ANALYSIS OF CONFORMATIONAL DYNAMICS IN HEPATITIS B CAPSID PROTEIN

by

Navid Movahed

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

March, 2015

DEDICATION

To my baby boy, Nima:

…your joyful presence changed our life in so many ways during this journey. You always inspire me with your tireless curiosity!

This dissertation is lovingly dedicated to my wife for her support, encouragement, constant love and patience,

And a special feeling of gratitude to my loving mother for instilling the importance of higher education.

ACKNOWLEDGEMENTS

I am most grateful to the members of my committee, Dr. Patrik Callis, Dr. Edward Dratz, Dr. Martin Lawrence, and Dr. Ross Carlson for their time and amazing feedback throughout my studies and the process of writing this document. A special thanks to Dr. Dratz for always pushing my thinking towards a deep understanding about science.

Special thanks to the chair of my committee, Dr. Valerie Copie for her exquisite attention to detail and for her demand for excellence.

I am extremely grateful to my amazing teacher, mentor and advisor, Dr. Brian Bothner for his consistent guidance, support, encouragement, and expertise throughout this project.

And I deeply appreciate my colleagues and friends for their help and support when I needed the most, especially Dr. Doreen Brown.

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF FIGURES

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

ABSTRACT

Hepatitis B virus (HBV) is a model system for investigating the principles of icosahedral capsid assembly and a major human pathogen. As detailed by the work presented herein, viral capsids are not simply a static container for the viral genome. Rather, they are highly functional molecular machines critical to the virus life cycle. The assembly process of the HBV capsid involves the concerted assembly of 120 homodimeric subunits to form a T=4 icosahedron, which has been shown to be affected by temperature, ionic strength, and small molecules in a manner consistent with models of allosteric regulation. Our lab has already completed rigorous measurements of the conformational equilibria for HBV protein using enzyme-mediated kinetic hydrolysis, where we investigated the role of potential molecular switches in capsid assembly. These studies have now been complemented with hydrogen deuterium exchange based mass spectrometry.

Hydrogen deuterium exchange mass spectrometry (HDX-MS) provides valuable insight into solution-phase protein conformation and structure. The resolution of protein structural information in HDX-MS measurements is primarily limited by the peptide coverage of the on-line pepsin proteolysis. We have realized near single amino acid resolution coverage maps by combining online proteolysis with rapid reverse-phase chromatography of highly rich peptide mixtures.

Through the use of differential HDX, I investigated the effect of temperature, salt and amino acid mutation on rate of uptake and protection. These effectors have proven to thermodynamically or/and kinetically target the capsid assembly. High resolution HDX-MS was used to investigate the effect of these effectors on the protein dynamics. This has allowed us to elucidate the allosteric mechanism involved in the capsid assembly. Together these results indicate that the conformational landscape of HBV can be remodeled by a range of factors. The ability to map protein motions by HDX on specifically selected conformational states has profound implications in revealing quasi-equivalent subunit associations and the design of antiviral therapies.

BACK GROUND

Introduction

Protein Dynamics/Function

The three dimensional structure of a protein is no longer considered to be sufficient to describe function, and the structure-function paradigm in protein science now includes dynamics. The native solution-phase state of many proteins is as an ensemble of interconverting conformations which are in-equilibrium [1]. Dynamics is crucial for the biological function of many proteins including how they interact with small molecules and proteins. It has long been recognized that conformational fluctuations are a critical component of biochemical phenomena such as enzymatic catalysis and signal transduction. Many enzyme catalysis reactions are associated with small structural motions in mobile loops and active site residues [2, 3]. On the other hand, many proteins require large scale motions. Extreme examples of this category are proteins that exist in partially or fully disordered states [4, 5]. The presence of unstructured regions is no longer considered a rarity. They are rather an important protein category spanning the proteome. Locally disordered or high mobility regions can be involved in molecular events such as recognition, assembly, and allosteric regulation. A more flattened energy landscape and spatiotemporal conformational diversity are some of biophysical characteristics of these proteins [6-8]. Regardless of whether a protein is structured or unstructured, the amino acid sequence dictates the ultimate conformation in

a given solution condition. A high net charge and polarity are common features of disordered regions and can be used to predict their presence [4, 5, 9].

Local protein disorder may also be involved in regional structured-to-unstructured transitions as part of the biological role. In other words, the biologically active form of a protein may be either the structured form or unstructured one. Proteins may couple these structural rearrangements required for long-range communication between different regions with binding and functional events [6, 9].

A protein in solution fluctuates between the many conformations of its energy landscape. The relative population of a particular state in the ensemble is based on energy minima and barriers between conformations [10, 11]. It is this structural 'dynamics' that is believed to have a role in protein function. Substantial efforts have been placed on measuring the rate of change and equilibria between conformations of interest in solution. To understand protein function in signaling, transport, recognition, and the assembly of complexes specific techniques for probing conformational change are needed. Knowledge about the location, range, frequency, and free energy of protein motions is critical to understanding function. However, technical hurdles make it difficult to study protein motion in the context of supramolecular complexes.

Biological and Biophysical Aspects of Viruses

Viral capsids are no longer considered to be static containers merely providing protection to the viral genome. Rather they are now understood to be complex molecular machines using coordinated movements triggered by microenvironment. Viruses are very efficient molecular machines with general life cycles involving self-assemble,

maturation, packaging specific material and genome, interacting with host cells, intra-host-cellular trafficking, entrance and release, and genome delivery events which all must be carefully orchestrated [12]. By studying viruses, much can be learned not only about virus biology, but that of the host cell as well.

Icosahedral viruses are model systems for studying protein dynamics in supramolecular complexes. Models of viral protein structural transitions have been captured by static techniques such as x-ray crystallography and biochemical approaches such as antigen exposure. Models developed from such studies provide information about protein plasticity, but often it is only the most thermodynamically stable conformers that can be studied. What is clear is that conformational change is part of lifecycle events and associated with quaternary rearrangements, distinct tertiary structures, and localized perturbations of the secondary structure. However, solution-phase equilibrium dynamics add another dimension to the dynamic nature of viruses as 'rapid equilibrium motions'. These motions have been described as protein breathing [13-15]. They may involve only a subset of the subunits in a capsid or be global in nature. Crystal packing forces in x-ray crystallography or low temperature in cryo-EM quench this motion. Breathing motions can involve different size and time scales, and can involve levels of protein structure from secondary to quaternary [13, 16].

A classic example of solution phase conformational dynamics associated with capsid proteins in solution involves poliovirus (a member of the Picornaviridea). Antibodies raised against intact virus particles were shown to interact with a region that is clearly internal based on the structural model from X-ray crystallography data [17].

Another example that demonstrated the critical role of conformational plasticity to the lifecycle of a virus is an antiviral drug from the WIN family which increases the thermostability of rhinovirus capsid (Picornaviruses), preventing transition between viral lifecycle steps such as uncoating [18-20]. Flockhouse virus (FHV) is another example for which conformational dynamics of viral capsid proteins have been demonstrated. When it was exposed to proteases, the first peptides to be generated mapped to the interior of the capsid, positioned next to the RNA [21]. Icosahedral capsids are good examples showing an ensemble of conformations of a viral capsid protein can exist in solution [12, 13].

Biological and Biophysical Aspects of HBV

Hepatitis B virus (HBV) is the major cause of hepatocellular carcinoma and cirrhosis, having infected two billion people leading to 500,000-600,000 deaths each year [22, 23]. HBV is an enveloped virus with an icosahedral protein shell at its core. After the assembly of the immature core and encapsulating the RNA pre-genome and reverse transcriptase, the linear single-stranded RNA pre-genome is reverse transcribed into the circular, partially double-stranded DNA genome of the infectious particle [22-24].

Icosahedral capsids are defined by their triangulation number (T). This corresponds to the numbers of capsid protein monomers (T × 60) in the structure [25]. Both *in vivo* and *in vitro*, HBV assembles into capsids with T=4 and T=3 morphologies. The predominant form of this capsid in solution has T=4 icosahedral symmetry having a diameter of ~ 350 Å, with ~ 130 A˚ inner radius and ~ 20 A˚ thickness of the shell. As a T=4 particle, it is composed of 240 copies of a homodimeric core protein (Cp) (Figure 1.1). The T=3 form is composed of 180 copies of Cp. The ratio of T=3 to T=4 particles

may be varied by the solution condition during assembly. Homodimers of Cp are the primary building blocks of the HBV capsid. Assembly is believed to proceed through a series of intermediates arising from interactions of trimers of dimers or dimers alone [24, 26-30].

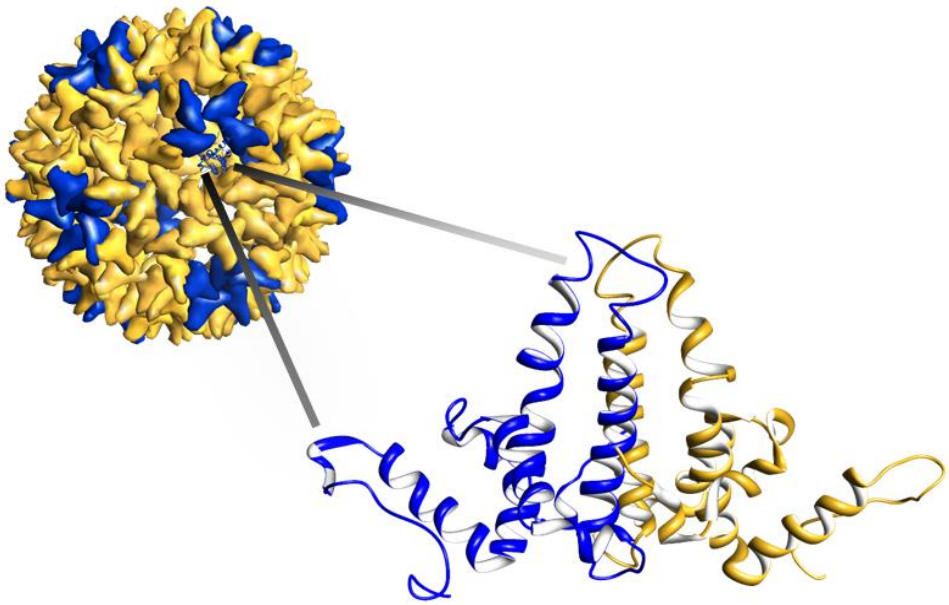

Figure 1.1: The structure of a Cp149-dimer within HBV T=4 capsids. Quasi-equivalent monomers within a HBV T=4 capsids; Blue shows monomers in fivefold vertex (PDB# 1QGT).

Biophysical Aspects of HB
Capsid Homodimeric Core Proteins (Cp)

HBV capsid building-blocks are homodimers of Cp, a 183-residue protein composed of a 149-residue assembly domain and a 34-residue RNA-binding C terminus. A helix-turn-helix (α-helical hairpins) from each subunit (half dimer) forms half of the four-helix bundle at the interface that holds the monomers together (amino acids 49-109). This four-helix bundle projects as a 30 Å spike from the plane of the capsid surface (Figure 1.2). The interaction between subunits in the dimer is very strong, and requires a

high concentration of denaturant agents to dissociate [16, 31]. Figure 1.2 also shows α-helical structural components of Cp149.

At the intradimer interface, a conserved cysteine from each monomer (C61) may form a disulfide cross-link. The presence of such a cross-link in the reducing condition of cytoplasm is uncertain [13, 23]. Helix-turn-extended structures (amino acids 111 to 143) of dimers around icosahedral five-fold and quasi-six-fold vertices overlap as shingles. The RNA-binding domain (amino acids 150 to 183), which is possibly disordered based on structural models and predictions has roles in signaling intracellular transport such as nuclear localization and particle secretion [32]. It is also involved in RNA packaging and reverse transcription. The phosphorylation status of the C-terminal domain is believed to be important for these functions [24, 33-35]. This RNA binding C-terminal domain is multifunctional and is believed to be transiently exposed to the exterior of capsid, although structural models localize it to the interior during transcription [24, 31]. Quantitative kinetic proteolysis studies for probing the dynamics of Cp149 defined this as a highly dynamic domain. Based on these studies, the internalized C-terminus is transiently exposed in solution [36].

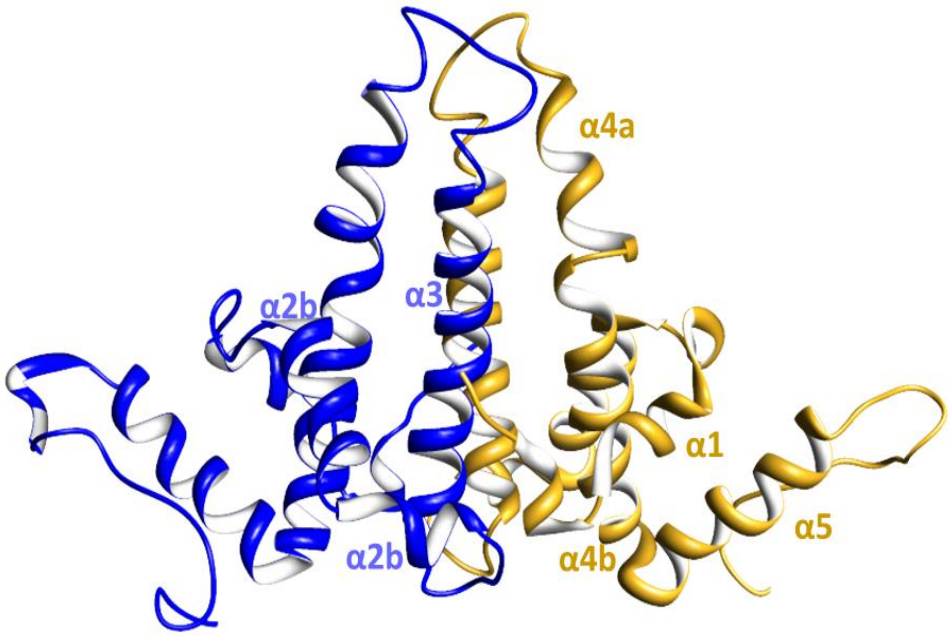

Figure 1.2: Structure of the Cp149 homodimer. The helix-turn-helix of every monomer contributes for a four-helix bundle dimerization. Labels of α-helical regions are shown based on the x-ray crystallography structure of HBV Cp149 within a T=4 capsid (PDB# 1QGT).

Every Cp149 monomer contains an extensive hydrophobic core. The presence of such a region is highly conserved among human viral proteins with a single polypeptide chain. The core is formed by residues Tyr-6, Phe-9, Leu-15, Leu-16, Phe-18, Leu-19, Phe-23, Phe-24, Trp-102, Phe-103, Phe-110, Val-115, Tyr-118, Leu-119, Phe-122, Trp-125, and Leu-140. These residues are mostly contributed by α1, α2, α4b and α5 (Figure 1.3). The hydrophobic core possibly play an important role in the stability of the folded monomer.[29].

