

Julia Eva Titz BSc

**CHARACTERIZING PROTEIN DYNAMICS BY
HYDROGEN DEUTERIUM EXCHANGE MASS
SPECTROMETRY**

Master Thesis

to achieve the university degree of

Diplom-Ingenieurin

Master's Degree Programme: Biotechnology

submitted to

Graz University of Technology

Supervisors

Dr. Klaus Rumpel

Boehringer Ingelheim, Vienna

&

Prof. Dr. Monika Oberer

Institute of Molecular Biosciences, Graz

Graz, April 2019

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used.

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Date/Datum

Signature/Unterschrift

Acknowledgment

First and foremost, I would like to express my gratitude to my supervisor Dr. Klaus Rumpel for his patient guidance, encouragement and advice throughout my time at Boehringer Ingelheim. I hope that I will keep the lessons that I have learned with me in my future career and life.

Moreover, I would like to thank my co-supervisor Prof. Dr. Monika Oberer for the helpful feedback and more importantly for inspiring me, to believe in myself and think bigger.

I would thank you my colleagues, Maximilian Scharnweber, Katja Hauer MSc, Michael Galant BSc, Sandra Winkler MSc and Dominik Gassner BSc, who welcomed me into their lab and helped me along the way.

I am grateful to Dr. Alex Muck, for providing me with fantastic lab training in mass spectrometry.

Lastly, I am infinitely grateful to my friends and family, for all their love, support and patience.

Mama, I always knew that you believed in me and wanted the best for me.

Abstract

Hydrogen deuterium exchange coupled with mass spectrometry (HDX MS) is useful for describing protein dynamics, protein folding, protein-protein interactions, and protein small molecule interactions. The exchange rate of hydrogen with deuterium by the peptide backbone is dependent on changes in protein structure, interaction with ligands, protein-protein interaction and environmental factors such as pH and temperature. These conformational changes can stabilize or destabilize specific regions in the protein, which increases or decreases the deuterium uptake level.

The main goals of this thesis were the establishment of the HDX MS technology and the study of protein dynamics and conformational changes in protein-ligand interactions. Chapter 1 describes the role of proteins in drug discovery and the general characterization of protein structures. Chapter 2 describes the optimization of the HDX MS workflow to generate robust and reproducible data. One part was to investigate how temperature can influence back exchange and the digestion efficiency of the protease pepsin. Results demonstrated that temperature conditions for digestions to obtain high sequence coverage and to keep the back exchange in the system at an acceptable level of <30% which is known from published data are around 20°C. Another aspect of this chapter was to analyze the reproducibility of HDX MS measurements. Experiments were performed to find out how well measurements of the protein phosphorylase b can be reproduced on different days. The obtained HDX MS data from the replicate measurements indicated that under defined conditions the deuterium uptake levels in all selected peptides shows minimal variation. Chapter 3 describes how ligand binding can stabilize or destabilize specific regions in the protein. The binding of ADP, GMP and the bi-substrate inhibitor AP5G to the *human* guanylate kinase (*hGMPK*) was studied. The obtained HDX MS results identified AP5G as a potent inhibitor for *hGMPK* and demonstrated that the protein undergoes large conformational changes. In the complex with AP5G the protein switched from an open to a closed conformation.

Chapter 4 investigates with HDX MS the dynamic effects of phosphoarginine and cyclomarin A binding to the N-terminal domain ClpC1 from *Mycobacterium tuberculosis*. The obtained HDX MS results showed that binding of cyclomarin A and phosphoarginine leads to rigidification of the binding sites with no effects on the overall dynamics. HDX MS data confirmed that simultaneous binding of both ligands is possible, which indicated that cyclomarin A does not influence the binding site of phosphoarginine. This means that the mode of action of cyclomarin A is not via changing the affinity of the phosphoarginine binding site, but it indicates that the antibiotic works in a different way.

Zusammenfassung

Wasserstoff-Deuterium-Austausch gekoppelt mit Massenspektrometrie (HDX MS) ist eine Methode, um die Konformationsänderungen von Proteinen in Lösung zu untersuchen. Die Austauschrate wird beeinflusst durch Änderungen in der Proteinstruktur, Interaktion mit Liganden, Protein-Protein-Wechselwirkungen und Umweltfaktoren wie pH-Wert und Temperatur. Durch strukturelle Veränderungen in einem Protein können bestimmte Regionen stabilisiert oder destabilisiert werden wodurch sich die Aufnahme von Deuterium erhöhen oder erniedrigen kann.

Die Hauptziele dieser Masterarbeit waren die Etablierung der Technologie HDX MS und die Untersuchung von Dynamiken des Proteins und Konformationsänderungen bei Protein-Ligand Wechselwirkungen. Kapitel 1 beschreibt die Charakterisierung von Proteinen im Allgemeinen und ihre Funktion in der Medikamentenforschung. Kapitel 2 beschreibt die Etablierung und Optimierung der Arbeitsschritte, um robuste und reproduzierbare HDX MS Daten zu erzeugen. Auswirkungen von Temperaturunterschieden hinsichtlich dem Rückaustausch und der Aktivität der Protease Pepsin wurden untersucht. Die Ergebnisse zeigten, dass Temperaturbedingungen um die 20°C optimal sind um eine hohe Sequenzabdeckung zu erreichen und den Rückaustausch im System auf einen akzeptablen Level von 30% zu halten der aus publizierten Daten bekannt ist. Weiteres wurde die Reproduzierbarkeit von HDX MS Messungen untersucht, indem das Protein Phosphorylase b an unterschiedlichen Tagen gemessen wurden. Die Untersuchungen zeigten, dass unter definierten Bedingungen nur minimale Schwankungen zwischen den Messungen entstehen. In Kapitel 3 wird die Methode HDX MS verwendet, um die Bindung von den Liganden ADP, GMP und AP5G an die *humane* Guanylatkinase (GMPK) zu untersuchen. Die Bindung des Inhibitors AP5G wies starke Konformationsänderungen auf. Das Protein wechselte von einer sehr offenen Struktur zu einer komplett geschlossenen Struktur.

In Kapitel 4 wird mit Hilfe von HDX MS die Bindung von Phosphoarginin und Cyclomarin A an die N-terminale Domäne ClpC1 von *Mycobacterium tuberculosis* untersucht. Die Ergebnisse zeigten, dass die Zugabe von Cyclomarin A und Phosphoarginin zu einer geringeren Aufnahme von Deuterium im Bereich der Bindungsstellen führte, jedoch keine globalen Veränderungen im Protein auslösten. HDX MS Daten bestätigten, dass eine gleichzeitige Bindung beider Liganden möglich ist. Dies zeigt, dass die Wirkungsweise von Cyclomarin A nicht die Affinität der Phosphoarginin Bindungsstellen beeinflusst.

Table of Contents

Chapter 1 Introduction	11
1. The Role of Proteins in Drug Discovery	11
2. Scope of the Thesis.....	12
2. Higher Order Structures (HOS) of Proteins.....	13
3. Methods for Characterization of Protein Structure and Dynamics	14
3.1. X-ray Crystallography	14
3.2. Nuclear Magnetic Resonance Spectroscopy	15
3.3. Optical Methods	16
3.4. Hydrogen deuterium exchange coupled with mass spectrometry (HDX MS)	18
3.4.1. Hydrogen deuterium exchange (HDX)	18
3.4.1.1. HDX Kinetics	20
3.4.1.2. Back Exchange	23
3.4.2. Mass spectrometry	23
3.4.2.1. Ion Source ESI	24
3.4.2.2. Mass Analyzers	26
3.4.3. General Workflow of HDX MS	28
3.4.4. HDX MS Data Processing and Analysis	29
3.4.5. HDX MS Application	30
Chapter 2 Quality and Reproducibility of HDX MS Data	31
1. Introduction	31
2. Materials & Methods.....	33
2.1. Experimental Part.....	35
2.1.1. Hydrogen Deuterium Exchange.....	35
2.1.2. UPLC Separation and Mass Spectrometry	35
2.1.3. Data Analysis	38
3. Results and Discussion	39
3.1. Sequence Coverage and Back Exchange	39
3.2. Reproducibility.....	44
4. Conclusion	46
Chapter 3 Characterization of the Open/Closed Conformations of Guanylate Kinase	47
1. Introduction	47
2. Materials and Methods	50
2.1. Microscale thermophoresis.....	51
2.1.1. Experimental Part.....	51

2.2. Intact Mass and HDX MS	52
2.2.1. Experimental Part	52
2.2.1.1. Intact Mass	52
2.2.1.2. Hydrogen Deuterium Exchange	53
2.2.2. UPLC Separation & Mass spectrometry	53
2.2.2.1. Intact Mass	54
2.2.2.2. Hydrogen Deuterium Exchange	55
2.2.3. Data Analysis	56
3. Results and Discussion	57
3.1. MST	57
3.2. Intact Mass	59
3.3. Sequence Coverage	59
3.4. HDX MS	60
4. Conclusion	66
Chapter 4 HDX MS Analysis of the N-terminal Domain of the ATP-dependent Clp protease ATP-binding subunit ClpC1	67
1. Introduction	67
2. Materials & Methods	70
2.1. Experimental Part	70
2.1.1. Intact Mass	71
2.1.2. Hydrogen Deuterium Exchange	71
2.2. UPLC Separation and Mass spectrometry	71
2.2.1. Intact Mass	71
2.2.2. Hydrogen Deuterium Exchange	71
2.3. Data Analysis	72
3. Results and Discussion	73
3.1. Intact Mass	73
3.2. Sequence Coverage	74
3.3. HDX MS	74
4. Conclusion	79
Chapter 5 Conclusion & Outlook	80
Appendix	81
References	82

List of Abbreviations

ACN	acetonitrile
ATP	adenosine triphosphate
ADP	adenosine diphosphate
ASM	auxiliary solvent manager
BEH	ethylene bridged hybrid
BSM	binary solvent manager
CD	circular dichroism
CID	collision induced dissociation
CyclA	cyclomarine A
DC	direct current
D ₂ O	deuterium oxide
DTT	dithiothreitol
ESI	electrospray ionization
Eqn	equation
E _{kin}	kinetic energy
E _{pot}	potential energy
FA	formic acid
FRET	Förster resonance energy transfer
GDP	guanosine diphosphate
GMP	guanosine monophosphate
GMPK	guanylate kinase
H ₂ O	hydrogen oxide
HDX	hydrogen deuterium exchange
HDX MS	hydrogen deuterium exchange mass spectrometry
K _D	dissociation constant
k _{int}	intrinsic rate
k _{op}	opening exchange rate
k _{cl}	closing exchange rate
K _{ep}	equilibrium constant
k _D	acid catalyzed exchange rate
k _{OD}	base catalyzed exchange rate
K ₂ HPO ₄	di-potassium phosphate
KH ₂ PO ₄	mono-potassium phosphate
LC	liquid chromatography

m/z	mass to charge ratio
m0%	water control
m100%	deuteration control
MALDI	matrix assisted laser desorption spectroscopy
MS	mass spectrometry
MST	microscale thermophoresis
NH	backbone amide hydrogens
NMR	nuclear magnetic resonance
NTD	N-terminal domain
PArg	phosphoarginine
PhosB	phosphorylase b
pg	picogram
RF	radio frequency
SPR	surface plasmon resonance
TCEP	tris(2-carboxyethyl)phosphine
TWEEN20	polyethylene glycol sorbitan monolaurate
Trp	tryptophan
Tyr	tyrosine
UPLC	ultra high performance liquid chromatography
UV-Vis	ultraviolet-visible
XRD	X-ray Diffraction
µg	microgram
µL	microliter
µM	micromolar

CHAPTER 1

INTRODUCTION

1. The Role of Proteins in Drug Discovery

Until the 1980s drug discovery and development in the pharmaceutical industry was focused on small molecule drugs. Major breakthroughs in molecular biology and biotechnology then made it feasible to produce recombinant proteins in large amounts. From this point biopharmaceuticals have developed into a second major area for pharmaceutical research and development. Nowadays, the research is based on a detailed understanding of molecular interactions in healthy and diseased organisms [1, 2]. During this whole process, structure-based procedures which include structure determination of the target proteins, prediction of hypothetical ligands by molecular modeling, chemical synthesis and biological testing of compounds play an integral role to discover and develop effective drugs. Furthermore, a deeper understanding of the relationship between structure and function is extremely useful for drug efficacy and safety. In this research area, a number of techniques are available to study high order structures (HOS) of proteins which include fundamental methods like nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography. However, only a few provide a comparative insight into protein dynamics. One of them is hydrogen deuterium exchange (HDX) coupled with mass spectrometry (MS) and has been emerged as an analytical tool in drug research and development for studying protein dynamics, protein folding, protein-protein interactions, and protein small molecule interactions in solution. In the pharmaceutical industry HDX MS has become a widely accepted analytical tool for the characterization of biological therapeutics, e.g. for mapping of antibody drug epitopes. HDX MS has the potential to play an important role in understanding the link between protein structure and activity [3-5].

2. Scope of the Thesis

The focus of this master thesis is the establishment and optimization of the method HDX MS to obtain high-quality and information-rich data and the investigation of protein dynamics and conformational changes of protein ligand interactions for potential drug targets.

In Chapter 2, the digestion step in an HDX MS workflow with different temperature conditions is analyzed to find out how temperature can influence the back exchange (loss of deuterium) and the digestion efficiency of the protease pepsin. The goal is to find the right balance between sequence coverage and back exchange for obtaining high-quality and information-rich data. Another aspect of this chapter is to analyze the reproducibility of HDX MS measurements. It is important to find out how well measurements can be reproduced with identical samples on different days, to ensure high data quality.

In Chapter 3, the effects of the binding of three different ligands ADP, GMP and AP5G to *human* guanylate kinase is studied using HDX MS. The goal is to find out how ligand binding can stabilize or destabilize certain areas of the protein which results in a decrease or increase of HDX rates.

In Chapter 4, the dynamic effects of phosphoarginine and cyclomarin A binding to the N-terminal domain of the protein ClpC1 from *Mycobacterium tuberculosis* are investigated on a HDX timescale which ranges from seconds to hours. The goal is to analyze how ligand binding influences the dynamics of the protein and if simultaneous binding of both ligands influences the binding site of phosphoarginine.

2. Higher Order Structures (HOS) of Proteins

Proteins are large biological molecules and are classified more generally as macromolecules or polymers. They are composed of 20 different naturally occurring monomer units called amino acids which are linked together chemically to form a peptide bond. This peptide bond modulates an element which is found in all proteins called the polypeptide backbone or chain. In general, all amino acids have the same backbone structure and a so-called side chain, which differ in size, shape, charge, hydrophobicity and reactivity (Figure 1). The characteristic chemical composition, in a protein, is the key element for its structural complexity which is associated with its functionality [6].

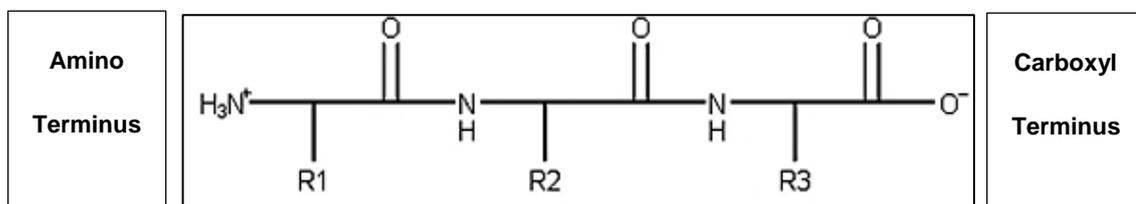


Figure 1: Primary structure. The linear sequence of amino acids, from the amino terminus to the carboxyl terminus (Image, MarvinSketch).

To understand how a protein gets its final shape and conformation, a fundamental knowledge of the four hierarchical structure levels which includes the primary, secondary, tertiary and quaternary structure is needed (Figure 2). The primary structure of a protein is the linear arrangement of amino acid residues that forms the polypeptide chain. The next level of protein structure is the secondary structure where parts of the polypeptide chain can assume several spatial arrangements. The most common types of secondary structure elements are the α helix and the β pleated sheet. These structures are stabilized through hydrogen bonds between the carbonyl oxygen of one amino acid and the amide proton of another. The next higher level of structure is the tertiary structure, the three-dimensional arrangement of secondary structural elements like helices, sheets, hairpins and loops. This structure is stabilized primarily by interactions between the side chains (R groups). R group interactions include hydrogen bonding, ionic bonding, dipole-dipole interactions, London dispersion forces and in some cases protein disulfide bonds [7].

For proteins with a single polypeptide chain, the tertiary structure is the highest-level of organization. Proteins with multiple polypeptide chains, also known as subunits, are organized in so called quaternary structures. In general, these structures are also stabilized by the same types of interactions that contribute to tertiary structure [3, 6].

3. Methods for Characterization of Protein Structure and Dynamics

In the pharmaceutical industry structural characterization of proteins to have a better understanding of structure-function relationship is of great importance. There are numerous fundamental methods for protein characterization available. Each method, however, has its advantages and limitations. A brief discussion of the common methods nuclear magnetic resonance spectroscopy (NMR), X-ray, UV-VIS, CD and fluorescence spectroscopy is given in the following paragraphs. Furthermore, the method hydrogen deuterium exchange coupled with mass spectrometry (HDX MS) will be discussed in more detail and is the main subject of this master thesis.

3.1. X-ray Crystallography

X-ray crystallography represents the gold standard for solving the 3D structure of proteins with atomic resolution. The data are collected by scattering X-rays from a single crystal which has an ordered arrangement of atoms [8]. By measuring the angles and intensities of these diffracted beams (X-rays), a 3D picture of the electron density within the crystal can be produced. With this electron density image, the mean positions of the atoms in the crystal can be determined (Figure 2). X-ray crystallography provides high resolution and very detailed information about the structure [9].

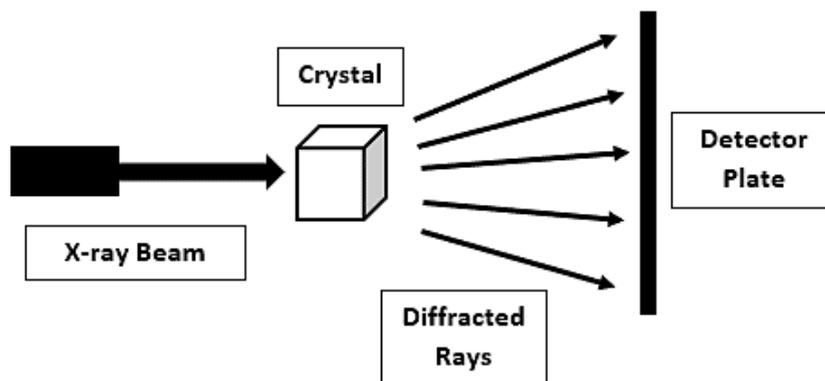


Figure 2: Schematic representation of the method X-ray crystallography. X-rays are fired onto a crystal and produce a pattern which is used to determine the arrangement of atoms inside the crystal.

The atomic coordinates derived from X-ray crystallography are usually accompanied by displacement parameters also known as B factors. These B factors describe the molecular disorder in the crystal which gives some information on the dynamics of the protein. But dynamic information on protein crystals is of limited value because proteins are not static structures and they usually exist as ensembles of different conformational states in solution. In X-ray crystallography, proteins are mostly modelled as a single conformation. The growth of crystals which are suitable for X-ray analysis can be challenging because not all proteins will crystalize readily and large amounts of purified proteins are needed [10].

3.2. Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is used in quality control and research to determine the content and purity of a sample as well as to determine its molecular structure. It is used to study physical properties at the molecular level such as conformational changes, phase changes, solubility and diffusion. In general, this technology is based on the interaction of nuclear spins when they are placed in an external magnetic field. In this strong magnetic field, the nuclei absorb particular radio frequencies [11]. For a nucleus to be NMR visible, it must have a non-zero spin. For biomolecules, the most important nuclei are H, O, C, N and P. NMR is a powerful technique to determine protein structures with high resolution and to provide detailed information about the structure and dynamics of protein-protein and protein-ligand interactions but this technique also has limitations.

NMR requires a relatively high concentration of protein what in some cases can lead to aggregation. Furthermore, NMR often requires isotopically labeled proteins and the measurement of proteins > 25kDa can become very difficult [12].

3.3. Optical Methods

Optical methods are widely used to study protein structures which include techniques like UV-Visible (UV-VIS) absorption spectroscopy [13], circular dichroism (CD) [14] and fluorescence spectroscopy [15].

