher uptake compared to the rest of the protein in the reference state. These areas lie around 90-100, 180-240, and 260-280 along the primary amino acid sequence. These areas include residues discussed earlier as being key players in the enzymatic activity of IPNS, and therefore, we see later some interesting dynamic behaviour when looking at the comparative states around these areas.

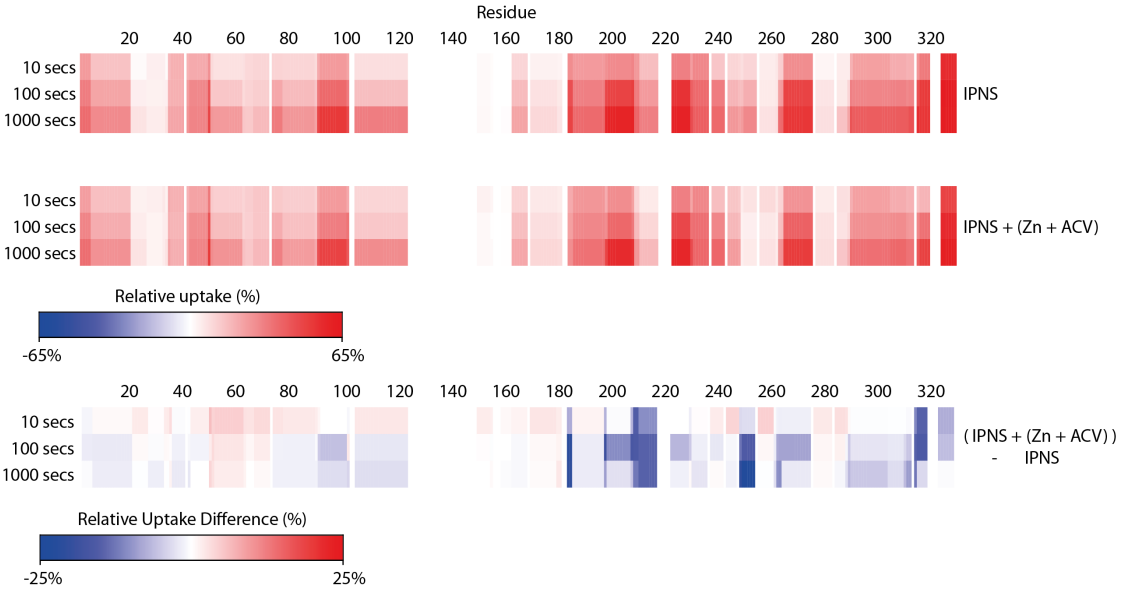


Figure 3.5: Heat maps from the raw HDX-MS data looking at the apo IPNS protein and the IPNS + (ACV and Zn) comparative state. Heat map shows the uptake as a coloured gradient at each labelling time as a 1D space along the x-axis marked residue. The lower portion is a comparative heat map showing the difference in uptake along the 1D space at each label time.

The heat maps provided represent two experimental states: the apo protein (labelled IPNS) and IPNS in the presence of both the Zn(II) metal ion and the tripeptide ACV. These maps display uptake across the primary sequence over time, using a colour-coded gradient where darker red signifies greater uptake. The differences between these states can be subtle, even imperceptible to the naked eye.

Therefore, a comparative analysis is performed by subtracting the apo state from the bound state and then normalising to the greatest differential value, resulting in a clearer comparative plot.

This normalised comparative plot reveals significant levels of protection (less uptake) in the bound state compared to the apo state, particularly around residues 210 to ~220 and 320 to ~330. These residues are known from other studies to be crucial for the catalytic activity of the enzyme IPNS[65, 77] These are areas that we discussed above in which we expected differences to occur because those are catalytically important for the enzymatic activity of IPNS. This method of analysis underscores the functional importance of these regions in response to metal ion binding and tripeptide interaction.

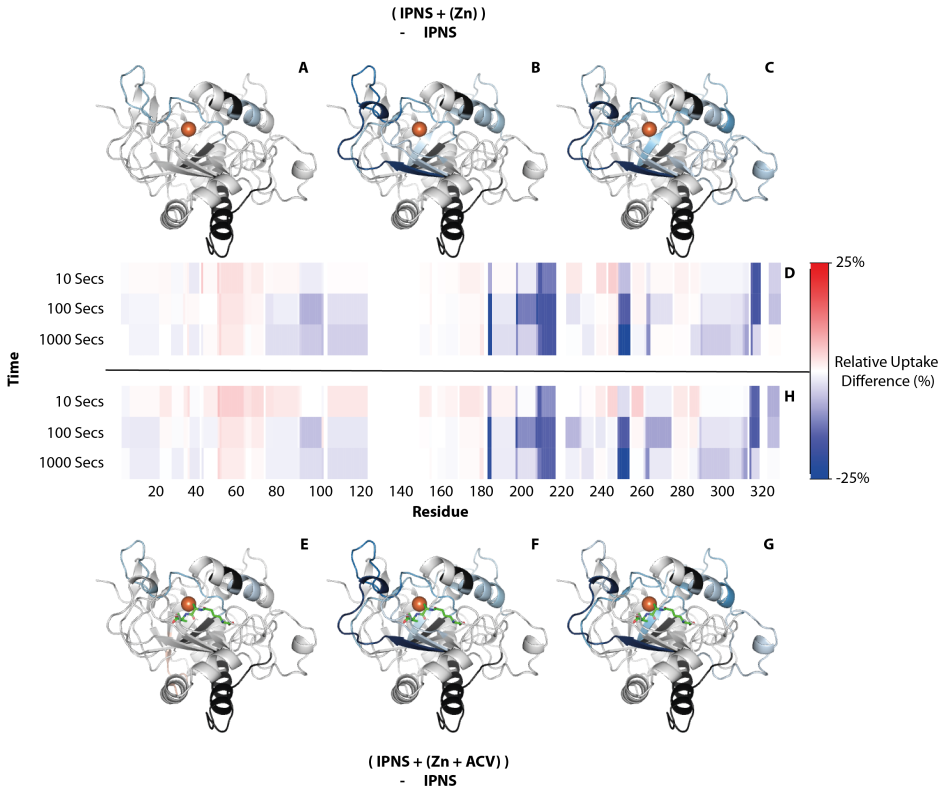


Figure 3.6: Figure showing two comparative heat maps, along with the coloured PDB files. PDB files (1BK0) have statistically different data coloured. **A-C** PDB file (1BK0) for the compared states of IPNS and (IPNS+Zn) at respective labelling times of 10, 100 and 1000 seconds. **D** Heat map for the IPNS and (IPNS+Zn) comparison.**E-G** PDB file (1BK0) for the compared states of IPNS and (IPNS+(Zn and ACV)) at respective labelling times of 10, 100 and 1000 seconds.**H** Heat map for the IPNS and (IPNS+(Zn and ACV)) comparison.Gradient colour bar shown to the right normalised across 0 to the max difference value.

Shown here are the comparative heat maps for the states IPNS + Zn and IPNS + (Zn + ACV), each compared to the apo state, labelled here as IPNS. These comparisons were performed using SAUSC, discussed in chapter 2. In both comparisons, clear differences occur around the residues known to affect the enzymatic activity of IPNS, namely His214, Asp216, and His270, indicating significant protection compared to the apo state. This protection arises from what we hypothesise to be the newly formed bonds to the metallic core. These formed bonds are essential for enzyme activity. Interestingly, when ACV is also present, there appears to be some additional protection around His270, another residue that is thought to be important when looking at the enzymatic activity of IPNS[65, 77]. However, this protection is more evident when the ACV substrate is present.

These differences have been tested for significance at the peptide level, as discussed in Chapter 2. The hybrid test discussed in chapter 2 was used to check the level of significance. α value of 0.01 was used in the analysis, which led to a confidence of 99%. This would mean that we would expect results found to be significant to have only a 1% chance of occurring by chance / error. The significant differences found are then mapped onto the 3-D structures for each labelling time, shown in panels A-C and E-F for the respective comparisons and labelling times. This mapping clearly demonstrates consistent protection around the metallic centre, which is made visible with the orange metallic ion in the PDB structure of IPNS.

3.4.1 Structural Relevance

These findings show upon binding of the metal ion that in this case the Zn(II) or Cd(II) ions were used to mimic Ferrous binding. There is an increase in the rigidity of IPNS that arises through the binding of that ion around the metal core; this is evident through the reduced uptake of peptides around the residues known to bind to these metallic ions. His214 and Asp216, two of the residues that were found to be of some importance when looking at the enzymatic activity of IPNS[66] exhibit deep protection - deep blue in colour. This leads one to assume that this change in conformation upon binding of the metal ion alters the dynamic structure of

the protein. Therefore, it would be of importance to the system for the catalytic purposes of the protein complex. His270 is also considered an important residue when dealing with enzymatic activity[66] however, yields a much less significant difference than in our experimental data, although this difference is more evident when the ACV substrate is present. Another area that yields a differential is that of residue Gln330, this again is likely due to the formation of electrostatic/hydrogen bonds to the metal core [65, 77] this however, is of a smaller margin when compared to other bound areas, more than likely due to it being at the terminal end of the protein and will likely maintain some level of solvent exposure and flexibility.

Figure 3.7 presents the cumulative data plots that compare the *apo* state of IPNS, our reference, with the IPNS + Cd state. These visualisations were generated using an internal script, detailed in Chapter 2. The significance of individual peptides was assessed by a hybrid form test, as described by Hageman and Weis[59]. The peptides found to be significant are consistently coloured across all panels **A-C** in Figure 3.7, providing a clear visual correlation between the different analyses. This uniform colouring aids in the interpretation of the data by linking the statistical analysis to the structural context within the protein.

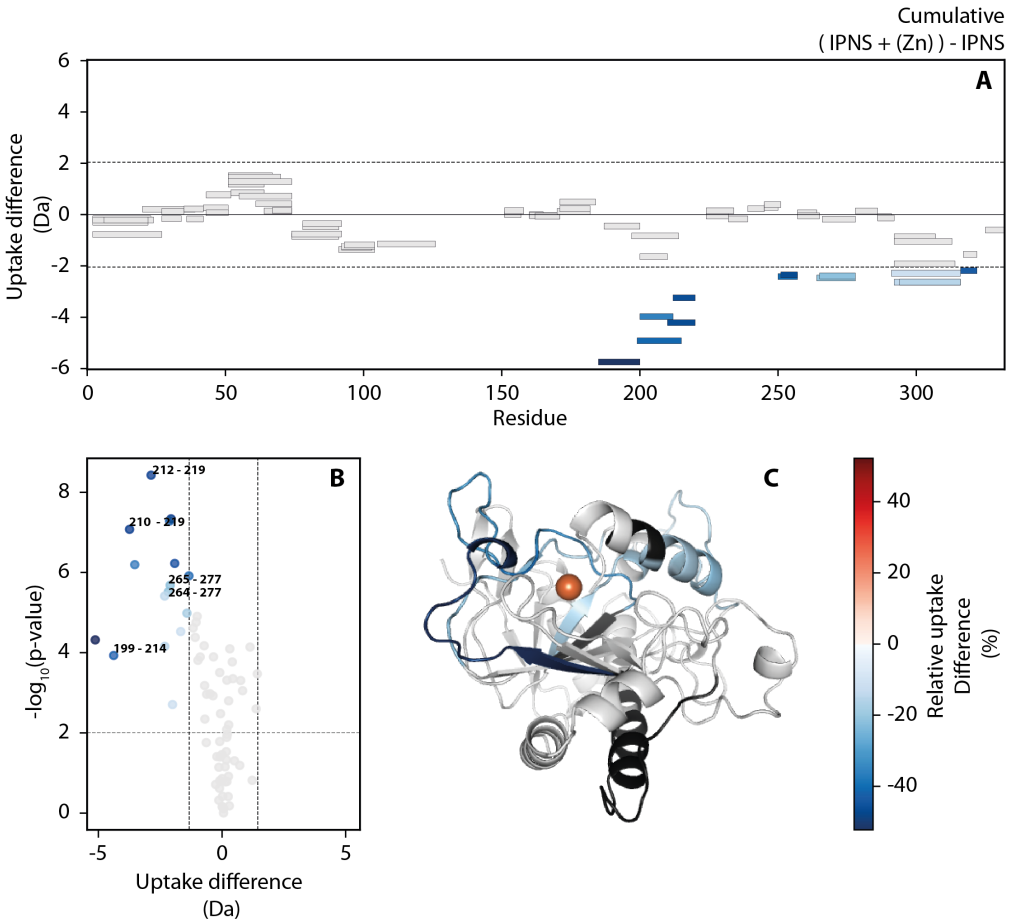


Figure 3.7: Here, we present cumulative plots and the subsequent structural mapping of significant regions onto the PDB file 1BLZ. **A** is a woods plot showing amino acid positions on the x-axis and uptake differences between two states on the y-axis. Common peptides identified in the analysis are represented as coloured rectangles. **B** is a volcano plot with peptides depicted as coloured circles, uptake differences on the x-axis, and -$\log_{10}$(p-value) on the y-axis. Peptides crucial for IPNS catalytic activity are annotated. **C** shows the PDB structure 1BLZ with significantly different peptides highlighted, using a consistent colour scheme across **A-C**. The gradient key is provided to the right.

We chose to annotate the peptides that were considered significant and spanned any of the residues previously mentioned to be of primary importance - His214, Asp216 and His270 - with regard to the catalytic activity of the IPNS complex[66]. Also analysed here are what we have described as the cumulative differences. This takes into account all data collected throughout the experiment. Through a summation process, we gain insight into the protein behaviour across all time points. We can see that the residues of His214 and Asp216 in the presence of the metallic core show evidence of strong protection. This is visually clear because

not only do the peptides spanning these residues have a greater uptake difference but also through the deeper blue colour that those peptides have designated - the colour is relative based on the length of the respective peptide - indicating a greater level of protection across those residues.

The annotated peptides also exhibit extremely low p-values, hence the -$\log_{10}$ respective values all show values around 4 or more. This means that we can confidently hypothesise that the differences that occur here can be attributed to the presence of the metal core and subsequent binding rather than to the fact that they occurred through chance or experimental error.

3.5 Discussion

3.5.1 Octahedral Coordination

The iron(II) centre in this context exhibits an octahedral coordination geometry, which provides several potential binding sites that are critical throughout the catalytic cycle. These sites are occupied sequentially by various molecular entities. These include the amide side-chain of Gln330, which interacts through hydrogen bonding or electrostatic interactions; one or two water molecules that contribute to the solvation environment; the sulphur atom from the substrate thiol group, pivotal in forming coordination bonds with the iron centre; the isopropyl group of the substrate d-valine, located close to the iron and interacting via van der Waals forces; molecular oxygen, which binds to the vacant coordination site to initiate oxygenation reactions; and reactive species generated from oxygen, such as superoxide or peroxide radicals, which participate in subsequent steps of the catalytic process[65, 77].

This coordinated interplay of molecules and ions in the iron(II) centre underscores its central role in facilitating enzymatic reactions, particularly in the context of penicillin biosynthesis as studied by McNeill et al. [77]. Understanding the dynamic changes in coordination and binding within the iron octahedral geometry provides crucial insights into the mechanistic steps involved in this biologically significant process.

It was clear from our findings that residues of importance when looking at the catalytic activity of IPNS, His214, Asp216, and His270 exhibit decreased levels of uptake compared to those of the reference state IPNS when in the presence of a metallic binding partner. This is what we expect as a result of changes (bonds formed) that affect the dynamic properties of the protein structure. All three of these residues have been found to be of great importance in terms of catalytic activity, and we can assume that, without the binding of these particular residues, the activity would not be preserved.

"Live as if you were to die tomorrow. Learn as if you were to live forever."

— Mahatma Gandhi

4

How does the presence of the C2 domain affect the structure and function of SHIP1 and SHIP2

Contents

4.1 Abstract 75
4.2 Introduction to SHIP1 and SHIP2 systems 76
4.2.1 Role of proteins in disease 77
4.3 SHIP 1 and 2 Structural make-up 78
4.4 Methods 81
4.4.1 HDX 81
4.5 Results 83
4.5.1 HDX 83
4.5.2 Activity assay 88
4.5.3 Molecular Dynamics 90
4.6 Discussion 93
4.6.1 C2 domain modulates flexibility 94

4.1 Abstract

In this study, we investigated the structural and functional dynamics of SHIP1 and SHIP2, focussing on the role of the C2 domain in the regulation of phosphatase activity. SHIP proteins are important in cell signalling, making them key therapeutic

targets. Using HDX-MS, molecular dynamics (MD) simulations, and activity assays, we explored how the C2 domain affects the flexibility and function of the phosphatase domain.

Our results revealed that removing the C2 domain increased the flexibility of key structural loops, particularly loop L4, with SHIP2 showing a more pronounced effect than SHIP1. This increased flexibility was correlated with a reduction in enzymatic activity, indicating that the C2 domain stabilises critical regions of the phosphatase domain. MD simulations further supported these findings, highlighting hydrogen bond disruptions when the C2 domain was absent.

In general, our study demonstrates that the C2 domain acts as a structural scaffold, stabilising the phosphatase domain and regulating SHIP activity. These insights could inform the development of selective inhibitors targeting SHIP proteins for therapeutic purposes.

4.2 Introduction to SHIP1 and SHIP2 systems

The pivotal role of SHIP1, an enzyme classified as inositol 5-phosphatase, within cellular signaling pathways underscores its importance in various physiological conditions[78]. Given its diverse implications, elucidating the structural intricacies of SHIP1 and devising specific small molecules tailored to target it present promising avenues in drug development. Using SHIP1 as a therapeutic target requires a comprehensive understanding of its molecular architecture and the strategic design of selective compounds, which marks a sophisticated approach to precision medicine [78]. Through a series of investigations involving activity assays, HDX-MS, and molecular dynamics analyses, we sought to unravel the mechanisms underlying the modulation of phosphatase domain activity by the C2 domains of both SHIP1 and its closely related counterpart, SHIP2. Our research reveals that the impact of the C2 domain on enzyme function is significantly more significant in the case of SHIP2 compared to SHIP1, highlighting the distinct regulatory nuances between these two variants [78].

4.2.1 Role of proteins in disease

SH2-containing inositol 5-phosphatases, namely SHIP1 and SHIP2, play an important role within the PI3K / AKT / mTOR signaling pathway. These enzymes are responsible for dephosphorylating phosphatidylinositol-3,4,5-trisphosphate (PI (3,4,5) P3) to produce phosphatidylinositol-3,4-bisphosphate (PI (3,4) P2). The delicate equilibrium between these two second messengers, along with related phosphatidylinositol phosphates, intricately regulates numerous cellular processes such as inflammation, autophagy, cell migration, metabolism, and cell growth[79]. Consequently, it is no surprise that SHIPs are involved in a myriad of conditions, including various cancers[80–82], diabetes[83], hypertension[84], and graft-versus-host disease[85].

4.2.1.1 Cancer

Within the complex landscape of the PI3K / AKT pathway, inositol 5-phosphatase SHIP1 occupies a pivotal role as a negative regulator. This pathway, known for its aberrant activation, particularly in a significant proportion (50-70%) of acute myeloid leukemias (AML), underscores the critical importance of SHIP1 in cellular homeostasis and disease pathogenesis. In fact, in the realm of AML, SHIP1 emerges as a focal point of investigation due to its substantial functional involvement[80].

In approximately 3% of patients with AML, a multitude of mutations have emerged through misense mechanisms, further accentuating the importance of SHIP1 in the pathogenesis of this hematologic malignancy. These mutations, which occur at various loci within the SHIP1 gene, underscore the multifaceted nature of SHIP1's functional impact and its intricate interplay within the context of AML pathophysiology.

The emergence of these mutations within the SHIP1 gene not only sheds light on the molecular underpinnings of AML but also underscores the dynamic nature of SHIP1's regulatory function within the PI3K/AKT pathway. Through meticulous genetic analyses and functional characterisation, researchers aim to

unravel the precise mechanisms through which SHIP1 mutations contribute to the pathogenesis of AML.

These findings not only deepen our understanding of the molecular mechanisms driving AML, but also offer promising avenues for targeted therapeutic interventions.

4.2.1.2 Alzheimer's

The INPP5D gene encodes the critical signaling molecule SHIP1, which plays a pivotal role in the regulation of cellular processes involved in health and disease, particularly the inflammatory process in the brain through cytokine release through signaling from TREM2 and DAP12[86–90].

The correlation between specific single nucleotide polymorphisms (SNP) and the INPP5D gene has garnered considerable attention, particularly in the context of late-onset Alzheimer's disease (LOAD). In particular, observed in individuals of Caucasian ethnicity, this association has been linked to an intriguing phenomenon: an increase in SHIP1 expression[91, 92]. This genetic predisposition, identified through meticulous genetic analyses and association studies, sheds light on the intricate interplay between genetic factors and disease susceptibility[93, 94].

The implications of these findings extend far beyond mere genetic correlations. They provide valuable insights into the molecular mechanisms underlying LOAD and provide a potential avenue for targeted therapeutic interventions. By elucidating the genetic underpinnings of SHIP1 expression modulation, researchers aim to unravel the intricate pathways that contribute to the pathogenesis of LOAD, paving the way for novel diagnostic and therapeutic strategies tailored to combat this debilitating neurodegenerative disorder.

4.3 SHIP 1 and 2 Structural make-up

The structural organisation of SHIP proteins encompasses several distinct domains, each contributing to their functional versatility. Upon initiation at the N-terminus, these proteins have a SH2 domain that is crucial to mediating interactions with phosphorylated membrane proteins [95]. Subsequently, a region of approximately

60 residues, characterised by its disorder, precedes a dimerisation / RhoA binding domain, which facilitates protein-protein interactions essential for signaling modulation[96, 97].

Centrally situated within the SHIP architecture lies a pleckstrin homology (PH) domain, recognised for its role in membrane targeting and phosphoinositide binding[98]. The following PH domain is the catalytic 5-phosphatase (Ptase) domain responsible for the dephosphorylation of phosphatidylinositol phosphates, a critical step in the regulation of intracellular signaling cascades[79, 99]. Adjacent to the Ptase domain lies a C2 domain, involved in calcium-dependent membrane targeting and protein-protein interactions, thus contributing to the intricate regulation of cellular processes.

Expanding further, SHIP proteins feature an additional disordered region spanning approximately 300 residues, the functional significance of which warrants further exploration[98]. It should be noted that SHIP2 exhibits two unique domains at the c-terminus [100]. These domains include the ubiquitin interacting motif (UIM) and the steril alpha motif (SAM), both of which are not present in SHIP1. The identities of the two complexes are very similar as they share the identity 57%, as shown in Figure B.1.

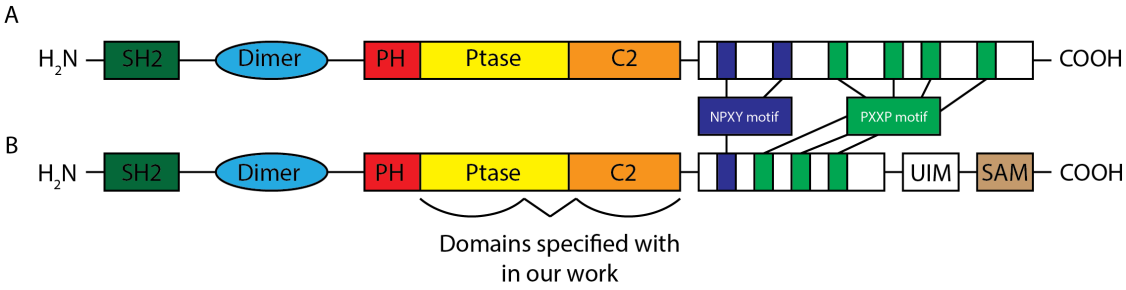


Figure 4.1: Schematic adapted from pedicone et al [100].**A** shows the domain make up of SHIP1 as a 2D representation from n to c terminus. **B** similarly shows a similar representation for SHIP2. Making clear the similarities and differences between the two.

In essence, the elaborate structural arrangement of SHIP proteins underscores their multifaceted roles in cellular signaling and highlights their potential as crucial regulators of various physiological and pathological processes.

The domains that we focus upon in this chapter are shown in Figure 4.1.