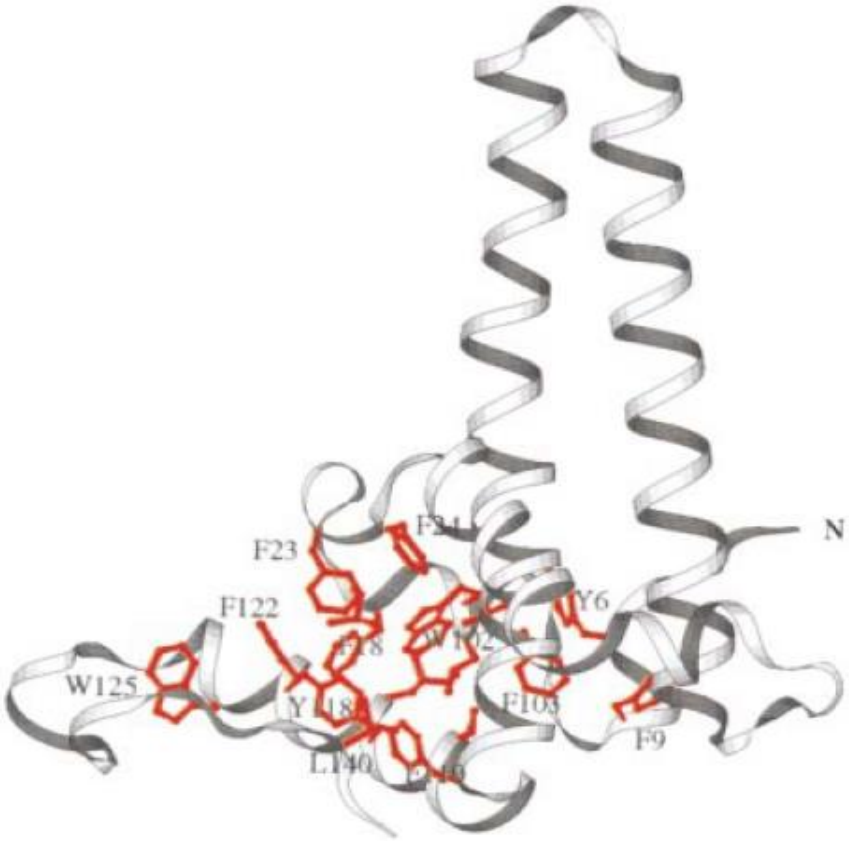

Figure 1.3: Hydrophobic core of HBV capsid protein. Sidechains have been added to show the region and amino acids involved in formation of the stabilizing hydrophobic core of the HBV Cp monomer. From crystal structure of capsid (PDB# 1QGT) [29].

HBV Capsid Assembly and Dynamics; Assembly Effectors

Viral capsid assembly is a carefully controlled process so that capsids containing the correct genetic material are formed. This process is a function of micro-environmental conditions such as pH, temperature, and ionic strength [37]. During HBV capsid assembly, the Cp undergoes a transition between free- and assembly- competent conformation. The homodimers of the 149-residue assembly domain (Cp149) self-assemble through a nucleation step as trimers of Cp149-dimers [38]. Zinc ions have been shown to bind Cp-dimers and stimulate assembly. The kinetics of the Zinc accelerated

reaction is consistent with an allosteric model. High levels of Zinc appear to drive assembly too quickly, promoting kinetic traps [39]. Heteroaryldihydropyrimidine (HAP), a family of assembly effectors, also activates assembly. Assembly of capsid at an inappropriate time or place has great potential as an antiviral strategy [40, 41]. HAPs overfill a cavity at the inter-dimer contact and thermodynamically misdirect assembly [26, 42]. They also kinetically activate the assembly by stabilizing an assembly-competent conformation [26, 42]. Assembly is a carefully coordinated process that represents a thermodynamic balance between distinct structural states of HBV core protein. Investigating the effect of variable micro-environmental parameters such as temperature, salts, pH, antiviral drugs, and particular HBV protein mutations have been used to characterize assembly. A substantial body of evidence now exists, indicating that conformational changes have a role in this and other stages of the HBV life cycle. However, the lack of quantitative data on protein dynamics for HBV, as well as other viruses, represents a major gap in our understanding at a mechanistic level. Quantitative investigations of capsid protein conformational change present an opportunity to address fundamental properties of allostery in a noncovalent complex.

In this work, Hepatitis B Virus (HBV) T=4 capsids have been used as a model icosahedral system for which the qualitative role of dynamics and the factors important for its assembly and stability are well documented. We applied hydrogen-deuterium exchange (HDX) coupled with LC-MS to investigate the location, frequency, range, and free energy of motions in HBV capsid protein. The qualification and quantification of protein motions by HDX can elucidate the interplay between local dynamics,

conformational plasticity, and global stability of HBV protein. Using automated high throughput HDX approaches we will investigate how solution condition alters the deuterium uptake by capsid protein.

Techniques to Study Dynamics

A variety of solution-phase biophysical techniques have been conducted to investigate equilibrium structural dynamics (Figure 1.4). Among these approaches, we utilized Hydrogen Deuterium Exchange (HDX) to investigate what we believe to be virus lifecycle relevant ranges of protein motions from angstroms to subnanometers, and from subseconds to hours.

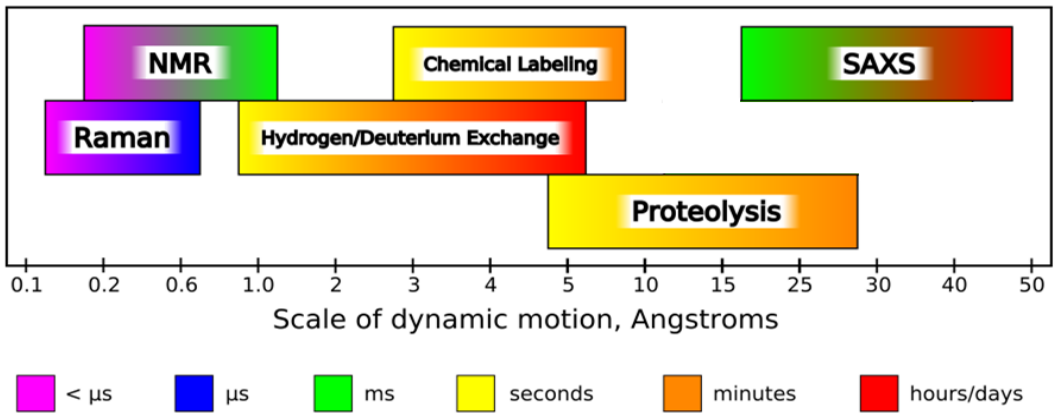

Figure 1.4: Utility of solution-phase approaches for study of protein dynamics. The approximate sensitivity of every approach based on either size or time scales protein motions (the horizontal axis and color keys, respectively). instrumental/methodological developments may vary the technical limitation of every approach [16].

Of the solution-phase techniques shown in Figure 1.4, kinetic proteolysis has proven to be a powerful approach to study the stability and dynamics of proteins and particularly virus particles [21, 43]. Carefully controlled proteolysis reactions coupled with SDS-PAGE and mass spectrometry have been used to mapped kinetically favored

site of cleavage which through inference are partially unfolded regions on the exterior of the particles [12, 16, 21]. The fundamental kinetic model of conformational equilibrium of a protein in solution which is based on cleavability by proteolytic enzymes, involves transitions between protected "closed" and cleavable partially "open" conformations:

$$Closed \underset{k_{close}}{\overset{k_{open}}{\rightleftharpoons}} Open \xrightarrow{k_{cleave}} Cleaved$$

The same two-state model of conformational equilibrium which involves transitions between "closed" and "open" conformations has been used to describe data arising from HDX experiments with the difference being that scales for motion and time are often different. The size and time scales of protein motions visible by HDX are usually smaller and expanded in a wider range compared to enzymatic hydrolysis (Figure 1.5). In other words, HDX is a more sensitive approach to track protein conformational movements. The minimum size of a locally unfold polypeptide required to be a potential substrate for proteolytic enzyme varies depending on the sequence and structure of the polypeptides and also the presence of neighbor steric protections by long-range secondary structure elements, although it cannot be less than 10-12 amino acids [44, 45]. To map the detailed structural motions across the HBV capsid protein, we utilized HDX technique as a more sensitive approach.

HYDROGEN DEUTERIUM EXCHANGE COUPLED TO LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY (HDX-LCMS)

Introduction

Hydrogen Deuterium Exchange and its Application in Protein Dynamics and Thermodynamics

The majority of biophysical techniques for studying protein dynamics reports only on specific "local" regions of the protein. For example, fluorescence and UV-Vis spectroscopy report about the micro-environment surrounding chromophores. Another powerful approach, electron paramagnetic resonance spectroscopy (EPR), provides accurate distance and range of motion measurements, but again only for specific reporter groups. In contrast, hydrogen deuterium exchange (HDX) can probe dynamics globally and locally. HDX is based on the observation that hydrogens attached to oxygen, nitrogen, and sulfur atoms in proteins (O–H, N–H, and S–H groups) continuously exchange with hydrogens from the solvent. The process follows the Lindstrom-Lang convention [46]. Nuclear magnetic resonance (NMR) and mass spectrometry are generally the analytical methods of choice for measuring exchange. Highly detailed mapping of dynamic regions for small proteins have been obtained by HDX coupled 1D and 2D NMR. However, protein solubility, molecular weight, and stability during analysis are all contribute to the suitability of a protein or system for HDX-NMR [47].

In mass spectrometry-based experiments, exchange between hydrogen and deuterium is observed as a change in mass. The 1 amu difference of mass gain is readily

observable using standard time-of-flight (TOF) mass analyzers, even when looking at proteins with 300-400 amino acids. Standard HDX protocol involves diluting the protein of interest by 10 fold or more into a buffered solution made with D_2O. Exchange is then tracked by measuring the protein mass at specific time intervals after dilution. For proteins, HDX involves exchange of amino acid side-chain and backbone amide hydrogens; however exchange of side-chain chemical groups is typically so rapid that this information is often lost during liquid chromatography. Amide hydrogen exchange is slower and can be controllable by low pH/low temperature [48-52].

The ability to control exchange rate is important because in the case of analysis by mass spectrometry, measurements are normally made under aqueous conditions. Exchange is an acid or based catalyzed reaction, with base (OH^-) dominating at neutral pH. The reaction can be slowed dramatically by lowering the pH. Figure 2.1 shows the exchange rate of amide hydrogen as a function of pH [52]. The half-life of back-exchange of peptide amide hydrogens in these slow-exchange (quench) conditions is normally about 30 minutes [47].

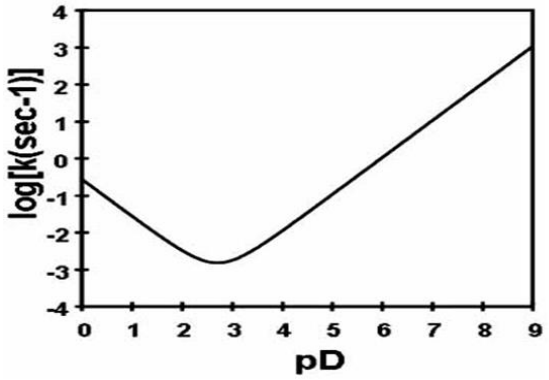

Figure 2.1: The rate of exchange of hydrogens as a function of pH. Hydrogens are located on peptide amide linkages (in polyalanine) [52].

Temperature also affects the exchange reaction. Therefore, the amide-HDX can be "quenched" by dropping the pH to 2.6 and the temperature to 0 ^{o}C. Under quench conditions, exchange is 10,000-100,000 times slower that at 25^{o}C, pH 7.

Measuring the exchange of amide hydrogens is the focus of HDX-experiments. Every amino acid residue has exchangeable amide hydrogen except for proline and the amino acid at the N-terminus of a polypeptide chain. Hydrogens that are involved in intra/intermolecular H bonds, buried inside the protein core, or located in a protein-protein interface are protected from exchange and therefore labeling by deuterium. The exchange of a backbone amid-hydrogen buried inside of protein is possible if there is access to bulk solvent molecules by either local or global unfolding (Figure 2.2). HDX is a powerful probe of protein structure and stability because amide hydrogens in the core of a folded protein are forming a hydrophobic core and are involved in stable hydrogen bonds. Therefore, the exchange rates of amide hydrogens can be used to investigate the thermodynamic and kinetic stability of protein conformations [53, 54]. Amide HDX is also a powerful technique for investigating dynamics, because structural fluctuations and local unfolding transiently expose labile hydrogens to D_2O-solvent. Exchange of this type is dependent on the size and time scale of fluctuations [48-52].

We coupled HDX with liquid chromatography (LC) and electrospray ionization mass spectrometry (ESI-MS). Every H to D exchange alters the mass of the protein by a Dalton. Monitoring protein mass change in a time-course experiment by mass spectrometry provides details on the kinetics of deuteration. We use continuous deuterium labeling; a protein in its native/equilibrium conformation is incubated in D_2O,

for a specific times, minutes to hours, and the mass shift of the intact protein is monitored as a function of incubation time [55]. This can be conducted as an ‘‘on-line’’ automated HDX-LC-MS experiment, making this a relatively high throughput and reproducible technique for monitoring the dynamic behavior of proteins under a variety of environmental conditions such as solvent composition, temperature and intermolecular interactions.

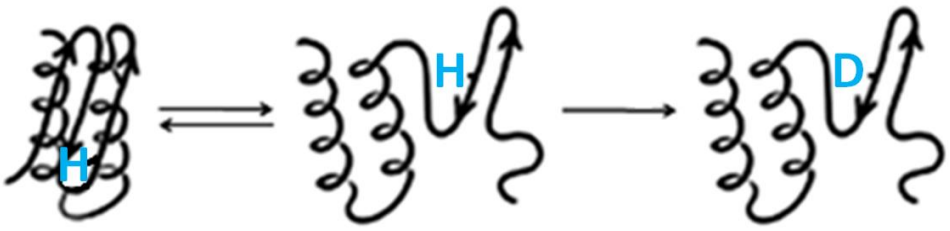

Figure 2.2: Local fluctuations in protein structure. Structural changes such as local conformational changes and unfolding events expose exchangeable hydrogens (H) to solvent. If H2O has been replaced with D2O, H will be replaced by deuterium (D).

Fast HDX-LC-MS for Intact Proteins

We have developed a fast HDX-LC-MS method to monitor the deuterium labeling of intact proteins. Rapid analysis has two important advantages; a greater number of data points can be collected in a given time frame and the time in which back-exchange can occur is minimized. However, care must be taken to assure that instrumental fluctuations in temperature, pH and chromatographic elution time do not occur. Therefore, careful considerations were taken to develop a method that was rapid and robust. Part of the process involved screening a large number reverse phase chromatography columns to find the optimal one. LC experiments were done under carefully controlled conditions (pH=2.6 and 0˚C) to minimize back exchange. The LC

method was optimized so that reproducible elution without the presence of carryover was realized. Chromatographic instability leads to fluctuations in the elution time of the protein. During HDX-LCMS, this will lead to changes in the back-exchange which will add error to the measurement of deuterium uptake. The carryover of proteins in the LC system (column, connections and tubing) is another concern as it can alter elution time and degrade the precision of uptake measurements. The carryover issue was given a great deal of consideration because running blank injections between experimental runs to clean the system would defeat the purpose of designing a high throughput approach.