In UV-VIS the absorption maxima for proteins are usually between 275 and 280nm which are caused by the absorbance of tryptophan (Trp), tyrosine (Tyr) and cysteine (disulfide bonds). The absorbance depends on the environment of their chromophores. In absorbance measurements the concentration of proteins can be easily and exactly determined. Furthermore the absorbance differences between the native and the unfolded forms of a protein can be investigated and are extremely useful for monitoring conformational changes of a protein [13]. Absorbance A is defined by:

$$A = -\log_{10}(I/I_0)$$

Eqn. 1

where I_0 is the intensity of the light before and I the intensity of the light after the passage through the protein solution. The absorbance A depends linearly on concentration, corresponding to the Lambert Beer law and is defined as:

$$A = \epsilon C d$$

Eqn. 2

where c is the concentration, d the path length of the cuvette and ϵ the molar absorption coefficient [13].

In CD spectroscopy the absorption differences of left and right circularly polarized light are measured. Characteristic CD signals are generated by different structures. This method is used to evaluate secondary structure, folding and binding properties of proteins. The advantage of CD is that data can be collected and analyzed in a few hours on solutions of samples containing only small amounts of protein in physiological buffer. However, CD has limitations in providing information about the specific location where the local conformational changes may occur. Further, there are limitations in the detection of tertiary structures [14].

Another optical method is fluorescence spectroscopy which is routinely used to analyze protein structures. In general, the proteins require a fluorescent label. If the protein of interest contains tryptophan its auto-fluorescence can be used for measurements but in many cases a specific fluorescence label has to be introduced. For probing molecular structures, the fluorescence effect Förster resonance energy transfer (FRET) is very useful. FRET depends on the molecular distances between a donor fluorophore and an acceptor chromophore. The distance R_0 for most FRET donor/acceptor pairs is between 1–7 nm. For measurements the FRET efficiency E is defined by:

$$E = \frac{I_A}{I_D + I_A}$$

Eqn. 3

where I_A and I_D are the total fluorescence intensities of the donor and acceptor. This powerful technique obtains specific and real-time information of biological processes. However, there are also limitations in the detection of tertiary structures [15].

3.4. Hydrogen deuterium exchange coupled with mass spectrometry (HDX MS)

The difficulties associated with characterization of protein structures and dynamics and their limitations in both high resolution and low resolutions methods necessitate the development of new methodologies. An alternative which could fill these needs can be hydrogen deuterium exchange (HDX) coupled with a suitable detection technology like mass spectrometry (MS). On its own, MS is unable to probe tertiary or quaternary structure directly like X-ray crystallography or NMR. Instead, MS relies on reagent molecules that can react with certain solvent-accessible sites on the protein and can measure the mass changes which are induced by different modifications. It can be used to measure solvent accessibility or the dynamics of labeled sites which can give an insight into changes of tertiary structures [16]. In general, HDX MS is not method for structure determination but it can be used to characterize protein structures [17].

3.4.1. Hydrogen deuterium exchange (HDX)

Hydrogen deuterium exchange (HDX) has emerged as a powerful tool for studying protein dynamics, protein folding, protein-protein interactions, and protein small molecule interactions in solution. The conformational dynamics of the protein structure in solution can be measured as a function of time ranging from seconds to hours [18].

In the early 1950s Kaj Ulrik Linderstrøm-Lang performed the first HDX experiments. He labelled proteins with D₂O and made measurements of protein deuteration by using density gradient tubes [19]. Subsequently, a number of different detection methods including UV-VIS, IR and NMR spectroscopy have been described [17]. Since the early 1990s the use of mass spectrometry (MS) has increased dramatically. Therefore, MS is now by far the most common detection method for HDX experiments [20].

In general, HDX is a method in which deuterium atoms present in buffer replace hydrogen atoms in the protein. From all hydrogens in a protein polypeptide, only hydrogen atoms in O-H, N-H and S-H groups can be replaced with deuterium atoms. With MS, only the mass shift of the exchangeable hydrogens in the amide linkage can be monitored because all other exchangeable hydrogens exchange too rapidly during sample handling (Figure 3).

Measurements can be performed at the intact protein level to study the overall deuterium incorporation or at the peptide level where the protein is enzymatically digested with a protease to obtain localized exchange information. This rate of deuterium incorporation can characterize the nature of a protein structure because the conformation and dynamics of a protein is affected by the amount of incorporation [23].

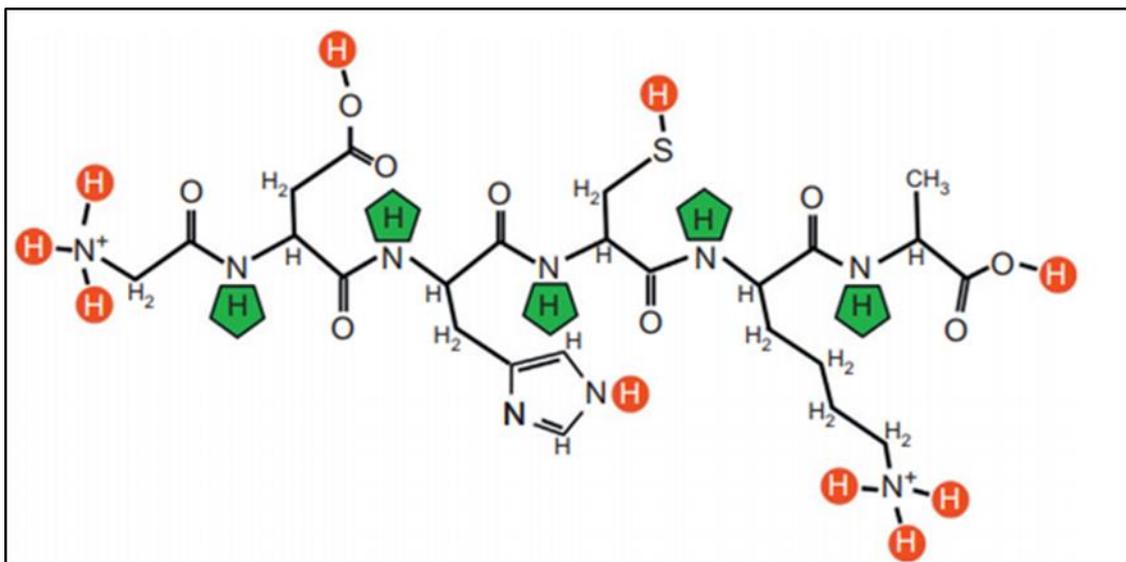
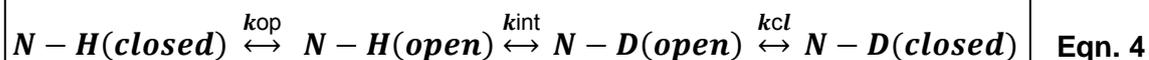


Figure 3: Exchangeability in an example peptide: glycine-aspartic acid-histidine-cysteine-lysine-alanine. Green: measurable backbone amide hydrogens Red: hydrogens that are not monitored with HDX MS due to their very rapid exchange rates. Black: numerous hydrogens that do not exchange due to their covalent binding with carbon. (Image, MarvinSketch)

3.4.1.1. HDX Kinetics

The deuterium incorporation for each peptide fragment over time yields exchange kinetics. This kinetics contains information about the stability of backbone amide hydrogens (NHs). Different backbone NHs can undergo exchange at different rates depending on factors like solvent accessibility, hydrogen bonding, temperature and pH [21]. The NH exchange in native folded proteins is the result of a combination of solvent accessibility and intramolecular hydrogen bonding. On the one hand, NHs in beta-sheets and alpha-helices are protected from the hydrogen deuterium exchange because the hydrogens are involved in hydrogen bonds within the secondary structure elements. On the other hand, loops and unstructured regions exchange relatively fast because they may be highly exposed to the solvent and their amide protons are not usually involved in hydrogen bonds [22, 23].

The overall backbone HDX exchange mechanism for native proteins under continuous labelling conditions can be described as:



where k_{op} , k_{cl} and k_{int} are opening, closing and intrinsic exchange rate constants for the backbone amide hydrogens and each amide hydrogen in a protein can have a unique combination of these constants [24].

This scheme leads to a measured exchange rate, k_{ex} , can be described as:

$$k_{ex} = \frac{k_{op}k_{int}}{k_{cl}+k_{int}+k_{op}} \quad \text{Eqn. 5}$$

In HDX kinetics there are two distinct HDX regimes called EX1 and EX2 which are commonly identified depending on the closing rate (protein refolding) (k_{cl}) and intrinsic exchange rate (k_{int}). EX1 is not common at physiological pH and it implies that the intrinsic rate of exchange is much greater than the closing rate ($k_{int} \gg k_{cl}$). The exchange mode can be promoted by denaturants and high temperatures [25]. The exchange rate constant (k_{ex}) can be defined as:

$$k_{ex} = k_{op}$$

Eqn. 6

In the EX1 regime, amide hydrogens exposed to solvent exchange immediately. EX1 mechanisms can be identified by the isotope peak distribution, which shows a bimodal mass distribution corresponding to a highly deuterated and a less deuterated species [18, 20].

EX2 is very frequently observed for nearly all proteins under physiological conditions or neutral pH in the absence of denaturants and is therefore the most common regime. This regime implies that the closing rate is much faster than the intrinsic rate of exchange ($k_{cl} \gg k_{int}$) [20, 26]. The exchange rate (k_{ex}) can be defined as:

$$k_{ex} = K_{eq}k_{int}$$

Eqn. 7

where K_{eq} is the equilibrium constant between the open and the closed state of the protein and is given by k_{op}/k_{cl} . The EX2 mechanism can be identified by isotope peak distribution; it shows a binomial mass distribution corresponding to the exchange. It leads to an increase in the average mass while the isotopic distribution remains nearly the same [26].

An important parameter that can also affect HDX is temperature. With the increase of temperature, the exchange rate increases exponentially. By decreasing the temperature from room temperature to 0 °C, the reaction slows by almost an order of magnitude. Therefore, it is important to have a temperature of 0°C after quenching to avoid the deuterium loss in the system [23]. The exchange kinetics for fully exposed amides is described as intrinsic rate (k_{int}) which depends on factors like temperature and can be determined by the Arrhenius equation. This is defined as:

$$\ln k_{int} = \ln A - \frac{E_a}{RT}$$

Eqn. 8

where A is the pre-exponential factor, E_a the activation energy, R the gas constant and T the temperature [27].

The pH is also critical because it can influence the exchange rate of backbone NHs enormously. The reaction can be catalyzed by acid (D30 +) or by base (OD -) [19, 28]. The rate constant can be described as:

$$k_{int} = k_H[D^+] + k_{OD}[OD^-]$$

Eqn. 9

where k_D and k_{OD} are rates of acid and base catalyzed exchange. From HDX MS publications it is known that the exchange rate has a minimum at pH 2.4 - 3. For example, at neutral pH (pH 7.0) the exchange rate is approximately 10,000 times faster than at pH 2.4. Therefore, the pH from quenching onwards should be at the minimum exchange rate to reduce the back exchange in the system [24].

3.4.1.2. Back Exchange

A major problem in HDX experiments is back exchange. The exchanged deuterium can exchange back to hydrogen because after the labelling step the experiments are performed in aqueous solution. Therefore, a successful HDX experiment requires careful design for minimizing the back exchange in the system because any loss in the original deuterium level means a loss of information on the dynamics of the protein under investigation. For this reason, it is essential to work as fast as possible at low pH and temperature after the quenching step [23]. Throughout this thesis, conditions at 0 °C and pH 2.4 are used for desalting and separation. For digestion, conditions between 15-20°C and pH 2.4 are used and will be discussed in more detail in Chapter 2.

3.4.2. Mass spectrometry

Mass spectrometry (MS) was used as the detection method for HDX. Therefore, the next paragraph will describe in more detail the instrumental set up and experimental approaches in mass spectrometry.

In general, MS requires that the proteins or peptides in solution or solid state be turned into an ionized form in the gas phase before they are analyzed in an electric or magnetic field (Figure 4). In proteomic research there are two primary ionization techniques to transfer the protein analytes into the gas phase, which include matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). Compounds with sizes ranging from a few Daltons (Da) to mega Daltons (MDa) can be investigated with this ionization technique [29, 30].

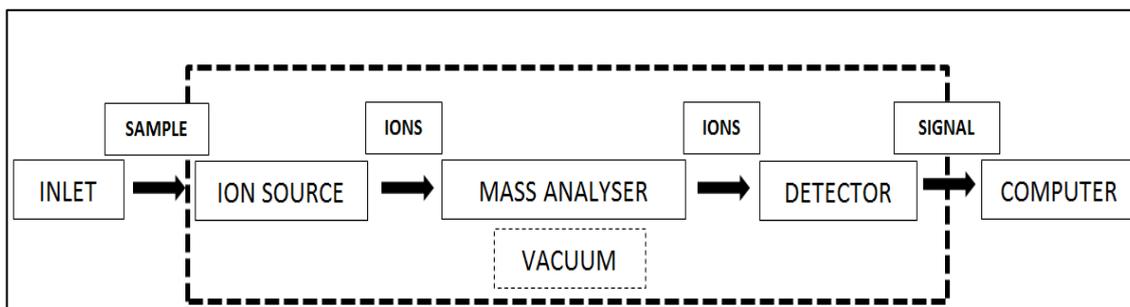


Figure 4: Schematic representation of a mass spectrometer. Ion source for producing the protein/peptide ions, mass analyser for separation and a detector for the registration of the ions.

After the ionization step, several mass analyzers can be employed. Mass analyzers like quadrupole mass filters, quadrupole ion trap, time of flight (TOF) and Fourier-Transform ion cyclotron resonance (FTICR) can be integrated into the instrument. Each mass analyzer addresses the need of a specific application like mass accuracy, resolution and sensitivity. In modern instruments it is common to use a combination of two mass analyzers (Tandem mass spectrometry) like a quadrupole-TOF (Q-TOF) or magnetic sector-TOF [31].

The ionization technique ESI and the mass analyzer quadrupole and TOF will be discussed in more detail because they are directly relevant for the experiments of this thesis.

3.4.2.1. Ion Source ESI

The "soft" ionization technique ESI transfers proteins or peptides from solution into the gas phase. ESI occurs at atmospheric pressure and uses electrical energy to assist the transfer of ions from solution into the gaseous phase. For charge separation where the electrons are removed from solution the voltage in positive mode is around 3kV. The major steps are: the production of charged droplets from the high-voltage capillary tip, evaporation of the solvent from the charged droplet followed by a droplet breakdown into much smaller droplets so called Columbian explosion and lastly the transfer of the ion into the gas phase. The potential difference between the tip of the capillary and the cone of the mass spectrometer ensures that the ions are effectively transferred into the MS system (Figure 5) [30, 31].

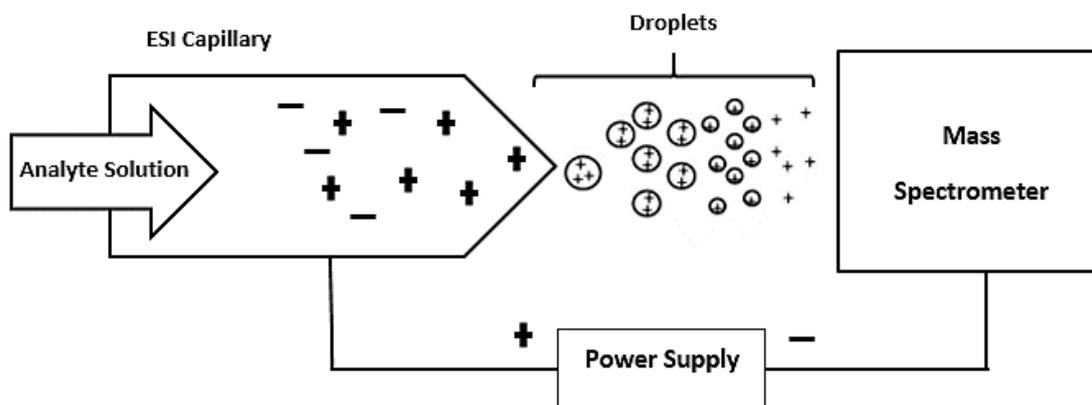


Figure 5: Schematic representation of the soft ionization technique ESI operated in positive mode.

For proteomic research, the ionization technique ESI has several advantages over MALDI. ESI produces multiply charged gaseous ions from solution phase samples therefore it can be easily coupled to a separation technology like Liquid chromatography (LC) [30, 32]. In contrast to ESI, MALDI ionizes proteins and peptides out of a dry, crystalline matrix and produces singly charged ions. This method is generally used to analyze simple peptides, while LC ESI MS systems are preferred for the analysis of more complex samples [29].

3.4.2.2. Mass Analyzers

The mass analyzer is used to separate charged analyte ions in a magnetic or electric field based on their mass to charge ratio (m/z) where m is equal to the mass of the ion and z is equal to the charge.

The most common mass analyzer is a quadrupole. The quadrupole consists of four parallel metal rods and each rod pair is connected together electrically (Figure 6) [33]. On one hand, the quadrupole can act as an ion guide, where a radio frequency (RF) voltage is applied and all generated ions from the ESI source can be transmitted to the detector. On the other hand, it can act as a mass filter where a combination of direct current (DC) voltage and radio frequency (RF) voltage is applied. Only ions with a certain m/z value can pass through the analyzer and do not collide with the rods on their unstable trajectories (Figure 6) [30].

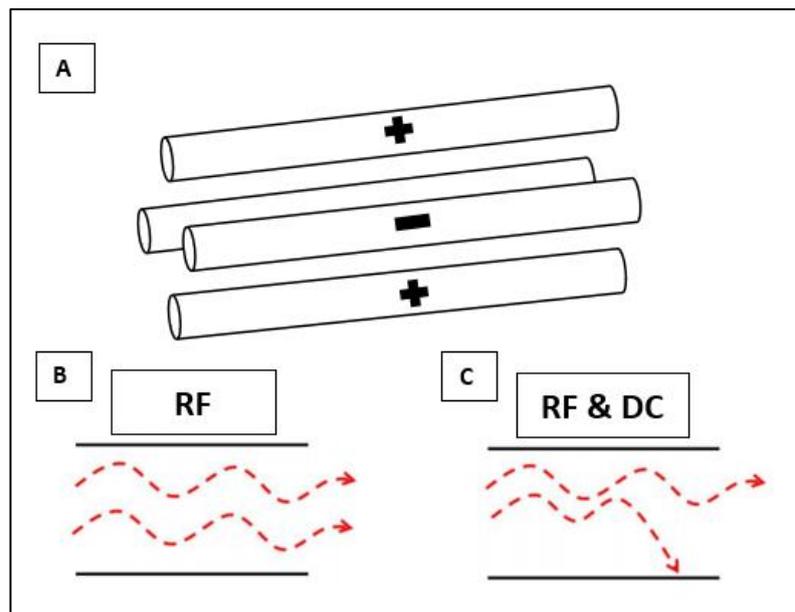


Figure 6: Schematic representation of a Quadrupole Mass Analyzer. A: Assembly of 4 parallel metal rods, B: RF only mode in which all ions can pass, C: DC and RF mode in which only ions with certain m/z can pass

Time of flight (TOF) mass analyzers can separate ions of different m/z values via their flight time in a field-free region called the flight tube. Ions are accelerated in the flight tube by an electric field where the potential energy of the ion is converted into kinetic energy. This results in ions with the same charge by the same kinetic energy. At the end of the analyzer a detector measures the arrival time of the ions. The arrival time depends on the velocity of the ion which means that ions with lower m/z arrive first then ions of greater m/z . The arrival time (t) for the ions can be calculated as:

$$E_{\text{pot}} = E_{\text{kin}}$$

Eqn. 10

$$t = \sqrt{\frac{m}{e} \frac{1}{z^2 U}}$$

Eqn. 11

where U , d , m , z and e represent the acceleration voltage, length of the flight tube, mass of the ion, charge state of the ion and velocity of the ion [34].

3.4.3. General Workflow of HDX MS

Over the last years different HDX MS protocols have been developed. The most common technique is “continuous labeling” where the native protein is exposed to a deuterated buffer solution and the deuterium incorporation is monitored over time [20]. A typical HDX MS experiment workflow with continuous labeling is illustrated in Figure 7.