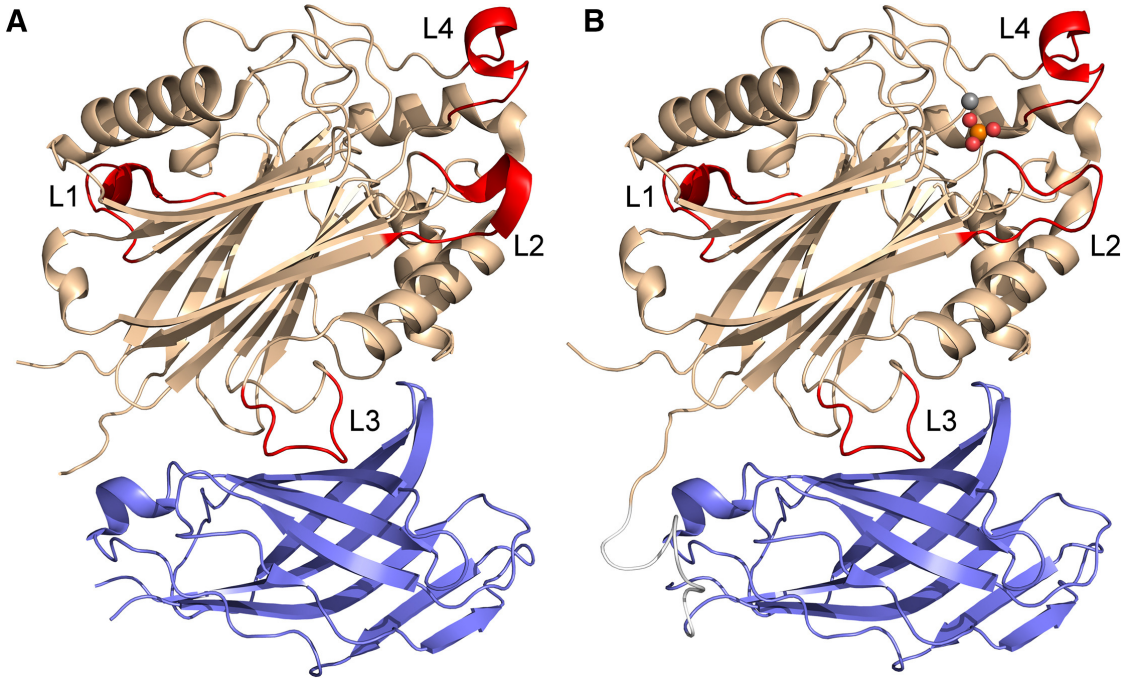


Figure 4.2: (A) The apo structure. The phosphatase domain is coloured brown and the C2 domain is blue. Loops 1 to 4 of the phosphatase domain, as numbered in Le Coq et al[95]for SHIP2, are coloured red.**(B)** The magnesium and phosphate bound structure, coloured as in A. A magnesium ion (grey) and a phosphate ion (orange and red) are shown bound to the active site

Although the catalytic site remains conserved across SHIP proteins, subtle structural disparities between them present promising opportunities for strategic exploitation in inhibitor development. A pivotal advancement in this pursuit stems from the comprehensive structural characterisation of SHIP2's phosphatase domain in conjunction with its C2 domain, which has yielded critical insights into their functional interplay. Le Coq et al. meticulously dissected four loops within the phosphatase domain, denoted L1 to L4, attributed functional significance to three of them[95].

A notable focus area lies within loop L2, which spans residues K510 to G518 in SHIP1 and K531 to G539 in SHIP2. It is postulated that this loop plays a pivotal role in facilitating membrane penetration and guiding substrate delivery to the catalytic site, thereby presenting a potential avenue for targeted intervention. In addition, loop L3 (which comprises residues D566 to N573 in SHIP1 and D587 to D594 in SHIP2) emerges as a crucial mediator in the interaction with the C2

domain. Its role is hypothesised to involve the transmission of allosteric signals to the catalytic site, exerting influence over enzyme activity.

Equally intriguing is loop L4 (spanning from T656 to N666 in SHIP1 and N674 to N684 in SHIP2), implicated in orchestrating the allosteric relay mechanism and contributing to the structural integrity of the active site. The proposed mechanism, defined by Le Coq et al., integrates a myriad of methodologies, including structural analyses, activity assays employing various mutants, and dynamic simulations.

This multidimensional approach not only provides a nuanced understanding of the regulatory intricacies governing SHIP protein function but also provides invaluable insight for the rational design of targeted inhibitors with therapeutic potential. By unraveling the structural nuances and functional dynamics of SHIP proteins, researchers can strategically exploit these molecular intricacies to develop novel therapeutic interventions tailored to combat diseases associated with deregulated PI3K signaling pathways.

4.4 Methods

In this work, I performed the HDX-MS portion. The rest of the work was completed by collaborators and the methods for these works can be found in the published paper 'Regulation of inositol 5-phosphatase activity by the C2 domain of SHIP1 and SHIP2' [78].

4.4.1 HDX

The study commenced with the preparation of the SHIP constructs. All constructs were systematically diluted to achieve a 12µM concentration using an equilibration buffer, specifically PBS at pH 7.4.

For labelling purposes, 5µL of this protein sample was further diluted to a cumulative volume of 50µL. This was achieved by integrating a labelling buffer, a concoction of PBS formulated in D_2O, with a pD value of 7.4 (pH read 7.0), resulting in a controlled 90% D_2O milieu. The labelling was executed over an array of time intervals.

Following the predetermined durations, the labelling reaction was immediately quenched. This was carried out through a 1:1 integration of an ice-cold quench buffer, comprising 50mM KH_2PO_4 and 4M Urea, calibrated at a pH of 2. This was done to slow both the forward and reverse exchange reactions as well as to aid in protein unfolding.

The following online protein digestion, facilitated using the Waters Enzymate BEH pepsin column, with the temperature held at 20°C. After digestion, the peptides were trapped on a Waters BEH C18 VanGuard precolumn. This stage lasted 3 minutes, with a consistent flow rate of 100 µL/min, employing buffer A, characterised by 0. 1% formic acid at an approximate pH of 2.5. Subsequent to this entrapment, the peptides were transferred to a Waters BEH C-18 analytical column.

The elution procedure was orchestrated through a linear gradient, using buffer B, which consists of 0.1% formic acid suspended in acetonitrile. This solution hovered around a pH value of 2.5. The elution maintained a flow rate of 40 µL/min. Importantly, to ensure precision and minimise back exchange, all trapping and peptide separation operations were carried out at a controlled temperature of approximately 0 ° C.

Data acquisition remained consistent across states and was gathered via the MSE workflow on a synapt G2-Si instrument. This workflow was used to aid with the fragmentation of precursor ions. The rigour of the experiment was amplified by obtaining all data points in triplicate. Furthermore, the MS data were subjected to an independent calibration using NaI, and the collected MS data were enhanced with a lock mass correction facilitated by LeuEnk.

The final data interpretation involved the assignment of peptides through the advanced ProteinLynx Global Server (PLGS, Waters Corporation, UK). Furthermore, the isotopic uptake corresponding to the identified peptides was accurately determined using the DynamX v3.0 software. Data were compared and analysed using the in-house script developed in the chapter 2.

4.5 Results

4.5.1 HDX

To assess the influence of the C2 domain on the flexibility of the phosphatase domain, we performed comprehensive HDX-MS analyses on the SHIP1 and SHIP2 constructs, comparing them with and without the C2 domain [78]. The HDX-MS technique allowed us to probe the dynamics of protein structures by measuring the exchange of labile hydrogen atoms with deuterium in solution over time.

For the SHIP1 constructs, we measured deuterium uptake (DU) at intervals of 1, 10, and 30 minutes. Meanwhile, for the SHIP2 constructs, DU measurements were taken at 1, 2, 5, and 10-minute intervals. These varied time points were chosen to capture different stages of the exchange process, providing insights into the kinetics of hydrogen exchange within the protein structures.

After digestion with pepsin, a total of 83 peptides were identified as common between the two SHIP1 constructs, representing approximately 88% coverage of the phosphatase domain. The raw uptake plots for the common peptides found in SHIP1 can be found in figures B.2 and B.3 in the appendix.

Similarly, for the SHIP2 constructs, 66 peptides were shared, covering approximately 86% of the phosphatase domain. This extensive coverage of the peptide ensured a thorough analysis of the regions encompassing the phosphatase domain, enabling us to assess the impact of the C2 domain on the flexibility of this critical region with high resolution. These raw uptake plots can also be found in the appendix in figures B.4 and B.5.

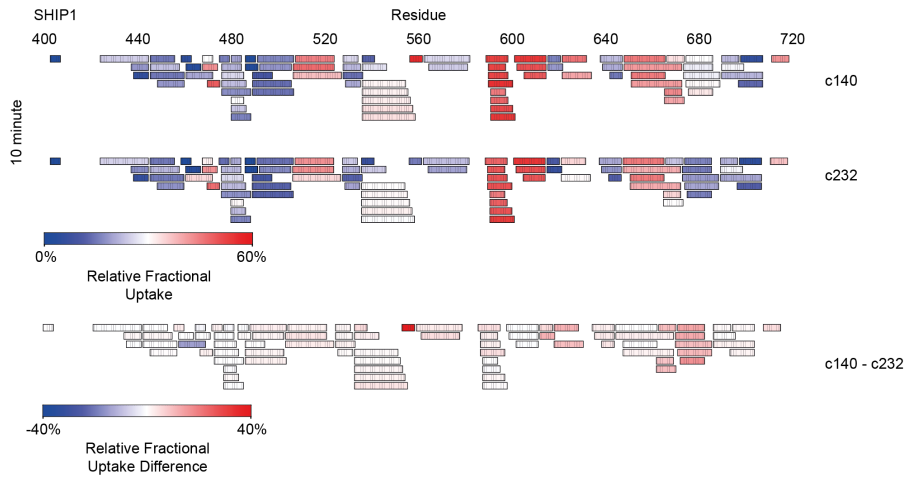


Figure 4.3: Shown here is the coverage map for the common peptides found across both constructs of SHIP1 with the uptake data from t=10m of each peptide mapped in colour on to the respective peptides. The constructs with and without the C2 domain are labelled here as c232 and c140 respectively. The lower has on the mapped relative difference across the two constructs.The top and bottom require different gradients.

Figure 4.3 shows the common peptides found in the c140 and c232 constructs of SHIP1, these are the constructs in the absence and presence of the C2 domain, respectively. Each peptide, rectangle along the x-axis, is coloured, respectively, with their relative uptake data from the labelling point t = 10 min. The gradient is set from the maximum value of both constructs, with dark blue indicating lesser uptake to dark red, which signifies greater levels of uptake. From the maps shown in figure 4.3 their are subtleties across the two constructs. These differences are clearer within the difference plot. We see a great difference around L3 - comprising residues 566 - 573 - which is known to serve as a pivotal moderator in the interaction of the ptase domain with the C2 domain. The L4 region, residues 656 - 666, also exhibits differences in uptake between the two constructs. Evident again, through the red colouring of the peptides in the difference map in figure 4.3. This region is believed to contribute to the structural integrity of the complex and could lead to the assumption that the C2 domain may help regulate this function; however, this increase in uptake could arise from the inherent lack of structure at the c-termini of the c140 construct as it approaches the terminal end [75].

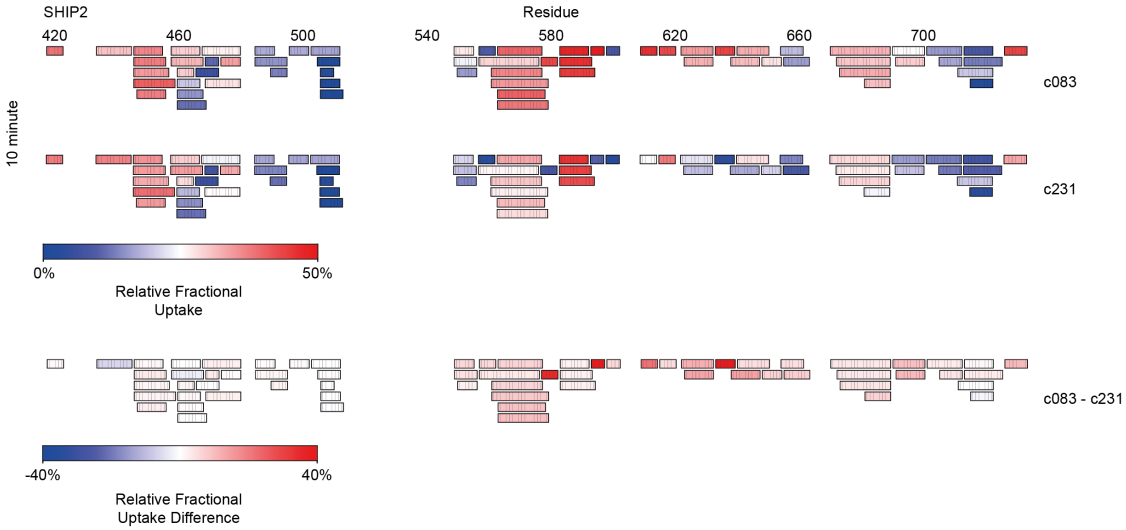


Figure 4.4: Shown here is the coverage map for the common peptides found across both constructs of SHIP2 with the uptake data from t=10m of each peptide mapped in colour on to the respective peptides. The constructs with and without the C2 domain are labelled here as c231 and c083 respectively. The lower has on the mapped relative difference across the two constructs.The top and bottom require different gradients.

Similarly to Figure 4.3, Figure 4.4 shows the data in the same way this time with the t = 10 minute data of SHIP2. In the difference plots, this time, we see some more areas portrayed in red. Aligned with SHIP1, both L3 and L4 again show increased uptake values when the C2 domain is absent. Again, perhaps suggesting that the C2 domain plays a crucial role in regulating these same structural roles. This is shown in figure 4.5 when we examine the cumulative plots provided by SAUSC.

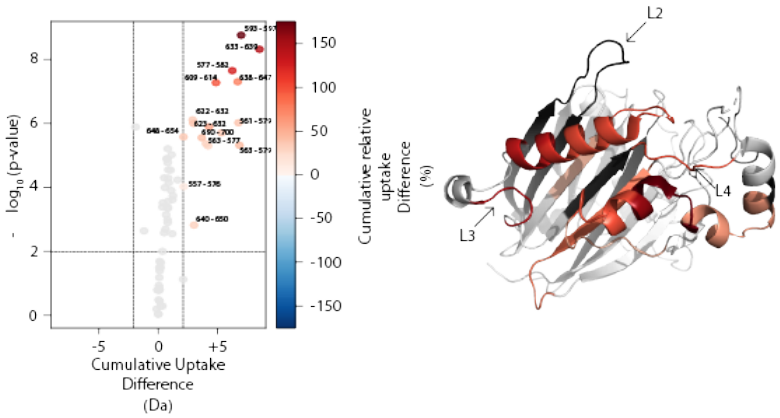


Figure 4.5: The cumulative volcano plot annotated with peptides that exhibited differences through our statistical testing. Along with the respective PDB (5OKM) file, coloured relatively as done by SAUSC from chapter 2. Areas of interest i.e L2,3 and 4 are highlighted and clearly marked on the PDB file.

By comparing the HDX-MS data obtained from the constructs with and without the C2 domain, our objective was to elucidate any differences in the dynamics and flexibility of the phosphatase domain induced by the presence or absence of the C2 domain. These findings would provide valuable information on the structural and functional implications of the C2 domain in regulating the activity and conformational dynamics of SHIP proteins.

From our HDX results, it is clear that there is a increase in flexibility in and around the loops mentioned above by Le Coq et al. when the C2 domain is not present, suggesting that the C2 domain could be central to the regulation of the Ptase domain.

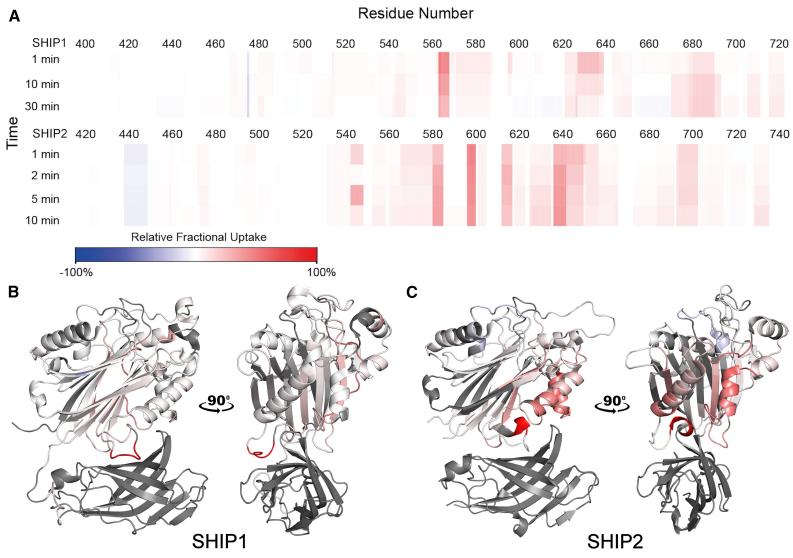


Figure 4.6: (A) The differential heatmaps visually represent variations in deuterium uptake (ΔDU) between constructs with and without the C2 domain. In the heat map, regions exhibiting greater DU for constructs containing the C2 domain are depicted in blue, while those with greater DU for constructs lacking the C2 domain are depicted in red. Notably, constructs devoid of the C2 domain display higher DU, with this effect more pronounced in SHIP2 compared to SHIP1. **(B)** The ΔDU at the 1-minute time point is depicted on the structure of SHIP1 (PDB ID: 6XY7). Regions where DU measurements were not observed, as well as those encompassing the C2 domain, are displayed in grey. **(C)** Similarly, the ΔDU at the 1-minute time point is mapped onto the structure of SHIP2 (PDB ID: 5OKM chain B). Again, regions not encompassing DU measurements and those corresponding to the C2 domain are depicted in grey for clarity.

First of all, the time points for both constructs differ. with 1, 10 and 30 being measured for SHIP1 constructs and 1,2,5, and 10 being measured for SHIP2

constructs. This was due to time constraints, and since it being a new complex of investigation, we were unsure what time period we would need to examine to see visible differences. If repeated, time points for both protein complexes would ideally match up and allow us to analyse fully across proteins. However, for the purpose of what we wanted to show, the work done here was adequate, despite not being ideal. With the focus of these proteins within the loops of the structure, shorter time points were suitable because those areas exhibited uptake at a higher rate and therefore any differences could be seen at earlier time points.

Although differences in deuterium uptake (DU) were discerned between Ptase and Ptase-C2, it is plausible that shorter time points could have elucidated clearer distinctions, especially in highly flexible regions such as L4. This observed phenomenon is rooted in the intricate interplay between uptake levels and specific protective factors at the atomistic level. Regions characterised by diminished rigidity or increased flexibility may exhibit uptake rates that, even at shorter time intervals, seem indistinguishable as a result of the rapid uptake of deuterium, obscuring potential differences in uptake kinetics.

Most of the DU variances observed demonstrated higher DU in the shorter construct, with discrepancies being more pronounced for SHIP2 compared to SHIP1. This implies that Ptase constructs inherently possess greater flexibility or solvent accessibility than Ptase-C2 constructs. This expectation aligns with regions typically buried in the interface between the two domains but not necessarily in more distant parts of the phosphatase domain. This suggests that the inclusion of the C2 domain may be of some intrinsic significance in regulating both the functional operation and structural integrity of the entire complex.

This suggestion hints at a nuanced interplay in which the presence of the C2 domain potentially exerts a multifaceted influence, impacting not only the functional aspects but also the overall structural robustness of the complex. This observation is further underscored by the data presented in figure. 4.7 and Tab. 4.1, which clearly illustrates a reduction in the activity of both SHIP1 and SHIP2 constructs in the absence of the C2 domain. The most significant disparities were observed in

residues approximately between 580 and 650 in SHIP2, which includes a part of α 4 (helix 4), L3 (interacting with the C2 domain), strand 9 (connecting L3 to the active site) and α5, α6, and α7. Although L3 and α 7 in SHIP1 also exhibited substantial deprotection, the impact on residues between these regions was less pronounced [78].

4.5.2 Activity assay

To gain an understanding of how the C2 domain affects phosphatase activity. In Bradshaw et al[78] we performed an activity assay to measure the phosphate produced by both the phosphatase domain alone and also in the presence of the C2 domain of both SHIP1 and 2 complexes shown in figure.4.7.These assays were carried out by collaborators and were done by utilising a phosphate sensor composed of a fluorescently labeled E. coli-derived inorganic phosphate-binding protein, specifically tagged with MDCC (N-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide), the assay showcased remarkable sensitivity. This sensitivity was highlighted by its lower limit of quantification, reaching into the picomolar range, a notable departure from conventional malachite green-based assays, which typically operate within the higher nanomolar range. This increased sensitivity proved instrumental in accurately quantifying phosphate levels, even when present at low substrate concentrations. Consequently, the assay demonstrated enhanced precision and reliability, particularly in situations that required meticulous quantification within a low concentration regime. The use of MDCC-labelled protein allowed the detection of subtle changes in phosphate concentration, thereby extending the applicability of the assay to a broader spectrum of experimental conditions requiring precise measurement in challenging concentration ranges.

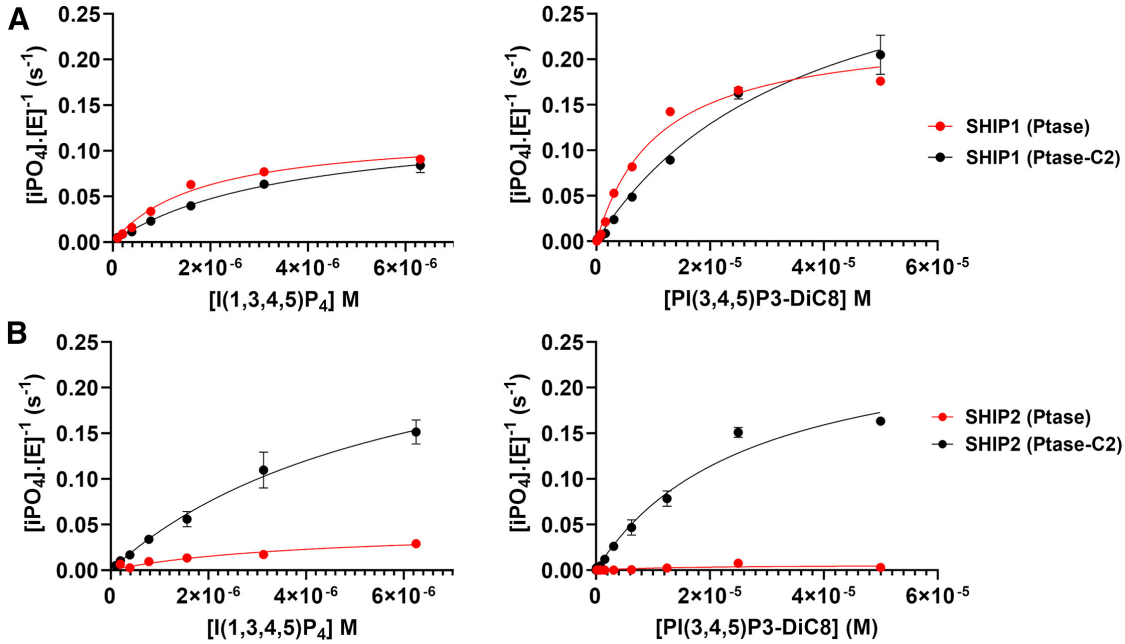


Figure 4.7: (A) The enzymatic activity of 50 nM SHIP1 was assessed against either I(1,3,4,5)P4 (on the left) or PI(3,4,5)P3-DiC8 (on the right). Initial rates were graphed as the amount of phosphate produced per second per molar concentration of enzyme. Notable differences in catalytic activity were not observed in the presence (depicted by black circles) or absence (depicted by red circles) of the C2 domain for either substrate. **(B)** The phosphatase activity of 50 nM SHIP2 was evaluated against I(1,3,4,5)P4 (on the left) or PI(3,4,5)P3-DiC8 (on the right). Consistent with previous observations, the catalytic efficiency of SHIP2 appeared to be enhanced in the presence of the C2 domain (illustrated by black circles) compared to the phosphatase alone (indicated by red circles). For I(1,3,4,5)P4, the concentration of SHIP2 (Ptase-C2) was adjusted to 15 nM. Data, represented as mean ± SEM, were obtained from three independent experiments

The presence of phosphate contamination in the substrates posed a significant challenge, as it restricted the attainable usable concentrations to values of limited scope. Consequently, the determination of K_m values became inherently challenging, particularly when dealing with low substrate concentrations. Both K_m and K_{cat} were determined for each construct for PI(3,4,5)P_3-DIC8 and I(1,3,4,5)P_4 using the phosphate sensor assay[78]. These two substrates are water-soluble alternatives that, although not fully replicating the behaviour of a membrane-bound substrate with longer acyl tails, are commonly used as analogs[101–103].