Evaluation of the Fast HDX-LC-MS Setup on a Supramolecular Protein Complex

To test our fast LCMS system for HDX, DPS-like protein (DPSL), which is a member of the ferritin superfamily was used. DPSL assembles into a dodecameric cage. We had two forms of the protein, one from mesophilic bacteria, and the other from a hyperthermophilic archaea. Because the protein has both tertiary and quaternary structure, it was gauged to be an excellent model for supramolecular complexes and a good system to use to test our HDX setup. Figure 2.3 shows the dodecamer cage of the thermophilic form of the protein and an isolated subunit from *Sulfolobus solfataricus* (SsDPSL) in comparison to its mesophilic form of *Bacteriodes fragilis* (BfDPSL) [56-58]. The structures of the two proteins are very similar, except for a slight variation at the C-terminus and their stable oligomeric states in solution. The SsDPSL forms a stable dodecamer cage, while BfDPSL protein remains as a dimer [57]. The fact that both

dimeric and assembled cages were present was highly relevant to our plan to use the HDX system to investigate dimeric and assembled capsids of HBV.

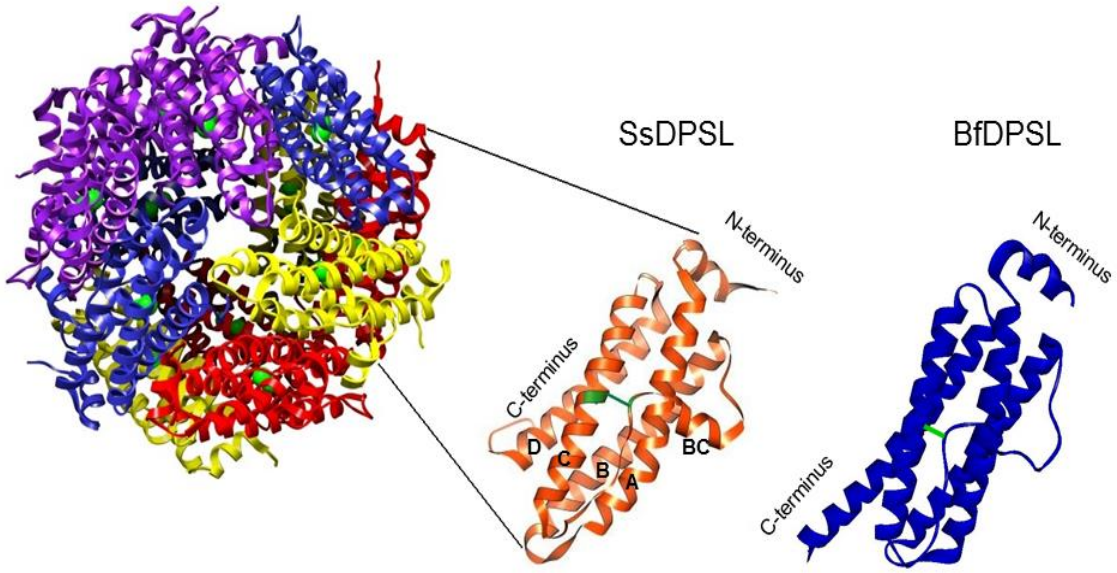

Figure 2.3: The structural model of the assembled SsDPSL dodecameric cage. Subunits are shown in different colors. Monomers from the two different species of SsDPSL and BfDPSL are shown on the right. The four helices forming the core helical bundle are decorated with two short helices of N-terminus and BC. The major difference in the protein fold between them is the extension of the D helix in BfDPSL. [PDB# 2CLB and 2VZB for SsDPSL and BfDPSL, respectively] [57, 58].

A disulfide bond in the SsDPSL acts as a staple, keeping the BC loop close to the core helical bundle (Figure 2.3) [58]. However, it is unknown if the same disulfide bond forms in the mesophilic protein [57]. We suggest that the thermostable DPSL cage has adapted to high temperature by forming this disulfide bond. Therefore we carried HDX on the dodecameric cage of C101S mutation of SsDPSL which did not form a disulfide bond to investigate this.

Materials and Methods

Methods

Hydrogen-Deuterium Exchange (HDX): experiments were conducted using continuous exchange methodology [55]. All deuterium labeling experiments were initiated by diluting concentrated DPSL (39:1) in D_2O-phosphate buffer (50mM, pH=6.5) at 37 ºC. Labeled proteins were injected into a highly controlled low temperature/low pH/fast reverse-phase liquid chromatographic system. The LC setup was coupled to mass spectrometry to monitor the mass gain of intact proteins as a function of time. LC was done by an Agilent 1100 LC system (Agilent Technologies, Waldbronn, Germany), a C4-macroTrap 3×8mm (Michrom Bioresources, Inc. Auburn, CA), water/acetonitrile binary gradient system in presence of 0.1% formic acid, flow rate 1ml/min, a temperature controlled micro-well plate sampler, and a column heater compartment. The electrospray ionization mass spectrometry (ESI-MS) system was a micrOTOF-MS system (Bruker Daltonik, Bremen, Germany). The mass spectrometer was operated in positive mode using a scan range from 300–3000 m/z. The system was tuned for optimum ion transfer/resolution of intact proteins.

Nuclear Magnetic Resonance (NMR): proton (^1H) NMR was recorded using a 300 MHz Bruker Avance DPX 300 spectrometer. All resonances are given in parts per million (δ).

Protein Expression and Purification: SsDPSL and C101S mutant expressed and purified as described in the paper by Gauss et al. [58] and were gifts from the lab of Dr

Trevor Douglas. BfDPSL expressed and purified as described in a paper by Gauss *et al*. [57] and were a gift from the lab of Dr Lawrence.

Differential Scanning Fluorometry (DSF): The thermostability assay was done by mixing proteins, PBS buffer pH=7.5 and Sypro-Orange dye (Invitrogen Inc.) in a quantitative PCR (qPCR) instrument (RG-3000; Corbett Research) [59].

Results and Discussion

Method Development

The goal in developing a high throughput method was to have the capability to investigate a wide range of proteins and experimental conditions, including both monomer and supramolecular complexes. To check the reproducibility of the approach across an entire HDX experiment, we overlaid total ion currents (TIC) of LC-chromatograms from different experiments (Figure 2.4).

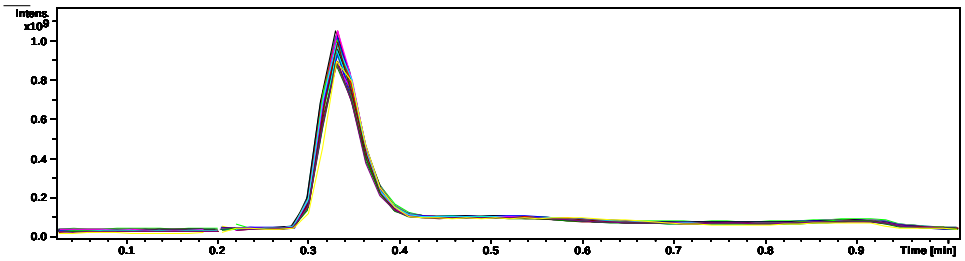

Figure 2.4: Overlay of 32 chromatograms of HDX-LC-MS experiments. The reproducibility of chromatograms is indicative of a stable LC system. Total Ion Chromatograms are showing a clean elution of the protein off salts and buffers in 20 seconds. The Figure shows the reproducibility of TIC-chromatograms; the elution-time and the TIC-chromatogram has a reproducible pattern.

We looked for the best performing organic solvent for fast and efficient elution of proteins. Methanol was shown to give the best chromatography versus acetonitrile. To investigate the contribution of protons of methanol as LC solvent in the exchange of protein deuterium uptakes back to the solvent, we studied the H-D exchange of methanol versus acetonitrile by ^{1}H NMR (Figure 2.5).

We conducted Proton (1H) NMR titration experiments of DMSO-D6+D2O by methanol and acetonitrile. Upon titration of a DMSO solution saturated with D2O by methanol, no methanol -OH shift (orange) was observable due to the rapid exchange of the D2O with the methanol-OHs (Figure 2.5 panel A). A rise in the DOH peak (sky blue) was however, observed as the results of the H-D exchange between methanol and D2O. The methanol-CH peak (red) was detectable and increased with increasing methanol. The DMSO peak (black) was adjusted to 1 in every step of titration and all integrals were referenced to the DMSO peak (Figure 2.5 panel A). The same trend was observed in the presence of formic acid (data are not shown).

Upon titration of the DMSO-D6+D2O solution by acetonitrile, the DOH peak (sky blue) remained nearly unchanged because there was no H-D exchange contributing to the DOH peak (Figure 2.5 panel B). The acetonitrile-CH peak (dark blue) was detectable and increased with increasing acetonitrile. The DMSO peak (black) was adjusted to 1 in every step of titration and all integrals were referenced to the DMSO peak (Figure 2.5 panel B). The same trend was observed in the presence of formic acid (data are not shown).

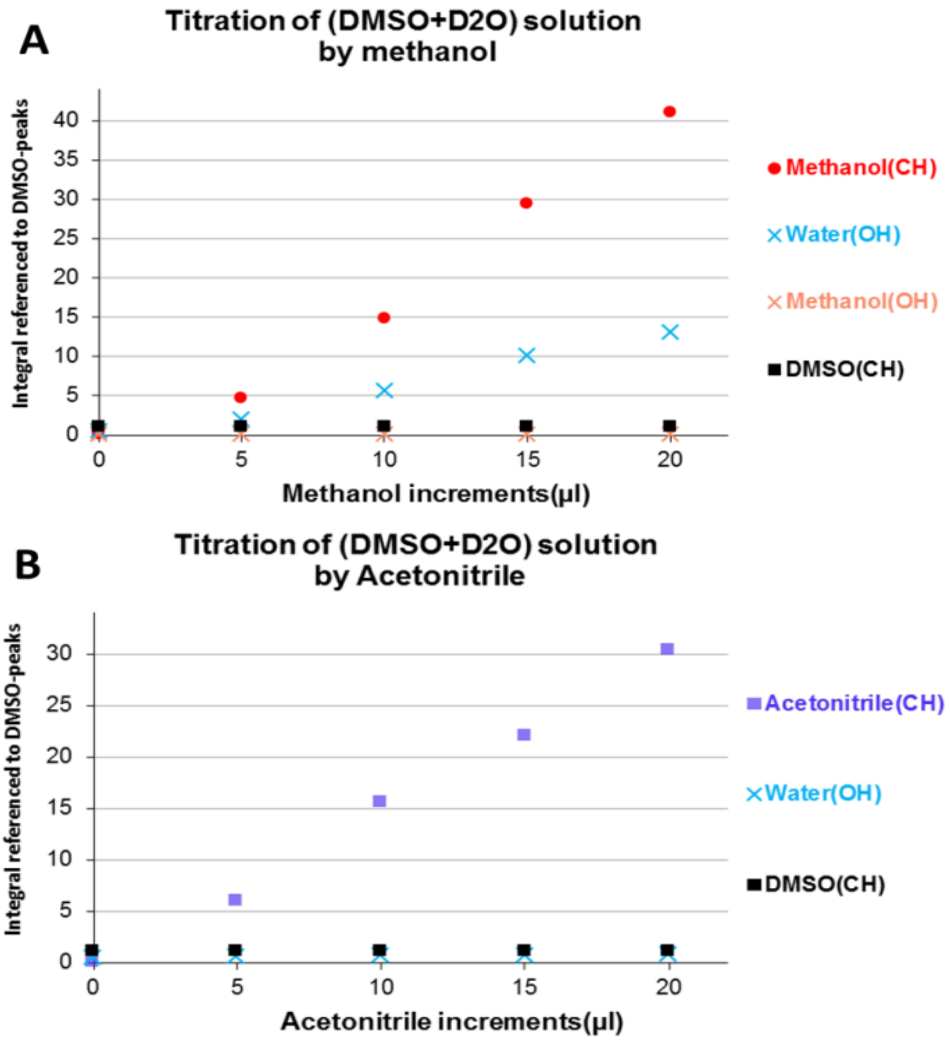

Figure 2.5: NMR studies on deuteration of methanol and acetonitrile. A) Upon titration of a DMSO solution saturated with D2O by methanol, no methanol -OH shift (orange) is observable due to the rapid exchange of the D2O with the methanol-OHs. A rise in the DOH peak (sky blue) is however, observed as the results of the H-D exchange between methanol and D2O. The methanol-CH peak (red) is detectable and increases with increasing methanol. The DMSO peak (black) is adjusted to 1 in every step of titration and all integrals are referenced to the DMSO peak. B) Upon titration of a DMSO solution saturated with D2O by acetonitrile, the DOH peak (sky blue) remains nearly unchanged because there is no H-D exchange contributing to the DOH peak. The acetonitrile-CH peak (dark blue) is detectable and increases with increasing acetonitrile. The DMSO peak (black) is adjusted to 1 in every step of titration and all integrals are referenced to the DMSO peak. Proton (1H) NMR was recorded using a 300 MHz Bruker Avance DPX 300 spectrometer, and all reports are given in parts per million (δ).

NMR experiments showed the immediate and complete exchange of methanol –OHs with D2O. In contrast, no exchange between acetonitrile and D_2O was observed which made the acetonitrile the solvent of choice for HDX experiments.

Investigation of Back-Exchange

Although there were no observable fluctuations in the liquid chromatography patterns (Figure 2.4), the back-exchange of deuterium uptakes was still the issue of concern. To investigate the dimension and contribution of the LC-related back-exchange in the final data analysis, we conducted back-exchange and forward -exchange experiments. Carefully manipulations of the LC protocol can be used to change the elution time of proteins from a reverse-phase column. By altering the elution time, the length of exposure of proteins to the bulk water content during chromatography and therefore back-exchange can be controlled (Figure 2.6 panel A). The same procedure was used with D_2O as the chromatography solvent to control the mass gain of proteins during chromatograph. Forward-exchange experiments add another dimension to understanding the nature of the hydrogen exchange in proteins during HPLC and represent an area for future investigation.

For forward-exchange experiments, the protein was pre-incubated in a saturated solution of protonated urea (H_4-urea + H_2O) for 1 hour. We used 30% D2O: 70% acetonitrile as the LC-solvent in forward-exchange experiments. We tracked the mass gain of proteins as a function of their elution time from the reverse-phase column by mass spectrometry (Figure 2.6 panel B).

Figure 2.6 panel A shows the reproducibility of the system even under LC-protocol manipulation. The observed average mass shift for DPSL based on elution-time from the reverse phase column was then determined (Figure 2.6 panel B). For comparison, the expected mass shift as the result of only amide-hydrogens contribution was added to the graph. The expected value was based on the half-life of exchange of peptide amide hydrogens under the slow-exchange conditions of pH=2.6 and low temperature which is ~30min [47].

