# Structural Mass Spectrometry Analysis of RNA Chaperones and Protein Therapeutics

KERENE BROWN

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO

JANUARY 2020

**Abstract**

Since the advent of soft ionization techniques, most notably electrospray ionization (ESI), the application of mass spectrometry (MS) analysis has expanded to intact biomolecules. ESI-MS gave rise to the field of structural MS (a relatively new tool in structural biology) which involves different types of MS techniques such as: Native, Ion Mobility, chemical labelling and proteomics. Conventional techniques such as Nuclear Magnetic Resonance (NMR) and X-ray crystallography provide high resolution structures of proteins. Although MS cannot provide the same level of resolution, it provides important structural information such as: binding stoichiometry, mass, size, subunit composition, ligand binding sites, stability, amino acid sequence and post-translational modifications (PTMs). In this work, structural MS was employed as a tool in the characterization of a crucial RNA chaperone, human La (hLa) protein, that is implicated in the processing of various types of cellular and viral RNAs. Many RNA chaperones such as hLa have large unstructured regions which make them difficult to study on a structural basis. Work done here revealed that the C-terminal domain (CTD) of hLa plays a role in binding RNA and that it binds structured and unstructured RNA using distinct dynamic modes. In addition, work was done to support a previous hypothesis about the presence of an interdomain interaction in hLa that affects nuclear trafficking and RNA binding. Additionally, the same structural MS techniques were used to assess the biosimilarity between a biological drug, Avastin, and its biosimilar in pre-clinical development. Work was done to test the structural similarity between the two drugs and the batch-to-batch variability of the biosimilar. Structural MS proved to be an effective technique in the rapid and facile characterization of biosimilarity and in the analysis of RNA binding activity of RNA chaperones.

A special thanks goes to my parents Veronica and Uriah Brown for being the most supportive parents throughout the 24 years I have spent in school and my family members (Mariah Brown, Kevin Hunter, Howard Thomas, Kaysia Thomas, Enid Murray, Shenice Douglas, Yanique Ellison). And of course, my loving husband Menroy McLean and my darling daughter Kyra McLean for providing emotional support through stressful times.

**Table of Contents**

# List of Tables

# List of Figures

# List of Abbreviations

BME – βeta Mercaptoethanol

CCS – Collision Cross Section

CEM – Chain Ejection Model

CID – Collision Induced Dissociation

CIU – Collision Induced Unfolding

CTD – Carboxy Terminal Domain

CRM – Charge Residue Model

DDA – Data Dependent Acquisition

DIA – Data Independent Acquisition

ECD – Electron Capture Dissociation

EDTA - Ethylenediaminetetraacetic acid

ESI – ElectroSpray Ionization

ETD – Electron Transfer Dissociation

EU – European Union

FDA – Food and Drug Administration

FMDV – Foot and Mouth Disease Virus

HAV – Hepatitis A Virus

HBV – Hepatitis B Virus

HCV – Hepatitis C Virus

HDX – Hydrogen Deuterium Exchange

hLa – human La

id – Inner Diameter

IEM – Ion evaporation Model

IgG – Immunoglobulin G

IMAC – Immobilized Metal Affinity Chromatography

IMS – Ion Mobility Spectrometry

IPTG - Isopropyl β-D-1-thiogalactopyranoside

IRES – Internal Ribosome Entry Site

LAM – La Motif

LARPs – La Related Proteins

LB - Luria Bertani

m/z – mass to charge

mAbs – Monoclonal Antibody

MS – Mass Spectrometry

NES – Nuclear Export Signal

NLS – Nuclear Localization Signal

NMR – Nuclear Magnetic Resonance

NRE -Nuclear Retention Element

NTD – N-Terminal Domain

OD – Optical density

od – Outer Diameter

PMMA – Polymethylmethacrylate

PTM – Post Translation Modification

PV – Poliovirus

RNA – Ribonucleic Acid

RRM – RNA Recognition Motif

SBM – Short Basic Motif

SDS-PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

TCEP - (tris(2-carboxyethyl)phosphine)

TRESI – Time Resolved Electrospray Ionization

TWIMS – Travelling Wave Ion Mobility Spectrometry

VEGF – Vascular Endothelial Growth Factor

# List of Publications

1. **Brown, K.A**., Lento, C., Rajendran, S., Dowd, J., and Wilson, D.J. (2020) Hydrogen deuterium exchange epitope mapping as a tool for early screening of bevacizumab: a pre-clinical biosimilar monoclonal antibody. *Manuscript in preparation.*

2. **Brown, K.A**., Rajendran, S., Dowd, J., Wilson, D.J. (2019) Rapid Characterization of Structural and Functional Similarity for a Candidate Bevacizumab (Avastin) Biosimilar using a Multipronged Mass Spectrometry-Based Approach. *Drug Testing and Analysis*, doi: 10.1002/dta.2609.

3. Marella, S.A., **Brown, K.A.,** Mansouri-Noori F, Porat J., Wilson, D.J. and Bayfield, M.A. (2018) An interdomain bridge influences RNA binding of the human La protein. *Journal of Biological Chemistry*, **294** (5), 1529-1540.

4. **Brown, K.A**. and Wilson, D.J. (2017) Bottom-up hydrogen deuterium exchange mass spectrometry-data analysis and interpretation. *Analyst*, **142**, 2874-2886.

5. **Brown, K.A**., Sharifi, S., Hussain, R., Donaldson, L., Bayfield, M.A. and Wilson, D.J. (2016) Distinct Dynamic Modes Enable the Engagement of Dissimilar Ligands in a Promiscuous Atypical RNA Recognition Motif. *Biochemistry*, **50** (51), 7141-7150.

6. Lento, C., Zhu, S., **Brown, K.A**., Knox, R., Liuni, P. and Wilson, D.J. (2017) Time-Resolved Electrospray Ionization Hydrogen-Deuterium Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *J Vis. Exp*. (122), e55464, doi: 10.3791/55464.

# Chapter 1: Introduction

A Version of this chapter was published in *Analyst* 2017, 142, 2874-2886, Copyright

**Brown, K.A**. and Wilson, D.J. (2017) Bottom-up hydrogen deuterium exchange mass spectrometry-data analysis and interpretation. *Analyst*, **142**, 2874-2886.

Manuscript prepared by **KAB** with editorial contributions by DJW.

Proteins are large biomolecules responsible for many critical cellular functions such as enzymatic catalysis, transport and storage, regulating cellular and physiological activities, immune functions and providing structure and support for cells. Protein structure is of great interest as function and structure are closely related. Proteins however are not static but rather dynamic in nature and their functions also greatly rely on conformational changes and binding interactions with their respective targets.[1] Currently, the vast majority of high resolution structures are derived from X-ray crystallography and Nuclear Magnetic Resonance (NMR) experiments. Although Mass Spectrometry (MS) cannot provide these high-resolution structures, it can facilitate these techniques by providing low resolution structural details. Structural MS comprises a variety of MS methods (Figure 1.1) that are used to determine structural information of biomolecules, most commonly proteins. These methods include chemical labelling MS, ion mobility MS, native MS and top down MS. [2–4] Native MS is the heart of structural MS, as it involves the successful transmission of the native protein into the gas phase for subsequent analysis.[5,6] Chemical labelling experiments such as hydrogen deuterium exchange (HDX) and oxidative labelling, allows for the determination of binding sites in protein/ligand and protein/protein complexes as well as the conformational changes that occur upon target engagement. [2] Ion mobility MS adds an extra dimension to conventional MS analysis as it rapidly separates gas phase ions based on their mass, charge and overall shape (collision cross section) and is often used in the structural analysis of protein and protein complexes. It can also be used to study the gas phase stability of a protein in the presence of a ligand.[2,5] Top Down MS involves the gas phase fragmentation of protein molecule to determine its primary structure and the location of post translation modifications (PTMs).[7] In this dissertation the structural MS techniques focused on is Native MS, IM-MS and hydrogen deuterium exchange (HDX) MS.
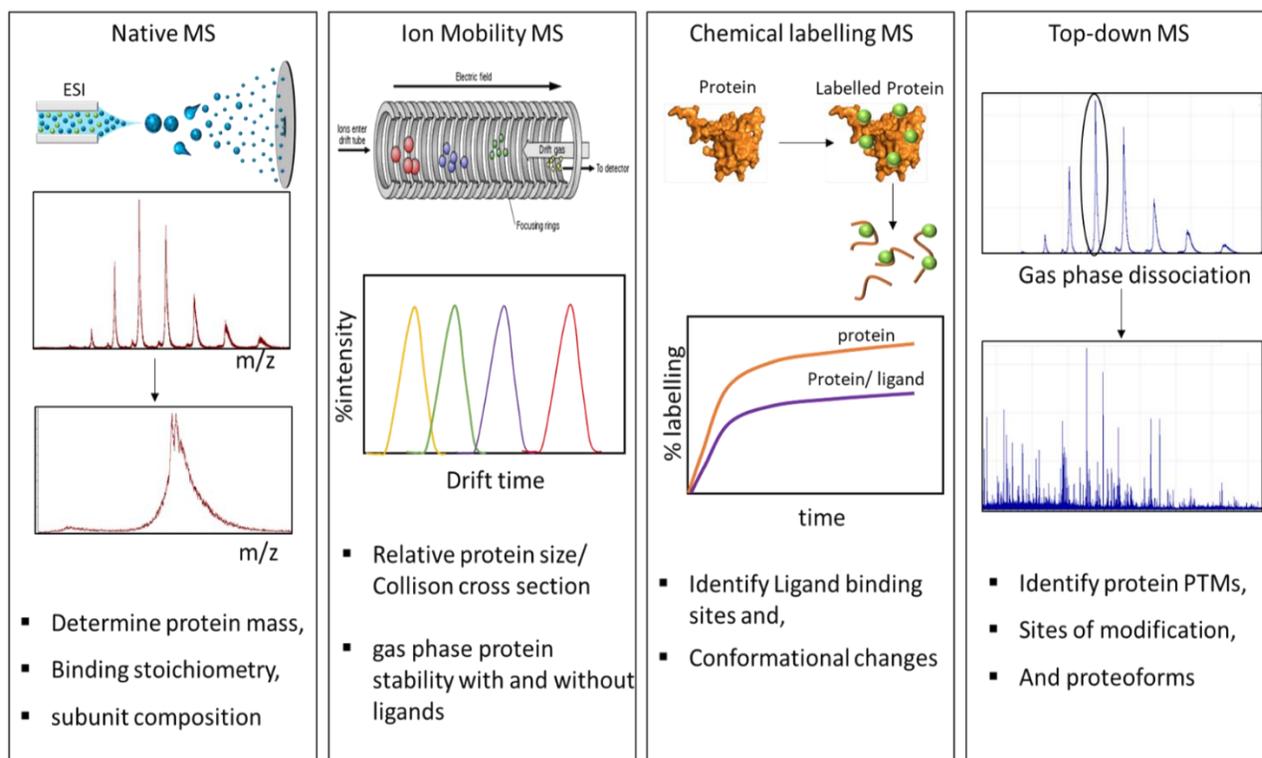
**Figure 1.1 Overview of structural mass spectrometry techniques and the information that can be obtained from each type of experiment.**

## 1.1 Native Mass Spectrometry

MS is an analytical technique used to measure the mass to charge ratio (m/z) of chemical compounds by first converting them into ions and then measuring their trajectory under a vacuum in an electric or magnetic field.[8–10] When first developed, MS was primarily used in the analysis of organic and inorganic chemical compounds as the early methods of ionization (such as electron impact ionization) were far too destructive to analyze biological molecules. In addition, early mass spectrometers did not have sufficient resolution required to analyze large biomolecules.[9] The development of "soft" ionization techniques have led to the widespread use of mass spectrometry analyses of intact biomolecules and biomolecular complexes.[11,12]

### 1.1.1 Electrospray Ionization

The most common "soft" ionization technique is electrospray ionization (ESI) that has the capabilities of transferring intact biological macromolecules into the gas phase. The development of ESI-MS was ground-breaking as a method was needed to transfer non-volatile compounds into the gas phase. It was originally developed by Malcolm Dole in the 1960's when Dole wanted to determine the mass of synthetic polymers by MS.[13] John Fenn and colleagues later expanded on this development by demonstrating the effectiveness of ESI-MS in the analysis of peptides and proteins [14] and that ESI-MS could be used to study analytes with mass range in the millions due to the production of multi-charged ions.[15] The development of ESI also lead to the combination of liquid chromatography (LC) and MS which was another important milestone in biomolecular MS analyses. [11]

During the electrospray process (Figure 1.2), there are three main steps to generate gas phase ions: 1) production of charged droplets at the tip of the capillary, 2) shrinkage of the charged droplet and 3) the production of the gas phase ions from the charged droplets. The resulting ions then enter the vacuum region of the interface leading to the mass spectrometer. This interface is typically a plate with an orifice or with a sampling capillary mounted to the plate leading to the mass spectrometer. The charged droplets are generated at the tip of the electrospray capillary. This is usually a metal capillary around 1 mm o.d. or less and located 1 − 3 cm away from the counter electrode (the plate with the orifice or sampling capillary). [16,17] There is a high electric field, $E_c$, generated at the tip of the electrospray capillary which depends on its radius $r_C$, the distance from the counter electrode, d, and the applied potential, $V_c$.[18] The relationship is given by:

$$E_c = \frac{2V_c}{r_c \times \ln\left(\frac{4d}{r_c}\right)} \tag{1.1}$$

The electric field is at its highest at the tip of the capillary resulting in the polarization of the liquid near the solution's meniscus. In positive ion mode, the spray capillary is at a positive voltage relative to the counter electrode resulting in positively charged droplets.[16] The electrolytes present in the solution will move under the influence of this electric field resulting in more positive ions at the surface of the meniscus and the negative ions will move away from the meniscus.[17] The downfield forces that are a result from the polarization causes a distortion in the meniscus forming a cone, commonly referred to as a Taylor cone.[19,20] If the electric field is high enough, the tip will become unstable and a fine jet emerges from the tip of the cone. The jet will then break up into small charged droplets due to charge repulsion. Because of the excess positive charge at the surface of the meniscus, the drops produced from the Taylor cone will be mostly positively charged. These charged droplets move toward the counter electrode and evaporation leads to shrinkage of the charged droplets. As droplets shrink and become smaller, the repulsion between the charges increase until the surface tension at the droplet surface is broken and causes a coulomb explosion. A cone and cone jet are formed on the charged droplets similar to what is observed at the tip of the capillary (the Taylor cone). This is what causes the formation of smaller droplets from the larger ones. This instability, referred to as a Coulomb fission occurs close to or at the Rayleigh limit given by the equation:[17]

$$Q_{Ry} = 8\pi(\varepsilon_0 \gamma R^3)^{1/2} \tag{1.2}$$

In which $Q_{Ry}$ is the charge on the droplet, $\gamma$ is the surface tension of the solvent, R the radius of the droplet and $\varepsilon_0$ the electrical permittivity. Repeated evaporation and fission of the charged droplets eventually lead to the formation of gas phase ions. [17]
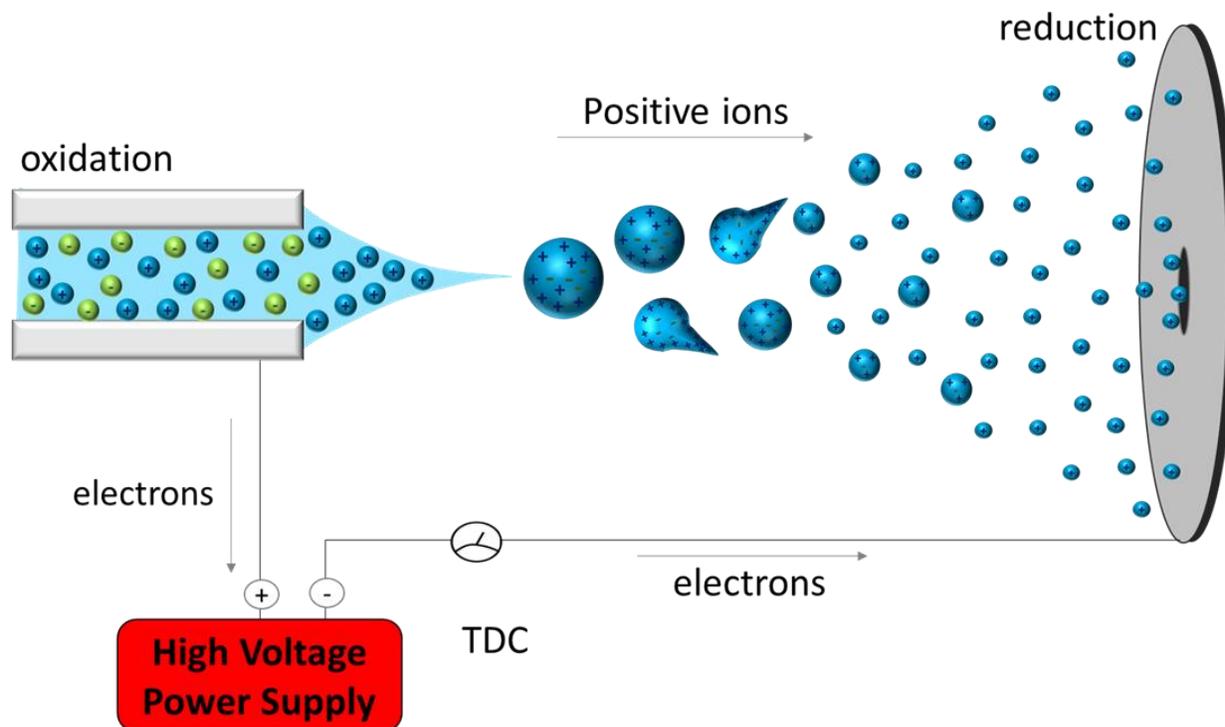
**Figure 1.2. Schematic representation of the electrospray process**. The electric field imposed on the solution results in an enrichment of positive ions near the meniscus causing a destabilization and subsequent formation of a Taylor cone and a jet charged with an excess of positive ions. This jet splits into droplets and evaporation brings the charges closer together. Columbic repulsion increases and causes a destabilization in the droplets resulting in the emission of charged progeny droplets. As evaporation continues the droplets get smaller until free gas phase ions are formed. TDC means total droplet current.

The electrospray spray device can be viewed as a special kind of electrolytic cell.[16] The oxidation reaction takes place at the tip of the capillary which leads to the excess of positive ions and production of electrons entering the metal. The reduction reaction on the other hand takes place at the counter electrode. It is referred to as a "special" electrolytic cell as the conversion of ions to electrons does not occur completely in solution as it does for a typical electrolytic cell. Instead, part of the ion transport takes place in the gas phase. [16]

6

### 1.1.2 Production of gas phase ions

There are three proposed mechanisms of how gas phase ions are produced from the small progeny droplets, the ion evaporation model (IEM),[21] the charge residue model (CRM) [22] and the chain ejection model (CEM) [23] (Figure 1.3).
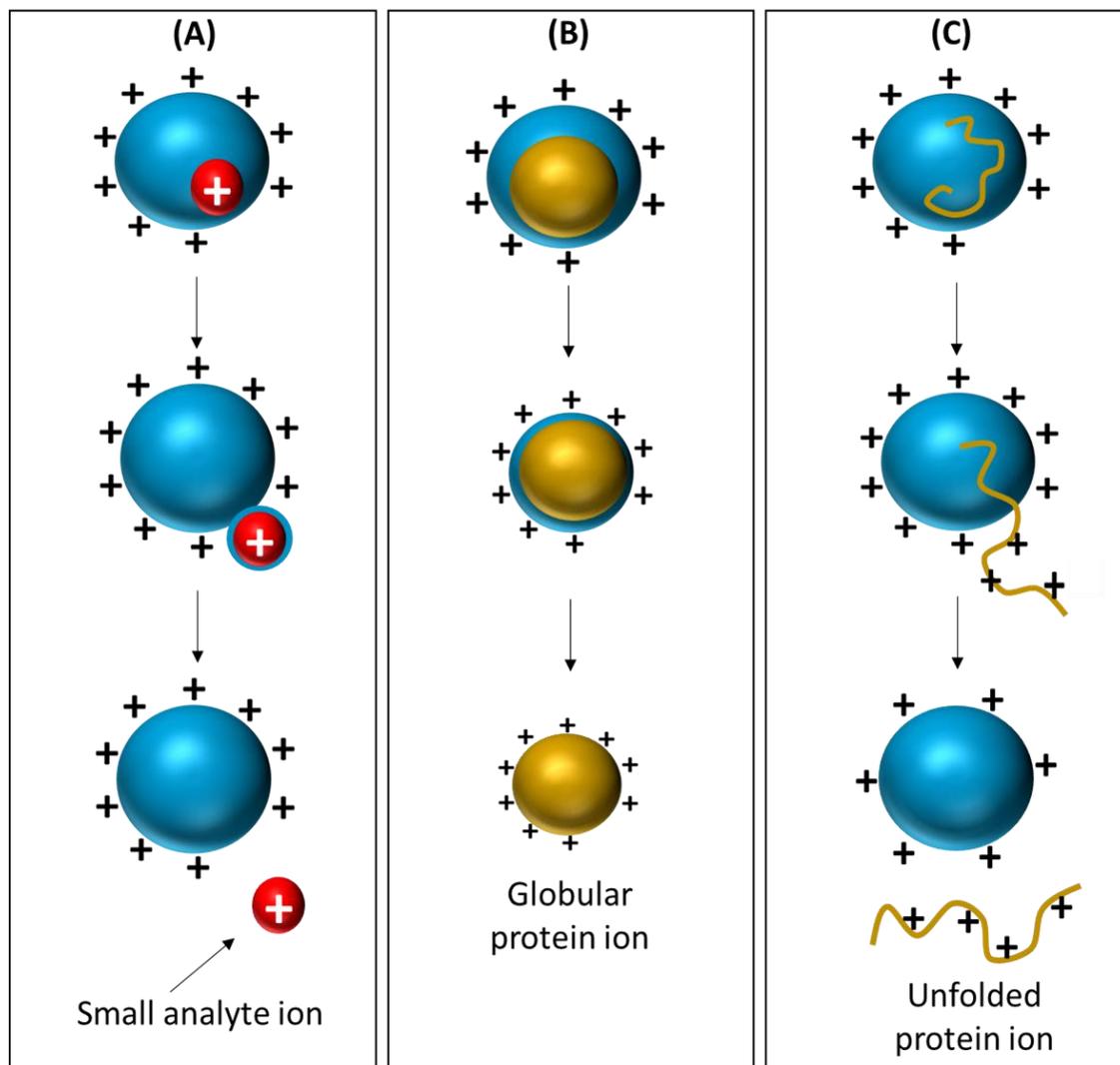


**Figure 1.3 Mechanism of ion formation during electrospray ionization**. (A) Ion evaporation model (IEM) in which small (in)organic molecular ions are produced through evaporation of the ion due to the large electric field applied to the surface of the droplet. (B) Charge residue model (CRM) for which globular protein ions are produced following the complete evaporation of the droplet and (C) the chain ejection model (CEM) in which unfolded protein ions are ejected from the droplet.

The IEM is more plausible and well supported experimentally for small inorganic and organic molecules. Ion evaporation is the emission of dissolved ions from highly charged liquid surfaces. Ions in an aqueous solution are unlikely to escape the solution if left to evaporate under normal conditions. This is because the ions are kept strongly in solution by electrostatic forces between the ionic charge and the dipole of the water molecules surrounding the ions. This allows the ions to move freely in the solution but there is an immense energy barrier when the ions approach the surface of the solution. However, when a larger enough electric field (1 -3 V/nm) is applied at the surface of the liquid, the potential energy of the ions at the surface can be increased enough to allow them to evaporate.[21] This occurs at the final stages of evaporation when the net charge on the droplet, combined with the small radius (approximately 10 nm) is enough to produce the electric field required for the ions to evaporate.[24,25] This mechanism though widely accepted for small molecules is very unlikely for large biomolecules.

It is expected that when native proteins are sprayed using ESI they remain in their natural folded state when sprayed from a neutral pH. However, unlike small (in)organic molecules, proteins typically have multiple different charges leading to a mass spectrum with multiple peaks each representing a different charge state. [17,26] The CRM assumes that these macro-ions are produced from small charged droplets containing a single macro-molecule. When the droplet completely evaporates the charges on the droplet are transferred to the molecule. As a result of this, it is common to observed dimers, trimers and even tetramers due to multiple macro-molecules in a single droplet when it completely evaporates. [13,22]

Although CRM is now widely accepted as the ESI mechanism for globular proteins, unfolded proteins are not thought to follow the same process, instead, the chain ejection model is proposed. In the CEM, the unfolded protein is driven to the surface of the droplet by electrostatic and

8

hydrophobic factors after which the protein undergoes gradual ejection via "tadpole-like" structures. Mobile $H^+$ migrates between the droplet and the protruding polypeptide tail causing the unfolded protein to depart as a highly charged ion. [23]

### 1.1.3 Adducts in Native MS

In Native MS analysis, it is important to have a mass spectrum that is as "clean" as possible with very little to no adducts. When proteins are to be analyzed the salt additive of choice is ammonium acetate in the millimolar concentration (typically 10 – 250 mM) as it produces very clean mass spectra of multiple charged proteins. One of the most common contaminants in analyzing proteins is sodium. Sodium adducts results in a mass shift which can result in misinterpretation of native spectrum especially when protein-ligand interactions are to be studied. At the final stages of evaporation in the ESI mechanism, the salt concentration increases resulting in ion pairing of positive and negative ions. The solute ions will also pair with charged residues near the surface of the protein. When sodium is present in the solution the sodium ion interacts with the ionized acidic residue or the amide group on the peptide backbone. This sodium bonding is extremely strong and collisional activation does not result in loss of the sodium ions. On the other hand when ammonium and acetate ions take part in the ion pairing reactions, when the protein is subjected to collisional activation, the ions are lost due to their relatively facile dissociation.[17,27]

### 1.1.4. What does the term "Native" represent?

After the development of ESI it became easier to analyze large biomolecular compounds such as intact proteins and protein complexes in pseudo-physiological conditions. This is referred to as Native MS. Although native MS has been in use since the 1990's the term "native MS" was only recently coined in 2004.[28] Terminology such as: non-denaturing, macromolecular, non-covalent or native spray is often used to describe this technique. Native MS should not be confused with

the biological definition of "native" that describes native as being the state of the protein in its natural environment, such as within the cell. MS is a gas phase measurement and cannot live up to this definition. The term "native" in native MS describes the biological state of the protein prior to the ionization event and entry into the mass spectrometer. Solution parameters are important to maintain such as the pH and ionic strength, so the protein maintains its native fold in solution. [28,29]

There has been ongoing debate in the mass spectrometry community as to whether, protein structures are even partially maintained in the gas phase for MS analysis. Assuming the CRM is true for proteins, during the ESI process, only solvent exposed basic residues are able to pick up a charge/proton. Proteins sprayed from an aqueous solvent at or near neutral pH exhibit a narrow charge state distribution whereas proteins sprayed from an organic solvent exhibited a wider charge state distribution with much higher charges. In organic solvents, a protein will have a more extended/unfolded structure, resulting in more charges being picked up.[28,29] This is evidence that protein structure is somewhat maintained through the ionization process.[30]

**1.2 Ion Mobility**

Ion mobility (IM) is an analytical technique that separates gas phase ions based on their mobility in a buffer gas. When combined with MS, it adds an additional dimension to the analysis. In IM-MS ions are separated by both their m/z as well as their collisional cross-section (CCS). The mobility of the ion (K) is related to its drift velocity ($V_D$) and the electric field (E). [31,32]

$$V_D = KE \tag{1.3}$$

The mobility, K, is typically reported as reduced mobility, $K_0$. This is the ion's mobility at standard temperature and pressure which is expressed as:

$$K_0 = \frac{P}{760} \times \frac{273.15}{T} \times K \tag{1.4}$$

The ion's mobility through a buffer gas depends on the ion's charge (q), the density of the buffer gas (N), the reduced mass of the ion-neutral complex (μ), the absolute gas temperature (T), and the collision cross section ($\Omega_0$) of the ions.[31] The mobility constant is then expressed as:

$$K = \frac{3}{16} \times \frac{q}{N} \times \left( \frac{1}{\mu} \times \frac{2\pi}{k_b T} \right)^{\frac{1}{2}} \times \frac{1}{\Omega_o} \tag{1.5}$$

In macromolecular analysis, the reduced mass of the ion approaches the mass of the buffer gas and as a result is treated as a constant. The mobility constant is therefore only dependent on the CCS of the gas phase ion. [31]

$$K \propto \frac{1}{\Omega_0} \tag{1.6}$$

### 1.2.1 Traveling Wave Ion Mobility Spectrometry

A different form of ion mobility known as traveling wave IMS (TWIMS) uses a strong electric field applied sequentially one segment at a time throughout out the IM cell. This results in the ions being moved in pulses as waves of the electric field passes through them. This technology has been commercialized by Waters corporation (Milford, USA) as the Synapt HDMS system depicted in Figure 1.4.[33]
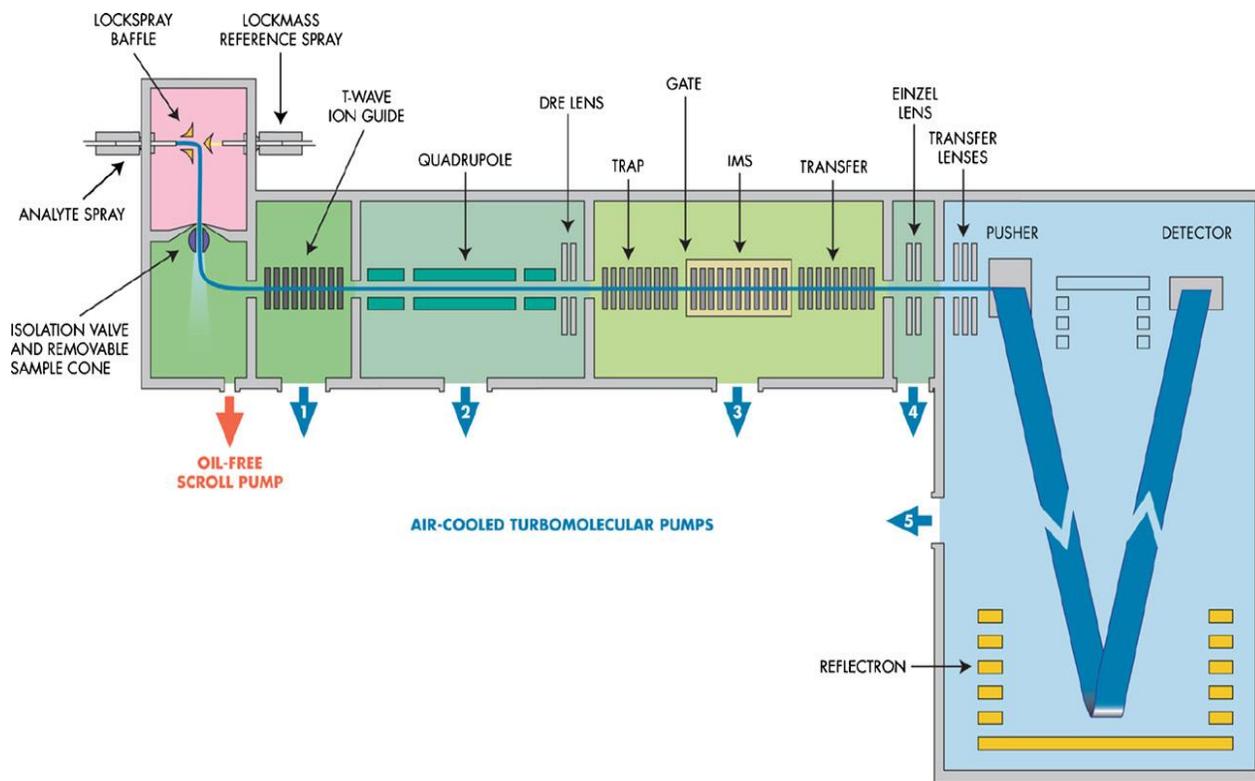
**Figure 1.4 Schematic diagram of the Synapt HDMS system**. The ion mobility cell is located between the quadrupole and the time of flight mass analyzer. Adapted with permission from Pringle *et al*., 2007. [33]

The ion guide in the TWIMS cell consists of a series of planar electrodes that are perpendicular to the ion transmission path with each adjacent electrode having opposite phases of RF voltage. To propel ions in the presence of background gas, for each pair of adjacent electrodes, a transient dc voltage is overlaid on the RF voltage in a repeating sequence along the length of the device (Figure 1.5). This creates a series of potential hills that are applied to the following pair of electrodes at regular time intervals. This creates a sequence of 'travelling waves'. The ions are driven from the potential hills and carried through the device with the waves.[33]
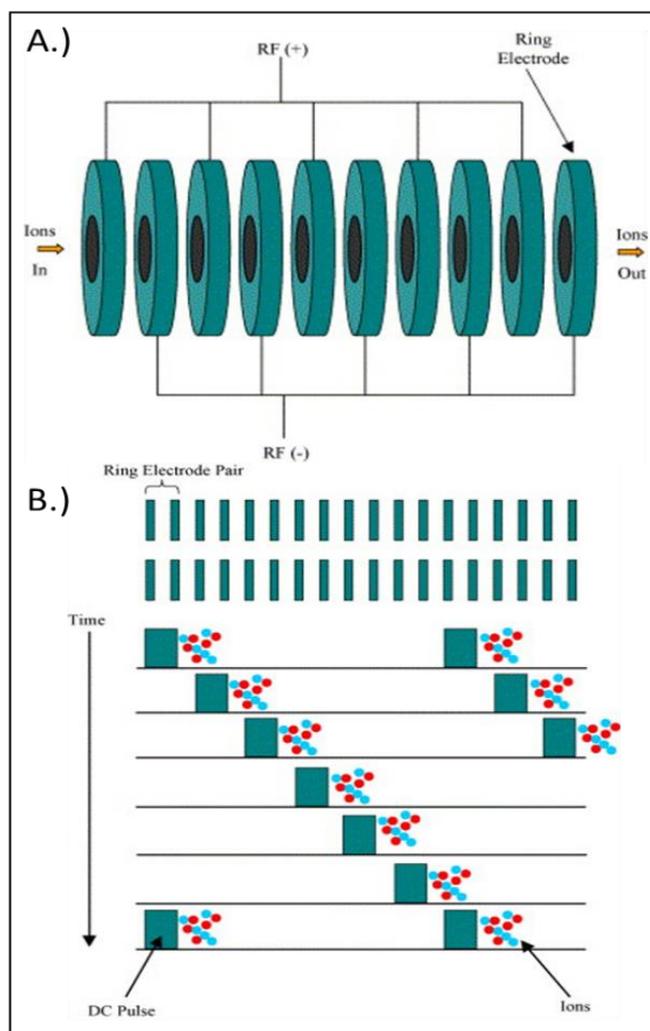
**Figure 1.5 Travelling wave ion mobility**. A.) A stacked ring ion guide and B.) Schematic representation of ion propulsion in the presence of background gas during the operation of a travelling wave ion guide. Adapted with permission from Pringle *et al*., 2007. [33]

The high pressures (0.5 – 1.0 mbar) used in the IM cell, results in lower mobility ions being overtaken by the travelling waves more often than higher mobility resulting in longer drift times. [33,34] Unlike conventional IMS, the electric field in TWIMS is not constant which makes it difficult to directly calculate $\Omega$ from $t_D$. It is easier to simply calibrate the TWIMS instrument using compounds with known $\Omega$ previously derived from conventional IM and subsequently determine the $\Omega$ of the ions of interest. [34]

**1.2.2 Collision Induced Unfolding**

With the use of an IM cell between mass analyzers, one can employ ion mobility in the study of protein folding intermediates and protein-ligand binding stability through collisional activation. When using the Synapt mass spectrometer, protein (or protein complex) ions are subjected to collisional activation through increasing collision energy in the trap cell prior to ion mobility and subsequent ToF separation. This technique is commonly referred to as collision induced unfolding (CIU). In CIU experiments (Figure 1.6), protein ions are subjected to increasing collisions with neutral gas molecules such as Argon or Nitrogen. During this collision process, the ion's kinetic energy is converted into internal energy [35,36]. In large proteins and protein complexes, these small heating events occur hundreds and even thousands of time allowing internal energy to accumulate. With enough internal energy built up, the 1$^{st}$ structural change that occurs is referred to as collision induced restructuring in which the overall structure arrangement becomes compromised. At this stage the structure is unlike the solution structure, but it is not yet unfolded or dissociated. With further increases in collision energy, intrasubunit interactions are compromised resulting in the unfolding of protein chains. [36] CIU experiments are monitored through observation of ion mobility spectra due to changes in the drift time as the experiment proceeds.

CIU can be used as a method of studying the stability of protein and protein-ligand complexes with the biggest drawback being the lack of absolute thermodynamic properties. Instead CIU can only provide relative or comparative data between proteins and their complexes.[37]
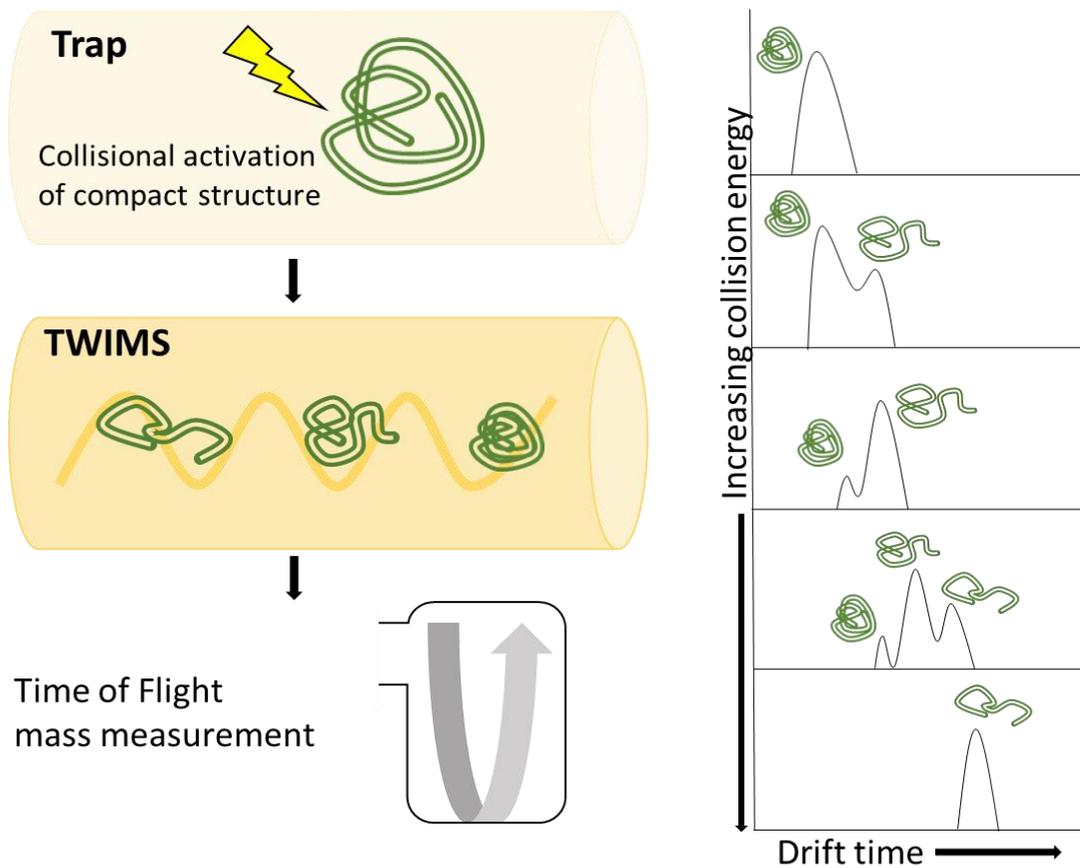
**Figure 1.6. Schematic representation of collision induced unfolding**. Collisional activation occurs in the trap region where the protein is subjected to increasing collision energies. With each increase in collision energy the drift time is measured in the ion mobility cell. The transition state between folded and unfolded states can then be determined and compared between a protein and protein/ligand complex.


## 1.3 Hydrogen Deuterium Exchange

Hydrogen Deuterium Exchange (HDX) allows for the study of solution phase protein conformations by monitoring the exchange of backbone amide hydrogens with deuterium.[38,39] The coupling of HDX to MS is beneficial as MS allows for the analysis of large and complex samples and eliminates the need for the incorporation of NMR active nuclei ($^{15}$N and $^{13}$C).[40] More recently,

the development of Time Resolved Electrospray Ionization (TRESI) HDX-MS has allowed for the structural analysis of intrinsically disordered proteins and proteins with intrinsically disordered regions which are difficult to accomplish using other structural techniques.[41,42]

The growth of HDX-MS can be attributed to its many different applications. One of the more recent applications is in the development of biopharmaceuticals, which comprises mostly monoclonal antibodies. HDX-MS has been used in the development of monoclonal antibodies, biosimilar antibodies, antibody drug conjugates and also in epitope mapping.[40,43–45] Another growing area of application for HDX-MS is the study of protein-membrane interactions. MS has the ability to differentiate between proteins and lipids and the areas of the protein that interacts with the membrane can be determined.[40,46] More common areas of application include, protein-ligand binding, homology-guided structural characterization of proteins, protein folding and protein-protein interactions.