Activity Assay - errors are SEM				
	PI(3,4,5)P_3-DIC8		I(1,3,4,5)P_4	
Phosphate sensor	K_m(uM)	K_{cat}(s^{-1})	K_m(uM)	K_{cat}(s^{-1})
SHIP1 Ptase	11.4 ± 1.6	0.24 ± 0.01	1.8 ± 0.2	0.12 ± 0.006
SHIP1 Ptase-C2	39.4 ± 5.7	0.38 ± 0.03	3.7 ± 0.4	0.13 ± 0.008
SHIP2 Ptase	19.3 ± 13	0.009 ± 0.003	6.9 ± 0.75	0.064 ± 0.003
SHIP2 Ptase-C2	26.6 ± 3.7	0.26 ± 0.02	2.7 ± 0.6	0.18 ± 0.02

Table 4.1: Table showing K_m and K_{cat} values determined using a Phosphate sensor for Ptase and Ptase-C2 constructs of both SHIP1 and SHIP2 to highlight the change in activity in the presence or abscence of the C2 domain.

The activity assay yields a conspicuous observation: a discernible reduction in enzymatic activity manifests itself in the absence of the C2 domain. This finding underscores the pivotal role of the C2 domain within the catalytically active segment of the Ptase domain. This empirical evidence not only accentuates the indispensibility of the C2 domain, but also underscores its intricate interplay within the enzymatic mechanism, shedding light on the nuanced functional dynamics governing this critical protein domain.

4.5.3 Molecular Dynamics

In our study Bradshaw et al. [78], MD simulations were performed to assess the comparative flexibility of the Ptase domains of SHIP1 and SHIP2, both in the presence and absence of the C2 domain. Initially, the intention was to replicate the simulations performed by Le Coq et al.[95] on SHIP2, starting with an L4-out conformation. However, our simulations began with SHIP2 in an L4-in conformation. Subsequently, a similar approach was employed for SHIP1, also starting with an L4-in conformation. In particular, attempts to replicate the simulations for the L4-out conformation were hampered by the unavailability of such a structural template. The analysis of these simulations unfolded through a meticulous dual-pronged approach. Initially, we embarked on discerning the root-mean-square fluctuations

(RMSFs) that govern the dynamic behaviour of C-alphas embedded within the phosphatase domain. This analytical endeavour afforded us the opportunity to unravel the nuanced influence exerted by the C2 domain on the intrinsic flexibility exhibited by the phosphatase domain, with a particular emphasis on scrutinising the behaviour of the critical L4 region.

Subsequently, our investigative gaze shifted towards a meticulous examination of intermolecular distances traversing the simulation trajectories. Specifically, we meticulously assessed the spatial relationships between residues K664/R682 (for SHIP1/SHIP2, respectively) nested within L4 and their proximate acidic counterparts (D592 and E597 in SHIP1, D613, and D615 in SHIP2). These acid residues, as posited by Le Coq et al.[95], have been hypothesised to play pivotal roles in stabilising the elusive L4 out conformation. In addition, they are conjectured to orchestrate vital signaling interactions that bridge the C2 domain and the active site.

Through these methodical analyses, our objective was to gain insight into the intricate interplay of conformational dynamics and structural determinants that govern the regulatory mechanisms inherent in the SHIP1 and SHIP2 proteins.

Our investigations of SHIP2 L4-in behaviour yielded results in alignment with the overarching trends elucidated by Le Coq et al.[95]. In particular, our simulations corroborated the notion that L4 exhibits heightened flexibility, and the stability of hydrogen bonds crucial for maintaining an L4-out conformation is compromised in the presence of the C2 domain. However, it should be noted that while our findings echoed these general trends, the observed effects were somewhat subdued compared to those reported in previous studies.

In contrast, simulations focussing on SHIP2 L4-out unveiled an intriguing reversal of dynamics. Here, we observed a scenario where L4 demonstrated greater flexibility, and the hydrogen bonds essential to maintaining an L4-out conformation showed increased susceptibility to disruption in the absence of the C2 domain.

Turning our attention to SHIP1 L4 in simulations, the outcomes presented a more refined narrative. Although the analysis of C-alpha RMSFs hinted at increased flexibility in L4 in the absence of the C2 domain, examination of the

minimum distances between K664 and D592 suggested that the presence or absence of the C2 domain does not significantly influence the formation of hydrogen bonds between these residues.

These comprehensive analyses offer valuable information on the interaction between the C2 domain and the dynamic behaviour of L4 across different SHIP isoforms, shedding light on the intricate regulatory mechanisms governing their functionality.

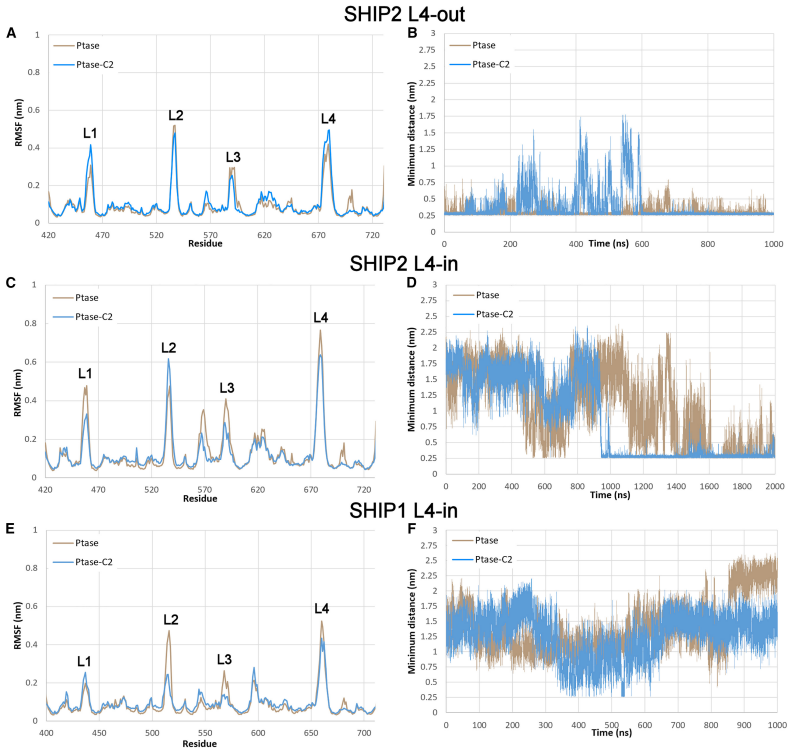


Figure 4.8: **(A)** SHIP2 in the L4-out state: The average root-mean-square fluctuations (RMSFs) derived from three iterations of simulations, as conducted by Le Coq et al., are depicted here. This representation corresponds to Figure 5A in Le Coq et al., with regions labeled from L1 to L4. **(B)** SHIP2 in the L4-out state: The minimum distance between R682 and D613/D615, obtained from replicates of simulations following the methodology outlined by Le Coq et al., is visualised. This graph reflects Figure 5D in Le Coq et al. and is indicative of three repeated experimental runs. **(C)** SHIP2 in the L4-in state: The average RMSFs resulting from simulations utilising 3NR831 and 3NR8 with the C2 domain extracted from chain B of 5OKM are illustrated. **(D)** SHIP2 in the L4-in state: A representative depiction of the minimum distance between R682 and D613/D615, obtained from simulations employing 3NR8 and 3NR8 with the C2 domain sourced from chain B of 5OKM, is provided. **(E)** SHIP1 in the L4-in state: The mean RMSFs derived from simulations utilising the current apo structure, 6IBD, are showcased. **(F)** SHIP1 in the L4-in state: A representative visualisation of the minimum distance between K664 and D592/E597 is presented.

4.6 Discussion

In this study, we embarked on a comprehensive exploration of the structural dynamics and functional implications of SHIP1 and SHIP2, with a particular focus on the interaction between their phosphatase and C2 domains. Our investigation was based on the recognition of the pivotal role that SHIP proteins play in cellular signaling pathways, making them attractive targets for therapeutic intervention in various pathological conditions.

The elucidation of the structural organisation of SHIP1 and SHIP2, which encompasses distinct domains such as SH2, dimerisation / RhoA binding, PH, ptase, and C2, provided a foundational understanding of their functional versatility and potential regulatory mechanisms. Through a multidimensional approach encompassing activity assays, HDX-MS and MD simulations, we aspired to unravel the intricate mechanisms that govern modulation of phosphatase domain activity by the C2 domains of SHIP1 and SHIP2.

Our HDX-MS analyses revealed distinct differences in the impact of the C2 domain on the flexibility of the phosphatase domain between SHIP1 and SHIP2. Specifically, we observed an increase in flexibility in regions that encompass critical loops, particularly loop L4, in the absence of the C2 domain, with SHIP2 exhibiting more pronounced effects compared to SHIP1. These findings were further corroborated by our MD simulations, which provided insight into the conformational dynamics of the phosphatase domains, shedding light on the subtle nuances governing their regulatory mechanisms.

Furthermore, our activity assays highlighted a discernible reduction in enzymatic activity in the absence of the C2 domain, underscoring the indispensable role that this domain plays in modulating the catalytic function of SHIP proteins. Importantly, the increased sensitivity and precision afforded by our phosphate sensor assay facilitated the accurate determination of kinetic parameters, providing valuable insight into the functional implications of the C2 domain.

The detailed analyses presented in this study provide a comprehensive understanding of the structural dynamics and functional implications of the SHIP1 and

SHIP2 proteins, paving the way for a deeper understanding of their roles in cellular signaling pathways and their potential as therapeutic targets. The elucidation of the intricate interplay between the phosphatase and C2 domains underscores the complexity of SHIP protein regulation and opens new avenues for the development of targeted inhibitors with therapeutic potential. Figures such as 4.6 and 4.8 provide visual representations of our findings, improving the clarity and comprehension of our results, highlighting the importance of the C2 domain in relation to the structural and functional importance of L3 and L4. Furthermore, the data presented in Table 4.1 provide quantitative insights into the changes in enzymatic activity observed in our assays, further corroborating our findings.

In summary, our study sheds light on the multifaceted roles of SHIP1 and SHIP2 proteins in cellular signaling pathways and provides a comprehensive framework for further exploration of their functional implications and therapeutic potential in various pathological conditions.

4.6.1 C2 domain modulates flexibility

The C2 domain plays a crucial role in modulating the flexibility of the phosphatase domain in both the SHIP1 and SHIP2 proteins. Through a combination of HDX-MS analyses and MD simulations, we observed distinct differences in the impact of the C2 domain on the structural dynamics of the phosphatase domain between SHIP1 and SHIP2.

Our HDX-MS analyses revealed that the absence of the C2 domain led to increased flexibility in regions encompassing critical loops, particularly loop L4, in both SHIP1 and SHIP2. This increased flexibility suggests that the C2 domain acts as a stabilising element, restricting dynamic movements of the phosphatase domain. In particular, SHIP2 exhibited more pronounced effects compared to SHIP1, indicating a differential regulatory role of the C2 domain between these isoforms.

Furthermore, our MD simulations provided insight into the conformational dynamics of the phosphatase domains in the presence and absence of the C2 domain. The simulations highlighted the intricate interplay between the C2 domain and

critical residues within loop L4, which are involved in stabilising the elusive L4-out conformation. In the absence of the C2 domain, disruptions in hydrogen bonding interactions were observed, leading to increased flexibility in loop L4.

In general, our findings suggest that the C2 domain acts as a structural scaffold, regulating the flexibility of the phosphatase domain in the SHIP1 and SHIP2 proteins. By stabilising critical regions such as loop L4, the C2 domain plays a pivotal role in maintaining the structural integrity and functional dynamics of these proteins, ultimately influencing their enzymatic activity and cellular signaling functions.

"We cannot solve our problems with the same thinking we used when we created them."

— Albert Einstein

5

Overcoming Glycan Heterogeneity Using an Adapted HDX Workflow

Contents

5.1 Abstract 98
5.2 Introduction 98
5.2.1 Inherent Heterogeneity 99
5.3 pH Problem 101
5.3.1 PNGase F 101
5.3.2 Enzyme Inactivity 103
5.3.3 Acidophiles - Potential Answers 104
5.4 Development of PNGase Rc Column 105
5.4.1 Addition to the Typical Workflow 105
5.4.2 Activity and Reproducibility 106
5.4.3 Methods 108
5.5 Results 110
5.5.1 Choice of Protein 110
5.5.2 HDX data post the addition of the PNGase Column . . 111
5.5.3 Column Performance 113
5.5.4 Individual Protein optimisation 114
5.5.5 Back exchange 115
5.6 Discussion 115

5.1 Abstract

This chapter presents the development and application of the novel PNGase Rc column for the online removal of N-linked glycans during HDX-MS of glycoproteins. By incorporating the column into a standard HDX-MS setup, we achieved improved peptide identification and sequence coverage, particularly in glycosylated regions previously inaccessible to analysis. The column was tested on several glycoproteins, including transferrin, C-Met, haptoglobin, and osteopontin, showing a notable improvement in coverage across all samples. The column demonstrated compatibility with typical HDX-MS quench conditions, including TCEP and Urea, with minimal impact on deuterium back-exchange. These findings suggest that the PNGase Rc column is a versatile tool to improve glycoprotein analysis in HDX-MS and other proteomics workflows requiring rapid deglycosylation.

5.2 Introduction

Common applications of HDX-MS include the analysis of conformational changes in proteins during function and regulation, such as membrane proteins, protein-ligand interactions, drug mechanisms, and epitope mapping for therapeutic antibodies. It is also valuable for studying post-translational modifications (PTMs) and their impact on protein dynamics and function[44, 104–106]. The glycosylation process, specifically N-linked glycosylation, constitutes a predominant facet within the realm of PTM[107–109], affecting numerous proteins. These proteins encompass a wide array of entities, including cell surface receptors, extracellular proteins, and biopharmaceutical agents, each serving diverse functional roles. The appended glycan moieties confer several crucial advantages, such as enhanced structural integrity, increased solubility, and sophisticated binding partner recognition mechanisms[110]. These attributes enable proteins to fulfil essential functions beyond the confines of cellular environments. Consequently, the intricate interplay of glycosylation makes it a focal point of great interest within the biological community, particularly when considering its unaltered native state[111].

Due to the intrinsically heterogeneous characteristics of glycans[112, 113], coupled with the intricacies involved in deciphering the data they generate, the study of glycoproteins by HDX-MS presents a formidable and multifaceted challenge[104, 112–114]. Within a single sample, diverse glycoforms of a protein may coexist, exerting a profound influence on the precision and amplitude of ion signals associated with those specific glycosylated residues. This intrinsic complexity extends further because the glycan moieties themselves can sequester deuterium during the labelling phases of our experiments. These subtleties, intrinsic to the glycoprotein landscape, make the HDX-MS investigation a complex undertaking, demanding meticulous attention to detail and rigorous data interpretation[104].

In pursuit of this, research has been directed towards the development of methodologies for enzymatically cleaving glycans while preserving biologically relevant native data[111]. Given the conventional experimental HDX-MS framework, this requires the identification of an enzyme endowed with the capacity to maintain its activity under exceedingly rigorous conditions, characterised by acidic pH levels and the presence of strong denaturants such as TCEP and Urea, especially due to the fact that glycoproteins are often extracellular, thus often containing a large number of disulphide bonds[115].

5.2.1 Inherent Heterogeneity

Protein glycosylation, an omnipresent phenomenon, plays a central role in numerous proteoforms by intricately linking complex metabolic pathways. Beyond its fundamental involvement in metabolic processes, glycosylation serves as a cornerstone in preserving functional and biological structures. Its ubiquity underscores its importance in modulating various cellular functions and underscores its essentiality in maintaining cellular homeostasis[116].

Therefore, it is imperative to expand our understanding of these complex glycosylation mechanisms, especially given their association with various diseases. Among these conditions are cancers, inflammatory diseases, and metabolic disorders. This underscores the importance of delving deeper into the intricate workings of

glycosylation processes discussed previously. By elucidating the molecular intricacies underlying these associations, we can pave the way for novel therapeutic interventions and targeted treatments to address these debilitating health conditions[116].

The intricate composition of glycans presents formidable obstacles during their analysis using HDX methodologies. This complexity manifests itself in various ways, such as signal depletion in peptides and the obscuring of identifiable features because of overlapping peaks. Moreover, the inherent heterogeneity of glycans introduces challenges in accurately assessing the deuterium uptake values, since the sugars themselves may undergo uptake, leading to skewed results. As researchers navigate these hurdles, the demand for innovative strategies to effectively manage glycan-related issues within the HDX environment continues to escalate, driving a surge of interest in this evolving field of study.

In the realm of mass spectrometry analysis of glycoproteomic complexes, a common practice involves cleaving the glycans to enable separate analysis using LC. This approach allows for the elucidation of the glycan profile, which is crucial for understanding the composition of these complexes. However, this cleavage process alters the inherent structure of the proteome, making it less biologically relevant for HDX studies. Maintaining the native structure of the proteome is paramount in HDX analysis to ensure that acquired information retains its biological significance.

To address this challenge, a method is needed to cleave the glycans after labelling and after quenching in HDX experiments. This would enable researchers to obtain information that is more relevant to draw meaningful conclusions. However, implementing such a method requires exploring alternatives to the typical proteomic approach, since the acidic pH of the quench buffer used in HDX experiments may affect enzymatic activity. As researchers seek solutions to this dilemma, innovative approaches that preserve the integrity of the proteome while enabling protein analysis post-quenching are essential for advancing our understanding of glycoproteomic complexes.

5.3 pH Problem

As outlined previously, the conventional approach to glycan analysis is not viable for obtaining relevant structural data in HDX studies due to pH-related problems. Consequently, researchers have explored alternative strategies to address this challenge.

One promising avenue involves investigating acidophiles as potential candidates for addressing the pH problem. Several research groups have conducted screenings to identify acidophilic organisms and assess their suitability for HDX experiments. Factors such as enzyme activity, specificity, and availability have been carefully evaluated to determine the most suitable candidates for further investigation. This meticulous selection process ensures that the chosen acidophiles are well-equipped to overcome the pH-related obstacles encountered in HDX studies, paving the way for innovative solutions in glycan analysis within this context.

Other work has previously been done to test different enzymes and their suitability to work within the typical HDX experimental setup. Some were more suited than others.Work started with looking at PNGase A. This work by Jensen et al. found that the differences in uptake compared to the cleaved and uncleaved areas were related to the glycosylations themselves, that is, the uptake from the glycan moiety rather than from the backbone after cleavage[112]. This led others to look at more acidophiles from this line of enzymes to find better suited enzymes that could potentially be attached onto a column for online work. This initial work was done in solution and requires a more hands-on experimental setup.

5.3.1 PNGase F

Peptide:N-glycosidase F, commonly known as PNGase F, is commonly used in glycoproteomics[117]. It is a popular amidase that works by cleaving between the innermost GlcNAc and asparagine residues of high-mannose, hybrid and complex oligosaccharides of N-linked glycoproteins and glycopeptides.

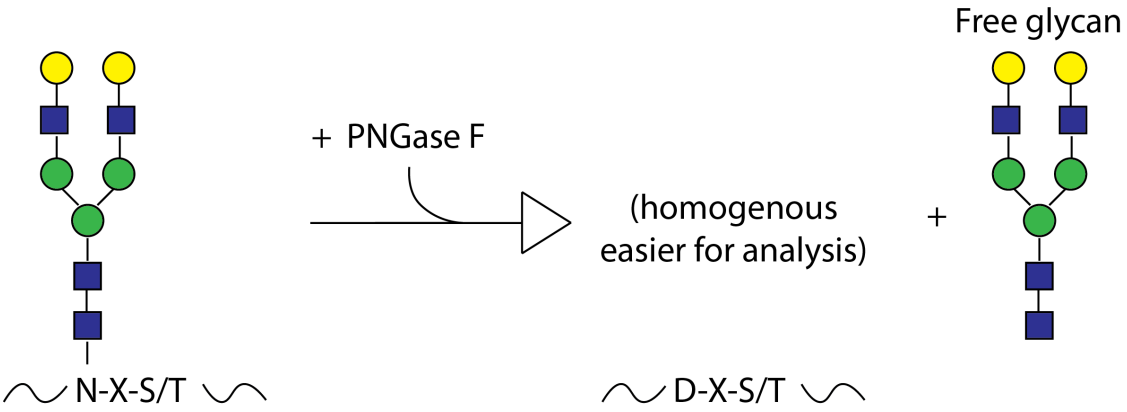


Figure 5.1: Typical Proteomic approach for cleaving glycans. Monosaccharide colour and shape align with figure C.1 in appendix C.

PNGase F maintains activity in the pH range of 6-9.5, which allows great use when studying glycoproteins by native mass spectrometry. This is a popular approach for glycomics as a result of the working activity of PNGase F being around the typical spraying pH in native mass spectrometry.

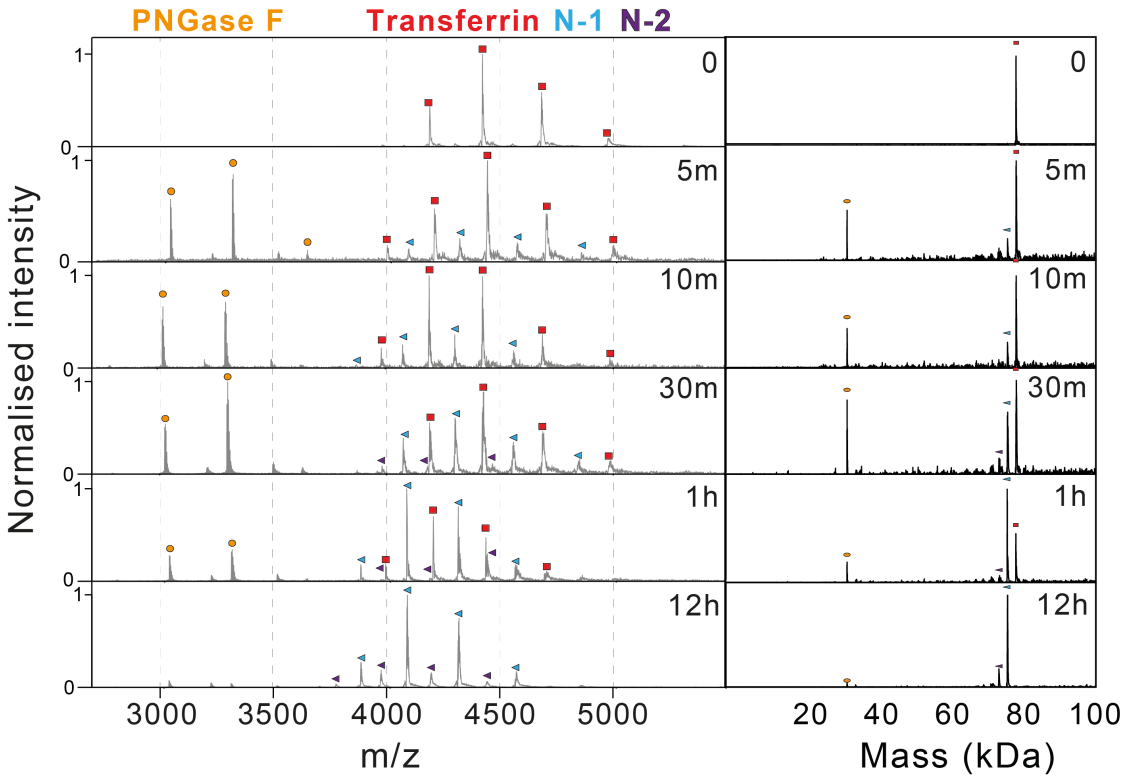


Figure 5.2: Time series of PNGase F cleaving apo Transferrin at pH 8. Mass Shifts in line with known glycan masses of N-1 and N-2 for Transferrin peaks.This can be seen clearer with the help of the deconvoluted mass histograms.