WtDPSL has 178 amide hydrogens and 169 side chain hydrogens that could potentially exchange. Therefore, the calculated maximum mass shift would be 347 amu. After the 1 hour pre-incubation in urea, we expected to see a deuterium uptake curve with a steep initial slope up to a couple of hundred deuteriums followed by a slower almost "plateau" zone. But what we observed was fast uptake of ~60 deuteriums (before the first collected data point at ~20 seconds) followed by a short slower almost "plateau" zone followed by another steep slope. The second steep slope in deuterium uptake at 1-2 minute time zone was an unexpected behavior. We concluded that the protein collapsed into a molten globule (or semi-denatured) state in which hydrogens are well protected from the bulk water content of LC solvents (the plateau zone) under the reverse phase column/LC conditions. We also concluded these molten structures have a significant conformational variation which contributes in the observed progressive forward-exchange mass gain (the steep slope of deuterium uptake starting at 1-2 minute time zone.

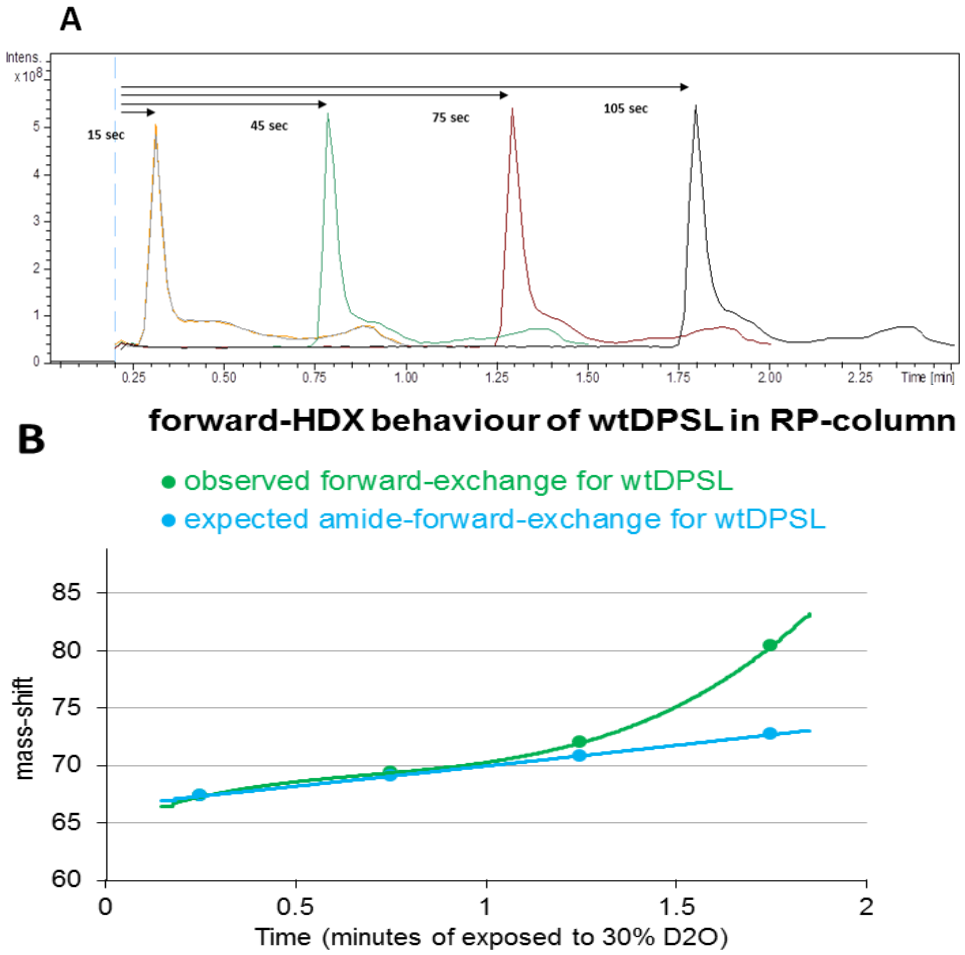

Figure 2.6: Effect of elution time on forward-exchange. Tightly controlled time-course experiments varying the elution-time of proteins from the reverse phase column were conducted. A) The overlay of TIC chromatograms of wtDPSL during forward-HDX experiments; B) Elution-times of 15, 45, 75 and 105 seconds (calculated stay time of proteins in the reverse phase column) of proteins from the reverse phase column shown in panel A are graphed versus their corresponding mass gain (green). The graph of observed mass gains in forward-exchange experiments is overlaid on the graph of expected amide contributions under the low pH and low temperature condition of these experiments (blue) (see text).

These forward-exchange experiments brought up the question of whether this was protein specific or if all proteins would exhibit similar degrees of unfolding in the denaturing condition of reverse-phase chromatography. If the latter case was true, we expect a similar back exchange behavior among different proteins which would mean we

could avoid correction for back exchange during final data analysis of proteins of interest. To address this question, the retention time of proteins in the reverse phase column was varied for the wtDPSL and the C101S mutant in time-course experiments as above. For back-exchange experiments, protein was pre-incubated in a saturated solution of deuterated urea (D_4-urea + D_2O) for 1 hour. We limited the incubation time in urea to 1 hour because the carbamylation process of proteins by urea complicates data analysis by mass spectrometry. We used 30% H_2O: 70% acetonitrile as the LC-solvent in back-exchange experiments. Such comparative back-exchange experiments on wtDPSL versus mutant C101S are shown in Figure 2.7.

The back-exchange behavior of wtDPSL and the mutant C101S during reverse phase liquid chromatography showed a similar loss of deuterium. For comparison, the expected mass shift as the result of only amide-hydrogens contribution was added to the graph. The expected value was based on the half-life of exchange of peptide amide hydrogens under the slow-exchange conditions of pH=2.6 and low temperature which is ~30min [47]. Comparing the expected amide contribution in back-exchange to observed ones shows side chains contribution in the observed mass gains. These resulted indicated that for these two proteins the effect of back-exchange would not be a factor in any comparative analysis.

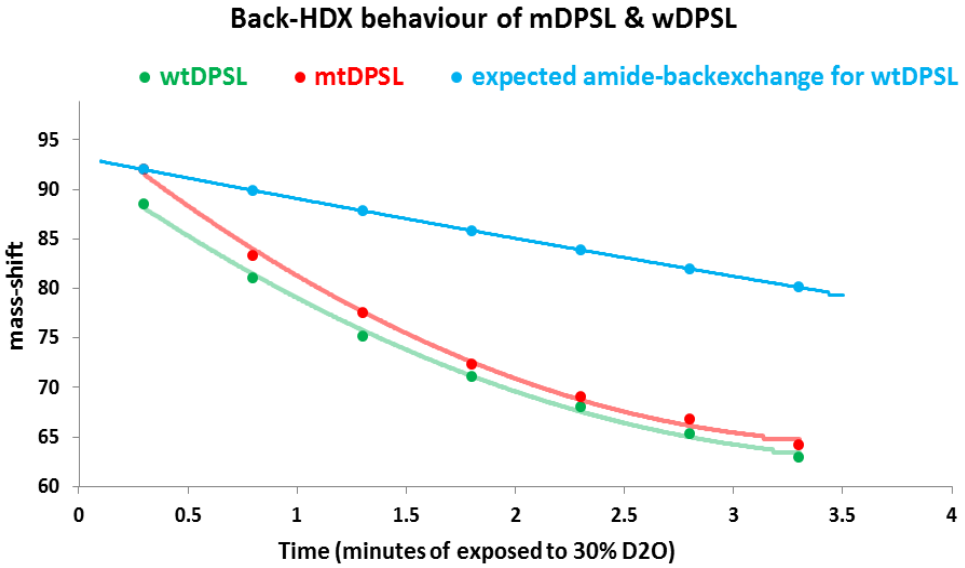

Figure 2.7: Measuring back-exchange of wtDPSL versus mtDPSL during reverse phase LCMS. Tightly controlled time-course experiments varying the elution-time of proteins from the reverse phase column were conducted. Elution-times of 15, 45, 75 and 105 seconds (calculated stay time of proteins in the reverse phase column) of wtDPSL from the reverse phase column similar to Figure 2.7 panel A are graphed versus their corresponding mass gain (green). For comparison, elution-times of mutant DPSL from the reverse phase column are also graphed versus their corresponding mass gain (red). These graphs of observed mass gains are overlaid on the graph of expected amide contributions under the low pH and low temperature condition of these experiments (blue) (see text).

Method Evaluation

Fast LC-MS allows for greater time resolution across the course of an exchange experiment as well as retaining more information by reducing back-exchange. The side chain deuteration is commonly lost in other HDX approaches. Figure 2.7 shows that ~30% of the observed deuteration in 0.3 minute experiment was lost in 3.3 minutes experiment. Therefore, the collected data by such a fast LC-MS setup provides more accurate data fitting for making estimates of the observed rates (Figure 2.8). A real/ideal deuterium uptake is multi-exponential because each site has a specific rate. This is described using a model such as a_1, a_2 …a_n and k_1, k_2…k_n as parameters (a_1, a_2, … and a_n

= numbers of hydrogens which exchange with respective rate constants of k_1, k_2, … and k_n; k_1, k_2, … and k_n). Rates of exchange depend on the amino acid sequence, secondary, tertiary, and quaternary structure. The HDX data for DPSL were fitted to a tri-exponent model: $D = N - a_1 \exp(-k_1 t) - a_2 \exp(-k_2 t) - a_3 \exp(-k_3 t)$, where N = the maximum number of exchangeable hydrogens; D = the observed number of deuterium atoms incorporated after t minutes. For example, HDX data of wtDPSL were fitted in these parameters: $a_1 = 42.80$; $a_2 = 24.22$; $a_3 = 105.69$; $K_1 = 215.12$; $k_2 = 1.32$; $k_3 = 0.02$. The equation will give us an approximate of N value for the HDX of the protein. The theoretical number of total exchangeable hydrogens, the sum of exchangeable backbone and sidechains hydrogens (347 for wtDPSL), is not considered as the N value.

The primary results of the HDX experiments on thermophilic and mesophilic forms of DPSL showed a globally higher flexibility for the dimeric BfDPSL. Based on the intact protein HDX data we cannot conclude any inherent structural differences between thermophilic cages and mesophilic dimeric forms of DPSL. This is because factors such as the cage-related structural constraints and the protection of protein-protein contacts in the cage could be reasons for the less deuterium uptake by thermophilic cages.

HDX data also revealed an important role for the C101-C126 disulfide bond in the observed global rigidity of SsDPSL compared to C101S. Figure 2.8 shows significantly more global flexibility for the C101S mutant form of DPSL. Based on hours long HDX experiments on both wtDPSL and its mutant C101S form, we can safely conclude that mtDPSL has a large protection loss compared to wtDPSL (Figure 2.8). The

reason can be the loose association of the short helix BC and BC-to-C loop with the bundle core in the mtDPSL which results in the exposure of many residues to solvent.

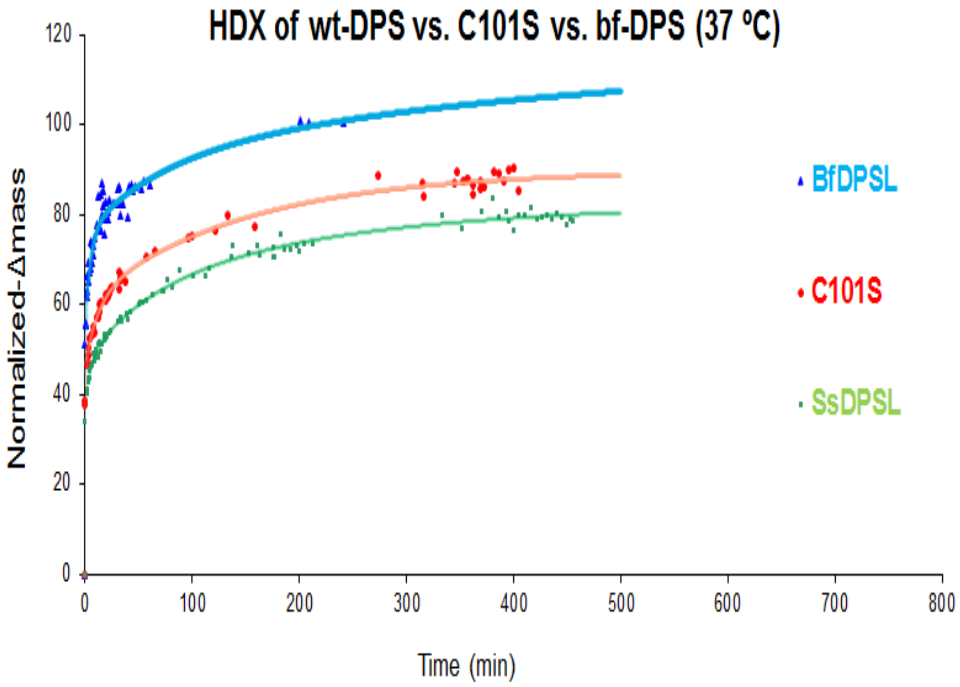

Figure 2.8: Deuteration progress curves for different forms of wtDPSL. Deuterium uptake curves of thermophilic DPSL-cages of SsDPSL (green) versus its C101S mutant form (red) versus the mesophilic dimeric form of BfDPSL (blue).

The thermostability of wtDPSL versus its mutant form by Differential Scanning Fluorometry (DSF) was conducted in our lab. Figure 2.9 shows the differential thermostability of wtDPSL and mtDPSL. The fluorescence data showed a higher thermostability of wtDPSL cages than C101S cages. A higher melting temperature of wtDPSL was consistent with the HDX experimental data suggesting a more compact structure for wtDPSL.

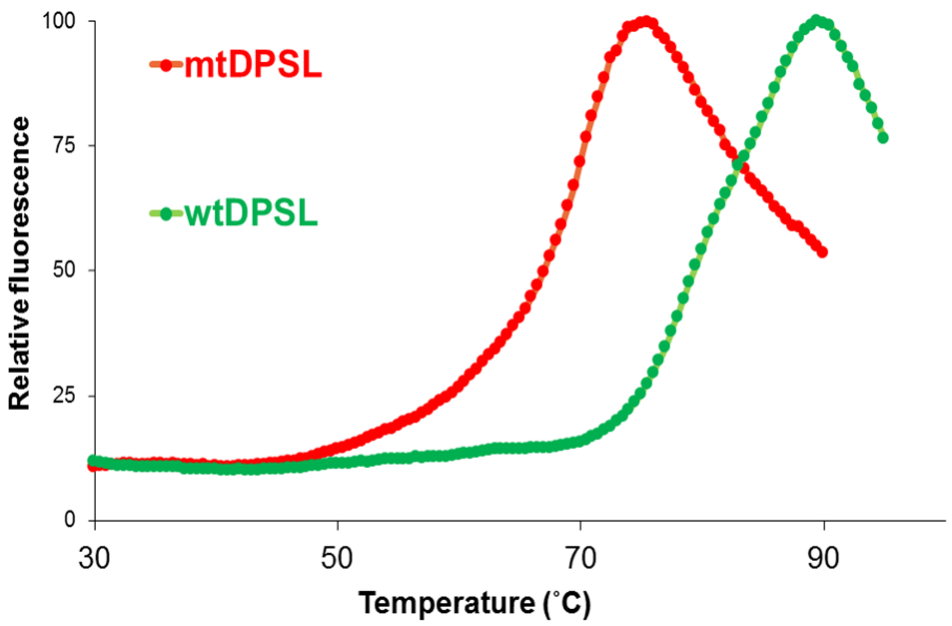

Figure 2.9: Differential scanning fluorometry of wtDPSL versus mtDPSL. WtDPSL (green) shows a significantly higher thermostability than its mutant form (red).