There are three approaches to local HDX-MS experiments: top-down, middle-down and the most common approach: Bottom-up. Top-down analyses, first introduced by Anderegg *et al* 1994, involved deuteration of the intact protein followed by gas phase fragmentation through collision induced dissociation (CID).[47] However, CID results in a high degree of hydrogen/deuterium scrambling prior to fragmentation which can significantly alter the results.[48–50] 'Non-ergodic' fragmentation techniques such as electron capture dissociation (ECD) [51,52] and electron transfer dissociation (ETD)[53] result in less or no hydrogen scrambling and are now widely used for top-down HDX-MS. Bottom-up HDX involves deuteration of the intact protein followed by acid quenching and acid protease digestion prior to ionization (usually by electrospray). The deuterium level of each peptide is determined by monitoring the change in the mass.[54] There is no issue with hydrogen scrambling as there is no gas phase fragmentation involved, however, the spatial

resolution is low as peptides are typically 4-10 amino acids long. On the other hand, unlike top-down analysis there is no size barrier and large proteins and protein complexes can be studied. To overcome the size barrier of top-down HDX, a middle-down HDX workflow has been developed, which involves a combination of the top-down and bottom-up approach, in which the intact protein is labelled, quenched and partially digested followed by gas phase ETD/ECD fragmentation of the resulting peptides.[55]

### 1.3.1 Factors affecting intrinsic exchange

HX (hydrogen exchange) rates of backbone amide hydrogens vary substantially depending on the presence of hydrogen bonding, solvent properties (pH, temperature), primary sequence and most importantly protein structure and dynamics.[38,56–58] In order to quantitatively asses HX in backbone amides, the observed HX rate should be referenced to the rate of the unprotected amide to determine its protection factor PF = $k_{int}/k_{obs}$, where $k_{obs}$ is the experimentally determined rate constant and $k_{int}$ is the intrinsic rate constant for a given backbone amide hydrogen.[58–61]

The intrinsic exchange rate of backbone amide hydrogens is highly sensitive to pH. The mechanism by which the backbone amide exchanges hydrogen with deuterium is primarily catalyzed by acid ($H_3O^+$) or base ($OH^-$) and in some cases there could be a small contribution by water ($H_2O$) catalysis. In base catalysis, which dominates in the pH range above 2.5, the amide group is deprotonated by the $OH^-$ ion producing an amidate anion which in turn is reprotonated by $H_2O$ (or $D_2O$) to regenerate $OH^-$. Acid catalysis can occur via N- or O-protonation. In the case of N-protonation, the mechanism is the reverse of base catalysis. The amide N is protonated by $H_3O^+$ (or $D_3O^+$), after which $H_2O$ (or $D_2O$) extracts a $H^+$ from the protonated intermediate.[62–64] On the other hand, O-protonation involves the protonation of the peptide bond oxygen which is the most basic site of the peptide bond leading to acidification of the NH proton. The NH proton is then

removed by water producing an imidic acid which is later protonated by $D_3O^+$.[65] Equation 1.7 below, represents the pseudo first order chemical exchange rate constant for unprotected amides ($k_{ch}$):

$$k_{ch} = k_{int,\ A}[H^+] + k_{int,\ B}[OH^-] + k_{int,\ W}[H_2O] \qquad (1.7)$$

in terms of the intrinsic second order rate constants for the acid catalyzed (A), base catalyzed (B) and water catalyzed (W) reactions. The contribution by water catalysis is exceptionally small in comparison to the base and acid catalysis and it is often omitted from equation 1.7.[66]

A plot of amino-acid-averaged log ($k_{ch}$) versus pH results in a V-shaped curve (Figure 1.7) in which the minimum pH is 2.5-3.0. At the pH minimum, the rate of acid and base catalysis is equal and the rate of exchange is at its lowest.[57,67] Deuterium labelling is often carried out under 'native-like' conditions at $\sim$ pH 7, where the reaction is primarily base catalyzed ($k_{ch}=k_{int,OH}[OH^-]$), followed by quenching by lowering the pH to ~2.5, to provide sufficient time for Liquid Chromatography (LC) separation and MS analysis.[57]
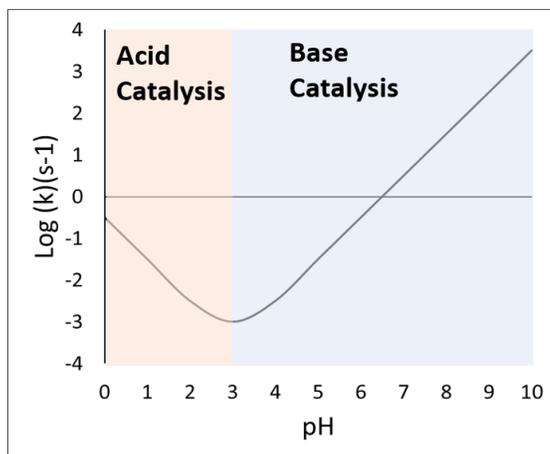


**Figure 1.7: The dependence of intrinsic rate of exchange (kch) on the solution pH in which the rate of exchange reaches a minimum at pH 2.5 - 3.** Acid and base catalysis is dominant below and above the pH minimum respectively. Reproduced from Brown and Wilson, 2017. [68]

For the base-catalyzed reaction, HX has an activation energy of approximately 17 kcal/mol, resulting in an increase in the HX rate by a factor of 3 for each 10 °C increment.[66,69] In order to predict HX rate constants at other temperatures, each rate constant in equation 1.7 should be modified according to equation 1.8. The theoretical HX rate constant ($k_{ch}$) as a function of temperature can be determined by:

$$k_{ch}(T) = k_{293}\exp\left(-\frac{E_a}{R}\left[\frac{1}{T} - \frac{1}{293}\right]\right)$$

(1.8)

In which $k_{293}$ is the rate constant $k_{int, H}$, $k_{int, OH}$, or $k_{int, H2O}$ at 293 K; $E_a$ is the activation energy corresponding to 14, 17 and 19 kcal mol$^{-1}$ for the acid, base and water-catalyzed exchange respectively; and R is the gas constant 0.001986 kcal mol$^{-1}$ K$^{-1}$. During the quench step in a bottom up HDX-MS experiment, along with decreasing the pH, the temperature is also typically decreased to 0 °C to minimize back-exchange.[64]

Another factor affecting amide HX rates is the primary sequence of the protein in which neighboring side chains can produce an inductive or steric effect. Inductive effects are caused by polar side chains, such as serine, which withdraw electron density from neighboring peptide groups resulting in an increase in their acidity. This increases the rate of deprotonation of the amide H by (OH$^-$ or OD$^-$) in the base catalyzed reaction whereas the overall rate of the acid-catalyzed reaction decreases. Therefore, at pH > pH$_{min}$, an amide H alongside serine residues will exchange 2.3 times faster than an amide H alongside a non-polar residue (*e.g.* Alanine).[58,61] Bulky side chains such as that of isoleucine decreases the amide HX rate by blocking solvent accessibility. Depending on pH this can decrease the rate of exchange by a factor of 10 to 20 if an amide is flanked by 2 isoleucines.[58] Both steric and inductive effects are additive and depend on the side chains of the nearest amino acid. In an unstructured peptide, the exchange rate constant of an amide H can be

calculated based on equation 1.9 in which $A_{left}$, $A_{right}$, and $B_{left}$, $B_{right}$ are side chain acid and base factors respectively.[58]

$$k_{ch} = k_{H^+}\left(A_{left} \cdot A_{right}\right)[H^+] + k_{OH^-}\left(B_{left} \cdot B_{right}\right)[OH^-] + k_{H2O}\left(B_{left} \cdot B_{right}\right) \quad (1.9)$$

### 1.3.2 Exchange Regimes

Although pH, temperature and primary sequence has a high impact on the HX rate in random coil polypeptides, in a folded protein the higher order structure has a greater impact on the HX rate. In order for a labile H to undergo exchange with $D_2O$, it needs to be exposed to the solvent (*i.e.* not buried in the core of the protein) and not be involved in any H bonding network.[38,70] The basic model for HX in folded proteins is depicted as:

$$NH_{cl} \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} NH_{op} \xrightarrow{k_{ch}} ND_{op} \underset{k_{op}}{\overset{k_{cl}}{\rightleftarrows}} ND_{cl} \quad (1.10)$$
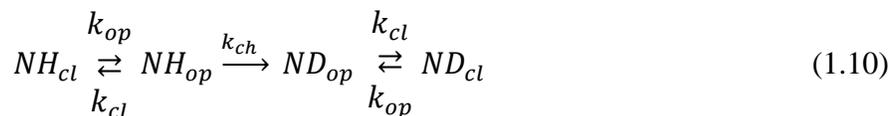
Due to the dynamic nature of proteins, buried or H-bonded backbone amide hydrogens do not always remain buried or H-bonded and can often come into an open state in which exchange can occur. In the mechanism depicted above, $k_{op}$ and $k_{cl}$ represents the rate constant of the opening and closing motions respectively. When in the open state, HDX occurs with the rate constant $k_{ch}$. As a result, the observed rate constant of exchange in folded proteins is a product of the rate constants of unfolding and refolding (i.e., 'opening' and 'closing') and the intrinsic rate constant.[70–72] The observed rate constant for exchange $k_{HX,}$ in a folded protein under native-like conditions can then be expressed as:[73]

$$k_{HX} = \frac{k_{op} \times k_{ch}}{k_{cl} + k_{ch}} \quad (1.11)$$

The conventional framework for analyzing HDX data is organized into two 'regimes'(EX1 and EX2) based on the relative magnitudes of $k_{ch}$ and $k_{cl}$. In the EX1 regime, during the opening event, all exposed labile hydrogens undergo exchange prior to the exchangeable sites returning to their

'closed' state (i.e., $k_{cl} \ll k_{ch}$) and the observed HX rate constant is equivalent to the rate constant of the opening reaction. Therefore equation 1.11 is simplified to:

$$k_{HX} = k_{op} \tag{1.12}$$

This type of exchange is not typical of proteins under physiological conditions (though it does occur) and can be induced through the addition of denaturants or a change in pH (usually an increase; pH > 9.0). Although this is not commonly representative of physiological protein dynamics, it can provide insights into protein folding.[72]

Although somewhat rare, cases of EX1 kinetics have been observed in proteins under physiological conditions.[74,75] One important benefit of coupling HDX with MS is the fact that EX1 kinetics can be easily identified in a mass spectrum through its characteristic bimodal distribution representing two populations (Figure 1.8B). The low mass population represent protein that has never undergone the EX1 conformational transition whereas the high mass population has undergone this conformational transition at least once. [72]

When the rate constant of the closing reaction is faster that the intrinsic rate of exchange ($k_{cl} \gg k_{ch}$) EX2 kinetics is observed, and equation (1.11) becomes: [70]

$$k_{HX} = \frac{k_{op}}{k_{cl}} \times k_{ch} \tag{1.13}$$

EX2 HDX results in a gradual increase in the mass over time in a single (binomial) distribution (Figure 1.8A). As most proteins are structurally stable at physiological pH, transient unfolding events in structured regions occur faster than $k_{ch}$, so EX2 is the most common exchange regime of a protein under native conditions.[72,74] EX2 is also the more informative of the two dominant exchange regimes, as it allows for the determination of the 'open'/'closed' equilibrium, which, when measured specific sites or segments, can be viewed as a local measurement of local thermodynamic structural stability.[76] It is important to be able to distinguish between EX1 and

EX2 because there are some cases in which EX1 is observed under native conditions, and this can confound automated data analysis software. Experimentally, HDX-MS provides the advantage of easily differentiating between EX1 and EX2, which is more difficult to do with NMR experiments.[75]
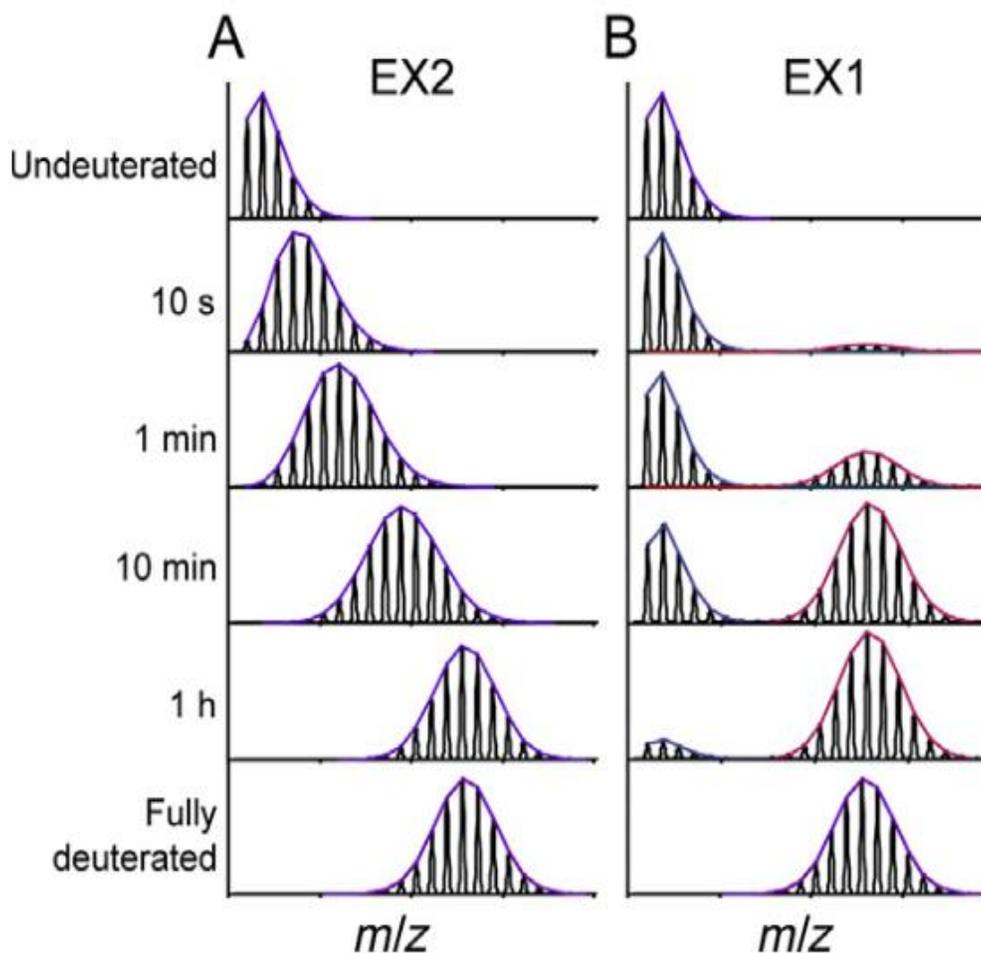


**Figure 1.8 The two major exchange regimes observed in hydrogen deuterium exchange measurements.** (A) EX2 regime in which peptides exchange deuterium gradually through fast local unfolding. (B) EX1 regime represents proteins that undergo much slower exchange due to large segments of the protein becoming exposed for longer periods of time. Adapted with permission from Gutman and Lee, 2016. [77]

EXX kinetics refer to an intermediate state where $k_{cl}$ is neither faster nor slower than the $k_{ch}$ ($k_{cl} \approx k_{ch}$). This phenomenon has been explored by Xiao *et al.*, in which they observed a bimodal distribution similar to that in the EX1 regime however each population drifted as observed in EX2 kinetics.[78]

### 1.3.3 Global HDX

A typical workflow for global HDX involves the protein of interest incubated in an excess of $D_2O$ for a period of time, after which the reaction is quenched by decreasing the pH to 2.5 and the temperature to $0^oC$ in order to minimize the exchange rate in what is (from quenching onwards) mainly protic solvent.[54] Separation is not typically required for global analysis so a trap column is often used in the LC step to desalt the protein sample prior to MS analysis. The LC step also eliminates deuterium that was incorporated into side chains (all ultra-fast exchangers) through complete back-exchange. As a result, by the time the sample is subjected to ESI, deuterium will have equilibrated with solvent for all but the backbone amides, which significantly simplifies data analysis and interpretation.[54,79]

As there is no protease digestion or gas phase fragmentation involved in global HDX analysis, it provides an overall picture of the conformational dynamics of a protein of interest. To determine the amount of deuterium incorporated in the protein, the mass of the non-deuterated protein is subtracted from the mass of the deuterated protein at each time point being studied. This can be done for individual charge states of the protein or by using deconvoluted masses. The data are then plotted as a change in mass or relative deuterium level (y axis) vs. time (x-axis).[79] The information that is gained from global HDX is limited and so local HDX is more common. However, global HDX can be useful to determine if the protein of interest is stable under HDX-MS conditions and serve as an initial guide as to how dynamic the protein is. The relative extent of disordered regions

in a protein can be estimated from the number of hydrogens that exchanged at early time points (10 s or less). Likewise, the number of hydrogens involved in more rigid or buried residues can be determined by subtracting the maximum observed exchange (after several hours of labelling) from the total number of exchangeable backbone hydrogens.[80] To ensure the accuracy of these estimates, an internal calibrant (such as an unstructured peptide) should be used to account for backbone amide back-exchange.
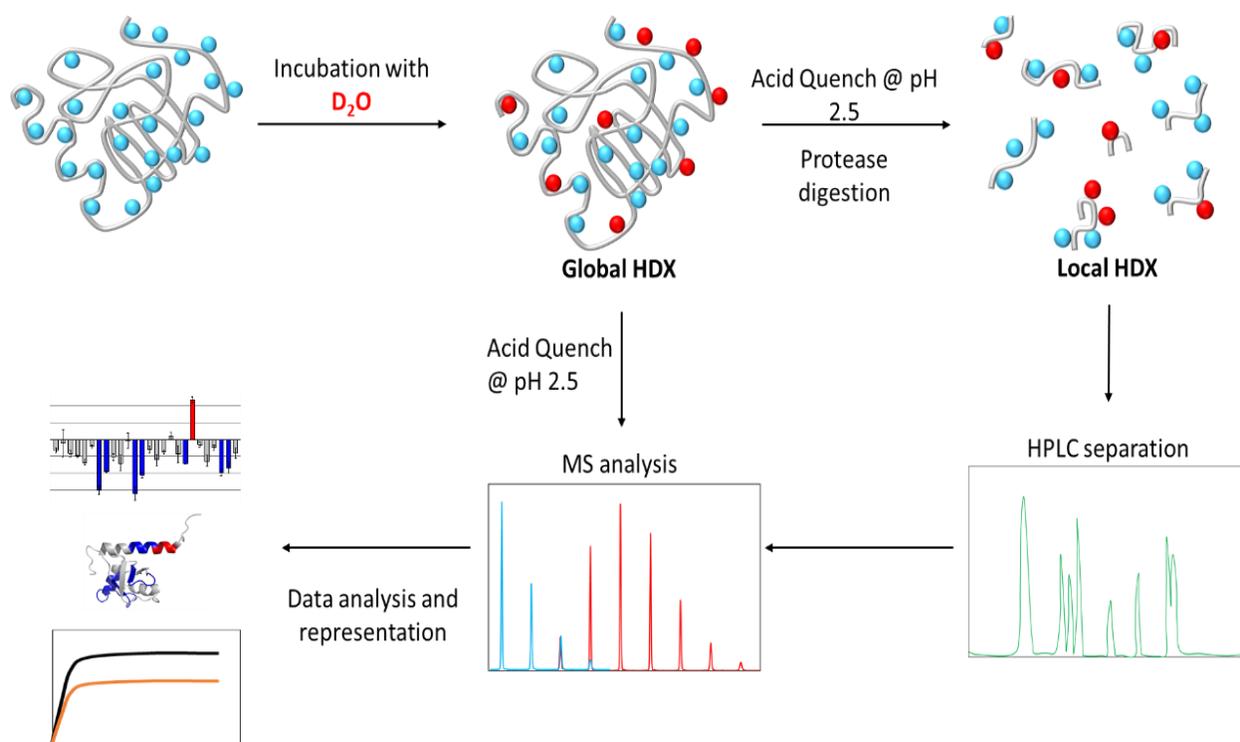


**Figure 1.9 Overview of Hydrogen Deuterium Exchange Mass Spectrometry Workflow**. Protein is first incubated with $D_2O$ for increasing periods of time ranging from seconds to hours and then quenched by lowering the pH to 2.5. In local HDX the deuterated protein is digested with an acid resistant protease followed by LC separation and subsequent MS analysis. In Global HDX, the deuterated protein is subjected to MS analysis. After MS analysis the data is analyzed and appropriately represented.

### 1.3.4 Local HDX

The experimental workflow for local HDX is essentially identical to global HDX with one additional step to digest the deuterated protein sample (Figure 1.9). The term "local" refers to the type of information one gets from the experiment, in which it is possible to measure the conformational or dynamic changes in localized regions of the protein, allowing for instance, for the determination of ligand binding sites.[81,44] There are two main types of local HDX workflows: bottom up and top down. Bottom-up local HDX is the more common of the two, which involves digestion of the deuterated sample with an acid resistant protease prior to ESI. This approach is in some ways more straightforward than top-down HDX and does not require specialized equipment but provides peptide-level resolution (typically 4 – 10 residues). The top down workflow involves gas phase fragmentation of the deuterated protein. Although top-down analysis can provide amino acid level resolution, there is an analyte size barrier and specialized equipment is required to prevent deuterium scrambling in the gas phase.[44,81,82]

In the bottom-up workflow, because the reaction is quenched with acid, an acid resistant protease must be used. Pepsin is the most commonly used protease in HDX experiments, [83] but others can be employed as well such as: aspergillopepsin (protease type XIII),[84] rhizopuspepsin (protease type XVIII) [84] and plasmepsin.[85] Non-specific proteases often result in multiple overlapping peptides and with careful analysis information on single residue amide hydrogens can theoretically be obtained but this should be done with caution as this could lead to over-interpretation of the results. A major disadvantage of protease digestion is incomplete sequence coverage or peptides that are too large in some cases where the protein does not get digested very well. This can sometimes be overcome with the use of a mix of different proteases or by adding a denaturant in

the quench buffer to improve the digestion.[84] After the protein is digested, the peptides are separated using reversed-phase LC prior to MS analysis.

With the addition of lengthy digestion and LC separation steps, back exchange may occur, but can usually be accounted for in data analysis. [54] To account for back exchange, a completely deuterated sample ($M_{100\%}$) is used as a control and the total deuterium uptake level in each peptide, D, is calculated at each time point using the equation:

$$D = \frac{M_t - M_{UND}}{M_{100\%} - M_{UND}} \times N \qquad (1.14)$$

Where $M_t$ and $M_{UND}$ is the mass of the deuterated peptide and mass of the un-deuterated peptide respectively and N is the maximum number of deuterium atoms that can be incorporated.[54] There has been debate over the years as to whether a back-exchange correction is useful as most HDX experiments measure relative rather than absolute deuterium uptake and studies have shown that back exchange correction does not affect the final results in relative measurement experiments (*i.e.*, where one state is being compared to another state of the same protein).[81,82] In addition, back exchange correction is more of an estimate, as it relies heavily on preparing a fully deuterated form of the sample of interest, which can be difficult to prepare and maintain.[82]

The 1st step in data analysis is the determination of the peptide sequences based on their *m/z* and LC retention time. Accurate mass is often not sufficient to determine the peptide sequences but it can be used to determine the possibilities.[86] To ensure accuracy in peptide peak assignments, tandem MS can be used to conduct classical MS/MS peptide sequencing.[87] This step is crucial as miss-identifying the peptides can lead to fundamental misinterpretation of the results. There are two main methods for peptide sequencing by LC MS/MS: Data Dependent Acquisition (DDA) and Data Independent Acquisition (DIA). DDA performs the MS/MS analysis on precursor m/z

selected from the MS survey scans. There is a selection criteria and specific m/z values can be included or excluded from the MS/MS analysis.[88] Conversely, DIA acquires the MS/MS scans using a wide isolation window with no specific precursor ion selected. [89]

After the peptide list is generated, deuterium incorporation can be quantified. There are two main ways in which this is done: through the change in the centroid mass or through the change in the isotopic distribution of the peptide. Through the former method, the centroid mass of the non-deuterated peptide ($M_{UND}$) and deuterated peptide ($M_t$) is calculated and the relative deuterium level ($D_t$) of each peptide is determined by the equation:

$$D_t = M_t - M_{UND} \tag{1.15}$$

This method is not a direct measure of uptake, but it can be sensitive enough in detecting changes in conformation. However, this method is not ideal for detecting subtle changes in deuterated peptide isotopic distributions associated with small levels of uptake and it may miss exchanges in the EX1 or EXX regime.[90,91]

Another method used to quantify deuterium uptake levels is to model and then fit the isotopic distribution of the deuterated peptide, which is often more accurate than using the centroid. This is particularly true at low deuteration levels or where there is overlap for some peaks in the distribution. In a non-deuterated peptide, the isotopic distribution observed is mostly as a result of $^{12}C$ and $^{13}C$. However, as deuterium is incorporated the distribution becomes a convolution of (mainly) the $^{12}C/^{13}C$ distribution and $^{1}H/^{2}H$ distribution.[86] The isotopic distribution after labelling can be modelled as a convolution of the natural isotope distribution given by the equation:

$$D = N * B \tag{1.16}$$

Where D is the isotope distribution after deuterium labeling, N is the natural isotope distribution and B is the binomial distribution. This binomial distribution is represented by:

$$B(p, n, k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad (1.17)$$

In which n is the total number of exchangeable hydrogens in the peptide of interest and k is the number of hydrogens that have undergone exchange (0,1,....n). The segment averaged level of deuteration (0 -100%) is represented by p. The actual deuteration level of each peptide can then be determined by fitting a simulated distribution B, to the experimental data.[92] It is important to emphasize that the best-fit value of p represents the average uptake for the peptide and does not provide information on the distribution of uptakes at individual sites on the peptide. For instance, in an 8 amino acid peptide, a highly skewed distribution (*e.g.*, 4 sites exchanging at 100% and 4 sites at 0%) would produce the same value for p as a flat distribution (*e.g.*, 8 sites exchanging at 50%), both of which are correct (*i.e.,* in both cases, a perfect fit to the data would yield p = 50% and the peptide is in fact 50% exchanged). There are several methods to fit theoretical distributions to the data. The most common and straightforward of these is the least-squares approach which minimizes the square of the difference between the normalized peak intensities of the theoretical and observed distributions. [90,93]
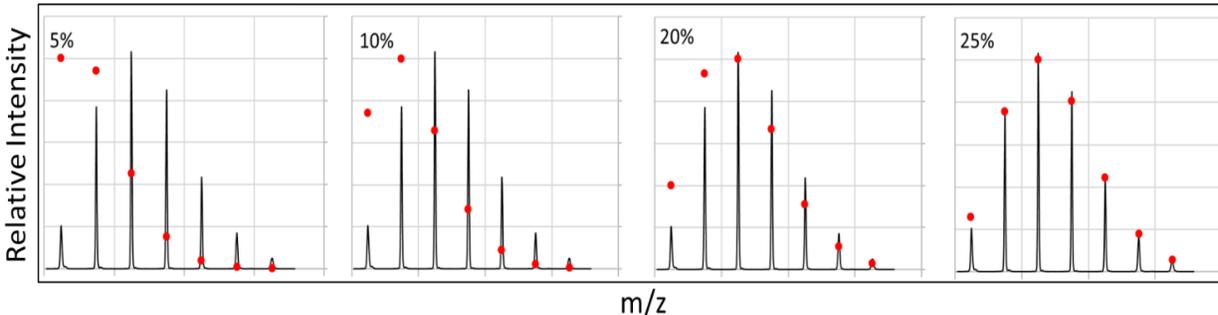
**Figure 1.10. Representation of how the theoretical distribution D (red dots) is fitted to the spectrum of a deuterated peptide to determine the percent deuterium incorporation.** Reproduced from Brown and Wilson, 2017. [68]

A major hurdle in the quantitation of deuterium levels, is the presence of over lapping peptides. Although the LC step after digestion separates the peptides, overlap still occurs often in samples from large proteins. One way in which this problem could be overcome is using a 2nd dimension of separation, such as ion mobility, that separates species based on their charge and conformation. This could essentially be used to separate overlapping peptides with the same LC retention time.[94] It is also noted that the 'distribution fitting' method described above requires only three overlap-free peaks within a distribution to accurately determine the uptake level. In reporting the relative deuterium level, it is also important to provide an error in these measurements. Errors are commonly estimated through standard deviations which are usually obtained from at least three technical replicates of the same labelling time point. It is equally important to use biological replicates where possible to test the reproducibility of the results.[86,95]

### 1.3.5 Time Resolved Electrospray Ionization Hydrogen Deuterium Exchange

HDX reveals a wealth of information on protein dynamics, protein/ligand and protein/protein interactions. One of the biggest drawbacks on conventional HDX however, is the lengthy reaction times. The most common reaction times range from minutes to hours although 20 second mixing

times can be achieved but it is often not accurate. Shorter reaction times, though not always necessary can reveal crucial information that could be missed in a conventional HDX set up.[96–98] For example, transient interactions are short lived and are often missed in conventional HDX because of the long reaction times. In addition, intrinsically disordered proteins[99] and proteins with intrinsically disordered domains[100] are not easily studied with conventional HDX as all the backbone amides exchange rapidly and will be fully exchanged within minutes.

Time-Resolved Electrospray Ionization Hydrogen Deuterium Exchange (TRESI-HDX) also referred to as millisecond HDX allows for shorter reaction times in the millisecond to second time scale. In addition, it offers a much simpler workflow with no LC separation and as a result very little back exchange is observed. The heart of millisecond HDX is the time resolved capillary mixer which was developed by Wilson and Konermann.[101] The capillary mixing device consist of two concentric capillaries. The inner glass capillary is placed inside a metal capillary with the end of the glass capillary sealed, and a notch cut 2 mm from the sealed end to allow solution to escape and facilitate proper mixing. A three-way mixing device is used to connect the capillary mixer, an acid line and the capillary leading to a microfluidic chip that houses a digestion chamber. The output channel of the microfluidic device also served as the electrospray ionization probe.[96] The capillary mixer can be used without the microfluidic chip for global HDX experiments in which digestion is not required. [102]
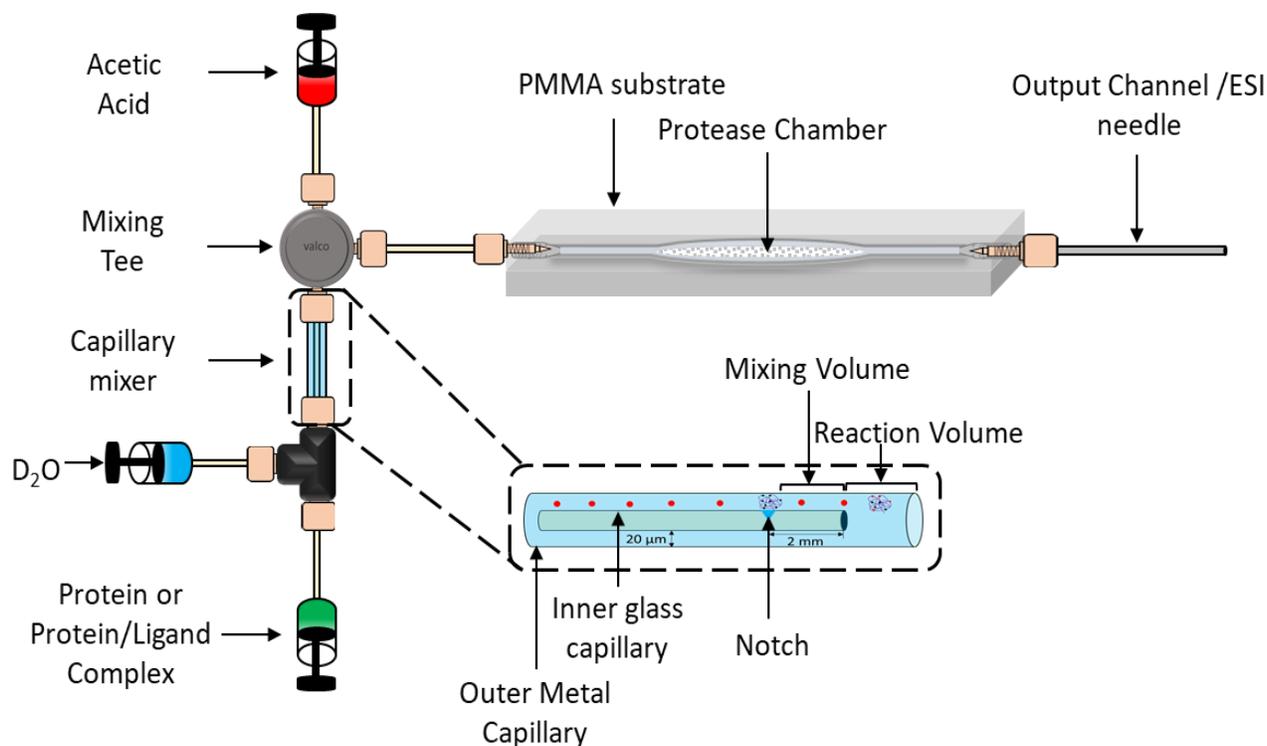
**Figure 1.11 Schematic representation of TRESI-HDX**. HDX reactions occur in the capillary mixer in which reaction volume and hence reaction time is increased by pulling back the inner capillary. The capillary mixer and acid channel are connected by a valco mixing tee in which the reaction is quenched before entering the protease chamber where digestion occurs. The output channel of the microfluidic device serves as the ESI needle.

### 1.3.6 Binding vs Allosteric Effects

HDX-MS is a fast-growing technique that continues to evolve with new applications, but the core process remains the same. HDX can be used to determine exchange kinetics of peptides and intact proteins and the effect ligand binding has on the kinetics.[72,74,103] Although most applications of HDX-MS involve the determination of binding sites, it can be used to determine allosteric changes under varying conditions although it remains difficult to distinguish binding sites from allosteric sites if the binding sites are not previously known.[104,105] If ligand binding results in an increase in deuterium uptake in a certain region of the protein, this change can be often (but not always) be attributed to an allosteric conformational change. However, if ligand binding results in decrease

deuterium uptake in an allosteric region, it becomes difficult to differentiate between the binding interactions and allosteric conformational changes.[105,106] If a crystal structure of the protein being studied is available that shows the protein in complex with its ligand, then HDX can be used to monitor conformational changes upon ligand binding to determine the allosteric hotspots.[107] Work done by Chandramohan *et al.* 2016, used HDX-MS to distinguish allosteric effects from ligand binding in Hsp90. The ligand interactions are known based on published high resolution structures of Hsp90 with 2 high affinity ligands (PDB ID:4EGK and 1YET). HDX-MS revealed significant decreases in deuterium uptake in the ligand binding regions. In addition to testing high affinity ligands, low affinity ligands (fragment 1 and 2) were tested and found to have more significant allosteric responses to ligand binding. The same 4 ligand binding sites were observed when using the high or low affinity ligands, but the high affinity ligands had a much more significant decrease in deuterium uptake. The fragments are smaller and are expected to have fewer interactions in the binding pocket.[105]

Both fragments had similar changes in the binding pocket but had significantly different allosteric responses. This was unexpected as both fragments have similar molecular weight and affinities and they belong to the same phenolic class. It was proposed that changes due to protection from HX as a result of H-bonding would show up in earlier reaction times whereas changes due to allosteric conformational changes would appear at longer times. Additional experiments would be needed in order to thoroughly test this hypothesis. Nonetheless, when used together, HDX-MS and X-ray crystallography can be used to distinguish between binding interactions and long-range allosteric changes.

## 1.4 RNA Chaperones

RNAs are among the most functionally versatile biological macromolecules within the cell. The roles of RNA include coding, transfer of genetic material, regulation of various cellular processes and catalysis of chemical reactions (such as cleavage and ligation). [108] In addition RNAs are also structurally versatile due to their four different bases that easily form stable helices that are not necessarily the RNA's native structure. Non-coding RNAs that control gene expression fold into a variety of different structures which often have the same stability as the native structure resulting in competition for the native fold. These alternate structures result in traps along the folding pathway and transient structures can become long-lived intermediates if the molecule is not able to escape.[108,109]
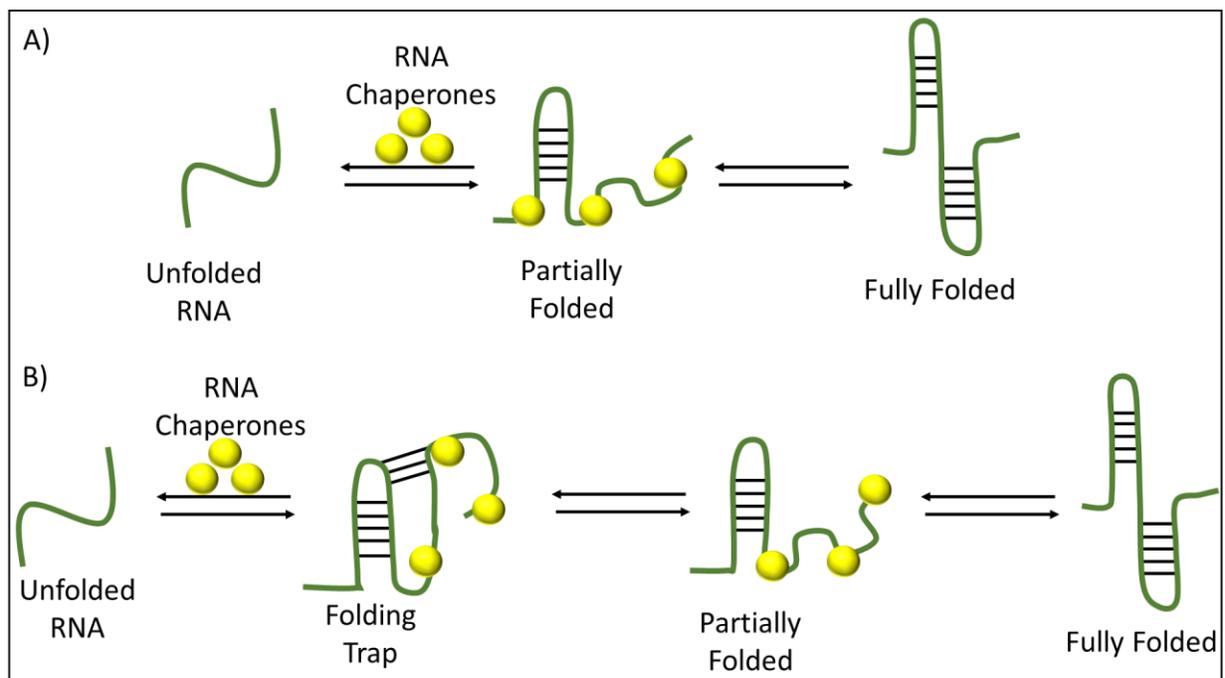


**Figure 1.12 Schematic representation of RNA chaperone activity**. A.) RNA chaperones are used to aid in the correct folding of RNA molecules and B.) to correct the mis-folding of RNA that fell into a folding trap.

Within the cell, RNA's are often found in complex with proteins that aid in the correct processing or folding of RNA molecules. In some cases, however, the RNA helps the protein to fold correctly. RNA chaperones are defined as proteins that prevent misfolding of RNAs and correct misfolded structures. RNA chaperones should not be confused with RNA helicases that perform the same function but require an external energy source. Most RNA chaperones are highly basic proteins and so they interact easily with the negatively charged RNA molecules. The interactions are often transient and the protein releases the folded RNA for it to carry on with its normal functions. [108,109] Structural disorder is common in RNA/protein recognition as the interactions typically involve conformational changes in structure. One of the most common example of this phenomenon is the La protein. [110]

Many RNA chaperones are key players in RNA translation and more importantly viral RNA translation. When a virus infects a cell, it essentially hijacks the cell machinery which includes the cellular RNA chaperones. This aids in the translation of viral proteins that are crucial for the life cycle of the virus.[111] The development of effective antiviral agents has proven difficult due to the rapid mutational rate of viral RNA which leads to drug resistance. A different approach to this problem could be to target an RNA chaperone that is crucial in the translation of the viral RNA. If this chaperone is rendered inactive, it could in turn inactivate the virus by preventing the translational of the RNA.[112]

Cellular capped mRNAs are translated using a cap-dependent mechanism. In the cap-dependent translation, the 40S ribosome binds to the 5' terminal cap structure on cellular mRNAs and scans the mRNA until it approaches the appropriate AUG codon.[113,114] However, uncapped viral RNAs are translated via a cap-independent mechanism in which all the canonical initiation factors required for cap-dependent translation are not needed. This cap-independent translation involves

internal entry of the ribosome within the 5' untranslated region (UTR) of the viral RNA and is referred to as internal ribosome entry site (IRES) mediated translation. Hepatitis C virus (HCV), Poliovirus (PV), foot-and-mouth disease (FMDV), Hepatitis A virus (HAV), classical swine fever virus are some examples of viruses with functional IRES elements. [114]

La protein was one of the 1st RNA chaperones identified to interact with IRES-elements, and it was found to promote IRES-mediated translation of PV and HCV virus. Although La is predominantly nuclear and aids in the processing of nascent RNA polymerase III transcripts, in times of cellular stress (such as viral infection) it relocates to the cytoplasm.[114] In addition, studies have shown that La binds and protects HBV from destruction by covering the RNA cleavage site. Recently anti-viral agents have been developed to target La which in turns inhibit HBV. [115] La also plays a role in cap-dependent translation, although the mechanism in which it does so has not been fully explored. La has however been shown to play a role in the translation of cellular mRNA through binding to the poly(A) tail.[116]
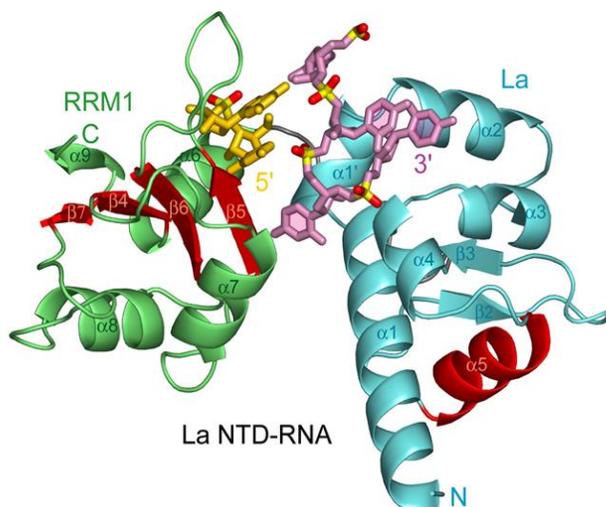


**Figure 1.13. The structure of human La-NTD in complex with a poly(U) RNA ligand**. The La motif and RRM1 forms a binding pocket in which the RNA ligand binds in between the two domains. The typical nucleic acid binding sites are highlighted in red. And the RNA ligand is highlighted in magenta. Reproduced with permission from Teplova *et al*., 2006. [117]

La plays such an important role in many aspects of RNA processing but there is a lack of structural data in the interaction of La and its various non-poly(U) RNA targets. Figure 1.13 shows the NTD of human La in complex with a poly(U) RNA ligand. Interestingly, this structure revealed that the Poly(U) RNA ligand does not come into contact with the typical nucleic acid binding sites highlighted in red.[117] Although mass spectrometry is a low-resolution structural technique, it still remains a powerful tool that can be used to shed some light on the La-RNA binding mechanisms.

## 1.5 Biological Drugs

Biological drugs are macromolecules (mostly proteins) made from living organisms. This class of drugs has been on the rise thanks to developments in molecular biology, genetics, protein engineering and cell sciences. [118] Biologics are substantially larger than traditional pharmaceutical drugs and they have a more complex manufacturing process. The majority of biological drugs are produced from three sources: *Escherichia coli*, yeast or Chinese hamster ovary cells. Insulin is one of the most common commercially available biological drugs that is used in the treatment of diabetes. Other examples of biological drugs include: human growth hormone, erythropoietin and monoclonal antibodies (mAbs).[119] mAbs and other antibody related therapeutics (such as antibody drug conjugates) are the fastest growing and the most promising biological drugs. This is due to the fact that antibodies bind to their target with high specificity efficacy and with fewer side effects than small molecule drugs. [2,120]

### 1.5.1 Therapeutic Antibodies.

Therapeutic mAbs are used to treat a variety of illnesses such as Crohn's disease, rheumatoid arthritis, Multiple sclerosis, asthma and various types of cancer. The first mAbs were murine (mouse origin) and when injected into patients produced an immune response. In addition, their efficacy was restricted as these murine antibodies did not interact correctly with the essential

components of the human immune system. Developments in molecular biology led to cloning of antibody genes which generated chimeric, humanized and human antibodies each of which consisted of fewer and fewer murine DNA (Figure 1.14).[121]
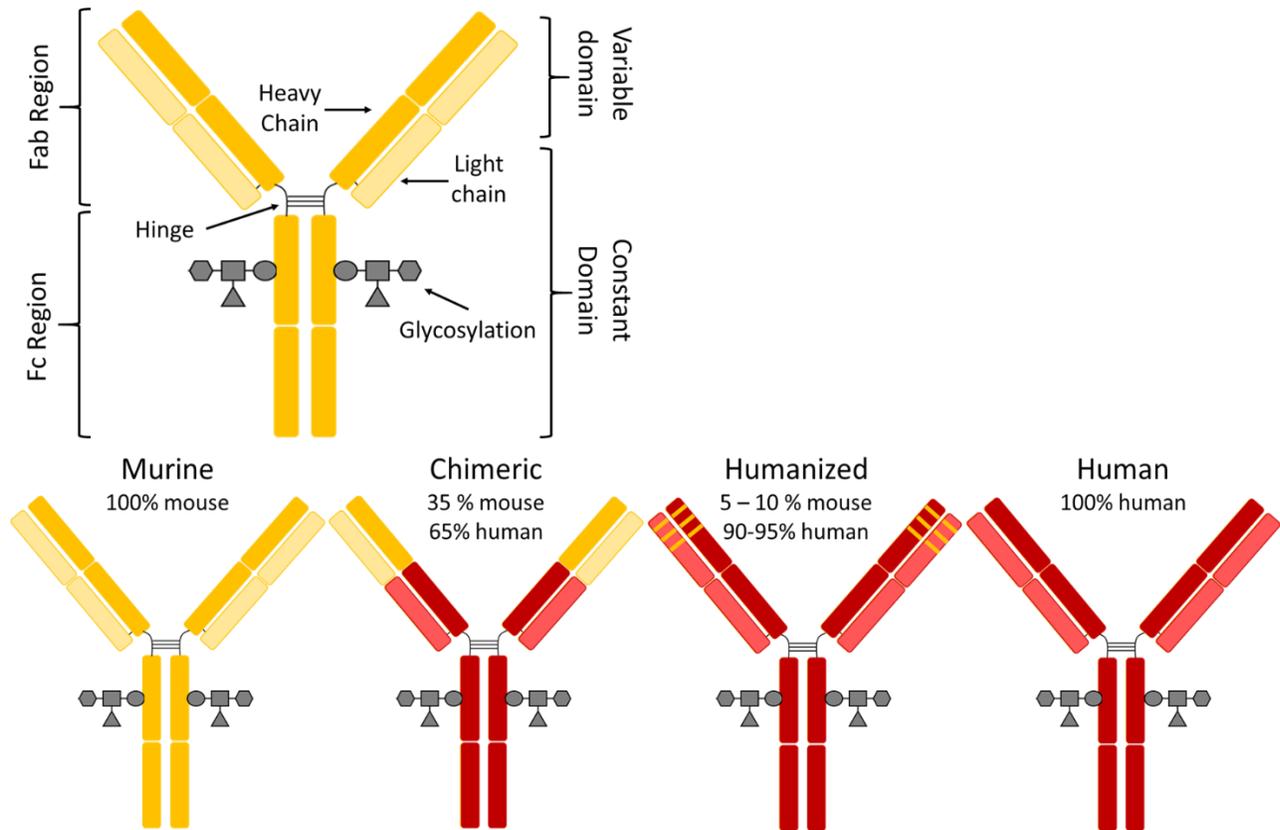


**Figure 1.14 Domain organization of recombinant IgG1 antibodies**. IgG1 antibodies consist of 2 heavy chains and 2 light chains connected by disulfide bonds in a "Y" shape. The Fab region is the antigen binding domain while the Fc region is the tail region that interacts with the cell surface receptors. Chimeric and humanized antibody is a combination of mouse and human DNA with chimeric being 65% human and humanized being 90-95% human.

Although therapeutic mAbs are extremely effective and growing more popular they have a few limitations, two of which are their costs to patients and lengthy and expensive manufacturing processes. mAbs are large proteins (150 kDa) consisting of multiple domains, post translational modifications and several disulfide bonds. The production of these molecules requires large

cultures of mammalian cells and extensive purification under good manufacturing practice (GMP) conditions. This along with the large doses required for clinical efficacy leads to high cost to patients.[118,121,122] Biosimilar development and production is on the rise and can reduce the cost to patients by providing another alternative.[123–125]

### 1.5.2 Biosimilar Drugs

Biosimilar drugs are biologically and clinically comparable to an innovator drug. Unlike generic drugs of small molecule pharmaceuticals, it is impossible to produce a drug that is 100% identical to a biological drug. This is due to the complex manufacturing process that is not readily available outside of the innovator company. In addition, glycosylation patterns in antibodies are extremely important as they affect the protein's three-dimensional structure, the overall stability and they might also have a role in cell to cell adhesions.[119] Once the patent for a particular biologic has expired, usually 20 years from the time of filing a submission, the development of biosimilars will be open to other companies. A key difference in the development of biologics vs biosimilars is that there is a larger emphasis placed on the preclinical development when it comes to biosimilars. This includes analytical, structural, in vitro functional and toxicology analysis. This large emphasis on the pre-clinical stage is in proving the similarity between the biosimilar and the innovator drug.[126]

### 1.6 Research Objectives

Structural mass spectrometry is a powerful tool for the characterization of proteins and protein ligand complexes, and it is becoming increasingly popular in structural biology and pharmaceutical sciences. RNA chaperones are of key interest as they aid in the folding and maturation of both cellular and viral RNAs. In addition, many RNA chaperones have a large unstructured region that is thought to be important for chaperone activity. An important example of such RNA chaperone

is the human La (hLa) protein that is essential in the translation of both cellular and viral RNAs (e.g. HCV, HBV and polioviral RNAs). Understanding how hLa engages viral RNA and cellular mRNA could potentially allow for the development of antiviral drugs that prevent the translation of viral RNA without comprising the translation of cellular mRNAs.