Shown here in figure 5.2 is a time series of the raw mass spectrometry data collected for Transferrin mixed with PNGase F (10:1) at pH 8, which is close to its known peak activity of pH 8.5. From the convoluted mass spectrum to the right, we observe that the initial mass of around 79 kDa decreases over time in stepwise increments of approximately 2.5 kDa. This pattern highlights the formation of N-1 and N-2 glycoforms (representing the loss of one and two glycans, respectively). However, at this pH, we did not achieve a fully deglycosylated state (N-2) for all transferrin molecules within a 12-hour period.

The high pH required for PNGase F activity, combined with its slow enzymatic rate under these conditions, poses significant challenges for HDX experiments. Typically, HDX quenching is performed under conditions that are not compatible with such high pH levels, making PNGase F an unattractive option for HDX studies.

5.3.2 Enzyme Inactivity

MS was used to conduct an investigation of the enzymatic activity of PNGase F across a spectrum of pH values. A notable decline in enzymatic activity was observed at pH 6, aligning with expectations. This observation would deem PNGase F an unsuitable enzymatic candidate within the HDX-MS workflow. Specifically, its diminished cleavage capability at post-quench pH values would therefore prevent the enzyme from effectively cleaving glycans.

It is imperative to note that the removal of glycoproteins before the quenching step would impart a transformation to the native structure of the proteins. Consequently, the ensuing data acquired would lack the biological relevance that one would like. Providing minimally valuable structural insights to researchers and analysts. In essence, the application of PNGase F in this context would compromise the fidelity of structural information retrieval, thereby necessitating the pursuit of alternative enzymatic solutions within the HDX-MS methodology.

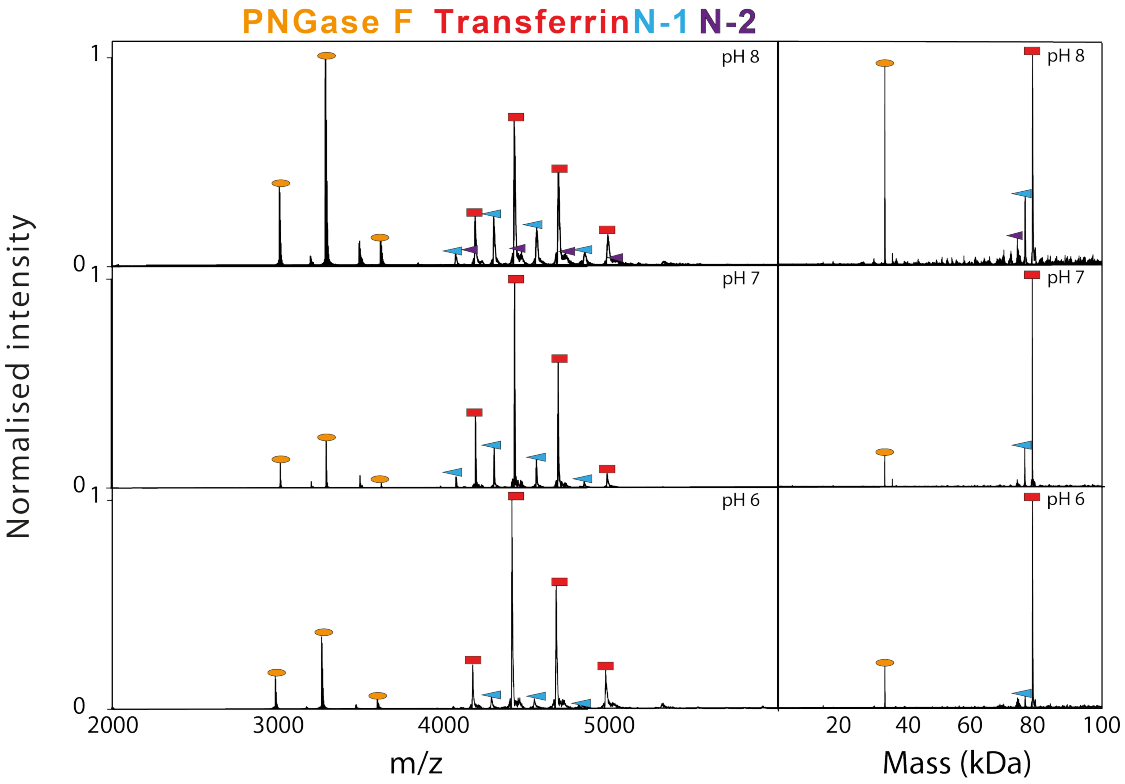


Figure 5.3: PNGase F activity assay at differing pH. Raw Mass spectrometry data supplemented with the deconvoluted spectra. After 10 minute incubation at 37 ^{o}C.

With its peak activity occurring at pH 8.5, the mass spectrometry data clearly indicates that PNGase F's activity diminishes as the pH deviates from this optimal point. Its activity is nearly negligible at pH 6, as evidenced by the minute abundance of the cleaved N-1 or N-2 peaks in the raw data. This further underscores its unsuitability as an enzyme in post-quench conditions.

5.3.3 Acidophiles - Potential Answers

A number of acidophiles have been examined as potential solutions to the glyan problem with HDX. As stated above, this work has become ever more important to delve into more complex protein problems and in investigating what was once very difficult to do, extending the scope of this already extensive and dynamic technique.

Through work done in 2022 by Guo et al.[114]. PNGase Rc was determined to be the best solution to answer the question as to which enzyme is best suited for cleaving glycans post-quench process. This is due to a lower optimal pH

value for activity and also a better resistance to additives typically used while quenching, such as TCEP and Urea.

Other enzymes that were considered potential candidates in the preliminary work were PNGase A, PNGase H^+ and PNGase Dj[118]. These were the first enzymes found to be candidates for what we were looking for. All of which maintain some form of activity around the low pH of the quench buffer.

5.4 Development of PNGase Rc Column

PNGase Rc was successfully expressed in *E. coli*, resulting in a high yield and purity [114]. Then it was immobilised onto POROSAL beads through its Lys residues. A review of the AlphaFold-predicted structure for PNGase Rc showed that all Lys residues are positioned away from the anticipated active site, drawing comparisons with the active sites in PNGase F and A. The degree of immobilisation, represented as ψ, was evaluated by measuring the concentration of PNGase Rc in the supernatant after the attachment of the beads. This was established using a calibration curve based on the UV 280 nm peak area from PNGase Rc standard dilutions, analysed using an HPLC C4 column. An enzyme-to-bead (E:B) ratio of 1:10 yielded roughly 39% in immobilisation efficiency. In contrast, a 1:20 E:B ratio produced striking 95%. This suggests that while the 1:20 E:B ratio facilitates almost complete enzyme attachment to beads, it might not saturate the entire available bead surface with the enzyme [119].

5.4.1 Addition to the Typical Workflow

The developed column is used downstream in the typical experimental setup that was previously seen in chapter 1 of the Quench/Digestion step. This allows for an increased level of deglycosylation of peptides that previously had glycan moieties attached to them. This also means that the labelling stage occurs prior to the glycan cleavage, so the data obtained are of relevance biologically, as it was untouched in its native state before quenching and slowing the exchange process.

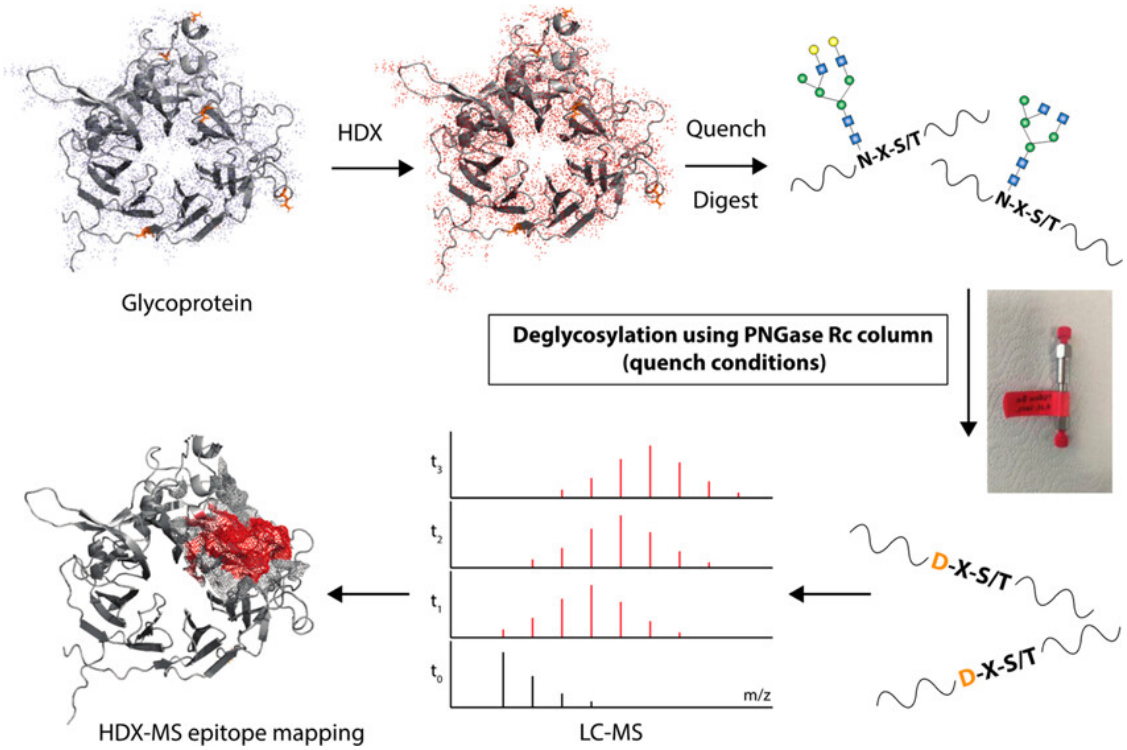


Figure 5.4: Adapted workflow of the typical HDX-MS experimental workflow. This visual representation portrays the enzymatic cleavage of glycans, which occurs sequentially downstream of the quenching process in harsh conditions.

The asparagine (N) residue that the glycan is attached to leaves behind once it is cleaved undergoes a deamidation step and leaves behind an aspartate (D) residue.

5.4.2 Activity and Reproducibility

The performance of the enzymatic column remained consistently robust. Repetitive usage continued for an extended observational period spanning more than 100 days. This remarkable resilience persisted even in the face of formidable challenges posed by the conventional quenching additives used in HDX-MS, including TCEP and Urea. Such sustained activity in the presence of these quenching agents magnifies the exceptional durability and reliability of the column, reaffirming its suitability for long and rigorous analytical aspirations [119].

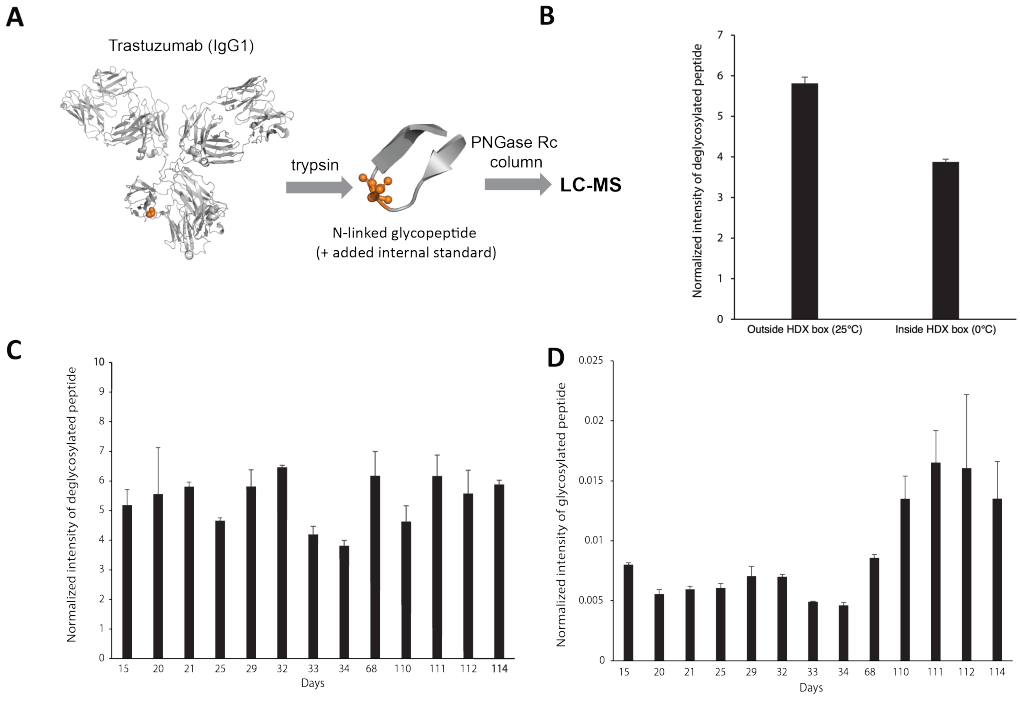


Figure 5.5: Evaluation of PNGase columns based on temperature-dependent trials. (A) Activity of the PNGase Rc column was gauged using an LC-MS assay by tracking intensities of the IgG1 N-linked glycopeptide EEQYNSTYR, set against the Angiotensin II internal standard. (B) Comparison of the column's activity when situated inside (0°C) and outside the HDX box (20°C) with a flow rate of 200µL/min. (C,D) Assessing the long-term reusability of the PNGase Rc column over 114 days using the LC-MS assay. Throughout this period, the column underwent routine usage in HDX-MS tests, coming into contact with TCEP and Urea (see table 5.1). Between day 0 to day 15, the column underwent an equilibration phase.

The durability and efficiency of the PNGase Rc column were evaluated using a specialised LC-MS assay. This assay used tryptic glycopeptides derived from IgG1 Trastuzumab as a substrate. To discern the activity of the column at various temperatures, assessments were performed both at 0 ° C and at ambient temperature. Interestingly, at room temperature, the deglycosylation efficiency soared to 99%. However, at 0°C, there was a notable decrease, with the efficiency dipping to 72%.

$$\text{Deglycosylation efficiency } (\%) = \left(\frac{I_0 - I}{I_0}\right) \times 100\% \tag{5.1}$$

It should be emphasised that throughout the observation period, the activity of the column exhibited remarkable consistency. This consistency underscores

not only the reliability of the column, but also its resilience and durability. This endurance becomes even more commendable considering the routine exposure of the column to rigorous HDX-MS processes, which often involve the use of aggressive additives such as TCEP and Urea [119].

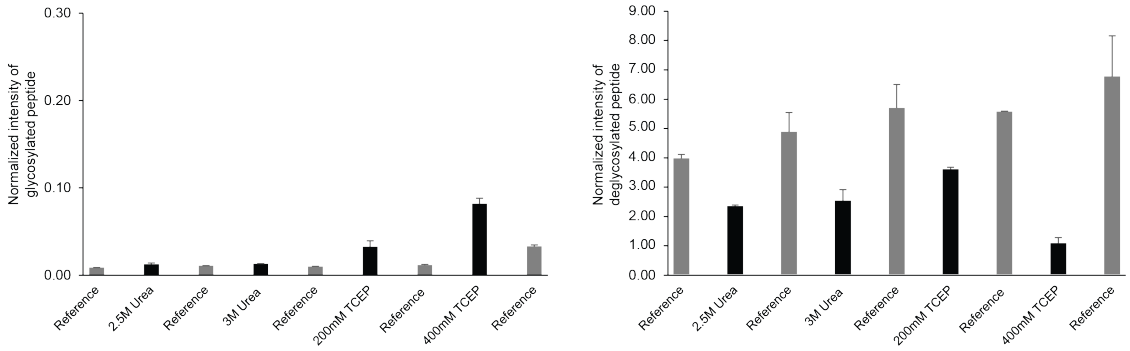


Figure 5.6: Activity of PNGase Column in the presence of typical HDX-MS additives. The abundances of the glycopeptide EEQYNSTYR were meticulously measured and compared. For accuracy and consistency in this evaluation, the abundances were normalised using an internal standard as a reference point.

As illustrated in Figure 5.6, the PNGase Rc column demonstrates exceptional performance, even under the rigorous influence of both TCEP and Urea. The resilience of the PNGase Rc column to these chaotropic and reducing agents is particularly noteworthy because they are essential for generating optimal HDX-MS results. In particular, the column effectively performed thorough deglycosylation across various glycoproteins in HDX-MS tests, even when quench buffers with significant concentrations of TCEP and Urea were employed. The impressive reusability of the column, combined with its commendable tolerance to Urea and TCEP, positions the PNGase Rc column as a valuable tool, enhancing the precision and efficiency of HDX-MS analysis for glycoproteins [119].

5.4.3 Methods

Activity and reusability Reusability and functionality of the PNGase Rc column were evaluated using a LC-MS assay utilising tryptic glycopeptides from IgG1 (Trastuzumab) as substrate. Trastuzumab (TZ) has an N-linked glycosylation at position N297 on the heavy chain Fc region, and the performance of the

PNGase Rc column was evaluated by tracking the degree of deglycosylation of this specific glycopeptide.

To explore the performance of the PNGase Rc column under varying temperature conditions, we assessed its enzymatic activity at both 0°C and room temperature. The PNGase Rc column exhibited a moderate decrease in activity (33.3% reduction) at 0°C. Despite this decline in activity when exposed to lower temperatures, it is important to note that the column still achieved a high level of deglycosylation, making it suitable for an HDX-MS experiment if needed.

Subsequently, in order to evaluate the long-term endurance of the PNGase Rc column, we consistently monitored the intensity of both the glycosylated and deglycosylated forms of the glycopeptide via LC-MS for a duration of 114 days. Throughout this extended period, the column was regularly used in numerous HDX-MS experiments, which involved exposure to TCEP and Urea. In sum, the column consistently demonstrated strong performance and maintained its stability.

HDX-MS of Transferrin The study began with the preparation of the apo-human transferrin sample, obtained from Sigma-Aldrich. It was methodically diluted to achieve a 10µM concentration using an equilibration buffer, specifically PBS at pH 7.4.

For labelling purposes, 5µL of this protein sample was further diluted to a cumulative volume of 50µL. This was achieved by integrating a labelling buffer, a concoction of PBS formulated in D_2O, with a pD value of 7.4 (pH read 7.0), resulting in a meticulously controlled 90% D_2O milieu. The labelling was executed over an array of time intervals.

Following the predetermined durations, the labelling reaction was immediately quenched. This was carried out through a 1:1 integration of an ice-cold quench buffer, comprising 50mM KH_2PO_4 and 4M Urea, calibrated at a pH of 2.

The subsequent phase involved online protein digestion, facilitated by the Waters Enzymate BEH pepsin column, with a uniform temperature of 20 ° C. After digestion, the peptides were momentarily trapped in a Waters BEH C18 VanGuard precolumn. This stage lasted 3 minutes, with a consistent flow rate of 100 µL/min,

employing buffer A, characterised by 0.1% formic acid at an approximate pH of 2.5. Subsequent to this entrapment, the peptides were transferred to a Waters BEH C-18 analytical column.

The elution procedure was orchestrated through a meticulously linear gradient, employing buffer B, which consists of 0.1% formic acid suspended in acetonitrile. This solution hovered around a pH value of 2.5. The elution maintained a flow rate of 40 µL/min. Importantly, to ensure precision and minimise back exchange, all trapping and peptide separation operations were carried out at an ambient temperature of approximately 0°C.

Data acquisition remained consistent across states, regardless of the presence or absence of the PNGase enzymatic column, and was gathered via the MSE workflow. The rigour of the experiment was amplified by obtaining all data points in triplicate. Furthermore, the MS data were subjected to an independent calibration using NaI, and the collected MS data were enhanced with a lock mass correction facilitated by LeuEnk.

The final data interpretation involved the assignment of peptides through the sophisticated ProteinLynx Global Server (PLGS, Waters Corporation, UK). Furthermore, the isotope uptake corresponding to the identified peptides was meticulously determined using DynamX v3.0 software.

The other proteins then tested followed a similar workflow. Differences in the preparation conditions for the sample and the quench conditions were observed which can be found in Table 5.1.

5.5 Results

5.5.1 Choice of Protein

A diverse set of proteins, ranging from simple to complex glycoproteins, was examined. Each protein had a unique number of glycosites. Transferrin, for example, had a relatively modest number of glycosites, with only two sites throughout its primary amino acid sequence. In contrast, C-met was a considerably larger protein, possessing three times the glycosites of Transferrin. This analysis underscored the

potential of the column to shed light on previously inaccessible areas, enhancing the peptide coverage of all the proteins evaluated.

5.5.2 HDX data post the addition of the PNGase Column

With the addition of the PNGase column, previously uncovered areas of transferrin, particularly those that spanned the glycan sites, were identified. This increased the overall coverage of Transferrin by 7.5%, from 65.5% to 73%. In particular, seven additional peptides were found at the two glycan sites. This enhancement allows for a more comprehensive analysis of biologically relevant regions of the protein that were previously undetected, while still preserving information from the native glycosylated state due to the post-quench cleavage.

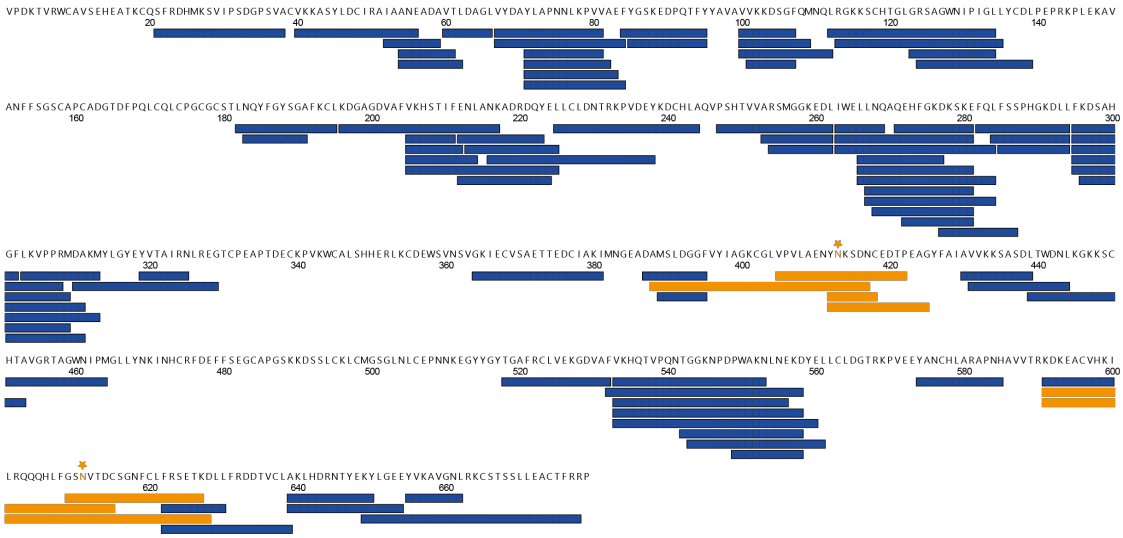


Figure 5.7: Transferrin sequence coverage map comparison. Blue peptides found without use of PNGase Column. Orange peptides are additional peptides found post cleavage. Orange stars mark point of glycosylation in the primary amino acid sequence

The coverage map in Figure 5.7 illustrates the peptides identified before and after cleavage. Peptides identified before treatment are coloured blue, and these were also detected after cleavage. The orange peptides represent the additional sequences discovered after the cleavage of glycoforms using the PNGase column. These orange peptides span the glycan sites, which are marked on the primary amino acid sequence with an orange star to indicate their positions.

To provide a visual aid, this colouring has been mapped onto the Transferrin PDB structure (2HAU), shown below in Figure 5.8. This mapping visually highlights the enhanced coverage and the specific regions where additional peptides were identified, highlighting the significance of the glycan sites.