Conclusion

The preliminary HDX data indicated that all three forms of DPSL protein in have fundamentally different dynamic properties. By these HDX experiments we could also access the developed HDX setup as an approach to probe the global dynamics of HBV capsid proteins under variable solution conditions.

PROBING THE GLOBAL DYNAMICS OF HBV DIMER IN FREE AND ASSEMBLED STATES UNDER VARIABLE SOLUTION CONDITIONS

Introduction

Automated HDX-LCMS

To investigate the effect of temperature, ionic strength, antiviral drugs, and amino acid mutations on HBV capsid protein dynamics, we sought to apply HDX experiments. Previously, it has been shown that these factors regulate the thermodynamics and kinetics of HBV capsid assembly [24, 38, 42, 60]. Preliminary studies indicated that the differences were likely to be subtle. Therefore, to minimize the loss of data due to back-exchange during reverse phase liquid chromatography, we improved upon our fast HDX approach, making it even faster. The results presented herein describe a comparative analysis of HBV Cp under different states of assembly, solvent conditions, and after mutation. Together, this data provides new insight into the mechanism behind the observed allosteric regulation of HBV T4 capsid assembly.

Materials and Methods

Methods

Automated Ultrafast Hydrogen Deuterium Exchange: A continuous deuterium labeling [55] followed by automated fast LC-MS. All labeling experiments were done by automated/or manually initiation of reactions upon 20 fold dilution of protein samples into a D_2O-tris + NaCl (and H_2O-tris for the nondeuterated reactions). The concentration

of NaCl, the temperature, and the pD of buffer were varied as part of the experimental design. The mass gain of the intact protein as a function of time was measured using available TOF mass spectrometry systems. LC was done by an Agilent 1100 LC, 1200 LC and 1290 uHPLC systems (Agilent Technologies, Waldbronn, Germany) with degassers, binary and quaternary pumps, temperature controlled micro well-plate samplers, Infinity diode array detector and a column heater compartment, a C4-macroTrap 3×8mm (Michrom Bioresources, Inc. Auburn, CA), water/acetonitrile binary gradient system in presence of 0.1% formic acid, flow rate 1.5ml/min. Electrospray ionization mass spectrometry (ESI-MS) system was a micrOTOF-MS system (Bruker Daltonik, Bremen, Germany) and Agilent 6538 Quadrupole TOF Accurate-Mass (Agilent Technologies, Waldbronn, Germany). The mass spectrometer was operated in positive mode in a scan range from 300–3000 m/z and was tuned for optimum ion transfer/resolution on intact protein. We used Agilent MassHunter Workstation Software Qualitative Analysis Version B.04.00, Bruker Daltonics DataAnalysis software package versions 3.2 and an in-lab script for automated deconvolution analysis.

Sample Preparation: Protein dimers were obtained from Dr Zlotnick lab. Truncated HBV capsid protein (Cp149) dimers and capsids were expressed and purified from E. coli at 37 °C as described by Zlotnick and et al. [61, 62]. HBV protein represented 5-15% of the total E. coli proteins. Cells were lysed by French press and sonication, capsids separated on sucrose gradient and $(NH_4)_2SO_4$ and purified using a Sepharose column and ultrafiltration membranes. Dimers were obtained by diluting capsids in pH 9.5 and 3.5 M urea and purified by a Superdex 75 column. Before HDX

experiments capsids were freshly formed at RT with high salt and pH=7.5. Both capsid and dimers were filtered and concentrated with pressure-based stirred cells (Amicon EMD Milipore) including 100kD ultrafiltration disk. The purity of dimers and capsid solution were checked by HPLC.

Results and Discussion

Method Development

After some major and minor improvements in the HDX setup configuration which was used for DPSL protein in chapter 2, we were able to develop a high throughput ultra-fast technique of HDX for reproducible data collection over-night experiments. The basic principle was the same; a deuterium labeling reaction was followed by automated sample injection, high pressure/high flow-rate liquid chromatography that included solvent exchange, multiple LC-gradients and electro spray mass spectrometry. The whole LC system, from solvents to ESI-MS, was maintained at pH=2.6 and temperature = -1 ˚C. To keep the MS system clean, we used high pressure splitters in the LC-ESI-MS configuration to direct the extra LC solvent to the waste instead of the mass spectrometer. This helped to keep the system clean and sensitive, even when operated continuously for 24 hours. To keep the column and LC system clean of capsid carryovers and prevent running blank injections between runs, we added a high pressure high flowrate step of the LC solvent (90% acetonitrile: 10% water) as a washing step in both back and forward directions. We also moved the column regeneration step to the end and during sample injection step of the next run. Although the new protocol for LC configuration increased

the length of each LC run from 20 to 30 seconds, it eliminated carryover, which was more of a problem with HBV than DPSL.

Method Evaluation

In the design and testing of our HDX method, four capabilities were kept in mind:

- The system should be capable of a thousand injections per day
- High-pressure, high flow-rate and cold liquid chromatography are needed
- Protein desalting must be rapid
- Sample carryover from plumbing and columns must not impact subsequent injections

The new setup proved to be highly reproducible. The capability to investigate a wide range of experimental conditions on different proteins from monomers to supramolecular complexes with high reproducibility was a major goal for the high throughput approach. To show the reproducibility of our new approach through the whole days-long HDX-data collection, we overlaid total ion currents (TIC) of LC-chromatograms from 150 different experiments under varying conditions (Figure 3.1). The stability of LC-MS system and its robustness, despite technical challenges, was high.

In addition to reproducible chromatography, the HDX setup also needed to capture exchange with an acceptable signal/noise ratio. Figure 3.2 shows the m/z spectrum quality of collected data on HBV capsid in a solution of 150mM NaCl, 100mM tris at pH=7.5.

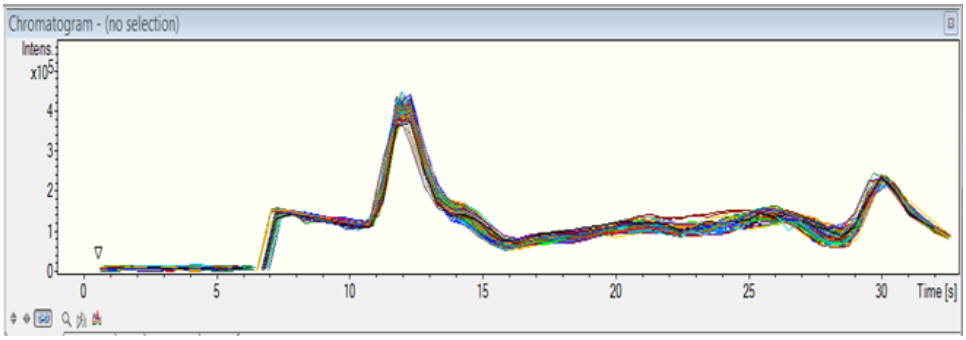

Figure 3.1: The Overlay of 150 TIC chromatograms of automated HDX-LC-MS experiments. The stability of LC-MS system provided a highly reproducible pattern among TIC-chromatograms.

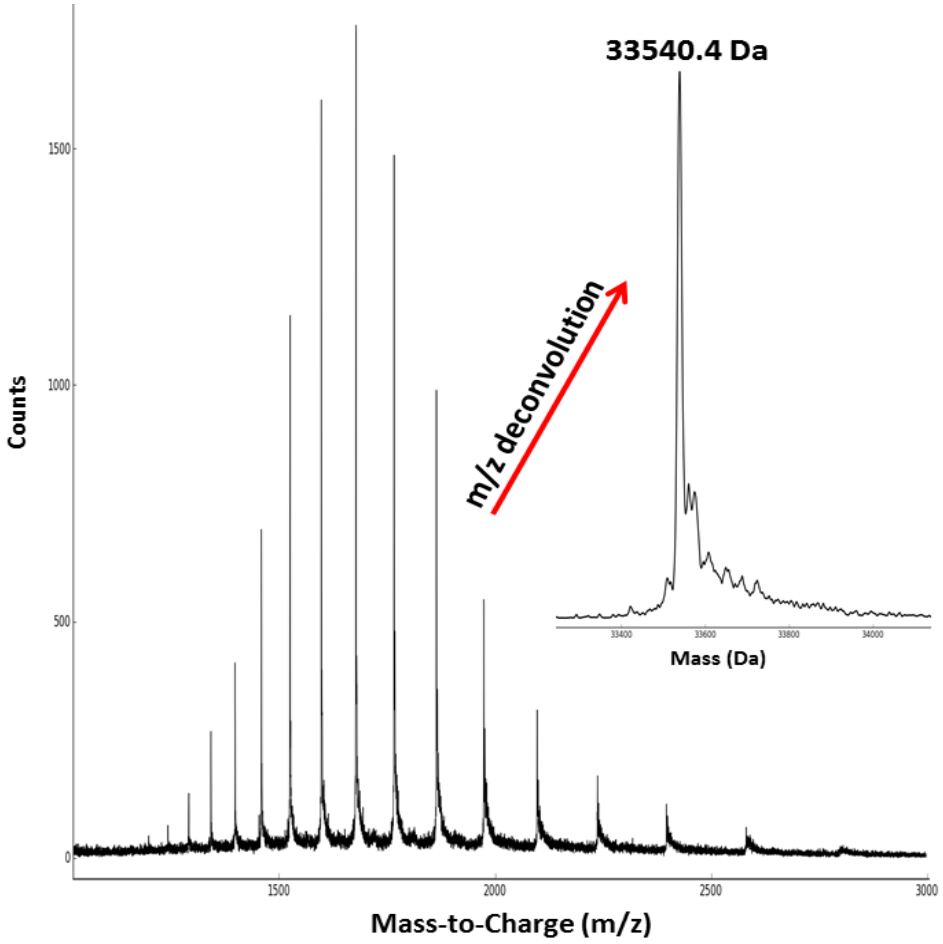

Figure 3.2: Charge envelope of HBV Cp149 in positive ionization ESI-MS mode. Despite high concentration of salt and buffer in the capsid solution, a high signal to noise m/z spectrum was obtained for HBV Cp149. Inset shows the deconvoluted spectrum. Data for deuterium uptake curves was obtained after deconvolution.

Hydrogen-Deuterium Exchange of HBV-Dimer Protein

Once the setup was shown to have acceptable reproducibility (Figure 3.1) and sensitivity, we tested the ability of the system to detect variations in the global and local flexibility of HBV Cp149 forms. At least two repeats were used for every deuteration reaction. Figure 3.3 shows the HDX-behavior of capsid versus dimer at 19°C. The very insignificant variation among repeated data points did not necessitate further statistical analysis.

A residue in an unstructured part of a protein has direct access to the solvent and a fast rate of deuterium uptake. Solvent exposed residues are defined as having a low protection factor meaning that a complete exchange will occur dependent on the inherent rate of the site. On the other hand, a residue involved in a stable hydrogen bond will show very slow or no exchange of H for D. This is observed for protein regions that are part of the hydrophobic core or stable secondary and tertiary structures such as helices, sheets and turns.

Upon assembly of Cp149 dimers into capsid, some residues exposed to the solvent are expected to become protected at the subunit interface. The observed rate of deuterium uptake for residues involved in protein-protein interaction would be expected to have slower uptake after assembly. These residues would therefore shift from to slow exchange regimes as well as showing a difference in protection. Figure 3.3 shows a significantly lower level of exchange per subunit for capsid. This was expected because the protein-protein contacts in capsid reduce global solvent accessibility. A second reason could be that conformational change is also constrained by capsid-related contacts. A

reduction in the number conformation or percentage of subunits in different conformations would decrease the diversity of the protein ensemble. We would term this as a decrease in conformational dynamics and would be consistent with a greater protection and slower rates of deuterium uptake.

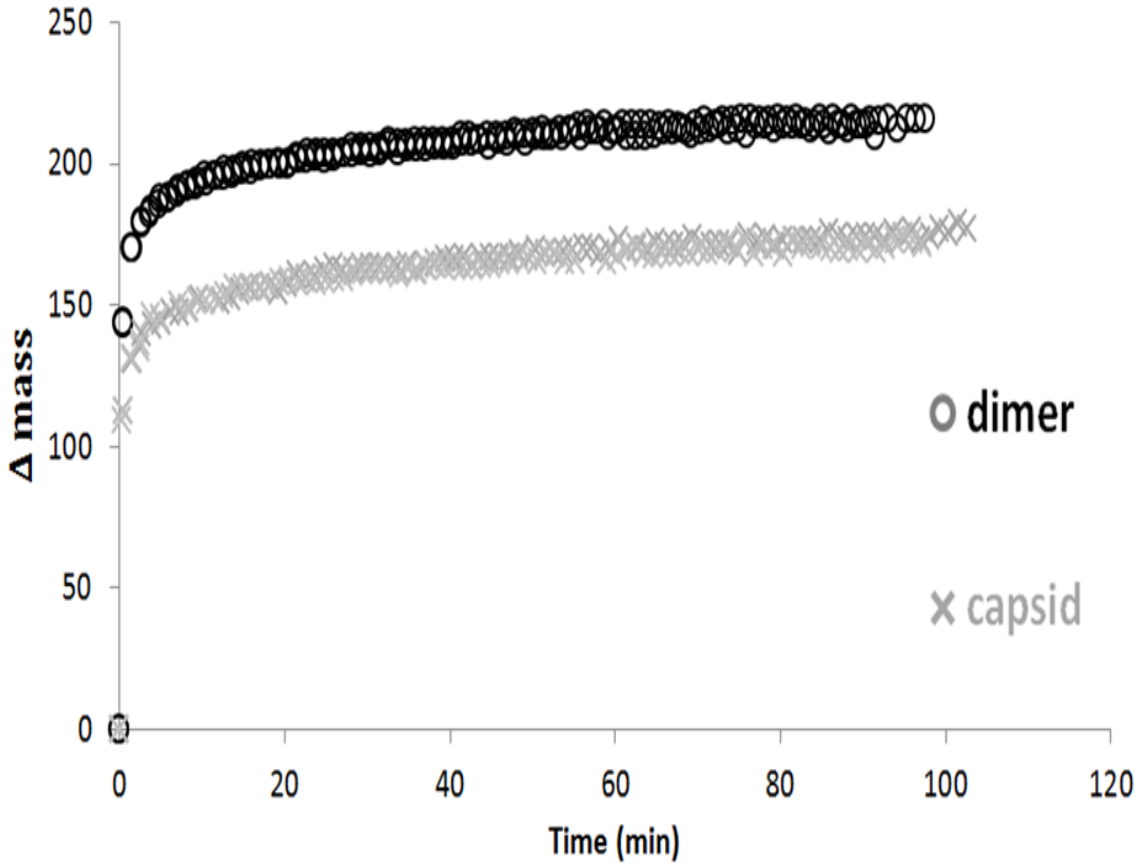

Figure 3.3: Deuteration uptake curves of HBV Cp149 dimer versus capsid at 19 ˚C. HDX experiments conducted as 2 and 4 replicates of deuteration reactions for capsid and dimer, respectively. Deuterium uptake curves show a significantly higher level of protection for subunits in capsid compared to the free dimer.