Chapter 2 focuses on the largely unstructured CTD of hLa and the mechanism in which it binds the stem loop domain IV of the HCV RNA in comparison to a 23-mer unstructured RNA.

La proteins are pre-dominantly nuclear but become cytoplasmic in times of cellular stress. Chapter 3 focuses on how alterations in the nuclear trafficking signals of the full length hLa affects RNA binding. Two RNA ligands are studied, a U10 and A20 RNA ligand.

Another emerging area of study in structural MS is in epitope mapping and biosimilar characterization. Chapter 4 focuses on using a variety of structural MS techniques in the structural characterization of a biosimilar mAb (bevacizumab) and its innovator drug (Avastin). Batch-Batch variation in the manufacturing process of the biosimilar drug is also examined to determine the reproducibility.

Chapter 5 focuses on epitope mapping HDX of Vascular endothelial growth factor (VEGF) when bound to bevacizumab and Avastin to compare epitope binding sites of both antibodies. This is done to confirm if both drugs bind VEGF in the same way.

# Chapter 2: Distinct Dynamic Modes Enable the Engagement of Dissimilar Ligands in a Promiscuous Atypical RNA Recognition Motif

**Brown, K.A**., Sharifi, S., Hussain, R., Donaldson, L., Bayfield, M.A. and Wilson, D.J.

Experiments were planned by **KAB**, SS, RH, LD, MAB and DJW. All mass spectrometry experiments, and data analyses were done by **KAB**. The manuscript was prepared by **KAB** with editorial contributions from MAB and DJW.

## 2.1 Summary

Conformational dynamics play a critical role in ligand binding, often conferring divergent activities and specificities even in species with highly similar ground-state structures. Obtaining a detailed picture of ligand binding-driven changes in conformational dynamics can be exceptionally challenging, particularly for transient, relatively weak interactions common in RNA chaperones. Here, we introduce Time-Resolved ElectroSpray Ionization Hydrogen-Deuterium eXchange (TRESI-HDX) as a powerful tool for characterizing the changes in dynamics that accompany transient oligonucleotide binding, using the atypical RNA recognition motif (RRM2) in human La (hLa) as a model. Using this approach, we reveal that RRM2 binding to a model 23-mer unstructured RNA and to the IRES domain IV of the hepatitis C viral (HCV) RNA are driven by different dynamic processes. In particular, binding of the ssRNA induces helical 'unwinding' in a region of the CTD previously hypothesized to play an important role in La and La-related protein associated RNA remodeling, while the same region becomes less dynamic upon engagement with the double-stranded HCV RNA. Our results demonstrate the ability of TRESI-HDX to provide insights into the dynamic underpinnings of RRM substrate binding, specificity and function.

## 2.2 Introduction

Obtaining a detailed picture of ligand-induced structure and dynamic changes that occur in RNA binding proteins can be challenging, especially in the case of the relatively weak, transient interactions commonly associated with RNA chaperones. [127] Classical structural techniques like structural Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography can in many cases provided highly detailed pictures of ground-state structures, but are generally ill-suited to the characterization of highly dynamic species or interactions. [128] Another common method of

structural analysis is hydrogen deuterium exchange (HDX) which relies on structure-dependent exchange between protein backbone amide hydrogens and solvent hydrogens (or deuterium).

When a protein is introduced into deuterated solvent, backbone amide hydrogens will exchange at a rate that depends principally on (**i**) their involvement in hydrogen bonding (*e.g.*, in secondary structures) and (**ii**) their degree of solvent exposure. To acquire local (spatially-resolved) deuterium uptake measurements by mass spectrometry, the HDX reaction is quenched by lowering the pH to 2.5, followed by digestion of the labelled protein by an acid protease. [129,130] Analysis of the resultant peptides provides information as to which parts of the protein were more 'dynamic'. A detailed review of this method is provided in references.[131,132] The 'time-resolved' aspect of our approach signifies millisecond-to-second $D_2O$ exposure times, accomplished through the use of a capillary mixer, which uniquely allows us to characterize structural changes in weakly-ordered regions of proteins and transient interactions. [133,134] The experiment is implemented on an electrospray mass spectrometry-coupled microfluidic chip as described in Chapter 1.3.5.

Human La (hLa) is a multifunctional phosphoprotein predominantly located in the nucleus of eukaryotic cells, [135,136] and was first discovered as an autoantigen in patients suffering from autoimmune diseases such as systemic lupus erythematosus, neonatal lupus and Sjogren's syndrome. [137,138] Genuine La proteins are present across several eukaryotic genomes including yeast, ciliates, flies, frogs and mammals.[139] The primary function of hLa is binding to nascent RNA polymerase III transcripts at the UUU-OH 3' end, protecting them from exonuclease digestion and ensuring correct folding and processing of these RNA molecules. [140–142] In addition, binding of hLa to small RNAs allows for nuclear retention and aids in the assembly of functional ribonucleoprotein (RNP) complexes. [143,144] The La protein is also known to bind not only to RNA

pol III transcripts but also to viral RNAs and mRNA transcripts and it is thought to play a role in their translation. [145–151]

The N-terminal domain (NTD) is the most conserved region of hLa whereas the C-terminal domain (CTD) increases in length and complexity in higher eukaryotes. [152] The NTD consists of a La motif (LaM) and an RNA recognition motif (RRM1) that acts in synergy as a single binding unit to bind the UUU-OH 3' region of RNA targets with high affinity and specificity. [153,154] The CTD of hLa consists of an unusual RRM (RRM2), a nuclear retention element (NRE), a highly charged, unstructured region known as the short basic motif (SBM), a casein kinase II phosphorylation site (S366) and a nuclear localization signal (NLS).[152,155–159] Although the RRM2 is not required for binding of hLa to poly (U) RNA, it has been suggested to play a role in recognition to non-poly(U) sequences, although the mechanism of RNA binding remains unclear. [157,160,161] The CTD of hLa has also been implicated in hLa dimerization; specifically, a region corresponding to roughly the C-terminal half of RRM2, as well as the NRE and SBM, was hypothesized to mediate hLa self-dimerization and associated poliovirus RNA translation,[162] although subsequent analysis was unable to find evidence for hLa dimers. [152] More recently, the CTD of hLa has been shown to be active in assays for RNA chaperone function,[163,164] consistent with other work highlighting the La and La-related proteins as RNA chaperones. [141,165–168]

Recently, a model for RNA binding to RRM2 of hLa was proposed based on the RNA binding of a similar RRM in the La-related protein p65 in *Tetrahymena thermophilia.*[169,170] p65 belongs to the LARP7 family and functions in the correct folding of telomerase RNA and the hierarchal assembly of the telomerase RNP. [171] Notably, the RRM2 of p65 and hLa share a similar structure which is different from the typical RRM (Figure 2.1). [152,169] The RRM is an abundant domain present in over 50% of RNA binding proteins (RBP), (reviewed in [172], and a typical RRM has a 4

stranded β-sheet and 2 α-helices arranged as β1α1β2β3α2β4. The RRM2s of hLa and p65 are atypical in that they have a β1α1β2β3α2β4`β4α3 fold, in which the α3 helix (corresponding to amino acids mapping to the NRE in hLa) lies across the canonical RRM-RNA binding site.
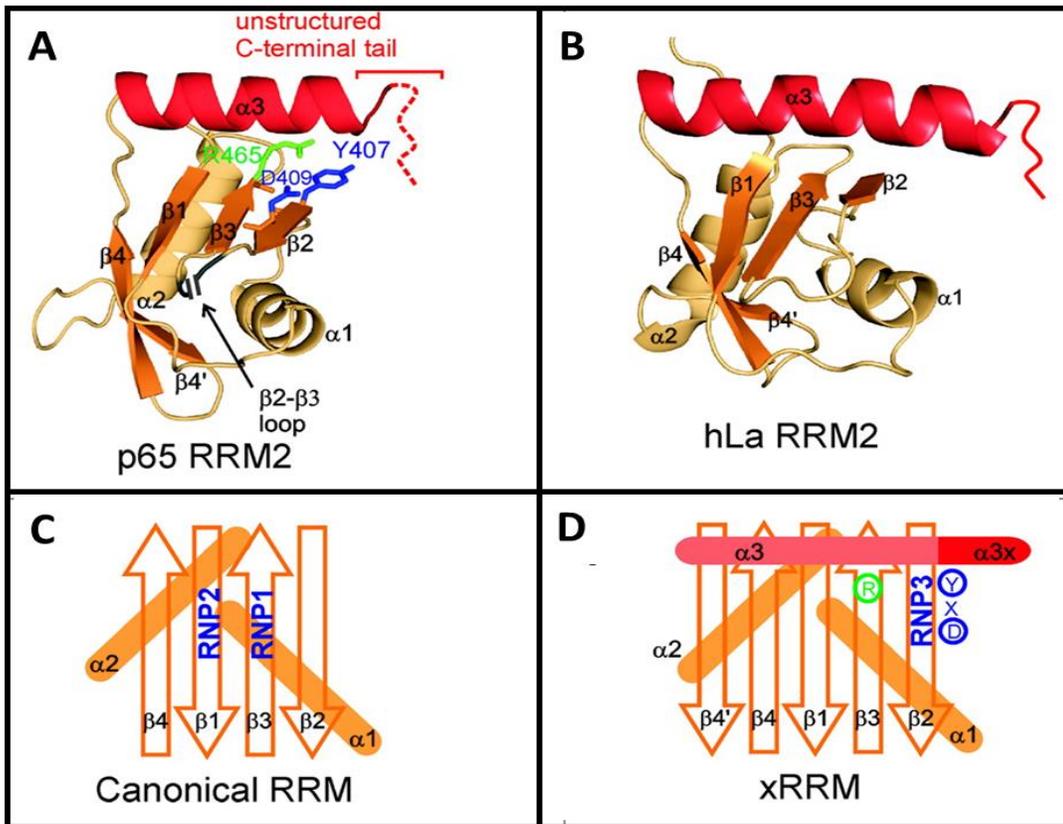


**Figure 2.1 Structure of the RNA Recognition Motif**. (A) The structure of p65 RRM2 domain (PDB ID 4EYT) and (B) the structure of the hLa RRM2 domain (PDB ID 1OWX). The non-canonical α3 helix is highlighted in red. (C) and (D) shows the schematic representation of a canonical and non-canonical RRM respectively. Reproduced from Singh *et al*., 2013.[170]

Recent work shows that when p65 binds to TER (telomerase) RNA, a previously disordered region C-terminal to RRM2 forms an α-helical extension to α3 that participates in RNA binding.[169] It was suggested that RRM2 in hLa would bind RNA in a similar mechanism due to the similarity in structure, however, efforts to acquire direct evidence *via* crystallization of the hLa CTD-RNA has not been successful.

Here we use TRESI-HDX-MS to probe the conformational changes in hLa-CTD to determine the structural differences when it binds to a single stranded RNA ligand and to the IRES domain IV of the hepatitis C viral (HCV) RNA.

## 2.3 Materials and Methods

### 2.3.1 Reagents

Deuterium oxide ($D_2O$; 99.99%), ammonium acetate ($C_2H_3O_2NH_4$; 99.99%), glacial acetic acid (99.85%) and pepsin-agarose beads were purchased from Sigma-Aldrich (Oakville, ON). Amicon Ultra-15 centrifugal filter units (10,000 molecular weight cut-off) were purchased form EMD Millipore (Etobicoke, ON) and Dialysis cassettes (10,000 molecular weight cut-off) were purchased from Fisher Scientific (Mississauga, ON). Nickel resin was purchased from GE healthcare (Mississauga, ON). A 23 mer ssRNA (sequence 5'-GAA UGU GGA AAA UCU CUA GCA CU-3' with a molecular weight of 7,395.5 Da) and the IRES domain IV of the hepatitis C viral RNA (5'-GCU GCA CCA UGA GCA CGA AUC CA-3' with a molecular weight of 7,328.5 Da) was obtained from IDT technologies (Kanata, ON).

### 2.3.2 Protein expression and purification

hLa 225-408 was expressed in *E. coli* BL21 DE3 cells and purified using immobilized metal affinity chromatography (IMAC). Briefly, plasmids encoding the CTD of hLa (pET28a hLa 225-408 in XL1) were obtained from laboratory stocks. The plasmid was extracted using a Life Technologies mini-prep kit and transformed in *E. coli* BL21 DE3 competent cells for protein expression. Bacterial cells were grown at 37°C to an optical density (OD) of 0.6 - 0.8 at 600 nm absorbance after which the bacterial culture was induced with 1 mM isopropyl beta-d-thiogalactopyranoside (IPTG) for an additional 3 hours. Cells were pelleted by centrifugation at 4226 x g for 15 minutes at 4°C. The pelleted cells were lysed, in 50 mM Tris-HCl (pH 7.6), 500

mM NaCl, 0.05% Igepal, 20 mM Imidazole, 5 mM β-mercaptoethanol (BME) and 100 μL Halt protease inhibitor cocktail, by sonication on ice for 7 minutes (pulse 15s on/15s off). The lysate was then centrifuged at 7234 x g for 30 minutes at 4°C. The supernatant was added to a gravity column packed with nickel resin and was incubated for at least 30 minutes on a shaker at 4°C. After incubation, the column was washed with 50 mM Tris-HCl (pH 7.6), 500 mM NaCl, 30 mM imidazole and 5 mM BME. The hLa-CTD was then eluted using 50 mM Tris-HCl (pH 7.6), 500 mM NaCl, 1 mM Ethylenediaminetetraacetic acid (EDTA), 200 mM imidazole and 5 mM BME. After purification, hLa-CTD was buffer exchanged using dialysis cassettes into 10 mM and 250 mM $C_2H_3O_2NH_4$, pH 8.1.

### 2.3.3 Native Mass Spectrometry and Ion Mobility.

For native hLa-CTD identification, 20 μM hLa-CTD in 10 mM and 250 mM $C_2H_3O_2NH_4$, pH 8.1 was identified using a quadrupole ion mobility time-of-flight mass spectrometer (Synapt G1, Waters) with an electrospray ionization (ESI) source. Nitrogen gas was used at the source whereas argon gas was used in the trap region of the mass spectrometer. In order to identify the protein/RNA complex, hLa-CTD in 10 mM and 250 mM $C_2H_3O_2NH_4$ (pH 8.1) and the 23 mer ssRNA or HCV RNA in ddH2O was mixed in a 1:1 ratio to a final concentration of 20 μM. The settings on the mass spectrometer were optimized in order to achieve ions without any fragmentation to maintain the native state of hLa-CTD and hLa-CTD/RNA complex. The instrument operated in positive ion mode with the capillary voltage at 2.5 kV, sampling cone at 100 V, extraction cone at 1.0 V and the source temperature at 100°C.

Similarly, an ion mobility spectrum was acquired for both hLa-CTD and the hLa-CTD/RNA complexes. The same MS conditions were used with an additional nitrogen gas source used for

ion mobility separation in the IMS T-wave cell of the Synapt G1 mass spectrometer. Wave velocity was set to 250 m/s at a wave height of 11.1 V.

### 2.3.4 Microfluidic chip fabrication.

The microfluidic device, used for HDX measurements, was made as previously described [173] with a few modifications. A proteolytic chamber and capillary channels were etched onto a poly methyl methacrylate (PMMA) block ($8.9 \times 3.8 \times 0.6$ cm) using a VersaLaser engraver (Universal Laser, Scottsdale, AZ). The TRESI mixer was made by inserting a glass capillary (outer diameter = 151 µm) into a metal capillary (inner diameter = 177.8 µm) which created an inter capillary space of 26.8 µm. Using the VersaLaser, the end of the glass capillary was sealed, and a notch was cut 2 mm from the sealed end. A mixing T with a dead volume of 51 nL was used to mix acid with the deuterated protein to quench the HDX reaction. The proteolytic chamber was filled with pepsin-agarose beads and a metal capillary was used as an outlet to the MS. A second PMMA block was placed below the constructed microfluidic device and a 3$^{rd}$ block with a sheet of silicon-rubber was placed on top to provide a liquid-tight seal which was secured using a custom-built pressure clamp. Gas tight Hamilton syringes were used to deliver the reagents through glass capillaries by using Syringe infusion pumps (Harvard Apparatus). The microfluidic device was then placed at the front end of a modified QSTAR Elite hybrid quadrupole time of flight (Q-TOF) mass spectrometer (Sciex) for HDX-MS in which a custom-built switch was used to stimulate the presence of the commercial ESI-source.

### 2.3.5 HDX of hLa 225-408.

The HDX reaction occurred inside the TRESI mixer where 70 µM hLa-CTD or hLa-CTD/RNA complex in 250 mM $C_2H_3O_2NH_4$ was allowed to flow through the inner glass capillary and the deuterium through the outer metal capillary both at 2 µL/min. Taking into account the total flow

rate in the TRESI mixer (4 µL/min), the inter capillary space and with the inner glass capillary pulled back 5mm, this created a mixing volume of 124 nL and HDX reaction time of 2.17 seconds. The instrument was operated at a source voltage of 4500 V in positive ion mode and the samples were scanned over a range of 400 to 1500 m/z. The experiment was done using 2 biological replicates with each having 3 technical replicates in which all the data was then averaged. A digestion profile of hLa-CTD and hLa-CTD/RNA complex was 1st acquired (without $D_2O$) in order to identify the peptides and a pepsin spectrum was acquired to eliminate any peptides produced from self -digestion of pepsin.

### 2.3.6 Data analysis.

The digests were analyzed using the FindPept tool on the ExPASy Proteomics server in order to determine the peptides and their sequence corresponding to each m/z. Peptides that could not be readily identified due to multiple 'best match', MS/MS was done on that particular m/z in order to fragment the peptide to produce a unique fragmentation pattern which was used to determine the correct peptide sequence. For each peptide, the experimental deuterium uptake was calculated using an in-house developed FORTRAN software which uses the change in the isotopic distribution to determine the amount of deuterium that is incorporated in each peptide.

### 2.3.7 NMR spectroscopy

hLa (225-332) protein for NMR spectroscopy was expressed in 1.5 L cultures of M9 media at 37 ˚C supplemented with 3 g/L ≥99% $^{13}C$ glucose, 1 g/L U-99%$^{15}N$ ammonium chloride and 1 g/L ≥98% $^{13}C,^{15}N$ algal extract (Cambridge Isotope Laboratories). At a cell density corresponding to A600 of 0.6, protein expression was induced with the addition of 1 mM IPTG and cultures were grown for a further 3 h. Cell pellets were resuspended in T300 buffer (20 mM Tris HCl, pH 7.8, 300 mM NaCl, 0.05 % $NaN_3$) + 2 M urea and lysed with a French press. The soluble cell extract

was applied to a 10 mL Ni-NTA affinity column (Qiagen) and washed with T300 buffer. After nonspecifically bound proteins were removed with a wash of T300 + 50 mM imidazole, hLa (225-332) protein was eluted with T300 + 20 mM EDTA. The solution was concentrated and subjected to gel filtration (Superdex HiLoad 16/60, GE Life Sciences) for further purification and desalting into buffer containing 10 mM Tris pH 7.8, 75 mM NaCl, 0.05% $NaN_3$. Protein solutions in this buffer were supplemented with 10% $D_2O$ and concentrated for NMR spectroscopy.

All experiments were performed with Bruker Avance 700 MHz NMR spectrometer equipped with a 5 mm TXI cryogenic probe at 300 K. Backbone chemical shifts assignments were achieved by a conventional triple resonance strategy (HNCACB, CBCACONH, HNCO, HNCACO) with non-uniform sampling using a 0.8 mM sample of hLa (225-332) alone, or in complex with a stoichiometric amount RNA. Datasets were processed with NMRpipe [174] and istHMS [175] and interpreted with CcpNmr Analysis. [176] Chemical shift differences were determined from $^1H,^{15}N$-HSQC spectra of uncomplexed and complexed samples according to the relationship:

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (0.14\Delta\delta_N)^2} \qquad (2.1)$$

## 2.4 Results and Discussion

### 2.4.1 RNA Binding to the hLa CTD.

In order to assess the importance of various regions of the hLa CTD in RNA binding, we performed electromobility shift assay (EMSA) using a variety of hLa-CTD mutants (Figure 2.2). While an RNA sequence specificity for the hLa-CTD has yet to be described, this region has been implicated in both the binding of single stranded pre-tRNA leaders [157] and a stem-loop located in the internal ribosome entry site (IRES) of the Hepatitis C virus (HCV). [177] In the case of the HCV RNA,

RRM2 was found to be important for binding a stem-loop with a 5' or 3' single stranded extension in the context of the rest of the NTD (*i.e.,* La motif/RRM1/RRM2 versus La motif/RRM1), but the RRM2 in isolation or with the disordered CTD was not tested. To test the importance of secondary structure in hLa CTD binding, we compared the HCV stem loop with a 6nt single stranded 3' extension (*HCV-RNA*; 23 nt total; estimated Tm of stem loop = 64°C) to a length matched 23nt sequence (*ssRNA*) hypothesized to form minimal secondary structure based on its previous use in annealing/strand dissociation assays (Naeeni *et al*., 2012; estimated Tm of most stable predicted secondary structure = 16°C).
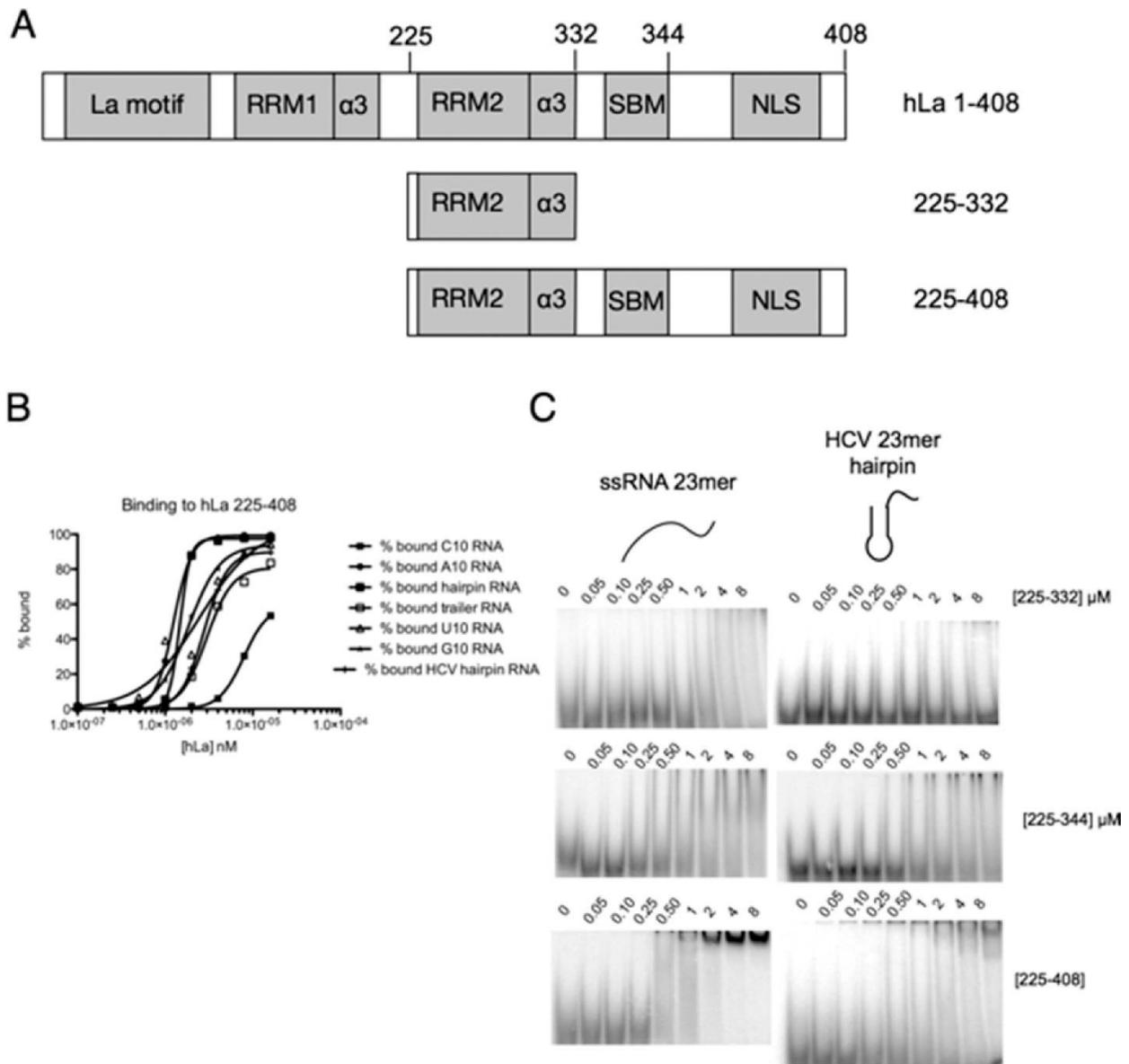
**Figure 2.2 Domain structure of hLa and EMSAs of the hLa CTD in complex with various ligands**. (A) Domain structure of hLa and the constructs used in this study. (B) EMSA-based affinity measurements for various hLa ligands. (C) Raw EMSAs for binding of length-matched single-stranded (ssRNA) or hairpin (HCV) RNAs to hLa CTD mutants carrying increasing portions of the disordered region C terminal to RRM2. Addition of the short basic motif (SBM; residues 225−344) or the entire CTD (residues 225−408) results in an increase in affinity for both RNAs compared to just the RRM2-α3 region (residues 225−332). Reproduced with permission from Brown *et al*., 2016. [178]

Our hLa CTD mutants contained increasing lengths of the disordered region C-terminal to RRM2, from just the RRM2 with its atypical $\alpha 3$ helix (225-332), to the inclusion of the short basic motif (225-344) implicated in pre-tRNA leader binding,[157] to the full length CTD (225-408). We observed that 225-332 was capable of binding some RNA at high concentration, but inclusion of progressively greater sections of the disordered CTD resulted in progressively higher affinity for both RNA substrates. For each CTD mutant, the ssRNA substrate was bound more tightly than the HCV-RNA, and inclusion of the full CTD (225-408) resulted in the formation of distinct, stable RNPs (especially for the ssRNA substrate) compared to more smeared RNP complexes with 225-332 and 225-344; the lack of discrete RNPs are often associated with higher $k_{off}$ rates.[179] These data are consistent with the disordered region C-terminal to RRM2 having an important function in hLa-CTD associated RNA binding and suggest that this region may bind single-stranded RNA more efficiently than RNA with secondary structure.

Native mass spectrometry was carried out in order to acquire insight into the global structure and dynamic nature of the protein. The CTD of hLa consists of a structured RRM domain and an unstructured polypeptide chain of about 81 amino acids (~44% of the CTD).[152] The native mass spectrum of hLa 225-408 acquired in 10 mM $C_2H_3O_2NH_4$ shows three distinct charge-state distributions (Figure 2.3A). While conclusions about liquid-phase protein structure based on ESI charge-state distributions should be drawn with a degree of caution, the 'trimodal' hLa distribution is strongly suggestive of multiple configurations and substantial conformational flexibility in solution.[180] In addition, low intensity peaks are observed between $m/z$ 2700 and 3200 with a deconvoluted mass of 45 kDa, corresponding to a CTD dimer. Acquired under higher salt conditions (250 mM $C_2H_3O_2NH_4$), there was a shift in the charge state distribution with only the putative 'folded' distribution being observed (Figure 2.4A). This was accompanied by loss of

dimer peaks suggesting hLa-CTD dimerizes exclusively in low salt conditions. These data provide

support for the notion that hLa can in some circumstances dimerize via its CTD, which has been

proposed previously, [162] however, the electrospray process can generate both false positives and

false negatives for biomolecular complexation and thus cannot provide unequivocal evidence for

dimerization.
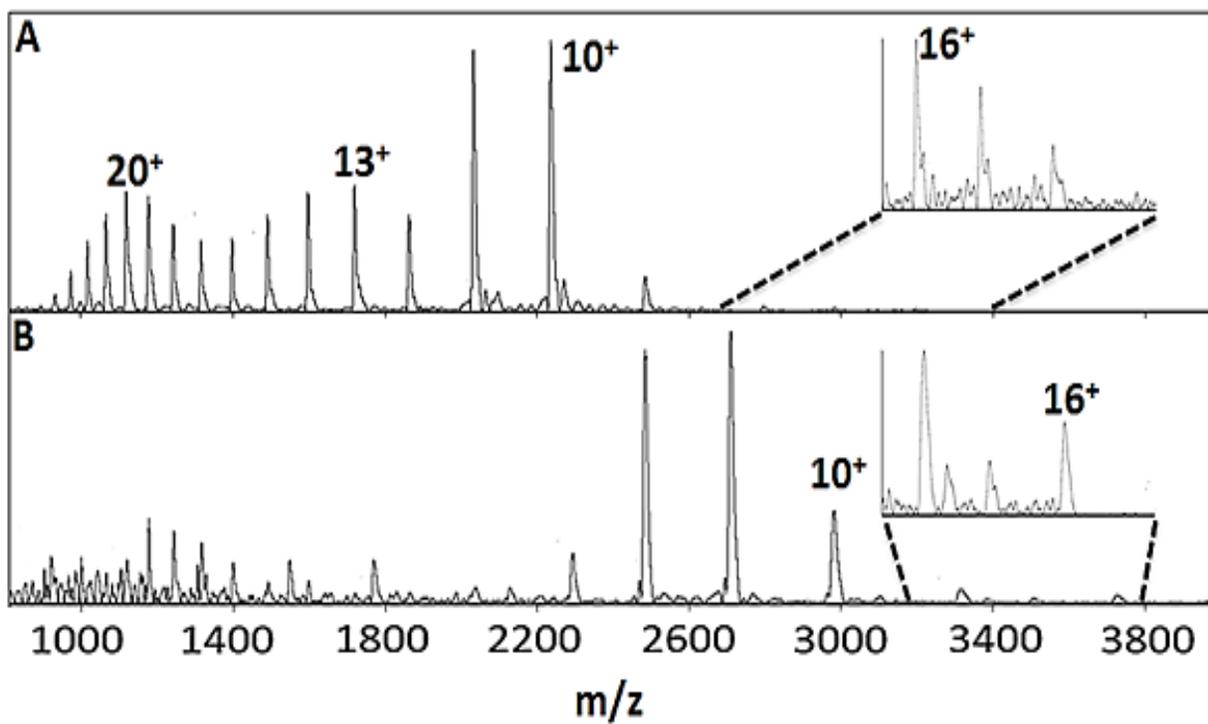


**Figure 2.3 Native mass spectra of the C-terminal domain of the human La protein**. (A) Spectrum of hLa (225–408) acquired in 10 mM $C_2H_3O_2NH_4$. Three distributions that represent an unfolded distribution (low $m/z$ value), an intermediate, and a folded distribution (high $m/z$ value) can be seen. Very low intensity peaks are observed between $m/z$ 2600 and 3200 that represent CTD dimers. (B) Spectrum of hLa (225–408) in 10 mM $C_2H_3O_2NH_4$ with the 23-mer ssRNA bound showing a folded distribution with dimer peaks observed between $m/z$ 3200 and 3800. The masses of these dimers correspond to two protein and two RNA molecules. Reproduced with permission from Brown *et al*., 2016. [178]

Native MS was also employed to observe global structural changes that occur upon RNA binding to hLa 225-408. In low salt (10 mM $C_2H_3O_2NH_4$), the hLa-CTD ssRNA complex exhibits only a folded distribution and a very low intensity unfolded distribution (Figure 2.3B), suggesting the protein becomes more structured upon binding to RNA. Peaks observed between m/z 3200 and 3800 yield a deconvoluted mass of 59.7 kDa, which corresponds to an hLa-CTD dimer with 2 RNA molecules bound. These peaks were substantially more intense than those observed for the CTD dimer in the absence of RNA indicating that CTD dimerization, when it occurs, is promoted by RNA binding. Electrosprayed from 250 mM $C_2H_3O_2NH_4$ the hLa-CTD ssRNA complex (Figure 2.4B) as well as the hLa-CTD HCV RNA (Figure 2.4C) complex exhibit only a 'folded' distribution with a deconvoluted mass of 29 kDa. The hLa-CTD HCV RNA complex also revealed, at low intensity, a set of peaks corresponding to a species with mass 37.2 kDa, which would correspond to a 1 hLa:2 HCV RNA complex. This complex is most likely an electrospray artifact resulting from non-specific adduction of HCV RNA to hLa, however, the fact that similar adducts are not observed for identical concentrations of ssRNA provides some support for the notion that the 'double HCV RNA' complex is specific and is a result of lower specificity of RRM2 for dsRNA targets.
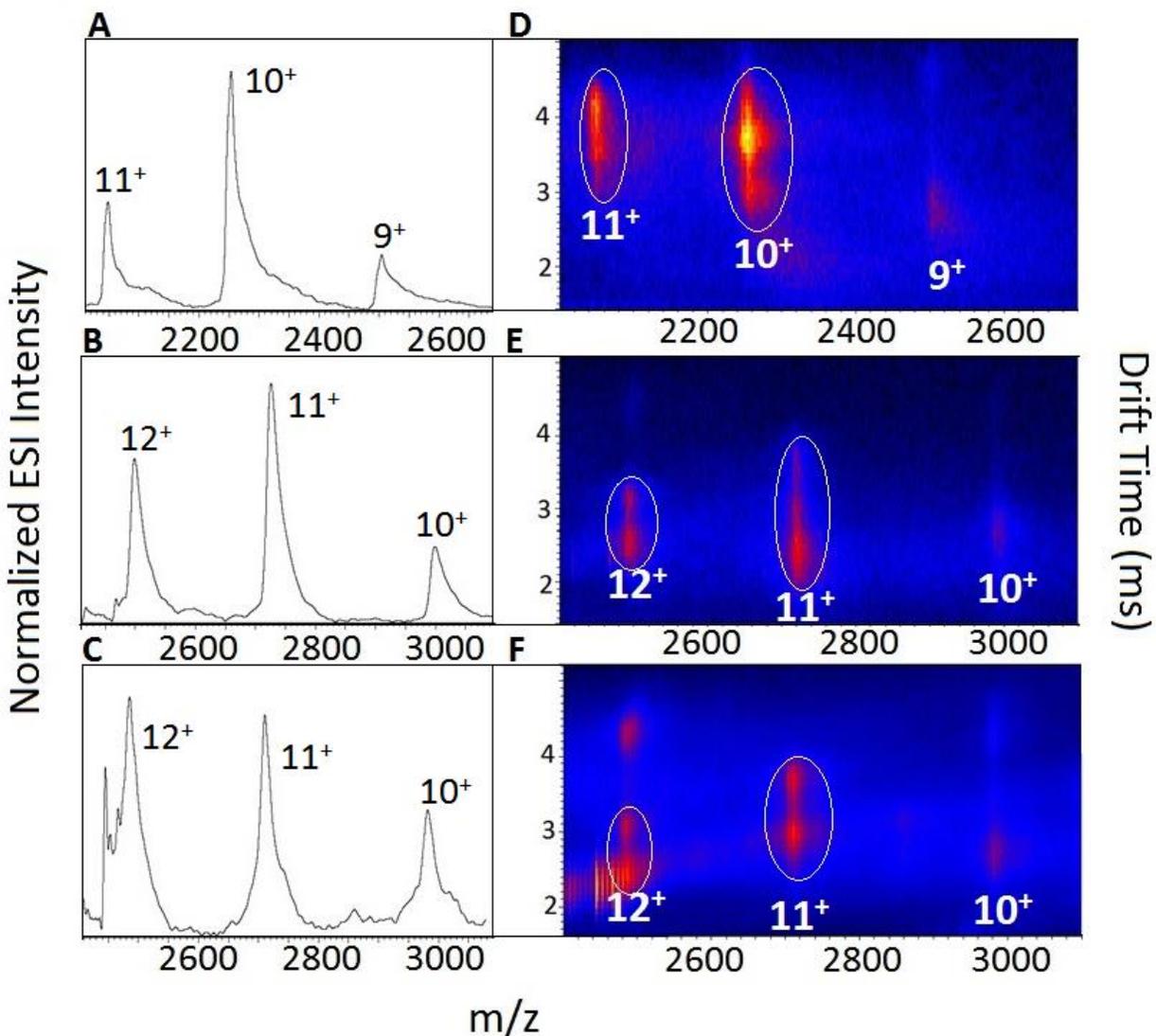
**Figure 2.4 Native mass spectra acquired in 250 mM C2H3O2NH4 (pH 8.1)** of (A) hLa(225–408), (B) hLa(225–408) bound to ssRNA, and (C) hLa(225–408) bound to HCV RNA. Ion mobility heat maps of (D) hLa(225–408), (E) hLa(225–408) bound to ssRNA, and (F) hLa(225–408) bound to HCV RNA. Reproduced with permission from Brown *et al.*, 2016. [178]

Ion mobility spectrometry-MS (IMS-MS) was used to provide further insight into the global conformational changes that occur upon RNA binding. IM-MS is effectively a gas-phase analogue of gel electrophoresis, separating ions in the mass spectrometer based on their size and shape. IMS is now a widely used technique for adding a 'structural' dimension to mass spectrometric

analyses.[181] There is substantial evidence that the gas-phase structures separated by some forms of IMS (including the 'travelling wave' setup used here) correspond very closely to structures populated in solution. [182,183] IMS-MS results in a three-dimensional dataset, usually represented as a heat map in which ions are *m/z* and size-delineated on a plane with an orthogonal 'heat' dimension representing intensity. The drift time, time it takes for ions traverse the ion mobility cell, is proportional to the collision cross section (a convolution of size and shape) of the gas phase ions. In the case of hLa-CTD, the RNA-free IMS-MS spectrum (Figure 2.4D) clearly shows the occurrence of two conformations at the $10^+$ charge state with a strongly dominant 'larger' conformation. Assuming a globular structure, the 0.3 ms difference in drift time between the two conformations, corresponds to an approximately 15% smaller diameter for the compact species. Upon RNA binding to the ssRNA (Figure 2.4E), at the $10^+$ charge state the more compact conformation remains, but the larger conformation disappears completely. Upon RNA binding to the HCV RNA (Figure 2.4F), the $10^+$ charge state shows multiple conformations. Although the most dominant conformation is the most compact, there is still a small population of the larger conformation that remains visible. The $11^+$ ssRNA-bound peak shows multiple conformations with a dominant compact state. In the case of HCV RNA binding, the more compact and the larger conformation are equally populated at $11^+$. These results together with the EMSA would suggest that the hLa-CTD binds more strongly to the ssRNA than it does to the HCV RNA in both the gas phase and solution phase. However, the appearance of a 'pre-existing' compact configuration similar to the RNA-bound conformation in the absence of RNA provides direct evidence that RRM2-mediated RNA binding in hLa occurs *via* conformer selection. [184]

### 2.4.2 Ligand- binding associated structure and dynamics changes in hLa via TRESI-HDX.

Much of what is predicted about RNA binding in the hLa CTD arises from comparison with the hLa homologue p65, for which there is a crystal structure of the RNA complex. Like hLa, p65 has an unusual RRM in the CTD in which there is a $3^{rd}$ α-helix that lies across the typical nucleic acid binding site. This atypical RRM is followed by an unstructured polypeptide chain, which is also present in hLa.[169,172] Previous structural work suggests that upon p65 RRM2 binding to telomerase RNA, the previously unstructured region C-terminal to RRM2 forms an extension to the α3 helix, and that this newly ordered region participates in RNA binding. Due to the similarity in structure with hLa RRM2, it was suggested that it would bind RNA in a similar manner. [169]

When the ssRNA ligand and HCV RNA bind to the RRM2 of hLa there is an overall decrease in the deuterium uptake (Figure 2.5), consistent with both RNA associated shielding and a general increase in structural compactness as indicated by the native MS and IMS-MS data. There are some regions of the CTD that show a more significant decrease in deuterium uptake, which could suggest possible RNA interaction sites. During canonical RRM-RNA binding, RNP motifs in the β1 and β3 strands often contain aromatic residues that stack between bases of the RNA target, [172] but this mode of binding is not expected for the hLa RRM2 due to the presence of the non-canonical α3 helix obscuring this site. [152] Our results show no significant decrease in deuterium uptake in the β1 strand when either RNA binds, however the peptide 271-274 in β3 shows a significant decrease in deuterium uptake (-20 %) upon binding to both the ssRNA and the HCV RNA (Figure 2.5). Adjacent to this peptide, the loop connecting the β2 and β3 strands (peptide 265-270) shows an even stronger decrease in deuterium uptake (-30 %) when the ssRNA is bound; the analogous region in RRM1 of hLa has been shown to have direct RNA contacts [167], and this

region is noted to vary in size considerably between LARP family members and between the RRM2 of hLa versus p65 [136,139,169]. Although this exact peptide (265-270) was not observed in the hLa-CTD HCV RNA complex, a peptide spanning the same loop and the β3 strand (265-274) did not have a significant decrease in deuterium uptake (Figure 2.5B). Because the β3 strand did show a significant decrease, it can be concluded that there are likely no significant contacts in the loop region connecting β2 and β3 strands when HCV RNA binds RRM2, which may account for its lower affinity.
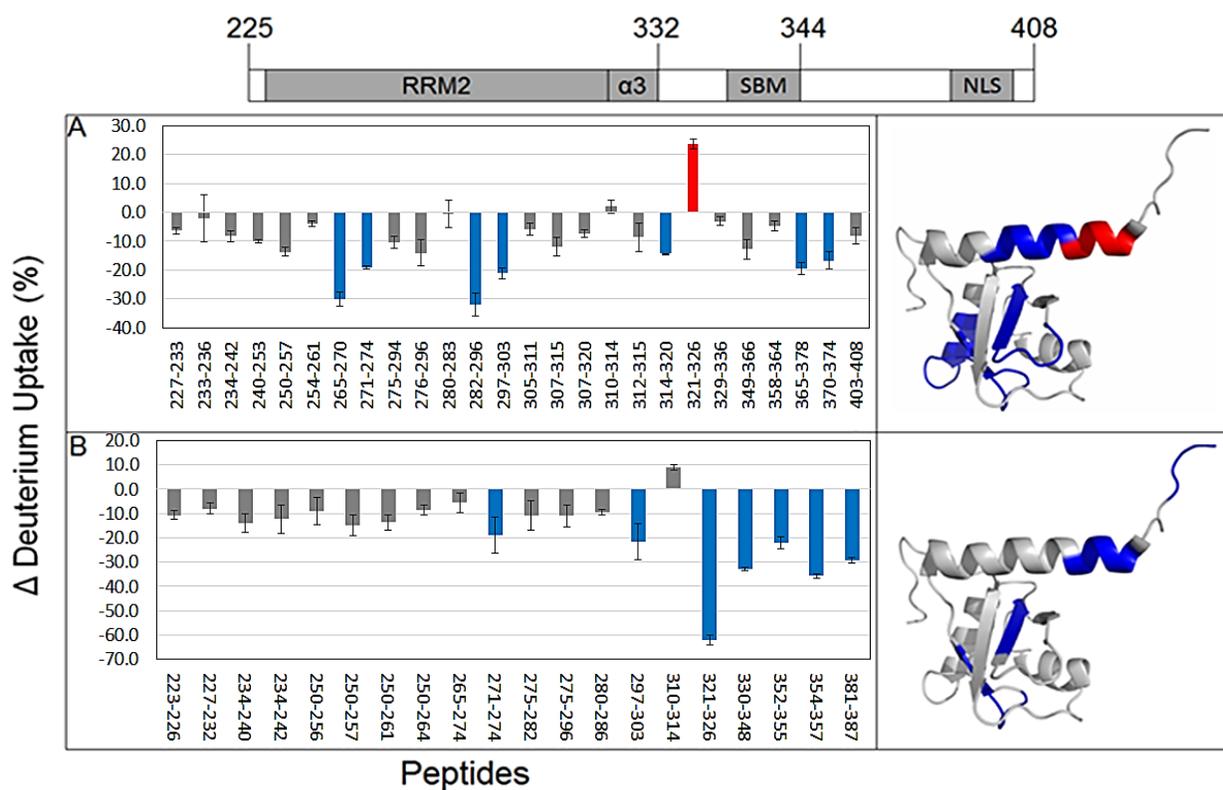


**Figure 2.5 Absolute differences in deuterium uptake when hLa-CTD is bound to** (A) the 23-mer ssRNA ligand and (B) the HCV RNA. The uptake differences are mapped onto Protein Data Bank entry 1OWX, representing the RRM2 domain of the hLa CTD, in which peptides colored blue correspond to the blue bars in the histogram representing significant decreases in the rate of deuterium uptake and the peptide colored red corresponds to the red bar in the histogram representing a significant increase. Error bars are generated from a minimum of three independent replicates ($n \geq 3$). Reproduced with permission from Brown *et al.*, 2016. [178]

58

An interesting region showing a significant decrease in deuterium uptake when bound to ssRNA comprises residues 314-320, located within the α3-helix. This region contains the two most critical residues associated with nuclear retention; point mutation of hLa K316 and K317 (or the analogous lysines in fission yeast La) to alanine result in aberrant nuclear export of La to the cytoplasm and dysfunctional La dependent pre-tRNA processing. [143] Thus, these data suggest that the mechanism of hLa nuclear retention may be linked to modes by which La engages RNA substrates. Decreases in deuterium uptake were not limited to RRM2 upon ssRNA binding. For instance, residues 365-378 in the unstructured polypeptide chain also displayed significantly lowered uptake. This region includes the major phosphorylation site for hLa, S366, whose phosphorylation is hypothesized to inhibit La binding to pre-tRNA leaders and control their processing by RNase P,[157] as well as a region recently hypothesized to function in CTD-associated RNA chaperone function.[164] Substantial and wide-ranging decreases in deuterium uptake were also observed in the unstructured polypeptide chain upon HCV RNA binding. These include peptides 352-357 and 381-387 which had a large decrease of -30%. These observations underscore the importance of the unstructured C-terminal region to RNA binding, in agreement with our EMSA data.