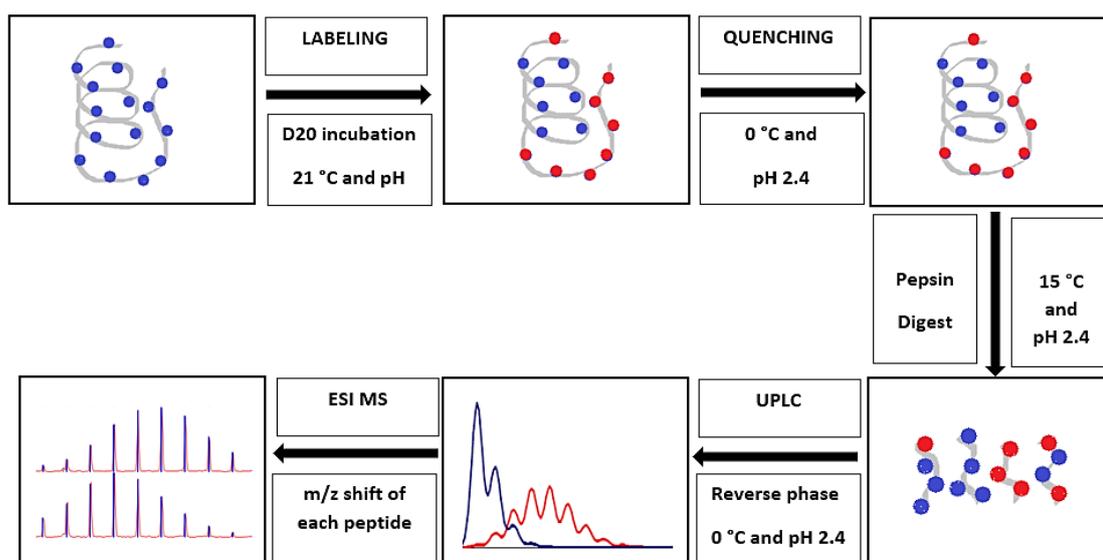


Figure 7: Schematic representation of a HDX MS workflow. Blue color symbolizes hydrogen (H_2O) and the red color symbolizes deuterium (D_2O).

In general, the target protein is purified and supplied in standard buffer. The protein is incubated with a 10-20-fold excess of D_2O at room temperature (21 °C) and pH 7.0 for a specific labelling time ranging from seconds to several hours. After time-dependent deuterium incorporation the reaction is quenched to 0°C and pH 2.4 to minimize the back exchange in the system [23]. After the labelling step the protein is again in aqueous buffer and unwanted back exchange of deuterons to protons will happen. After quenching the labeled protein is enzymatically digested with pepsin at a temperature of 15°C and a pH 2.4. Pepsin is a preferred protease because of its maximal activity at low pH and temperature [35]. The resulting peptides are retained and desalted on a trapping column before they are transferred onto a reversed-phase column for separation at 0°C.

The peptides are usually eluted by using a 0–60% acetonitrile gradient. Subsequently, ESI MS is used to record the mass to charge ratio (m/z) of the peptide ions at each incubation time point [30]. A non-deuterated control with H_2O instead of D_2O is also performed to determine with MS the natural isotope distribution pattern of all peptides.

3.4.4. HDX MS Data Processing and Analysis

For HDX the data must be corrected for natural isotope distribution (e.g., ^{13}C , ^{15}N) and for back exchange which means the loss of deuterium during the experiment. Therefore, controls with 0% and 100% D_2O are integrated into an experiment [18, 28]. The protein for the 0% control is incubated with H_2O instead of D_2O to determine the natural isotope distribution pattern. For the 100% control the protein is incubated with D_2O to incorporate all possible backbone amides with deuterium, to measure the maximum amount of deuterium incorporation. For data analysis a chromatogram sums up the MS intensities of hundreds of peptides. The mass spectra show the isotope envelope pattern for each peptide. The shift in the isotope distribution pattern to a higher mass number describes the increase in deuterium incorporation over time for all peptides. Subsequently the centroid of each isotope distribution is determined. The calculation of the number of deuterons (D) which are incorporated into each peptide at each time point can be described as:

$$D = N \frac{(m_t - m_{0\%})}{(m_{100\%} - m_{0\%})}$$

Eqn. 12

where N describes the total number of exchangeable amide hydrogens corrected for proline residues and the N-terminal amide [36]. The $m_{0\%}$, $m_{100\%}$ and m_t are the centroid masses for the 0% control, 100% control and the incubation time [18].

The corrected amount of deuterium (D) in each peptide is then plotted versus time and fit to the sum of first-order rate. It can be described as:

$$D = N - \sum_{i=1}^N \exp(-k_i t)$$

Eqn. 13

where N is the number of amide protons that exchange at a given rate constant k_i , during the time (t) allowed for isotopic exchange [18].

The software DynamX (Waters) uses the experimental data to perform theoretical isotope fits using a set of algorithms. The software requires the input of protein sequence, the peptide list along with the respective retention times and MS data for undeuterated and deuterated samples from each time point [37].

3.4.5. HDX MS Application

HDX MS has emerged as a powerful tool for the development process of small molecule therapeutics and biological pharmaceuticals. Through the development of a fully automated HDX MS platform which allows rapid acquisition and processing of experimental data, HDX MS has become more and more interesting for the pharmaceutical industry to bring certain drugs successfully, faster and safer to the market. It is envisaged that HDX MS will play an increasing role for understanding the link between protein structure and activity [5, 22].

From a broader perspective, HDX MS will continue to act as a complementary technique to other structural techniques like X-ray crystallography and NMR to handle challenging questions in protein science which require multiple approaches. For these reasons there are exciting times for the technique HDX MS and the field is still growing in academia and the industrial sector [4].

CHAPTER 2

QUALITY AND REPRODUCIBILITY OF HDX MS DATA

1. Introduction

There are several key components including labelling, quenching, digestion, separation and m/z detection in a typical HDX MS workflow. To generate robust and reproducible HDX MS data it is important that all parameters like temperature, pH and runtime are fully optimized. Figure 8 shows that for each step individual parameters can be chosen to maximize the quality and information content of the data. After the labelling step the exchange is quenched with aqueous buffer and unwanted back exchange of deuterons to protons will happen, which means that deuteration is not a one-way street and back exchange will occur at a slow rate even when the reaction is quenched. Any loss in the original deuterium level means a loss of information on the dynamics of the protein under investigation. For this reason, it is essential to minimize the back exchange as much as possible. As shown in Chapter 1, HDX is drastically slowed down at low pH and temperature. Therefore, the conditions used in all HDX MS publications from quenching onwards are pH 2.4 and a temperature of 0°C for desalting and separation. For the digestion step, the preferred protease is pepsin because it is fully active at pH 2.4. However, it is 'designed' by nature to optimally work at a temperature of 37°C. Therefore, a challenge associated with HDX MS is to find optimal conditions in the digestion step. On the one hand, it is important for HDX MS to obtain very high sequence coverage for peptide mapping to get detailed information and avoid the loss of some potentially interesting regions in the protein. But on the other hand it is also important to work as fast as possible at low pH and temperature to minimize the back exchange in the system. This chapter describes the optimization of the digestion step and how temperature can influence both back exchange and the digestion efficiency of the protease pepsin. The goal was to find the right balance between sequence coverage and back exchange for obtaining high-quality and information-rich data.

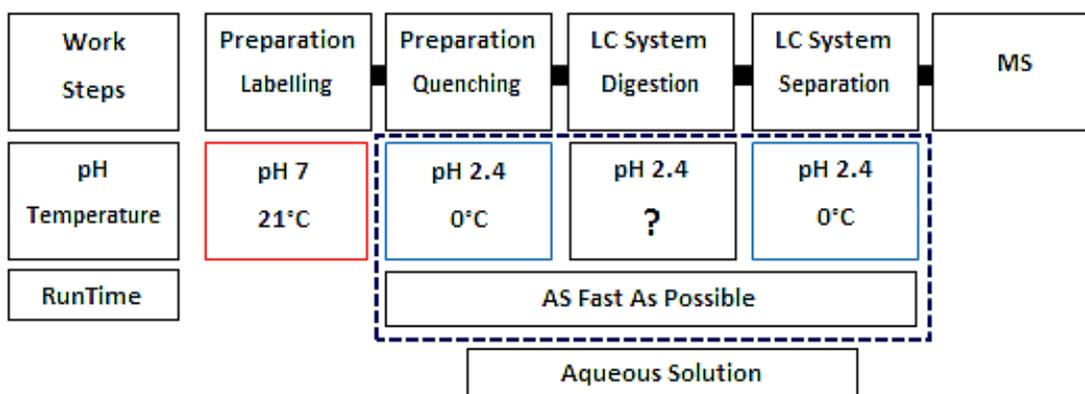


Figure 8: Defined parameters for a typical HDX MS workflow. Labelling: time depended incubation at pH 7 and room temperature. Quenching: reaction is quenched to pH 2.4 and 0°C to stop the labelling process and avoid back exchange. Digestion: specific protease which is highly active at pH 2.4 and low temperature. Separation: use a UPLC system to work as fast as possible at pH 2.4 and 0°C to avoid back exchange.

Another aspect of this chapter was to analyze the reproducibility of HDX MS measurements. Experiments were performed to find out how well measurements can be reproduced with identical samples on different days. Variation of data can be caused by the experimental procedure, by any component of the measurement apparatus and even from the sample itself. For the direct comparison of two or more data sets conditions like pH, temperature and runtime should be kept as constant as possible, because variation can influence the reproducibility of HDX MS data enormously. As an example, HDX rates slow down by a factor of ten for a drop of pH by one unit [38]. Each process step must be reviewed for potential problems that influence the sample or the measurement of HDX MS data [39].

The pioneering work of the John Engen group, Boston [38], established phosphorylase b as standard protein for the optimization and evaluation of HDX MS systems. Phosphorylase b was therefore used for testing the efficiency of deuteration, degree of back exchange, digestion efficiency, separation and reproducibility by analyzing several phosphorylase b peptides.

2. Materials & Methods

Chemicals, materials and devices which were used in nearly every HDX MS workflow during this master thesis are listed in Table 1 and Table 2.

Table 1: General chemicals for an HDX MS workflow.

Chemicals	Company	Comments
Phosphorylase b from rabbit muscle	Waters	UniProt P00489
Predigested phosphorylase b	Waters	Tryptic Digest
Deuterium Oxide, (D ₂ O)	Cambridge Isotope Laboratories	D, 99.96%
KH ₂ PO ₄ ACS grade ≥ 99%	Merck KGaA	MW 136.09 g/mol
K ₂ HPO ₄ ACS grade ≥ 99%	Merck KGaA	MW 174.18 g/mol
Sodium deuterioxide (NaOD)	Cambridge Isotope Laboratories	D, 99.50%
Deuterium chloride (DCI)	Cambridge Isotope Laboratories	D, 99.50%
Hydrogen chloride (HCl)	Merck KGaA	
Water (H ₂ O)	Thermo Fisher Scientific	LC-MS compatible
Acetonitrile (ACN)	Thermo Fisher Scientific	LC-MS compatible
Methanol (MeOH)	Thermo Fisher Scientific	LC-MS compatible
Isopropanol	Thermo Fisher Scientific	LC-MS compatible
Formic acid	Merck KGaA, Darmstadt	LC-MS compatible
Leucine Enkephalin (LeuEnk)	Waters	Standard
Glu-1-Fibrinopeptide B (GluFib)	Waters	Standard

Table 2: General Materials and Devices for an HDX MS workflow.

Materials	Company	Comments
ACQUITY UPLC Protein BEH C4 Column, 300Å, 1.7 µm, 2.1 mm X 50 mm	Waters	Analytical column
ACQUITY UPLC BEH C18 1.7µm, 1mm x 100mm column	Waters	Analytical column
ACQUITY UPLC® BEH C18 VanGuard 1.7µm, 2.1 x 5.0mm pre-column	Waters	Trapping column
ACQUITY UPLC Protein BEH C4 Column, 300Å, 1.7 µm, 2.1 mm X 50 mm	Waters	Digestion column
Devices	Company	Comments
nanoACQUITY UPLCTM	Waters	LC System
Synapt G2-Si HDMS	Waters	MS System
Robot, Autosampler	Leap	
Sonorex Ultrasonic baths	Bandelin	
Centrifuge 5430	eppendorf	

Eluents and washes which were used for the LC system and the Leap Robot during this master thesis are listed in Table 3.

Table 3: Eluents and washes for the LC system and Robot. ASM: Auxiliary Solvent Manager, BSM: Binary Solvent Manager, LEAP: robot system for HDX experiments

Application	Buffer
AUX A1	Water + 0.2% Formic Acid
nanoAcquity BSM A1 (Pump)	Water + 0.2% Formic Acid
nanoAcquity BSM B1 (Pump)	Acetonitrile + 0.2% Formic acid
Seal Wash	90% Water : 10% Acetonitrile
Wash solvent 1 LEAP	Water + 0.2% Formic acid
Wash solvent 2 LEAP	25% Methanol : 25% Acetonitrile : 25% Isopropanol : 25% Water

2.1. Experimental Part

The intact standard protein phosphorylase b (Sequence, Appendix) had a concentration of 0.62mg powder per vial. The protein is 843 amino acids long and the theoretical molecular mass is 97,289Da.

2.1.1. Hydrogen Deuterium Exchange

A stock solution of intact phosphorylase b was prepared to a concentration of 8 μ M with potassium phosphate buffer [5mM KH₂PO₄, 5mM K₂HPO₄ in H₂O, pH 7.00]. Further, phosphorylase b was diluted 16-fold with potassium phosphate buffer in H₂O [5mM KH₂PO₄, 5mM K₂HPO₄, pH 7.00], for reference experiments (no deuteration), or a potassium phosphate buffer in D₂O [5mM KH₂PO₄, 5mM K₂HPO₄ in D₂O, pD 6.6], for deuterated experiments. Measured pH values of D₂O solutions were all adjusted to the corresponding pD values by using the equation: pD = pH_{read} + 0.40 [40]. In the next step, the samples were incubated at room temperature (21°C) for different time periods. Time range for the undeuterated experiment was 0min and for the deuterated experiment 2 and 120minutes (min).

At the end of the incorporation time, the samples were quenched 1:1 with ice-cold potassium buffer [50mM KH₂PO₄, 50mM K₂HPO₄ in H₂O, pH 2.4]. To avoid back exchange, it was important to make sure that the solution has exactly pH 2.4 after quenching.

2.1.2. UPLC Separation and Mass Spectrometry

The digestion takes place at 10, 20 and 37°C and the peptide desalting and separation at 0°C. Phosphorylase b with a concentration of 10pmol in 50 μ L (loop volume) was passed through the BEH pepsin column, at 70 μ L/min for 3min in H₂O with 0.2% formic acid, pH 2.4 for digestion. The generated peptides were desalted in the HDX manager by using an BEH C18 pre-column. In the next step the peptides were eluted and separated in an BEH C18 column at 0°C. Separation was performed with an acetonitrile gradient containing 0.2% formic acid at 40 μ L/min (Table 4).

Table 4: Gradient for HDX. A (%) H₂O with 0.2% FA, B (%) ACN with 0.2% FA.

Time [min]	Flow rate [μ l/min]	% A	% B
Initial	40.0	95.0	5.0
6.0	40.0	65.0	35.0
7.0	40.0	60.0	40.0
8.0	40.0	5.0	95.0
10.0	40.0	5.0	95.0
11.0	40.0	95.0	5.0
12.0	40.0	5.0	95.0
13.0	40.0	95.0	5.0

A schematic representation of the fluidics connections in the LC system for digestion, desalting and separation is shown in Figure 9 and 10.

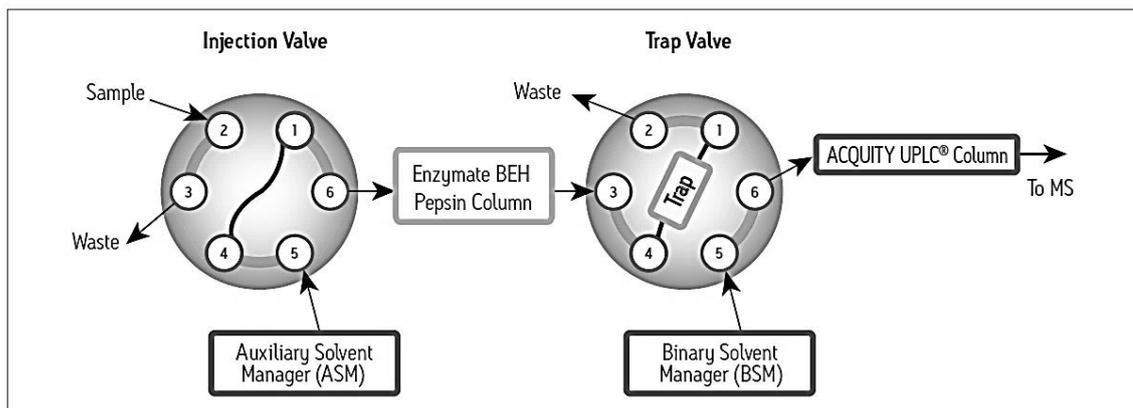


Figure 9: Protein digestion in trapping mode. The pump system from the Auxiliary Solvent Manager (ASM) delivers mobile phase [0.2% formic acid in water, pH2.5] for pepsin digestion, trapping and desalting. The pump system from the Binary Solvent Manager (BSM) drives the reversed phase gradient [0.2% formic acid in water and 0.2% formic acid in acetonitrile] for peptide separation (Image Waters, <http://www.waters.com>).

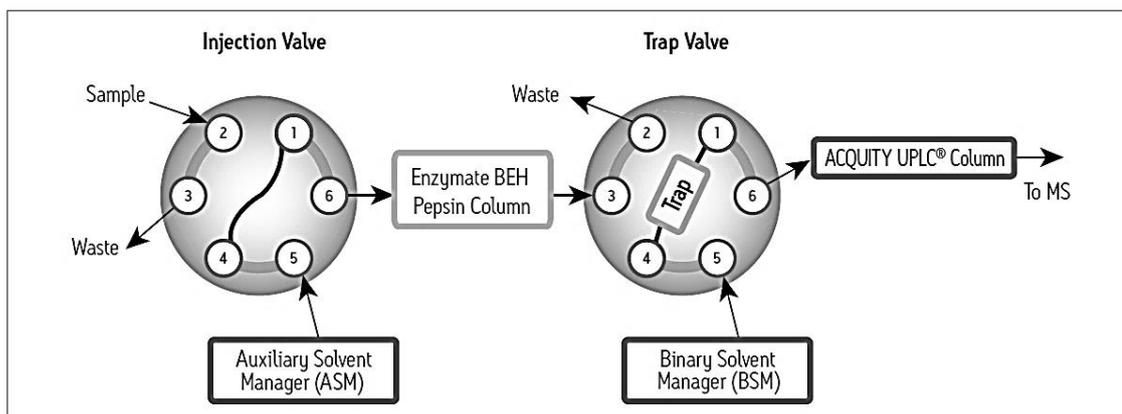


Figure 10: Separation mode. After proteins are digested on the Enzymatic BEH Pepsin Column, the valve switched, and the peptic peptides are separated using a reversed-phase column (Image Waters, <http://www.waters.com>).

From the LC system, the eluent was introduced into a Synapt G2-Si Mass spectrometer (Waters). Ionization was performed with Electrospray Ionization and the peptide leucine enkephalin (LeuEnk) with a concentration of 50pg/ μ L (flowrate: 10 μ L/min) was used for lock mass correction. Calibration with leucine enkephalin was performed each day to ensure high mass accuracy.

The source and desolvation temperatures of the instrument were set to 80 and 250°C, cone voltage 30V, and the ESI voltage 3kV. MS conditions were ESI positive, resolution mode and the mass spectra were acquired in MS_E mode [41] over the m/z range 50-2000. For a detailed overview of the MS conditions see Table 5.

Table 5: MS condition for HDX. ESI positive, Resolution mode, Lock Mass: 100pg/ μ L Leucine Enkephalin at 10L/min

Source	Parameters
Capillary (KV)	2.5
Sample Cone (V)	30
Source Offset (V)	40
Temperature	
Source ($^{\circ}$ C)	80
Desolvation ($^{\circ}$ C)	250
Gas Flows	
Cone Gas (L/h)	100
Desolvation Gas (L/h)	800
Nebuliser (Bar)	6.5

2.1.3. Data Analysis

Peptides were identified by using the software Protein Lynx Global Server (PLGS). Identified peptides were plotted onto the phosphorylase b sequence. The deuterium incorporation levels for each peptide were automatically calculated by using the software DynamX. The data are expressed as relative deuterium uptake (Da), which was calculated by dividing the deuterium level (Da) by the total number of backbone amide hydrogens that could have become deuterated (equal to the number of amino acids, minus proline residues minus 1 for the N-terminal amide) (see Chapter 1).

3. Results and Discussion

Three individual aspects of an HDX MS experiment were tested. This included pepsin digestion efficiency, back exchange and reproducibility, to find conditions which ensure an HDX MS workflow obtains high-quality and information-rich data. Phosphorylase b (PhosB) from rabbit muscle was used for testing the HDX MS system.

3.1. Sequence Coverage and Back Exchange

Experiments were performed at 10, 20 and 37°C in the digestion step, to investigate how individual temperature conditions can influence the back exchange and the digestion efficiency of pepsin. The expected sequence coverage for standard digestion conditions (20°C) for phosphorylase b should be around 75%, with circa 170 peptides known from published data [38]. After separate digestions with pepsin, 135 (10°C), 189 (20°C) and 220 (37°C) peptides were identified. Sequence coverages were 73% (10°C), 81% (20°C) and 84% (37°C) for phosphorylase b (Table 6).