For the next experiment, the deuteration of dimers in free and assembled states at different temperatures was monitored. The effect of temperature on protein-HDX has two

major components: (1) a temperature effect on chemical kinetics of the HDX reaction which results in changing the rate of exchange for every residues; (2) a temperature effect on protein-dynamics such as the thermal motions and structural fluctuation which may in turn shift residues from a slow-exchange group to a fast-exchange (solvent accessibility). Therefore higher deuterium uptake level upon temperature increase is expected for every typical protein. The graphs in Figure 3.4 show overlays of the deuterium uptake curves at different temperatures of free and assembled dimer with at least three repeats for every sample. Increasing temperature can significantly increase the deuterium uptake for both free dimer and capsid. The overall pattern of overlaid curves in both free dimer and capsid is consistent with the general mechanisms of temperature effect on protein deuterium uptake (as was mentioned earlier).

We hypothesized that if the temperature is an activator for capsid assembly, then the effect of temperature effect on deuterium uptake would be different for free and assembled states. As it was discussed earlier, proteins are expected to exhibit higher deuterium level upon temperature increase. That would be because first, HDX reactions have faster kinetics in higher temperature, and second, overall protein structural movements are expected to be larger and faster in higher temperatures. What if temperature has a local structural effect which acts as a molecular switch? In the other words, structured parts of a protein can become unstructured upon a temperature change. Therefore, such a transition from structured to unstructured of a local region would be expected to contribute differently in corresponding deuterium uptake curves of before and after temperature change.

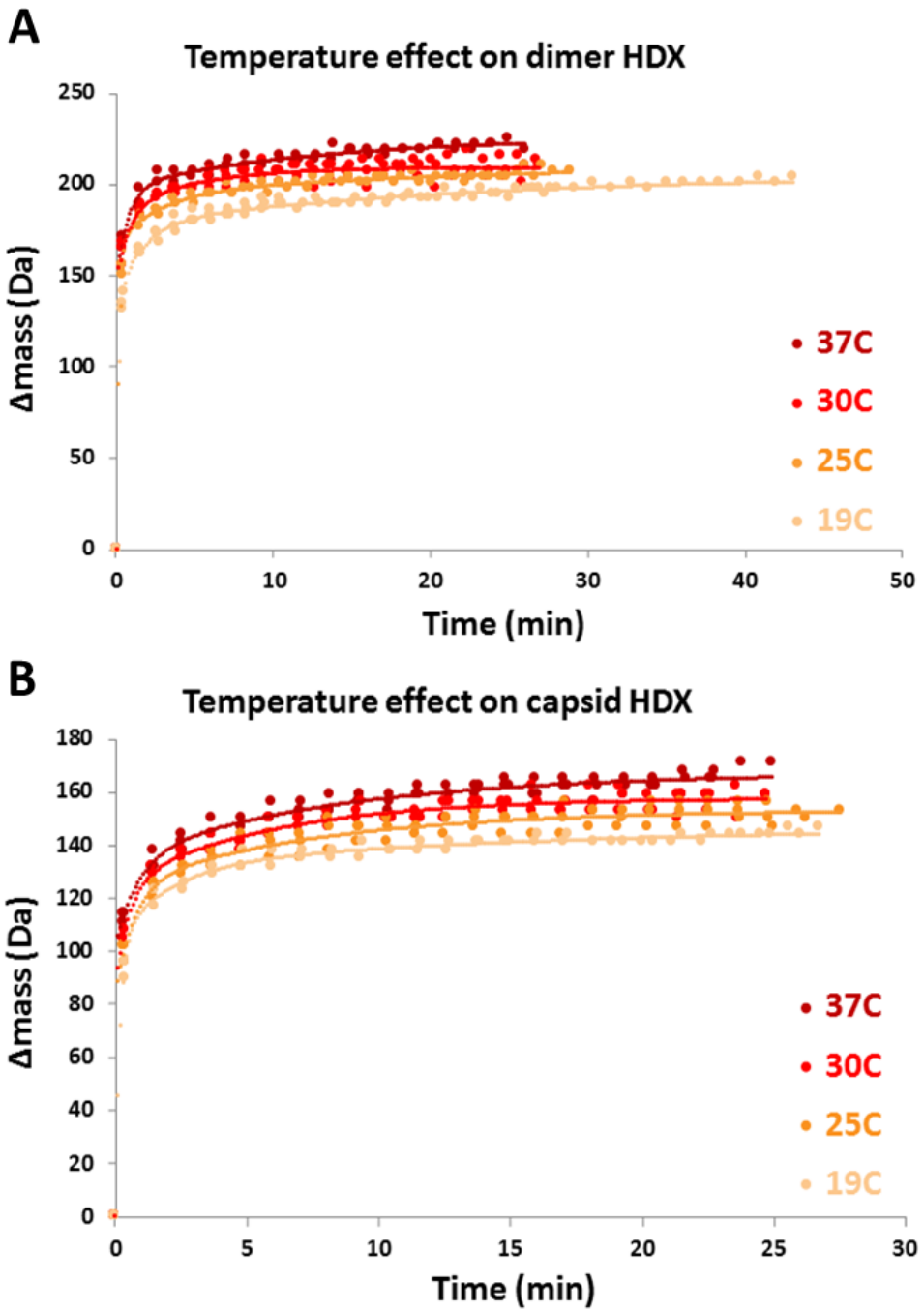

Figure 3.4: Temperature effect on deuterium uptake of HBV CP149 dimer and capsid. Deuterium uptake was monitored at 19 ˚C, 25 ˚C, 30 ˚C and 37 ˚C. Uptake curves for free dimer (A) and capsid (B) show an increase in uptake with temperature. (Data were collected as three replicates for every protein form at any given temperature).

To investigate the basis for the change in exchange that was observed, we looked at the temperature effect on HBV Cp149 dimer and capsid more deeply. Several temporal region or zones were selected across the deuterium uptake curve. Each zone has one dominant rate of exchange that contributes to the final observed uptake curve. A whole, the protein has a number of different rates that can defined as k_1, k_2 …k_n, with n being the number of different exchangeable groups. In practice, exchange rates are grouped into categories such as slow, medium, and fast. The rate of exchange at any point on the uptake curve can then be said to be mostly influenced by fast, medium, or slow exchanging groups. Factors determining the rate of exchange at a given protein structure site are solvent accessibility, dynamics, and chemical characteristics of the exchangeable hydrogen. To investigate the temperature effect on HBV capsid protein, we selected time zones of 1, 3, 5, 10, and 24 minutes with the corresponding rates of $k_{1minute} > k_{3minute} > k_{5minute} > k_{10minute} > k_{24minute}$. The protein mass shift per minute for every time zone is not the rate of exchange, however it can be considered as a rate-equivalent for that time domain. For every time zone, there is also a contributions from other time zones. For example the mass gain in the 3 min time zone is a summation of almost the total mass gain before the 3 min time zone, almost half of the mass gain during 3 min time zone, and some minor contribution from the 5 minute time zone. We can use a time zone rate-equivalent and graph it as a function of temperature. The model and slope of this graph is the representative of the temperature effect on the exchange behavior of hydrogens in that time zone. Using HDX data on the effect of temperature on global exchange of HBV capsid, we graphed each time zone separately. Figure 3.5 shows these five curves and

their slopes corresponding to the temperature effect on HBV Cp149 capsid. We combined 3 replicates of the HDX-reactions and this is shown in Figure 3.4 panel B. Data was fit using a triple exponential to represent the slow, medium, and fast exchanging groups of residues. Higher orders of multi-exponentials were tried but did not improve the simulation process. We then used the simulated data as the average uptake curve at each temperature. Data in Figure 3.5 are calculated from simulated curves.

To understand this data, we begin by looking at the 1 minute time zone. As it was mentioned before, the temperature increase leads to an increase in the observed exchange. The chemical mechanism is works by either increasing the collision rate which improves the chemistry of labeling or by increasing thermal motions and local fluctuations (short-range breathing) for all molecules in the solution resulting in more solvent accessibility. If the chemical effect is the only factor, there should be a gradual increase in the rate-equivalent of the 1minute time zone as a function of temperature (positive slope). In the other word, the graph of the rate-equivalent in a given time zone versus temperature is expected to be a line (or a quasi-line) with a positive slope. Linear graphs with positive slopes for time zones of 1, 3, 5, 10 and 24 minutes were obtained. As it was pointed before, rate-equivalents are cumulative and the 5 min rate-equivalent is a summation of 3 min, half of 5 min and a minor part of 10 min rate-equivalents. If this is the case we expect to see a gradual increase in the slopes of the lines as we move from earlier to later zones. Figure 3.5 shows such a gradual increase in slopes of graphs of rate-equivalents versus temperature from HDX data on capsid. In the other words, the temperature effect on capsid structure is a general effect resulting in an increase in chemical reaction rate

and molecular motions. If we put these explanations in a more dynamics-related context, it relates to decreased protection, i.e. higher temperature = increased flexibility = less protection against exchange. A gradual protection loss as a function of a 'gradual' temperature increase results in a 'gradual' increase in exchange rates of residues.

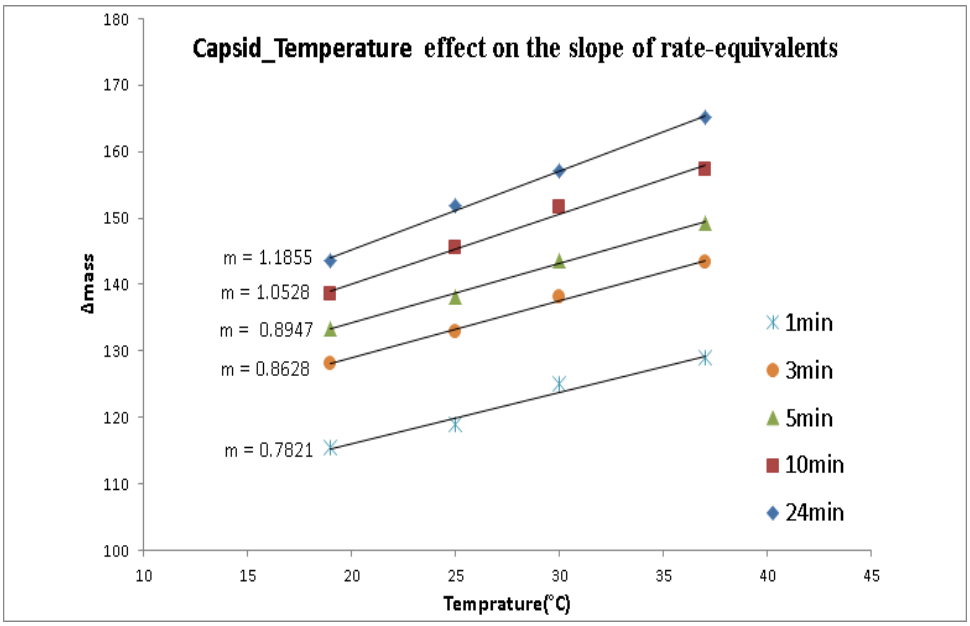

Figure 3.5: Graphs of the rate-equivalent versus temperature for time zones across the capsid deuterium uptake curve. The model and slope of every graph is the representative of the temperature effect on the exchange behavior of hydrogens in that specific time zone.

The next step consisted of repeating the analysis for HBV dimer. A similar set of slopes for rate-equivalents versus temperature was obtained for dimer Cp149 (Figure 3.6). We combined 3 replicates of the HDX-reactions and this is shown in Figure 3.4 panel A. Data was fit using a triple exponential to represent the slow, medium, and fast exchanging groups of residues. We then used the simulated data as the average uptake curve at each temperature. Data in Figure 3.6 are calculated from simulated curves.

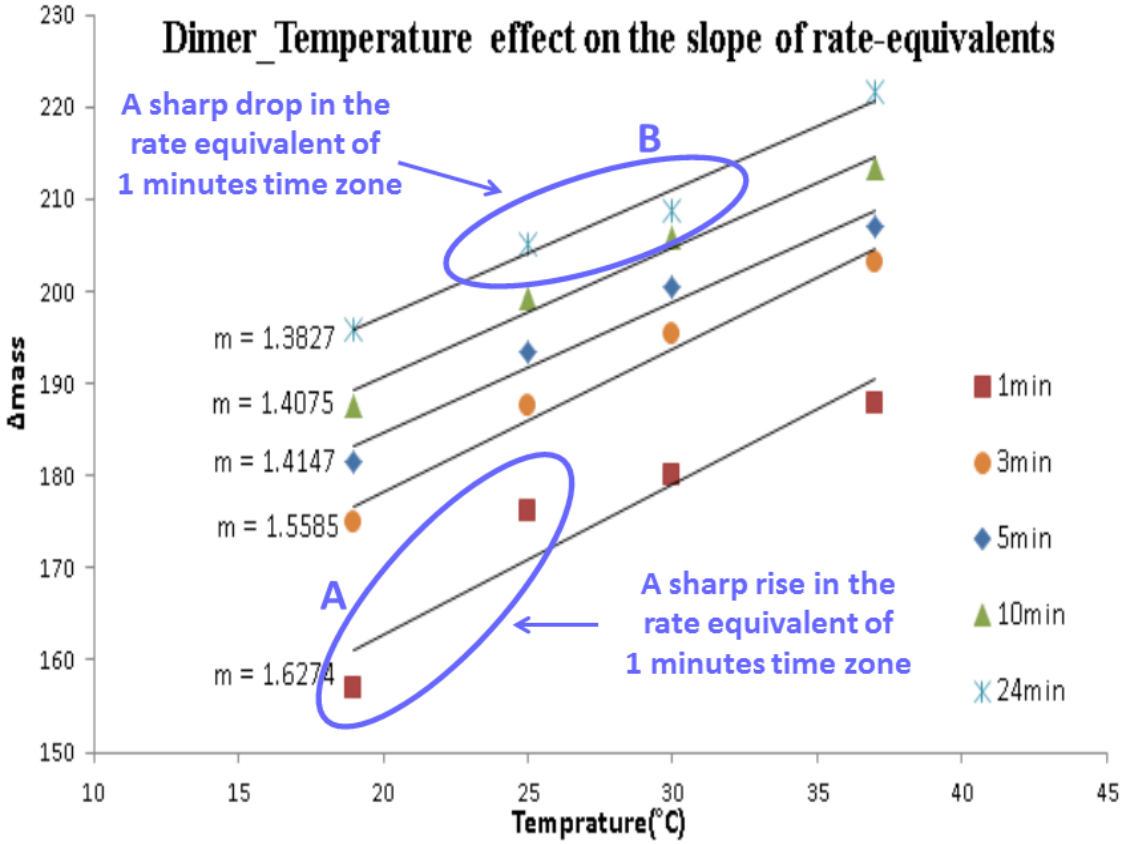

Figure 3.6: Graphs of the rate-equivalent versus temperature for time zones across the dimer deuterium uptake curve. The model and slope of every graph is the representative of the temperature effect on the exchange behavior of hydrogens in that specific time zone. The highlighted area A shows a sharp rise in the rate-equivalent for the 1 minute time zone data from 19 ˚C to 25 ˚C. The highlighted area B shows a sharp drop in the rate-equivalent of 24 minute time zone upon a temperature increase from 25 ˚C to 30 ˚C.