While the dynamic changes described above indicate somewhat distinct binding modes for ss- and HCV RNA in terms of involvement of the RRM2 and C-terminal regions, the most striking difference was localized to residues 321-326, the transition region between the α3 helix and the disordered region C-terminal to RRM2. Interestingly, we observed that this peptide becomes significantly less structured upon ssRNA binding, as evidenced by a large increase in deuterium uptake (20%) (Figure 2.5A) whereas it becomes significantly more structured upon engagement with HCV RNA, which is evident by the large decrease in deuterium uptake (-60%) (Figure 2.5B). Previous NMR work indicates that the final turn of this helix extends to residue S325 but starts to

become ill-defined starting at residue Q322.[152] We observed that when bound to the ssRNA, the amino acids following this region (329-336) showed no significant change in deuterium uptake, indicating an absence of the α3 helical extension predicted from the p65 analogue. Thus, the data are consistent with RNA-induced structural changes that occur in an equivalent region to p65, but are opposite in nature (*i.e*., helical unwinding as opposed to helical extension). However, when bound to the HCV RNA, the peptide 330-348 (following the α3 helix) shows a substantial decrease in deuterium uptake (-32%) (Figure 2.5B), suggestive of an extension in the α3 helix as proposed by the p65 model.

### 2.4.3 Comparison to Δδ NMR

Of the limited set of experimental approaches that can provide detailed structural and/or dynamic information in the context of ligand binding, the NMR chemical shift change analysis (Δδ) is most commonly used for protein-RNA interactions. Compared to MS-enabled HDX, this approach has the advantage of being site specific (MS-based HDX measurements are usually segment averaged), however it shares the disadvantage that it is not always possible to distinguish between true interactions and allosteric effects. In contrast to MS-based HDX, it is often not possible to determine from Δδ data whether a particular shift corresponds to an increase or decrease in dynamics, although the development of stable secondary structure can be tracked using chemical shift index analysis. [185] In addition, Δδ experiments are subject to the limitations of NMR as an analytical method, such as relative insensitivity and a fundamental analyte size limit. There is also a significant 'blind spot' associated with conformational or binding/unbinding transitions that occur on the millisecond timescale owing to line broadening. This limitation is particularly felt when examining weakly structured regions of proteins or high on/off rate interactions.
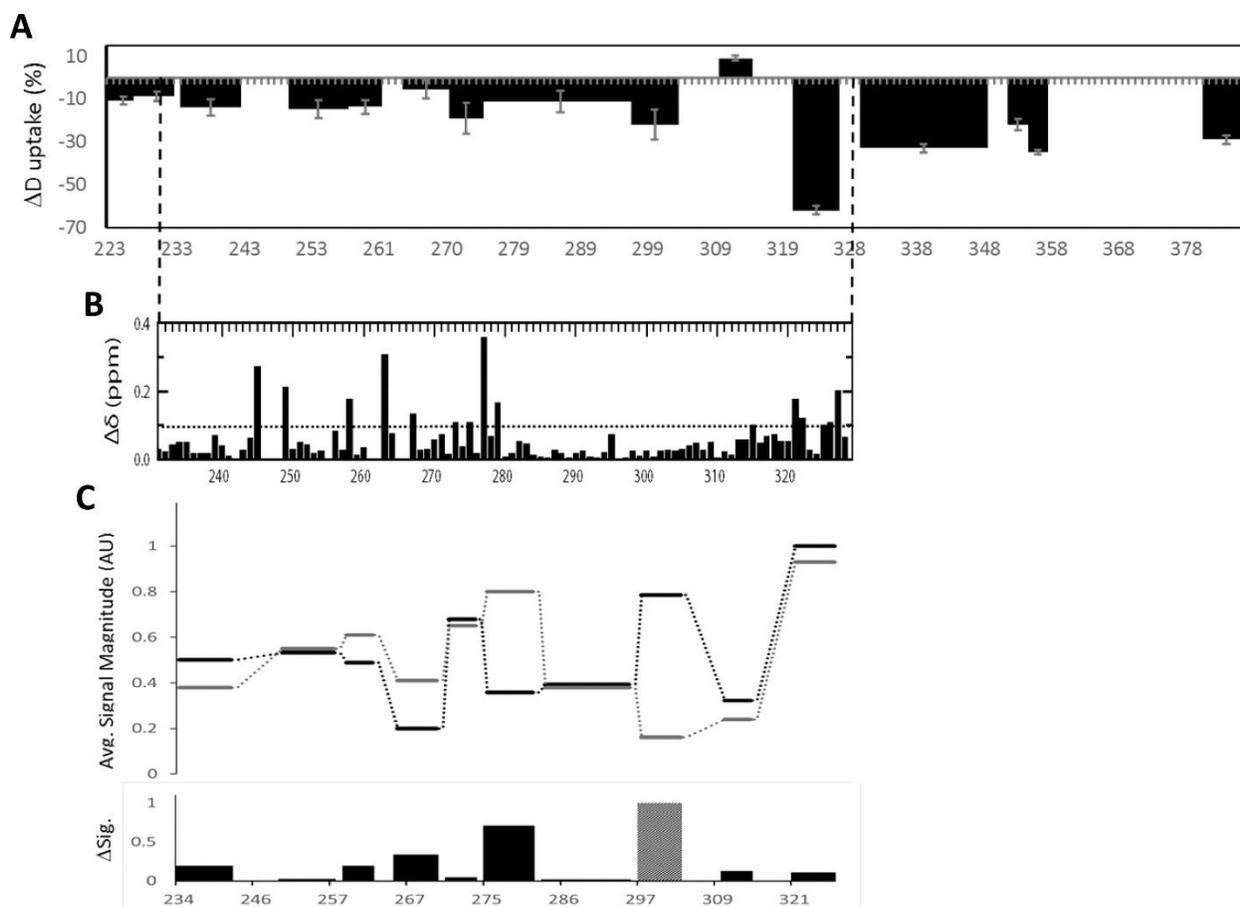
**Figure 2.6 Comparison between TRESI-HDX and NMR data** (A) TRESI-HDX uptake differences and (B) NMR peak shift data for the HCV RNA binding interaction. The horizontal line in the NMR plot indicates the significance threshold. In the TRESI-HDX plot, statistical significance is given by the standard deviation from at least three independent runs ($n \geq 3$). The data are aligned by sequence to allow a direct comparison between observations derived from solvent access and backbone amide hydrogen bonding averaged over regions (TRESI-HDX) and site-specific measurements of changes in the electronic environment of the backbone amides (NMR). (C) Comparison between the normalized HDX "signal magnitude" (i.e., the magnitude of the difference in the rate of HDX uptake between bound and unbound hLa) and the normalized $\Delta\delta$ "signal magnitude" (i.e., the average $\Delta\delta$ for segments corresponding to the HDX profile). A "residuals" plot is provided below, showing the magnitude of the difference between the averaged, normalized $\Delta\delta$, and HDX profiles. The gray bar indicates a region in which HDX exhibits substantially higher sensitivity to conformational and dynamic changes. Reproduced with permission from Brown *et al*., 2016. [178]

Assuming that peak assignments can be made, it is sometimes possible to overcome these limitations using more sophisticated NMR methods such as relaxation dispersion, $T_2$ relaxation and TROSY.

A direct comparison of $\Delta\delta$ and TRESI-HDX analyses for the HCV RNA is provided in Figure 2.6. In general, there is limited agreement between the two datasets. Sizeable $\Delta\delta$s at glutamic acid 277 and aspartic acid 263 appear to have had little impact on the corresponding regions of the MS data, for instance, and several regions identified by TRESI-HDX as undergoing a substantial change in dynamics do not contain any residues with $\Delta\delta$s above the threshold (e.g., regions around 239 and 300 and 311). Part of this lack of agreement can be attributed to the 'segment averaged' nature of MS-based HDX data versus site specific NMR measurements. There is substantially improved agreement between the TRESI-HDX profile and the $\Delta\delta$ profile when the latter is normalized and averaged over the same segments measured in the HDX data (Figure 2.6C). Of the remaining regions with significant disagreement, two result from a substantially stronger NMR signal and one (highlighted as a gray bar in the bottom plot of Figure 2.6C) results from a substantially stronger HDX signal. These regions highlight the fact that, in contrast to HDX measurements that use MS or NMR detection, HDX and $\Delta\delta$ NMR measurements are probing fundamentally different phenomena and are therefore not fully comparable. One region of good agreement between the TRESI-HDX and $\Delta\delta$-NMR profiles, with both indicating significant structural perturbation, corresponds to the C-terminus of the α3 helix, which is the site of helical extension in our model for HCV RNA binding.

**2.5 Conclusions**

Taken together, our data indicate significantly different binding modes for unstructured ssRNA versus structured HCV RNA binding in the atypical RRM motif of hLa. Unstructured ssRNA binds with high relative affinity via a mechanism that involves close contacts with the RRM2 domain, destabilization of the α3 helix, and engagement with select segments of the unstructured C-terminal region. Structured RNA binds with a lower affinity via a mechanism that involves fewer contacts in the RRM2 binding pocket, extension of the α3 helix, and more widespread engagement with the unstructured C-terminal region. This second mechanism has some commonalities with the structural changes that are known to occur upon binding of hLa RRM2 homologue p65 to telomerase RNA.[169,186] TRESI-HDX provides a powerful alternative for probing conformational and dynamic changes that occur in ligand binding, offering in this case an expanded view of the interaction that includes the unstructured C-terminus (which was not resolved in the NMR data). TRESIHDX and $\Delta\delta$ NMR are complementary probes of ligand binding, sensitive to different fundamental processes that occur in regions of significant conformational or dynamic change.

## Chapter 3: An interdomain bridge influences RNA binding of the human La protein

Stefano A. Marrella, **Kerene A. Brown**, Farnaz Mansouri-Noori, Jennifer Porat, Derek J. Wilson, and Mark A. Bayfield.

All Mass Spectrometry Experiments were conducted and analyzed by **KAB** and Electrophoretic mobility shift assays (in the original paper) were conducted by **SAM.** Only the mass spectrometry experiments from the original paper is shown in this chapter.

## 3.1 Summary

Human La (hLa) is a multifunctional RNA chaperone with varying functions relating to its subcellular location. In the nucleus, hLa is a crucial factor in the processing of RNA polymerase III transcripts, whereas in the cytoplasm, hLa is involved in the translation of cellular and viral mRNAs. Chapter 2 provided some insight on the binding mechanism of the C-terminal domain (CTD) of hLa that was previously unknown. It has been known however to not take part in Poly (U) RNA binding which is characteristic of the N-terminal domain (NTD). Here we further explore how hLa interacts with other non-poly (U) RNA and how mutations in the nuclear trafficking signals affect RNA binding. The nuclear retention element (NRE) and nuclear export signal (NES) have been previously shown to control the accumulation of hLa in the cytoplasm and nucleus. Point mutations to residues in the NRE and NES that are associated with localization, affects binding of hLa to various RNA substrates. Here, we show evidence of an interaction between the NRE and NES, spanning across the RRM1 and RRM2 domains of the protein.  In addition, the conformational dynamics that occur when this interaction is intact or disrupted are studied via collision induced unfolding and time-resolved hydrogen-deuterium exchange mass spectrometry. Our data suggest that the intracellular distribution of La is specifically linked to its RNA binding mode and provides the first evidence for a direct interdomain interaction in La proteins.

## 3.2 Introduction

Human La (hLa) is a highly abundant and multifunctional RNA chaperone.[187] Although hLa protein is predominately nuclear,[188] it shuttles between the nucleus and the cytoplasm in which the cytoplasmic levels increase in times of cellular stress and viral infections.[189,190] Within the nucleus it binds to nascent RNA polymerase III transcripts preventing exonuclease digestion. This mechanism of recognition and binding is well studied and is characteristic of the N-terminal

domain (NTD). [191,192] In the cytoplasm however, hLa associate with the translation of cellular and viral coding RNAs. Human La was recognized as the first cellular factor crucial for the translation of viral internal ribosome entry site (IRES), [193,194] and in chapter 2 we shed some light as to how the C-terminal domain (CTD) engages the IRES domain of the hepatitis C viral RNA.[178]

The transportation of hLa between the nucleus and cytoplasm is controlled through a nuclear localization signal (NLS) and a nuclear retention element (NRE) [195]. Previous work done by Bayfield *et al*., revealed two highly conserved residues within the NRE which are identified to be crucial for its activity, Lysine-316 and Lysine-317. When the entire NRE (residues 316 -332) is deleted hLa becomes cytoplasmic and as a result pre-tRNA processing is suppressed.[196] These lysine resides when substituted with alanine mimics the activity of the hLAΔNRE mutant. Deletion of the RRM1 restored nuclear accumulation due to the removal of the nuclear export signal (NES) which was mapped to two conserved residues glutamic acid 132 and aspartic acid 133. The lysines in the NRE are hypothesized to form an interdomain salt bridge with glutamic acid 132 and aspartic acid 133 in RRM1. E132 and D133 in the RRM1 when substituted with alanines in a hLa K316A317A mutant were found to be nuclear and restored pre-tRNA processing. This NES is functionally suppressed by the NRE however there is no direct evidence for any physical NES-NRE interaction. Therefore, the mechanism by which the NRE and NES contribute to the intracellular distribution of hLa is poorly understood.

Recently it has been demonstrated that La-mRNA binding (which takes place in the cytoplasm) occurs in part through sequence specific recognition of poly(A) tail.[116] The altered localization of hLa variants with mutations to their NRE and/or NES could be related to an altered capacity of said mutants to engage La RNA target(s). We therefore tested hLa and hLaΔNRE for changes in structural dynamics when bound to UUU-3'OH containing substrate (U10) and a poly(A) substrate

(A20). We also tested for a direct interaction between the NRE and NES by making compensatory mutations to these two regions and assessing these mutants for their ability to bind these same RNAs. Our data demonstrate that mutations to the NRE and NES indeed affect the association of hLa with RNA, and provide evidence of a direct, physical interaction between the NES on RRM1 and the NRE on RRM2. Altered engagement of NRE mutants with RNA may thus play a role in their altered subcellular localization.
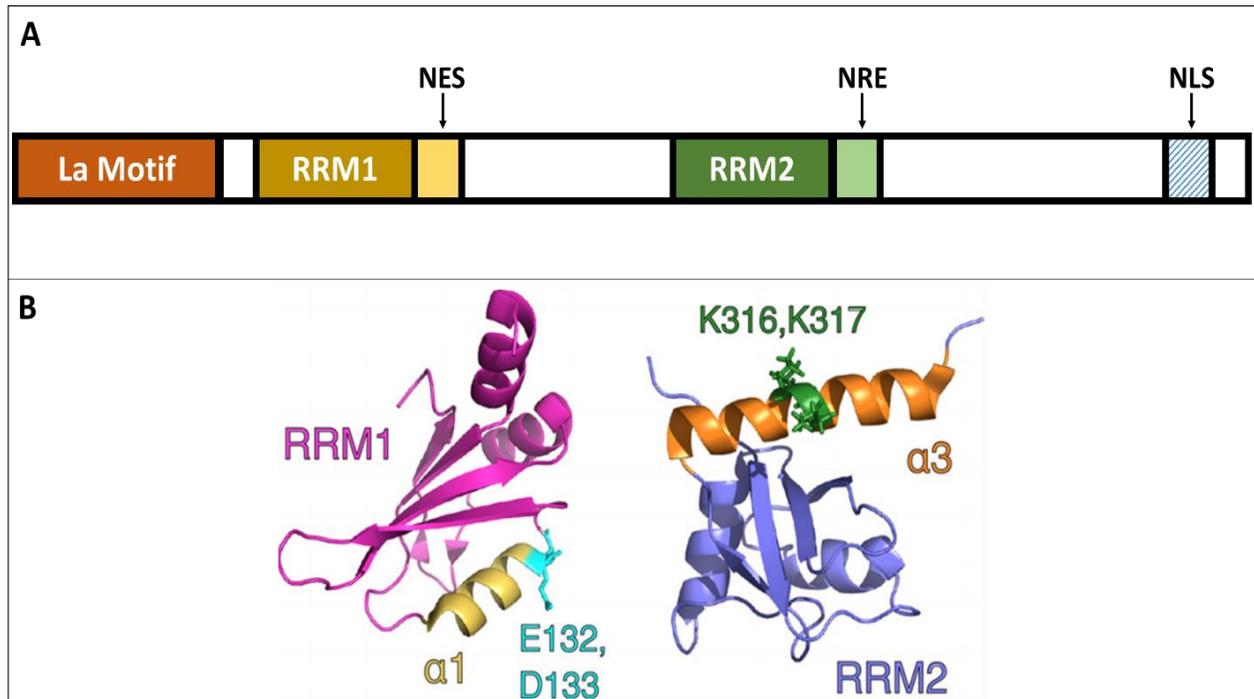


**Figure 3.1 Domain organization of human La protein (hLa).** A.) Domain organization of hLa showing the locations of Nuclear Export Signal (NES), Nuclear Retention Element (NRE), and the Nuclear Localization signal (NLS). B.) Structural representations of RRM1 and RRM2 domains of human La. RRM1 is in pink with RRM1-α1 in yellow, and Glu-132 and Asp-133 are in cyan. RRM2 is in violet, with RRM2-α3 in orange, and Lys-316 and Lys-317 are in green. Structures were generated from PDB codes 2VOO and 1OWX.

### 3.3 Materials and Methods

#### 3.3.1 Protein Expression and Purification

The pET 28a vector in *E. coli* BL21 DE3 encoding hLa, hLaΔNRE, hLa E132K/D133K/K316E/K317E and hLa K316E/K317E were provided by the Bayfield Lab (York University). Protein expression and purification were carried out with induction of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) to over express the proteins of interest followed by purification using immobilized metal affinity chromatography (IMAC) as described in chapter 2. A second purification step was added, using a 5 mL HiTrap Heparin column pre-packed with Heparin Sepharose, to eliminate bound nucleic acids that may have co-purified with hLa. The heparin column was equilibrated with a buffer containing: 20 mM Tris-HCL, pH 7.6, 1.5 mM $MgCl_2$, 100 mM KCL, and 0.2 mM Ethylenediaminetetraacetic acid (EDTA). Using a syringe, the elution from the IMAC purification was loaded onto the heparin column. The column was then washed with increasing concentrations of KCL (200 mM, 400mM and 800 mM) in which the protein was eluted at 800 mM KCL. Purified hLa was then concentrated and buffer exchanged, using centrifugal filter units, into 500 mM $C_2H_3O_2NH_4$, pH 8.0 for subsequent mass spectrometry analysis.

#### 3.3.2 Collision Induced Unfolding

Samples were buffer exchanged into 250 mM $C_2H_3O_2NH_4$, pH 8 by ultrafiltration and analyzed using a quadrupole-ion mobility-time-of-fight mass spectrometer instrument (Synapt G2-S HDMS, Waters, Milford, MA). Protein ions were generated using an electrospray ionization (ESI) source in positive mode. The source operated with a capillary voltage of 3.0 kV and sampling cone of 100 V. To generate ion mobility separation, the travelling wave ion mobility cell operated at a pressure of 3.79 mbar and series of DC voltage waves (wave height of 40 V with a velocity of 652

m/s) were used. The ToF-MS operated over the m/z range of 1000 – 5000 at a pressure of 1.09 x10$^{-6}$ mbar.

To perform CIU experiments, collision energy was applied to the ions in the ion trap located prior to the ion mobility cell. The collision energy voltage was ramped from 10 to 100 V in 5 V increments to construct a CIU fingerprint. The CIU fingerprints were generated using PULSAR software.[197]

### 3.3.3 Microfluidic chip fabrication

The microfluidic device was constructed as previously described in chapter 2 with some modifications. The microfluidic chip was made smaller for added convenience. Briefly, a proteolytic chamber and capillary channels were etched onto a poly methyl methacrylate (PMMA) substrate (dimensions: 2.1 cm x 5.1 x 0.6) using a VersaLaser engraver (Universal Laser, Scottsdale, AZ). Another blank PMMA substrate with the same dimensions was used and the two were welded together using acrylic glue and left to dry for 24 hours. A drill press and a 10 32 NF Bit was used to drill both sides of the substrate to a depth of 9 mm. A smaller drill bit, 1/16" was used to drill both sides to depth of 20 mm. A 10 32 NF bottoming tap was then used to manually add threads. A 10 cm glass capillary with an outer diameter of 151 µm and a 10 cm 33-gauge stainless steel metal capillary screwed into the chip via double-winged nuts were used as the input and output channel respectively. The output channel also served as the ESI capillary.

The TRESI mixer was made by inserting a glass capillary (Outer Diameter = 151 µm) into a metal capillary (inner diameter = 177.8 µM) creating an inter capillary space of 26 µm. The end of the glass capillary was sealed, and a notch was cut 2 mm from the sealed end using the Vera Laser. The device was placed at the front end of a modified NanoSpray platform on a quadrupole-ion

mobility-time-of-fight mass spectrometer instrument (Synapt G2-S HDMS, Waters, Milford, MA).

### 3.3.4 Time-Resolved ElectroSpray Ionization Hydrogen Deuterium eXchange (TRESI-HDX)

The HDX reaction took place inside the TRESI mixer in which gas tight Hamilton syringes were used to deliver the reagents through the capillaries using Syringe infusion pumps (Harvard Apparatus). Human La at 5 µM in 250 mM $C_2H_3O_2NH_4$ flowed through the inner capillary while deuterium flowed through the outer capillary at 2 µL/min. The total flow rate in the TRESI mixer (4 µL), the inter-capillary space and the inner capillary pulled back at 1 mm created a reaction volume of 25.8 nL and HDX reaction time of 0.59 s. The instrument operated at a voltage of 2.25 kV in positive ion mode with a sampling cone voltage of 25 V. The samples were scanned over a range of 400 to 1500 m/z. The experiment was done with two biological replicates each having three technical replicates.

A digestion profile was acquired prior to the HDX reaction to identify the peptides. The m/z and charge of each peptide was recorded and ExPASY findpept was used to find the "best-match" for each m/z. Tandem MS was then used to fragment the peptides to obtain a fragmentation pattern that was used to verify the correct identity of each peptide. Data Analysis was carried out using the MS studio software[198] in which the deuterium levels for each peptide in free protein and protein/RNA complex was calculated. Differences in deuterium uptake that fell with 2 sigmas (95% confidence interval) was considered to be significant. The significant differences were then mapped onto the X-ray crystallography and NMR structures.

## 3.4 Results and Discussion

### 3.4.1 Native Mass Spectrometry Analysis of human La protein

Native mass spectrometry analysis of wild type human La protein reveals the mass to be 47.5 kDa. Three structural populations were observed in the native spectrum (Figure 3.2). The folded protein distribution was the most prominent with charge states $12^+$ to $17^+$. A highly charged unfolded distribution was observed at lower m/z values. This is expected, although hLa has three structured domains, as it has a large proportion of unstructured loops in its structure. This is the reason there is no high-resolution structural data for the entire protein, only the isolated structured domains. [117,199–201] In addition a small population of hLa dimers were observed at higher m/z, but this may be an ESI artifact and is not sufficient evidence for the presence of hLa dimers.
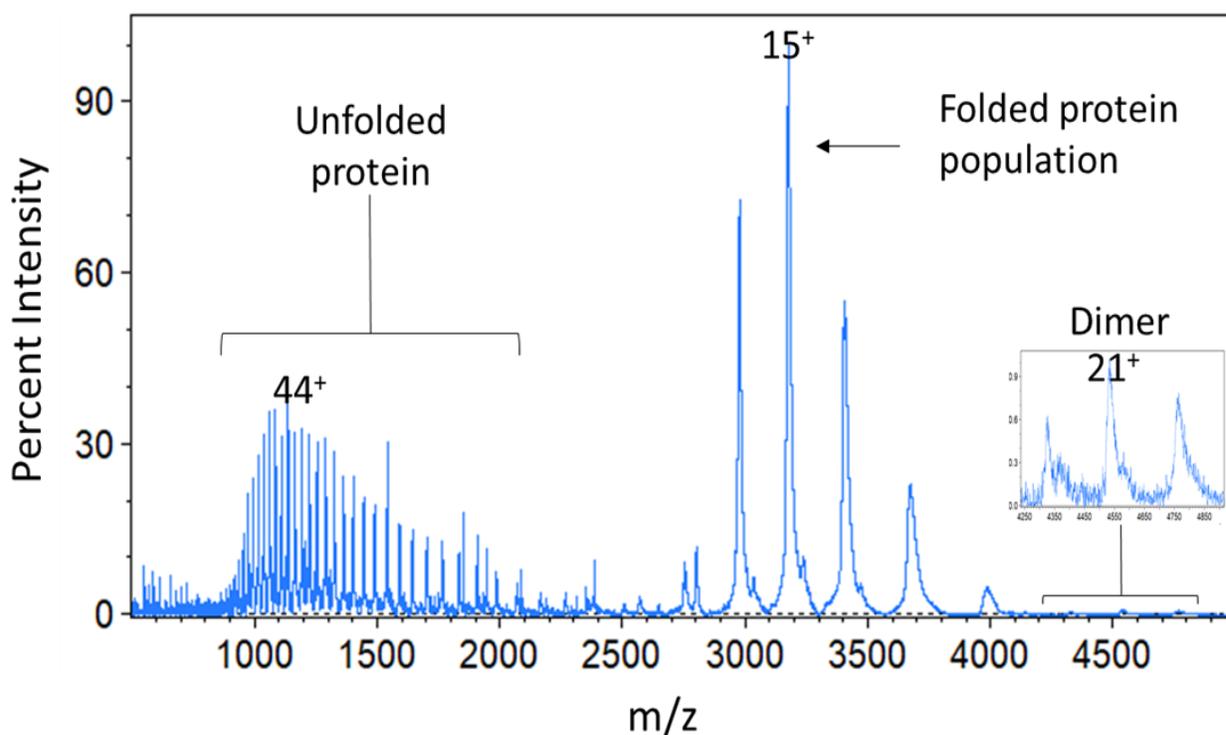


**Figure 3.2. Native Mass Spectrum of the wild type human La (hLa), showing a folded and unfolding population as well as a low intensity dimer population.**

For simplicity, only the folded distribution is highlighted when observing RNA binding. The native

spectra in figure 3.3 reveal a 1:1 binding ratio of hLa when bound to both U10 and A20 RNA
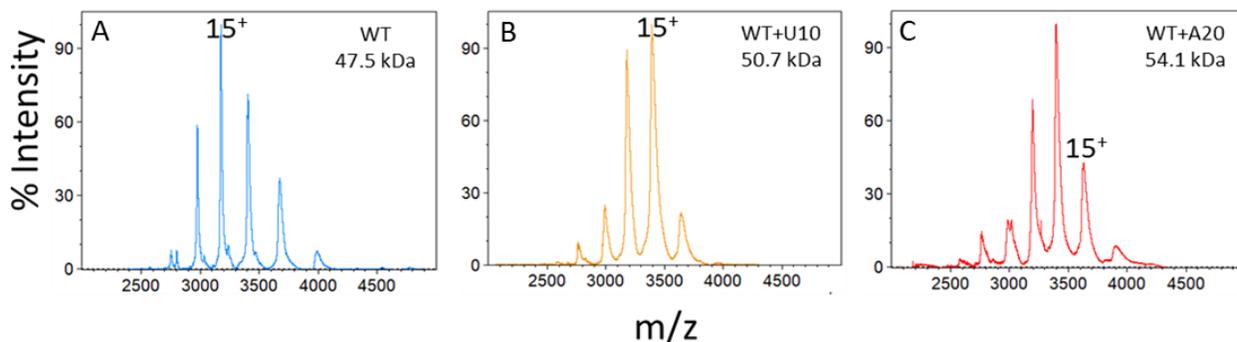


**Figure 3.3. Native Mass Spectra of hLa-RNA complexes.** A. hLa with a mass of 47.5 kDa B. hLA-U10 complex with a mass of 50.7 kDa and C. hLa-A20 complex with a mass of 54.1 kDa.

### 3.4.2 Collisional Induced Unfolding reveals structural differences and RNA binding patterns in hLa and the NRE deficient mutant.

To study the changes in structural dynamics in hLa and the NRE deficient mutant, we 1st employed

a technique referred to as Collision Induced Unfolding (CIU). CIU uses Ion Mobility Mass

Spectrometry (IM-MS) to study the gas phase stability of proteins in the presence and absence of

ligands.[202,203] IM-MS is a gas phase analog of electrophoresis, as ions are separated based on their

size and charge. It measures the drift time of ions that can be related back to the collisional cross

section of the protein, which has been proven to correlate well with the X-ray crystallography

structures in many instances.[34,204] In the CIU experiment, protein ions are subjected to increasing

collision energies (voltage) which results in unfolding of the protein and the accompanying

increase in drift time is observed and compared between hLa and hLaΔNRE with and without U10

RNA. Three major protein conformations with two stages of unfolding were observed. In hLa, the

native conformation remained intact until 35 V, were a distinct transition state is observed and

many conformations begin to arise leading to the 1st unfolded state. At 45 V a single, unfolded

conformation is observed which is maintained until 95 V where another transition begins, and a

small population of a more unfolded species became apparent. Likewise, in hLaΔNRE, at 35 V

there was a transition to the 1st unfolded state, however, there was an almost immediate transition

to the 2nd unfolded state, unlike what was observed in hLa. This data suggests a significant

reduction in the stability of hLa in the absence of the NRE.
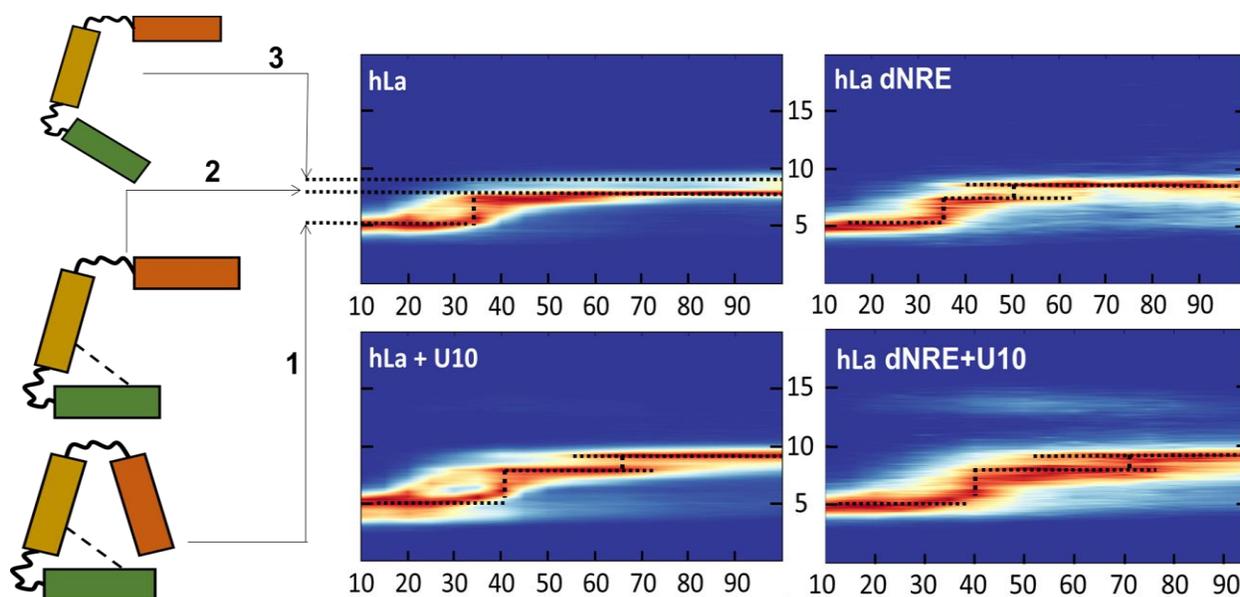


**Figure 3.4. CIU experiments on WT hLa (A), hLaΔNRE (B), WT hLa (C) bound to U10, and hLaΔNRE bound to U10 (D), where the unfolding and conformational changes of hLa are tracked over increasing collision energy**. The left panel shows a representation of the most likely structural conformations that populate each unfolding state. The La motif, RRM1, and RRM2 are represented by the red, yellow, and green squares, respectively, in which the proposed interdomain interaction between RRM1 and RRM2 is represented by the dashed line. Reproduced with permission from Marrella *et al*., 2019. [205]

To determine if RNA binding would have an effect on the stability of hLa, a U10 RNA ligand was

used as it had the most stable gas phase structure. When bound to hLa, an increase in stability was

observed between the folded state and the 1$^{st}$ unfolded state which is characterized by an increase in the collision energy of the 1$^{st}$ transition (from 35 V to 40 V). Interestingly, there was a decrease in stability when transitioning between the 1$^{st}$ and 2$^{nd}$ unfolded states, as there was a drastic decrease in collision energy (from 95 V to 55 V) required for this transition. When U10 was allowed to bind to hLaΔNRE, there was an increase in stability in both transition states, unlike what was observed for hLa. This CIU model would suggest that the 1$^{st}$ unfolded state is the unfolding of the La motif, as this unfolded state is stabilized by U10 in the WThLa. The 2$^{nd}$ unfolded state would likely be the disruption of the RRM1-RRM2 interaction in the WT which is not stabilized by U10 binding as poly(U) trailer does not interact with the CTD. This provided some evidence for the potential of an interdomain interaction but to gain further insight we employed TRESI-HDX experiments to get a more detailed look at the structural dynamics of RNA binding to hLa.

### 3.4.3 Time-Resolved Hydrogen Deuterium Exchange on the hLaΔNRE mutant

Although there is no binding expected in the CTD, RNA binding at the NTD could have large effects on the dynamics in the CTD. HDX-MS was used to shed some light on what was observed in the CIU experiments previously mentioned but unlike IMS-MS, HDX studies solution phase structures. In these experiments, the deuterium uptake in the free protein was compared to the deuterium uptake in the protein/RNA complex to determine the difference in relative deuterium levels of each peptide post RNA binding. The changes observed are mapped on to the PDB structures, where blue and red represent significant decreases and increases in deuterium uptake respectively in the protein/RNA complex.
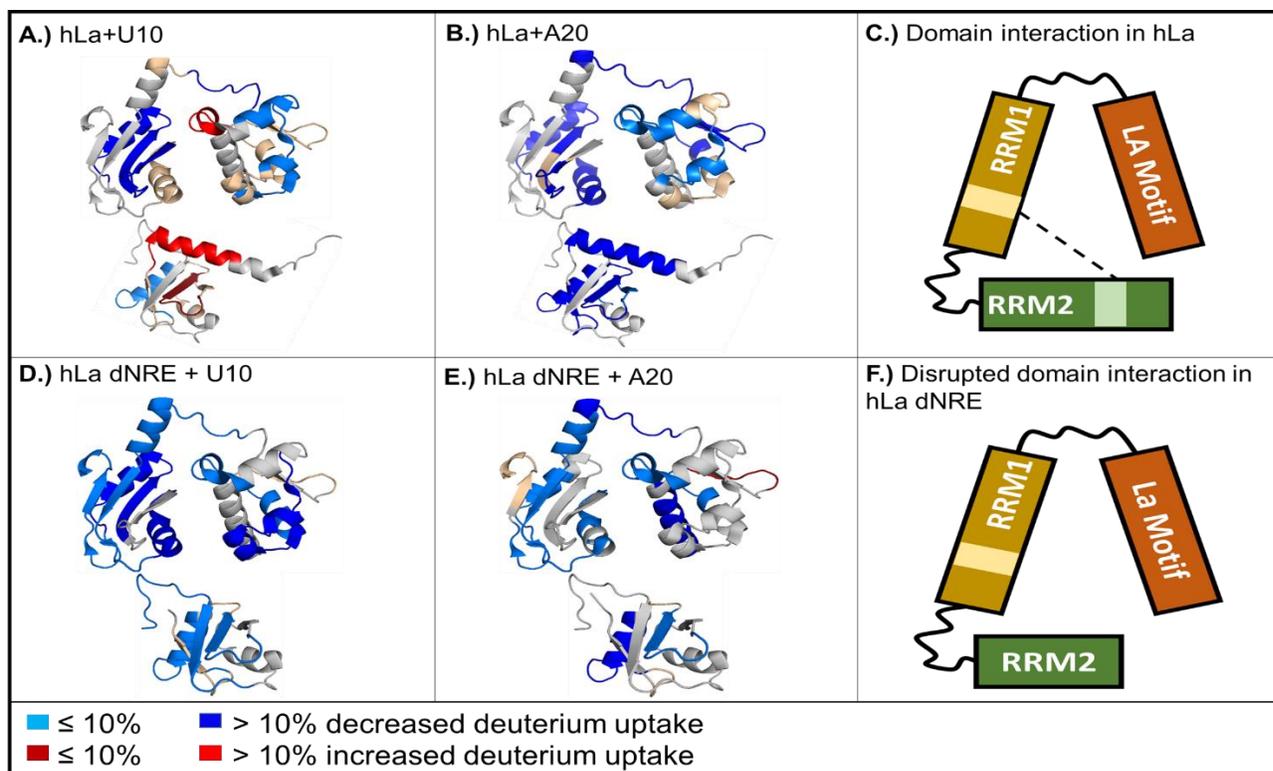
**Figure 3.5. Differences in deuterium uptake mapped onto the structures obtained from PDB (codes 1OWX and 1YTY).** Changes in deuterium uptake are in the WT hLa U10 (A) and A20 complexes (B), in which C is a representation of the likely domain organization prior to RNA binding. Changes in deuterium uptake are in the hLaΔNRE U10 (D) and A20 complexes (E), in which F is the proposed domain organization prior to RNA binding. The areas of the protein with decreased, increased, or no change in deuterium uptake are highlighted in blue, red, and tan, respectively. The areas colored gray mean there were no observed peptides from that area. Reproduced with permission from Marrella *et al*., 2019. [205]

When hLa binds to a U10 RNA ligand, significant decreases in deuterium uptake are observed in the NTD where poly(U) RNA is known to bind from previous studies. Interestingly, the CTD revealed significant increases in deuterium levels in the NRE helix when U10 is bound. In most HDX experiments, when a ligand binds to a protein, decreases in deuterium levels are observed in the complex. It is far less common to observe increases in deuterium levels in a stable complex as increased deuterium uptake represent a decreased in stability as the hydrogen bonding network becomes weaker which allows for more exchange with deuterium. If the proposed interaction

between the NRE helix in the CTD and the NES in the NTD holds true, when U10 engages hLa, this interaction would likely be broken to allow for U10 to bind to the NTD. This broken interaction would likely weaken the hydrogen bonding network in the NRE helix resulting in the decreased stability that is observed. The NES however, showed no significant changes in deuterium levels in the complex which could be attributed to the fast mixing time (700 ms) used in this experiment or the increased dynamics of the helix could be offset by RNA binding. When bound to A20, hLa showed decreased deuterium levels across the protein, most notably at the NRE helix and the NES. This is opposite of what was observed when hLa was bound to U10. This difference is likely due to the fact that A20 makes contact with all three domains of hLa, so even if the interaction between NRE helix and the NES is broken, the A20 binds and prevents any destabilization in the hydrogen bonding network that was observed when U10 binds. When hLaΔNRE binds to U10 and A20 the results are almost indistinguishable. The destabilization observed in the hLa CTD when bound to U10 is not observed in hLaΔNRE. Instead there was an overall decrease in deuterium uptake in the hLaΔNRE/U10 and A20 complexes with no significant differences observed between the two complexes. This further supports the notion of the interdomain interaction, as deletion of the NRE helix resulted in almost identical HDX profiles for both U10 and A20 complexes.

**Time-Resolved Hydrogen Deuterium Exchange on the point mutants hLaKKEE and hLaEE**

When hLa K316E/K317E was bound to U10, there was a deviation from the hLa-U10 complex observed. There was an overall decrease in the deuterium levels in the complex similar to what was observed for the hLaΔNRE-U10 complex. So, although the interdomain interaction would be broken, U10 would likely still interact with the NTD. However, with the lack of the interdomain interaction, the dynamics of the CTD completely changed. However, when hLa K316E/K317E was bound to A20, there were not many significant changes in deuterium level observed and there

76

were significant increases in deuterium uptake in the NRE helix. If A20 binds all three domains and the 316K/317K is a potential RNA binding surface, when altered it likely reformed the way in which A20 binds if it can no longer bind to that helix.
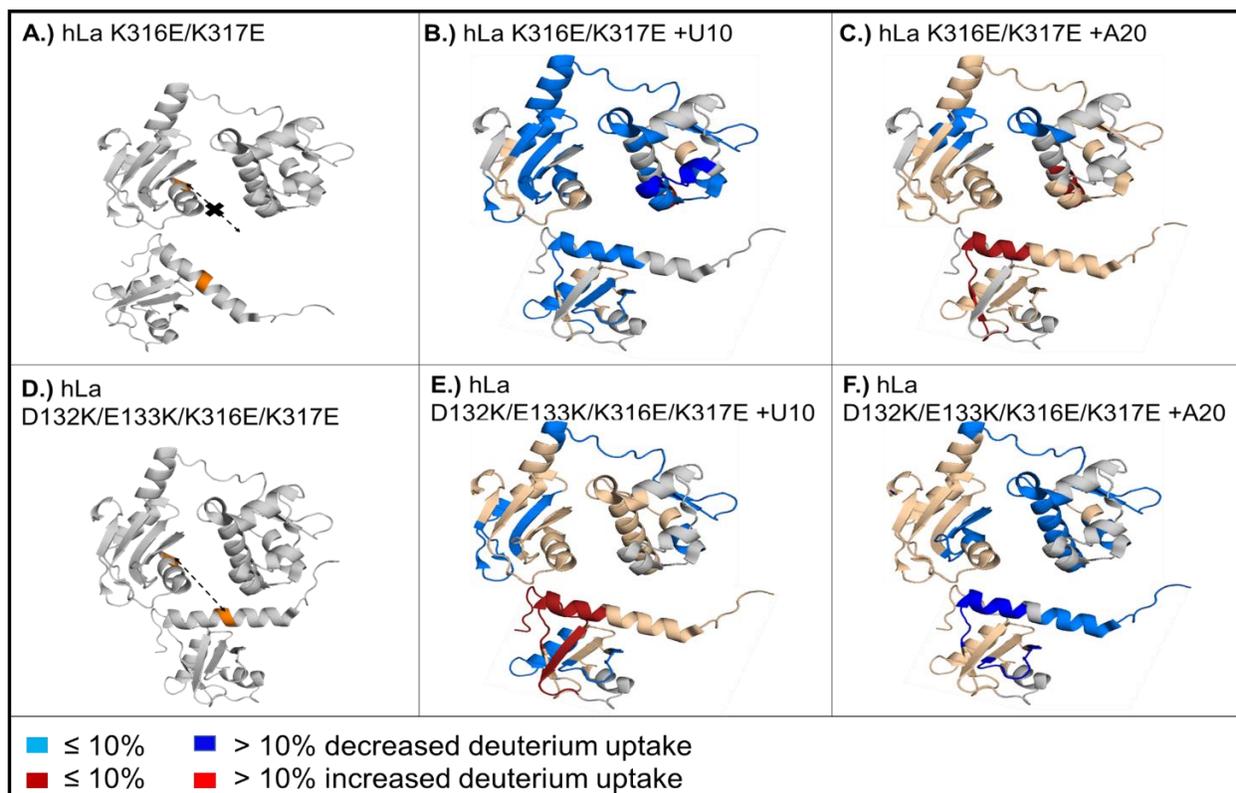


**Figure 3.6 Differences in deuterium uptake mapped onto the structures obtained from PDB (codes 1OWX and 1YTY).** (A) shows the hypothesized disrupted interdomain interaction in hLa K316E/K317E in which the key residues are highlighted in orange. Changes in deuterium uptake are in the hLa K316E/K317E U10 (B) and A20 complexes (C).(D)shows the re-establishment of the inter-domain interaction in hLa D132K/E133K/K316E/K317E in which the key residues are highlighted in orange. Changes in deuterium uptake are in the hLa D132K/E133K/K316E/K317E U10 (E) and A20 complexes (F). B, C, E, and F, areas of the protein with decreased, increased, or no change in deuterium uptake are highlighted in blue, red and tan, respectively. The areas colored gray means there were no observed peptides from that area. Reproduced with permission from Marrella *et al.*, 2019. [205]

Unlike the hLaΔNRE mutant that showed many significant decreases, with the removal of the entire NRE, this exposed the typical RNA binding site in which A20 could possibly engage. When hLa E132K/D133K/K316E/K317E was bound to U10 and A20, the results were very similar to the hLa/U10 and A20 complexes suggesting the proposed interaction between the NRE and NES was restored. The increased dynamics in the NRE helix was observed when bound to U10 and the decreased dynamics in the NRE helix and NES was observed when bound to A20. This data supports the concept of an inter-domain interaction between RRM1 and RRM2.