Table 6: Results from the pepsin digestion 10, 20 and 37°C. Indicates total peptides, sequence coverage and redundancy.

Protein	Temp. (°C)	Peptides Total	Coverage (%)	Redundancy
PhosB	10	135	72.8	2.20
PhosB	20	189	80.6	2.95
PhosB	37	220	83.8	3.10

*Coverage data were taken from DynamX and were the average of four chromatographic replicates. Redundancy of covered amino acids, which is the average number of peptides each residue is found in.

It is clearly visible that sequence coverage increases with digestion temperature. Between 10°C and 37°C the coverage difference was 11% which means that digesting at 10°C will have a negative impact on the amount of structural information that can be obtained from the protein. Between 20°C and 37°C there is only a minimal difference which indicates that pepsin activity does not significantly increase above 20°C.

Figure 11 shows a selected sequence area of phosphorylase b. The peptide pattern at digestion temperatures of 20°C and 37°C are nearly identical but there is a significant decrease in digestion efficiency at 10°C. For example, at 10°C the region between amino acids 530 and 580 is only covered by two peptides and there are two major gaps in this area. At 20°C and 37°C the whole region is covered by peptides. In general, at 37°C the number of overlapping peptides increases and therefore the structural information but it has also consequences for the back exchange level which means the loss of deuterium (see below, Table 7). Therefore, it is important to have a good balance between sequence coverage and back exchange to generate high-quality and information-rich data.

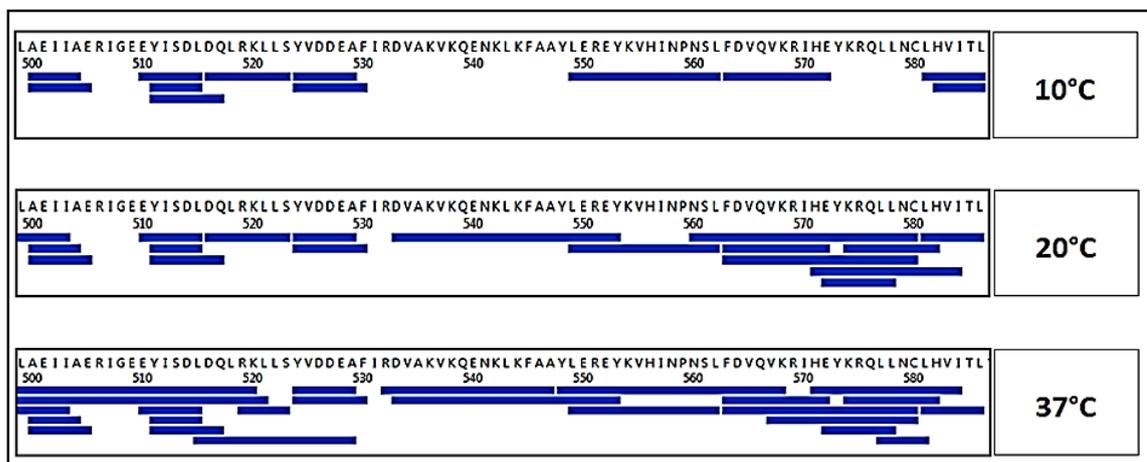


Figure 11: Comparison of the coverage map of peptic peptides from digestion of phosphorylase b under different temperature conditions. Selected phosphorylase b sequence area from 499-586.

In table 7 the deuterium uptake of phosphorylase b peptides at the different digestion temperatures (10, 20 and 37°C) is shown. Deuteration exposure times were 2 and 120 minutes. The data are represented as deuterium uptake in Dalton (Da), which was calculated by dividing the deuterium level in Dalton, by the total number of backbone amide hydrogens that could have become deuterated (equal to the number of amino acids, minus proline residues minus 1 for the N-terminal amide). Eight peptides were selected from phosphorylase b to investigate how temperature can influence the back exchange over a specific time range. It is clearly visible that the back exchange significantly increases at higher temperature (Table 7, Figure 12).

After 120 minutes in seven of the eight selected peptides a significant deuterium loss due to the back exchange between 10°C and 37°C was observed. The experiment demonstrates also that different peptides exhibit a different range of back exchange. This can be expected since the amide HX rate in a peptide varies with amino acid type and nearest neighbor [23]. The peptide from 338 to 349 shows no significant deuterium uptake at all temperature conditions. This could be explained by the fact that this peptide is part of a region of phosphorylase b that is highly structurally ordered and therefore it will incorporate deuterium only slowly. In general, these results show that a digestion temperature around 20°C is optimal to get good sequence coverage, redundancy and an acceptable amount of back exchange for phosphorylase b.

Table 7: Deuterium uptake levels (in Dalton, which indicates the mass differences between undeuterated and deuterated sample, see Chapter 1) plotted versus the deuterium incorporation time (in minutes). The experimental uncertainty of measuring deuterium uptake was $\pm 0.5\text{Da}$ as described in the next paragraph. Digestion temperatures are 10, 20 and 37°C. RT: retention time in minutes (min).

RT [min]	Peptide	Sequence	Digestion Temp. [°C]	Uptake 2min [Da]	Uptake 120min [Da]
4.10	16-32	RGLAGVENVTELKKNFN	10.00	7.10	7.16
4.10	16-32	RGLAGVENVTELKKNFN	20.00	6.45	6.84
4.10	16-32	RGLAGVENVTELKKNFN	37.00	5.39	5.39
5.39	125-139	IEEDAGLGNGGLGRL	10.00	5.69	6.83
5.39	125-139	IEEDAGLGNGGLGRL	20.00	5.33	6.53
5.39	125-139	IEEDAGLGNGGLGRL	37.00	3.65	4.31
4.89	244-252	WSAKAPNDF	10.00	2.03	2.85
4.89	244-252	WSAKAPNDF	20.00	2.09	2.83
4.89	244-252	WSAKAPNDF	37.00	1.90	2.30
5.89	338-349	NDTHPSLAPEL	10.00	0.53	0.69
5.89	338-349	NDTHPSLAPEL	20.00	0.53	0.65
5.89	338-349	NDTHPSLAPEL	37.00	0.38	0.43
5.78	510-515	EYISDL	10.00	1.10	1.95
5.78	510-515	EYISDL	20.00	1.10	1.88
5.78	510-515	EYISDL	37.00	1.04	1.42
3.52	605-618	IGGKAAPGYHMAKM	10.00	4.27	5.51
3.52	605-618	IGGKAAPGYHMAKM	20.00	4.16	5.51
3.52	605-618	IGGKAAPGYHMAKM	37.00	3.47	4.59
5.01	750-758	FSPKQPDLF	10.00	1.78	2.33
5.01	750-758	FSPKQPDLF	20.00	1.68	2.37
5.01	750-758	FSPKQPDLF	37.00	1.42	1.84
5.22	831-839	RQRLPAPDE	10.00	3.19	3.64
5.22	831-839	RQRLPAPDE	20.00	3.22	3.90
5.22	831-839	RQRLPAPDE	37.00	2.88	3.12

* Deuterium uptake data were taken from DynamX and are the average of four chromatographic replicates and all charge states.

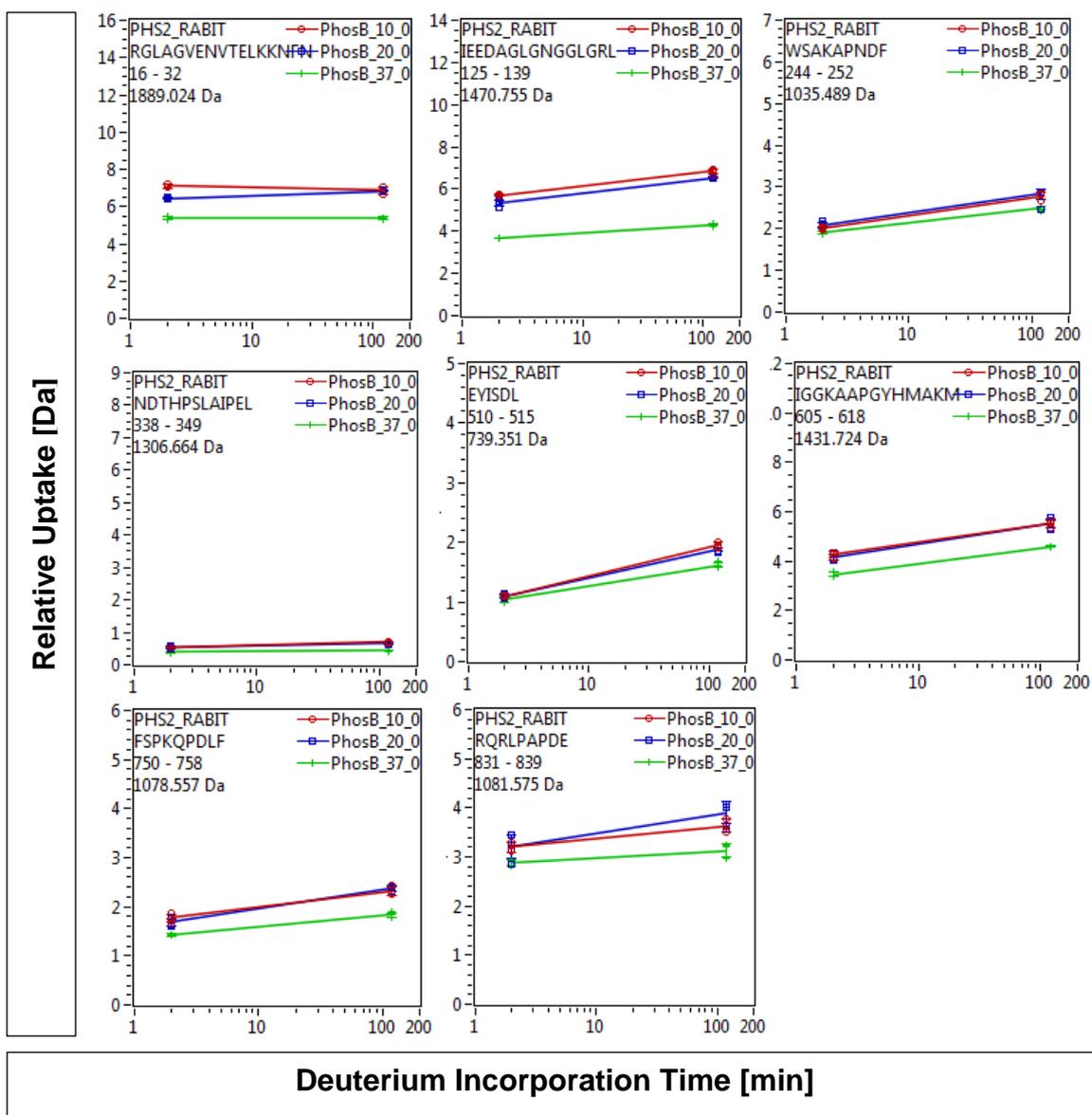


Figure 12: Relative Deuterium Uptake from eight selected phosphorylase b peptides. Deuterium uptake in Dalton (Da) on the y-axis. Deuterium incorporation time of 2 and 120min on the x-axis. RED: 10°C digestion temp., BLUE: 20°C digestion temp., and GREEN: 37°C digestion temp.. Deuterium uptake data were taken from DynamX and are the average of four chromatographic replicates and all charge states.

3.2. Reproducibility

Reproducibility describes the level of variation among replicates. Nearly every change of solution environment like pH and temperature can affect the exchange kinetics in HDX. Therefore, the evaluation of reproducibility should include a collection of measurements. This study was based on replicate measurements of phosphorylase b, compared from two days. At each day five individual replicates with a deuteration exposure time of 10min were prepared. The protein was analyzed like in a normal HDX MS experiment, to investigate the reproducibility of the measurements. Figure 13 shows the deuterium uptake and standard deviation (SD) of fifteen selected peptides.

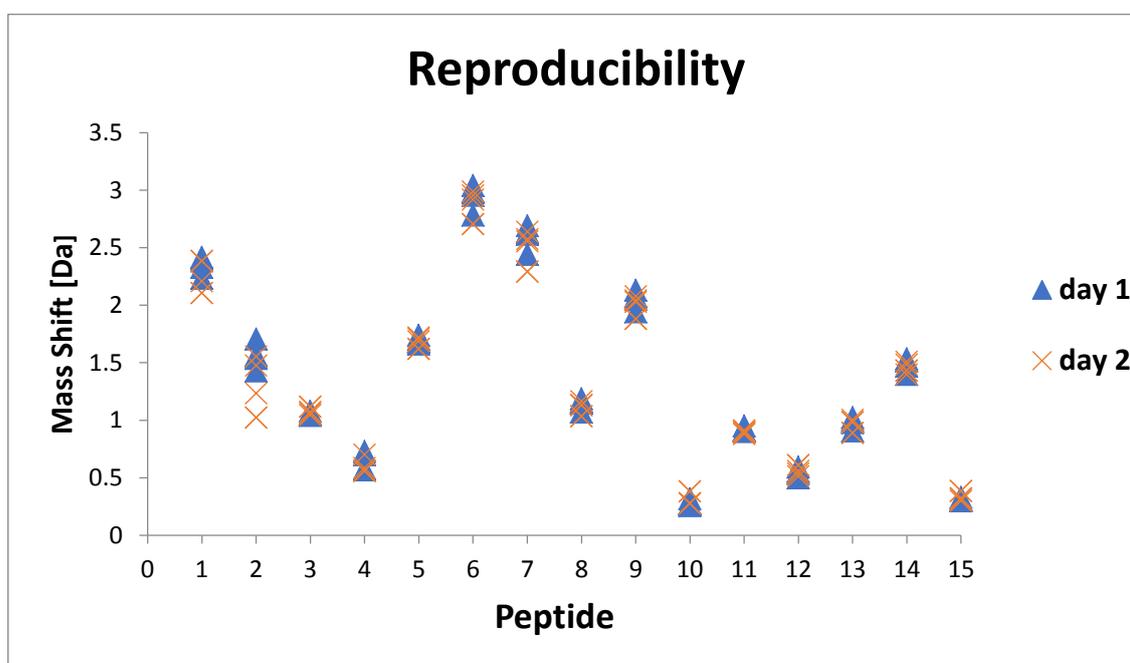


Figure 13: Reproducibility of ten individual replicates from two days. Each symbol presents one measurement. Selected peptides of PhosB: Pep.1: 16-32, Pep.2: 40-52, Pep.3: 53-67, Pep.4: 62-67, Pep.5: 90-99, Pep.6: 125-139, Pep.7: 128-139, Pep.8: 244-252, Pep.9: 275-285, Pep.10: 392-405, Pep.11:411-425, Pep.12: 472-491, Pep.13: 625-640, Pep.14: 713-721, Pep.15: 820-825. Standard deviation for each peptide: Pep.1: 0.12Da, Pep.2: 0.24Da, Pep.3: 0.03Da, Pep.4: 0.07Da, Pep.5: 0.04Da, Pep.6:0.12Da, Pep.7: 0.15Da, Pep.8:0.06Da, Pep.9: 0.09Da, Pep.10: 0.05Da, Pep.11: 0.02Da, Pep.12: 0.04, Pep.13: 0.05Da, Pep.14: 0.05Da, Pep.15: 0.03Da.

The data are expressed as relative deuterium uptake in Dalton (for calculation see Chapter 1). Based on data obtained from the HDX MS experiment, the highest experimental uncertainty of measuring a deuterium uptake was found to be $\pm 0.24\text{Da}$ from peptide 2. This variability is well within the generally accepted range of $\pm 0.5\text{Da}$ for HDX difference measurements [22], which means that if the difference in HDX level for any peptide at any labeling time exceeds 0.5Da then it is a significant difference. For all further experiments (Chapter 3 and 4), the published difference value of $\pm 0.5\text{Da}$ instead of $\pm 0.24\text{Da}$ was therefore used.

In general, the obtained deuterium uptakes for all selected peptides show minimal variation. Only for peptide 2 there are two outliers. A reason could be that the variability is influenced by a contamination in the system. If a peptide elutes exactly at the same time as a specific contamination, it can influence the HD exchange or ion intensity for further evaluations [38]. Additionally, the lesser the standard deviation the lesser the uncertainty and thus the higher the confidence in the experiment. Therefore, it is important for obtaining the most reproducible results to understand each variation in the system and try to minimize it as much as possible.

4. Conclusion

For high-quality, information-rich data and the direct comparison of two or more data sets without variation, conditions like pH, temperature and runtime should be thoroughly optimized in an HDX MS workflow for a specific protein of interest. Each process step should be reviewed for potential problems that might contribute to the variability of HDX MS data. In this study it was shown that digestion temperature significantly influences the proteolytic efficiency of pepsin. On the one hand pepsin is a protease of choice because it is highly active at low pH but on the other hand the digestion efficiency decreases significantly at lower temperature. Therefore, a challenge associated with HDX MS is to find optimal conditions in the digestion step to obtain good sequence coverage and to keep the back exchange in the system at an acceptable level of <30% [23]. The results demonstrate that temperature conditions for digestion that are still acceptable are around 20°C. For phosphorylase b, the optimal digestion temperature for acceptable sequence coverage was at 20°C. In chapter 3 the chosen digestion temperature was at 15°C instead of 20°C. At this lower temperature the sequence coverage was still excellent for the protein *hGMPK* and the back exchange could thus be further minimized. Therefore, the optimal digestion temperature also depends on the protein. One aspect for improving the digestion step in an HDX MS workflow could even be to find a protease which tolerates acidic conditions and is highly active at temperatures lower than 10°C. Obtained data from the replicate measurements of phosphorylase b indicated that under optimal conditions for pH, temperature and runtime the deuterium uptake levels in all selected peptides show minimal variation.

CHAPTER 3

CHARACTERIZATION OF THE OPEN/CLOSED CONFORMATIONS OF GUANYLATE KINASE

1. Introduction

Guanylate kinase is a member of the nucleoside monophosphate (NMP) kinase family, which catalysis the reversible phosphoryl transfer from ATP to GMP to produce ADP and GDP (Figure 14). The kinase is necessary for providing the required adenine and guanidine nucleotide levels in the cell and it also activates numerous antiviral and anticancer prodrugs. Therefore, structural information on guanylate monophosphate kinases (GMPKs) is important for the design of antiviral and antineoplastic agents [42].



Figure 14: Reversible phosphoryl transfer.

GMPKs have been studied from several organisms, and many functional and structural details are known, like the X-ray structure of *E.coli* GMPK (eGMPK) (Figure 15) [43]. Analysis of this structure indicates that eGMPK contains three subdomains, the CORE, LID and NMP binding region. The CORE region contains the ATP binding P- loop, the NMP-binding region binds the (d)NMP substrate and the LID region contributes general catalytical parts of residues for the reaction. In the absence of any substrate, the structure of the protein has a completely open conformation. If a substrate binds to the protein it undergoes large conformational changes and switches from an open to a closed conformation. In the presence of GMP the NMP-binding domain is partially closed. The ATP-binding site is partially blocked by inter-subunit interactions in all published crystal structures. This indicates that local conformational changes are required to open the ATP binding pocket [43, 44]. Arginine residues play an important role for transition state stabilization: R138 interacts with the α and β phosphates of ADP and bridges the ADP leaving group and the transferred phosphates at the transition state (TS), whereas R45 binds the α phosphate of GMP to bridge the GMP to the transferred phosphoryl group in the TS [45].

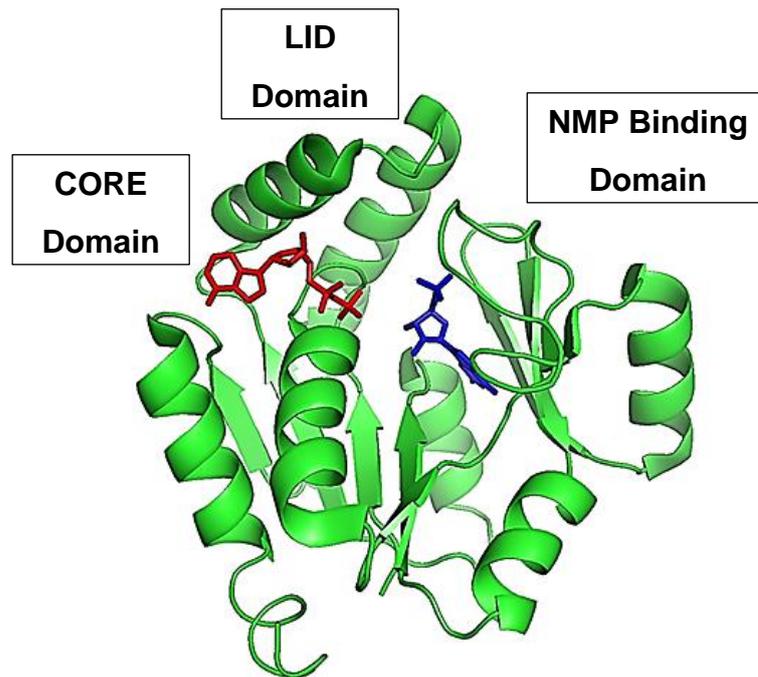


Figure 15: Guanylate Kinase from *E.coli*: with its three subdomains CORE (ATP-binding P-loop), LID and NMP binding region. The structural figure was generated with PyMOL by using the Protein Data Bank (PDB) accession code 1lvj [43].