Here is can be seen that the two forms of Cp149 (dimer and capsid) had different levels of total deuteration (total protein mass gain). This makes the normalization process more complicated. Therefore, the data shown in Figures 3.5 and 3.6 were not normalized. For the remainder of the analysis, we will consider only the pattern of slopes rather than their absolute values.

All the curves in Figure 3.6 have positive slope which is what we expect to see based on the temperature effect on every protein. However, there are some unexpected

behaviors. The first unusual behavior is that slopes corresponding to the temperature effect on dimer in Figure 3.6 show a different pattern compared to the observed pattern for capsid in Figure 3.5. Slopes corresponding to dimer show a gradual decrease from the 1 minute time zone through the last at 24 minutes. An explanation of this data based on a simple scenario is that structured parts of a protein that contribute primarily to slower zones become more mobile at higher temperature, shifting their contribution to faster zones. What if the protein is already highly flexible (or highly unstructured) with a high rate of exchange for most of its hydrogens? Even at low temperatures, most of the residues would exchange quickly for such a protein which means the majority of residues exchange in early zones (for example 1 and 3 min). However a small population of residues in highly structured parts of the protein has a slower exchange rate. If higher temperature converts the last remaining structured regions into unstructured ones (lacking stable H-bonds), these would shift to faster exchange zones and almost all the residues would exchange in early time zones. Such a curve is observed with the Cp149 dimer in Figure 3.7. The deuterium uptake curve of such a protein at higher temperature would have a steep uptake curve initially (the highlighted area A in Figure 3.7) and would then level off after just a short time (the highlighted area B in Figure 3.7). The figure has been simplified and only shows uptake curve of dimer at 30 ˚C and 19 ˚C. Data fitting using three independent exponential rates (as described above) matched the data well (simulation data are not shown). Based on these simulations, upon changing the slow exchange rate of a group of residues into a fast exchange rate, the shape change of the uptake curve showed the same pattern of the observed shape change for dimer upon the

temperature change from 19 ˚C to 30 ˚C (Figure 3.7). Based on the same simulation studies, upon adding the same quantity to all of exchange rates, the shape change of the uptake curve showed the same pattern of the observed shape change for capsid upon the temperature change from 19 ˚C to 30 ˚C (Figure 3.7). The change in the shape of the uptake curve of dimer at higher temperatures (Figure 3.7) is the reason for the observed gradual decrease of slopes in Figure 3.6.

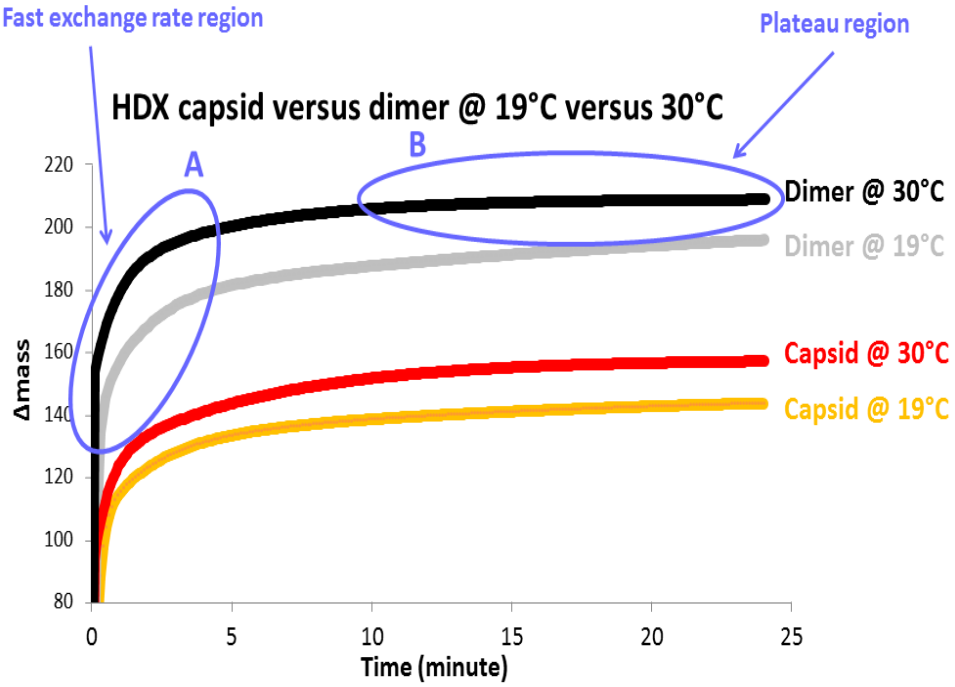

Figure 3.7: Temperature effect on the shape of deuterium uptake curve of capsid versus dimer. Uptake curves are based on statistically weighted average of at least three replicates of HDX reactions for every temperature. Deuterium uptake at 19 ˚C and 30 ˚C for capsid and dimer correspond to the color code in the legend. The inset figure A shows the region of fast exchanging residues in the uptake curve of dimer. The inset B shows the plateau region of the uptake curve of dimer.

A second interesting behavior was a sudden rise in the rate-equivalent of 1 minute time zone upon the temperature increase from 19 ˚C to 25 ˚C (the highlighted area A in Figure 3.6). The observed sharp rise is consistent with a sudden change of exchange rates

for a group of residues from very slow to very fast as was hypothesized in earlier paragraphs and was supported by data simulation on a theoretical uptake curve model. In other words, a local temperature dependent structured-to-unstructured transition can be the reason for different patterns of temperature effect on dimer versus capsid.

What is the source of these slow exchanging residues which moved into the 1 minute time zone during the 19-25 ˚C transition? The highlighted area B in Figure 3.6 shows a sharp drop of the rate equivalent corresponding to the 24 minute time zone during the 25-30 ˚C transition. We can hypothesize that a structured part contributing in the slow exchange time zone of 24 minutes becomes unstructured upon the 19-25 ˚C temperature effect and its corresponding residues acquire the exchange rate of 1 minute time zone. For next temperature increase (25-30 ˚C), these shifted residues are no longer contributing in deuterium exchange of 24 minutes time zone which results in a sharp drop of the rate equivalent corresponding to the 24 minute time zone during the 25-30 ˚C transition.

We graphed the slopes of the data of Figures 3.5 and 3.6 versus time zones for a better visualization. Figure 3.8 shows the comparative temperature effect on capsid versus dimer. The graphs of "slopes versus time zones" of capsid and dimer in Figure 3.8 show totally different patterns. As we hypothesized earlier, the effect of temperature as an activator of capsid assembly can be differentiated between free and assembled states by a sensitive HDX-LSMS approach. The distinctive pattern of the graph of "slopes versus time zones" for dimer in comparison with capsid in Figure 3.8 is consistent with results from the lab of Adam Zlotnick. They have looked extensively at the ability of

temperature to activate capsid assembly [36, 60] and have shown that temperatures above 20 ˚C are a strong trigger for capsid assembly. If temperature is a strong activator of assembly, it means that assembly is an entropically driven process and that a structured-to-unstructured transition is acting as a thermodynamic switch to regulate the kinetics of assembly [60].

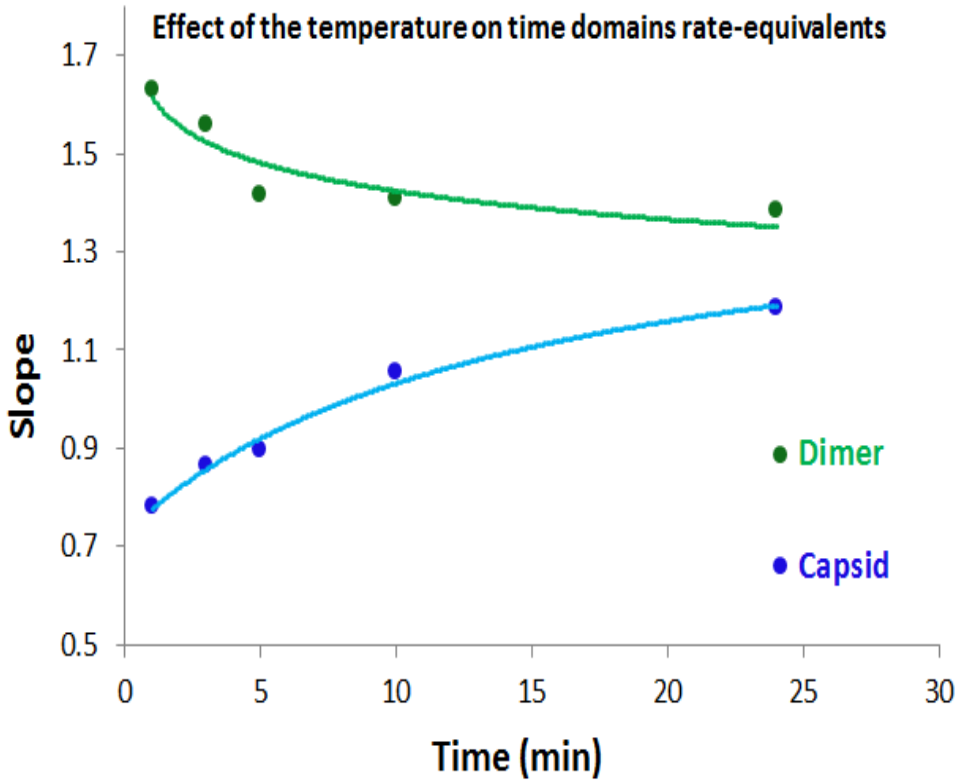

Figure 3.8: Graph of temperature effect on slopes of rate-equivalents versus time zones for capsid and dimer. The comparative temperature effect on rate-equivalents of time zones of capsid (blue) versus dimer (green). Every data point is the slope of the temperature effect from 19 ˚C to 37 ˚C on the rate-equivalent of a given time zone of deuterium uptake curves. The different pattern of temperature effect on dimer compared to capsid is indicative of the distinctive nature of conformational changes between them. (Explanation on the statistical approach for this graph is in the text).

Structural models of the Cp149 dimer in solution suggest that the conformation is strained [63, 64]. This strain originates from what have been termed “frustrated” protein regions that lack contacts in solution. In the assembled capsid they would be involved in inter- and intra-protein contacts [65, 66]. These strains can act as an energetic compromise for capsid assembly [24, 38]. Higher temperature may increase these strains by disordering structural parts, therefore improving the thermodynamic and the kinetics of assembly. Low temperature or mutation at position Y132 can increase the energy barriers necessary for assembly by relieving the protein from these strains. Such a relaxed structure is believed to decrease the favorable entropic contribution to the assembly process. An entropic difference between “chaotic disordered” state (Cp149 as free dimer) and “buried in hydrophobic area” state (Cp149 in capsid) potentially plays an important role in capsid assembly. [The basis for kinetically enhanced assembly with respect to disordered regions will be discussed further in chapter 4]. We can define four basic steps during the inter-protein interaction event: random diffusion, directional diffusion, encounter complex, final stable complex. If extended unstructured parts are involved at inter-dimer contact regions, they increase the directional flexibility, reducing steric hindrance, facilitating the ‘encounter complex’ [67]. This mechanism was recognized early on as part of the reason that disordered regions are common in protein docking domains [6, 8].

The next step was to investigate the effect of ionic strength on Cp149. This was done by varying the concentration of NaCl salt in HDX experiments of capsid, dimer, and the mutant Y132A. The mutation of interest was at the position Y132, because it has

been shown to have an important on the assembly process. The mutant Y132A is an assembly defective dimer [24, 38]. Increased ionic strength in the form of salt is required for assembly. Salt can have a general effect on protein structure and dynamics by screening (weakening) electrostatic interactions and therefore reducing the thermodynamic cost (energy barrier) of conformational changes [60, 63, 68]. This mechanism is consistent with what has been reported for HBV [23, 28, 37, 39, 60, 68].

We carried out HDX on Y132A, the assembly defective mutant, with and without 100mM NaCl. Figure 3.9 shows the effect of salt on Y132A. HDX data showed no significant salt effect on global conformational flexibility of Y132A (or any effect that could be detected by HDX). Because we are comparing the total deuterium uptake of intact proteins, we use the term ‘global’ unless the data can fit a specific model for HBV.

The next step was to investigate the effect of salt on Cp149 dimer. This was done by conducting HDX experiments on Cp149 dimer with and without 100 mM NaCl. Based on HDX data on the salt effect on the mutant Y132A, we were not surprised to see no significant change in the deuterium uptake curve of HBV Cp149 dimer by salt. This was because Cp149 dimer and Y132A have highly similar structure in both crystal and solution [24, 38]. However the HDX data showed a very significant conformational change of HBV Cp149 dimer in the presence of 100 mM NaCl salt.