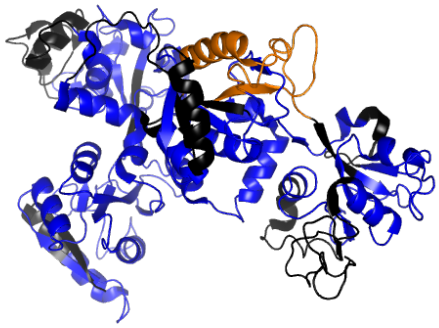


Figure 5.8: Transferrin PDB file(2HAU) coloured in line with the coverage map in Figure 5.7.Black additionally here indicating areas of no coverage.

The coloured PDB structure enables us to visualise the newly identified regions of the 1-D primary sequence within the higher-order protein structure. This enhanced analysis capability particularly highlights areas around the glycan sites, which are often of significant functional importance. The ability to now analyse these previously inaccessible regions allows us to extract relevant data, offering new insights into the roles and functions of the glycoforms. This advancement opens up new avenues for research, enhancing our understanding of the protein's structural and functional dynamics.

5.5.2.1 Discerning between N and O-linked Glycans

A persistent challenge lies in the solution of O-linked glycans. The immense heterogeneity of O-glycans, resulting from the absence of a consistent glycan core, complicates their analysis[120].

Osteopontin represents a fairly complex structure to use and test the PNGase column. Osteopontin consists of a number of O-linked glycans and just two N-linked glycans [121]. Osteopontin has been identified in various human tissues,

including bone, kidney, macrophages, urinary stones, and human milk. Recent studies have increasingly highlighted the crucial role of this protein. In particular, a recent study suggests that osteopontin significantly influences cell surface binding and bone catabolism[121].

However, the PNGase Rc column retains the ability to discern between a mix of N- and O-linked sites. Its application notably improved the coverage of Osteopontin from 75.2% to 81.5%, identifying seven additional peptides spanning two N-Linked sites.

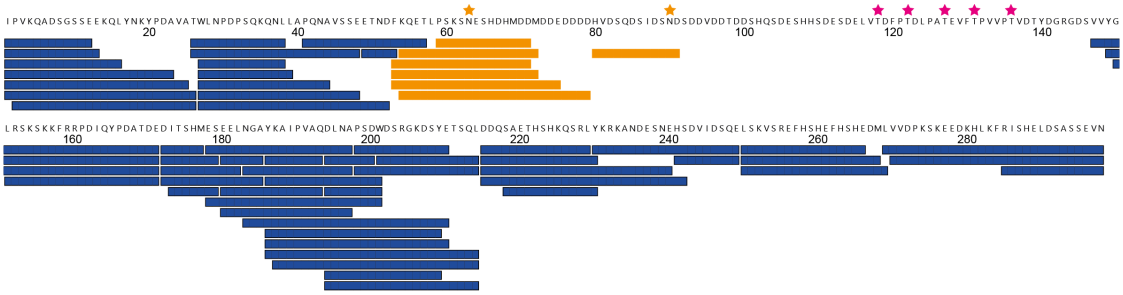


Figure 5.9: Osteopontin Sequence coverage with and without the use of PNGase column. As above, blue peptides are peptides identified without the use of the PNGase Rc column.Orange are additional peptides identified after the use of the PNGase column. glycosites (N-linked) are indicated by orange stars. The pink stars indicate sites of O-linked glycans.

As in the colour scheme used in the Transferrin coverage map in Figure5.7, we have applied the same scheme to the Osteopontin coverage map. N-linked glycans are marked with an orange star, and the additional peptides identified in these regions are also coloured orange. Additionally, we designated the positions of the known O-linked glycans with a pink star within the primary amino acid sequence. No peptides were identified in the O-linked glycan regions, which is consistent with expectations because PNGase Rc does not cleave O-linked glycans, so we expect heterogeneity to be maintained and difficulty in identification and assignment to be sustained.

5.5.3 Column Performance

The PNGase column was tested on numerous different Glycoproteins, all of differing complexity. In all investigated systems, we observed an increase in both the number of peptides identified and the coverage obtained across the respective sequences[119].

This empirical observation underscores the versatility and efficacy of the PNGase column in the context of glycoprotein analysis. Its ability to improve both the identification of peptides and the coverage of the sequence in a wide spectrum of glycoprotein structures further substantiates its utility as an indispensable tool in the field of structural proteomic studies[119].

Glycoprotein	**Sequence Coverage without PNGase Rc column (%)**	**No. of peptides without PNGase Rc column**	**Sequence coverage with PNGase Rc column (%)**	**No. of peptides with PNGase Rc column**	**No. of N-linked glycosites**	**No. of deglycosylated N-linked glycosites**	**Quench Conditions (after 1:1 dilution)**
C-met	91	721	98	755	6	6	125mM TCEP and 1.7M Urea, pH 3
Haptoglobin 1-1 (beta chain)	84.5	66	98	81	4	4	125mM and 2M Urea, pH 2.3
Transferrin	65.5	86	73	93	2	2	100mM TCEP and 2M Urea, pH 2
VISTA	73	50	97	110	6	6	200mM TCEP and 2M Urea, pH 2.6

Table 5.1: Table of different systems tested using the PNGase Rc column. Showing number of peptides found and the sequence coverage obtained with and with out the usage of the enzymatic column.

5.5.4 Individual Protein optimisation

The optimisation of the experimental setup for each individual protein was an essential prerequisite, mirroring the typical procedure inherent to any HDX-MS

investigation. The column in use exhibited the capacity to accommodate customary operational flow rates, which extended up to 200 uL/min. To accommodate the addition of an extra column and connecting tubing, thus introducing an unavoidable extension in dead volume space, the trapping time was extended. This deliberate extension in trapping time allowed for the seamless flow of samples through the augmented system. But how would that affect the back-exchange of the identified peptides across the system?

5.5.5 Back exchange

The dead volume added in the system left a question as to how that would affect the back exchange of each peptide across the entirety of the experimental data set. The back exchange was found to be affected only minimally, which means that the columns used were capable of effectively cleaving and obtaining useful data.

$$\text{Back exchange } (\%) = \left(1 - \frac{m_{100} - m_0}{N \times D_{\text{frac}}}\right) \times 100\% \tag{5.2}$$

The equation above was used to calculate the back-exchange of each peptide, allowing for an average calculated value across the entire data set. 'm_{100}' here is the centroid mass value of the data 100%, 'm_0' is the centroid mass a the time point 't_0', N is the number of exchangable amide hydrogens in the peptide in question and 'D_{frac}' is the fractional deuterium environment within the labelling buffer[44].

When used in conjugate of a typical HDX-MS set up, back exchange was found to increase only by around 2% in all protein systems tested[119].

5.6 Discussion

In this study, we present the novel PNGase Rc column as an effective tool for the online removal of N-linked glycans during HDX-MS investigations of glycoproteins. The seamless integration of the PNGase Rc column into the standard HDX-MS setup proved to be straightforward, requiring its placement downstream from the pepsin column to facilitate efficient online digestion and deglycosylation. Our research included the analysis of four complex glycoproteins: Hp 1–1, VISTA, Transferrin, and

c-Met, all of which required strict quench conditions, including high concentrations of Urea and TCEP. Through incorporation of the PNGase Rc column into our experimental protocol, we observed a marked improvement in the sequence coverage, enabling the measurement of HDX in previously inaccessible glycosylated regions.

Furthermore, our study highlights the versatility of the PNGase Rc column in elucidating epitopes on challenging glycoprotein drug targets, such as c-Met. Notably, control experiments conducted to assess the impact of employing the PNGase Rc column during HDX-MS analysis revealed minimal increases in deuterium label loss (back-exchange).

Moreover, our investigation demonstrated the robustness of the PNGase Rc column, even in the presence of common quench buffer additives such as TCEP and Urea. Although a moderate reduction in activity was observed at elevated concentrations of TCEP and Urea, this reduction was shown to be fully reversible. Furthermore, the PNGase Rc column consistently maintained its performance in multiple sample injections containing TCEP and Urea, exhibiting stability for more than 100 days.

In the future, we anticipate that the seamless integration of the PNGase Rc column into conventional HDX-MS setups will significantly enhance the analytical capabilities for studying the conformational dynamics and ligand interactions of challenging glycosylated proteins. This advancement holds promise in unlocking insight into numerous attractive and previously unexplored drug targets. Furthermore, the versatility of the PNGase Rc column extends beyond HDX-MS experiments to other bottom-up proteomic workflows requiring rapid online deglycosylation of peptides or proteins prior to LC separation and MS analysis. Notably, the commercial availability of the PNGase Rc column offers a convenient solution for researchers looking to improve the effectiveness of their HDX-MS investigations without compromising reliability or efficiency.

"I have not failed. I've just found 10,000 ways that won't work."

— Thomas Edison

6 Overview

Contents

6.1 Overview/Summary 117
6.1.1 SAUSC 117
6.1.2 Ligand binding HDX 118
6.1.3 Dynamics of SHIP constructs 119
6.1.4 Dealing with glycoforms 120
6.1.5 Conclusion 121

6.1 Overview/Summary

6.1.1 SAUSC

Efforts were made in the development of SAUSC to build on and maximise the works started by Lau et al., in the development of Deuteros [51]. We made a concise effort to take the statistics used in Deuteros, which were put forward in the work of Hageman and Wise through the conjugate hybrid test [59] to help present the data obtained throughout our experiments. We altered and played with the premise of the plots presented, particularly the volcano plots presented in Deuteros [51]. The way in which we did this allows us to alter the axes and the gradient-based colour scheme to give us a consistent output through both plots and 3D representations. This allows for a complete analysis and aids us in our

endeavours to answer the biological question at hand. We are able to edit the figures and manipulate the data in such a way that we can present our findings clearly. These plots act as aids to show and present differences that have occurred in our analysis and have been found to be statistically relevant. Enhance the speed with which we can analyse and extract significant differences.

In summary, through our efforts, SAUSC can be used as a simplistic approach to aid us not just in analysing but also in understanding the differences made upon ligation or any other stimuli. This understanding gives us an indication on a mechanistic level through its output of in-depth figures and visual aids of how the protein responds to different stimuli and how the dynamics alter throughout this process.

6.1.2 Ligand binding HDX

Using IPNS, a non-heme iron-dependent oxidase, we investigated the ligand-binding process to gain biologically relevant insights into protein dynamics. The active site of IPNS features a metallic iron core, which is essential to convert ACV into IPN. For this experiment, our focus was on the structural dynamic changes associated with ligand binding, rather than product formation. To achieve this, we substituted Fe(II) with Zn(II) and Cd(II), effectively mimicking ligand binding while preventing product turnover.

Residues known to be critical for the catalytic activity of IPNS in other studies demonstrated significant protection when exposed to the metal cation and ACV. This indicates that the binding of the metal and the tripeptide to the active site induces alterations in the hydrogen-bonding network. The extent of this protective effect is evident in Figures 3.5 and 3.6. Our findings on the structural differences align with previous studies, confirming the specific binding mode of the metal at the active site of IPNS. These results provide valuable dynamic insights into the mechanistic process of metal and tripeptide binding, enhancing our understanding of ligand interactions within the enzyme.

SAUSC was used to analyse the data and generate figures that visually represent our findings, highlighting the significant differences observed at the structural level. These visual aids were crucial in clearly communicating how these differences impact the protein's conformation and dynamics, allowing us to better understand the effects of ligand binding. By integrating these analytical tools, we sought to gain biologically relevant insights into how binding interactions influence the structure and function of the enzyme, deepening our knowledge of the underlying mechanisms at play.

6.1.3 Dynamics of SHIP constructs

To investigate the regulatory role of the C2 domain in the flexibility of the phosphatase domain in SHIP proteins, we performed detailed Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) analyses on SHIP1 and SHIP2 constructs, both with and without the C2 domain [78]. HDX-MS enabled us to examine dynamic structural changes by measuring deuterium uptake (DU) over time, providing insights into structural integrity and functional impact of the C2 domain.

For SHIP1, DU was measured at 1, 10, and 30 minutes, while for SHIP2, measurements were taken at 1, 2, 5, and 10 minutes to capture varying stages of hydrogen exchange. The coverage of the phosphatase domain was extensive, with 83 common peptides identified in SHIP1 (88% coverage) and 66 in SHIP2 (86% coverage), allowing a high resolution analysis of structural changes.

Comparative HDX-MS data revealed that in both SHIP1 and SHIP2, the absence of the C2 domain led to increased flexibility and solvent accessibility in key regions, particularly around L3 and L4, which are believed to be critically involved in the interaction and stability of the phosphatase domain. These differences suggest that the C2 domain plays a critical role in maintaining structural integrity and regulating the flexibility of these regions.

The findings showed a general trend of increased deuterium uptake in shorter constructs lacking the C2 domain, indicating that the C2 domain contributes to the structural and functional regulation of the phosphatase domain. This was particularly evident in SHIP2, where regions typically buried at the phosphatase-C2

interface exhibited higher uptake when the C2 domain was absent. The influence of the C2 domain extends beyond structural stabilisation, affecting the overall activity of protein complexes, as demonstrated by reduced activity in the absence of the C2 domain [78].

In general, this study provides valuable information on the dynamic regulatory effects of the C2 domain on SHIP proteins, highlighting its role in structural stability and functional modulation. These results contribute to a broader understanding of protein-protein interactions and the biological significance of protein domain architecture in regulatory mechanisms.

6.1.4 Dealing with glycoforms

PNGase Rc was successfully expressed in E. coli with high yield and purity, then immobilised onto POROSAL beads through Lys residues, ensuring that active sites remained accessible. The immobilised PNGase Rc column was integrated downstream of the Quench/Digestion step in HDX-MS workflows, significantly enhancing the deglycosylation of glycopeptides while preserving biological relevance, as labelling occurred prior to glycan cleavage.

The column demonstrated exceptional stability and reusability, maintaining activity for 100 days despite exposure to harsh quenching agents such as TCEP and urea, confirming its robustness and suitability for rigorous HDX-MS applications [119]. As illustrated in Figure 5.6, the PNGase Rc column efficiently deglycosylated various glycoproteins under challenging conditions, including high concentrations of urea and TCEP. Its use increased the sequence coverage of Transferrin from 65.5% to 73%, as well as great increases for other complexes, revealing previously undetected glycosylated regions and improving the overall precision of the analysis.

The integration of the PNGase Rc column into HDX-MS workflows marks a significant advancement for us as researchers because it allows efficient, rapid, and consistent deglycosylation of glycoproteins under stringent experimental conditions. This allows us to probe previously inaccessible glycosylated regions, enhancing sequence coverage and providing a clearer understanding of protein structure and

dynamics without disrupting native states prior to analysis. The ability to maintain enzyme activity for extended periods, even in the presence of challenging quench agents, adds robustness and reliability to HDX-MS studies.

These advances facilitate deeper insights into the conformational changes and ligand interactions of glycosylated proteins, expanding our capacity to investigate complex biological questions and explore new drug targets. This development could lead to breakthroughs in understanding protein function, disease mechanisms, and therapeutic interventions, particularly for glycoproteins that were previously difficult to analyse comprehensively.

6.1.5 Conclusion

In conclusion, these advances significantly improve our ability to analyse protein structure and function, especially within challenging biological contexts. SAUSC, developed by building on earlier statistical frameworks - such as Deuteros - streamlines the visualisation of complex datasets, allowing faster and clearer identification of biologically meaningful differences in protein behaviour.

The ligand binding study with IPNS provided valuable insights into the structural changes triggered by metal and peptide interactions, deepening our understanding of the active site dynamics without confounding product formation. The use of alternative metals in this set-up offered a fresh perspective on the binding mechanisms that are key to enzyme functionality.

HDX-MS analysis of SHIP proteins revealed how domain interactions, particularly those involving the C2 domain, influence both structural flexibility and functional activity. These findings underscore the importance of the protein domain architecture in modulating protein stability and enzymatic performance, particularly in regulatory processes.

Finally, the introduction of the immobilised PNGase Rc column into HDX-MS workflows marks a practical breakthrough that improves the analysis of glycoproteins by ensuring efficient deglycosylation under harsh conditions. This development allows for a broader sequence coverage and deeper investigation into previously

inaccessible glycosylated regions, offering new opportunities to study protein conformational dynamics and interactions.

Together, these innovations improve the precision and depth of protein analysis, providing critical tools to explore complex biological mechanisms, and facilitating future breakthroughs in therapeutic research.

Appendices

A

Investigating the affects of metal binding to Iso-Penicillin-N-synthase

Contents

A.1 Raw Uptake Plots 125
A.1.1 IPNS vs IPNS + Zn(II) 126
A.1.2 IPNS vs IPNS + Zn(II) and ACV 128

A.1 Raw Uptake Plots

Throughout the investigations presented in Chapter 3, we conducted an in-depth analysis of the complex enzyme IPNS. Our focus was on understanding the dynamic structural changes that occur upon binding of a metal ion and the ACV tripeptide at the active site. By examining these interactions, we aim to shed light on the mechanisms that drive these conformational changes. This analysis relied heavily on the comparison of HDX-MS data, which provided insights into the differences in protein behaviour under various conditions. Quantitative evaluation of these differences was aided by statistical tools that improved the reliability and interpretability of the data.

In this section, we present the raw data obtained from the experiments. These raw uptake curves serve as the foundation for all subsequent analyses, capturing the deuterium incorporation patterns over time for individual peptides under different conditions. By visualising this data, we can observe the direct impact of ligand binding on the structural dynamics of IPNS, allowing for a comprehensive understanding of its behaviour.

A.1.1 IPNS vs IPNS + Zn(II)

The first set of comparisons investigates the changes in deuterium uptake between IPNS in its default state and IPNS bound to Zn(II).

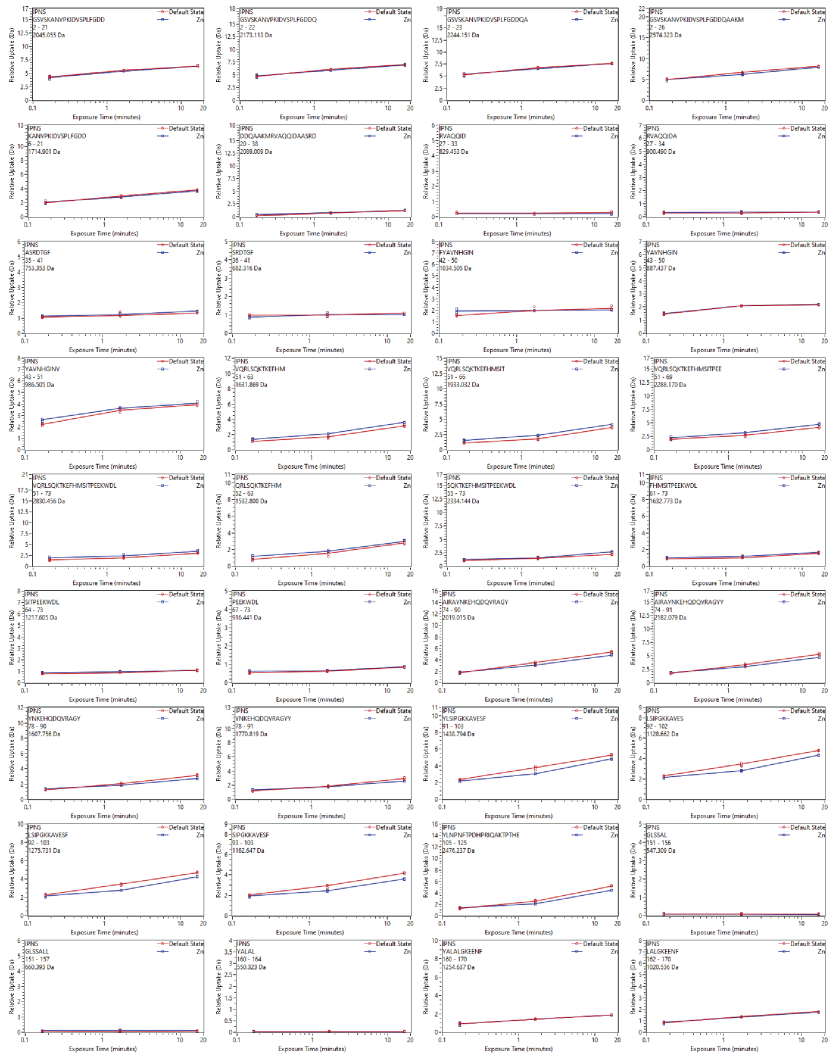


Figure A.1: Raw uptake plots of peptides 1-36 of IPNS and IPNS + Zn(II). IPNS in the plots is labelled as Default state, and IPNS + Zn(II) is labelled Zn.

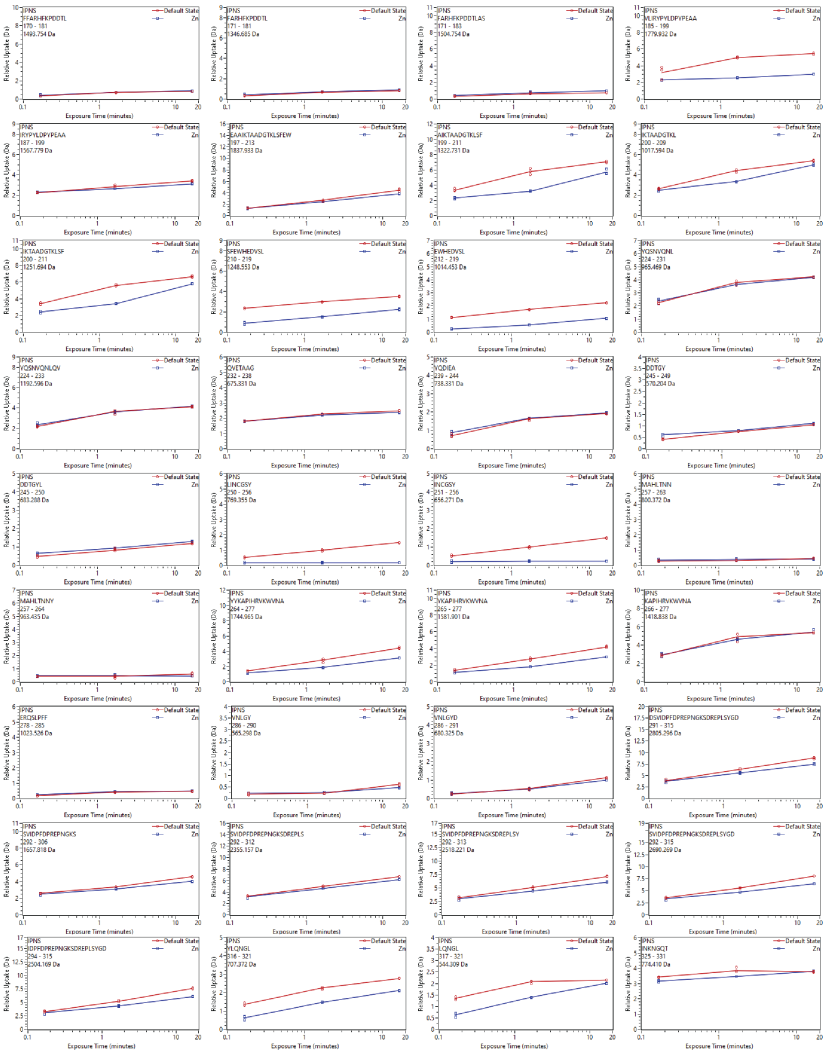


Figure A.2: Raw uptake plots of peptides 37-72 of IPNS and IPNS + Zn(II). IPNS in the plots is labelled as Default state, and IPNS + Zn(II) is labelled Zn.

Figures A.1 and A.2 display raw data for the deuterium uptake patterns of peptides that span residues 1-72 in the presence and absence of Zn(II). These plots highlight the changes in structural dynamics caused by the interaction of IPNS with the metal ion. Each curve represents a specific peptide, showing its deuterium incorporation over time within the experimental time window. Observing these curves provides insight into how Zn(II) binding influences the flexibility and stability of different regions of the protein.

A.1.2 IPNS vs IPNS + Zn(II) and ACV

In the next stage of our analysis, we compare the deuterium uptake of IPNS in its default state with that of IPNS bound to both Zn(II) and the ACV tripeptide. This step allowed us to evaluate the combined effects of metal ion binding and substrate interaction on protein structural dynamics.