## 3.5 Conclusion

Although MS is a low-resolution structural technique, it has provided a wealth of information regarding RNA interactions in hLa and the effects of altering the nuclear trafficking elements. hLa is a highly flexible protein and it has been hypothesized that the structured domains within the protein are not in a fixed orientation with respect to each other.[201] This flexibility has made it difficult to determine RNA binding modes using more high-resolution structural techniques. This structural flexibility however has been hypothesized to be important for the capacity of hLa to bind a variety of RNA substrates. Binding to UUU-3'OH end of nascent RNA polymerase III transcripts has been extensively studied and high-resolution structures of the RNA-hLa complexes are available.[117,200,201] hLa has also been implicated in the binding of poly(A) sequences [116] found on mRNAs as well as binding to more structured RNAs such as stem loops[161,178] and the main body of pre-tRNAs.

CIU and TRESI-HDX–MS based analysis provided evidence in which the NRE and the α1 helix region of RRM1, which includes amino acids Glu-132 and Asp-133, make a direct protein–protein interaction that either breaks or forms depending on the identity of the bound RNA ligand. The bridge is thought to be broken upon binding to U10 and intact when bound to A20 in the wild type.

The hLa K316E/K317E mutant that disrupts the interdomain interaction results in change in the dynamics of the RRM2 with respect to the WT when bound to U10. This further strengthens the hypothesis of the interdomain interaction. It also shows that the NRE is likely a crucial binding surface for poly(A) RNA as there was very little decreases in deuterium uptake (representative of very few possible contacts) when bound to A20 in the hLa K316E/K317E mutant. The compensatory mutant hLa E132K/D133K/K316E/K317E further proves the proposed interaction between NRE and NES as the interaction is likely restored due to the similar results that were observed in the wild type. Although this data is consistent with electrophoretic mobility shift assays (EMSAs) provided by the Bayfield Lab showing RNA binding patterns of the WT hLa, hLa ΔNRE, hLa E132K/D133K/K316E/K317E and hLa K316E/K317E bound to various RNA substrates, high resolution structures are required to get a firmer grasp on the nature of the interdomain interaction.

# Chapter 4: Rapid Characterization of Structural and Functional Similarity for a Candidate Bevacizumab (Avastin) Biosimilar using a Multipronged Mass Spectrometry-Based Approach

**Brown, K.A**., Rajendran, S., Dowd, J. and Wilson, D.J.

## 4.1 Summary

The ongoing shift from small molecule drugs to protein therapeutics in the pharmaceuticals industry presents a considerable challenge to 'generic' drug developers who are increasingly required to demonstrate biosimilarity for biological macromolecules; a task that is decidedly more complex than doing the same for small molecule drugs. In this work, we demonstrate a multipronged mass spectrometry-based workflow that allows rapid and facile molecular characterization of antibody-based protein therapeutics, applied to biosimilars development. Specifically, we use a combination of native mass spectrometry (MS), ion mobility spectrometry (IMS) and global time-resolved electrospray ionization (TRESI) hydrogen deuterium exchange (HDX) to provide an unambiguous assessment of the structural, dynamic and chemical similarity between Avastin (bevacizumab) and a biosimilar in the late stages of pre-clinical development. Minor differences between the biosimilar and Avastin, and between lots of the biosimilar, were tested for functional relevance using Surface Plasmon Resonance (SPR) -derived kinetic and equilibrium binding parameters. Our results revealed minor differences in glycosylation between Avastin and its biosimilar which translated to statistically significant differences in structural dynamics observed in global HDX. However, no significant differences were observed using IMS. Differences, however, were minor and had no effect on target engagement as determined by SPR analysis.

## 4.2 Introduction

Protein therapeutics, and in particular monoclonal antibodies (mAbs) are now widely used in the treatment of cancer and other illnesses owing to the high selectivity and sensitivity to their respective targets.[206,207] However, therapeutic mAbs can be tremendously expensive, partly due to the high-cost, high-precision processes associated with consistent production of therapeutic

proteins, but also due to a lack of market alternatives.[207] Increased availability of biosimilar antibodies might therefore significantly reduce the costs of therapeutic mAbs, some of which are currently charged at more than $200,000 per year per patient.[208] In contrast to conventional small molecule drugs, the complex correlation between biomacromolecular structure, dynamics and biological activity heightens the challenge of unambiguously demonstrating molecular and functional biosimilarity in therapeutic mAbs.[209]

Differences between a mAb biosimilar and its reference product stem from the fact they are invariably generated by different manufacturing processes, usually from different host cell clones. In addition to formulation issues, varied mAb manufacturing processes can result in distinct post-translational modifications (PTMs) (*e.g.*, glycosylation, glycation *etc.*) as well as other minor variations in structure that could impact potency, clearance and immunogenicity and thus clinical efficacy. Such differences can also have a substantial effect on degradative processes affecting shelf-life, such as susceptibility to oxidation, deamidation and aggregation.[210–214] There are many analytical methods that can be used to assess the similarity between biosimilars and their reference products in the early stages of development. These methods can be classified according to quality attributes that they measure such as structural characterization (primary and higher order), purity (product and manufacturing process related), assessment of glycan profile and biological activity (*e.g.*, binding assays).[215]

In this report, we focus on the comparison of multiple physical attributes between bevacizumab, an FDA-approved therapeutic mAb marketed as Avastin, and a candidate biosimilar using a unique mass spectrometry-based workflow. Bevacizumab is a humanized IgG1 monoclonal antibody developed as a target to the Vascular Endothelial Growth Factor (VEGF).[216,217] It binds and neutralizes all isoforms of VEGF-A preventing tumor angiogenesis and progression.[218] It is used

along with chemotherapy in the treatment of metastatic colorectal cancer,[219] metastatic non-small-cell lung cancer,[220] metastatic renal cell carcinoma,[221] ovarian cancer,[222] and metastatic breast cancer.[223]

Mass spectrometry instrumentation has evolved over the last two decades, providing ever-increasing sensitivity (low picomolar sensitivity is now routine), resolution and mass accuracy. Additional innovations in coupled separation techniques, particularly liquid chromatography (LC-MS) have made MS an exceedingly powerful tool for peptide-level analyses of mAbs, which allows for sequencing and PTM analysis.[211,215,224] Arguably less well-recognized is the rapidly developing capability of MS instruments to analyze large, intact biomolecules under native-like conditions.[2,28,225–228] This, combined with coupled techniques like ion mobility spectrometry (IMS)[5,229] and hydrogen/deuterium exchange (HDX) [68,230,231] allows mass spectrometry to support higher order structure analysis, in some cases at single residue resolution.[232,233]

IMS is a relatively recent addition to the MS-based analysis toolbox, with a number of commercial instruments now including IMS technologies. IMS is essentially a gas-phase analogue of electrophoresis, allowing for (usually pre-mass spectrometric) separation of analytes on the basis of size, shape and charge.[234] IM-MS techniques such as collision induced unfolding (CIU) can easily be used to differentiate between the sub-classes of IgG antibodies.[235] In principle, this approach could be used to identify even relatively subtle structural heterogeneity in mAb drug candidates, including conformational differences that arise from variances in PTM profiles (*e.g.*, differential glycosylation in mAbs).[236] IM-MS also has the sensitivity to monitor domain exchanges in antibodies.[237]

The argument can always be made, however, that the gas-phase conformational properties monitored by IMS do not adequately reflect conformational properties in solution. To allay these

concerns, IMS analyses can be complemented by global hydrogen deuterium exchange (HDX), which provides similar information (essentially a measure of relative global conformational stability and heterogeneity) based on isotope labeling in solution.[232,238] In this work, we used a variation of HDX that involves short (ms timescale) isotope labeling times, called 'time-resolved' HDX. Time resolved HDX has been proven to be beneficial in the analysis of disordered proteins[239] or the weakly structured regions of proteins[178,240] and, most importantly for the present work, highly sensitive to subtle changes in conformational dynamics.[241] The time-resolved set up (Figure 4.1) is more straightforward than conventional LC-based systems as the mixer also acts as the electrospray capillary, eliminating the need for a quenching buffer and a trap column.[68]

Together, native MS, IMS and global HDX provide a convenient, integrated platform for multiple attribute monitoring and early-stage characterization of candidate protein therapeutics. Here, we demonstrate this approach for the rapid and facile comparison of biosimilars to reference products and in lot-to-lot variability analysis. Physical differences revealed by this approach are then correlated to changes in function *via* Surface Plasmon Resonance (SPR) measurements.
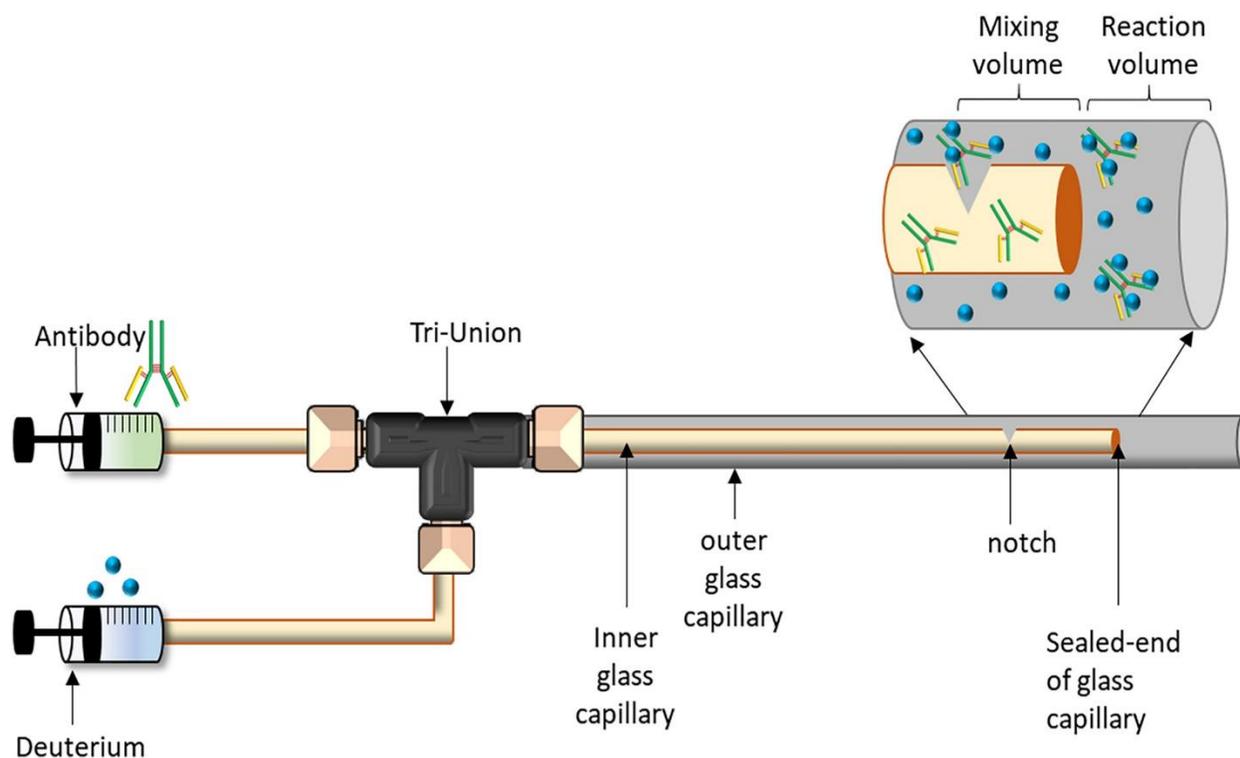
**Figure 4.1 Capillary mixer used for time-resolved electrospray ionization hydrogen deuterium exchange mass spectrometry (TRESI−HDX−MS)**. A glass capillary carrying the antibody is inserted into a metal capillary through which deuterium is passed. The distal end of the inner glass capillary is sealed, and a notch is cut 2 mm from the seal to allow the antibody to escape into the narrow annular space between the two capillaries (where mixing takes place). The HDX labeling time is adjusted by pulling back the inner glass capillary within the outer capillary thereby increasing the reaction volume. Reproduced with permission from Brown *et al*., 2019. [242]

## 4.3 Methods and Materials

### 4.3.1. Materials

Bevacizumab (Avastin, 149.2 kDa) and all lots of the candidate biosimilars (149.4 kDa) were provided by Apobiologix (North York, ON). Deuterium oxide ($D_2O$, 99.9%, 151882) and high purity acetic acid (>99.7%, 695092) were purchased from Sigma-Aldrich (Oakville, ON). HPLC-grade ammonium acetate and water were purchased from Fisher Chemical (Mississauga, ON).

### 4.3.2. Native and Ion Mobility Mass Spectrometry

Prior to analysis, each antibody was diluted to 1 µM and buffer exchanged into 250 mM ammonium acetate solution, pH 7.4 using zeba spin desalting columns (7 MWCO) (89890, 7K, Thermo) on Allegra 25R centrifuge (Beckman, Mississauga, ON) by centrifugation at 1000 g/min at 4 °C. The antibodies were subjected to electrospray ionization (ESI) Mass Spectrometry (MS) using a Synapt G2-S mass spectrometer (Waters, Milford, MA) at a flow rate of 10 µL/min. The tuning parameters were set to optimize resolution while preventing gas phase unfolding or fragmentation. The tuning parameters were set as follows: the capillary voltage was set to 3.0 kV with a source temperature of 120°C and a sampling cone voltage at 100 V. To enable desolvation, the desolvation temperature was set to 250°C with the gas flow set to 800 L/Hr. The collision energy in the trap and transfer cell was set to 15 V and 10 V respectively with a gas flow of 2.5 mL/min in the trap cell. To better enhance ion mobility separation the helium and IMS gas flow was set to 180 mL/min and 70 mL/min respectively. The IMS wave velocity and wave height was adjusted to 800 m/s and 40 V respectively. The native and IMS data were analyzed using the MassLynx software package, v. 4.1.

### 4.3.3 Global Time-Resolved Hydrogen Deuterium Exchange Mass Spectrometry

The capillary mixing device (Figure 4.1) was used for time resolved global HDX measurements to provide information on the dynamics of the antibodies on the millisecond to second timescale. The capillary mixer was interfaced directly with the Synapt G2-S mass spectrometer using an in-house built platform on the nano-ESI source. 10 µM of Bevacizumab flowed through the inner capillary at a flow rate of 4 µL/min and 99.9% deuterium oxide flowed through the outer capillary at 16 µL/min. This resulted in mixing times of 0.146, 0.218, 0.29, 0.368, 0.44, 0.518, 0.89, 1.63 and 3.87 seconds to generate a deuterium uptake profile for each antibody, with three technical

replicates per lot. The deuterium uptake was calculated as the change in mass ($\Delta M$) using the equation:

$$\Delta M = M_D - M_{UND} \tag{4.1}$$

Where $M_D$ is the mass of the deuterated protein and $M_{UND}$ is the mass of the undeuterated protein. The deuterium uptake curves were fit in SigmaPlot 14.0 using the double exponential with 5 parameters function.

### 4.3.4 pH-Stability Test

The antibodies were buffer exchanged into 250 mM ammonium acetate with pH 3.5, 5.5 and 7.4. The antibodies were allowed to incubate in the buffer solutions for 5 mins, 5 days and 5 weeks at 4°C after which they were subjected to ESI-IMS-MS as described above. Drift time heat maps were generated using the Protein Unfolding for Ligand Stabilization and Ranking (PULSAR) software.[197]

### 4.3.5 VEGF Binding by SPR

The SPR VEGF binding assay was carried out on the Biacore platform (GE Healthcare). Briefly, Avastin or a biosimilar is immobilized on a Protein A-coated sensor chip, together with a relevant standard (Avastin or a validated biosimilar internal standard) in a separate flow channel, so that the Fab regions are oriented toward the bulk solution. In this orientation, VEGF binding occurs independently at each of the Fab binding sites, which allows data fitting using a 1:1 Langmuir model[243] that incorporates mass transfer.[244] VEGF is then passed over the anti-VEGF mAb-functionalized surfaces (sample and control), allowing association and dissociation processes to be monitored ultimately *via* the change refractive index (and thus the resonance frequency of the surface plasmon) that occurs when the target molecule in the bulk-phase is adsorbed onto the

functionalized surface. The association rate constant ($k_{on}$) and the dissociation constant ($k_{off}$) are estimated from fits to the data generated over several different VEGF concentrations (0.8 nM to 25 nM), ultimately allowing for calculation of the equilibrium dissociation constant $K_D$ for the complex.

## 4.4 Results

### 4.4.1 Native Mass Spectrometry of Avastin and Biosimilar

Native MS was used to determine the molecular weight of Avastin and its biosimilar. The native mass spectrum showed identical charge distributions (Figure 4.2A and 4.2B) with Avastin having an average molecular weight of 149.2 kDa and an average molecular weight of 149.4 kDa for the biosimilar. A key difference observed in the native spectrum is a small but significant population of a lower molecular weight species in the Avastin reference, which was absent in the biosimilar (discussed below). Although the native spectra look broadly similar, closer inspection of individual charge states reveals substantially different peak patterns, which can be attributed to differing glycoform distributions (Figure 4.2C and 4.2D). The resolution of individual glycosylation state peaks is far from what can be achieved with extensive sample preparation and higher resolution instrumentation,[245,246] however, these data demonstrate that even with minimal optimization, it is possible to assign a number of the observed sub-peaks to known common glycosylation states of IgG1. For Avastin, the first major peak most likely represents the antibody with two 'G0F' glycosylations (Figure 4.2E). This also suggests that the low intensity degradation product observed in Avastin (Figure 4.2A, immediately left of each main charge peak), representing a loss of 1447 Da, corresponds to a monoglycosylated form with a single G0F modification.[211,247]

The two peaks following the main peak represent additions of ~58 Da and ~116 Da (Figure 4.2C), which we attribute to the adduction of one and two acetate ions, respectively. Covalent acetylation

is also a possibility but lacks a plausible mechanism in the manufacturing and sample handling processes used and has not previously been observed for Avastin. Glycation has been proposed to occur in Avastin[248], however, we see no direct evidence for this phenomenon in the data (*i.e.*, it may account for some of the complexity in the peak distribution, but we cannot resolve individual peaks that would correspond to glycation). The next common glycosylation, G1F, involves the addition of a galactose to one of the N-acetylglucosamine termini (Figure 4.2E). The glycoform G0F/G1F is observed as a shoulder in the Avastin sample and is the most populated configuration in the biosimilar (Figure 4.2C and 4.2D). A population with two G1F glycosylations (G1F/G1F) and/or the addition of galactose to both branches of one G0F glycosyl chain (G0F/G2F) is also observed as a significant population in the biosimilar and as part of the peak tail in Avastin (Figure 4.2C and 4.2D). Additional glycoforms (*e.g*., G1F/G2F, G2F/G2F) were clearly also present as minor populations in the peak tails, with a higher proportion of heavier 'mature' species in the biosimilar. To test whether these differences had a substantial impact on the global physical properties of the antibodies, *i.e.*, on structure or conformational dynamics, we conducted a series of IMS-MS and global-HDX based analyses.
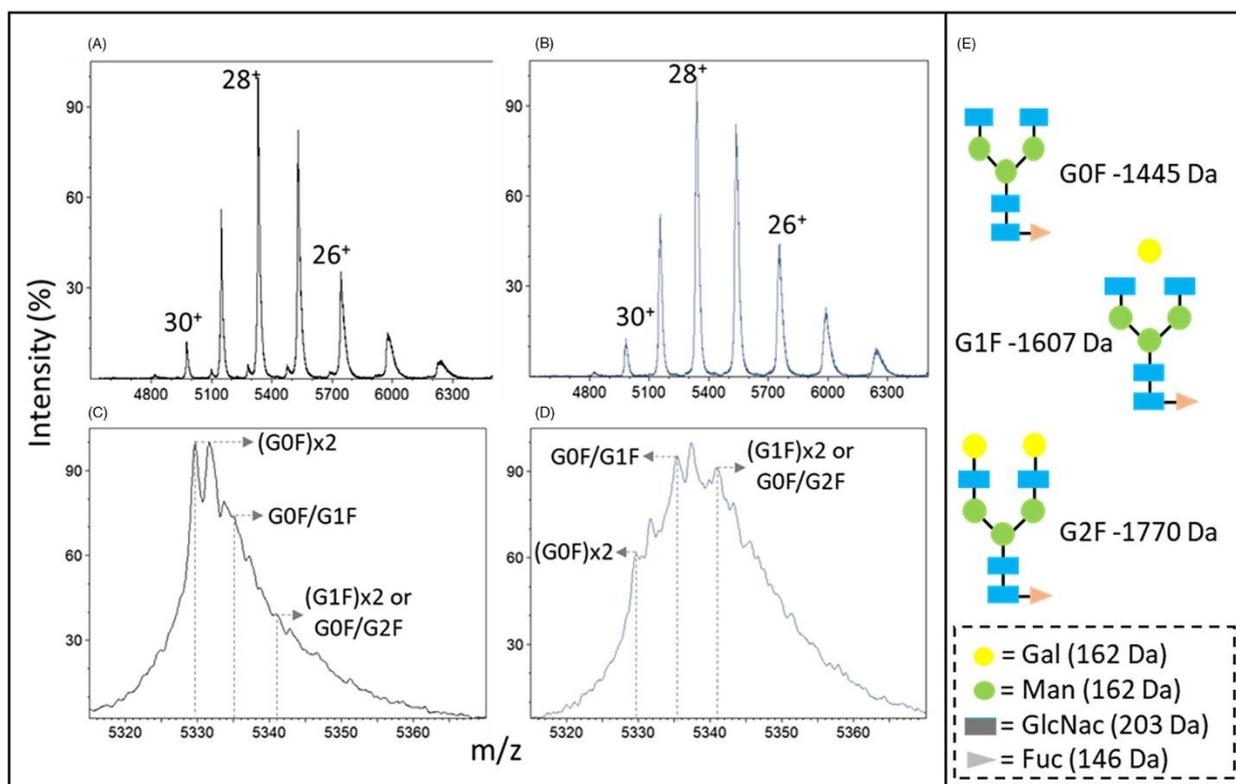
**Figure 4.2 Native mass spectrum of (A) Avastin and (B) biosimilar D** revealing the same charge state distribution for both antibodies and a small population of a lower mass species correlating to a monoglycosylated form of the antibody in the Avastin sample. A close-up view of the $28^+$ charge state peaks revealing the glycosylation patterns of (C) Avastin and (D) the biosimilar. Panel (E) shows the composition of the main glycoforms thought to be present, G0F, G1F, and G2F. The theoretical mass of bevacizumab is 149 kDa which is in line with the measured mass of 149.2 and 149.4 kDa for Avastin and its biosimilar, respectively. Reproduced with permission from Brown *et al*., 2019. [242]

### 4.4.2 Global Dynamics of Avastin and Biosimilar

IMS separates gas phase ions based on their charge and collision cross section, which is associated with the time required for ions to traverse the ion mobility cell (drift time). Essentially, ions with smaller collisional cross sections will have shorter drift times than ions with the same charge and larger collisional cross sections.[32,34,249] Our IMS data shows that Avastin and the biosimilar have statistically identical average drift times (15.94 ± 0.19 ms and 15.85 ± 0.18 ms, respectively) and

Full Width Half Maximum (FWHM) peak widths ($2.93 \pm 0.07$ and $2.98 \pm 0.08$ ms, respectively) suggesting very similar cross sections and gas phase conformational stability (Figure 4.3A). This is as expected, since the observed differences in glycosylation (described above) are unlikely to significantly impact protein structure, nor substantially change the cross section in themselves. In both cases, a single conformation was observed, indicating that neither Avastin nor the biosimilar have alternate structures that are significantly populated in the 'native state'. Such structural intermediates can play a role in protein stability and, in some cases, biological activity.[250,251]

IMS analyses can provide critical information on global protein structure and dynamics, but it can always be argued that gas-phase structure may not correlate to structure in solution. Global HDX provides a complementary view of protein structure and dynamics based on labeling in solution and in a similarly straightforward experiment. In the current study, we use millisecond global HDX, which provides added sensitivity to dynamic differences in weakly structured regions of proteins.[68,252] The initial deuterium uptake at the first time point, referred to as the burst phase, corresponds to all exchangeable side chains and fast exchanging backbone amides. Avastin had an initial uptake of $786 \pm 35$ Da while the biosimilar had an initial uptake of $711 \pm 9$ Da (Figure 4.3B). Based on the difference in the sums of the first 7 $\Delta$mass values, where weakly structured regions of the protein dominate exchange, there is a small but statistically significant difference in deuterium uptake ($> 2\sigma$, $n = 6$), between the reference and the biosimilar, with lower uptake in the biosimilar. This agrees with the very small IMS drift time difference noted above, but it is not clear that such small differences could be measured for biological replicates (error estimates for the current measurements are from technical replicates). At 4 seconds, corresponding to the end of the burst phase, Avastin and the biosimilar had statistically identical maximum uptakes of $1263 \pm 35$ Da and $1294 \pm 45$ Da, respectively. To test biological replicability and to explore the effects of

different manufacturing processes on biosimilar samples, we carried out the same IMS/Global

HDX-based in a batch-to-batch variability assay.
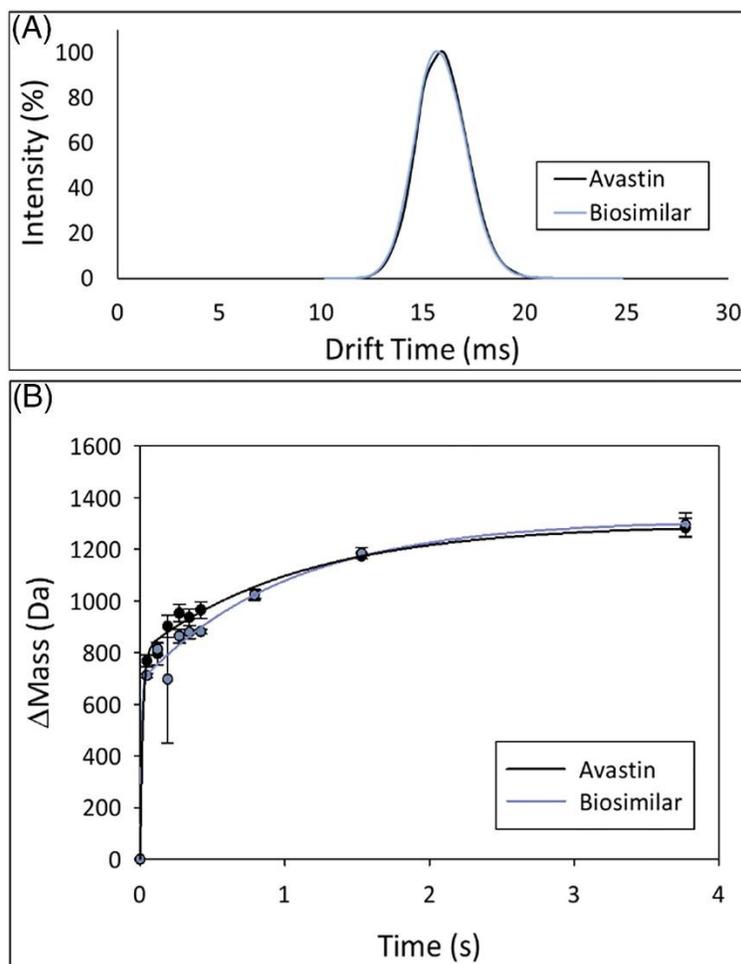


**Figure 4.3 (A) Ion mobility drift time chromatogram of charge state 28+ of Avastin and biosimilar D** revealing almost identical drift times of $15.94 \pm 0.19$ ms and $15.85 \pm 0.18$ ms respectively. (B) Global HDX of Avastin and biosimilar revealing very similar solution dynamics. Error bars from technical triplicates. Reproduced with permission from Brown *et al.*, 2019. [242]

### 4.4.3 Batch-to-Batch Variability of Biosimilar

Five different batches of the biosimilar (labelled biosimilar A-E), each manufactured in a different

run, were provided by Apobiologix. Native MS revealed identical spectra for all the biosimilars

(Figure 4.4A), with no occurrence of the monoglycosylated species. This is consistent with the

generally more mature glycosylation of the biosimilar. In analyzing a single charge state to get an overall view of the glycosylation patterns (Figure 4.4B), it was observed that the biosimilar batches were, as designed, significantly more mature in glycosylation than the Avastin reference and highly self-similar, indicating that consistent run-to-run manufacturing had largely been achieved. However, minor differences can be noted; for instance, there is a clear order of lots with respect to maturity of glycosylation. Using G0F/G0F as an indicator of low maturity glycosylation, the lot order is Avastin >> C > D > A ≈ E > B. For the biosimilar lots, this order is maintained for the G0F/G1F modification, and as one might expect, it is essentially reversed for higher maturity glycoforms.
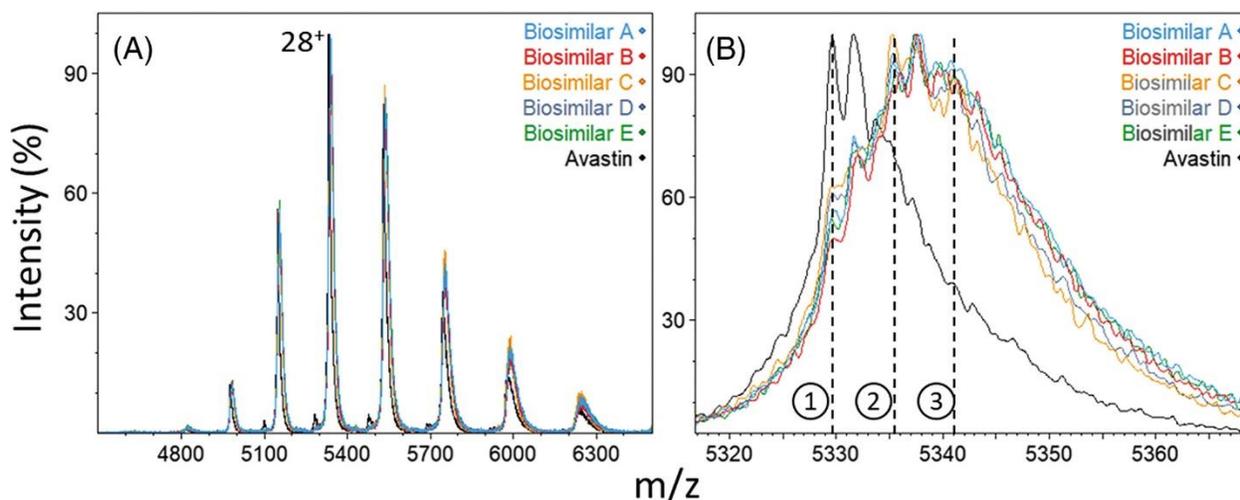


**Figure 4.4 (A) Native mass spectrum of the five biosimilar lots and Avastin**. (B) A close-up look at charge state 28+ revealing the glycosylation patterns where 1 represents (G0F)x2, 2 represents G0F/G1F and 3 represents (G1F)x2 or G0F/G2F. Reproduced with permission from Brown *et al*., 2019. [242]

To measure potential batch-to-batch differences in global conformation and dynamics in biosimilar lots, we again used IMS-MS and global HDX. For each lot, a single IMS peak was observed indicating no significantly populated alternate conformations in the gas phase (Figure 4.5A). In

terms of the average IMS drift time, all lots were identical to the avastin reference to within experimental error. This demonstrates that, at least in a coarse-grained sense, structual similarity to the reference product and consistent batch-to-batch manufacturing had been achieved.
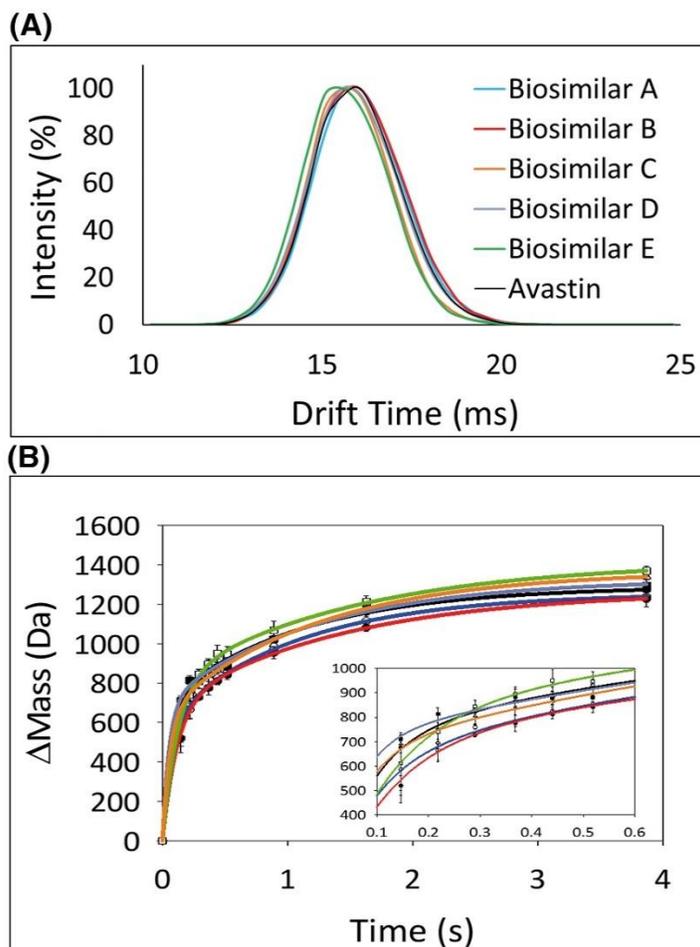


**Figure 4.5 Ion mobility and global HDX comparison of all the biosimilar lots and Avastin**. (A) Drift time chromatograms of all the biosimilar batches and Avastin. (B) Deuterium uptake kinetics plots from global HDX (with $n = 3$) of all biosimilar batches compared to an Avastin reference. The inset shows the burst phase (early time points) representative of the fast exchanging amides. Reproduced with permission from Brown *et al*., 2019. [242]

Global HDX is essentially a measure of the strength of the hydrogen-bonding network that maintains secondary and tertiary structure in 'native' proteins, averaged over the molecule. This measurement, and particularly the millisecond measurement used here, is expected to be more

sensitive to subtle changes in structure and dynamics than IMS (where reliable difference measurements require a substantial change in collisional cross section). Our millisecond HDX analysis revealed small but statistically significant differences in the 'burst phase' deuterium uptake between lots, corresponding to an order of D > Avastin > C > E > A > B (Figure 4.5B). At the end of the second deuterium uptake phase (4 s), which is associated exclusively with backbone amides, the uptake order was: E > C > D > Avastin > A $\approx$ B. What is striking about the current data is therefore the high degree to which the global conformational and dynamic properties of Avastin and the biosimilar are similar, despite arising from separate manufacturing runs.

### 4.4.4 pH-Stability test of Biosimilar

A pH-stability test was used to compare the stability of Avastin and the biosimilar (lot E) at three pHs corresponding to 7.4, 5.5 and 3.5. These pHs were selected to mimic the pH conditions to which the mAbs are exposed at various points in the manufacturing process. The antibodies were incubated over a five-week period at 4°C and analyzed at 5 mins, 5 days and 5 weeks and the *m/z vs*. drift time heat maps were compared at each time and for each pH (Figure 4.6). The antibodies were found to be stable at pH 7.4 and 5.5 throughout the five weeks, although at the end of the 5-week period, a very small (below the heat map threshold) population of aggregated antibody was observed at both pHs.

At pH 3.5 a few structural changes were observed over the time period studied. One interesting phenomenon was a shift in the 'folded' protein distribution to lower charge states (*i.e*., to the right) compared the higher pH data. A possible explanation for this observation is that the protein is undergoing deamidation, resulting in charge reduction, prior to unfolding and aggregation. Based on the magnitude of the shift (*i.e*., two charges on average) it is likely that deamidation is occurring to a significant degree at two positions. These deamidations occur within 5 min of exposure to pH

3.5 under our conditions, but clearly do not occur during manufacturing. Also, at pH 3.5, both Avastin and biosimilar samples exhibited a low population of highly charged, likely unfolded, protein at all timepoints (barely visible in the heat map to the left of the main peak distributions). By day 5, these 'unfolded' peaks were convoluted with a broad 'hump' at low $m/z$ (not visible in the heatmap until 5 weeks), which we attribute to aggregation. After 5 weeks at pH 3.5, the signal for the aggregated protein dominated the spectrum, although there remained a set of peaks corresponding to the unfolded monomer.
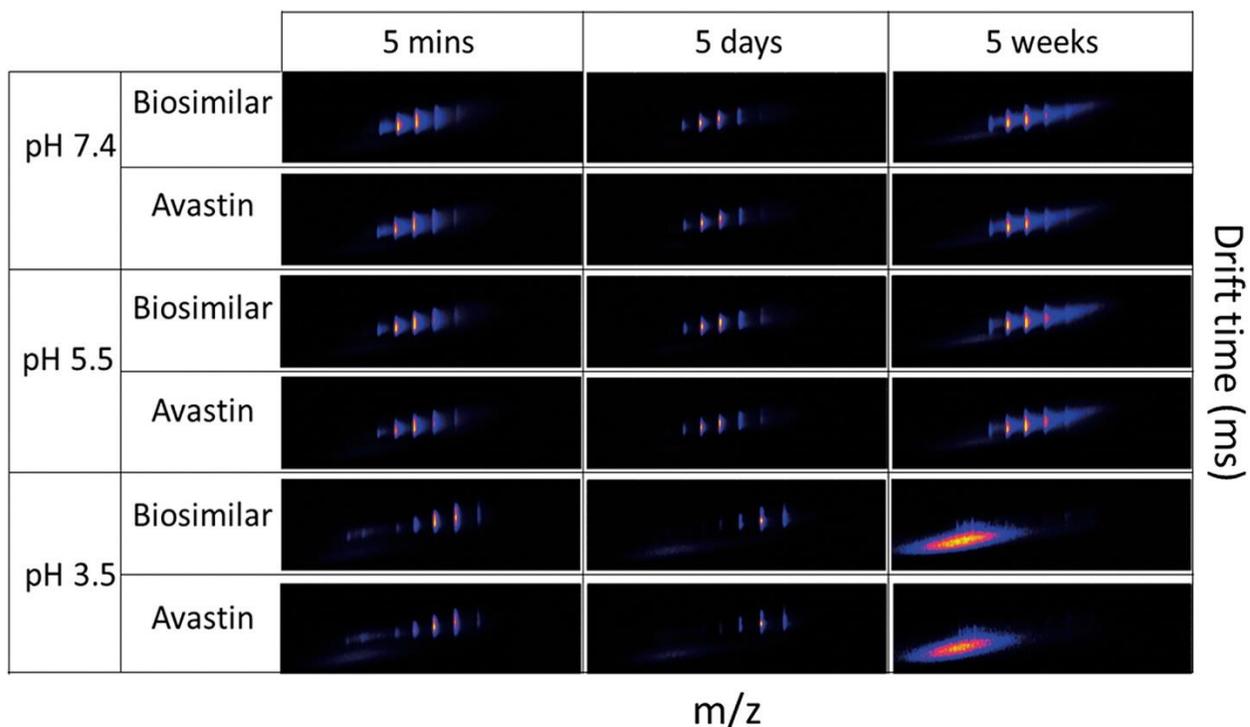


**Figure 4.6 Drift time vs. m/z heatmaps of Avastin and biosimilar at pH 7.4, 5.5, and 3.5 over a five-week period**. Reproduced with permission from Brown *et al*., 2019. [242]

### 4.4.5 VEGF Binding by SPR

To determine if the minor physical differences observed between Avastin and the Biosimilar could be correlated to functional changes *in vitro*, each batch was subjected to an SPR-based binding affinity assay. The results of this analysis, shown in Table 4.1, indicate a high degree of similarity in the kinetic parameters $k_{on}$ and $k_{off}$, as well as the dissociation constant $K_D$, between lots and with the Avastin reference.

Table 4.1. SPR analysis of Avastin and a biosimilar (lots A-E) binding to VEGF. Reproduced with permission from Brown *et al.*, 2019. [242]

| Lot | $k_{on}$ (M·s⁻¹) | $k_{off}$ (s⁻¹) | $K_D$ (nM) | $K_D$ (nM) for Assay Reference Std. | | |
|---|---|---|---|---|---|---|
| | | | | Avastin US | Avastin EU | Int. Std. |
| **A** | 1.80E+06 | 2.40E-04 | 0.133 | - | - | *0.126* |
| **B** | 5.83E+05 | 5.94E-05 | 0.102 | *0.128* | *0.086* | - |
| **C** | 1.50E+06 | 1.28E-04 | 0.086 | *0.103* | *0.095* | - |
| **D** | 1.48E+06 | 1.29E-04 | 0.087 | | | |
| **E** | 1.81E+06 | 2.38E-04 | 0.132 | - | - | *0.126* |
| **Avastin** | 2.04E+06 | 2.53E-04 | 0.124 | - | - | *0.123* |

The dissociation constants $K_D$ for the biosimilar batches were in the range of 0.086 to 0.133 nM, with a mean of 0.108 ± 0.023 nM. Avastin $K_D$ measurements exhibited similar dispersal, ranging from 0.086 to 0.128 nM with a mean of 0.105 ± 0.016 nM, with the Avastin reference used in the structural analyses exhibiting a $K_D$ of 0.124 nM. In keeping with International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) method validation standards (ICH Q2R1)[253], these assays were run in parallel with reference standards corresponding to Avastin from the US or EU market, or a biosimilar that has been qualified to be

an internal standard, in order to detect and control systematic errors. In this case, no significant systematic variations were observed.

**4.5 Discussion**

In the production of mAb biosimilars, significant differences from the reference product can arise from a host of factors, ranging from the cell lines used for production to subtle differences in the manufacturing process.[246,254,255] Establishing biosimilarity for protein therapeutics is a challenging and multifaceted process.[215] Although high resolution techniques such as Nuclear Magnetic Resonance (NMR)[256,257] and X-ray crystallography[256] are the methods of choice to obtain high resolution protein structures, these methods are not always feasible in the structural characterization of biosimilars.[209] NMR has an inherent analyte size limit that is usually substantially lower than the size of an intact antibody and it is not always possible to obtain high quality crystals for X-ray crystallography. Both approaches are typically quite slow and labor intensive, requiring extensive optimization and highly specialized equipment. For routine testing during development and production, approaches that are rapid, straightforward and yet sufficiently informative are needed.

MS-based analyses of intact proteins is increasingly used in the context of early-stage protein therapeutic development to unambiguously identify targets *via* accurate mass,[258] to detect oligomerization[259,260] and to characterize protein complexes.[261] Post translational modifications such as glycosylations can be identified on the intact protein, although determination of the positions of these modifications typically requires digestion.[211,246] While non-native LC-MS remains the preferred method for intact protein analysis, this approach requires sample handling under relatively harsh conditions, sometimes resulting in artifactual modifications in the protein

and, almost invariably, a complete loss of non-covalent interactions. For intact mAb samples in particular, commonly used LC-MS conditions, especially low-pH running buffers, can induce aggregation.[211,262] Native MS, in which proteins are electrosprayed under native-like conditions (usually 10 – 150 mM $NH_4Ac$ buffer at pH 7), provides characterizations of protein analytes that, while more technically challenging to achieve, are often more reflective of the protein in its biological context: Observed PTMs are more likely to reflect those originally present in the sample and the charge-state envelope provides a general picture of structural stability.[28] Native MS analysis have been successfully used to compare biosimilar antibodies, [263] analyze antibody drug conjugates,[264,265] and provide qualitative and quantitative analysis of antibody glycosylation profiles.[266]

In this work, we implemented a suite of mass spectrometry-based assays that allow rapid and highly informative evaluation of physical similarities between biosimilar candidates and originator products, which may ultimately serve to explain differences in function (or as in this case, lack thereof). Native MS was used to determine the mass of the antibodies and to highlight any differences that arise from PTMs. The difference in average mass between the antibodies was approximately 200 Da, which is attributable to the generally more mature glycosylation states exhibited by the biosimilar (Figure 4.2). Although the same glycosylations were observed to be present in both Avastin and the biosimilar, the distributions of glycoforms were substantially different (Figure 4.2C and 4.2D). Understanding these glycosylation profiles can be critical in downstream development, since such differences may ultimately result in differences in biological activity and/or immunogenicity of the biosimilar compared to the reference product.[267,268] Although the biosimilar candidate and Avastin were manufactured using the Chinese Hamster Ovary (CHO) cell line, differences in glycoform profile is not uncommon since culture process,

and culture media are known to have an influence on glycosylation. Also, the glycan composition does not relate to the mechanism of action of bevacizumab, hence the differences observed are not expected to have any clinical impact.

Although IMS-MS and global HDX do not provide high resolution structural information, they do offer insights on protein dynamics that can be critical for protein function and stability.[209] IMS-MS and time-resolved global HDX were used to determine any differences in protein dynamics. Broadly, the results from these analyses indicate that there is no substantial difference in the global structure or dynamics between Avastin and any of the biosimilar lots tested (Figure 4.3). This is perhaps unsurprising given that the only molecular difference detected in native MS was a relatively minor shift in glycosylation (corresponding mostly to the presence or absence of a galactose residue). Very small systematic variances in the HDX data do indicate that there is a statistically significant difference in the deuterium uptake profiles between Avastin and most of the biosimilar lots, corresponding to subtle differences in conformational dynamics, but was not clear that such subtle differences should translate to functional variation.