NMPKs can also be inhibited by bi-substrate analogs. AP5G has been described as an inhibitor for eGMPK, which occupies the binding positions of both substrates (GMP and ATP), with a linker of five phosphates in between (Figure 16) [44].

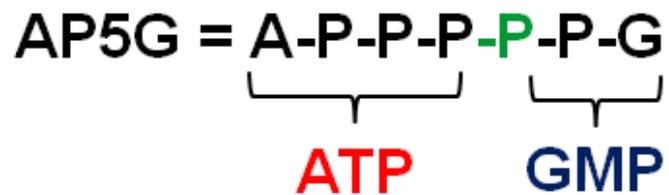


Figure 16: Bi-substrate inhibitor AP5G with a linker of five phosphates in between.

HDX MS data can provide insights into the conformational dynamics of a protein. In this chapter HDX MS was used to study the effects of ligand binding by the *human* guanylate kinase (*hGMPK*). The rate of deuterium incorporation for *hGMPK* in the absence of any ligand was measured and compared to the results observed for protein-ligand complexes. The binding of three different ligands ADP, GMP and AP5G to *hGMPK* was studied. At first the dissociation constants (K_D) for the three ligands were measured by MicroScale Thermophoresis (MST). This technique is based on thermophoresis and detects the directed movement of molecules in a temperature gradient, which strongly depends on a variety of molecular properties [46]. With this information the protein-ligand ratios for HDX MS experiments were calculated. If the K_D for the protein-ligand interaction is known, the percentage of the complex from the solution can be calculated with the following equation:

P_F = Free Protein
L_F = Free Ligand
P_T = Total Protein
L_T = Total Ligand

$$P_F + L_F \frac{K_A}{K_D} PL \quad \text{Eqn. 14}$$

$$PL = \frac{-[K_D + L_T + P_T] - \sqrt{[K_D + L_T + P_T]^2 - 4[(P_T)(L_T)]}}{2}$$

$$\%P_B = \frac{PL}{P_T} \quad \%L_B = \frac{PL}{L_T}$$

HDX MS was used to investigate how ligand binding can stabilize or destabilize certain areas of the protein which results in a decrease or increase of HDX rates. Another aspect was to compare the acquired HDX MS data with the published X-ray structure. Up to now, there is no structural data available for *hGMPK* therefore the X-ray structure of *eGMPK* was used.

2. Materials and Methods

General chemicals, materials and devices which were used for the master thesis project are listed in Table 1, 2 and 3. Chemicals that are only used for this chapter are listed in Table 8.

Table 8: Chemicals which were used for the whole *hGMPK* experiment. For the sequences of *hGMPK* see Appendix.

Chemicals	Company	Comments
<i>hGMPK</i> tagged	Boehringer Ingelheim	Uncleaved: MW of 24500 kD
<i>hGMPK</i> untagged	Boehringer Ingelheim	Cleaved: MW of 22307 kD
Adenosine 5' -diphosphate sodium salt	Sigma-Aldrich (Merck) A2754_SIGMA	
Guanosine 5'-monophosphate disodium salt hydrate	Sigma-Aldrich (Merck) G8377_SIGMA	
P1-(5'-Adenosyl) P5-(5'-guanosyl) pentaphosphate, Triethylammonium salt	Jena Bioscience, GmbH	
Natriumchloride (NaCl)	Sigma-Aldrich (Merck)	
Kaliumchloride (KCl)	Sigma-Aldrich (Merck)	
Disodium phosphate (Na ₂ HPO ₄)	Sigma-Aldrich (Merck)	
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Sigma-Aldrich (Merck)	
TRIS Buffer	Lonza	
Sodium Chloride	Lonza	
Tween 20	Sigma-Aldrich (Merck)	
PBS P+ Buffer	GE Healthcare	
Monolith His-Tag Labeling Kit	NanoTemper	MO-L008
RED-tris-NTA	Technologies GmbH	

2.1. Microscale thermophoresis

Materials and devices which were used for the MST experiment are listed in Table 9.

Table 9: Materials and Devices for MST.

Materials	Company	Comments
Monolith NT.Automated Capillary Chips	NanoTemper Technologies GmbH	MO-AK002
Assay Plate	Corning	384 Well
Devices	Company	Comments
Monolith NT.Automated	NanoTemper Technologies GmbH	
Centrifuge 5430	Eppendorf	

2.1.1. Experimental Part

The *hGMPK* with a hexahistidin-Tag (HIS-Tag) was overexpressed by Boehringer Ingelheim internally (Sequence, Appendix).

For the microscale thermophoresis (MST) experiment an assay buffer [20mM TRIS, 100mM NaCl and 0.05% Tween 20, pH 7.4] was prepared. In the first step the protein was centrifuged for 10min at 15,000g, 4°C to remove any aggregated protein. After centrifugation the protein was diluted with the assay buffer to a concentration of 0.2µM. For labeling a dye (RED-tris-NTA) was diluted with PBS+T buffer [10mM PBS, 0.05% Tween 20, pH 7.4] to a concentration of 5µM and subsequently diluted with the assay buffer to a concentration of 0.1µM. It was then mixed in a 1:1 ratio with the protein and incubated for 30min at room temperature (21°C). After incubation the labeled protein was centrifuged for 10min at 15,000 g, 4°C. As the first step a serial 1:1 dilution of the ligands (ADP [start concentration: 2mM], GMP [start concentration: 2mM] and AP5G [start concentration: 50µM]) was made in the assay buffer. In the next step, labeled protein was added in a 1:1 ratio to all the serially diluted compounds (ADP, GMP and AP5G) and incubated for 30min at room temperature (21°C).

Microscale thermophoresis was measured using a Monolith NT.Automated instrument (NanoTemper). Capillaries (MO-AK002) were filled manually through capillary action. For the MST measurements a red laser setting was chosen. Fluorescence was recorded for 5 seconds before activating the IR laser, then for 15 seconds and after turning off the IR laser for 1 more second. Each sample of the serial dilution was measured from the lowest to the highest concentration of ligand and each dilution series was measured two times. For data analysis the software NT Analysis 1.5.41 (NanoTemper) was used.

2.2. Intact Mass and HDX MS

Chemicals, materials and devices which were used for the HDX MS experiment are listed in Table 1, 2, 3 and 7.

2.2.1. Experimental Part

Untagged *hGMPK* was overexpressed by Boehringer Ingelheim internally and its mass confirmed by ESI-MS (Sequence, Appendix). The protein had a concentration of 18.88mg/mL and was stored in phosphate buffered saline (PBS) [137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄ pH 7.4] with 2mM TCEP at -80°C.

2.2.1.1. Intact Mass

The protein was diluted to a concentration of 20nM with H₂O and 0.1% formic acid. The intact mass was measured to check the quality and mass of the untagged *hGMPK*

2.2.1.2. Hydrogen Deuterium Exchange

For the experiment in absence of ligands, *h*GMPK was diluted to a concentration of 12 μ M in PBS [137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, pH 7.4] with 2mM TECEP. For the experiments with ligands, the stock solution (*h*GMPK) was also diluted to a concentration of 12 μ M in PBS with 2mM TECEP, pH 7.40 and incubated for 20 minutes with ADP [final concentration of 3mM], GMP [final concentration of 3mM] or AP5G [final concentration of 0.3mM].

The HDX MS experiment was performed in individual parts: GMPK unbound, *h*GMPK bound with the ADP, GMP and AP5G. *h*GMPK in absence of any ligand and *h*GMPK complexes were diluted 20-fold with PBS, 2mM TECEP in H₂O, pH 7.4 for reference experiments (no deuterium), or PBS, 2mM TECEP in D₂O, pD 7.0 for exchange experiments. In the next step, the samples were incubated at room temperature (21°C) for different time periods. In the undeuterated experiment for 0min and in the deuterated experiment for 10sec, 1, 10, 60 and 240min. At the end of the incorporation time the samples were quenched 1:1 with ice-cold potassium buffer [100mM KH₂PO₄, 100mM K₂HPO₄ in H₂O, pH 2.4]. Wash solution with 1.5M guanidine HCl, 4% acetonitrile and 0.8% formic acid, pH 2.5 was used to clean the system between each sample injection to avoid peptide carryover from the previous run. All HDX measurements were conducted in duplicates.

2.2.2. UPLC Separation & Mass spectrometry

Eluents and washes which were used for the LC system and the Leap Robot during this experiment are listed in Table 3.

2.2.2.1. Intact Mass

Guanylate kinase was directly injected into the LC system at a concentration of 20nM. The temperature of the UPLC system was 25°C. BEH C4 pre-column and a BEH C4 column were used for desalting and separation. The protein was eluted by using an acetonitrile gradient (Table 10) containing 0.2% formic acid with a flow rate of 20µL/min.

Table 10: Gradient for intact Masses. A (%) H₂O with 0.2% FA, B (%) ACN with 0.2% FA.

Time (min)	Flow rate(µL/min)	% A	% B
Initial	20	95.0	5.00
1.00	20	95.0	5.00
3.00	20	5.00	95.0
5.00	20	5.00	95.0
5.10	20	95.0	5.00
6.00	20	95.0	5.00

From the LC system the eluent was directed into a Synapt G2-Si mass spectrometer. The ionization was performed with electrospray ionization (ESI). Source and desolvation temperatures of the instrument were 130°C and 400 °C, cone voltage 30V, and the ESI voltage 3kV. MS conditions were ESI positive, sensitivity mode and the mass spectra were acquired in MS mode over the m/z range 400-2000 (Table 11).

Table 11: MS condition for HDX. ESI positive, sensitive mode and lock Mass: 100pg/ μ L Leucine Enkephalin at 10L/min

Source	Parameters
Capillary (KV)	3
Sample Cone (V)	30
Source Offset (V)	40
Temperature	
Source ($^{\circ}$ C)	130
Desolvation ($^{\circ}$ C)	400
Gas Flows	
Cone Gas (L/h)	100
Desolvation Gas (L/h)	800
Nebuliser (Bar)	6.5

2.2.2.2. Hydrogen Deuterium Exchange

The *h*GMPK was passed through the BEH pepsin column at 80 μ L/min for 3 min in H₂O with 0.2% formic acid, pH 2.4 with a concentration of 15pmol/50 μ L (loop volume). The digestion takes place at 15 $^{\circ}$ C in a separate chamber of the HDX Manager. The generated peptides were desalted by using a BEH C18 pre-column at 0 $^{\circ}$ C. Subsequently the peptides were eluted and separated on a BEH C18 column at 0 $^{\circ}$ C. An acetonitrile gradient (Table 12) with 0.2% formic acid at a flow rate of 40 μ L/min was used for the peptide separation.

Table 12: Gradient for HDX. A (%) H₂O with 0.2% FA, B (%) ACN with 0.2% FA.

Time (min)	Flow rate (μ L/min)	% A	% B
Initial	40.0	95.0	5.0
6.0	40.0	65.0	35.0
7.0	40.0	60.0	40.0
8.0	40.0	5.0	95.0
10.0	40.0	5.0	95.0
11.0	40.0	95.0	5.0

From the LC system the eluent was introduced into the Mass spectrometer. For detailed parameters for the Synapt G2-Si Mass spectrometer see Chapter 2.

2.2.3. Data Analysis

The mass of the intact protein was determined by using the tool MaxEnt1 (Waters) [47], which automatically finds the molecular masses from a set of multiply charged species. Peptides were identified by using the software Protein Lynx Global Server (PLGS). Identified peptides were then plotted onto the *hGMPK* sequence. The deuterium incorporation levels for each peptide were automatically calculated by using the software DynamX. The data are expressed as relative deuterium uptake (Da), as described in Chapter 1.

3. Results and Discussion

3.1. MST

The binding affinities of the nucleotides ADP, GMP as well as the inhibitor AP5G for *hGMPK* were measured. The calculated K_D values from four replicates for each ligand are summarized in Table 13. For all ligands K_D values could be determined.

Table 13: Results of the MST measurements. The K_D values and signal amplitudes for AP5G, ADP and GMP binding to *hGMPK*.

Target	Ligand	K_D [μM]	Amplitude [ΔF_{norm}]
<i>hGMPK</i>	AP5G	0.4 ± 0.03	22 ± 0.5
<i>hGMPK</i>	ADP	10 ± 2.0	13 ± 0.5
<i>hGMPK</i>	GMP	110 ± 12	22 ± 0.5

* K_D values are the average of four measurements.

The binding affinity for the inhibitor AP5G was determined to be $0.4\mu\text{M}$. This value matches the published data of $0.5\mu\text{M}$ [48]. For the two nucleotides the binding affinities were determined to be $10\mu\text{M}$ for ADP and $110\mu\text{M}$ for GMP.

Figure 17 shows the binding curves of four measurements for each protein-ligand interaction. The obtained values confirmed good data reproducibility. With the known K_D values for the protein-ligand interaction, the percentage of the complex in solution was calculated. For all complexes a ligand concentration was used to obtain a binding ratio of approximately 99%.

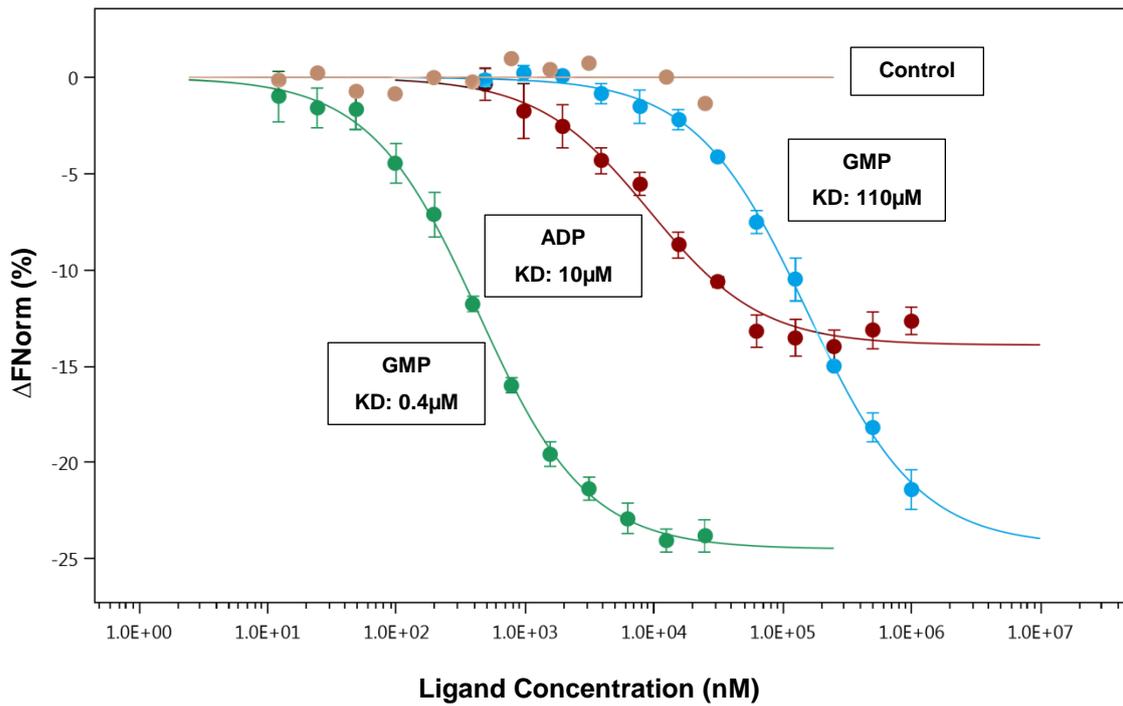


Figure 17: Plotted binding curves for *hGMPK*. Green: AP5G KD of 0.4 μM , Red: ADP KD of: 10 μM , Blue: GMP KD of 110 μM and Brown: Control without a Ligand. Each protein-ligand interaction (AP5G, ADP and GMP) was measured four times.

3.2. Intact Mass

The measured molecular weight of 22,307 Da (Figure 18) matches exactly the expected molecular weight and no other protein species were detected in the spectrum. It was important that the measured mass was identical to the specific sequence of the protein for further HDX MS experiments.

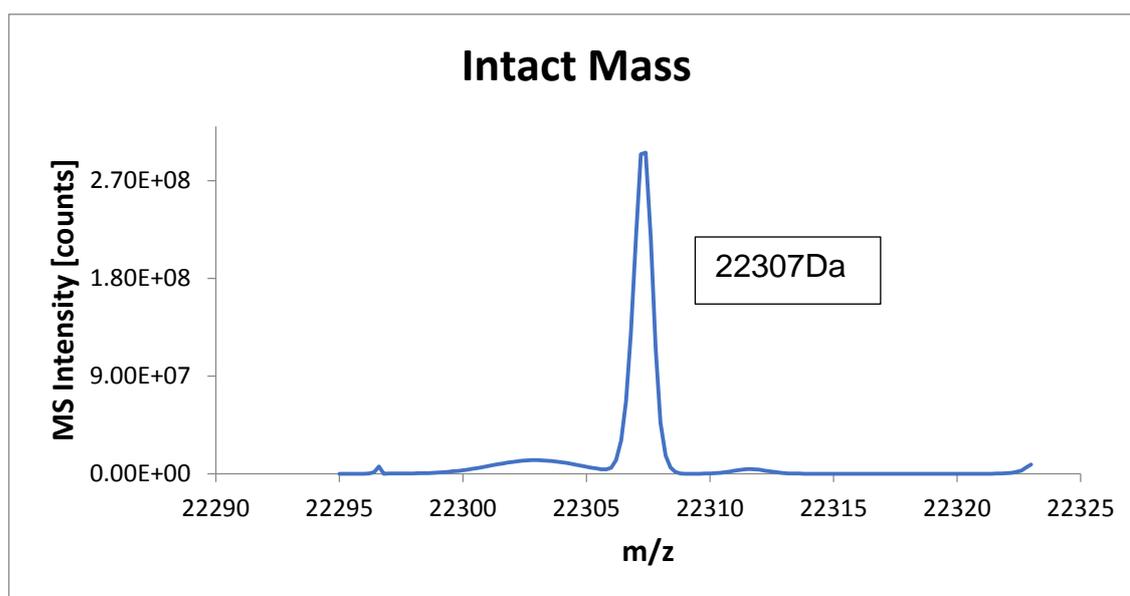


Figure 18: Intact Mass of *hGMPK*. ESI positive, sensitivity mode and the mass spectra were acquired in MS mode.

3.3. Sequence Coverage

A peptide map for *hGMPK* was created by using the program PLGS and DynamX (Waters). The peptide map of *hGMPK* shows very good sequence coverage of 98.5% with 74 peptides and a redundancy of 4.51 (Figure 19). For *hGMPK* the chosen temperature for pepsin digestion was at 15°C instead of 20°C which was chosen for phosphorylase b (see Chapter 2). The results show that 15°C is high enough for *hGMPK* to obtain high sequence coverage (98.5%) and to further minimize the back exchange. Only three amino acids were not covered in the sequence. It is evident from the data that the optimal digestion temperature also depends on the protein.

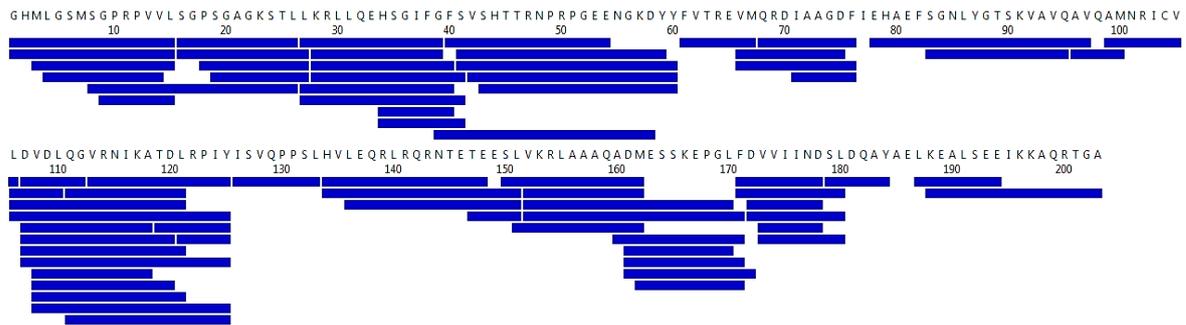


Figure 19: Coverage map of pepsine digestion. Online pepsin digestion of *hGMPK* resulted in of 98.5% sequence coverage with 74 peptides and 4.51 redundancy. Each blue bar represents an identified peptic peptide.