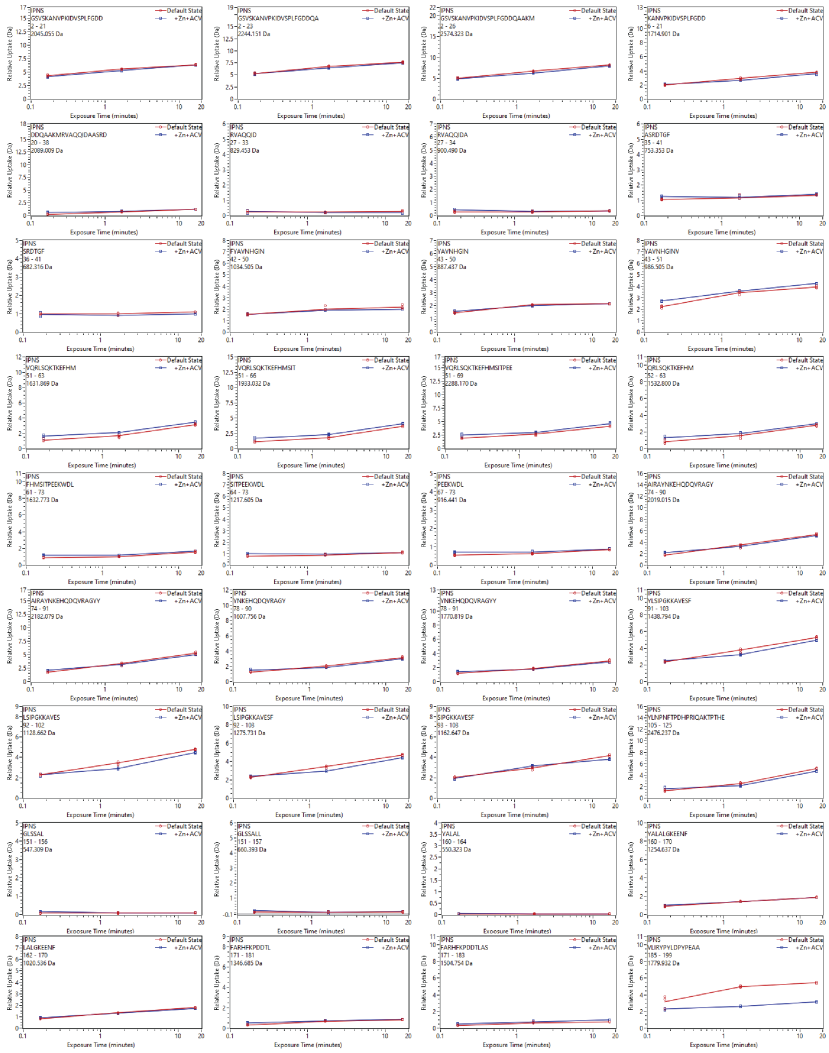


Figure A.3: Raw uptake plots of peptides 1-36 of IPNS and IPNS + Zn(II) and ACV. IPNS in the plots is labelled as Default state, and IPNS + Zn(II) and ACV is labelled + Zn + ACV.

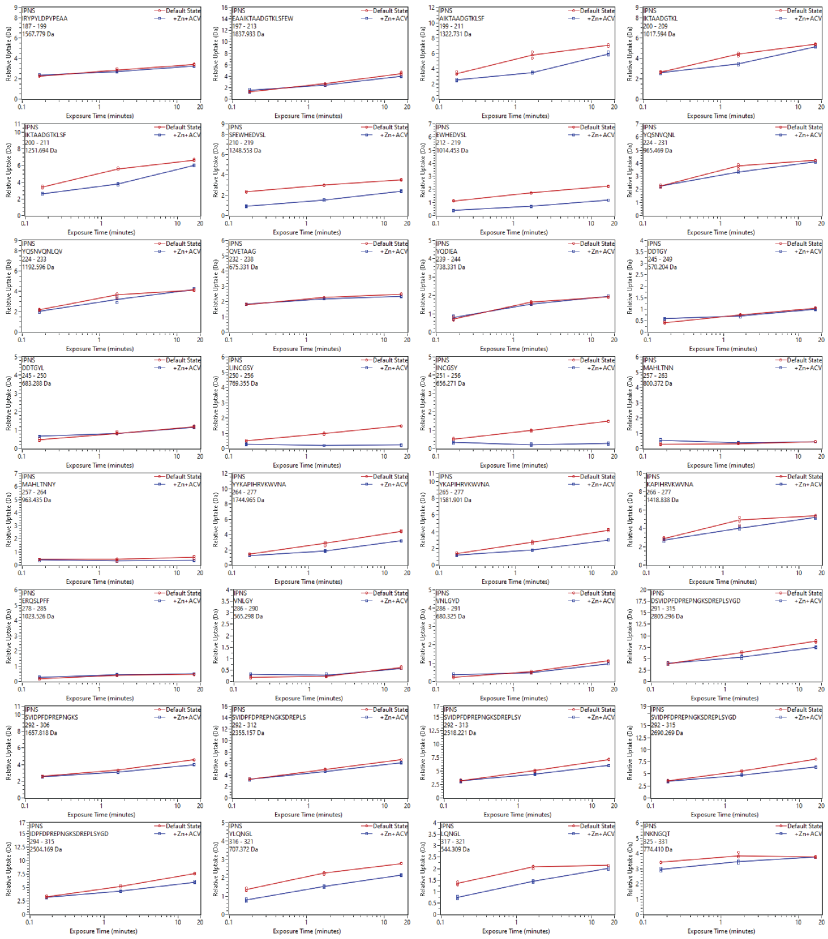


Figure A.4: Raw uptake plots of peptides 37-68 of IPNS and IPNS + Zn(II) and ACV. IPNS in the plots is labelled as Default state, and IPNS + Zn(II) and ACV is labelled + Zn + ACV.

Figures A.3 and A.4 show the deuterium uptake curves for peptides spanning residues 1-68. These plots compare the structural dynamics of IPNS alone with its behaviour when bound to both Zn(II) and ACV. By examining these curves, we can observe how the addition of the ACV tripeptide, in combination with Zn(II), further modulates the conformational behaviour of the protein. The resulting data provide critical insights into the interplay between metal ion and substrate binding in regulating the activity and structure of IPNS.

Each of the data sets presented here forms the backbone of the broader conclusions drawn in Chapter 3. These raw uptake plots provide a clear and unprocessed view of the dynamic changes underpinning the IPNS function under different experimental conditions.

B

How does the presence of the C2 domain affect the structure and function in SHIP1 and SHIP2

Contents

B.1 Sequence Identification of SHIP1 and SHIP2 132
B.1.1 Sequence Alignment of SHIP1 and SHIP2 132
B.2 Raw Uptake Plots 134
B.2.1 SHIP1 134
B.2.2 SHIP2 137
B.2.3 Full SAUSC output for SHIP2 139

B.1 Sequence Identification of SHIP1 and SHIP2

As discussed in Chapter 4, SHIP1 and SHIP2 are closely related phosphatases that play crucial roles as mediators in various diseases. Their functional similarities and involvement in critical signalling pathways have made them subjects of considerable interest. Given their close relationship, a detailed comparison of their sequences is essential to understand their shared and distinct functions.

The sequence identity between SHIP1 and SHIP2 is significant, with a match 57% between the primary sequences of the two proteins. This level of similarity suggests conserved functional roles while also leaving room for potential differences in activity, regulation, or interaction partners.

B.1.1 Sequence Alignment of SHIP1 and SHIP2

To further investigate the relationship between SHIP1 and SHIP2, we performed a sequence alignment using the online BLASTp tool. The analysis focused specifically on the Ptase and C2 protein domains, as these regions are critical for the enzymatic and regulatory functions of proteins.

Figure B.1 highlights the conserved and variable regions between SHIP1 and SHIP2 within the Ptase and C2 domains. The alignment reveals a high degree of conservation, particularly in regions directly involved in catalytic activity or domain-specific interactions. These findings underscore the importance of these domains in maintaining the shared core functions of SHIP1 and SHIP2, while also pointing to areas that may contribute to their unique roles in cell signaling.

```
Identities:262/458(57%), Positives:342/458(74%), Gaps:5/458(1%)

SHIP2  2    PEPDMISVFIGTWNMGSVPPPKNVTSWFTSKGLGKTLDEVTVTIPHDIYVFGTQENSVGD  61
            PEPDMI++FIGTWNMG+ PPPK +TSWF SKG GKT D+    IPHDIYV GTQE+ + +
SHIP1  5    PEPDMITIFIGTWNMGNAPPPKKITSWFLSKGQGKTRDDSADYIPHDIYVIGTQEDPLSE  64

SHIP2  62   REWLDLLRGGLKELTDLDYRPIAMQSLWNIKVAVLVKPEHENRISHVSTSSVKTGIANTL  121
            +EWL++L+  L+E+T + ++ +A+ +LWNI++ VL KPEHENRISH+ T +VKTGIANTL
SHIP1  65   KEWLEILKHSLQEITSVTFKTVAIHTLWNIRIVVLAKPEHENRISHICTDNVKTGIANTL  124

SHIP2  122  GNKGAVGVSFMFNGTSFGFVNCHLTSGNEKTARRNQNYLDILRLLSLGDRQLNAFDISLR  181
            GNKGAVGVSFMFNGTS GFVN HLTSG+EK  RRNQNY++ILR L+LGD++L+ F+I+ R
SHIP1  125  GNKGAVGVSFMFNGTSLGFVNSHLTSGSEKKLRRNQNYMNILRFLALGDKKLSPFNITHR  184

SHIP2  182  FTHLFWFGDLNYRLDM---DIQEILNYISRKEFEPLLRVDQLNLEREKHKVFLRFSEEEI  238
            FTHLFWFGDLNYR+D+   + + I+  I ++++  LL  DQL  ER + KVFL F EEEI
SHIP1  185  FTHLFWFGDLNYRVDLPTWEAETIIQKIKQQQYADLLSHDQLLTERREQKVFLHFEEEEI  244

SHIP2  239  SFPPTYRYERGSRDTYAWHKQKPTGVRTNVPSWCDRILWKSYPETHIICNSYGCTDDIVT  298
            +F PTYR+ER +RD YA+ KQK TG++ N+PSWCDR+LWKSYP  H++C SYG T DI+T
SHIP1  245  TFAPTYRFERLTRDKYAYTKQKATGMKYNLPSWCDRVLWKSYPLVHVVCQSYGSTSDIMT  304

SHIP2  299  SDHSPVFGTFEVGVTSQFISKKGLSKTSDQAYIEFESIEAIVKTASRTKFFIEFYSTCLE  358
            SDHSPVF TFE GVTSQF+SK G      Q  IEF    A +KT S+TKF++EF+S+CLE
SHIP1  305  SDHSPVFATFEAGVTSQFVSKNGPGTVDSQGQIEFLRCYATLKTKSQTKFYLEFHSSCLE  364

SHIP2  359  EYKKSFENDAQSSDNINFLKVQWSSRQLPTLKPILADIEYLQDQHLLLTVKSMDGYESYG  418
             + KS +       +   L V++    LP LKPI++D EYL DQH+L+++KS D  ESYG
SHIP1  365  SFVKS-QEGENEEGSEGELVVKF-GETLPKLKPIISDPEYLLDQHILISIKSSDSDESYG  422

SHIP2  419  ECVVALKSMIGSTAQQFLTFLSHRGEETGNIRGSMKVR  456
            E  +AL+     T     T L+H GE TG+ +G +K++
SHIP1  423  EGCIALRLEATETQLPIYTPLTHHGELTGHFQGEIKLQ  460
```

Figure B.1: The aligned sequences of the SHIP1 and SHIP2 complexes across the Ptase and C2 domains.

B.2 Raw Uptake Plots

HDX-MS was used to examine the structural dynamics of SHIP1 and SHIP2 under various conditions. The raw uptake data, which form the basis of the analysis, provide a direct view of the incorporation of deuterium for individual peptides. This section presents the results for both proteins, highlighting the behaviour of constructs with and without the C2 domain.

B.2.1 SHIP1

For SHIP1, the uptake plots are divided into two groups, spanning peptides 1-48 and 49-83. These regions correspond to common peptides identified within the Ptase domain across both constructs.

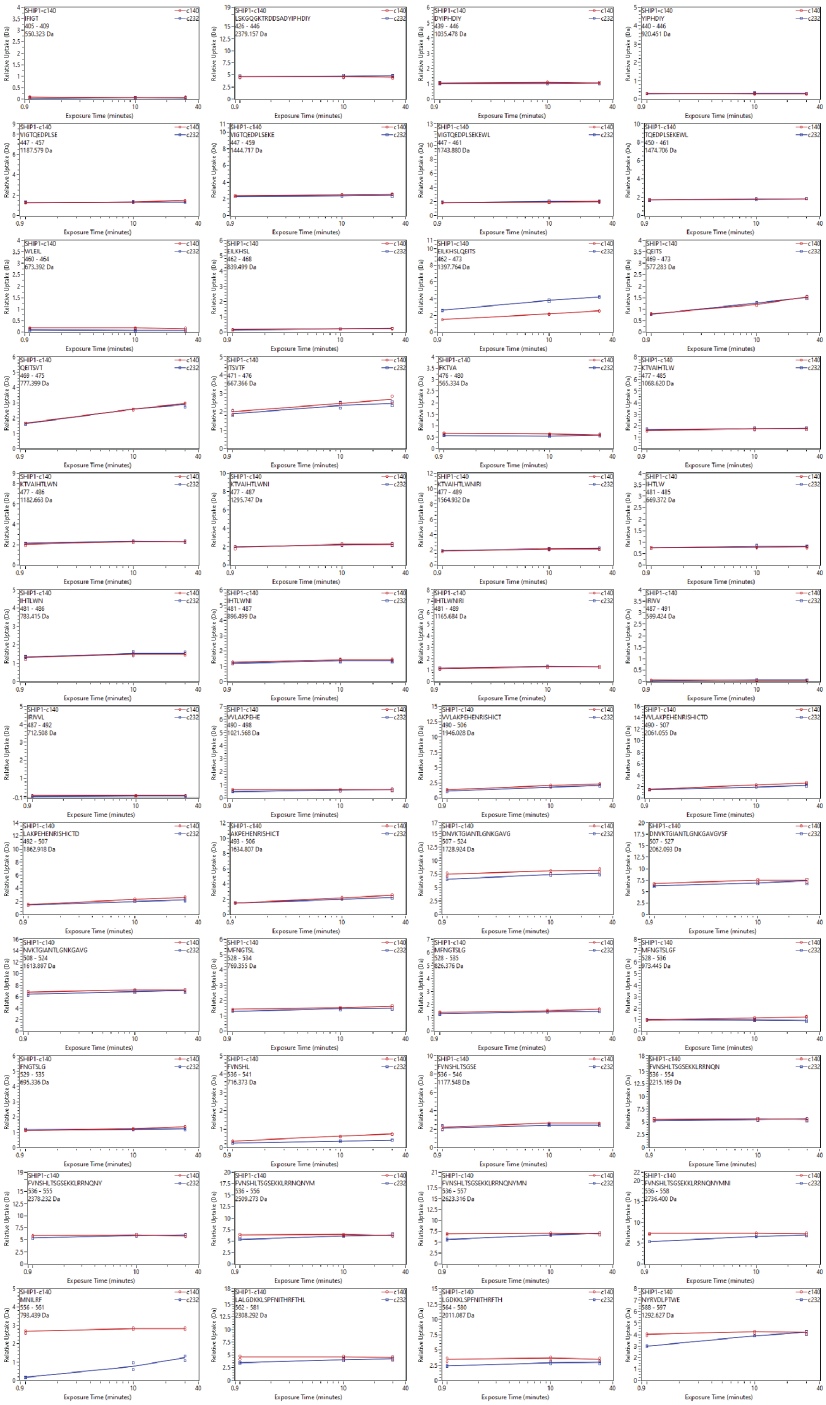


Figure B.2: Raw uptake plots of peptides 1-48 of SHIP1. Constructs with and without the C2 domain are labelled c232 and c140, respectively. The plots show common peptides within the Ptase domain.

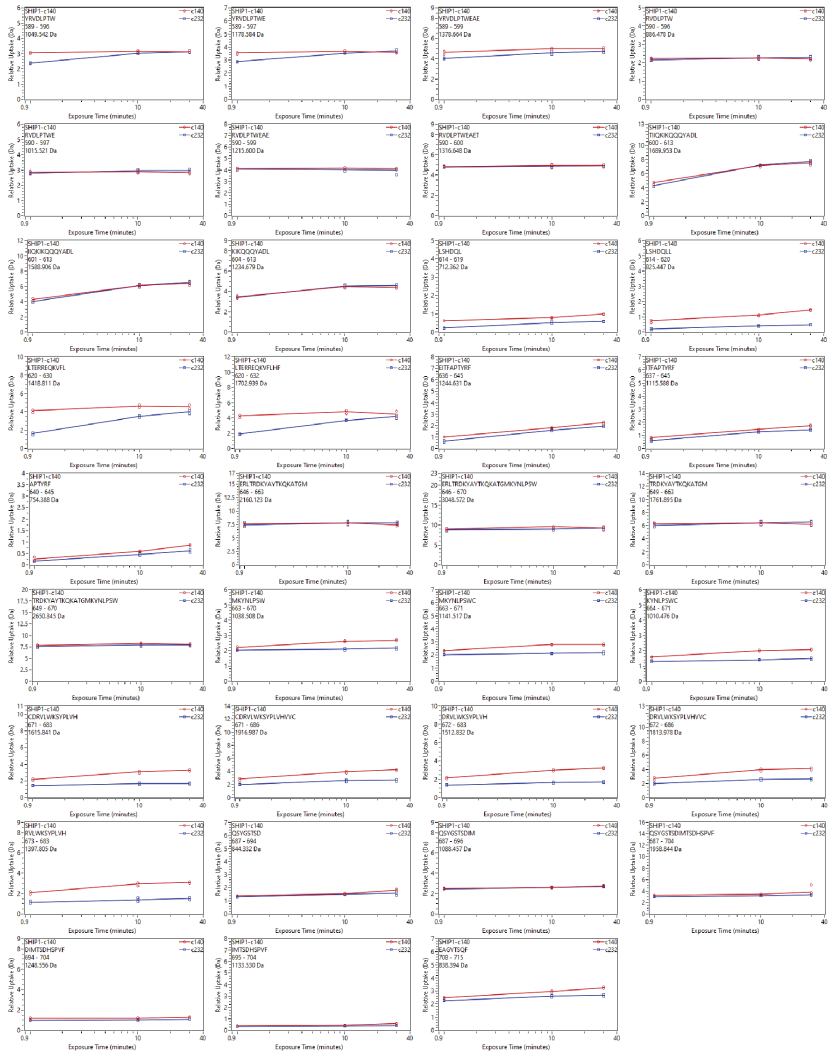


Figure B.3: Raw uptake plots of peptides 49-83 of SHIP1. Constructs with and without the C2 domain are labelled c232 and c140, respectively. The plots show common peptides within the Ptase domain.

Figures B.2 and B.3 present the raw deuterium uptake curves for peptides throughout the SHIP1 Ptase domain. The two constructs, c232 (with the C2 domain) and c140 (without the C2 domain), are compared to assess the influence of the C2 domain on protein flexibility and dynamics. Differences in uptake patterns reveal regions of structural modulation caused by the presence or absence of the C2 domain. Data was analysed using the SAUSC tool, as discussed in detail in Chapter 2, to quantify and interpret these differences.

B.2.2 SHIP2

Similarly, the raw uptake plots for SHIP2 are presented in two sections, covering peptides 1-36 and 37-66. These peptides represent common regions within the Ptase domain identified across the constructs.

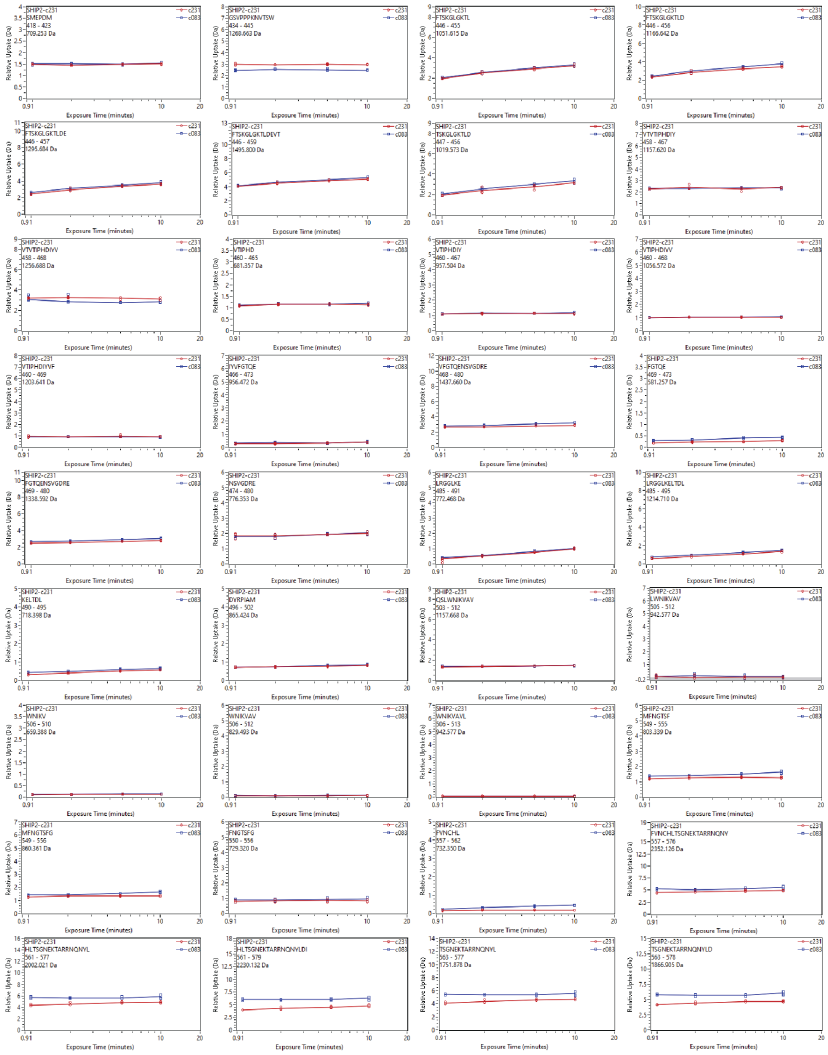


Figure B.4: Raw uptake plots of peptides 1-36 of SHIP2. Constructs with and without the C2 domain are labelled c231 and c083, respectively. The plots show common peptides within the Ptase domain.

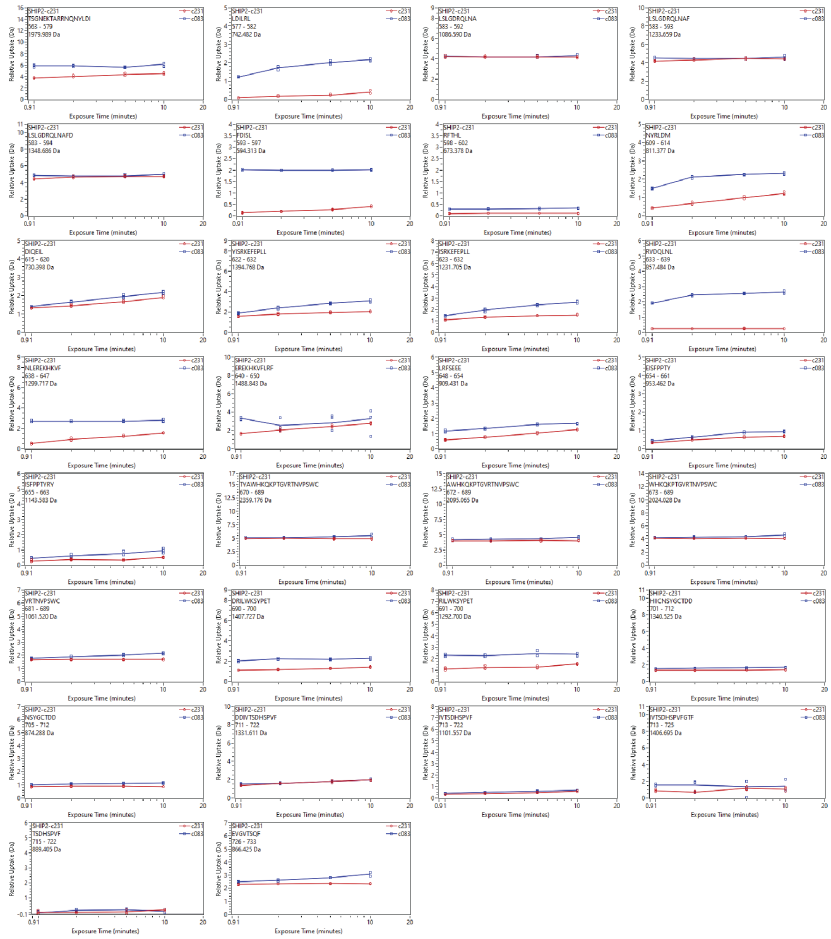


Figure B.5: Raw uptake plots of peptides 37-66 of SHIP2. Constructs with and without the C2 domain are labelled c231 and c083, respectively. The plots show common peptides within the Ptase domain.