Functional biosimilarity *in vitro* was established from conventional SPR measurements, validated using a Tier 2 "Quality Range" approach. The approach is adapted from the FDA draft guidance for analytical similarity assessment for biosimilars (UCM576786) in which analytical similarity is demonstrated if the measurement for each biosimilar lot, or a target percentage of lots (*e.g.*, 90%), fall within $x \times \sigma$ of the reference product mean, where $x$ is determined by variability in the same parameter of the originator product (in this case $K_D$).[269] Since all of the originator $K_D$ values fall within $2\sigma$ of the mean, the analytical similarity criterion was selected as $2\sigma$, and this condition is met by the biosimilar (see Table 4.1).

The detection of small functional variance between all biosimilar lots and Avastin agrees with the detection of limited structural and dynamic differences in the MS-based assays. Nonetheless, there are some correlations between the structural and functional data that can be drawn. For instance, the biosimilar lots with the highest $K_D$ values (A, B and E, corresponding to 0.133, 0.102 and 0.132 nM, respectively) were those with the most mature glycosylation distributions (Figure 4.4) and lowest deuterium uptake in the 'slow' phase of global HDX (Figure 4.5B). In addition, biosimilar lot B, which showed the lowest rate of deuterium uptake among all biosimilar batches in global HDX, exhibited substantially lower kinetic binding parameters $k_{on}$ and $k_{off}$ in the SPR analysis. However, given the subtle nature of the physical differences observed among all biosimilar lots, additional studies with other systems will be needed to definitively establish causal relationships between physical and properties measured using the MS-based methods employed here and functional parameters.

Batch-to-batch variability is a particular concern in protein therapeutic production, given the inherent complexity of the process and the sensitivity of the product to even small changes in manufacturing procedures.[255] Our work demonstrates that the combination of native MS, IMS and global HDX can provide a detailed picture of batch-to-batch reproducibility. In this case, native MS showed a generally high degree of similarity in different biosimilar batches with respect to glycosylation, while at the same time clearly defining some minor differences among batches (Figure 4.4). These results could be used to optimize manufacturing for lower variability in glycosylation and toward a distribution that better favors a desired glycoform. Similarly, the data acquired in the pH stability analysis provide a detailed picture of product stability, including the detection of alternate conformations of the monomeric protein that are associated with aggregation. Applied to the sample at different stages of manufacturing, these results can generate an 'early

warning' that a particular process is producing conformational instability that may cause aggregation or artifactual modification. The 'stability' of the native configuration, as determined by the pH conditions needed to perturb it, may also have direct implications for product shelf-life.

## 4.6 Conclusions

In this work, we demonstrated a rapid, facile approach for detailed characterization of molecular similarity in 'biosimilar' candidate drugs using a suite of MS-based tools. Using this approach, we were able to demonstrate a high degree of molecular similarity between a commercially available mAb, Avastin, and a biosimilar in development, with substantial differences occurring only with respect to the glycoform distribution. Neither these differences, nor the small (but statistically significant) differences in conformational dynamics observed appeared to have a substantial impact on target engagement *in vitro* as indicated by SPR measurements, nor on pH stability as demonstrated via an IMS-MS based analysis. Nonetheless, some trends in the data suggest a potential correlation between glycosylation maturity, conformational dynamics and binding affinity, which can be explored in a broader set of mAbs systems in future work. Both IMS-MS and HDX-MS are relatively new techniques in the field of molecular biology and more specifically protein therapeutics. IM has the ability to rapidly distinguish isoforms of mAbs and could be a useful tool in early screening in biosimilar development.[2] In addition, HDX-MS provides additional (solution phase) information that could sometimes be missed in IMS-MS (gas phase) measurements and so these techniques work well together. Although local HDX provides more detailed information on the dynamics of proteins, global HDX serves as a more rapid screening method to quickly identify differences which can then be later explored using local HDX experiments if necessary.[68] Ultimately, our results here demonstrate the utility of native MS, IM-

MS and global HDX-MS  as a facile, rapid and reliable toolkit for establishing molecular similarity

of mAb biosimilars.

**Chapter 5: Hydrogen deuterium exchange epitope mapping as a tool for early screening of bevacizumab: a pre-clinical biosimilar monoclonal antibody.**

**Brown, K.A**., Lento, C., Rajendran, S., Dowd, J., and Wilson, D.J.

## 5.1 Summary

Vascular Endothelial Growth Factor A (VEGF-A) plays a crucial role in angiogenesis and it is upregulated in various forms of cancer. As a result, VEGF-A has been an important target in the development of cancer treatment therapies. Bevacizumab, a monoclonal antibody, was developed as a cancer treatment tool as it binds and neutralizes VEGF. Biologics, however, are extremely expensive and the development of biosimilars is an effective tool in reducing cost to patients. Here we use Time Resolved ElectroSpray Ionization Hydrogen Deuterium Exchange Mass Spectrometry (TRESI-HDX-MS) as a tool to compare the epitope binding sites of bevacizumab (Avastin) and a biosimilar in pre-clinical development when bound to $VEGF_{165}$. The epitope site determined in our experiments agree well with the known epitope determined from the co-crystal of the Fab domain of bevacizumab and VEGF. The epitope site was also determined to be identical for both bevacizumab and its biosimilar. However, there were short lived allosteric conformational changes observed for both complexes (Bevacizumab/VEGF and Biosimilar/VEGF) in distinct regions of VEGF suggestive of slightly different binding modes.

## 5.2 Introduction

A well-developed vascular system in humans ensures each cell gets adequate and consistent supply of oxygen. This extensive vascular system is maintained through angiogenesis, the growth and development of new blood vessels from pre-existing ones.[270] Vascular Endothelial Growth Factor A (VEGFA_Human, UniProtKB P15692) is an important regulator of angiogenesis and it was found that the deletion of the VEGF allele led to abnormal blood vessel growth and even death in embryos.[271,272] VEGF is likewise a key player in pathological angiogenesis and the VEGF mRNA is over expressed in the majority of human tumors that have so far been studied. As a result, VEGF has been an important target for cancer treatment studies.[217]

Alternate exon splicing of the VEGF-A gene gives rise to seven isoforms in which the four main VEGF isoforms are (VEGF$_{121}$, VEGF$_{165}$, VEGF$_{189}$, VEGF$_{206}$) of which VEGF$_{165}$ is the predominant molecular species. [273] Humanized Bevacizumab (Avastin) is an IgG1 monoclonal antibody that binds and neutralizes all isoforms of VEGF-A, preventing VEGF from binding to its receptors (VEGFR-1 and VEGFR-2) on endothelial cell surfaces which subsequently prevents further signal transduction events leading to angiogenesis. As a result it is now used as an effective cancer treatment tool as it prevents tumor angiogenesis. [217]

The availability of biosimilars is expected to significantly reduce the cost to patients, [125] but due to their complex structures their development can be difficult, especially in proving similarity. [274] A few biosimilar mAbs have been approved by the FDA with Bevacizumab-awwb (Mvasi) being the 1st biosimilar approved for cancer treatment in 2017. [248] Although new biologics put a larger emphasis on clinical trials, biosimilar development requires a more extensive pre-clinical development phase. As a result pre-clinical development of biosimilars is a lengthy process requiring several different analytical techniques. [126] In this study, we use Hydrogen Deuterium eXchange Mass Spectrometry (HDX-MS) to perform epitope mapping experiments to compare the interactions between Avastin (Genentech) and VEGF to that of Bevacizumab (Apobiologix) and VEGF.

HDX-MS have been progressively used for studying the dynamics of protein therapeutics, as it is not limited by size or crystallization properties. It can be used to study protein conformational dynamics on a wide time scale and it can be used to study complex protein systems. [231,233,275] Epitope mapping using HDX-MS is a powerful tool in early stage biosimilar development as it provides a rapid method for the determination of binding epitopes which is critical for the assessment of bioequivalence. [276] Here we use Time Resolved Electrospray Ionization (TRESI)

HDX-MS which provides a faster and simpler workflow in comparison to conventional HDX-MS and experiments are conducted on faster HDX reaction times.[277–279] The HDX experiment set up used is described in detail in chapters 1.3.5 and 3.3.4.

## 5.3 Materials and Methods

### 5.3.1 Materials

Avastin (MW = 149.2 kDa) and the candidate biosimilar (MW = 149.4 kDa) were provided by Apobiologix (North York, ON). Recombinant Human Vascular Endothelial Growth Factor 165 ($VEGF_{165}$) was purchased from R and D systems (Minneapolis, MN). Deuterium oxide ($D_2O$, 99.9%, 151882) and high purity acetic acid (>99.7%, 695092) were purchased from Sigma-Aldrich (Oakville, ON). HPLC-grade ammonium acetate and water were purchased from Fisher Chemical (Mississauga, ON).

### 5.3.2 Time-Resolved ElectroSpray Ionization Hydrogen Deuterium eXchange (TRESI-HDX)

The microfluidic device was constructed as described in chapter 3.3.4. The HDX reaction occurred within the TRESI mixer in which gas tight Hamilton syringes were used to transport the reagents through the capillaries using Syringe infusion pumps (Harvard Apparatus). Two types of reactions were monitored (Figure 5.1). In the 1st type of experiment, which is referred to as an equilibrium experiment, the antibody and antigen were allowed to equilibrate for 30 minutes prior to HDX analysis. Afterwards, 2 µM $VEGF_{165}$ (or Antibody/$VEGF_{165}$ complex) in 150 mM $C_2H_3O_2NH_4$, pH 6.7 flowed through the inner capillary while deuterium flowed through the outer capillary at 2 µL/min. In the 2nd type of experiment, which we refer to as a 'kinetic' experiment, the antibody and antigen were not pre-equilibrated. In this experiment, VEGF flowed through the inner capillary and the antibody in deuterium flowed through the outer capillary. As a result, the antibody-antigen interaction occurred concurrently with the deuterium labelling. In both types of experiment, the

HDX reaction times were achieved by increasing the space between the end of the inner capillary

and the outer capillary by 2, 5, 10 and 20 mm which corresponded to HDX reaction times of 1, 2,

4 and 8 s respectively. The instrument operated with a capillary voltage of 2.5 kV in positive ion

mode with a sampling cone voltage of 20 V. The samples were scanned over a range of 400 to

1500 m/z.
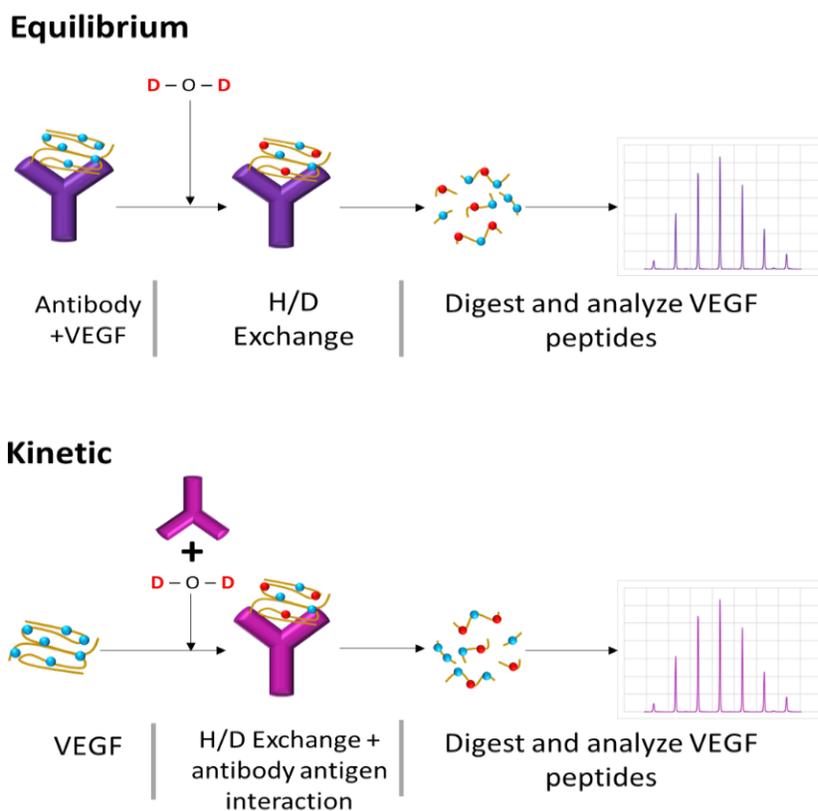


**Figure 5.1. Schematic representation of the equilibrium and kinetic HDX experiment workflow.** In the equilibrium reactions, VEGF and Bevacizumab were allowed to pre-equilibrate for 30 minutes prior to the HDX reaction. In the kinetics reaction, VEGF and bevacizumab interact simultaneously with the HDX reaction.

A digestion profile was acquired prior to the HDX reaction and tandem MS was used to determine

the sequence of each peptide. Data Analysis was carried out using the MS studio software [198] in

which the deuterium levels for each peptide in free VEGF$_{165}$ and Antibody/VEGF$_{165}$ complex was calculated.

**5.4 Results and Discussion**

The receptor binding domain of VEGF which spans residues 34-135 of the canonical sequence (VEGF$_{206}$) has been previously co-crystalized with the Fab domain of a neutralizing antibody. There are 19 VEGF residues at the binding interface with the Fab fragment: residues 43, 47, 71, 74, 105-110 and 112-120. However, the functional epitope includes only six residues; Met107, Arg108, Ile109, Gly114, Gln115 and Gly118.[216] Although we are not able to receive this kind of high resolution data with our method, we are able to provide an overall picture of the dynamics of VEGF when bound to Avastin and the biosimilar in order to perform a rapid comparison. And because HDX-MS is fast and reproduceable this could be a reliable tool in early screening of potential biosimilar antibodies in the primary stages of development.

Here, we used the full construct of VEGF$_{165}$ (residues 27-191 of the canonical sequence). Epitope mapping HDX analysis was done on the VEGF peptides at four time points: 1, 2, 4 and 8 s. The graphs in figure 5.2, represent a difference plot, in which the percent deuterium uptake of the free VEGF is subtracted from the VEGF-Antibody complex. A negative value represents a decrease in deuterium uptake upon binding to the antibody whereas a positive value represents an increase. In HDX experiments, decreased deuterium uptake is often the result of direct binding interactions or reduced solvent accessibility at a binding interface. Decreased deuterium could also represent structural changes due to allosteric conformational changes upon binding. It is often difficult to distinguish the two without prior knowledge of binding interactions. On the other hand, in most cases an increase in deuterium uptake is as a result of an allosteric conformational change. [68,280] To determine what changes are considered significant, three replicates were conducted, and an

error analysis was done in which changes greater than 5% within two sigmas (95% C.I.) were considered significant.

In the VEGF-Avastin complex a peptide (108-115) in the binding interface began to show a significant decrease in deuterium uptake at the initial time point (1 s), whereas no significant decreases at the epitope region was observed in the VEGF-biosimilar complex until 2 s. Interestingly, a small but significant increase in deuterium uptake outside of the binding region was observed in both complexes, indicative of an allosteric conformational change. However, this was observed in different regions of each complex. In the VEGF-Avastin complex this was observed in the C-terminal region (peptide 154-162) that has been shown to be a heparin binding site. Interestingly, it was found that loss of this heparin binding site decreased the cell mitogenic activity of VEGF substantially.[281] On the other hand, in the VEGF-Biosimilar complex, this increase was observed in the N-terminal region (peptide 27-30). These increases, however, were only observed at the 1st time point, and it is important to note that these differences although significant are very small and disappeared over time. This could be indicative of a slight difference in binding modes, but more experiments would be required to explore this further. Another potential allosteric conformational change was observed in both complexes at 4 s. In the VEGF-Avastin complex a small but significant decrease was observed just outside the binding interface at peptide 91-103. Whereas in the VEGF-Biosimilar complex it was observed at 154-162 (within the heparin binding domain). The final time point (8s) only showed significant decreases in the epitope binding region in both complexes. This suggests that the biosimilar has the same binding interface as Avastin. Unlike any short lived allosteric conformational changes observed, the decreases observed at the binding site remained as the HDX reaction time increased.
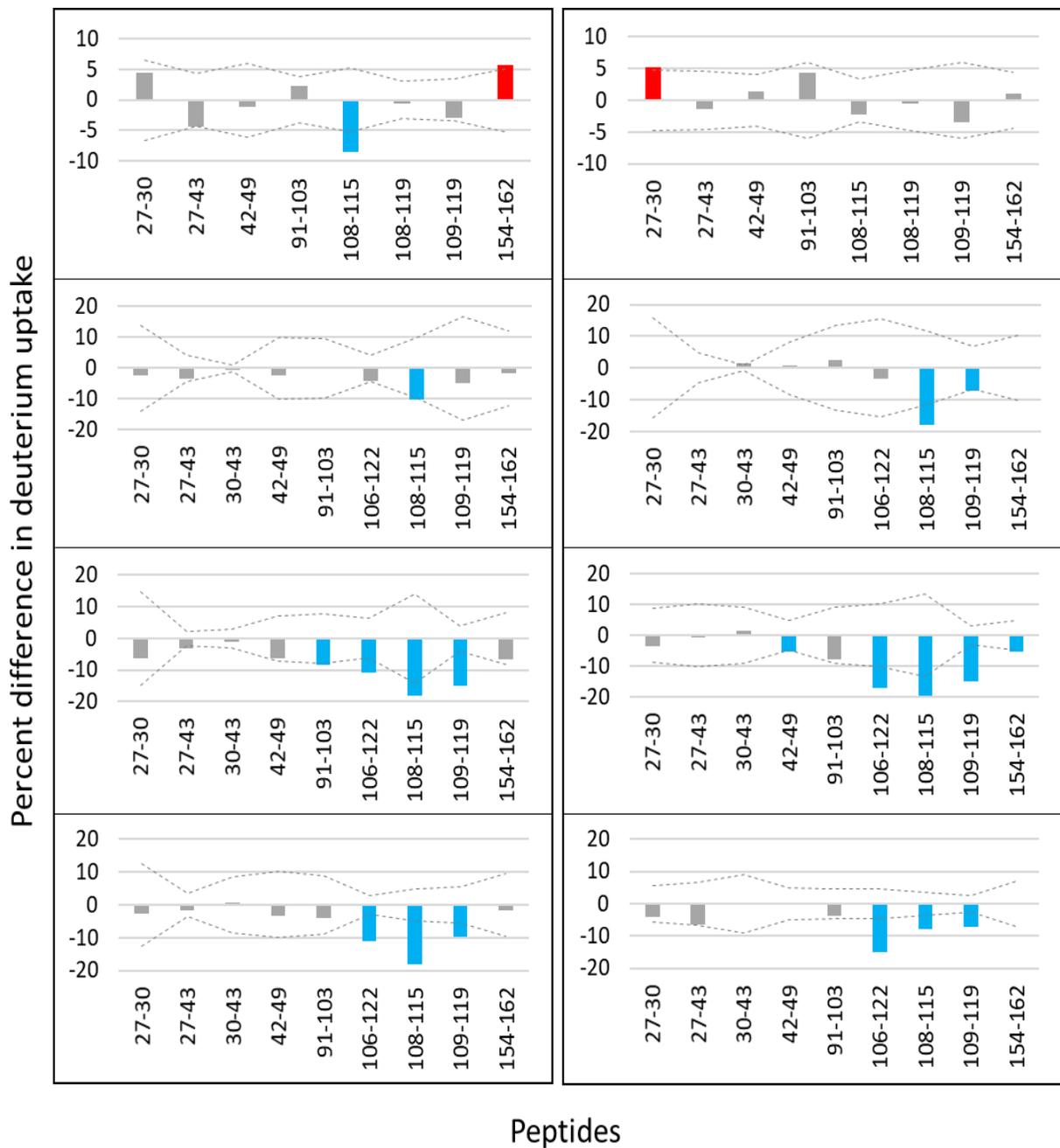
**Figure 5.2. Equilibrium HDX experiment.** Percent difference in deuterium uptake in VEGF-Avastin (left) and VEGF-Biosimilar complex (right) obtained at HDX reaction times (from top to bottom) of 1, 2, 4 and 8 s. Dashed lines represent a two-sigma cut off for differences considered to be significant. Significant decreases and increases are highlighted blue and red respectively.

A major advantage and drawback of HDX is the sensitivity to structural changes within a protein upon ligand binding. It can become difficult to differentiate between binding interactions and allosteric conformational changes when decreased deuterium uptake is observed in multiple regions of a protein. [280] Previous studies have suggested that monitoring the changes in the differences in deuterium over time could potentially shed some light as to what changes correspond to binding or allosteric conformational changes. The hypothesis is that significant decreases due to binding is observed at the earliest HDX reaction times and long range conformational changes that may occur as a result of binding are likely to become apparent at longer HDX reaction times.[279,280,282] This is due to the fact that at the binding interface the decreased deuterium uptake is as a result of H-bonding of the amide backbone with a ligand or protection from solvent at the binding interface. In allosteric regions the rate of exchange is likely to depend on the rate of conformational change (which often happens after ligand binding) and the rate of intrinsic exchange at the allosteric site. Therefore, a decrease observed at the allosteric site at later HDX reaction times implies the intrinsic rate is slower than the rate of conformational change. [280,282]

To further test this hypothesis and to get a clearer differentiation between binding site and allosteric conformational change, a 'kinetic' HDX experiment was done (Figure 5.3). In this experiment, instead of incubating VEGF and Bevacizumab prior to the HDX reaction, the deuterium labelling and antibody-antigen interaction were carried out concurrently. In the VEGF-Avastin complex, at the initial time point (1 s) only the peptide within the epitope site (peptide 108-115) showed a significant decrease in deuterium uptake. However, in the VEGF-Biosimilar complex peptide 108-115 along with another peptide (141-154) showed a significant decrease in deuterium uptake.
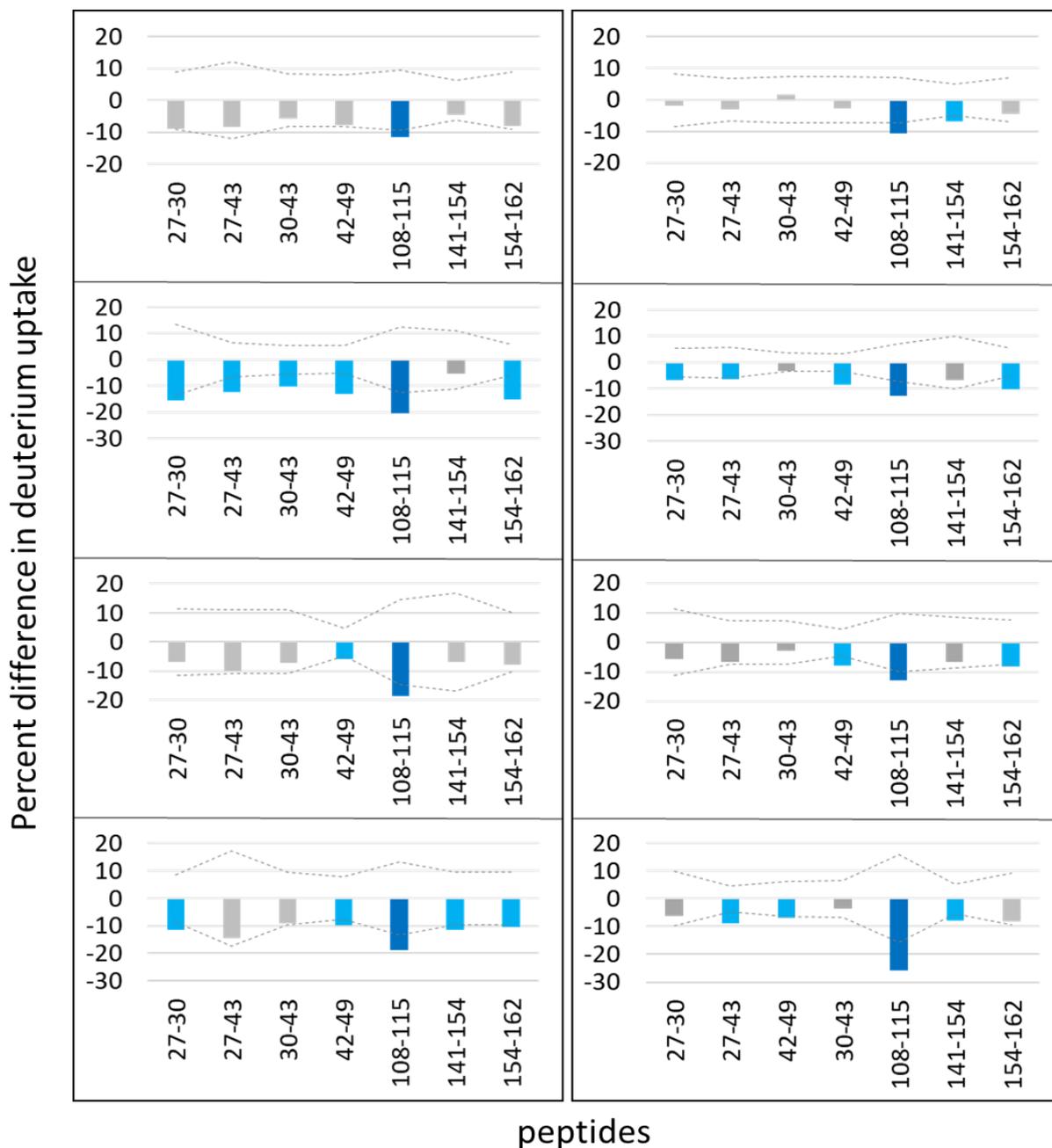
**Figure 5.3. Kinetic HDX experiment**. Percent difference in deuterium uptake in VEGF-Avastin (left) and VEGF-Biosimilar complex (right) obtained at HDX reaction times (from top to bottom) of 1, 2, 4 and 8 s. Dashed lines represent a two-sigma cut off for differences considered to be significant. Significant decreases are highlighted blue and the epitope peptides are highlighted dark blue.

Based on the aforementioned hypothesis, one would conclude that there was an additional epitope site in the VEGF-Biosimilar complex. However, upon further analysis, this decrease at residues 141-154 disappeared at 2 s and 4 s and reappeared again at 8 s. In this 'kinetic' analysis, the known epitope consistently showed a significant decrease at all time points studied, whereas any long-range conformational changes did not show consistent changes in deuterium uptake across the HDX reaction times. The epitope site determined from the HDX reaction was mapped onto the co-crystal of VEGF and the Fab domain of bevacizumab for better visualization of the results. Figure 5.4A shows the binding interface determined from the X-ray co-crystal, in which all the amino acids involved in the binding interface is highlighted green. Figure 5.4B and 5.4C shows the epitope site in the VEGF-Avastin and VEGF biosimilar complex identified from the HDX results. The HDX results are comparable to what was observed in the X-ray crystallography experiments. Moreover, the epitope site is identical in the VEGF-Avastin and VEGF-Biosimilar complex.
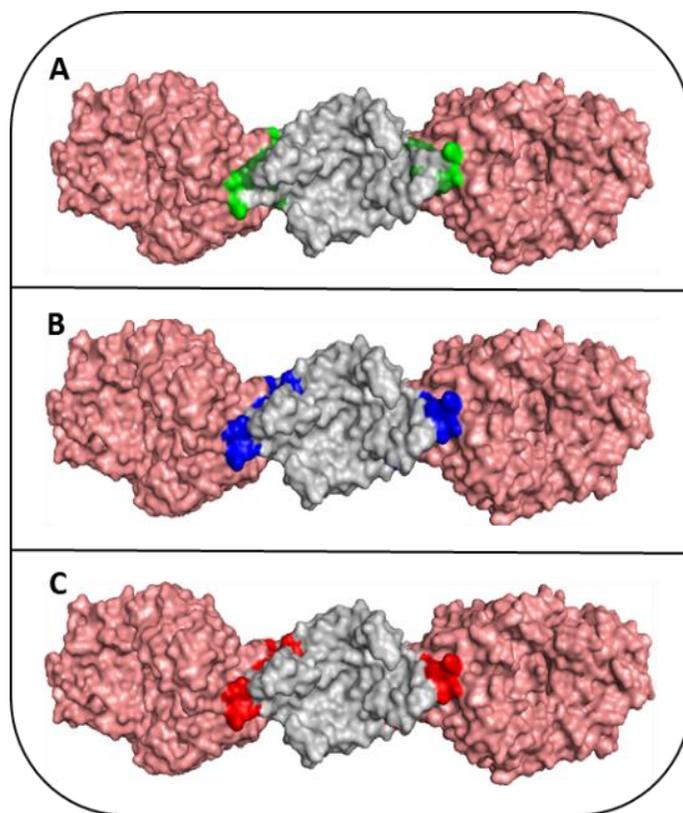
**Figure 5.4. X-ray structure of VEGF (gray) and the Fab domain of bevacizumab (pink) PDB ID 1BJ1**. (A) shows the binding interface determined from the X-ray structure highlighted in green. (B) shows the binding interface between VEGF and Avastin determined from the HDX experiments and (C) shows the binding interface of VEGF and the biosimilar determined from the HDX experiments.

## 5.5 Conclusions

Development of biosimilars is important and can greatly reduce the cost to patients. However, the development can be challenging, especially in proving similarity. In this work, we show that HDX-MS can be used in the pre-clinical development of biosimilars as an early screening tool. It remains challenging to differentiate binding interactions and long-range conformational changes using HDX-MS experiments. However, when used alongside X-ray crystallography it becomes easier to study these allosteric changes. Nevertheless, careful analysis of multiple HDX reaction times can aid in the differentiation between binding sites and allosteric conformational changes[280,282]

especially in epitope mapping experiments that can easily result in false positives when epitope sites are to be determined.[279] Based on these experiments, Avastin and the biosimilar studied have identical epitope sites. However, differences were observed in the allosteric conformational changes which could be indicative of difference in binding mechanisms. Additional experiments would be required to explore this further.

Future work involves epitope mapping experiment of several different batches of biosimilars to determine the effectiveness of HDX-MS as an early screening tool. In addition, more HDX reaction times need to be studied in both the equilibrium and "kinetic" experiments. Both shorter and longer HDX reaction times need to be studied to further test the hypothesis for the differentiation between binding and allosteric conformational changes. Additional experiments would be required to further explore the differences in the binding mechanisms such as Molecular Dynamics (MD) Simulation which can be a useful tool in gaining insights into the conformational changes that result from binding. MD simulations can be used as a guide to aid in better interpretation of the HDX results.[283]

**Chapter 6: Conclusions and Future Work**

**6.1 Conclusions**

Structural MS is a powerful analytical tool that can be used to analyze structural and binding characteristics of protein and protein ligand complexes.[2,3] Here, it was used to study the binding characteristics of hLa, an important RNA chaperone, to various RNA ligands. In addition, it was used as a comparison tool to characterize the biosimilarity between a biological drug, Avastin, and a biosimilar in preclinical development.

Chapter 2 aimed to discover the RNA binding mode of the CTD of hLa which was previously unknown. The Binding mode of hLa to 3'UUU-OH RNA is well studied and there are high resolution structures (generated through the use of X-ray crystallography) available for the complex.[117,201] However, due to the unstructured nature of the CTD there are currently no high-resolution structures available for the CTD (which is ~44% unstructured) or for the entire protein. Native MS revealed that in low salt concentrations (10 mM) hLa-CTD is highly unstructured which is evident by the trimodal distribution in the native spectrum. Upon binding to the 23 mer ssRNA, the folded distribution increased in intensity while the unfolded distribution decreased to almost zero. This suggests that RNA binding significantly changes the dynamics causing the protein to potentially become more structured. IM-MS revealed that RRM2 mediated RNA binding in hLa occurs through conformer selection. The IM spectrum revealed two conformations in the $10^+$ charge state (in the folded distribution) of the free protein, where the more extended conformation is the more dominant, but a small intensity compact conformation can be seen. This compact conformation becomes more populated in the hLa-RNA complex. HDX-MS data revealed that hLa-CTD binds a ssRNA and a more structured RNA (the HCV RNA) via

117

significantly different binding modes. Binding of hLa to the unstructured ssRNA resulted in a destabilization in RRM2 α3 helix that blocks the typical RNA binding site. The structured RNA however, resulted in a stabilization of this α3 helix and in addition, several more decreases in deuterium uptake are observed in the unstructured C-terminal tail suggestive of more widespread engagement with the unstructured region.

Chapter 3 aimed to look at hLa in its entirety to understand the role each domain plays in RNA binding. In addition, alterations in the nuclear trafficking signals that allow hLa to shuttle between the cytoplasm and nucleus were explored. It was previously hypothesized by Bayfield *et al*., that there is the possibility of an interdomain bridge between RRM1 and RRM2 that control the shuttling of hLa between the nucleus and the cytoplasm.[196] Native MS analysis revealed a 1:1 binding ratio of hLa with U10 and A20 RNA. CIU was then used to determine how the stability of hLa and hLaΔNRE are affected in the presence of U10 RNA. With two clear stages of unfolding observed in which the 1st stage is stabilized by the presence of U10 RNA but the 2nd stage being destabilized suggest this 1st stage of unfolding is representative of the disruption of the la motif and the 2nd stage represents the breaking of the RRM1-RRM2 interdomain interaction. In hLaΔNRE, due to the absence of the α3 helix, and therefore the RRM1-RRM2 interaction, the transition between the 2nd and 3rd unfolded states was almost immediate. TRESI-HDX revealed distinct binding mechanisms of hLa when bound to U10 and A20 RNA. the most notable difference is the increased dynamics of the α3 helix in the RRM2 when hLa is bound to RNA whereas decreased dynamics is observed when bound to A20. If an interaction is present, it would likely be disrupted for U10 RNA to bind within the known binding site in the La motif. A20 contacts all three domains, so even if the interaction is broken the A20 binding may still prevent the α3 helix from destabilizing. Deletion of the NRE helix in the hLaΔNRE mutant resulted in almost identical

HDX profiles for U10 and A20 complexes which further supports the presence of an interdomain interaction. The lack of the interdomain interaction in hLaK316E/K317E completely changed the dynamics of the CTD when hLaK316E/K317E was bound to U10. When bound to A20, not many significant changes in deuterium uptake was observed suggesting hLa316K/317K may have been an important RNA binding surface or that the interaction itself is important for A20 RNA binding. In the compensatory mutant (hLa D132K/E133K/K316E/K317E), in which the interaction is restored, binding to U10 and A20 was very similar to hLa suggesting the interaction has been restored.

The development of biosimilar antibodies has been on the rise for the past few years as a way to potentially reduce the cost of therapeutic antibodies to patients. Due to the complex manufacturing processes however, it is impossible to create a biosimilar that is 100% identical to its innovator drug. None the same, the biosimilar must be chemically similar to the innovator drug while providing the same clinical effect.[124,126] Chapter 4 focused on characterizing and comparing the structural dynamics of Avastin and its biosimilar, bevacizumab. Structural MS proved to be a rapid and facile approach to characterize the molecular similarity. Native MS of the antibodies revealed a similar average molecular weight with a difference of only 200 Da. A Major difference observed between the Avastin and its biosimilar, is the presence of a low intensity degradation product (loss of 1447 Da) in Avastin, representing a monoglycosylated form with a single G0F modification. In addition, the biosimilar had more mature glycosylations than Avastin. The differences in glycosylations however did not seem to have a large impact on the structural dynamics and target engagement *in vitro,* as indicated by SPR measurements. The complex manufacturing process in mAbs can also lead to batch-to-batch variability which is of concern in protein therapeutic production. Native-MS revealed high similarity between different batches of the biosimilar in

regard to the glycosylations, although minor differences could be seen. These results obtained through native MS could be used to optimize manufacturing processes to potentially lower the variability between batches. Additionally, IM-MS and global HDX revealed only minor differences between batches.

Chapter 5 further explored the biosimilarity of Avastin and its biosimilar through HDX-MS epitope mapping to its target, VEGF. The epitope site was previously determined through the analysis of a co-crystal of the Fab domain of a neutralizing antibody bound to the receptor binding domain of VEGF. It was discovered that the functional epitope comprised of residues 107-109, 114, 115 and 118. Due to the availability of this information it was possible to differentiate binding interactions from allosteric conformational changes in the HDX-MS experiments. The equilibrium HDX experiments done at 4 time points (1, 2, 4 and 8 s) revealed the same epitope sites, for both Avastin and the biosimilar, peptides 106-122, 108-115 and 109-119. However, different allosteric conformational changes were observed in both complexes which might be indicative of slightly different binding modes. 'Kinetic' HDX experiments were done to get a clearer differentiation between binding site and allosteric conformational changes. The same epitope site, peptide 108-115, was determined for Avastin and its biosimilar but the allosteric conformational changes were again different between the two which further suggests that the two antibodies may have slightly different binding modes.

## 6.2 Outlook and Future Work

Structural MS is becoming an increasingly popular technique in structural biology and in the discovery and development of therapeutic antibodies. With the advancements in MS instrumentation that allows for increased resolution, sensitivity and coupling of ion mobility separation, larger and more complex protein structures can be analyzed.[2–4] Most notably highly

flexible proteins with little to no structure can be more easily analyzed by MS than other conventional structural techniques and with increased sensitivity of new instrumentation, very little sample is required for analysis. [42,98,100,178]

The work done in chapter 2 and 3 shows that structural MS experiments has the ability to screen protein-RNA interactions to determine RNA binding sites. This work can help us to better understand the processing of nascent RNAs and translation of cellular and viral RNAs. As RNA chaperones often have highly flexible domains, they are difficult to study with X-ray crystallography to determine absolute binding sites. Another structural MS technique that can be used to provide additional information on the changes in dynamics upon RNA binding is chemical cross linking MS. Crosslinking is often used to study protein structure as it provides proximity information of multi-domain proteins and protein complexes.[284–286] Although MS is a low-resolution structural technique, it has the ability to provide details on the changes in dynamics when RNA is bound as well as how the structure and stability is affected by RNA binding.

To gain more information on the structure of the protein and protein/RNA complexes, molecular modelling could be used alongside HDX-MS results to determine possible structures and how these structures change when the protein is bound to RNA. In the case of hLa, there are high resolution structures available for the La motif,[117,200,201] RRM1[153,200,201] and RRM2[199] obtained by X-Ray crystallography and NMR experiments. Homology modelling can be used to determine the structure of the loop regions between La Motif and RRM1 after which MD simulations[283] can be used to determine the combined structure of all 3 domains. Finally, the structure of the long unstructured C-terminal tail can be added to the rest of the structure. To determine RNA binding interactions, docking experiments can then be conducted to determine possible RNA binding sites.

Chapters 4 and 5 shows how structural MS can be used in the analysis of therapeutic antibodies and biosimilars. Biopharmaceuticals is a fast-growing class of therapies used for the treatment of a wide range of illnesses from cancers to autoimmune diseases in which monoclonal antibodies are the most common and fastest growing sub-class. This is due to the specificity and selectively of mAbs for their respective targets.[209,230,276] The work done in chapter 5 can be extended to study more HDX reaction times in both kinetic and equilibrium experiments to further understand any differences in binding mechanisms. The major limitation in local HDX experiments is the ability to distinguish between binding interactions and long range allosteric conformational changes. Although several groups have proposed a link between HDX reaction time and the ability to distinguish binding and allostery,[105,106,287] more research is needed to determine if there is truly a link. Without prior knowledge of binding sites, it is difficult to determine what changes in local HDX experiments correspond to binding interactions.

Mass Spectrometry is a powerful technique in structural biology that can provide information on protein primary structure, binding stoichiometry, subunit composition, protein size, gas phase stability, ligand binding sites and conformational changes. There are however many challenges left to be overcome but with the continuing development of Structural MS it will remain a strong tool in the analysis of biomacromolecules.

# References

(1) Yang, L.-Q., Sang, P., Tao, Y., Fu, Y.-X., Zhang, K.-Q., Xie, Y.-H., and Liu, S.-Q. (2014) Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn. 32*, 372–393.

(2) Tian, Y., and T. Ruotolo, B. (2018) The growing role of structural mass spectrometry in the discovery and development of therapeutic antibodies. *Analyst 143*, 2459–2468.

(3) Allison, T. M., and Bechara, C. (2019) Structural mass spectrometry comes of age: new insight into protein structure, function and interactions. *Biochem. Soc. Trans. 47*, 317–327.

(4) Liko, I., Allison, T. M., Hopper, J. T., and Robinson, C. V. (2016) Mass spectrometry guided structural biology. *Curr. Opin. Struct. Biol. 40*, 136–144.

(5) Konijnenberg, A., Butterer, A., and Sobott, F. (2013) Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta BBA - Proteins Proteomics 1834*, 1239–1256.

(6) Heck, A. J. R. (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Meth 5*, 927–933.

(7) Artigues, A., Nadeau, O. W., Rimmer, M. A., Villar, M. T., Du, X., Fenton, A. W., and Carlson, G. M. (2016) Protein Structural Analysis via Mass Spectrometry-Based Proteomics. *Adv. Exp. Med. Biol. 919*, 397–431.

(8) Bleakney, W., and Hipple, J. A. (1938) A New Mass Spectrometer with Improved Focusing Properties. *Phys. Rev. 53*, 521–529.

(9) Wiley, W. C., and McLaren, I. H. (1955) Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum. 26*, 1150–1157.

(10) Nier, A. O. (1947) A Mass Spectrometer for Isotope and Gas Analysis. *Rev. Sci. Instrum. 18*, 398–411.

(11) Whitehouse, C. M., Dreyer, R. N., Yamashita, Masamichi., and Fenn, J. B. (1985) Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem. 57*, 675–679.

(12) Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science 246*, 64–71.

(13) Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., and Alice, M. B. (1968) Molecular Beams of Macroions. *J. Chem. Phys. 49*, 2240–2249.

(14) Yamashita, M., and Fenn, J. B. (1984) Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem. 88*, 4451–4459.

(15) Nohmi, T., and Fenn, J. B. (1992) Electrospray mass spectrometry of poly(ethylene glycols) with molecular weights up to five million. *J. Am. Chem. Soc. 114*, 3241–3246.

(16) Blades, A. T., Ikonomou, M. G., and Kebarle, Paul. (1991) Mechanism of electrospray mass spectrometry. Electrospray as an electrolysis cell. *Anal. Chem. 63*, 2109–2114.

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

(18) Loeb, L. B., Kip, A. F., Hudson, G. G., and Bennett, W. H. (1941) Pulses in Negative Point-to-Plane Corona. *Phys. Rev. 60*, 714–722.

(19) Taylor Geoffrey Ingram. (1964) Disintegration of water drops in an electric field. *Proc. R. Soc. Lond. Ser. Math. Phys. Sci. 280*, 383–397.

(20) Taylor, G. I., and McEwan, A. D. (1965) The stability of a horizontal fluid interface in a vertical electric field. *J. Fluid Mech. 22*, 1–15.

(21) Loscertales, I. G., and Fernández de la Mora, J. (1995) Experiments on the kinetics of field evaporation of small ions from droplets. *J. Chem. Phys. 103*, 5041–5060.

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

(23) Metwally, H., Duez, Q., and Konermann, L. (2018) Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry. *Anal. Chem. 90*, 10069–10077.

(24) Iribarne, J. V., and Thomson, B. A. (1976) On the evaporation of small ions from charged droplets. *J. Chem. Phys. 64*, 2287–2294.

(25) Thomson, B. A., and Iribarne, J. V. (1979) Field induced ion evaporation from liquid surfaces at atmospheric pressure. *J. Chem. Phys. 71*, 4451–4463.

(26) Kebarle, P., and Tang, L. (1993) From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. *Anal. Chem. 65*, 972A-986A.

(27) Verkerk, U. H., and Kebarle, P. (2005) Ion-Ion and ion-molecule reactions at the surface of proteins produced by nanospray. Information on the number of acidic residues and control of the number of ionized acidic and basic residues. *J. Am. Soc. Mass Spectrom. 16*, 1325–1341.

(28) Heuvel, R. H. van den, and Heck, A. J. (2004) Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol. 8*, 519–526.

(29) Leney, A. C., and Heck, A. J. R. (2017) Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom. 28*, 5–13.

(30) Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C. M., Hoey, K., and Change, C.-M. (1996) Mass spectrometry and viral analysis. *Chem. Biol. 3*, 45–48.

(31) Verbeck, G., Ruotolo, B., Sawyer, H., Gillig, K., and Russell, D. (2002) A fundamental introduction to ion mobility mass spectrometry applied to the analysis of biomolecules. *J. Biomol. Tech. JBT 13*, 56–61.

(32) Kanu, A. B., Dwivedi, P., Tam, M., Matz, L., and Hill, H. H. (2008) Ion mobility–mass spectrometry. *J. Mass Spectrom. 43*, 1–22.

(33) Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T., and Scrivens, J. H. (2007) An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom. 261*, 1–12.

(34) Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E., and Ashcroft, A. E. (2009) Deciphering drift time measurements from travelling wave ion mobility spectrometry- Mass spectrometry studies. *Eur. J. Mass Spectrom. 15*, 113–30.

(35) Shukla, A. K., and Futrell, J. H. (2000) Tandem mass spectrometry: dissociation of ions by collisional activation. *J. Mass Spectrom. 35*, 1069–1090.

(36) Benesch, J. L. P. (2009) Collisional activation of protein complexes: Picking up the pieces. *J. Am. Soc. Mass Spectrom. 20*, 341–348.

(37) Dixit, S. M., Polasky, D. A., and Ruotolo, B. T. (2018) Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Curr. Opin. Chem. Biol. 42*, 93–100.

(38) Woodward, C., Simon, I., and Tüchsen, E. (1982) Hydrogen exchange and the dynamic structure of proteins. *Mol. Cell. Biochem. 48*, 135–160.