3.4. HDX MS

The interaction between protein and ligand leads to changes in the backbone dynamics of the protein. These conformational changes can be observed by HDX MS via changes in deuterium incorporation for each peptide in a protein. Uptake plots for each pepsin generated peptide were created. The deuterium uptake for each peptide was plotted as a function of time. The fits for *hGMPK* were compared to the corresponding fits for *hGMPK* in complex with ADP, GMP and AP5G. In general, the data for ADP, GMP and AP5G binding to *hGMPK* shows that ligand interaction leads to a complex pattern of stabilized and destabilized regions in the protein. Figure 20 shows some interesting peptides from *hGMPK*. All four measurements (protein and protein-ligand) were integrated into one uptake plot for each peptide. The first row (A) shows some peptides that are not affected by nucleotide binding at all. In the second row (B) it is clearly visible that for some peptides AP5G binding results in strong protection. The third row (C) shows an interesting trend of deuterium incorporation: these peptides are part of the ATP binding site; both AP5G and ADP show strong protection.

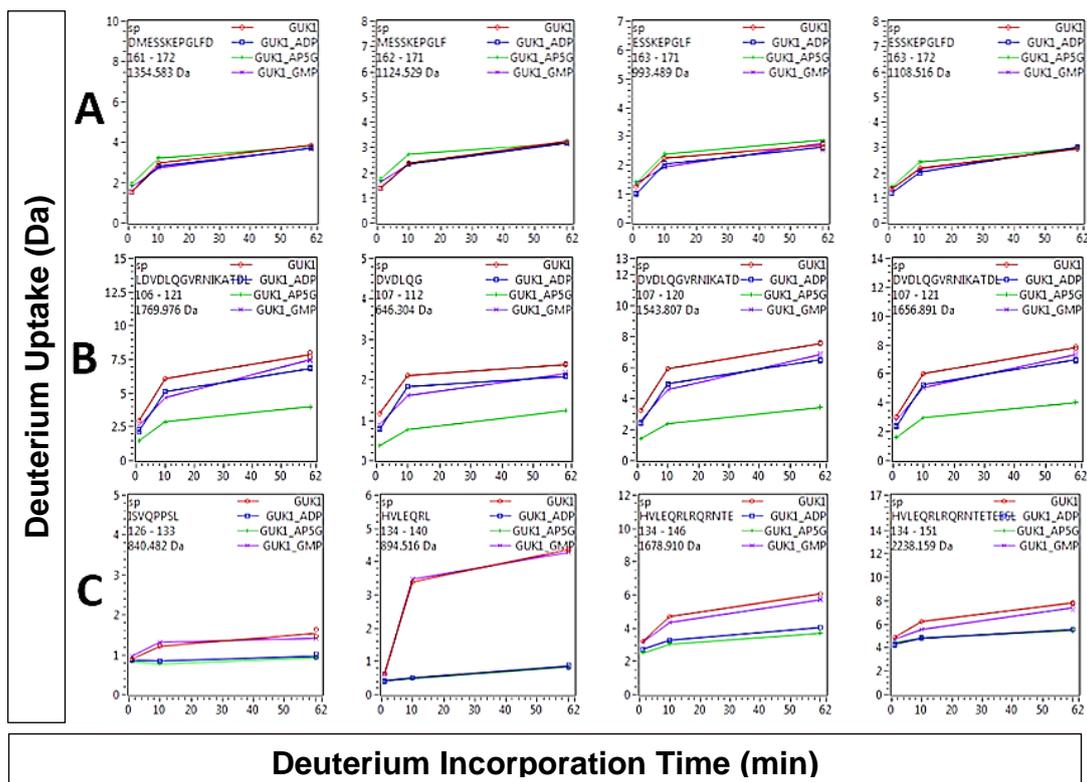


Figure 20: Uptake Plot with selected sequence areas of all four measurements. Deuterium uptake in Dalton (Da) on the y-axis and the deuterium incorporation time of 0, 17, 1, 10 and 60min on the x-axis. The red colour symbolises *h*GMPK on its own, blue colour *h*GMPK bound with ADP, violet *h*GMPK with GMP and green *h*GMPK with AP5G. A: sequence region from 161-172, B: sequence region from 106-121 and C: sequence region from 126-151.

John Engen's group, Boston, introduced the 'chicklet plot' for a simple visualization of HDX MS data Ref??. This plot uses color-coding to show the percentage of deuterium uptake for each generated peptide plotted as a function of time. It gives a global view of deuterium uptake for the whole protein. Figure 21 indicates the percentage of deuterium uptake for each generated peptide of *h*GMPK in absence of any ligand. The *h*GMPK reaches a deuteration level of almost 60% at the end of the experimental time window of 1 hour. This can be interpreted in a way that the protein is highly flexible in solution. Interestingly show the kinetics for the sequence area around amino acids 61-100 a particularly high rate of deuterium incorporation. This area is part of the GMP binding site and it is not surprising to see that this area is highly dynamic in the absence of the nucleotide GMP.

Peptide 100-106 is part of the hinge region of *hGMPK* and shows only a minimal amount of deuterium uptake over the whole time of incorporation. Apparently, this area is highly stable even in the absence of any ligand. The sequence area around amino acids 126-170 shows also a high rate of deuterium incorporation, which is known to be the interaction site for the

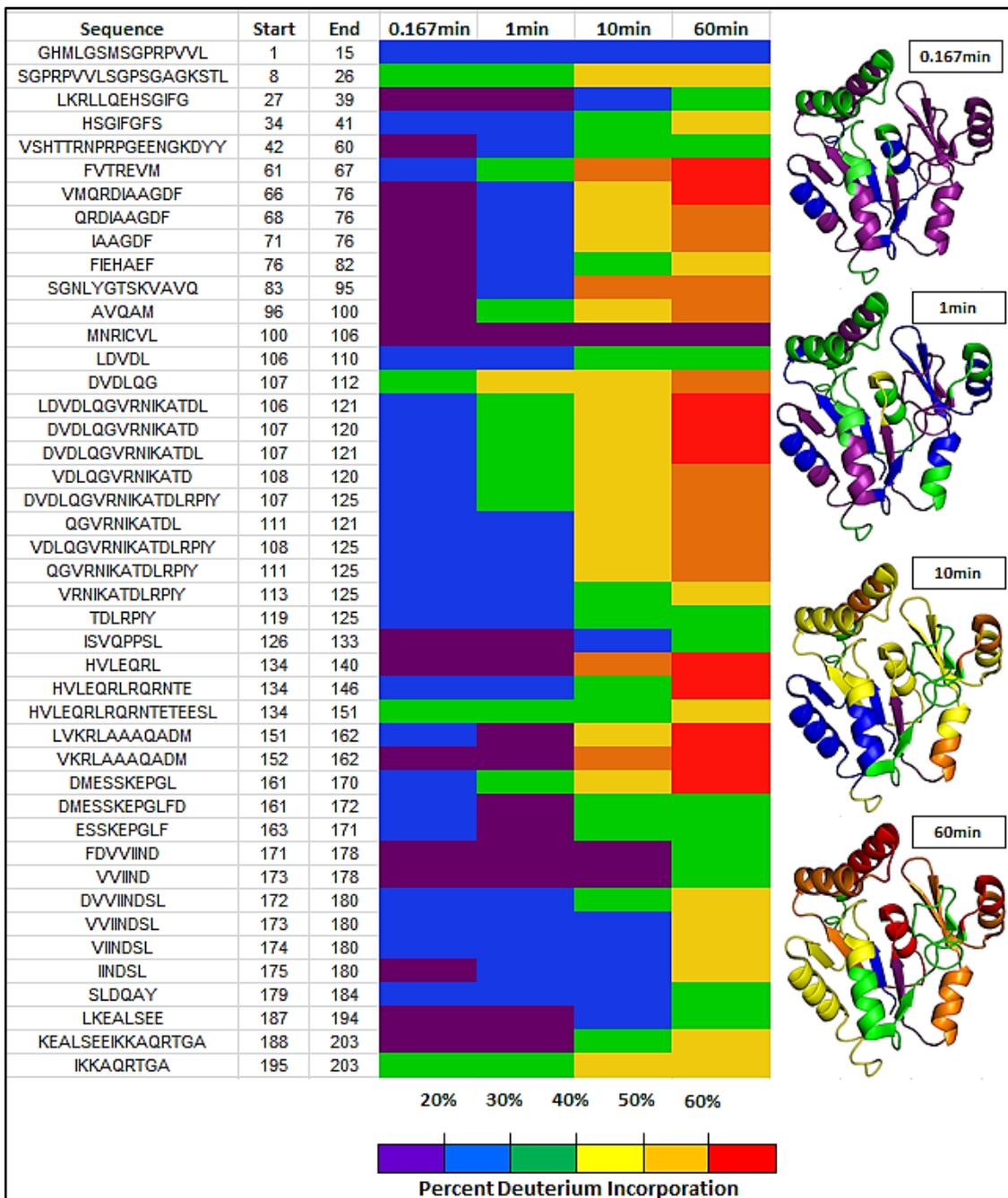


Figure 21: Chicklet Plot and cartoon structure representation of *hGMPK*. Indicates the Percentage of deuterium uptake for each generated peptide. Time range of deuterium incubation 0.167, 1, 10, 60 minutes. Violet: under 20%, Blue: 20-30%, Green: 30-40%, Yellow: 40-50%, Orange: 50-60% and Red: above 60%. The structural figures were generated with PyMOL by using the PDB accession code 1lvj [43].

Figure 22 depicts the uptake differences between *hGMPK* and the *hGMPK* complexes with ADP, GMP and AP5G. Time dependent data were acquired for the time range from 10 seconds to 1 hour. For evaluation, the generally accepted experimental uncertainty of ± 0.5 Da for HDX difference measurements was used. Therefore, deuterium uptake differences up to 0.5Da were not considered and colored in grey in the “chicklet plot”. The blocks which are marked in red represent areas close to the known ligand binding sites. In the presence of the inhibitor AP5G which occupies both binding sites, many regions of *hGMPK* experience strong protection and the whole protein appears to become rather rigid. In the case of ATP and GMP only areas close to one of the two binding sites show additional protection. AP5G has two additional phosphates compared to ADP and GMP together. These phosphates operate as “clamps” between the two binding sites and freezing the dynamics of the two subdomains completely which leads to a very low deuterium uptake level. Another interesting area in the protein is the C-terminus, which shows in the presence of GMP and AP5G a higher amount of deuterium incorporation than in absence of any ligand. Possibly the binding of GMP changes the dynamics of the C-terminus in a way that interaction with potential protein binding partners in the cell can be modified.

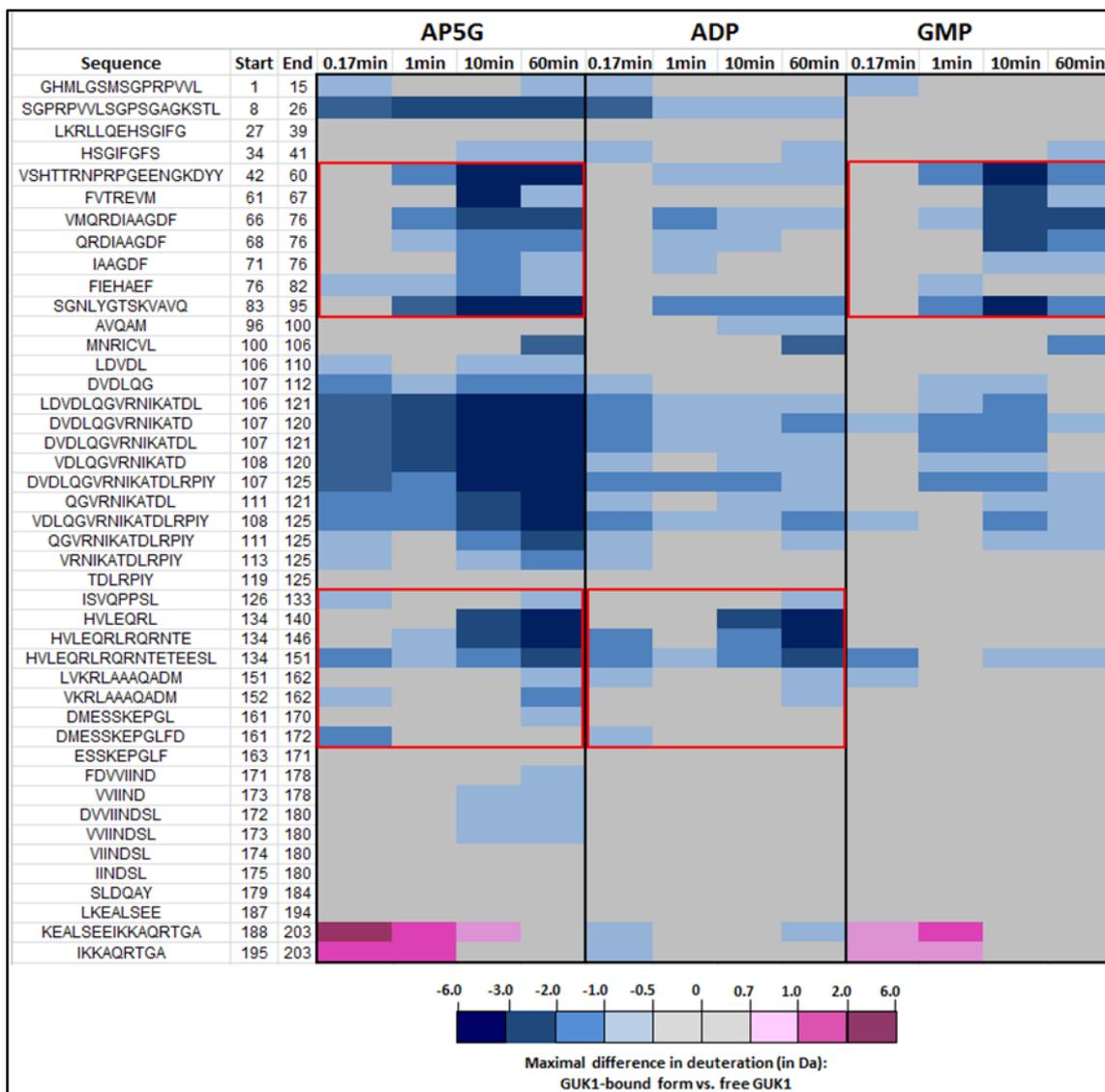


Figure 22: Deuterium uptake differences in Da between *h*GMPK to the corresponding *h*GMPK interaction with AP5G, ADP or GMP. Blue colour code: between -0.6 and -0.5Da, Grey: no significant differences, Pink colour code: between 0.7 and 6.0Da.

Figure 23 shows the observed major differences in deuterium uptake between *h*GMPK and its complexes with AP5G, ADP or GMP mapped onto the published 3D structure from *e*GMPK. Up to now, there is no structural data available for *h*GMPK and therefore the X-ray structure of *E. coli* GMPK (*ec*GMPK) was used. With the binding of AP5G the region in the core domain is protected more strongly compared to ADP and GMP together. It indicates that the phosphate linker of AP5G influences the core domain of the protein even more. The regions of protection observed in the two subdomains align well with the currently available crystal structure from *E. coli*.

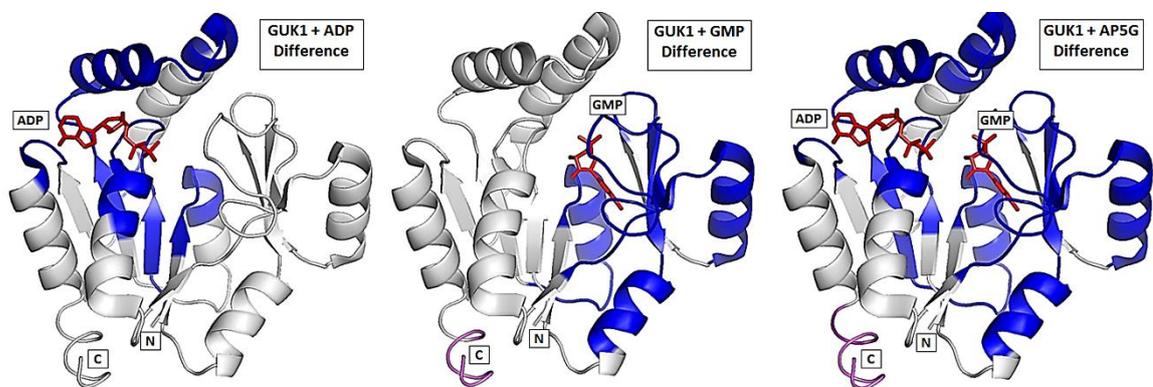


Figure 23: Differences of *h*GMPK with ADP, GMP and AP5G mapped onto cartoon structure (*E. coli*). Blue colour represents the major changes in the present of a substrate and Grey: no changes. The structural figures were generated with PyMOL by using the PDB accession code 1lvj [43].

4. Conclusion

This chapter demonstrates that HDX MS can be effectively used to study protein-ligand interactions. HDX MS can be applied to determine the regions of interaction as well as to gain information on protein dynamics in solutions which are not available through traditional methods such as X-ray crystallography. HDX MS data identified AP5G as a potent inhibitor and confirmed the published data [44]. The results demonstrate that *h*GMPK undergoes large conformational changes. In the complex with AP5G the protein switches from a completely open to a closed conformation. Overall, the obtained HDX MS data match very well with the published 3D structure of *E. coli* GMPK. The areas of known interaction sites overlap nearly exactly with the major changes of the deuterium uptake levels. With this work the system was established and could now be used to characterize changes in protein dynamics resulting from potential drug candidates with *h*GMPK.

CHAPTER 4

HDX MS ANALYSIS OF THE N-TERMINAL DOMAIN OF THE ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPC1

1. Introduction

Eukaryotes as well as prokaryotes have degradation mechanisms for non-functional, misfolded proteins. The ubiquitin-proteasome system present in eukaryotes is already well understood. In this system proteins destined for degradation by the 26S proteasome are marked with a polyubiquitin chain [49]. An analogous system to the eukaryotic 26S proteasome has been described for gram positive bacteria where a prokaryotic ubiquitin-like protein-tag (pup tag) is added to the protein to mark it for degradation [50]. Very recently, a second degradation mechanism for gram positive bacteria has been described. In this system, phosphorylated arginine residues serve as a tag for degradation of proteins by the ATP-dependent caseinolytic protease complex ClpC-ClpP. However, until now little is known about the detailed mechanism of protein degradation by ClpC-ClpP [51]. In general, the regulated protein “shredders” in gram-positive bacteria contain proteolytic sites within a chamber which are only accessible through entrance gates. These hexameric gates like ClpC are regulated by AAA+ ATPases which use the energy of ATP hydrolysis for unfolding and translocation of protein substrates into the associated protease domain ClpP for degradation (Figure 24). The specificity of AAA+ for protein substrates is often regulated by associated adaptor proteins [52].

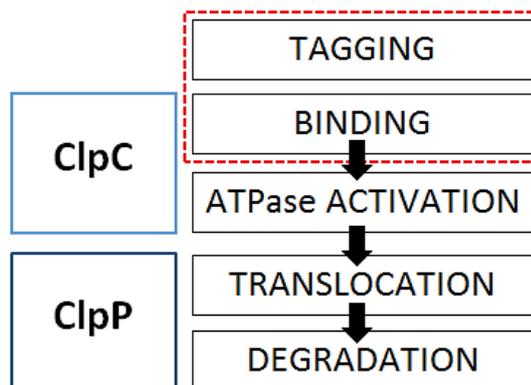


Figure 24: Bacterial proteolytic complex ClpC-ClpP. Binding of substrates to the NTD stimulates the ATPase activity of ClpC, leading to the translocation of the substrate into the ClpP protease and to subsequent protein degradation.

The group of Tim Clausen [53] discovered the tagging and binding system of ClpC and demonstrated that *Bacillus subtilis* ClpC specifically recognizes proteins with arginine residues that are phosphorylated by the arginine kinase McsB (Figure 25). With this knowledge a new way for selective degradation of misfolded proteins in bacteria was discovered. It is known that arginine residues become accessible for phosphorylation if they are located in unfolded or misfolded areas of proteins. Unfortunately, no ClpC structure could be obtained with phosphorylated substrate bound to ClpC1 but a structure could be solved with phosphoarginine acting as surrogate for a substrate protein [52]. However, until now it is unclear how substrate binding triggers the process of protein degradation in the ClpC-ClpP complex. Recently, NMR data were published which demonstrated that ligand binding to the N-terminal domain of ClpC is linked to changes in the dynamics of the protein in the millisecond range. The data indicated that the binding does not result in unfolding or oligomerization. It was shown that phosphoarginine binding leads to increased dynamics in the NTD of ClpC which could facilitate the next steps of unfolding and translocation to the protease ClpP. Another ligand that was investigated was cyclomarin A which is a natural antibiotic and also targets the NTD of ClpC. From published NMR data is known that cyclomarin A binding completely abolishes the millisecond-dynamic movements of the N-terminal domain that were observed in the presence of phosphoarginine [54]. However, it is still unclear how the changed dynamics of the NTD influence the next steps that lead to protein degradation.