Figures B.4 and B.5 show the raw deuterium uptake data for SHIP2 peptides in constructs with (c231) and without (c083) the C2 domain. These plots allow us to visualise how the structural dynamics of SHIP2 change depending on the presence of the C2 domain. The data provides crucial insights into the role of the C2 domain in modulating the behaviour of the Ptase domain, with potential implications for the protein's function in signalling pathways.

By comparing the raw data for SHIP1 and SHIP2, we can better understand the conserved and divergent features of these closely related proteins. This foundational data underpin the broader analyses and conclusions presented throughout this work.

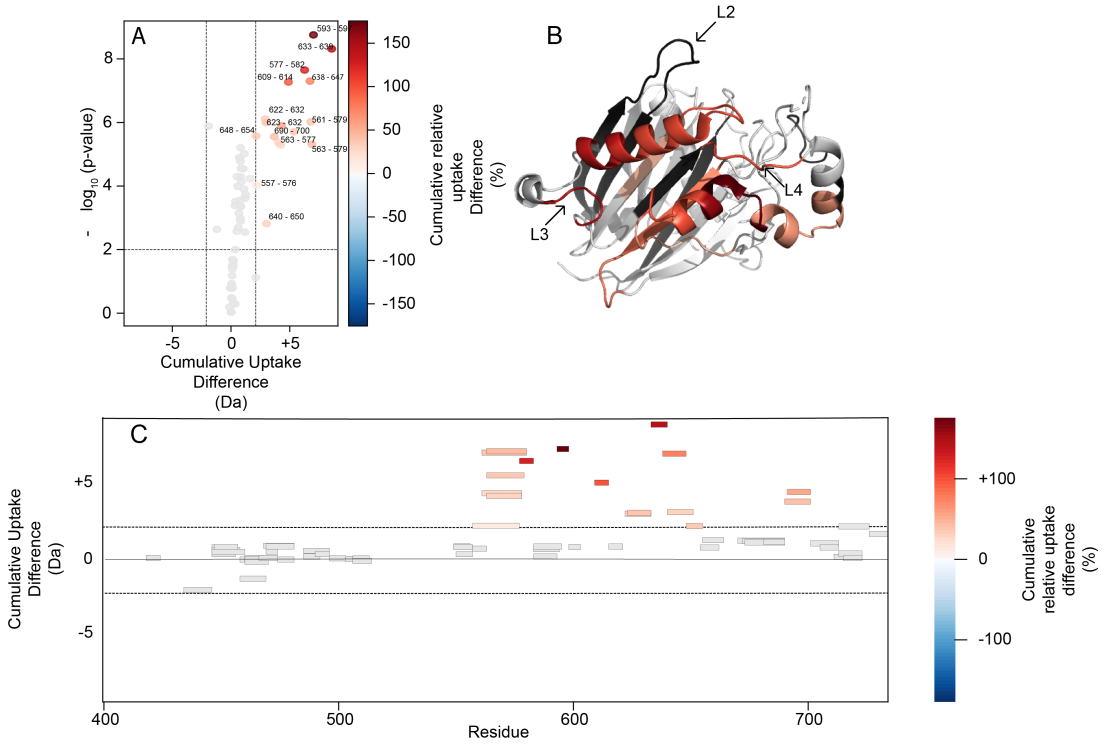


Figure B.6: Full SAUSC output for SHIP2. Shown here is the SAUSC output for SHIP2. Data examined is the cumulative data. The volcano plot in **A** is annotated with the terminal residue of the peptides. **B** is the coloured PDB structre and **C** is the respective woods plot. All coloured with the same gradient.

B.2.3 Full SAUSC output for SHIP2

Figure B.6 presents a slightly modified version of Figure 4.5. The adjustments made in this version aim to emphasise specific areas of interest while retaining the integrity of the original data set. The analysis of these data was performed using the SAUSC methodology. This robust statistical approach allows for the identification and quantification of differences in the uptake of deuterium under the experimental conditions.

The SAUSC analysis provides a comprehensive view of the variability within the dataset, allowing us to pinpoint regions where structural changes are most pronounced. These differences are illustrated with clarity in the output, highlighting specific peptides or domains where the dynamic behaviour of the protein varies significantly under different experimental conditions. By focussing on these distinct areas, we gain deeper insights into the structural and functional implications

of the observed changes.

The altered figure aligns with the overarching goals of this work by offering a more targeted representation of the data, aiding in the visualisation of the key regions of interest. This improves our understanding of how experimental variables, such as the presence of the C2 domain, influence the structural dynamics of SHIP2. Through these analyses, we can better interpret the implications of these findings in the broader context of SHIP2's functional functions.

C

Overcoming Glycan Heterogeneity Using an Adapted HDX Workflow

Contents

C.1 Glycan nomenclature 142
C.2 Additional coverage maps of tested proteins 143
C.2.1 Hp 1-1 143
C.2.2 VISTA 144

C.1 Glycan nomenclature

Standardisation of the representation of glycan structures is critical for clear and effective communication within the scientific community. The tools and methodologies developed for this purpose have been widely embraced by researchers as they provide a consistent and universally understood framework for depicting these complex molecules. This standardised approach simplifies the interpretation and comparison of glycan-related data across various studies and disciplines, ensuring that the information is conveyed accurately and without ambiguity.

By adopting this standardised system, researchers can rely on a shared visual language that transcends individual preferences or laboratory-specific conventions. The symbols used in these representations are carefully designed to be intuitive, unambiguous, and easily distinguishable, making them suitable for integration into scientific manuscripts, presentations, databases, and educational materials.

SHAPE	White	Blue	Green	Yellow	Orange	Pink	Purple	Light Blue	Brown	Red
Filled Circle	Hexose	Glc	Man	Gal	Gul	Alt	All	Tal	Ido	
Filled Square	HexNAc	GlcNAc	ManNAc	GalNAc	GulNAc	AltNAc	AllNAc	TalNAc	IdoNAc	
Crossed Square	Hexosamine	GlcN	ManN	GalN	GulN	AltN	AllN	TalN	IdoN	
Divided Diamond	Hexuronate	GlcA	ManA	GalA	GulA	AltA	AllA	TalA	IdoA	
Filled Triangle	Deoxyhexose	Qui	Rha		6dGul	6dAlt		6dTal		Fuc
Divided Triangle	DeoxyhexNAc	QuiNAc	RhaNAc			6dAltNAc		6dTalNAc		FucNAc
Flat Rectangle	Di-deoxyhexose	Oli	Tyv		Abe	Par	Dig	Col		
Filled Star	Pentose		Ara	Lyx	Xyl	Rib				
Filled Diamond	Deoxynonulosonate		Kdn				Neu5Ac	Neu5Gc	Neu	Sia
Flat Diamond	Di-deoxynonulosonate		Pse	Leg		Aci		4eLeg		
Flat Hexagon	Unknown	Bac	LDManHep	Kdo	Dha	DDManHep	MurNAc	MurNGc	Mur	
Pentagon	Assigned	Api	Fruc	Tag	Sor	Psi				

Figure C.1: Typical nomenclature of monosaccharide shape and colours

Figure C.1 shows the standard symbols and colours used to represent many

of the most common monosaccharides. These symbols are part of the Symbol Nomenclature for Glycans (SNFG), a system designed to make glycan structures easier to understand and share [122, 123].

Each monosaccharide is shown with a specific shape and colour, which makes it easy to distinguish them in diagrams. For example, glucose is represented as a blue circle, mannose is a green circle, and galactose is a yellow circle. This clear and visual coding helps avoid confusion and makes even complex structures more straightforward to interpret [122].

The combination of shapes and colours is practical and effective and works well for everything from scientific papers to presentations and teaching materials. Using the same symbols consistently, researchers can communicate more clearly and collaborate more easily [123].

In short, the standard shapes and colours in Figure C.1 are a simple but powerful tool for representing glycans, making it easier for everyone to work with and understand these intricate molecules [124, 125].

C.2 Additional coverage maps of tested proteins

In chapter C we show in Table 5.1 different proteins used to test the ability and performance of the PNGase column to cleave glycans to help us identify peptides in areas where we were previously unable to do so. The sequence coverage of some of the proteins mentioned is shown here.

C.2.1 Hp 1-1

Human Hp 1-1 was used to test the performance of the columns. With four glycan sites, we saw a great increase in coverage. An additional 15 peptides were found at these four sites, leading to an increase of 13.5% of sequence coverage, this can be seen in figure C.2.

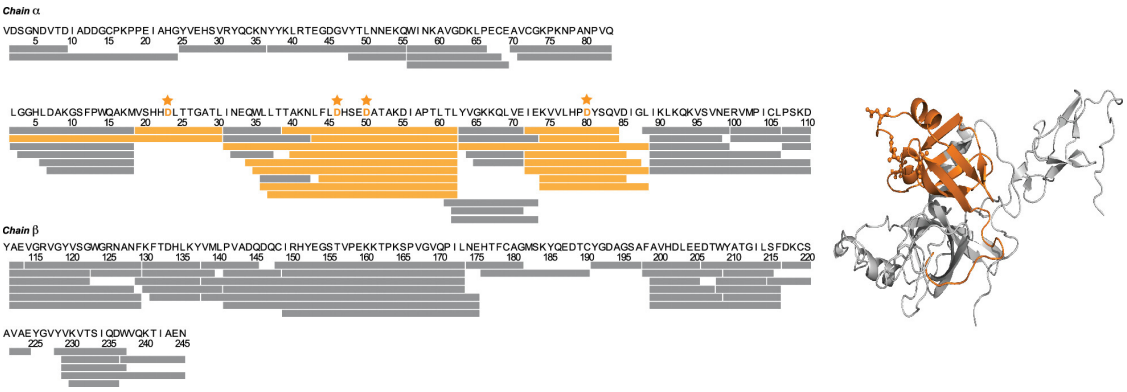


Figure C.2: Coverage map of Hp 1-1. Additional peptides found through the use of the PNGase Rc column are shown in orange with the peptides identified without the use of the PNGase column in grey. Glycan site identified with the use of an orange star over letter designated in the primary sequence. This colouring was then mapped onto a PDB file to show the increased coverage in a more visual way.

Peptides that were identifiable prior to the use of the PNGase column are shown in grey. The additional peptides found after the use of the column are shown in orange. The glycan sites are shown with the use of an orange star over a single letter designation in the primary sequence.

C.2.2 VISTA

Similarly, VISTA was used to test the column performance, and we saw huge increases in the number of peptides that we found. The sequence coverage increased by 24% and additional peptides were found spanning the six glycan sites. The coverage map is shown below in figure C.3.

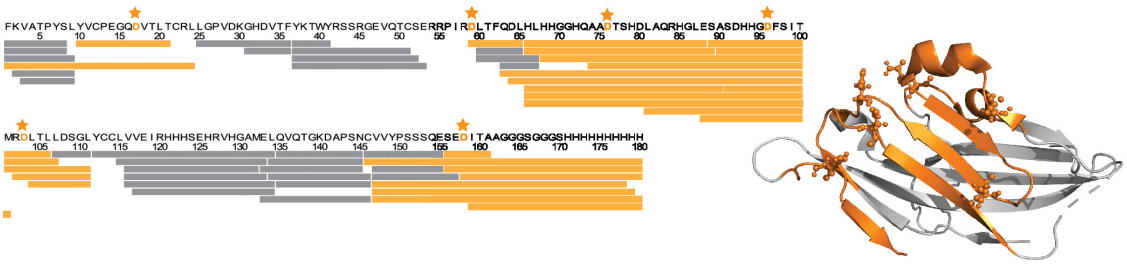


Figure C.3: Coverage map of VISTA. Additional peptides found through the use of the PNGase Rc column are shown in orange with the peptides identified without the use of the PNGase column in grey. Glycan site identified with the use of an orange star over letter designated in the primary sequence. This colouring was then mapped onto a PDB file to show the increased coverage in a more visual way.

In the same way as shown in figure C.2 the newly found peptides after the use of the PNGase column are shown here in orange. with the glycan site shown by the use of the same orange stars.

The first kind of intellectual and artistic personality belongs to the hedgehogs, the second to the foxes ...

— Sir Isaiah Berlin [**berlin_hedgehog_2013**]

References

[1] Justin LP Benesch and Brandon T Ruotolo. "Mass spectrometry: come of age for structural and dynamical biology". In: *Current Opinion in Structural Biology* 21.5 (Oct. 2011), pp. 641–649. URL: http://dx.doi.org/10.1016/j.sbi.2011.08.002.

[2] Joseph A. Loo. "Studying noncovalent protein complexes by electrospray ionization mass spectrometry". In: *Mass Spectrometry Reviews* 16.1 (1997), pp. 1–23. URL: http://dx.doi.org/10.1002/(SICI)1098-2787(1997)16:1%3C1::AID-MAS1%3E3.0.CO;2-L.

[3] Manuel Baca and Stephen B. H. Kent. "Direct observation of a ternary complex between the dimeric enzyme HIV-1 protease and a substrate-based inhibitor". In: *Journal of the American Chemical Society* 114.10 (May 1992), pp. 3992–3993. URL: http://dx.doi.org/10.1021/ja00036a066.

[4] K. J. Light-Wahl, B. E. Winger, and R. D. Smith. "Observation of the multimeric forms of concanavalin A by electrospray ionization mass spectrometry". In: *Journal of the American Chemical Society* 115.13 (June 1993), pp. 5869–5870. URL: http://dx.doi.org/10.1021/ja00066a083.

[5] Karen J. Light-Wahl, Brenda L. Schwartz, and Richard D. Smith. "Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry". In: *Journal of the American Chemical Society* 116.12 (June 1994), pp. 5271–5278. URL: http://dx.doi.org/10.1021/ja00091a035.

[6] Eric C. Huang et al. "Application of electrospray mass spectrometry in probing protein-protein and protein-ligand noncovalent interactions". In: *Journal of the American Society for Mass Spectrometry* 4.8 (Aug. 1993), pp. 624–630. URL: http://dx.doi.org/10.1016/1044-0305(93)85026-T.

[7] Justin L. P. Benesch et al. "Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics". In: *Chemical Reviews* 107.8 (July 2007), pp. 3544–3567. URL: http://dx.doi.org/10.1021/cr068289b.

[8] Andrej Sali et al. "From words to literature in structural proteomics". In: *Nature* 422.6928 (Mar. 2003), pp. 216–225. URL: http://dx.doi.org/10.1038/nature01513.

[9] Laura A. Lane et al. "Mass Spectrometry Reveals Stable Modules in holo and apo RNA Polymerases I and III". In: *Structure* 19.1 (Jan. 2011), pp. 90–100. URL: http://dx.doi.org/10.1016/j.str.2010.11.009.

[10] Anton Poliakov et al. "Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies". In: *Journal of Structural Biology* 157.2 (Feb. 2007), pp. 371–383. URL: http://dx.doi.org/10.1016/j.jsb.2006.09.003.

[11] John R. Engen and David L. Smith. “Peer Reviewed: Investigating Protein Structure and Dynamics by Hydrogen Exchange MS”. In: *Analytical Chemistry* 73.9 (May 2001), 256 A–265 A. URL: http://dx.doi.org/10.1021/ac012452f.

[12] Michal Sharon and Carol V. Robinson. “The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes”. In: *Annual Review of Biochemistry* 76.1 (June 2007), pp. 167–193. URL: http://dx.doi.org/10.1146/annurev.biochem.76.061005.090816.

[13] Timothy M. Allison and Mark T. Agasid. “Native Protein Mass Spectrometry”. In: *Protein Nanotechnology*. Springer US, Oct. 2019, pp. 287–299. URL: http://dx.doi.org/10.1007/978-1-4939-9869-2_15.

[14] B RUOTOLO and C ROBINSON. “Aspects of native proteins are retained in vacuum”. In: *Current Opinion in Chemical Biology* 10.5 (Oct. 2006), pp. 402–408. URL: http://dx.doi.org/10.1016/j.cbpa.2006.08.020.

[15] Brian E. Winger et al. “High-resolution accurate mass measurements of biomolecules using a new electrospray ionization ion cyclotron resonance mass spectrometer”. In: *Journal of the American Society for Mass Spectrometry* 4.7 (July 1993), pp. 566–577. URL: http://dx.doi.org/10.1016/1044-0305(93)85018-S.

[16] Masamichi Yamashita and John B. Fenn. “Electrospray ion source. Another variation on the free-jet theme”. In: *The Journal of Physical Chemistry* 88.20 (Sept. 1984), pp. 4451–4459. URL: http://dx.doi.org/10.1021/j150664a002.

[17] Malcolm Dole et al. “Molecular Beams of Macroions”. In: *The Journal of Chemical Physics* 49.5 (Sept. 1968), pp. 2240–2249. URL: http://dx.doi.org/10.1063/1.1670391.

[18] Matthias S. Wilm and Matthias Mann. “Electrospray and Taylor-Cone theory, Dole’s beam of macromolecules at last?” In: *International Journal of Mass Spectrometry and Ion Processes* 136.2–3 (Sept. 1994), pp. 167–180. URL: http://dx.doi.org/10.1016/0168-1176(94)04024-9.

[19] J. Fernández De La Mora and I. G. Loscertales. “The current emitted by highly conducting Taylor cones”. In: *Journal of Fluid Mechanics* 260 (Feb. 1994), pp. 155–184. URL: http://dx.doi.org/10.1017/S0022112094003472.

[20] Matthias Wilm. “Principles of Electrospray Ionization”. In: *Molecular amp; Cellular Proteomics* 10.7 (July 2011), p. M111.009407. URL: http://dx.doi.org/10.1074/mcp.M111.009407.

[21] Shibdas Banerjee and Shyamalava Mazumdar. “Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte”. In: *International Journal of Analytical Chemistry* 2012 (2012), pp. 1–40. URL: http://dx.doi.org/10.1155/2012/282574.

[22] M. Gamero-Castaño and J. Fernández de la Mora. “Mechanisms of electrospray ionization of singly and multiply charged salt clusters”. In: *Analytica Chimica Acta* 406.1 (Feb. 2000), pp. 67–91. URL: http://dx.doi.org/10.1016/S0003-2670(99)00596-6.

[23] J. V. Iribarne and B. A. Thomson. “On the evaporation of small ions from charged droplets”. In: *The Journal of Chemical Physics* 64.6 (Mar. 1976), pp. 2287–2294. URL: http://dx.doi.org/10.1063/1.432536.

[24] Elnaz Aliyari and Lars Konermann. "Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry". In: *Analytical Chemistry* 92.15 (July 2020), pp. 10807–10814. URL: http://dx.doi.org/10.1021/acs.analchem.0c02290.

[25] Haidy Metwally, Quentin Duez, and Lars Konermann. "Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry". In: *Analytical Chemistry* 90.16 (July 2018), pp. 10069–10077. URL: http://dx.doi.org/10.1021/acs.analchem.8b02926.

[26] Martijn W. H. Pinkse et al. "Macromolecular assembly of Helicobacter pylori urease investigated by mass spectrometry". In: *Journal of Mass Spectrometry* 38.3 (Jan. 2003), pp. 315–320. URL: http://dx.doi.org/10.1002/jms.443.

[27] Sarah Sanglier et al. "Comparative ESI-MS study of 2.2 MDa native hemocyanins from deep-sea and shore crabs: from protein oligomeric state to biotope". In: *Journal of the American Society for Mass Spectrometry* 14.5 (May 2003), pp. 419–429. URL: http://dx.doi.org/10.1016/S1044-0305(03)00131-4.

[28] Mark A. Tito et al. "Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid". In: *Journal of the American Chemical Society* 122.14 (Mar. 2000), pp. 3550–3551. URL: http://dx.doi.org/10.1021/ja993740k.

[29] Stefani N. Thomas. "Mass spectrometry". In: *Contemporary Practice in Clinical Chemistry*. Elsevier, 2019, pp. 171–185. URL: http://dx.doi.org/10.1016/B978-0-12-815499-1.00010-7.

[30] P. Labastie and M. Doy. "A high mass range quadrupole spectrometer for cluster studies". In: *International Journal of Mass Spectrometry and Ion Processes* 91.2 (July 1989), pp. 105–112. URL: http://dx.doi.org/10.1016/0168-1176(89)83001-0.

[31] Michaela Scigelova and Alexander Makarov. "Orbitrap Mass Analyzer – Overview and Applications in Proteomics". In: *PROTEOMICS* 6.S2 (Sept. 2006), pp. 16–21. URL: http://dx.doi.org/10.1002/pmic.200600528.

[32] Jürgen H Gross. *Mass Spectrometry: A Textbook*. Springer Science & Business Media, 2011.

[33] Qizhi Hu et al. "The Orbitrap: a new mass spectrometer". In: *Journal of Mass Spectrometry* 40.4 (Apr. 2005), pp. 430–443. URL: http://dx.doi.org/10.1002/jms.856.

[34] David Armbruster and Terry Pry. "Limit of Blank, Limit of Detection and Limit of Quantitation". In: *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists* 29 Suppl 1 (Sept. 2008), S49–52.

[35] F. Dubois, R. Knochenmuss, and R. Zenobi. "An ion-to-photon conversion detector for mass spectrometry". In: *International Journal of Mass Spectrometry and Ion Processes* 169–170 (Dec. 1997), pp. 89–98. URL: http://dx.doi.org/10.1016/S0168-1176(97)00210-3.

[36] Philip Lössl, Michiel van de Waterbeemd, and Albert JR Heck. "The diverse and expanding role of mass spectrometry in structural and molecular biology". In: *The EMBO Journal* 35.24 (Oct. 2016), pp. 2634–2657. URL: http://dx.doi.org/10.15252/embj.201694818.

[37] Oliver Ozohanics and Attila Ambrus. “Hydrogen-Deuterium Exchange Mass Spectrometry: A Novel Structural Biology Approach to Structure, Dynamics and Interactions of Proteins and Their Complexes”. In: *Life* 10.11 (Nov. 2020), p. 286. URL: http://dx.doi.org/10.3390/life10110286.

[38] Hui Wei et al. “Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications”. In: *Drug Discovery Today* 19.1 (Jan. 2014), pp. 95–102. URL: http://dx.doi.org/10.1016/j.drudis.2013.07.019.

[39] Vanesa Vinciauskaite and Glenn R. Masson. “Fundamentals of HDX-MS”. In: *Essays in Biochemistry* 67.2 (Mar. 2023). Ed. by Hannah Britt, Rebecca Beveridge, and Antonio Calabrese, pp. 301–314. URL: http://dx.doi.org/10.1042/EBC20220111.

[40] Thomas E. Wales and John R. Engen. “Hydrogen exchange mass spectrometry for the analysis of protein dynamics”. In: *Mass Spectrometry Reviews* 25.1 (Oct. 2005), pp. 158–170. URL: http://dx.doi.org/10.1002/mas.20064.

[41] Richard S. Johnson and Kenneth A. Walsh. “Mass spectrometric measurement of protein amide hydrogen exchange rates of apo- and holo-myoglobin”. In: *Protein Science* 3.12 (Dec. 1994), pp. 2411–2418. URL: http://dx.doi.org/10.1002/pro.5560031224.

[42] Zhongqi Zhang and David L. Smith. “Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation”. In: *Protein Science* 2.4 (Apr. 1993), pp. 522–531. URL: http://dx.doi.org/10.1002/pro.5560020404.