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

(40) Pirrone, G. F., Iacob, R. E., and Engen, J. R. (2015) Applications of Hydrogen/Deuterium Exchange MS from 2012 to 2014. *Anal. Chem. 87*, 99–118.

(41) Lento, C., Audette, G. F., and Wilson, D. J. (2014) Time-resolved electrospray mass spectrometry — a brief history. *Can. J. Chem. 93*, 7–12.

(42) Zhu, S., Shala, A., Bezginov, A., Sljoka, A., Audette, G., and Wilson, D. J. (2015) Hyperphosphorylation of intrinsically disordered tau protein induces an amyloidogenic shift in its conformational ensemble. *PloS One 10*, e0120416.

(43) Deng, B., Lento, C., and Wilson, D. J. (2016) Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development – A review. *Anal. Chim. Acta 940*, 8–20.

(44) Pan, J., Zhang, S., and Borchers, C. H. (2016) Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry. *Biochim. Biophys. Acta BBA - Proteins Proteomics 1864*, 1801–1808.

(45) Prądzińska, M., Behrendt, I., Astorga-Wells, J., Manoilov, A., Zubarev, R. A., Kołodziejczyk, A. S., Rodziewicz-Motowidło, S., and Czaplewska, P. (2016) Application of amide hydrogen/deuterium exchange mass spectrometry for epitope mapping in human cystatin C. *Amino Acids 48*, 2809–2820.

(46) Vadas, O., Jenkins, M. L., Dornan, G. L., and Burke, J. E. (2017) Chapter Seven - Using Hydrogen–Deuterium Exchange Mass Spectrometry to Examine Protein–Membrane Interactions, in *Methods in Enzymology* (Gelb, M. H., Ed.), pp 143–172. Academic Press.

(47) Anderegg, R. J., Wagner, D. S., and Stevenson, C. L. (1994) The mass spectrometry of helical unfolding in peptides. *J. Am. Soc. Mass Spectrom. 5*, 425–433.

(48) Harrison, A. G., and Yalcin, T. (1997) Proton mobility in protonated amino acids and peptides. *Int. J. Mass Spectrom. Ion Process. 165*, 339–347.

(49) Johnson, R. S., Krylov, D., and Walsh, K. A. (1995) Proton mobility within electrosprayed peptide ions. *J. Mass Spectrom. 30*, 386–387.

(50) Kenny, P. T. M., Nomoto, K., and Orlando, R. (1992) Fragmentation studies of peptides: The formation of Y ions. *Rapid Commun. Mass Spectrom. 6*, 95–97.

(51) Breuker, K., Oh, H., Lin, C., Carpenter, B. K., and McLafferty, F. W. (2004) Nonergodic and conformational control of the electron capture dissociation of protein cations. *Proc. Natl. Acad. Sci. U. S. A. 101*, 14011–14016.

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

(53) Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A. 101*, 9528–9533.

(54) Zhang, Z., and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci. Publ. Protein Soc. 2*, 522–531.

(55) Pan, J., Zhang, S., Chou, A., and H. Borchers, C. (2016) Higher-order structural interrogation of antibodies using middle-down hydrogen/deuterium exchange mass spectrometry. *Chem. Sci. 7*, 1480–1486.

(56) McBride-Warren, P. A., and Mueller, D. D. (1972) Hydrogen-Deuterium Exchange of Lysozyme. I. Rate Constants and pH Dependence. *Biochemistry 11*, 1785–1792.

(57) Matthew, J. B., and Richards, F. M. (1983) The pH Dependence of Hydrogen Exchange in Protein. *J. Biol. Chem. 258*, 3039–3044.

(58) Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins Struct. Funct. Bioinforma. 17*, 75–86.

(59) Englander, J. J., Calhoun, D. B., and Englander, S. W. (1979) Measurement and calibration of peptide group hydrogen-deuterium exchange by ultraviolet spectrophotometry. *Anal. Biochem. 92*, 517–524.

(60) Connelly, G. P., Bai, Y., Jeng, M.-F., and Englander, S. W. (1993) Isotope effects in peptide group hydrogen exchange. *Proteins Struct. Funct. Bioinforma. 17*, 87–92.

(61) Molday, R. S., Englander, S. W., and Kallen, R. G. (1972) Primary structure effects on peptide group hydrogen exchange. *Biochemistry 11*, 150–158.

(62) Leichtling, B. H., and Klotz, I. M. (1966) Catalysis of Hydrogen-Deuterium Exchange in Polypeptides*. *Biochemistry 5*, 4026–4037.

(63) Berger, A., Loewenstein, A., and Meiboom, S. (1959) Nuclear Magnetic Resonance Study of the Protolysis and Ionization of N-Methylacetamide1. *J. Am. Chem. Soc. 81*, 62–67.

(64) Klotz, I. M., and Frank, B. H. (1965) Deuterium-Hydrogen Exchange in Amide N—H Groups1. *J. Am. Chem. Soc. 87*, 2721–2728.

(65) Perrin, C. L. (1989) Proton exchange in amides: Surprises from simple systems. *Acc. Chem. Res. 22*, 268–275.

(66) Englander, S. W. (2006) Hydrogen Exchange and Mass Spectrometry: A Historical Perspective. *J. Am. Soc. Mass Spectrom. 17*, 1481–1489.

(67) Nielsen, S. O., Bryan, W. P., and Mikkelsen, K. (1960) Hydrogen-deuterium exchange of small peptides in aqueous solution. *Biochim. Biophys. Acta 42*, 550–552.

(68) Brown, K. A., and Wilson, D. J. (2017) Bottom-up hydrogen deuterium exchange mass spectrometry: data analysis and interpretation. *The Analyst 142*, 2874–2886.

(69) Englander, S. W., Downer, N. W., and Teitelbaum, H. (1972) Hydrogen Exchange. *Annu. Rev. Biochem. 41*, 903–924.

(70) Ferraro, D. M., Lazo, N. D., and Robertson, A. D. (2004) EX1 Hydrogen Exchange and Protein Folding. *Biochemistry 43*, 587–594.

(71) Witten, J., Ruschak, A., Poterba, T., Jaramillo, A., Miranker, A. D., and Jaswal, S. S. (2015) Mapping Protein Conformational Landscapes under Strongly Native Conditions with Hydrogen Exchange Mass Spectrometry. *J. Phys. Chem. B 119*, 10016–10024.

(72) Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M., and Ten Eyck, L. F. (2006) Identification and Characterization of EX1 Kinetics in H/D Exchange Mass Spectrometry by Peak Width Analysis. *J. Am. Soc. Mass Spectrom. 17*, 1498–1509.

(73) Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys. 16*, 521–655.

(74) Wales, T. E., and Engen, J. R. (2006) Partial Unfolding of Diverse SH3 Domains on a Wide Timescale. *J. Mol. Biol. 357*, 1592–1604.

(75) Engen, J. R., Smithgall, T. E., Gmeiner, W. H., and Smith, D. L. (1997) Identification and Localization of Slow, Natural, Cooperative Unfolding in the Hematopoietic Cell Kinase SH3 Domain by Amide Hydrogen Exchange and Mass Spectrometry †. *Biochemistry 36*, 14384–14391.

(76) Hollien, J., and Marqusee, S. (1999) Structural distribution of stability in a thermophilic enzyme. *Proc. Natl. Acad. Sci. U. S. A. 96*, 13674–13678.

(77) Guttman, M., and Lee, K. K. (2016) Chapter Fifteen - Isotope Labeling of Biomolecules: Structural Analysis of Viruses by HDX-MS, in *Methods in Enzymology* (Kelman, Z., Ed.), pp 405–426. Academic Press.

(78) Xiao, H., Hoerner, J. K., Eyles, S. J., Dobo, A., Voigtman, E., Mel'čuk, A. I., and Kaltashov, I. A. (2005) Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment. *Protein Sci. Publ. Protein Soc. 14*, 543–557.

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

(80) Morgan, C. R., and Engen, J. R. (2009) Investigating solution-phase protein structure and dynamics by hydrogen exchange mass spectrometry. *Curr. Protoc. Protein Sci. 58*.

(81) Kaltashov, I. A., Bobst, C. E., and Abzalimov, R. R. (2009) H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: Is there a need for a top-down approach? *Anal. Chem. 81*, 7892–7899.

(82) Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev. 25*, 158–170.

(83) Rosa, J. J., and Richards, F. M. (1979) An experimental procedure for increasing the structural resolution of chemical hydrogen-exchange measurements on proteins: Application to ribonuclease S peptide. *J. Mol. Biol. 133*, 399–416.

(84) Cravello, L., Lascoux, D., and Forest, E. (2003) Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. *Rapid Commun. Mass Spectrom. 17*, 2387–2393.

(85) Marcoux, J., Thierry, E., Vivès, C., Signor, L., Fieschi, F., and Forest, E. (2010) Investigating Alternative Acidic Proteases for H/D Exchange Coupled to Mass Spectrometry: Plasmepsin 2 but not Plasmepsin 4 Is Active Under Quenching Conditions. *J. Am. Soc. Mass Spectrom. 21*, 76–79.

(86) Wales, ThomasE., Eggertson, MichaelJ., and Engen, JohnR. (2013) Considerations in the Analysis of Hydrogen Exchange Mass Spectrometry Data, in *Mass Spectrometry Data Analysis in Proteomics* (Matthiesen, R., Ed.), pp 263–288. Humana Press.

(87) Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., and Hauer, C. R. (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. 83*, 6233–6237.

(88) Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., and Yates, J. R. (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods 1*, 39–45.

(89) Plumb, R. S., Johnson, K. A., Rainville, P., Smith, B. W., Wilson, I. D., Castro-Perez, J. M., and Nicholson, J. K. (2006) UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom. 20*, 1989–1994.

(90) Chik, J. K., Vande Graaf, J. L., and Schriemer, D. C. (2006) Quantitating the Statistical Distribution of Deuterium Incorporation To Extend the Utility of H/D Exchange MS Data. *Anal. Chem. 78*, 207–214.

(91) Zhang, J., Ramachandran, P., Kumar, R., and Gross, M. L. (2013) H/D Exchange Centroid Monitoring is Insufficient to Show Differences in the Behavior of Protein States. *J. Am. Soc. Mass Spectrom. 24*, 450–453.

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

(93) Zhang, Z., Guan, S., and Marshall, A. G. (1997) Enhancement of the effective resolution of mass spectra of high-mass biomolecules by maximum entropy-based deconvolution to eliminate the isotopic natural abundance distribution. *J. Am. Soc. Mass Spectrom. 8*, 659–670.

(94) Iacob, R. E., Murphy, J. P., and Engen, J. R. (2008) Ion mobility adds an additional dimension to mass spectrometric analysis of solution-phase hydrogen/deuterium exchange. *Rapid Commun. Mass Spectrom. 22*, 2898–2904.

(95) Engen, J. R., and Wales, T. E. (2015) Analytical Aspects of Hydrogen Exchange Mass Spectrometry. *Annu. Rev. Anal. Chem. 8*, 127–148.

(96) Rob, T. (2009) A versatile microfluidic chip for millisecond time-scale kinetic studies by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom. 20*, 124–130.

(97) Liuni, P., Rob, T., and Wilson, D. J. (2010) A microfluidic reactor for rapid, low-pressure proteolysis with on-chip electrospray ionization. *Rapid Commun Mass Spectrom 24*, 315–320.

(98) Rob, T., Liuni, P., Gill, P. K., Zhu, S., Balachandran, N., Berti, P. J., and Wilson, D. J. (2012) Measuring Dynamics in Weakly Structured Regions of Proteins Using Microfluidics-Enabled Subsecond H/D Exchange Mass Spectrometry. *Anal Chem 84*, 3771–3779.

(99) Zhu, S., Shala, A., Bezginov, A., Sljoka, A., Audette, G., and Wilson, D. J. (2015) Hyperphosphorylation of intrinsically disordered tau protein induces an amyloidogenic shift in its conformational ensemble. *PLoS ONE 10*, e0120416.

(100) Zhu, S., Khatun, R., Lento, C., Sheng, Y., and Wilson, D. J. (2017) Enhanced Binding Affinity via Destabilization of the Unbound State: A Millisecond Hydrogen–Deuterium Exchange Study of the Interaction between p53 and a Pleckstrin Homology Domain. *Biochemistry 56*, 4127–4133.

(101) Wilson, D. J., and Konermann, L. (2003) A Capillary Mixer with Adjustable Reaction Chamber Volume for Millisecond Time-Resolved Studies by Electrospray Mass Spectrometry. *Anal Chem 75*, 6408–6414.

(102) Liuni, P., Jeganathan, A., and Wilson, D. J. (2012) Conformer Selection and Intensified Dynamics During Catalytic Turnover in Chymotrypsin. *Angew Chem 124*, 9804–9807.

(103) Lísal, J., Kainov, D. E., Lam, T. T., Emmett, M. R., Wei, H., Gottlieb, P., Marshall, A. G., and Tuma, R. (2006) Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage. *Virology 351*, 73–79.

(104) Sowole, M. A., Simpson, S., Skovpen, Y. V., Palmer, D. R. J., and Konermann, L. (2016) Evidence of Allosteric Enzyme Regulation via Changes in Conformational Dynamics: A Hydrogen/Deuterium Exchange Investigation of Dihydrodipicolinate Synthase. *Biochemistry 55*, 5413–5422.

(105) Chandramohan, A., Krishnamurthy, S., Larsson, A., Nordlund, P., Jansson, A., and Anand, G. S. (2016) Predicting Allosteric Effects from Orthosteric Binding in Hsp90-Ligand Interactions: Implications for Fragment-Based Drug Design. *PLOS Comput. Biol. 12*, e1004840.

(106) Krishnamurthy, S., Veesler, D., Khayat, R., Snijder, J., Huang, R., Heck, A. J. R., Johnson, J., and Anand, G. S. (2014) Distinguishing direct binding interactions from allosteric effects in the protease–HK97 prohead I δ domain complex by amide H/D exchange mass spectrometry. *Bacteriophage 4*, e959816.

(107) Rob, T., Gill, P. K., Golemi-Kotra, D., and Wilson, D. J. (2013) An electrospray ms-coupled microfluidic device for sub-second hydrogen/deuterium exchange pulse-labelling reveals allosteric effects in enzyme inhibition. *Lab. Chip 13*, 2528–2532.

(108) Semrad, K. (2010) Proteins with RNA Chaperone Activity: A World of Diverse Proteins with a Common Task—Impediment of RNA Misfolding. *Biochem. Res. Int. 2011*, e532908.

(109) Woodson, S. A. (2010) Taming free energy landscapes with RNA chaperones. *RNA Biol 7*, 677–686.

(110) Tompa, P., and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J 18*, 1169–1175.

(111) Zúñiga, S., Sola, I., Cruz, J. L. G., and Enjuanes, L. (2009) Role of RNA chaperones in virus replication. *Virus Res. 139*, 253–266.

(112) Geller, R., Vignuzzi, M., Andino, R., and Frydman, J. (2007) Evolutionary constraints on chaperone-mediated folding provide an antiviral approach refractory to development of drug resistance. *Genes Dev. 21*, 195–205.

(113) Richter, J. D., and Sonenberg, N. (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature 433*, 477–480.

(114) Dasgupta, A., Das, S., Izumi, R., Venkatesan, A., and Barat, B. (2006) Targeting internal ribosome entry site (IRES)-mediated translation to block hepatitis C and other RNA viruses. *FEMS Microbiol. Lett. 234*, 189–199.

(115) Tong, S., Pan, J., and Tang, J. (2019) Study on the structure optimization and anti-hepatitis B virus activity of novel human La protein inhibitor HBSC11. *J. Med. Virol. 91*, 1818–1829.

(116) Vinayak, J., Marrella, S. A., Hussain, R. H., Rozenfeld, L., Solomon, K., and Bayfield, M. A. (2018) Human La binds mRNAs through contacts to the poly(A) tail. *Nucleic Acids Res. 46*, 4228–4240.

(117) Teplova, M., Yuan, Y.-R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A., and Patel, D. J. (2006) Structural Basis for Recognition and Sequestration of UUUOH 3′ Temini of Nascent RNA Polymerase III Transcripts by La, a Rheumatic Disease Autoantigen. *Mol. Cell 21*, 75–85.

(118) Elvin, J. G., Couston, R. G., and van der Walle, C. F. (2013) Therapeutic antibodies: Market considerations, disease targets and bioprocessing. *Int. J. Pharm. 440*, 83–98.

(119) Dranitsaris, G., Amir, E., and Dorward, K. (2011) Biosimilars of Biological Drug Therapies. *Drugs 71*, 1527–1536.

(120) Fuh, G. (2007) Synthetic antibodies as therapeutics. *Expert Opin. Biol. Ther. 7*, 73–87.

(121) Chames, P., Van Regenmortel, M., Weiss, E., and Baty, D. (2009) Therapeutic antibodies: successes, limitations and hopes for the future. *Br J Pharmacol 157*, 220–233.

(122) Bakhtiar, R. (2012) Therapeutic Recombinant Monoclonal Antibodies. *J Chem Educ 89*, 1537–1542.

(123) Beck, A., and Reichert, J. M. (2013) Approval of the first biosimilar antibodies in Europe. *mAbs 5*, 621–623.

(124) Declerck, P. J. (2013) Biosimilar monoclonal antibodies: a science-based regulatory challenge. *Expert Opin. Biol. Ther. 13*, 153–156.

(125) Mulcahy, A. W., Hlavka, J. P., and Case, S. R. (2018) Biosimilar Cost Savings in the United States. *Rand Health Q. 7*.

(126) Bui, L. A., Hurst, S., Finch, G. L., Ingram, B., Jacobs, I. A., Kirchhoff, C. F., Ng, C.-K., and Ryan, A. M. (2015) Key considerations in the preclinical development of biosimilars. *Drug Discov. Today 20*, 3–15.

(127) Woodson, S. A. (2010) Taming free energy landscapes with RNA chaperones. *RNA Biol. 7*, 677–686.

(128) Jones, S., Daley, D. T. A., Luscombe, N. M., Berman, H. M., and Thornton, J. M. (2001) Protein–RNA interactions: a structural analysis. *Nucleic Acids Res. 29*, 943–954.

(129) Resetca, D., and Wilson, D. J. (2013) Characterizing rapid, activity-linked conformational transitions in proteins via sub-second hydrogen deuterium exchange mass spectrometry. *FEBS J. 280*, 5616–5625.

(130) Rob, T., Gill, P. K., Golemi-Kotra, D., and Wilson, D. J. (2013) An electrospray ms-coupled microfluidic device for sub-second hydrogen/deuterium exchange pulse-labelling reveals allosteric effects in enzyme inhibition. *Lab. Chip 13*, 2528–2532.

(131) Konermann, L., Vahidi, S., and Sowole, M. A. (2014) Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules. *Anal. Chem. 86*, 213–232.

(132) Konermann, L., Pan, J., and Liu, Y.-H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev. 40*, 1224–1234.

(133) Rob, T., Liuni, P., Gill, P. K., Zhu, S., Balachandran, N., Berti, P. J., and Wilson, D. J. (2012) Measuring Dynamics in Weakly Structured Regions of Proteins Using Microfluidics-Enabled Subsecond H/D Exchange Mass Spectrometry. *Anal. Chem. 84*, 3771–3779.

(134) Lento, C., Audette, G. F., and Wilson, D. J. (2014) Time-resolved electrospray mass spectrometry — a brief history. *Can. J. Chem. 93*, 7–12.

(135) Wolin, S. L., and Cedervall, T. (2002) The La protein. *Annu. Rev. Biochem. 71*, 375–403.

(136) Bayfield, M. A., Yang, R., and Maraia, R. J. (2010) Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). *Biochim. Biophys. Acta 1799*, 365–378.

(137) Alspaugh, M. A., and Tan, E. M. (1975) Antibodies to cellular antigens in Sjögren's syndrome. *J. Clin. Invest. 55*, 1067–1073.

(138) Mattioli, M., and Reichlin, M. (1974) Heterogeneity of RNA protein antigens reactive with sera of patients with systemic lupus erythematosus description of a cytoplasmic nonribosomal antigen. *Arthritis Rheum. 17*, 421–429.

(139) Bousquet-Antonelli, C., and Deragon, J.-M. (2009) A comprehensive analysis of the La-motif protein superfamily. *RNA N. Y. N 15*, 750–764.

(140) Yoo, C. J., and Wolin, S. L. (1997) The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell 89*, 393–402.

(141) Huang, Y., Bayfield, M. A., Intine, R. V., and Maraia, R. J. (2006) Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nat. Struct. Mol. Biol. 13*, 611–618.

(142) Copela, L. A., Fernandez, C. F., Sherrer, R. L., and Wolin, S. L. (2008) Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. *RNA N. Y. N 14*, 1214–1227.

(143) Bayfield, M. A., Kaiser, T. E., Intine, R. V., and Maraia, R. J. (2007) Conservation of a masked nuclear export activity of La proteins and its effects on tRNA maturation. *Mol. Cell. Biol. 27*, 3303–3312.

(144) Brenet, F., Dussault, N., Borch, J., Ferracci, G., Delfino, C., Roepstorff, P., Miquelis, R., and Ouafik, L. (2005) Mammalian peptidylglycine alpha-amidating monooxygenase mRNA expression can be modulated by the La autoantigen. *Mol. Cell. Biol. 25*, 7505–7521.

(145) Meerovitch, K., Pelletier, J., and Sonenberg, N. (1989) A cellular protein that binds to the 5'-noncoding region of poliovirus RNA: implications for internal translation initiation. *Genes Dev. 3*, 1026–1034.

(146) Ali, N., Pruijn, G. J., Kenan, D. J., Keene, J. D., and Siddiqui, A. (2000) Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation. *J. Biol. Chem. 275*, 27531–27540.

(147) Kim, Y. K., and Jang, S. K. (1999) La protein is required for efficient translation driven by encephalomyocarditis virus internal ribosomal entry site. *J. Gen. Virol. 80 ( Pt 12)*, 3159–3166.

(148) Ray, P. S., and Das, S. (2002) La autoantigen is required for the internal ribosome entry site-mediated translation of Coxsackievirus B3 RNA. *Nucleic Acids Res. 30*, 4500–4508.

(149) Kim, Y. K., Back, S. H., Rho, J., Lee, S. H., and Jang, S. K. (2001) La autoantigen enhances translation of BiP mRNA. *Nucleic Acids Res. 29*, 5009–5016.

(150) Sommer, G., Dittmann, J., Kuehnert, J., Reumann, K., Schwartz, P. E., Will, H., Coulter, B. L., Smith, M. T., and Heise, T. (2011) The RNA-binding protein La contributes to cell proliferation and CCND1 expression. *Oncogene 30*, 434–444.

(151) Trotta, R., Vignudelli, T., Candini, O., Intine, R. V., Pecorari, L., Guerzoni, C., Santilli, G., Byrom, M. W., Goldoni, S., Ford, L. P., Caligiuri, M. A., Maraia, R. J., Perrotti, D., and Calabretta, B. (2003) BCR/ABL activates mdm2 mRNA translation via the La antigen. *Cancer Cell 3*, 145–160.

(152) Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P. D., Curry, S., and Conte, M. R. (2003) Structure of the C-Terminal Domain of Human La Protein Reveals a Novel RNA Recognition Motif Coupled to a Helical Nuclear Retention Element. *Structure 11*, 833–843.

(153) Teplova, M., Yuan, Y.-R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A., and Patel, D. J. (2006) Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol. Cell 21*, 75–85.

(154) Kotik-Kogan, O., Valentine, E. R., Sanfelice, D., Conte, M. R., and Curry, S. (2008) Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein. *Struct. Lond. Engl. 1993 16*, 852–862.

(155) Simons, F. H., Broers, F. J., Van Venrooij, W. J., and Pruijn, G. J. (1996) Characterization of cis-acting signals for nuclear import and retention of the La (SS-B) autoantigen. *Exp. Cell Res. 224*, 224–236.

(156) Goodier, J. L., Fan, H., and Maraia, R. J. (1997) A carboxy-terminal basic region controls RNA polymerase III transcription factor activity of human La protein. *Mol. Cell. Biol. 17*, 5823–5832.

(157) Fan, H., Goodier, J. L., Chamberlain, J. R., Engelke, D. R., and Maraia, R. J. (1998) 5' processing of tRNA precursors can Be modulated by the human La antigen phosphoprotein. *Mol. Cell. Biol. 18*, 3201–3211.

(158) Intine, R. V., Dundr, M., Misteli, T., and Maraia, R. J. (2002) Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs. *Mol. Cell 9*, 1113–1123.

(159) Schwartz, E. I., Intine, R. V., and Maraia, R. J. (2004) CK2 is responsible for phosphorylation of human La protein serine-366 and can modulate rpL37 5'-terminal oligopyrimidine mRNA metabolism. *Mol. Cell. Biol. 24*, 9580–9591.

(160) Horke, S., Reumann, K., Rang, A., and Heise, T. (2002) Molecular characterization of the human La protein.hepatitis B virus RNA.B interaction in vitro. *J. Biol. Chem. 277*, 34949–34958.

(161) Martino, L., Pennell, S., Kelly, G., Bui, T. T. T., Kotik-Kogan, O., Smerdon, S. J., Drake, A. F., Curry, S., and Conte, M. R. (2012) Analysis of the interaction with the hepatitis C virus mRNA reveals an alternative mode of RNA recognition by the human La protein. *Nucleic Acids Res. 40*, 1381–1394.

(162) Craig, A. W., Svitkin, Y. V., Lee, H. S., Belsham, G. J., and Sonenberg, N. (1997) The La autoantigen contains a dimerization domain that is essential for enhancing translation. *Mol. Cell. Biol. 17*, 163–169.

(163) Naeeni, A. R., Conte, M. R., and Bayfield, M. A. (2012) RNA chaperone activity of human La protein is mediated by variant RNA recognition motif. *J. Biol. Chem. 287*, 5472–5482.

(164) Kuehnert, J., Sommer, G., Zierk, A. W., Fedarovich, A., Brock, A., Fedarovich, D., and Heise, T. (2015) Novel RNA chaperone domain of RNA-binding protein La is regulated by AKT phosphorylation. *Nucleic Acids Res. 43*, 581–594.

(165) Chakshusmathi, G., Kim, S. D., Rubinson, D. A., and Wolin, S. L. (2003) A La protein requirement for efficient pre-tRNA folding. *EMBO J. 22*, 6562–6572.

(166) Belisova, A., Semrad, K., Mayer, O., Kocian, G., Waigmann, E., Schroeder, R., and Steiner, G. (2005) RNA chaperone activity of protein components of human Ro RNPs. *RNA N. Y. N 11*, 1084–1094.

(167) Bayfield, M. A., and Maraia, R. J. (2009) Precursor-product discrimination by La protein during tRNA metabolism. *Nat. Struct. Mol. Biol. 16*, 430–437.

(168) Hussain, R. H., Zawawi, M., and Bayfield, M. A. (2013) Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. *Nucleic Acids Res. 41*, 8715–8725.

(169) Singh, M., Wang, Z., Koo, B.-K., Patel, A., Cascio, D., Collins, K., and Feigon, J. (2012) Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol. Cell 47*, 16–26.

(170) Singh, M., Choi, C. P., and Feigon, J. (2013) xRRM: a new class of RRM found in the telomerase La family protein p65. *RNA Biol. 10*, 353–359.

(171) Prathapam, R., Witkin, K. L., O'Connor, C. M., and Collins, K. (2005) A telomerase holoenzyme protein enhances telomerase RNA assembly with telomerase reverse transcriptase. *Nat. Struct. Mol. Biol. 12*, 252–257.

(172) Maris, C., Dominguez, C., and Allain, F. H.-T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J. 272*, 2118–2131.

(173) Rob, T. (2009) A versatile microfluidic chip for millisecond time-scale kinetic studies by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom. 20*, 124–130.

(174) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.

(175) Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H., and Wagner, G. (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J. Biomol. NMR 52*, 315–327.

(176) Skinner, S. P., Goult, B. T., Fogh, R. H., Boucher, W., Stevens, T. J., Laue, E. D., and Vuister, G. W. (2015) Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallogr. D Biol. Crystallogr. 71*, 154–161.

(177) Martino, L., Pennell, S., Kelly, G., Bui, T. T. T., Kotik-Kogan, O., Smerdon, S. J., Drake, A. F., Curry, S., and Conte, M. R. (2012) Analysis of the interaction with the hepatitis C virus

mRNA reveals an alternative mode of RNA recognition by the human La protein. *Nucleic Acids Res. 40*, 1381–1394.

(178) Brown, K. A., Sharifi, S., Hussain, R., Donaldson, L., Bayfield, M. A., and Wilson, D. J. (2016) Distinct Dynamic Modes Enable the Engagement of Dissimilar Ligands in a Promiscuous Atypical RNA Recognition Motif. *Biochemistry 55*, 7141–7150.

(179) Laniel, M.-A., Beliveau, A., and Guerin, S. L. (2001) Electrophoretic mobility shift assays for the analysis of DNA-Protein interactions, in *DNA-Protein Interactions* 2nd ed., p 638. Humana Press.

(180) Konermann, L., and Douglas, D. J. (1998) Unfolding of proteins monitored by electrospray ionization mass spectrometry: a comparison of positive and negative ion modes. *J. Am. Soc. Mass Spectrom. 9*, 1248–1254.

(181) Scarff, C. A., Thalassinos, K., Hilton, G. R., and Scrivens, J. H. (2008) Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom. 22*, 3297–3304.

(182) Ruotolo, B. T., Hyung, S.-J., Robinson, P. M., Giles, K., Bateman, R. H., and Robinson, C. V. (2007) Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chem. Int. Ed Engl. 46*, 8001–8004.

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

(184) Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol. 5*, 789–796.

(185) Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*, 1647–1651.

(186) Maris, C., Dominguez, C., and Allain, F. H.-T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J. 272*, 2118–2131.

(187) Wolin, S. L., and Cedervall, and T. (2002) The La Protein. *Annu. Rev. Biochem. 71*, 375–403.

(188) Fok, V., Friend, K., and Steitz, J. A. (2006) Epstein-Barr virus noncoding RNAs are confined to the nucleus, whereas their partner, the human La protein, undergoes nucleocytoplasmic shuttling. *J. Cell Biol. 173*, 319–325.

(189) Meerovitch, K., Svitkin, Y. V., Lee, H. S., Lejbkowicz, F., Kenan, D. J., Chan, E. K., Agol, V. I., Keene, J. D., and Sonenberg, N. (1993) La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. *J. Virol. 67*, 3798–3807.

(190) Zhang, J., Dinh, T. N., Kappeler, K., Tsaprailis, G., and Chen, Q. M. (2012) La autoantigen mediates oxidant induced de novo Nrf2 protein translation. *Mol. Cell. Proteomics MCP 11*, M111.015032.

(191) Kotik-Kogan, O., Valentine, E. R., Sanfelice, D., Conte, M. R., and Curry, S. (2008) Structural Analysis Reveals Conformational Plasticity in the Recognition of RNA 3′ Ends by the Human La Protein. *Structure 16*, 852–862.

(192) Teplova, M., Yuan, Y.-R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A., and Patel, D. J. (2006) Structural Basis for Recognition and Sequestration of UUUOH 3′ Temini of Nascent RNA Polymerase III Transcripts by La, a Rheumatic Disease Autoantigen. *Mol. Cell 21*, 75–85.

(193) Ali, N., Pruijn, G. J. M., Kenan, D. J., Keene, J. D., and Siddiqui, A. (2000) Human La antigen is required for the hepatitis C virus internal ribosome entry site (IRES)-mediated translation. *J. Biol. Chem.*

(194) Meerovitch, K., Pelletier, J., and Sonenberg, N. (1989) A cellular protein that binds to the 5'-noncoding region of poliovirus RNA: implications for internal translation initiation. *Genes Dev. 3*, 1026–1034.

(195) Intine, R. V., Dundr, M., Misteli, T., and Maraia, R. J. (2002) Aberrant Nuclear Trafficking of La Protein Leads to Disordered Processing of Associated Precursor tRNAs. *Mol. Cell 9*, 1113–1123.

(196) Bayfield, M. A., Kaiser, T. E., Intine, R. V., and Maraia, R. J. (2007) Conservation of a masked nuclear export activity of La proteins and its effects on tRNA maturation. *Mol. Cell. Biol. 27*, 3303–3312.

(197) Allison, T. M., Reading, E., Liko, I., Baldwin, A. J., Laganowsky, A., and Robinson, C. V. (2015) Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nat. Commun. 6*, 8551.

(198) Rey, M., Sarpe, V., Burns, K. M., Buse, J., Baker, C. A. H., van Dijk, M., Wordeman, L., Bonvin, A. M. J. J., and Schriemer, D. C. (2014) Mass Spec Studio for Integrative Structural Biology. *Structure 22*, 1538–1548.

(199) Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P. D., Curry, S., and Conte, M. R. (2003) Structure of the C-Terminal Domain of Human La Protein Reveals a Novel RNA Recognition Motif Coupled to a Helical Nuclear Retention Element. *Structure 11*, 833–843.

(200) Alfano, C., Sanfelice, D., Babon, J., Kelly, G., Jacks, A., Curry, S., and Conte, M. R. (2004) Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nat Struct Mol Biol 11*, 323–329.

(201) Kotik-Kogan, O., Valentine, E. R., Sanfelice, D., Conte, M. R., and Curry, S. (2008) Structural Analysis Reveals Conformational Plasticity in the Recognition of RNA 3′ Ends by the Human La Protein. *Structure 16*, 852–862.

(202) Hopper, J. T. S., and Oldham, N. J. (2009) Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *J. Am. Soc. Mass Spectrom. 20*, 1851–1858.

(203) Eschweiler, J. D., Rabuck-Gibbons, J. N., Tian, Y., and Ruotolo, B. T. (2015) CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Anal. Chem. 87*, 11516–11522.

(204) Scarff, C. A., Thalassinos, K., Hilton, G. R., and Scrivens, J. H. (2008) Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom. 22*, 3297–3304.

(205) Marrella, S. A., Brown, K. A., Mansouri-Noori, F., Porat, J., Wilson, D. J., and Bayfield, M. A. (2019) An interdomain bridge influences RNA binding of the human La protein. *J. Biol. Chem. 294*, 1529–1540.

(206) Bakhtiar, R. (2012) Therapeutic Recombinant Monoclonal Antibodies. *J. Chem. Educ. 89*, 1537–1542.

(207) Chames, P., Van Regenmortel, M., Weiss, E., and Baty, D. (2009) Therapeutic antibodies: successes, limitations and hopes for the future. *Br. J. Pharmacol. 157*, 220–233.

(208) McCamish, M., and Woollett, G. (2011) Worldwide experience with biosimilar development. *mAbs 3*, 209–217.

(209) Berkowitz, S. A., Engen, J. R., Mazzeo, J. R., and Jones, G. B. (2012) Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat. Rev. Drug Discov. Lond. 11*, 527–40.

(210) Patel, T. P., Parekh, R. B., Moellering, B. J., and Prior, C. P. (1992) Different culture methods lead to differences in glycosylation of a murine IgG monoclonal antibody. *Biochem. J. 285*, 839–845.

(211) Beck, A., Sanglier-Cianférani, S., and Van Dorsselaer, A. (2012) Biosimilar, Biobetter, and Next Generation Antibody Characterization by Mass Spectrometry. *Anal. Chem. 84*, 4637–4646.

(212) Serrato, J. A., Hernández, V., Estrada-Mondaca, S., Palomares, L. A., and Ramírez, O. T. (2007) Differences in the glycosylation profile of a monoclonal antibody produced by hybridomas cultured in serum-supplemented, serum-free or chemically defined media. *Biotechnol. Appl. Biochem. 47*, 113–124.

(213) Zhang, J., and Robinson, D. (2005) Development of Animal-free, Protein-Free and Chemically-Defined Media for NS0 Cell Culture. *Cytotechnology 48*, 59–74.

(214) Beck, A., Wagner-Rousset, E., Bussat, M.-C., Lokteff, M., Klinguer-Hamour, C., Haeuw, J.-F., Goetsch, L., Wurch, T., Van Dorsselaer, A., and Corvaïa, N. (2008) Trends in glycosylation,

glycoanalysis and glycoengineering of therapeutic antibodies and Fc-fusion proteins. *Curr. Pharm. Biotechnol. 9*, 482–501.

(215) Cho, I. H., Lee, N., Song, D., Jung, S. Y., Bou-Assaf, G., Sosic, Z., Zhang, W., and Lyubarskaya, Y. (2016) Evaluation of the structural, physicochemical, and biological characteristics of SB4, a biosimilar of etanercept. *mAbs 8*, 1136–1155.

(216) Muller, Y. A., Chen, Y., Christinger, H. W., Li, B., Cunningham, B. C., Lowman, H. B., and de Vos, A. M. (1998) VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 å resolution and mutational analysis of the interface. *Structure 6*, 1153–1167.

(217) Presta, L. G., Chen, H., O'Connor, S. J., Chisholm, V., Meng, Y. G., Krummen, L., Winkler, M., and Ferrara, N. (1997) Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res. 57*, 4593–4599.

(218) Ferrara, N., Hillan, K. J., Gerber, H.-P., and Novotny, W. (2004) Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat Rev Drug Discov 3*, 391–400.

(219) Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., Berlin, J., Baron, A., Griffing, S., Holmgren, E., Ferrara, N., Fyfe, G., Rogers, B., Ross, R., and Kabbinavar, F. (2004) Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med. 350*, 2335–2342.

(220) Johnson, D. H., Fehrenbacher, L., Novotny, W. F., Herbst, R. S., Nemunaitis, J. J., Jablons, D. M., Langer, C. J., DeVore, R. F., Gaudreault, J., Damico, L. A., Holmgren, E., and Kabbinavar, F. (2004) Randomized Phase II Trial Comparing Bevacizumab Plus Carboplatin and Paclitaxel With Carboplatin and Paclitaxel Alone in Previously Untreated Locally Advanced or Metastatic Non-Small-Cell Lung Cancer. *J. Clin. Oncol. 22*, 2184–2191.

(221) Escudier, B., Pluzanska, A., Koralewski, P., Ravaud, A., Bracarda, S., Szczylik, C., Chevreau, C., Filipek, M., Melichar, B., Bajetta, E., Gorbunova, V., Bay, J.-O., Bodrogi, I., Jagiello-Gruszfeld, A., and Moore, N. (2007) Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *The Lancet 370*, 2103–2111.

(222) Burger, R. A., Brady, M. F., Bookman, M. A., Fleming, G. F., Monk, B. J., Huang, H., Mannel, R. S., Homesley, H. D., Fowler, J., Greer, B. E., Boente, M., Birrer, M. J., and Liang, S. X. (2011) Incorporation of Bevacizumab in the Primary Treatment of Ovarian Cancer. *N. Engl. J. Med. 365*, 2473–2483.

(223) Robert, N. J., Diéras, V., Glaspy, J., Brufsky, A. M., Bondarenko, I., Lipatov, O. N., Perez, E. A., Yardley, D. A., Chan, S. Y. T., Zhou, X., Phan, S.-C., and O'Shaughnessy, J. (2011) RIBBON-1: Randomized, Double-Blind, Placebo-Controlled, Phase III Trial of Chemotherapy With or Without Bevacizumab for First-Line Treatment of Human Epidermal Growth Factor Receptor 2–Negative, Locally Recurrent or Metastatic Breast Cancer. *J. Clin. Oncol. 29*, 1252–1260.

(224) Beck, A., Debaene, F., Diemer, H., Wagner-Rousset, E., Colas, O., Dorsselaer, A. V., and Cianférani, S. (2015) Cutting-edge mass spectrometry characterization of originator, biosimilar and biobetter antibodies. *J. Mass Spectrom. 50*, 285–297.

(225) Beck, A., Wagner-Rousset, E., Ayoub, D., Van Dorsselaer, A., and Sanglier-Cianférani, S. (2013) Characterization of Therapeutic Antibodies and Related Products. *Anal. Chem. 85*, 715–736.

(226) Thompson, N. J., Rosati, S., and Heck, A. J. R. (2014) Performing native mass spectrometry analysis on therapeutic antibodies. *Methods 65*, 11–17.

(227) Tassi, M., Vos, J., Chatterjee, S., Sobott, F., Bones, J., and Eeltink, S. (2018) Advances in native high-performance liquid chromatography and intact mass spectrometry for the characterization of biopharmaceutical products. *J. Sep. Sci. 41*, 125–144.

(228) Felitsyn, N., Peschke, M., and Kebarle, P. (2002) Origin and number of charges observed on multiply-protonated native proteins produced by ESI. *Int. J. Mass Spectrom. 219*, 39–62.

(229) Ferguson, C. N., and Gucinski-Ruth, A. C. (2016) Evaluation of Ion Mobility-Mass Spectrometry for Comparative Analysis of Monoclonal Antibodies. *J. Am. Soc. Mass Spectrom. 27*, 822–833.

(230) Houde, D., Berkowitz, S. A., and Engen, J. R. (2011) The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. *J. Pharm. Sci. 100*, 2071–2086.

(231) Huang, R. Y.-C., and Chen, G. (2014) Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry. *Anal. Bioanal. Chem. 406*, 6541–6558.

(232) Pan, J., Zhang, S., and Borchers, C. H. (2016) Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry. *Biochim. Biophys. Acta BBA - Proteins Proteomics 1864*, 1801–1808.

(233) Pan, J., Zhang, S., Chou, A., and H. Borchers, C. (2016) Higher-order structural interrogation of antibodies using middle-down hydrogen/deuterium exchange mass spectrometry. *Chem. Sci. 7*, 1480–1486.

(234) Atmanene, C., Wagner-Rousset, E., Malissard, M., Chol, B., Robert, A., Corvaïa, N., Dorsselaer, A. V., Beck, A., and Sanglier-Cianférani, S.-. (2009) Extending Mass Spectrometry Contribution to Therapeutic Monoclonal Antibody Lead Optimization: Characterization of Immune Complexes Using Noncovalent ESI-MS. *Anal. Chem. 81*, 6364–6373.

(235) Tian, Y., Han, L., Buckner, A. C., and Ruotolo, B. T. (2015) Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Anal. Chem. 87*, 11509–11515.

(236) Damen, C. W. N., Chen, W., Chakraborty, A. B., van Oosterhout, M., Mazzeo, J. R., Gebler, J. C., Schellens, J. H. M., Rosing, H., and Beijnen, J. H. (2009) Electrospray Ionization Quadrupole Ion-Mobility Time-of-Flight Mass Spectrometry as a Tool to Distinguish the Lot-to-Lot Heterogeneity in N-Glycosylation Profile of the Therapeutic Monoclonal Antibody Trastuzumab. *J. Am. Soc. Mass Spectrom. 20*, 2021–2033.

(237) Watanabe, Y., Vasiljevic, S., Allen, J. D., Seabright, G. E., Duyvesteyn, H. M. E., Doores, K. J., Crispin, M., and Struwe, W. B. (2018) Signature of Antibody Domain Exchange by Native Mass Spectrometry and Collision-Induced Unfolding. *Anal. Chem. 90*, 7325–7331.

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

(239) Zhu, S., Shala, A., Bezginov, A., Sljoka, A., Audette, G., and Wilson, D. J. (2015) Hyperphosphorylation of intrinsically disordered tau protein induces an amyloidogenic shift in its conformational ensemble. *PloS One 10*, e0120416.

(240) Rob, T., Liuni, P., Gill, P. K., Zhu, S., Balachandran, N., Berti, P. J., and Wilson, D. J. (2012) Measuring Dynamics in Weakly Structured Regions of Proteins Using Microfluidics-Enabled Subsecond H/D Exchange Mass Spectrometry. *Anal. Chem. 84*, 3771–3779.

(241) Liuni, P., Jeganathan, A., and Wilson, D. J. (2012) Conformer Selection and Intensified Dynamics During Catalytic Turnover in Chymotrypsin. *Angew. Chem. 124*, 9804–9807.

(242) Brown, K. A., Rajendran, S., Dowd, J., and Wilson, D. J. (2019) Rapid characterization of structural and functional similarity for a candidate bevacizumab (Avastin) biosimilar using a multipronged mass-spectrometry-based approach. *Drug Test. Anal. 11*, 1207–1217.

(243) Langmuir, I. (1918) THE ADSORPTION OF GASES ON PLANE SURFACES OF GLASS, MICA AND PLATINUM. *J. Am. Chem. Soc. 40*, 1361–1403.

(244) Schuck, P. (1997) Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. *Annu. Rev. Biophys. Biomol. Struct. 26*, 541–566.

(245) Yang, Y., Wang, G., Song, T., Lebrilla, C. B., and Heck, A. J. R. (2017) Resolving the micro-heterogeneity and structural integrity of monoclonal antibodies by hybrid mass spectrometric approaches. *mAbs 9*, 638–645.

(246) Xie, H., Chakraborty, A., Ahn, J., Yu, Y. Q., Dakshinamoorthy, D. P., Gilar, M., Chen, W., Skilton, S. J., and Mazzeo, J. R. (2010) Rapid comparison of a candidate biosimilar to an innovator monoclonal antibody with advanced liquid chromatography and mass spectrometry technologies. *mAbs 2*, 379–394.

(247) Liu, L. (2015) Antibody Glycosylation and Its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins. *J. Pharm. Sci. 104*, 1866–1884.

(248) Seo, N., Polozova, A., Zhang, M., Yates, Z., Cao, S., Li, H., Kuhns, S., Maher, G., McBride, H. J., and Liu, J. (2018) Analytical and functional similarity of Amgen biosimilar ABP 215 to bevacizumab. *mAbs 0*, 1–14.

(249) Iacob, R. E., Murphy, J. P., and Engen, J. R. (2008) Ion mobility adds an additional dimension to mass spectrometric analysis of solution-phase hydrogen/deuterium exchange. *Rapid Commun. Mass Spectrom. 22*, 2898–2904.