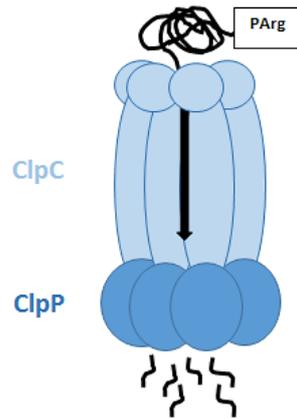


Figure 25: Bacterial proteolytic complex ClpC-ClpP. Tetradecameric protease domain ClpP and the hexameric ATPase domain ClpC.

In this chapter HDX MS was used to investigate the effects of phosphoarginine and cyclomarin A binding on the dynamics of the N-terminal domain ClpC on the HDX timescale which ranges from seconds to hours. In NMR data were only obtained on a millisecond time range [54]. This study used the N-terminal domain of ClpC1 from *M. tuberculosis*, because over the last few years the proteolytic complex in *M. tuberculosis* which consists of the proteins ClpP1 and ClpP2 and their hexameric regulatory ATPase ClpC1 have emerged as potential antibacterial tuberculosis (TB) target for drug development. Tuberculosis (TB) is a serious infectious disease that mainly affects the lungs. Each year ten million people are infected and two million do not survive the infection. Therefore, it is of great interest to find a potential tuberculosis (TB) drug target [54].

2. Materials & Methods

General chemicals, materials and devices which were used for the master thesis project are listed in Table 1, 2 and 3. Chemicals that are only used for this Chapter are listed in Table 14.

Table 14: Chemicals which were used for the whole ClpC1 experiment. For the sequences of the NTD ClpC1 from *Mycobacterium tuberculosis* see Appendix.

Chemicals	Company	Comments
ClpC1 NTD <i>M.tuberculosis</i>	Boehringer Ingelheim, RCV GmbH & Co KG, Vienna	MW:17360 kD
Phosphoarginine (PArg)	Sigma-Aldrich (Merck) A2754_SIGMA	
Cyclomarin A	Cooperation Porject, IMP	
Natriumchloride (NaCl)	Sigma-Aldrich (Merck)	
2-Amino-2-(hydroxymethyl)propane-1,3-diol	Sigma-Aldrich (Merck)	

2.1. Experimental Part

ClpC1 NTD from *M. tuberculosis* with a HIS-TAG was overexpressed by Boehringer Ingelheim internally and with ESI-MS a molecular weight of 17,360 Da was determined (Sequence, Appendix). The protein had a concentration of 30mg/mL and was stored in 50mM TRIS and 150mM NaCl, pH 8.0 at -80°C. ClpC1 NTD was diluted to a concentration of 13µM in 50mM TRIS, 150mM NaCl, pH 8.0. For protein-ligand complexes, ClpC1 NTD with a concentration of 13µM was incubated for 20 minutes with phosphoarginine [concentration of 10mM] or cyclomarin A [concentration 20µM]. Published K_D values [55] with the method SPR (Surface plasmon resonance) of 4.6µM for phosphoarginine and 0.2nM for cyclomarine A were used. With the known K_D values for the protein-ligand interaction, the percentage of the complex in solution was calculated. For all complexes a ligand concentration was used to obtain a binding ratio of approximately 99%.

2.1.1. Intact Mass

The protein was diluted to a concentration of 20nM with H₂O and 0.1% formic acid. LC-MS was used to check the quality and mass of ClpC1 NTD (*M. tuberculosis*).

2.1.2. Hydrogen Deuterium Exchange

ClpC1 NTD (protein and protein-ligand) was diluted 16-fold with 50mM TRIS, 150mM NaCl in H₂O, pH 8.0 for reference experiments (no deuteration), or 50mM TRIS, 150mM NaCl in D₂O, pD 7.6 for deuterated experiments. For D₂O labeling the samples were incubated at room temperature (21°C) for different time periods. Time points were 10sec, 1, 10, 60 and 240 minutes. At the end of the incorporation time the samples were quenched 1:1 with ice-cold potassium buffer [100mM KH₂PO₄, 100mM K₂HPO₄ in H₂O, pH 2.4]. Wash solution with 1.5 M guanidine HCl, 4% acetonitrile and 0.8% formic acid, pH 2.5 were used to clean the system between each sample injection to avoid peptide carryover from the previous run. All HDX measurements were conducted in duplicates.

2.2. UPLC Separation and Mass spectrometry

Eluents and washes which were used for the LC system during these experiments are listed in Table 3.

2.2.1. Intact Mass

For the LC MS procedure of intact proteins see chapter 3.

2.2.2. Hydrogen Deuterium Exchange

ClpC1 NTD was passed through the BEH pepsin column at 80 µL/min for 3 min in H₂O with 0.2% formic acid, pH 2.5 with a concentration of 15pmol/50µL (loop volume). The digestion takes place at 20°C in a separate chamber of the HDX Manager. The generated peptides were desalted by using a BEH C18 pre-column at 0 °C.

Subsequently the peptides were eluted and separated at a BEH C18 column at 0 °C. An acetonitrile gradient (Table 15) with 0.2% formic acid at a flow rate of 40 μ L/min was used for the peptide separation.

Table 15: Gradient for HDX. A (%) H₂O with 0.2% FA, B (%) ACN with 0.2% FA.

Time (min)	Flow rate (μ L/min)	% A	% B
Initial	40.0	95.0	5.0
6.0	40.0	65.0	35.0
7.0	40.0	60.0	40.0
8.0	40.0	5.0	95.0
9.0	40.0	95.0	5.0

The source and desolvation temperature of the instrument was 80 and 120°C, cone voltage 30V, and the ESI voltage 3kV. MS conditions were ESI positive, resolution mode and the Mass spectra were acquired in MS_E mode over the m/z range 50-2000.

2.3. Data Analysis

The mass of the intact protein was defined by using the tool MaxEnt1 [47], which automatically finds the molecular masses from a set of multiply charged species. Peptides were identified by using the software Protein Lynx Global Server (PLGS). Identified peptides were plotted onto the ClpC1 NTD sequence. The deuterium incorporation levels for each peptide were automatically calculated by using the software DynamX. The data are expressed as relative deuterium uptake (Da), as described in Chapter 1.

3. Results and Discussion

Conformational changes and protein dynamics can be an important factor in ligand binding and in the drug mechanism of action. HDX MS was used to study the effects of the interaction of two ligands, phosphoarginine and cyclomarin A in the N-terminal domain ClpC1.

3.1. Intact Mass

The intact mass of the tagged NTD ClpC1 from *M. tuberculosis* was measured with ESI-MS and evaluated with the software MaxEnt1. The measured molecular weight of 17,360Da matches exactly the expected molecular weight and no other protein species were detected in the spectrum (Figure 26). It was important that the measured mass was identical to the specific sequence of the protein for further HDX MS experiments.

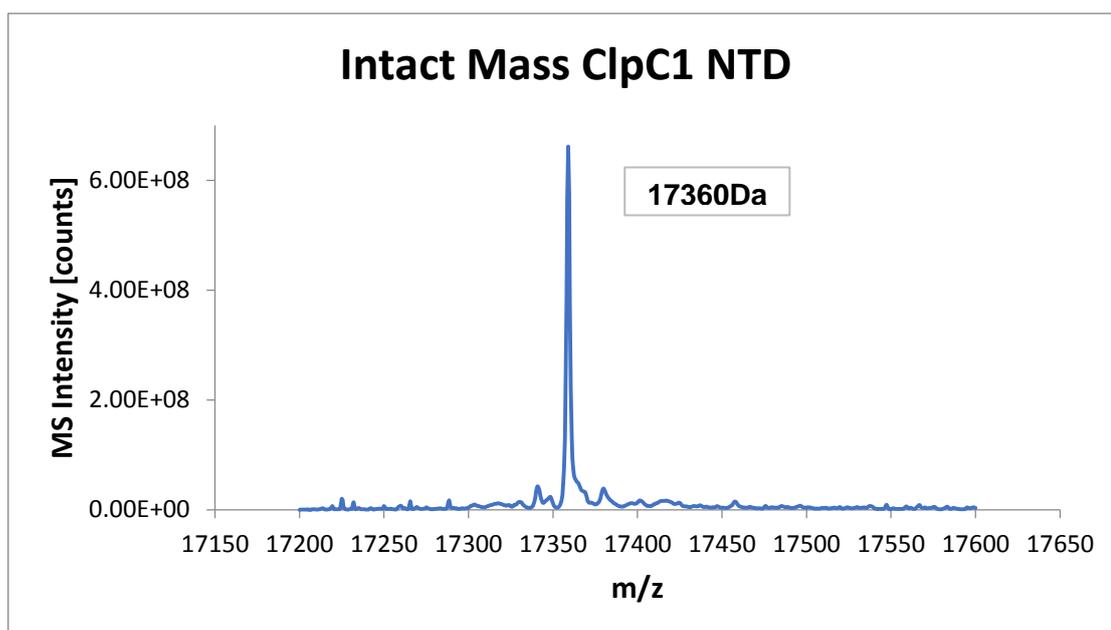


Figure 26: Intact Mass of ClpC1 NTD from *M. tuberculosis*. ESI positive, sensitivity mode and the mass spectra were acquired in MS mode.

3.2. Sequence Coverage

A peptide map for ClpC1 NTD was created by using the program PLGS and DynamX (Waters). The peptide map of ClpC1 shows good sequence coverage of 87.2% with 30 peptides and a redundancy of 2.71 (Figure 27). For this protein the chosen temperature for pepsin digestion was 20°C. The peptides corresponding to the C-terminus could not be detected, potentially due to the highly hydrophilic nature of the (hexahistidine tag).

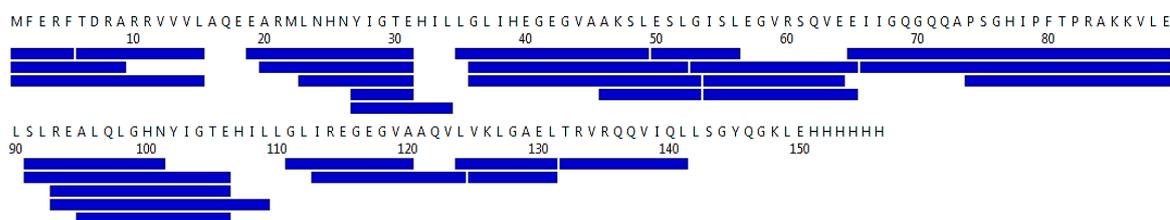


Figure 27: Coverage map for pepsin digestion. Online pepsin digestion of ClpC1 NTD resulted in of 87.2% sequence coverage with 30 peptides and 2.71 redundancy. Each blue bar represents an identified peptic peptide.

3.3. HDX MS

A 'chicklet plot' with the obtained HDX MS data from ClpC1 NTD [56] in absence of any ligand was generated to visualize the percentage of deuterium uptake for each generated peptide plotted as a function of time. In the absence of any ligand, ClpC1 NTD undergoes rapid deuteration and reaches a deuteration level of almost 60% at the end of the experimental time window of 4 hours. The kinetics for the sequence area around amino acids 2-37 shows a low rate of deuterium incorporation (Figure 28, A). Interestingly, by mapping the data onto the 3D structure, it is clearly visible that this area matches exactly the helices in the core region of the protein. This region could be part of a scaffold which provides stability for the ligand binding sites (Figure 28, B). The whole sequence area between amino acids 36-131 shows high deuterium incorporation after a few minutes of incubation. This whole area is part of the known cyclomarin A and phosphoarginine binding sites and it is not surprising to see that this area is highly dynamic in the absence of any ligand.

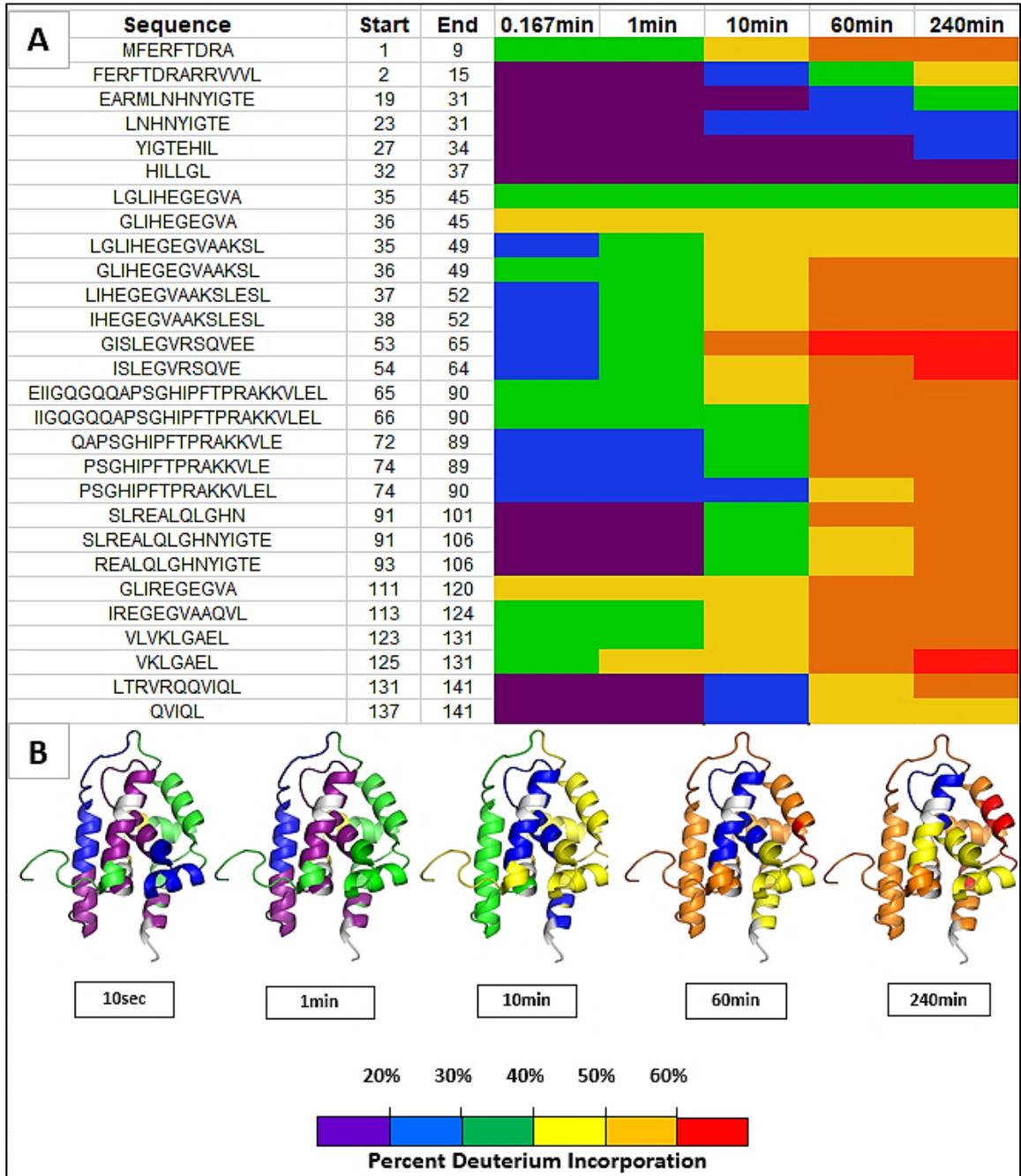


Figure 28: Chicklet Plot and cartoon structure representation of ClpC1 NTD. A: Percentage of deuterium uptake for each generated peptide. Time range of deuterium incubation 0.167, 1, 10, 60 minutes. B: Percentage of deuterium uptake mapped onto cartoon structure. Violet: under 20%, Blue: 20-30%, Green: 30-40%, Yellow: 40-50%, Orange: 50-60% and Red: above 60%. The structural figures were generated with PyMOL by using the PDB accession code 3wdb [56].

Figure 29 demonstrates the uptake differences between NTD ClpC1 and the NTD ClpC1 complexes with cyclomarin A and phosphoarginine. Time dependent data were acquired for the time range from 10 seconds to 4 hours. For evaluation the generally accepted experimental uncertainty of ± 0.5 Da for HDX difference measurements were used. Therefore, deuterium uptake differences up to 0.5Da were not considered and colored in grey in the “chicklet plot”. The blocks which are marked in red symbolize areas close to the known cyclomarin A binding site and the green blocks symbolize areas close to the known phosphoarginine binding sites. In the presence of cyclomarin A (A) areas close to the binding site which also includes the N-terminus show strong protection. Surprisingly, a small part in the area from one of the two phosphoarginine binding sites shows a higher amount of deuterium incorporation. Possibly the binding of cyclomarin A leads to a dis-ordering of this area which would make it more difficult for phosphoarginine to bind in the presence of cyclomarin A. In the presence of phosphoarginine (B) areas close to the known binding sites and areas of the cyclomarin A binding site show strong protection. A reason could be that the phosphoarginine binding site (1) partially overlaps with the cyclomarin A binding site. The results from the ClpC1 complex with cyclomarin A and phosphoarginine (C) demonstrate that the ordering effects upon binding of the two ligands are additive.

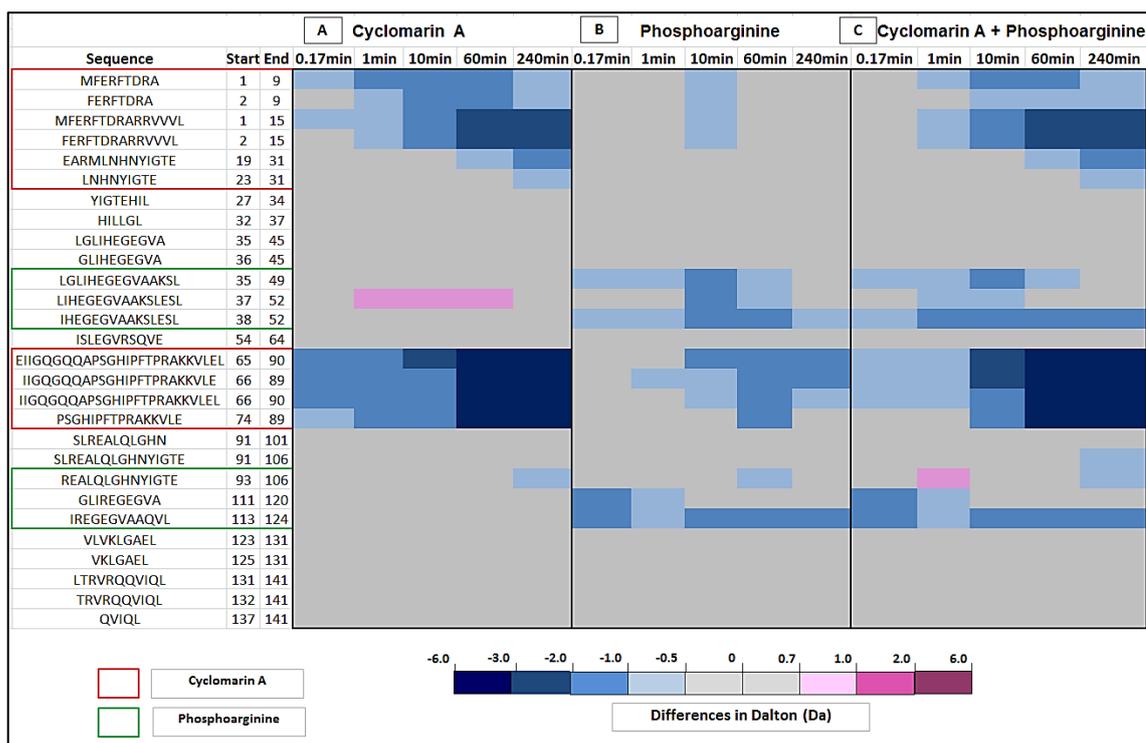


Figure 29: Differences in Da between ClpC1 to the corresponding ClpC1 interaction with cyclomarin A and phosphoarginine. Blue colour code: between -0.6 and -0.5Da, Grey: no significant differences, Pink colour code: between 0.5 and 6.0Da. Red Block: Cyclomarin A binding site and the Green Block: phosphoarginine binding site.