[43] David L. Smith, Yuzhong Deng, and Zhongqi Zhang. “Probing the Non-covalent Structure of Proteins by Amide Hydrogen Exchange and Mass Spectrometry”. In: *Journal of Mass Spectrometry* 32.2 (Feb. 1997), pp. 135–146. URL: http://dx.doi.org/10.1002/(SICI)1096-9888(199702)32:2%3C135::AID-JMS486%3E3.0.CO;2-M.

[44] Glenn R. Masson et al. “Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments”. In: *Nature Methods* 16.7 (June 2019), pp. 595–602. URL: https://doi.org/10.1038/s41592-019-0459-y.

[45] Yuzhong Deng, Zhongqi Zhang, and David L. Smith. “Comparison of continuous and pulsed labeling amide hydrogen exchange/mass spectrometry for studies of protein dynamics”. In: *Journal of the American Society for Mass Spectrometry* 10.8 (Aug. 1999), pp. 675–684. URL: http://dx.doi.org/10.1016/S1044-0305(99)00038-0.

[46] Houjun Yang and David L. Smith. “Kinetics of Cytochrome c Folding Examined by Hydrogen Exchange and Mass Spectrometry”. In: *Biochemistry* 36.48 (Dec. 1997), pp. 14992–14999. URL: http://dx.doi.org/10.1021/bi9717183.

[47] Kerene A. Brown and Derek J. Wilson. “Bottom-up hydrogen deuterium exchange mass spectrometry: data analysis and interpretation”. In: *The Analyst* 142.16 (2017), pp. 2874–2886. URL: http://dx.doi.org/10.1039/c7an00662d.

[48] Tajim Sheikh and Anil Shinde. “A Review And Recent Update On Lcms”. en. In: (2018). URL: https://zenodo.org/record/1446412.

[49] Andrew N. Hoofnagle, Katheryn A. Resing, and Natalie G. Ahn. "Practical Methods for Deuterium Exchange/Mass Spectrometry". In: *MAP Kinase Signaling Protocols*. Humana Press, 2004, pp. 283–298. URL: http://dx.doi.org/10.1385/1-59259-671-1:283.

[50] Sean R. Marcsisin and John R. Engen. "Hydrogen exchange mass spectrometry: what is it and what can it tell us?" In: *Analytical and Bioanalytical Chemistry* 397.3 (Mar. 2010), pp. 967–972. URL: http://dx.doi.org/10.1007/s00216-010-3556-4.

[51] Andy M Lau et al. "Deuteros 2.0: peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry". In: *Bioinformatics* 37.2 (Jan. 2021). Ed. by Arne Elofsson, pp. 270–272. URL: http://dx.doi.org/10.1093/bioinformatics/btaa677.

[52] Sanmin Liu et al. "HDX-Analyzer: a novel package for statistical analysis of protein structure dynamics". In: *BMC Bioinformatics* 12.S1 (Feb. 2011). URL: http://dx.doi.org/10.1186/1471-2105-12-S1-S43.

[53] Miklos Guttman et al. "Analysis of Overlapped and Noisy Hydrogen/Deuterium Exchange Mass Spectra". In: *Journal of the American Society for Mass Spectrometry* 24.12 (Sept. 2013), pp. 1906–1912. URL: http://dx.doi.org/10.1007/s13361-013-0727-5.

[54] John K. Chik, Jaclyn L. Vande Graaf, and David C. Schriemer. "Quantitating the Statistical Distribution of Deuterium Incorporation To Extend the Utility of H/D Exchange MS Data". In: *Analytical Chemistry* 78.1 (Nov. 2005), pp. 207–214. URL: http://dx.doi.org/10.1021/ac050988l.

[55] Paul H. C. Eilers. "Parametric Time Warping". In: *Analytical Chemistry* 76.2 (Dec. 2003), pp. 404–411. URL: http://dx.doi.org/10.1021/ac034800e.

[56] David Bouyssié et al. "HDX-Viewer: interactive 3D visualization of hydrogen–deuterium exchange data". In: *Bioinformatics* 35.24 (July 2019). Ed. by Arne Elofsson, pp. 5331–5333. URL: http://dx.doi.org/10.1093/bioinformatics/btz550.

[57] Véronique Hourdel et al. "MEMHDX: an interactive tool to expedite the statistical validation and visualization of large HDX-MS datasets". In: *Bioinformatics* 32.22 (July 2016), pp. 3413–3419. URL: http://dx.doi.org/10.1093/bioinformatics/btw420.

[58] Yoav Benjamini and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (Jan. 1995), pp. 289–300. URL: http://dx.doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[59] Tyler S. Hageman and David D. Weis. "Reliable Identification of Significant Differences in Differential Hydrogen Exchange-Mass Spectrometry Measurements Using a Hybrid Significance Testing Approach". In: *Analytical Chemistry* 91.13 (May 2019), pp. 8008–8016. URL: http://dx.doi.org/10.1021/acs.analchem.9b01325.

[60] Matthew D. Lieberman and William A. Cunningham. “Type I and Type II error concerns in fMRI research: re-balancing the scale”. In: *Social Cognitive and Affective Neuroscience* 4.4 (Dec. 2009), pp. 423–428. URL: http://dx.doi.org/10.1093/scan/nsp052.

[61] Scott Bezeau and Roger Graves. “Statistical Power and Effect Sizes of Clinical Neuropsychology Research”. In: *Journal of Clinical and Experimental Neuropsychology* 23.3 (June 2001), pp. 399–406. URL: http://dx.doi.org/10.1076/jcen.23.3.399.1181.

[62] Patrick Rabe et al. “Roles of 2-oxoglutarate oxygenases and isopenicillin N synthase in -lactam biosynthesis”. In: *Natural Product Reports* 35.8 (2018), pp. 735–756. URL: http://dx.doi.org/10.1039/C8NP00002F.

[63] Peter J. Rutledge. “Isopenicillin N Synthase”. In: *2-Oxoglutarate-Dependent Oxygenases*. The Royal Society of Chemistry, May 2015, pp. 414–424. URL: http://dx.doi.org/10.1039/9781782621959-00414.

[64] Refaat B. Hamed et al. “The enzymes of -lactam biosynthesis”. In: *Nat. Prod. Rep.* 30.1 (2013), pp. 21–107. URL: http://dx.doi.org/10.1039/C2NP20065A.

[65] Peter L. Roach et al. “Structure of isopenicillinN synthase complexed with substrate and the mechanism ofpenicillin formation”. In: *Nature* 387.6635 (June 1997), pp. 827–830. URL: http://dx.doi.org/10.1038/42990.

[66] Ilya Borovok et al. “Ferrous Active Site of Isopenicillin N Synthase: Genetic and Sequence Analysis of the Endogenous Ligands”. In: *Biochemistry* 35.6 (Jan. 1996), pp. 1981–1987. URL: http://dx.doi.org/10.1021/bi951534t.

[67] Lawrence Que and Raymond Y. N. Ho. “Dioxygen Activation by Enzymes with Mononuclear Non-Heme Iron Active Sites”. In: *Chemical Reviews* 96.7 (Jan. 1996), pp. 2607–2624. URL: http://dx.doi.org/10.1021/cr960039f.

[68] Christy C. Bridges and Rudolfs K. Zalups. “Molecular and ionic mimicry and the transport of toxic metals”. In: *Toxicology and Applied Pharmacology* 204.3 (May 2005), pp. 274–308. URL: http://dx.doi.org/10.1016/j.taap.2004.09.007.

[69] Amy L. Pochodylo, Annaleise R. Klein, and Ludmilla Aristilde. “Metal-binding selectivity and coordination dynamics for cyanobacterial microcystins with Zn, Cu, Fe, Mg, and Ca”. In: *Environmental Chemistry Letters* 15.4 (May 2017), pp. 695–701. URL: http://dx.doi.org/10.1007/s10311-017-0639-x.

[70] Michael J Chalmers et al. “Differential hydrogen/deuterium exchange mass spectrometry analysis of protein–ligand interactions”. In: *Expert Review of Proteomics* 8.1 (Feb. 2011), pp. 43–59. URL: http://dx.doi.org/10.1586/epr.10.109.

[71] Nicole C. Chapman and Peter J. Rutledge. “IsopenicillinN Synthase: Crystallographic Studies”. In: *ChemBioChem* 22.10 (Mar. 2021), pp. 1687–1705. URL: http://dx.doi.org/10.1002/cbic.202000743.

[72] Yawen Bai et al. “Primary structure effects on peptide group hydrogen exchange”. In: *Proteins: Structure, Function, and Bioinformatics* 17.1 (Sept. 1993), pp. 75–86. URL: http://dx.doi.org/10.1002/prot.340170110.

[73] Lars Konermann, Xin Tong, and Yan Pan. "Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches". In: *Journal of Mass Spectrometry* 43.8 (June 2008), pp. 1021–1036. URL: http://dx.doi.org/10.1002/jms.1435.

[74] S. Walter Englander. "Hydrogen exchange and mass spectrometry: A historical perspective". In: *Journal of the American Society for Mass Spectrometry* 17.11 (Nov. 2006), pp. 1481–1489. URL: http://dx.doi.org/10.1016/j.jasms.2006.06.006.

[75] Etai Jacob and Ron Unger. "A tale of two tails: why are terminal residues of proteins exposed?" In: *Bioinformatics* 23.2 (Jan. 2007), e225–e230. URL: http://dx.doi.org/10.1093/bioinformatics/btl318.

[76] Giada Marino, Ulrich Eckhard, and Christopher M. Overall. "Protein Termini and Their Modifications Revealed by Positional Proteomics". In: *ACS Chemical Biology* 10.8 (July 2015), pp. 1754–1764. URL: http://dx.doi.org/10.1021/acschembio.5b00189.

[77] Luke A. McNeill et al. "Terminally Truncated Isopenicillin N Synthase Generates a Dithioester Product: Evidence for a Thioaldehyde Intermediate during Catalysis and a New Mode of Reaction for Non-Heme Iron Oxidases". In: *Chemistry – A European Journal* 23.52 (Aug. 2017), pp. 12815–12824. URL: http://dx.doi.org/10.1002/chem.201701592.

[78] William J. Bradshaw et al. "Regulation of inositol 5-phosphatase activity by the C2 domain of SHIP1 and SHIP2". In: *Structure* (Feb. 2024). URL: http://dx.doi.org/10.1016/j.str.2024.01.005.

[79] Matthew J. Eramo and Christina A. Mitchell. "Regulation of PtdIns(3, 4, 5)P3/Akt signalling by inositol polyphosphate 5-phosphatases". In: *Biochemical Society Transactions* 44.1 (Feb. 2016), pp. 240–252. URL: http://dx.doi.org/10.1042/BST20150214.

[80] Helena Brauer et al. "Leukemia-associated mutations in SHIP1 inhibit its enzymatic activity, interaction with the GM-CSF receptor and Grb2, and its ability to inactivate PI3K/AKT signaling". In: *Cellular Signalling* 24.11 (Nov. 2012), pp. 2095–2101. URL: http://dx.doi.org/10.1016/j.cellsig.2012.07.017.

[81] Wendy M. Kandell et al. "MicroRNA-155 governs SHIP-1 expression and localization in NK cells and regulates subsequent infiltration into murine AT3 mammary carcinoma". In: *PLOS ONE* 15.2 (Feb. 2020). Ed. by Fabrizio Mattei, e0225820. URL: http://dx.doi.org/10.1371/journal.pone.0225820.

[82] Todd W Miller et al. "Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer". In: *Breast Cancer Research* 13.6 (Nov. 2011). URL: http://dx.doi.org/10.1186/bcr3039.

[83] Hajime Ishihara et al. "Molecular Cloning of Rat SH2-Containing Inositol Phosphatase 2 (SHIP2) and Its Role in the Regulation of Insulin Signaling". In: *Biochemical and Biophysical Research Communications* 260.1 (June 1999), pp. 265–272. URL: http://dx.doi.org/10.1006/bbrc.1999.0888.

[84] A. C. B. Marcano et al. “Genetic association analysis of inositol polyphosphate phosphatase-like 1 (INPPL1, SHIP2) variants with essential hypertension”. In: *Journal of Medical Genetics* 44.9 (May 2007), pp. 603–605. URL: http://dx.doi.org/10.1136/jmg.2007.049718.

[85] Kim H. T. Paraiso et al. “Induced SHIP Deficiency Expands Myeloid Regulatory Cells and Abrogates Graft-versus-Host Disease”. In: *The Journal of Immunology* 178.5 (Mar. 2007), pp. 2893–2900. URL: http://dx.doi.org/10.4049/jimmunol.178.5.2893.

[86] Yingyue Zhou, Tyler K. Ulland, and Marco Colonna. “TREM2-Dependent Effects on Microglia in Alzheimer’s Disease”. In: *Frontiers in Aging Neuroscience* 10 (July 2018). URL: http://dx.doi.org/10.3389/fnagi.2018.00202.

[87] Qisheng Peng et al. “TREM2- and DAP12-Dependent Activation of PI3K Requires DAP10 and Is Inhibited by SHIP1”. In: *Science Signaling* 3.122 (May 2010). URL: http://dx.doi.org/10.1126/scisignal.2000500.

[88] Samantha L. Rosenthal and M. Ilyas Kamboh. “Late-Onset Alzheimer’s Disease Genes and the Potentially Implicated Pathways”. In: *Current Genetic Medicine Reports* 2.2 (Mar. 2014), pp. 85–101. URL: http://dx.doi.org/10.1007/s40142-014-0034-x.

[89] Heather M. Wilkins et al. “Bioenergetic Dysfunction and Inflammation in Alzheimerâ€™s Disease: A Possible Connection”. In: *Frontiers in Aging Neuroscience* 6 (Nov. 2014). URL: http://dx.doi.org/10.3389/fnagi.2014.00311.

[90] Maud Gratuze, Cheryl E. G. Leyns, and David M. Holtzman. “New insights into the role of TREM2 in Alzheimer’s disease”. In: *Molecular Neurodegeneration* 13.1 (Dec. 2018). URL: http://dx.doi.org/10.1186/s13024-018-0298-9.

[91] Jean-Charles Lambert et al. “Meta-analysis of 74, 046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease”. In: *Nature Genetics* 45.12 (Oct. 2013), pp. 1452–1458. URL: http://dx.doi.org/10.1038/ng.2802.

[92] Brian W. Kunkle et al. “Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates A, tau, immunity and lipid processing”. In: *Nature Genetics* 51.3 (Feb. 2019), pp. 414–430. URL: http://dx.doi.org/10.1038/s41588-019-0358-2.

[93] Hua Jing et al. “INPP5D rs35349669 polymorphism with late-onset Alzheimer’s disease: A replication study and meta-analysis”. In: *Oncotarget* 7.43 (Oct. 2016), pp. 69225–69230. URL: http://dx.doi.org/10.18632/oncotarget.12648.

[94] Diana J. Zajac et al. “Expression of INPP5D Isoforms in Human Brain: Impact of Alzheimer’s Disease Neuropathology and Genetics”. In: *Genes* 14.3 (Mar. 2023), p. 763. URL: http://dx.doi.org/10.3390/genes14030763.

[95] Johanne Le Coq et al. “Structural basis for interdomain communication in SHIP2 providing high phosphatase activity”. In: *eLife* 6 (Aug. 2017). URL: http://dx.doi.org/10.7554/eLife.26640.

[96] Katsuhiro Kato et al. “The inositol 5-phosphatase SHIP2 is an effector of RhoA and is involved in cell polarity and migration”. In: *Molecular Biology of the Cell* 23.13 (July 2012). Ed. by Tamotsu Yoshimori, pp. 2593–2604. URL: http://dx.doi.org/10.1091/mbc.E11-11-0958.

[97] Heng Zhou et al. "Expression, purification and characterization of the RhoA-binding domain of human SHIP2 in E.coli". In: *Protein Expression and Purification* 180 (Apr. 2021), p. 105821. URL: http://dx.doi.org/10.1016/j.pep.2021.105821.

[98] Christophe Erneux et al. "SHIP2 multiple functions: A balance between a negative control of PtdIns(3, 4, 5)P3 level, a positive control of PtdIns(3, 4)P2 production, and intrinsic docking properties". In: *Journal of Cellular Biochemistry* 112.9 (Aug. 2011), pp. 2203–2209. URL: http://dx.doi.org/10.1002/jcb.23146.

[99] Dennis R. Viernes et al. "Discovery and Development of Small Molecule SHIP Phosphatase Modulators". In: *Medicinal Research Reviews* 34.4 (Dec. 2013), pp. 795–824. URL: http://dx.doi.org/10.1002/med.21305.

[100] Chiara Pedicone et al. "Targeting SHIP1 and SHIP2 in Cancer". In: *Cancers* 13.4 (Feb. 2021), p. 890. URL: http://dx.doi.org/10.3390/cancers13040890.

[101] Johanne Le Coq et al. "A new layer of phosphoinositide-mediated allosteric regulation uncovered for SHIP2". In: *The FASEB Journal* 35.8 (July 2021). URL: http://dx.doi.org/10.1096/fj.202100561R.

[102] Jone Paesmans et al. "A structure of substrate-bound Synaptojanin1 provides new insights in its mechanism and the effect of disease mutations". In: *eLife* 9 (Dec. 2020). URL: http://dx.doi.org/10.7554/eLife.64922.

[103] Lionel Trésaugues et al. "Structural Basis for Phosphoinositide Substrate Recognition, Catalysis, and Membrane Interactions in Human Inositol Polyphosphate 5-Phosphatases". In: *Structure* 22.5 (May 2014), pp. 744–755. URL: http://dx.doi.org/10.1016/j.str.2014.01.013.

[104] Esben Trabjerg, Zeinab E. Nazari, and Kasper D. Rand. "Conformational analysis of complex protein states by hydrogen/deuterium exchange mass spectrometry (HDX-MS): Challenges and emerging solutions". In: *TrAC Trends in Analytical Chemistry* 106 (Sept. 2018), pp. 125–138. URL: http://dx.doi.org/10.1016/j.trac.2018.06.008.

[105] Rane A Harrison and John R Engen. "Conformational insight into multi-protein signaling assemblies by hydrogen–deuterium exchange mass spectrometry". In: *Current Opinion in Structural Biology* 41 (Dec. 2016), pp. 187–193. URL: http://dx.doi.org/10.1016/j.sbi.2016.08.003.

[106] Oscar Vadas and John E. Burke. "Probing the dynamic regulation of peripheral membrane proteins using hydrogen deuterium exchange–MS (HDX–MS)". In: *Biochemical Society Transactions* 43.5 (Oct. 2015), pp. 773–786. URL: http://dx.doi.org/10.1042/BST20150065.

[107] Gary Walsh and Roy Jefferis. "Post-translational modifications in the context of therapeutic proteins". In: *Nature Biotechnology* 24.10 (Oct. 2006), pp. 1241–1252. URL: https://doi.org/10.1038/nbt1252.

[108] Tala M. Bakheet and Andrew J. Doig. "Properties and identification of human protein drug targets". In: *Bioinformatics* 25.4 (Jan. 2009), pp. 451–457. URL: http://dx.doi.org/10.1093/bioinformatics/btp002.

[109] Dalit Shental-Bechor and Yaakov Levy. "Folding of glycoproteins: toward understanding the biophysics of the glycosylation code". In: *Current Opinion in Structural Biology* 19.5 (Oct. 2009), pp. 524–533. URL: http://dx.doi.org/10.1016/j.sbi.2009.07.002.

[110] Ajit Varki. "Biological roles of glycans". In: *Glycobiology* 27.1 (Aug. 2016), pp. 3–49. URL: http://dx.doi.org/10.1093/glycob/cww086.

[111] A. Varki. "Evolutionary Forces Shaping the Golgi Glycosylation Machinery: Why Cell Surface Glycans Are Universal to Living Cells". In: *Cold Spring Harbor Perspectives in Biology* 3.6 (Apr. 2011), a005462–a005462. URL: http://dx.doi.org/10.1101/cshperspect.a005462.

[112] Pernille Foged Jensen et al. "Removal of N-Linked Glycosylations at Acidic pH by PNGase A Facilitates Hydrogen/Deuterium Exchange Mass Spectrometry Analysis of N-Linked Glycoproteins". In: *Analytical Chemistry* 88.24 (Nov. 2016), pp. 12479–12488. URL: http://dx.doi.org/10.1021/acs.analchem.6b03951.

[113] Gerard Comamala et al. "Deglycosylation by the Acidic Glycosidase PNGase H+ Enables Analysis of N-Linked Glycoproteins by Hydrogen/Deuterium Exchange Mass Spectrometry". In: *Journal of the American Society for Mass Spectrometry* 31.11 (Sept. 2020), pp. 2305–2312. URL: http://dx.doi.org/10.1021/jasms.0c00258.

[114] Rui-Rui Guo et al. "PNGase H variant from iRudaea cellulosilytica/i with improved deglycosylation efficiency for rapid analysis of eukaryotic iN/i-glycans and hydrogen deuterium exchange mass spectrometry analysis of glycoproteins". In: *Rapid Communications in Mass Spectrometry* 36.21 (Sept. 2022). URL: https://doi.org/10.1002/rcm.9376.

[115] Matthias J Feige and Linda M Hendershot. "Disulfide bonds in ER protein folding and homeostasis". In: *Current Opinion in Cell Biology* 23.2 (Apr. 2011), pp. 167–175. URL: http://dx.doi.org/10.1016/j.ceb.2010.10.012.

[116] Di Wu and Carol V. Robinson. "Understanding glycoprotein structural heterogeneity and interactions: Insights from native mass spectrometry". In: *Current Opinion in Structural Biology* 74 (June 2022), p. 102351. URL: http://dx.doi.org/10.1016/j.sbi.2022.102351.

[117] Asif Shajahan et al. "Glycomic and glycoproteomic analysis of glycoproteins—a tutorial". In: *Analytical and Bioanalytical Chemistry* 409.19 (June 2017), pp. 4483–4505. URL: https://doi.org/10.1007/s00216-017-0406-7.

[118] Gerard Comamala et al. "Hydrogen/Deuterium Exchange Mass Spectrometry with Integrated Electrochemical Reduction and Microchip-Enabled Deglycosylation for Epitope Mapping of Heavily Glycosylated and Disulfide-Bonded Proteins". In: *Analytical Chemistry* 93.49 (Nov. 2021), pp. 16330–16340. URL: http://dx.doi.org/10.1021/acs.analchem.1c01728.

[119] Thomas Lambert et al. "Development of a PNGase Rc Column for Online Deglycosylation of Complex Glycoproteins during HDX-MS". In: *Journal of the American Society for Mass Spectrometry* (Sept. 2023). URL: https://doi.org/10.1021/jasms.3c00268.

[120] Hayden Wilkinson and Radka Saldova. "Current Methods for the Characterization of iO/i-Glycans". In: *Journal of Proteome Research* 19.10 (Sept. 2020), pp. 3890–3905. URL: https://doi.org/10.1021/acs.jproteome.0c00435.

[121] Kaewta Rattanapisit et al. "Recombinant human osteopontin expressed in Nicotiana benthamiana stimulates osteogenesis related genes in human periodontal ligament cells". In: *Scientific Reports* 7.1 (Dec. 2017). URL: http://dx.doi.org/10.1038/s41598-017-17666-7.

[122] Ajit Varki et al. "Symbol Nomenclature for Graphical Representation of Glycans". In: *Glycobiology* 25.12 (2015), pp. 1323–1324. URL: https://doi.org/10.1093/glycob/cwv091.

[123] Sriram Neelamegham et al. "Updates to the Symbol Nomenclature for Glycans guidelines". In: *Glycobiology* 29.9 (2019), pp. 620–624. URL: https://doi.org/10.1093/glycob/cwz045.

[124] Kai Cheng, Yaoqi Zhou, and Sriram Neelamegham. "DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information". In: *Glycobiology* 27.3 (2017), pp. 200–205. URL: https://doi.org/10.1093/glycob/cww118.

[125] Anuja Y. Mehta and Richard D. Cummings. "GlycoGlyph: a glycan visualizing, drawing and naming application". In: *Bioinformatics* 36.11 (2020), pp. 3613–3614. URL: https://doi.org/10.1093/bioinformatics/btaa200.