(250) Bleiholder, C., Dupuis, N. F., Wyttenbach, T., and Bowers, M. T. (2011) Ion mobility–mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation. *Nat. Chem. 3*, 172–177.

(251) Smith, D. P., Giles, K., Bateman, R. H., Radford, S. E., and Ashcroft, A. E. (2007) Monitoring Copopulated Conformational States During Protein Folding Events Using Electrospray Ionization-Ion Mobility Spectrometry-Mass Spectrometry. *J. Am. Soc. Mass Spectrom. 18*, 2180–2190.

(252) Wilson, D. J., and Konermann, L. (2003) A Capillary Mixer with Adjustable Reaction Chamber Volume for Millisecond Time-Resolved Studies by Electrospray Mass Spectrometry. *Anal. Chem. 75*, 6408–6414.

(253)  ICH Official web site : ICH.

(254) Fang, J., Doneanu, C., Jr, W. R. A., Yu, Y. Q., Beck, A., and Chen, W. (2016) Advanced assessment of the physicochemical characteristics of Remicade® and Inflectra® by sensitive LC/MS techniques. *mAbs 8*, 1021–1034.

(255) Colbert, A., Umble-Romero, A., Prokop, S., Chow, V. F.-S., Wong, T., DeSimone, D., Zhou, L., and Pederson, S. (2014) Bioanalytical strategy used in development of pharmacokinetic (PK) methods that support biosimilar programs. *mAbs 6*, 1178–1189.

(256) Marsh, J. A., and Teichmann, S. A. (2015) Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem. 84*, 551–575.

(257) Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., and van Gunsteren, W. F. (1985) A protein structure from nuclear magnetic resonance data: lac Repressor headpiece. *J. Mol. Biol. 182*, 179–182.

(258) Bondarenko, P. V., Second, T. P., Zabrouskov, V., Makarov, A. A., and Zhang, Z. (2009) Mass Measurement and Top-Down HPLC/MS Analysis of Intact Monoclonal Antibodies on a Hybrid Linear Quadrupole Ion Trap–Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom. 20*, 1415–1424.

(259) Sanglier, S., Leize, E., Dorsselaer, A., and Zal, F. (2003) Comparative ESI-MS study of ~2.2 MDa native hemocyanins from deep-sea and shore crabs: from protein oligomeric state to biotope. *J. Am. Soc. Mass Spectrom. 14*, 419–429.

(260) Zal, F., Chausson, F., Leize, E., Van Dorsselaer, A., Lallier, F. H., and Green, B. N. (2002) Quadrupole Time-of-Flight Mass Spectrometry of the Native Hemocyanin of the Deep-Sea Crab Bythograea thermydron. *Biomacromolecules 3*, 229–231.

(261) Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., Xie, Y., Kaufman, S. L., and Chernushevich, I. V. (2005) Electrospray Ionization Mass Spectrometry and Ion Mobility Analysis of the 20S Proteasome Complex. *J. Am. Soc. Mass Spectrom. 16*, 998–1008.

(262) Rosati, S., Yang, Y., Barendregt, A., and Heck, A. J. R. (2014) Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc. 9*, 967–976.

(263) Pisupati, K., Tian, Y., Okbazghi, S., Benet, A., Ackermann, R., Ford, M., Saveliev, S., Hosfield, C. M., Urh, M., Carlson, E., Becker, C., Tolbert, T. J., Schwendeman, S. P., Ruotolo, B. T., and Schwendeman, A. (2017) A Multidimensional Analytical Comparison of Remicade and the Biosimilar Remsima. *Anal. Chem. 89*, 4838–4846.

(264) Marcoux, J., Champion, T., Colas, O., Wagner-Rousset, E., Corvaïa, N., Dorsselaer, A. V., Beck, A., and Cianférani, S. (2015) Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate. *Protein Sci. 24*, 1210–1223.

(265) Campuzano, I. D. G., Netirojjanakul, C., Nshanian, M., Lippens, J. L., Kilgour, D. P. A., Van Orden, S., and Loo, J. A. (2018) Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem. 90*, 745–751.

(266) Rosati, S., Bremer, E. T. van den, Schuurman, J., Parren, P. W., Kamerling, J. P., and Heck, A. J. (2013) In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *mAbs 5*, 917–924.

(267) Zheng, K., Bantog, C., and Bayer, R. (2011) The impact of glycosylation on monoclonal antibody conformation and stability. *mAbs 3*, 568–576.

(268) Jefferis, R. (2009) Recombinant antibody therapeutics: the impact of glycosylation on mechanisms of action. *Trends Pharmacol. Sci. 30*, 356–362.

(269) Lemery, S. J., Ricci, M. S., Keegan, P., McKee, A. E., and Pazdur, R. (2017) FDA's Approach to Regulating Biosimilars. *Clin. Cancer Res. 23*, 1882–1885.

(270) Hoeben, A., Landuyt, B., Highley, M. S., Wildiers, H., Oosterom, A. T. V., and Bruijn, E. A. D. (2004) Vascular Endothelial Growth Factor and Angiogenesis. *Pharmacol. Rev. 56*, 549–580.

(271) Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K. S., Powell-Braxton, L., Hillan, K. J., and Moore, M. W. (1996) Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature 380*, 439–442.

(272) Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C., Declercq, C., Pawling, J., Moons, L., Collen, D., Risau, W., and Nagy, A. (1996) Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature 380*, 435–439.

(273) Ferrara, N., and Davis-Smyth, T. (1997) The Biology of Vascular Endothelial Growth Factor. *Endocr. Rev. 18*, 4–25.

(274) Schellekens, H. (2005) Follow-on biologics: challenges of the 'next generation.' *Nephrol. Dial. Transplant. 20*, iv31–iv36.

(275) Wei, H., Mo, J., Tao, L., Russell, R. J., Tymiak, A. A., Chen, G., Iacob, R. E., and Engen, J. R. (2014) Hydrogen/Deuterium Exchange Mass Spectrometry for Probing Higher Order Structure of Protein Therapeutics: Methodology and Applications. *Drug Discov. Today 19*, 95.

(276) Deng, B., Lento, C., and Wilson, D. J. (2016) Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development – A review. *Anal. Chim. Acta 940*, 8–20.

(277) Rob, T. (2009) A versatile microfluidic chip for millisecond time-scale kinetic studies by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom. 20*, 124–130.

(278) Rob, T., Gill, P. K., Golemi-Kotra, D., and Wilson, D. J. (2013) An electrospray ms-coupled microfluidic device for sub-second hydrogen/deuterium exchange pulse-labelling reveals allosteric effects in enzyme inhibition. *Lab. Chip 13*, 2528–2532.

(279) Deng, B., Zhu, S., Macklin, A. M., Xu, J., Lento, C., Sljoka, A., and Wilson, D. J. (2017) Suppressing allostery in epitope mapping experiments using millisecond hydrogen / deuterium exchange mass spectrometry. *mAbs 9*, 1327–1336.

(280) Chandramohan, A., Krishnamurthy, S., Larsson, A., Nordlund, P., Jansson, A., and Anand, G. S. (2016) Predicting Allosteric Effects from Orthosteric Binding in Hsp90-Ligand Interactions: Implications for Fragment-Based Drug Design. *PLOS Comput. Biol. 12*, e1004840.

(281) Keyt, B. A., Berleau, L. T., Nguyen, H. V., Chen, H., Heinsohn, H., Vandlen, R., and Ferrara, N. (1996) The Carboxyl-terminal Domain(111165) of Vascular Endothelial Growth Factor Is Critical for Its Mitogenic Potency. *J. Biol. Chem. 271*, 7788–7795.

(282) Krishnamurthy, S., Veesler, D., Khayat, R., Snijder, J., Huang, R., Heck, A. J. R., Johnson, J., and Anand, G. S. (2014) Distinguishing direct binding interactions from allosteric effects in the protease–HK97 prohead I δ domain complex by amide H/D exchange mass spectrometry. *Bacteriophage 4*, e959816.

(283) McAllister, R. G., and Konermann, L. (2015) Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective. *Biochemistry 54*, 2683–2692.

(284) Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010) Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Mol. Cell. Proteomics 9*, 1634–1649.

(285) Petrotchenko, E. V., and Borchers, C. H. (2010) Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev. 29*, 862–876.

(286) Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev. 25*, 663–682.

(287) Deng, B., Zhu, S., Macklin, A. M., Xu, J., Lento, C., Sljoka, A., and Wilson, D. J. (2017) Suppressing allostery in epitope mapping experiments using millisecond hydrogen / deuterium exchange mass spectrometry. *mAbs 9*, 1327–1336.

# Appendices

## Appendix A: Supplementary information for Chapter 2
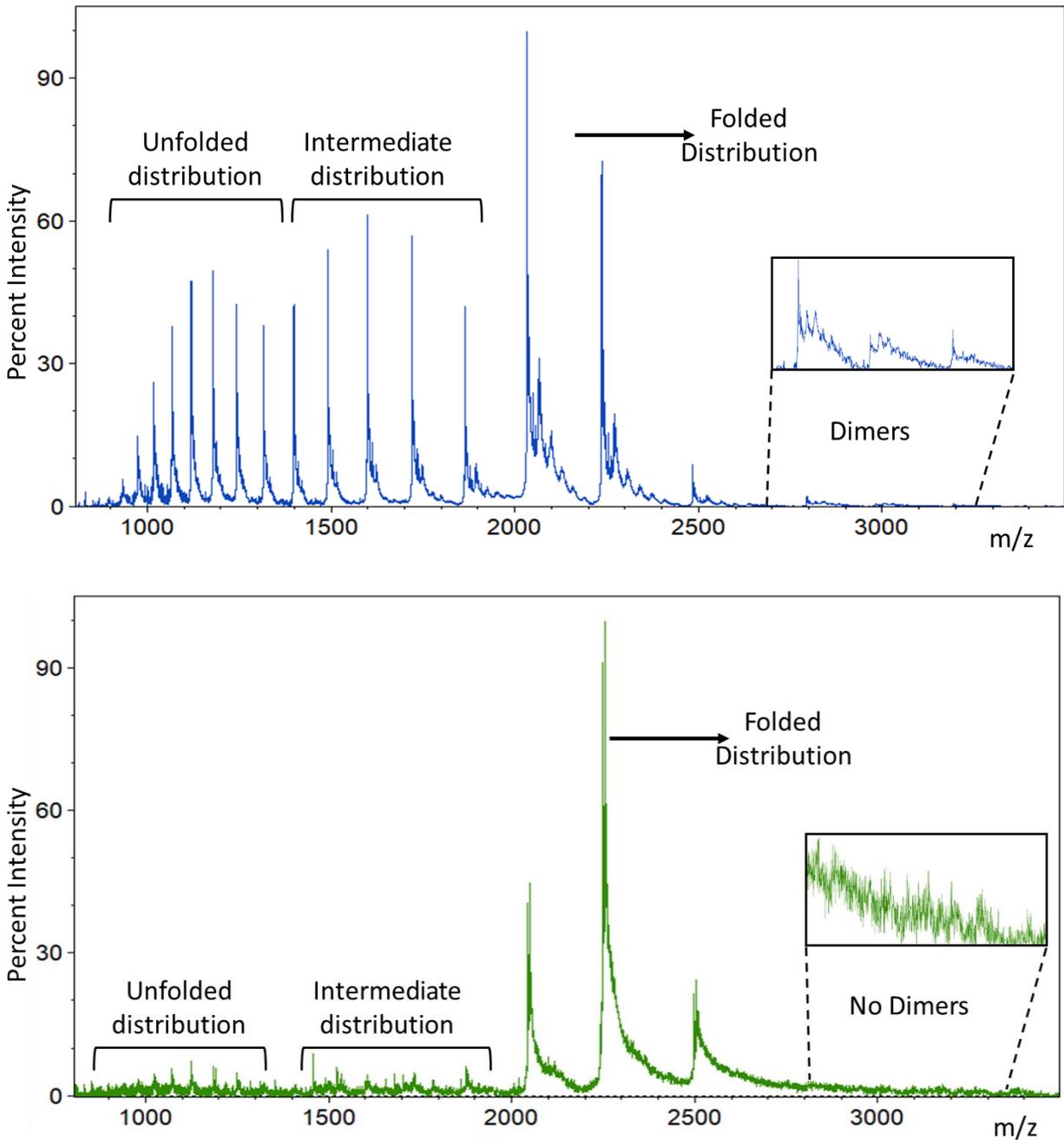


Figure A1. Native Mass Spectrum of hLa-CTD acquired in 10 mM $CH_3COONH_4$ (top) and 250 mM $CH_3COONH_4$ (bottom).
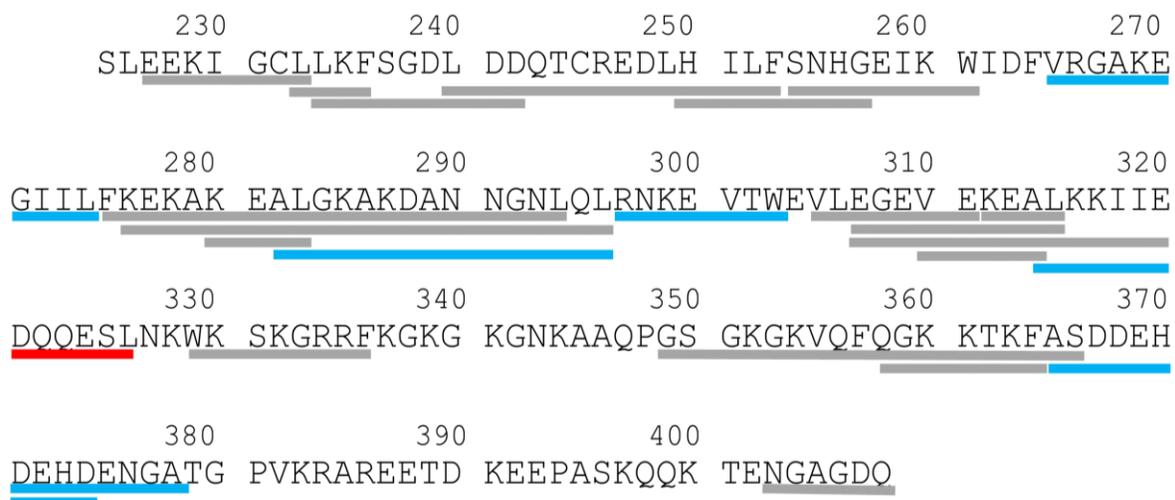
Figure A2. Sequence coverage map of hLa-CTD in complex with 23-mer ssRNA with a 76% sequence coverage (top) and IRES domain IV of HCV RNA with a 64% sequence coverage (bottom). Blue lines represent peptides with significant decreases upon RNA binding, red lines represent significant increases upon RNA binding and grey lines represent no changes.

```
              10         20         30         40         50
MAENGDNEKM AALEAKICHQ IEYYFGDFNL PRDKFLKEQI KLDEGWVPLE

              60         70         80         90        100
IMIKFNRLNR LTTDFNVIVE ALSKSKAELM EISEDKTKIR RSPSKPLPEV

             110        120        130        140        150
TDEYKNDVKN RSVYIKGFPT DATLDDIKEW LEDKGQVLNI QMRRTLHKAF

             160        170        180        190        200
KGSIFVVFDS IESAKKFVET PGQKYKETDL LILFKDDYFA KKNEERKQNK

             210        220        230        240        250
VEAKLRAKQE QEAKQKLEED AEMKSLEEKI GCLLKFSGDL DDQTCREDLH

             260        270        280        290        300
ILFSNHGEIK WIDFVRGAKE GIILFKEKAK EALGKAKDAN NGNLQLRNKE

             310        320        330        340        350
VTWEVLEGEV EKEALKKIIE DQQESLNKWK SKGRRFKGKG KGNKAAQPGS

             360        370        380        390        400
GKGKVQFQGK KTKFASDDEH DEHDENGATG PVKRAREETD KEEPASKQQK
```

hLa
hLadNRE
hLa K316E/K317E
hLa D132K/E133K/K316E/K317E

```
TENGAGDQ
```

Figure B1. Sequence coverage map of hLa, hLaΔNRE, hLa K316E/K317E and hLa D132K/E133K/K316E/K317E used in HDX experiments. Covered regions included at least one peptide unambiguously identified by MS/MS.

hLa + U10 RNA - 0.59 s



hLa + A20 RNA - 0.59 s

hLa+U10 - 2.09 s

hLa + A20 - 2.09 s
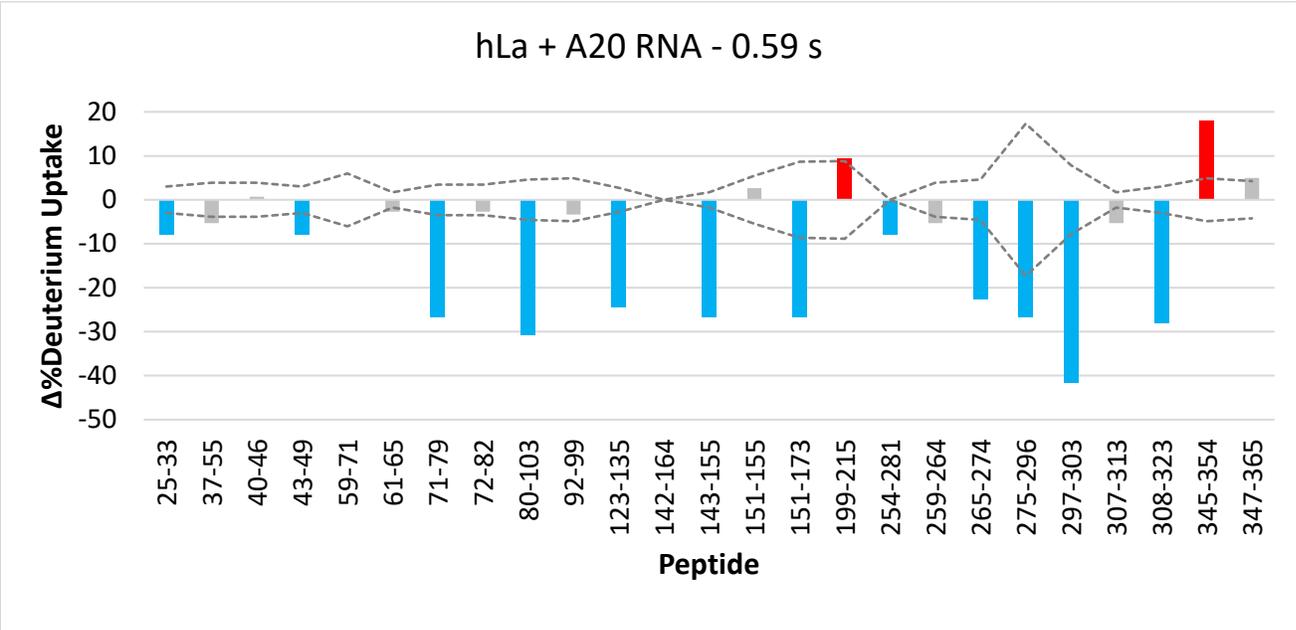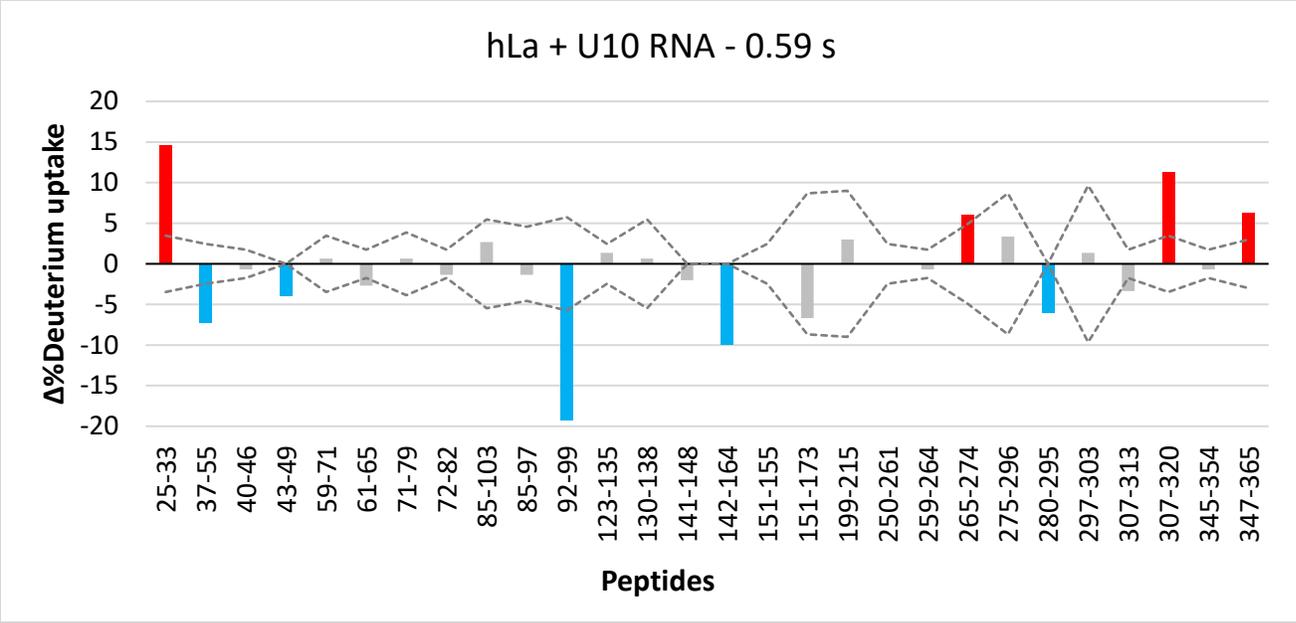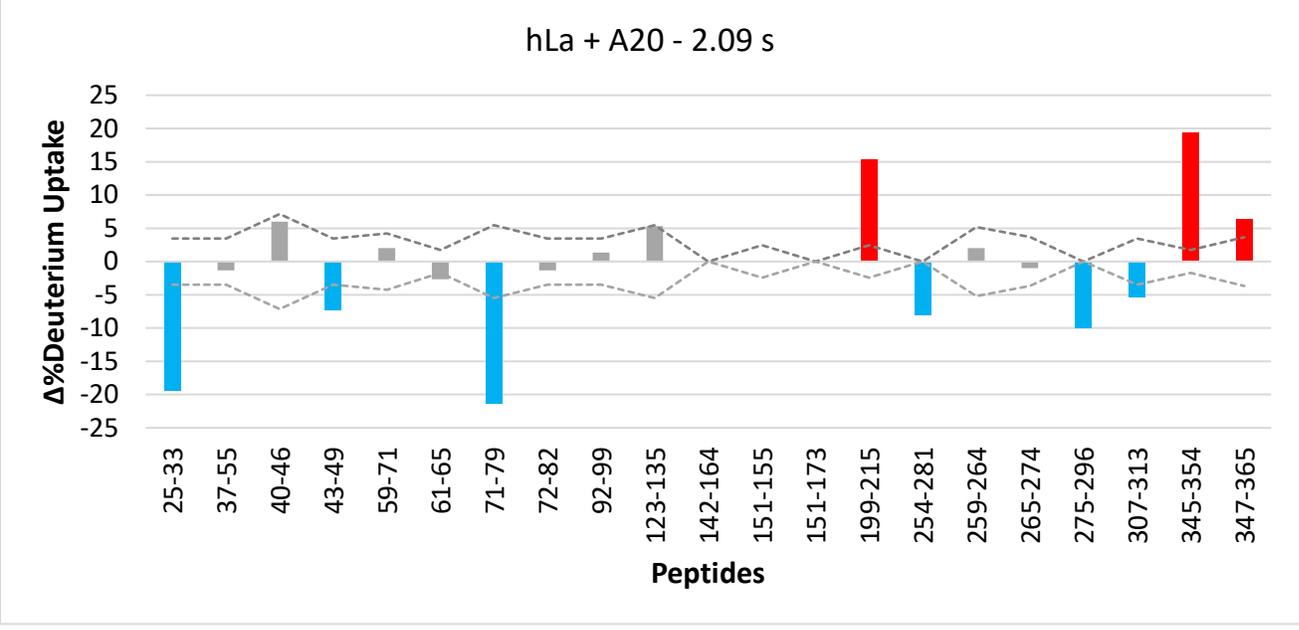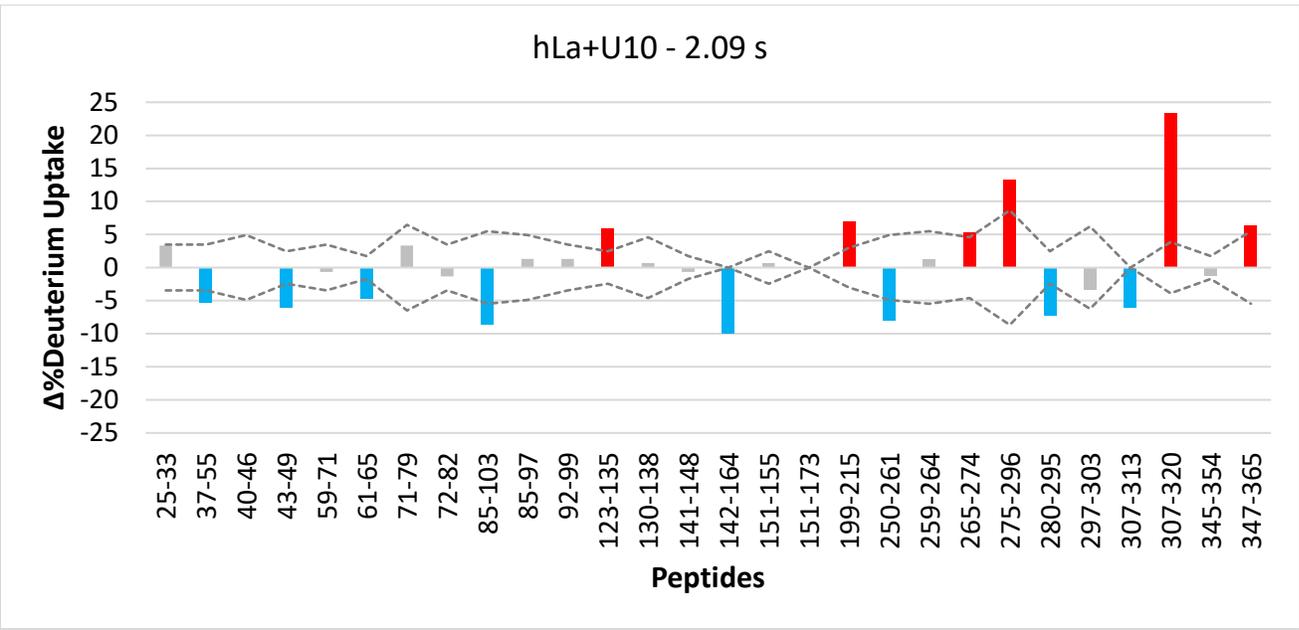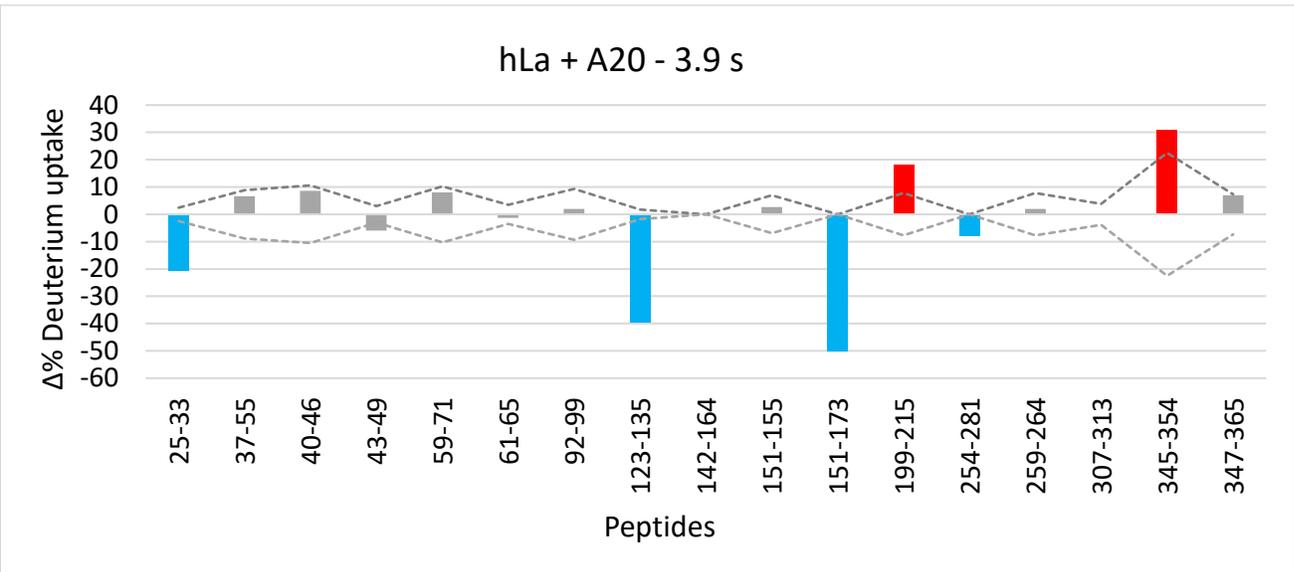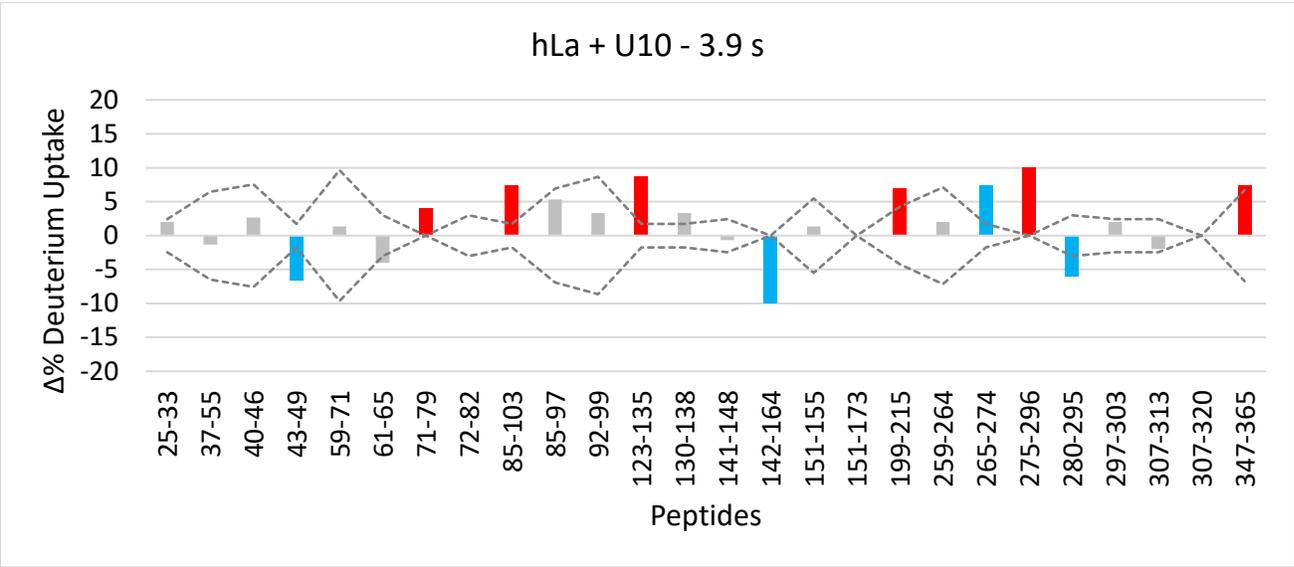
Figure B2: Preliminary HDX results of hLa in complex with U10 and A20 RNA at three different HDX reaction times.
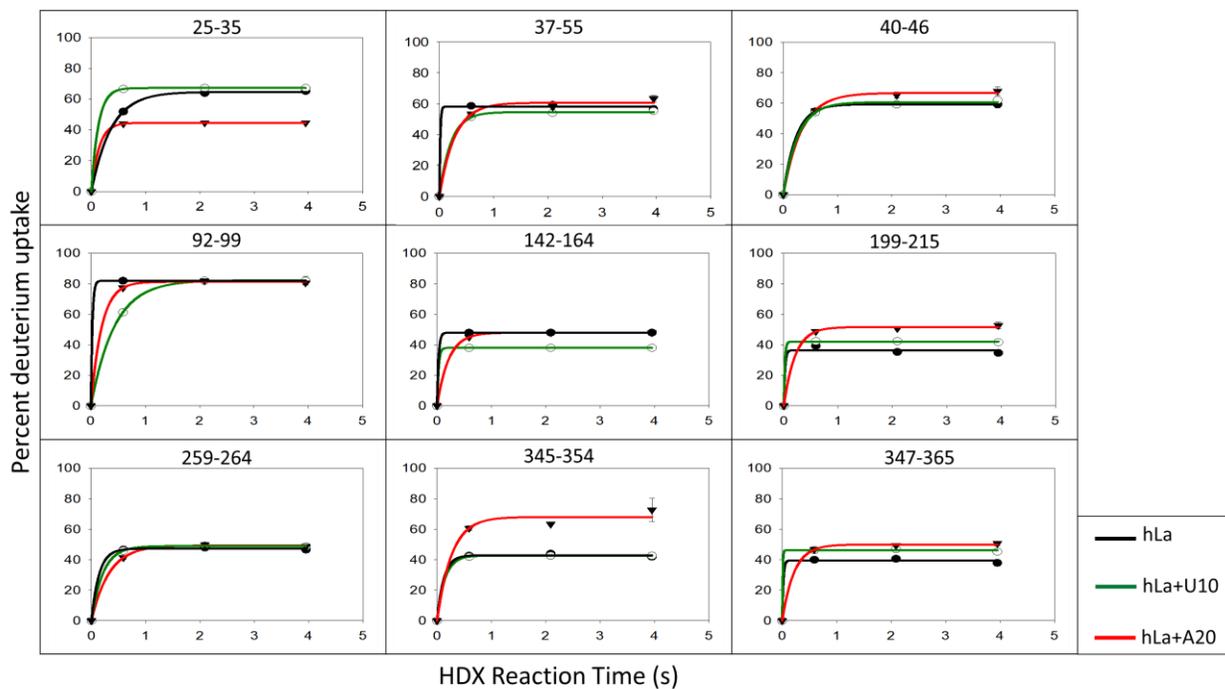
Figure B3: Deuterium uptake kinetics for selected peptides in hLa binding interactions. Black trace (filled circles), green trace (open circles) and red trace (filled triangles) correspond to free hLa, U10-bound hLa and A20-bound hLa, respectively.

**Appendix C: Supplementary information for Chapter 4**
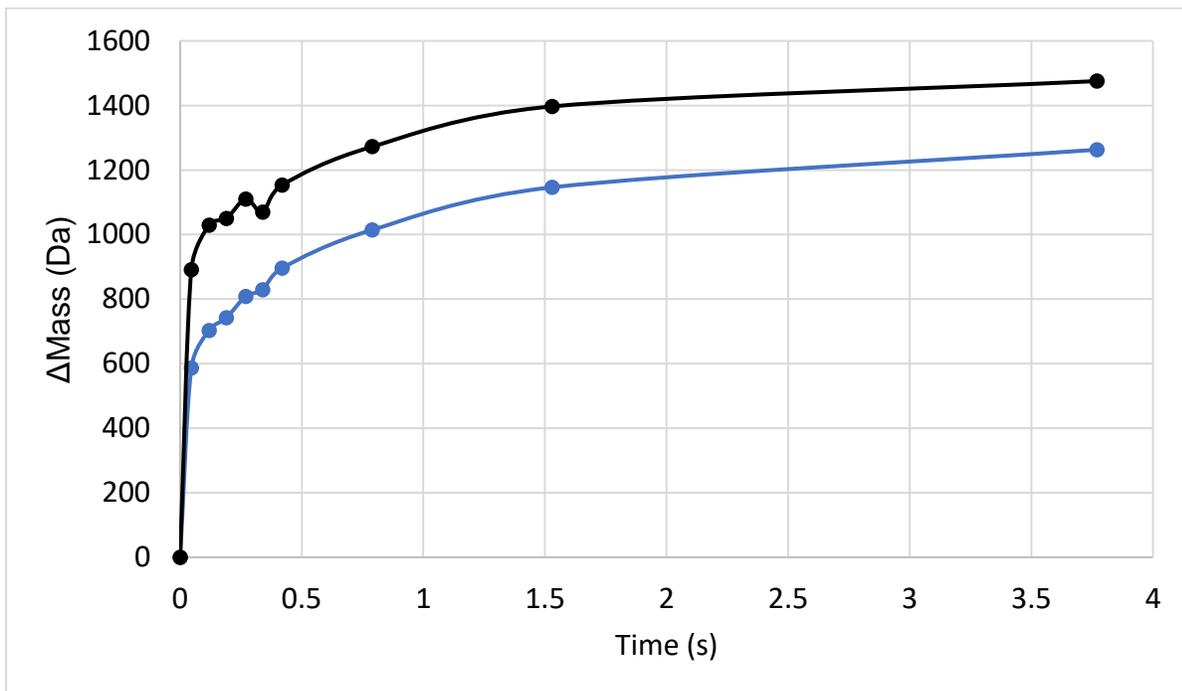


Figure C1. Global HDX of two different batches of Avastin showing batch to batch variation.

**Appendix D: Supplementary information for Chapter 5**
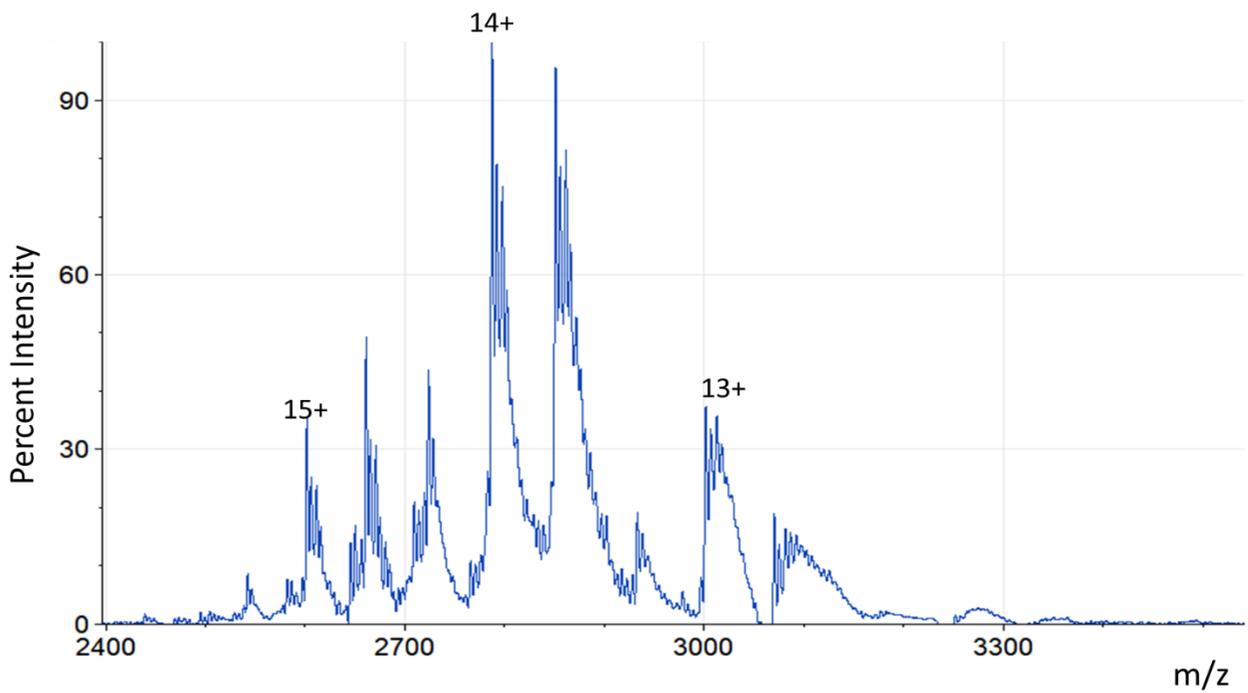


Figure D1. Native mass spectrum of Vascular Endothelial Growth Factor A-165 sprayed from 250 mM $C_2H_3O_2NH_4$, pH 6.7. The mass was revealed to be 39 kDa with additions of 894 Da and 2040 Da likely to be glycosylations on the protein.
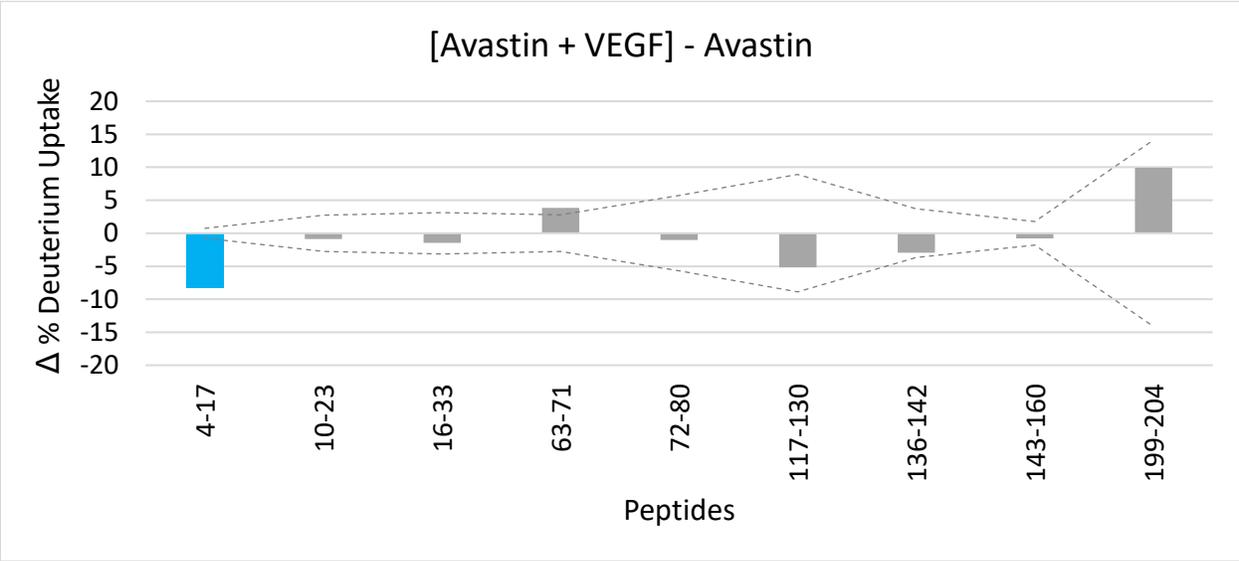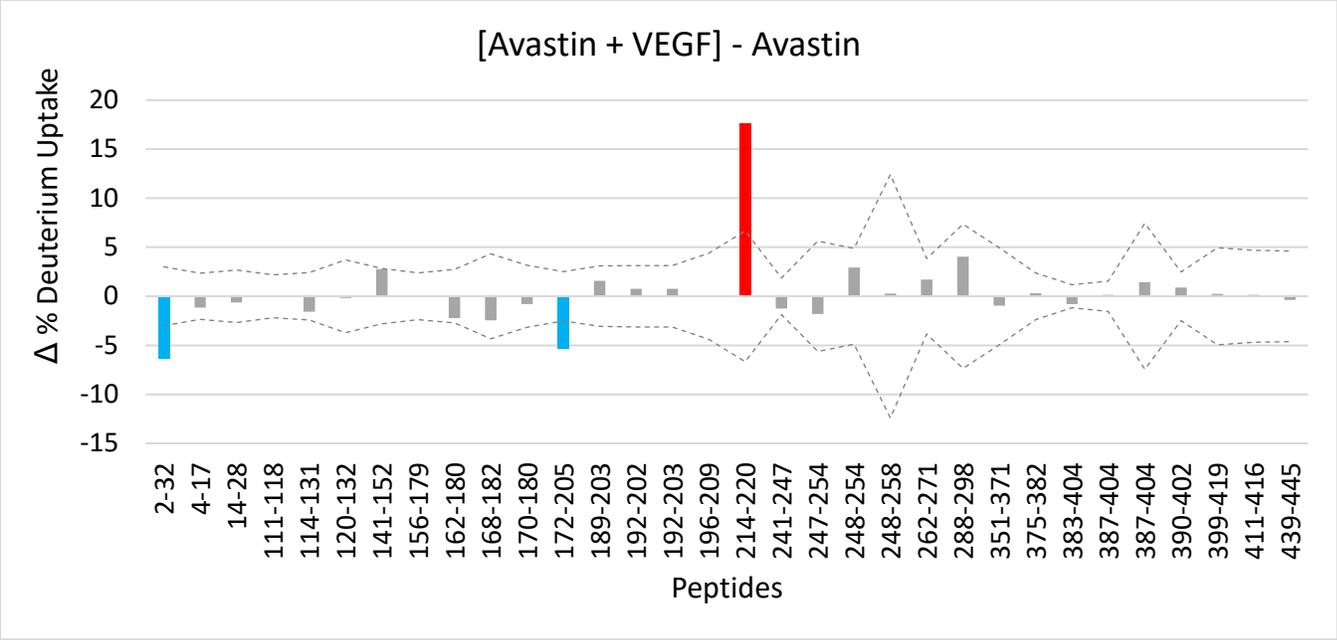
Figure D2: Paratope Mapping HDX analysis of Avastin when bound to VEGF. The top graph shows the heavy chain peptides in Avastin whereas the bottom graph shows light chain peptides in Avastin.