Figure 30 shows the observed major differences in deuterium uptake between ClpC1 and its complexes with cyclomarin A and phosphoarginine mapped onto the published 3D structure. Up to now there are only structural data available for the NTD ClpC1 from *Mycobacterium tuberculosis* bound with cyclomarin A. For displaying the binding of phosphoarginine the X-ray structure of *Bacillus subtilis* NTD ClpC was used because no structure is available for the ClpC1 NTD in complex with phosphoarginine. For displaying the binding sites of both ligands in one structure the two structures were overlaid in PyMol. The sequence alignment (ExpASy) shows a sequence identity of 60.8% between ClpC (*Bacillus subtilis*) and ClpC1 (*Mycobacterium tuberculosis*). In contrast to the NMR data, HDX MS shows dynamic differences on a timescale from seconds to hours. However, the findings of NMR could not be confirmed on the timescale that can be observed by HDX MS.

In the presence of cyclomarlin A only areas close to the binding site show strong protection and not as expected from the NMR data a complete rigidification of the N-terminal domain. In the case of phosphoarginine the protein shows protection around the binding sites and does not show signs of increased flexibility. Furthermore, the obtained HDX MS data confirmed that simultaneous binding of both ligands is possible, which indicates that cyclomarlin A does not influence the binding site of phosphoarginine.

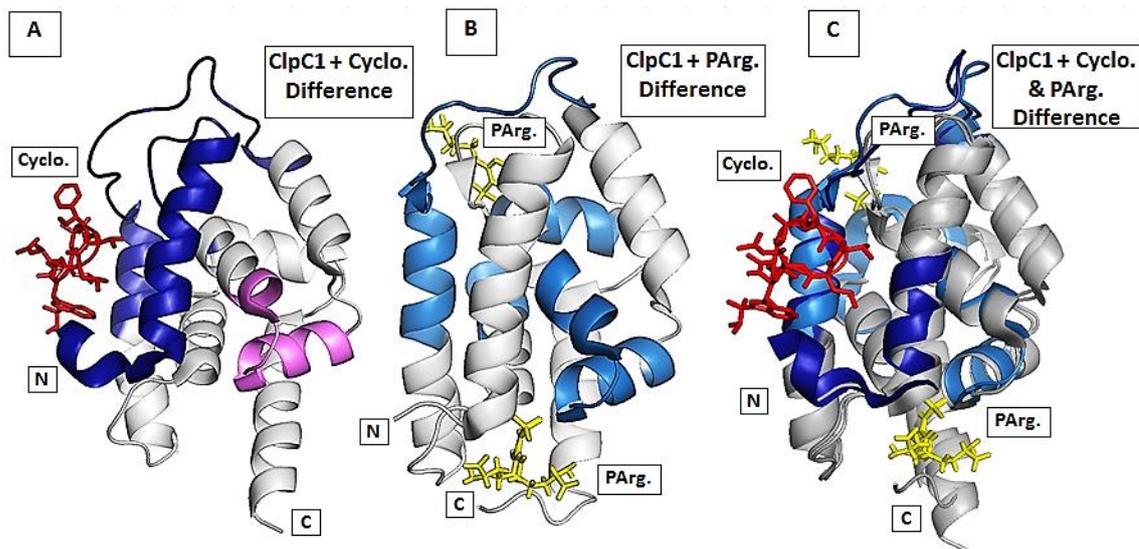


Figure 30: Differences of ClpC1 with cyclomarlin A and phosphoarginine mapped onto cartoon structure. A: representation of ClpC1 with cyclomarlin A, B: ClpC1 with phosphoarginine and C: ClpC1 with cyclomarlin A and phosphoarginine. Blue colour represents the major changes in the present of a substrate and Grey: no changes. So far, an X-ray structure is only available for the NTD *Mtb*ClpC1 but not for the whole *Mtb*ClpC1 complex. The structural figures were generated with PyMOL by using the PDB accession code 3wdc (ClpC1 with cyclomarlin A) [56] and 5hbn (ClpC1 with phosphoarginine) [53].

4. Conclusion

Energy-dependent proteases are essential to ensure that only correctly folded proteins are present in the cell. Very recently a degradation system was found in gram-positive bacteria where phosphorylated arginine residues serve as a tag for degradation of proteins by the ClpC-ClpP proteolytic complex. In this chapter HDX MS was used to study the dynamics of the N-terminal domain of ClpC1 which is part of the bacterial ClpC-ClpP proteolytic complex. Here, the potential dynamic effects caused by ligand binding were investigated. In general, it is of great interest to understand how ligand-induced changes in protein dynamics can influence the subsequent steps towards protein degradation.

The obtained HDX MS results show that the addition of cyclomarin A and phosphoarginine leads to rigidification of their binding sites with no effects on the overall dynamics which were observed on the timescale of the published NMR measurements. HDX MS data confirmed that simultaneous binding of both ligands is possible, which indicates that cyclomarin A does not influence the binding site of phosphoarginine. This means that the mode of action of cyclomarin A does not affect the affinity of the phosphoarginine binding sites for phosphoarginine but it indicates that the antibiotic works in a different way. It could for example lead to dissociation of the ClpC1 hexamer which further leads to the loss of function. It has also been suggested that cyclomarin A binding might lead to bacterial cell death by hyper-activating the ATPase activity of ClpC1. However, the bacterial ClpC1 complex from *Mycobacterium tuberculosis* is a novel potential drug target. It will take more detailed studies to understand the mechanism by which the N-terminal domain of ClpC1 initiates and facilitates protein degradation in the ClpCP1P2 proteasome.

CHAPTER 5

CONCLUSION & OUTLOOK

In the first chapter, the study describes the optimization of the system for getting robust and reproducible HDX MS data. Temperature conditions were optimized to perform pepsin digestion with high sequence coverage and minimal back exchange. The experimental uncertainty in deuterium uptake measurements was investigated to obtain the limit of differences in deuterium uptake that can be measured and reported with confidence (Chapter 2). The following chapters described the application of the technology to study the effects of protein-ligand interaction on the protein dynamics. The studies from *hGMPK* demonstrated that the protein undergoes large conformational changes in the presence of ligands. The inhibitor AP5G blocks the dynamics of the protein completely and it switched from an open to a closed conformation. With this work the system was established and could now be used to characterize changes in protein dynamics resulting from potential drug target candidates like *hGMPK* (Chapter 3). The N-terminal domain of ClpC1 from *M. tuberculosis* which is also a potential target for drug discovery was investigated by HDX MS for the first time. The data provided an insight into the dynamic system of the NTD ClpC1 in the presence of the ligands cyclomarin A and phosphoarginine. HDX MS results showed specific protection on the binding sites but did not indicate that ligand binding leads to global changes in the dynamics of the protein (Chapter 4). It would be highly interesting to perform further HDX MS experiments with the arginine phosphorylated substrate, β -casein. From the paper of Tim Clausen [53], co-workers demonstrated that arginine phosphorylated substrate, β -casein can be produced but unfortunately no structure could be solved for the complex.

Appendix

Phosphorylase b, rabbit muscle (MW: 97,289)

MSRPLSDQEKRKQISVRGLAGVENVTELVKDRNVATPRDYYFALAHT
VRDHLVGRWIRTQQHYEYKDPKRIYYLSLEFYMGRTLQNTMVNLALLENACDEATYQLGL
DMEELEEIEEDAGLGNGLGRLAACFLDSMATLGLAAYGYGIRYEFGI FNQKICGGWQM
EEADDWLRYGNPWEKARPEFTLPVHFYGRVEHTSQGAKWVDTQVVLAMPYDTPVPGYRN
NVVNTMRLWSAKAPNDFNLKDFNVGGYIQAVLDRNLAENISRVLYPNDNFFEGKELRLK
QEYFVVAATLQDI IRRFKSSKFGCRDPVRTNFDAFPDKVAIQLNDRHPSLAIPPELMRVL
VDLERLDWDKAWEVTVKTCAYTNHTVLPALERWPVHLLLETLLPRHLQIIYEINQRFLN
RVAAAFPGDVRRLRMSLVEEGAVKRINMAHLCAIAGSHAVNGVARIHSEILKKTIFKDF
YELEPHKFQNKTNGITPRRWLVLCNPGLAEIIAERIGEEYISDLDQLRKLKLSYVDDEAF
IRDVAKVKQENKLFKFAAYLREYKVINPNSLFDVQVKRIHEYKRQLLNCLHVITLYNR
IKKEPNKFVVPRTVMIGGKAAPGYHMAKMI IKLITAI GDVVNHDPVVGDRLRVIFLENY
RVSLAEKVI PAADLSEQISTAGTEASGTGNMKFMLNGALTIGTMDGANVEMAE EAGEEN
FFIFGMRVEDVDRLDQRGYNAQEYYDRIPELRQII EQLSGFFSPKQPDLFKDIVNMLM
HHDRFKVFADYEEYVKCQERVSALYKNPREWTRMVI RNIATSGKFS SDRTIAQYAREIW
GVEPSRQRLPAPDEKIP

GMPK, human (MW: 22,307 Da)

GHMLGSMGPRPVVLSGSPGAGKSTLLKRLQLQEHSGIFGFSVSHTRNPRPGEENGKDY
YFVTREVMQRDIAAGDFIEHAEFSGNLYGT SKVAVQAVQAMNRCVLDVDLQGVNRKA
TDLRPIYISVQPPSLHVLEQRLRQRNTEETESLVKRLAAAQADMESSEKPEGLFDVVI IN
DSLQAYAEELKEALSEEIKKAQTGA

GMPK with an HIS-TAG, human (MW 24,500 Da)

MGSSHHHHHSSGENLYFQGHMLGSMGPRPVVLSGSPGAGKSTLLKRLQLQEHSGIFGF
SVSHTRNPRPGEENGKDY YFVTREVMQRDIAAGDFIEHAEFSGNLYGT SKVAVQAVQA
MNRICVLDVDLQGVNRKATDLRPIYISVQPPSLHVLEQRLRQRNTEETESLVKRLAAA
QADMESSEKPEGLFDVVI INDSLQAYAEELKEALSEEIKKAQRTGA

ClpC1 NTD, Mycobacterium tuberculosis (MW 23,300 Da)

MFERFTDRARRVVVLAQEEARMLNHNYIGTEHILLGLIHEGEGVAAKSLESLSLEGV
RSQVEEIIIGQQQAPSGHIPFTPRAKKVLELSLREALQLGHNYIGTEHILLGLIREGEG
VAAQVLVKLGAELTRVRQQVIQLLSGYQGKLEHHHHHH

References

1. Houde, D.J. and S.A. Berkowitz, *Biophysical Characterization of Proteins in Developing Biopharmaceuticals*. 2014, Oxford, NETHERLANDS: Elsevier.
2. Ganellin, C.R., R. Jefferis, and S.M. Roberts, *Introduction to Biological and Small Molecule Drug Research and Development : Theory and Case Studies*. 2013, London, UNITED KINGDOM: Elsevier.
3. Bonetta, R., et al., *Role of Protein Structure in Drug Discovery*. XJENZA, 2000. **2016**: p. 03.
4. Masson, G.R., M.L. Jenkins, and J.E. Burke, *An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery*. *Expert Opinion on Drug Discovery*, 2017. **12**(10): p. 981-994.
5. Marciano, D.P., V. Dharmarajan, and P.R. Griffin, *HDX-MS guided drug discovery: small molecules and biopharmaceuticals*. *Current opinion in structural biology*, 2014. **28**: p. 105-111.
6. Berk, A., S. Zipursky, and H. Lodish, *Molecular Cell Biology 4th edition*. 2000, National Center for Biotechnology Information's Bookshelf.
7. Huang, R.Y.-C. and G. Chen, *Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry*. *Analytical and Bioanalytical Chemistry*, 2014. **406**(26): p. 6541-6558.
8. Picknett, T.M. and S. Brenner, *X-Ray Crystallography*, in *Encyclopedia of Genetics*, S. Brenner and J.H. Miller, Editors. 2001, Academic Press: New York. p. 2154.
9. Ryu, W.-S., *Chapter 2 - Virus Structure*, in *Molecular Virology of Human Pathogenic Viruses*, W.-S. Ryu, Editor. 2017, Academic Press: Boston. p. 21-29.
10. Wlodawer, A., et al., *Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures*. *The FEBS journal*, 2008. **275**(1): p. 1-21.
11. Gerothanassis, I.P., et al., *NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY: BASIC PRINCIPLES AND PHENOMENA, AND THEIR APPLICATIONS TO CHEMISTRY, BIOLOGY AND MEDICINE*. *Chemistry Education Research and Practice*, 2002. **3**(2): p. 229-252.
12. Cavanagh, J., et al., *Protein NMR spectroscopy: principles and practice*. 1995: Elsevier.
13. Schmid, F.X., *Biological macromolecules: UV-visible spectrophotometry*. e LS, 2001.
14. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure*. *Nature protocols*, 2006. **1**(6): p. 2876-2890.
15. Ghisaidoobe, A.B.T. and S.J. Chung, *Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques*. *International journal of molecular sciences*, 2014. **15**(12): p. 22518-22538.
16. Borotto, N.B., *The Application of Hydrogen/Deuterium Exchange and Covalent Labeling Coupled with Mass Spectrometry to Examine Protein Structure*. 2016.
17. Engen, J.R. and T.E. Wales, *Analytical Aspects of Hydrogen Exchange Mass Spectrometry*. *Annual review of analytical chemistry (Palo Alto, Calif.)*, 2015. **8**: p. 127-148.

18. Busenlehner, L.S. and R.N. Armstrong, *Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry*. Archives of Biochemistry and Biophysics, 2005. **433**(1): p. 34-46.
19. Englander, S.W., et al., *Hydrogen exchange: the modern legacy of Linderstrøm-Lang*. Protein science : a publication of the Protein Society, 1997. **6**(5): p. 1101-1109.
20. Konermann, L., J. Pan, and Y.-H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics*. Chemical Society Reviews, 2011. **40**(3): p. 1224-1234.
21. Krishna, M.M.G., et al., *Hydrogen exchange methods to study protein folding*. Methods, 2004. **34**(1): p. 51-64.
22. Ahn, J., *Hydrogen–Deuterium Exchange Mass Spectrometry: An Emerging Biophysical Tool for Probing Protein Behavior and Higher-Order Structure*. 2013.
23. Walters, B.T., et al., *Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment*. Journal of the American Society for Mass Spectrometry, 2012. **23**(12): p. 2132-2139.
24. Skinner, J.J., et al., *Protein dynamics viewed by hydrogen exchange*. Protein science : a publication of the Protein Society, 2012. **21**(7): p. 996-1005.
25. Mayne, L., *Hydrogen Exchange Mass Spectrometry*. Methods in enzymology, 2016. **566**: p. 335-356.
26. Konermann, L., X. Tong, and Y. Pan, *Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches*. Journal of Mass Spectrometry, 2008. **43**(8): p. 1021-1036.
27. Suchanova, B. and R. Tuma, *Folding and assembly of large macromolecular complexes monitored by hydrogen-deuterium exchange and mass spectrometry*. Microbial cell factories, 2008. **7**: p. 12-12.
28. Maier, C.S. and M.L. Deinzer, *Protein Conformations, Interactions, and H/D Exchange*, in *Methods in Enzymology*. 2005, Academic Press. p. 312-360.
29. Jonsson, A., *Mass spectrometry for protein and peptide characterisation*. Cellular and Molecular Life Sciences CMLS, 2001. **58**(7): p. 868-884.
30. Ho, C.S., et al., *Electrospray ionisation mass spectrometry: principles and clinical applications*. The Clinical biochemist. Reviews, 2003. **24**(1): p. 3-12.
31. de Hoffman, E. and V. Stroobant, *Tandem mass spectrometry*. Mass spectrometry principles and applications, 3rd Edition, John Wiley & Sons Ltd., West Sussex, England, 2007: p. 189-215.
32. Lane, C.S., *Mass spectrometry-based proteomics in the life sciences*. Cellular and Molecular Life Sciences CMLS, 2005. **62**(7): p. 848-869.
33. Honour, J., *Benchtop mass spectrometry in clinical biochemistry*. Vol. 40. 2003. 628-38.
34. Chernushevich, I., A. V. Loboda, and B. Thomson, *An Introduction to quadrupole-time-of-flight mass spectrometry*. Vol. 36. 2001. 849-65.
35. Wales, T.E. and J.R. Engen, *Hydrogen exchange mass spectrometry for the analysis of protein dynamics*. Mass Spectrometry Reviews, 2006. **25**(1): p. 158-170.
36. Busenlehner, L.S., et al., *Structural Elements Involved in Proton Translocation by Cytochrome c Oxidase as Revealed by Backbone Amide Hydrogen–Deuterium Exchange of the E286H Mutant*. Biochemistry, 2008. **47**(1): p. 73-83.

37. Wei, H., et al., *Using hydrogen/deuterium exchange mass spectrometry to study conformational changes in granulocyte colony stimulating factor upon PEGylation*. Journal of the American Society for Mass Spectrometry, 2012. **23**(3): p. 498-504.
38. Ahn, J., et al., *Pepsin immobilized on high-strength hybrid particles for continuous flow online digestion at 10,000 psi*. Analytical chemistry, 2012. **84**(16): p. 7256-7262.
39. Hudgens, J.W., R.Y.-C. Huang, and E. D'Ambro, *Method validation and standards in hydrogen/deuterium exchange mass spectrometry*.
40. Li, J., et al., *Hydrogen–Deuterium Exchange and Mass Spectrometry Reveal the pH-Dependent Conformational Changes of Diphtheria Toxin T Domain*. Biochemistry, 2014. **53**(43): p. 6849-6856.
41. Plumb, R.S., et al., *UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation*. Rapid Communications in Mass Spectrometry, 2006. **20**(13): p. 1989-1994.
42. Topalis, D., et al., *Novel Antiviral C5-Substituted Pyrimidine Acyclic Nucleoside Phosphonates Selected as Human Thymidylate Kinase Substrates*. Journal of Medicinal Chemistry, 2011. **54**(1): p. 222-232.
43. Sekulic, N., et al., *Structural Characterization of the Closed Conformation of Mouse Guanylate Kinase*. Journal of Biological Chemistry, 2002. **277**(33): p. 30236-30243.
44. Jain, R., et al., *Insights into open/closed conformations of the catalytically active human guanylate kinase as investigated by small-angle X-ray scattering*. European Biophysics Journal, 2016. **45**(1): p. 81-89.
45. Hible, G., et al., *Calorimetric and Crystallographic Analysis of the Oligomeric Structure of Escherichia coli GMP Kinase*. Journal of Molecular Biology, 2005. **352**(5): p. 1044-1059.
46. Jerabek-Willemsen, M., et al., *MicroScale Thermophoresis: Interaction analysis and beyond*. Journal of Molecular Structure, 2014. **1077**: p. 101-113.
47. Ferrige, A.G., et al., *The Application of Maxent to Electrospray Mass Spectrometry*, in *Maximum Entropy and Bayesian Methods: Seattle, 1991*, C.R. Smith, G.J. Erickson, and P.O. Neudorfer, Editors. 1992, Springer Netherlands: Dordrecht. p. 327-335.
48. Hible, G., et al., *Crystal structures of GMP kinase in complex with ganciclovir monophosphate and Ap5G*. Biochimie, 2006. **88**(9): p. 1157-1164.
49. Kleiger, G. and T. Mayor, *Perilous journey: a tour of the ubiquitin-proteasome system*. Trends in cell biology, 2014. **24**(6): p. 352-359.
50. Burns, K.E. and K.H. Darwin, *Pupylation versus ubiquitylation: tagging for proteasome-dependent degradation*. Cellular microbiology, 2010. **12**(4): p. 424-431.
51. Malik, I.T. and H. Brötz-Oesterhelt, *Conformational control of the bacterial Clp protease by natural product antibiotics*. Natural Product Reports, 2017. **34**(7): p. 815-831.
52. Marsee, J.D., et al., *Mycobacterium tuberculosis ClpC1 N-Terminal Domain Is Dispensable for Adaptor Protein-Dependent Allosteric Regulation*. International journal of molecular sciences, 2018. **19**(11): p. 3651.
53. Trentini, D.B., et al., *Arginine phosphorylation marks proteins for degradation by a Clp protease*. Nature, 2016. **539**: p. 48.

54. Weinhäupl, K., et al., *The antibiotic cyclomarin blocks arginine-phosphate-induced millisecond dynamics in the N-terminal domain of ClpC1 from Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 2018: p. jbc.RA118. 002251.
55. Thöny, V., *Biophysical studies of bacterial proteasome proteins*. 2018, Vienna.
56. Vasudevan, D., S.P.S. Rao, and C.G. Noble, *Structural basis of mycobacterial inhibition by cyclomarin A*. *The Journal of biological chemistry*, 2013. **288**(43): p. 30883-30891.