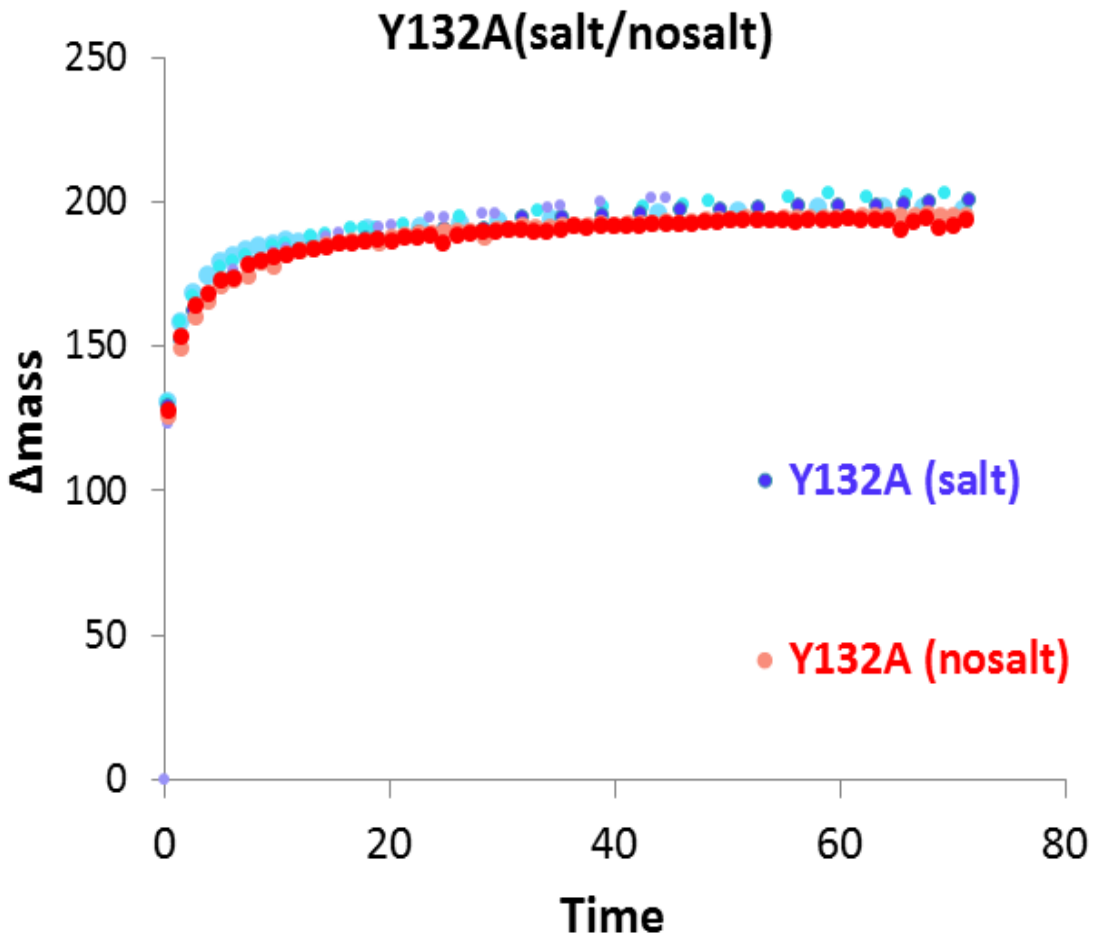

Figure 3.9: Comparative deuterium uptake of Y132A in the presence and lack of NaCl. Replicate uptake curves are shown in the presence of 100 mM NaCl salt (blue and cyan) and with no salt condition (red and pink). Salt does not have a significant effect on global exchange of Y132A mutant. This suggests that ionic strength is not altering the structure or dynamic properties of the mutant. (Data were collected as three replicates for any given salt condition).

For a better comparison between the effects of salt and mutation at the position Y132 on the global conformational flexibility Cp140 dimer, we overlaid the HDX data of Y132A at high ionic strength (100 mM NaCl) on the HDX curves of Cp149 dimer under salt (100 mM NaCl) and no salt conditions. Figure 3.10 shows the comparative HDX data on HBV Cp149 dimer versus Y132A in the presence and lack of salt. The mutant Y132A

had less global exchange (greater rigidity) than wildtype Cp149 when both were tested in 100mM NaCl. This result is consistent with previous work that pointed out that Y132A structure was slightly different from Cp149 in capsid especially in some subdomains around the hydrophobic core such as inter-dimer contact and spike regions [24]. If unstructured-to-structured switches are thermodynamically coupled to protein-protein contact events, a more compact structure would reduce the role of these molecular switches. If so, the Y132A which has a more compact structure than Cp149 cannot couple such an unstructured-to-structured switch to assembly. This may be the reason for its assembly defectiveness. The Y132A mutation is relaxing the conformation into a less strained form [63]. The alleviated structural strain in Y132A is possibly coupled with some α-helix stability which can be interpreted as conformational compactness as was observed in the HDX data as less deuterium uptake. However HDX data predicts a very rigid conformation for Cp149 dimer with no salt which is also considered an assembly-defective form of Cp149. This suggests that conformational compactness is responsible for capsid assembly deficiency of Y132A. In other words, is some conformational freedom a necessary condition for capsid assembly?

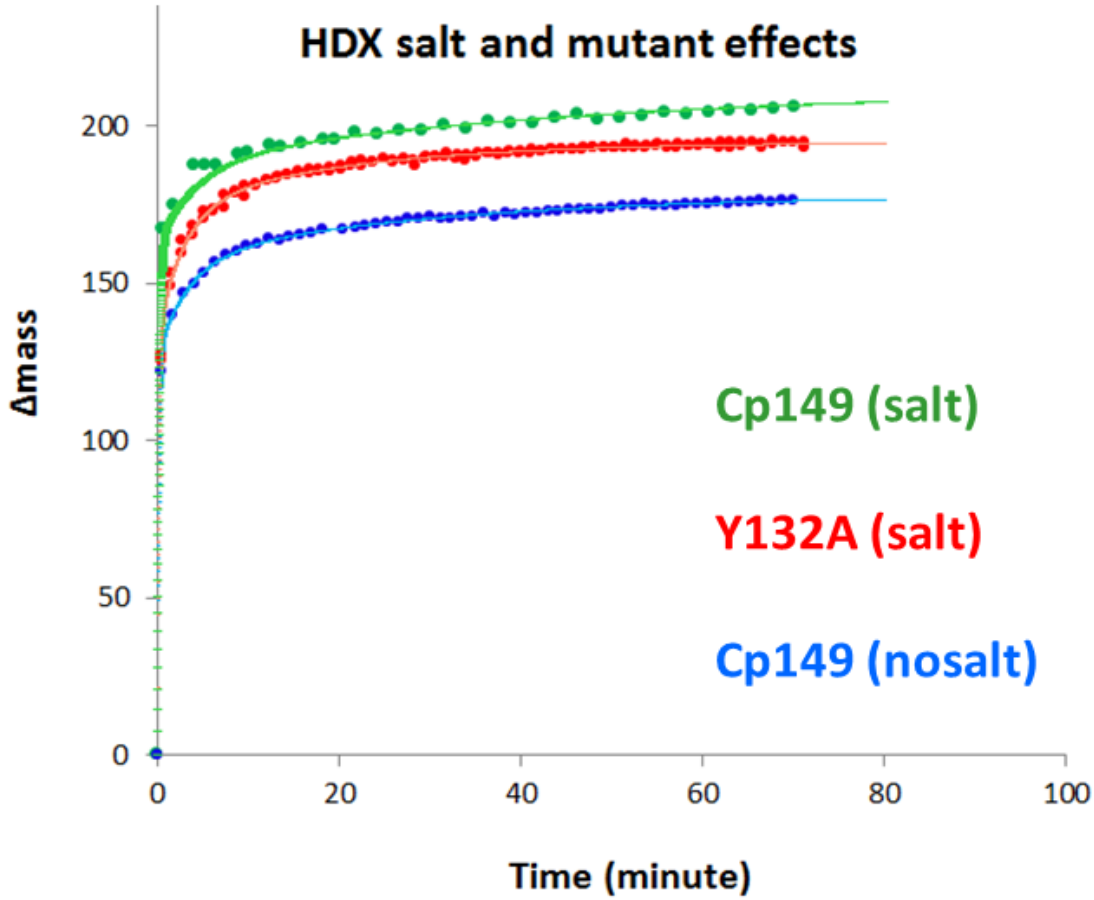

Figure 3.10: The comparative effects of salt versus a mutation at Y132 on the conformation of Cp149 dimer. HDX data shows that salt significantly increases the global exchange of Cp149. We can consider the conformation of Cp149 in salt as the native conformation. The mutation Y132A or lowering of NaCl decreases the natural flexibility of Cp149. As shown above, salt did not have a significant effect on the global exchange of Y132A. (Data were collected as three replicates for every protein form at any given salt condition).

To further investigate structural properties of HBV Cp 149, analytical HPLC based size exclusion chromatography (SEC) was used. SEC showed a significantly earlier elution time for Y132A compared to Cp149. This implies a larger apparent hydrodynamic radius in solution for Y132A (Figure 5.6). HDX data showed a more compact conformation for Y132A (Figure 3.10). One scenario that is consistent with both SEC and HDX data is that two ‘compact’ monomers of Y132A form a structurally

extended dimer compared to Cp149 dimer. If so, either the structure of every monomer is extended or the dimer subunit contact area is disrupted. In the latter case, changes to the dimer subunit contact area could destabilize four-helix bundle which in turn expected to reduce the global stability of the dimer. To test this hypothesis, Differential Scanning Fluorometry (DSF) was conducted to investigate the thermostability of Y132A versus Cp149. DSF data showed a significantly lower thermostability for Y132A dimer (Figure 5.7). A possible explanation can be the disruption of dimer subunit contact area which could destabilize the four-helix bundle jeopardizing the integrity and stability of the dimer structure.

Similar to the temperature effect, we graphed rate-equivalents from the 5 time zones at different salt concentration. To be visually comparable to temperature effect graphs, we considered no salt equal to the minimum temperature (19 ˚C) and 100 mM NaCl equal to the maximum temperature (37 ˚C). We also graphed rate-equivalents for wild type Cp149 and the Y132A mutant using the same approach with Y132A equal to 19 ˚C and the wild type Cp149 under salt condition equal to 37 ˚C. Here again all slopes should be positive. Figure 3.11 shows slopes of rate-equivalents of time zones across the deuterium uptake curve of Cp149 dimer in response to 100 mM NaCl and mutation at Y132. Similar to the temperature effect on Cp149 dimer, slopes of salt show a gradual decrease across time zones from 1 min to 3min and etc. Similar to the temperature effect on capsid, slopes of Y132A mutation effect show a gradual increase across time zones from 1 min to 3min and etc.

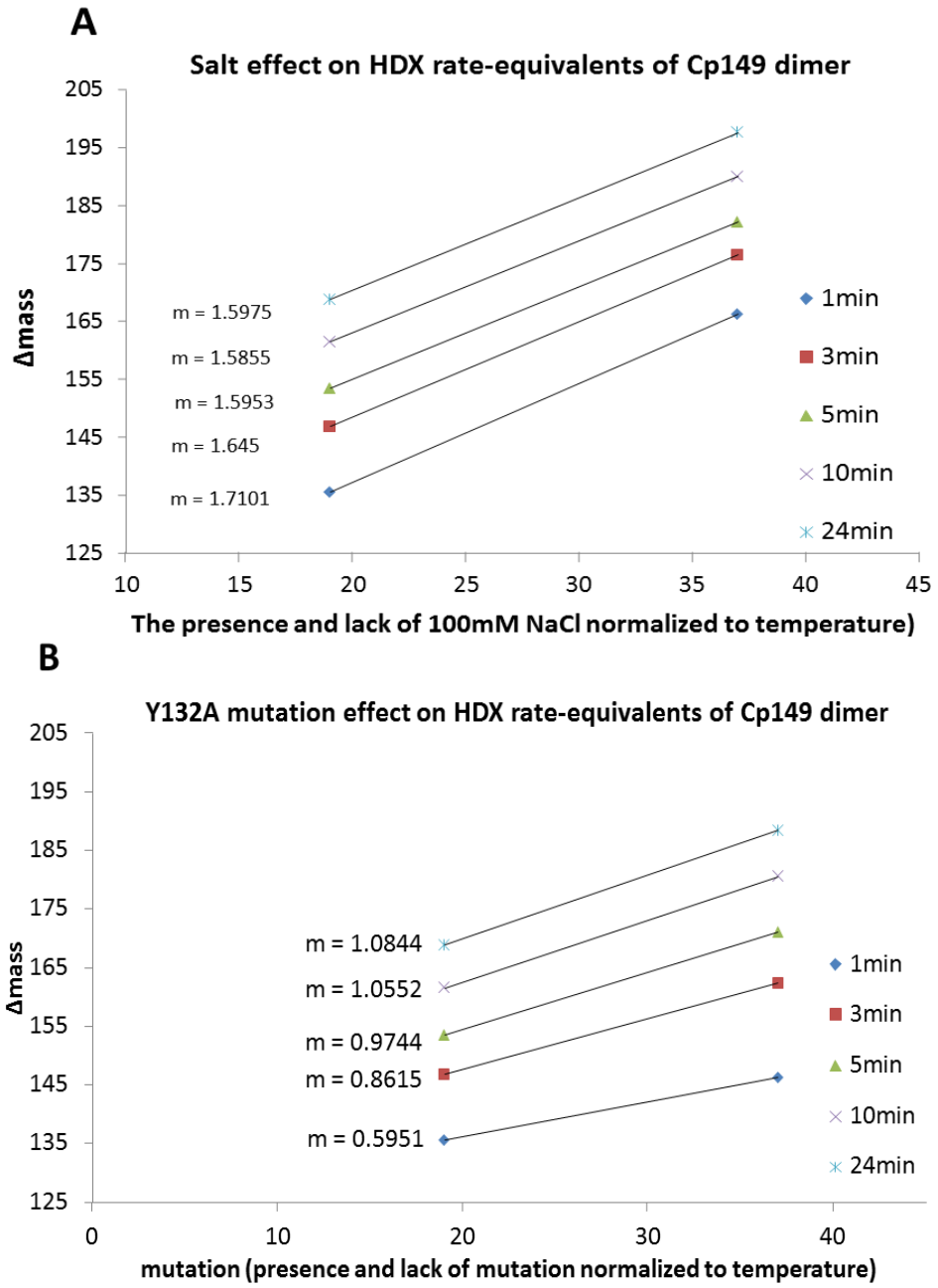

Figure 3.11: Graphs of rate-equivalent versus salt and Y132 mutation for time zones across the deuterium uptake curve. (Explanation on the statistical approach for this graph is in the text). The slope of every graph is the representative of the salt and Y132 mutation effect on the exchange behavior of hydrogens in that time zone. A) Graph of rate-equivalents versus salt. The presence and lack of 100mM NaCl were normalized to temperatures 19 and 37 ˚C, respectively. B) Graph of rate-equivalents versus mutation at Y132. The presence and lack of Y132A mutation were normalized to temperatures 19 and 37 ˚C, respectively.

We combined 2-4 replicates of the HDX-reactions and this is shown in Figures 3.9 and 3.10. Data was fit using a triple exponential to represent the slow, medium, and fast exchanging groups of residues. We then used the simulated data as the average uptake curve of either Cp149 or Y132A at each salt concentration. Data in Figure 3.6 are calculated from simulated curves.

For better visualization of these slopes of rate-equivalents, we graphed them as slope versus time zones in Figure 3.12. This is similar to what we did for the temperature effect, in Figure 3.8. Figure 3.12 panel A shows a gradual decrease in rate-equivalents upon salt. As we observed for the temperature effect on dimer such a gradual decrease in slopes of rate-equivalents could be explained by a local conformational change in a form of a local structured-to-unstructured switch. Regarding the strong effect of salt as an assembly activator, we did not investigate the effect of higher concentrations of NaCl by HDX. We conducted only two sets of experiments on Cp149 under salt and no salt condition depending on the presence or lack of 100 mM NaCl, respectively. However we cannot rule out that a factor such as a structured to-unstructured transition may be involved in the observed pattern of slopes corresponding to the salt effect. HDX data suggests a significant structural disorderness, either local or global, upon adding 100 mM NaCl to Cp149 dimer. Figure 3.12 panel B shows a gradual increase in rate-equivalents upon replacing alanine (A) with tyrosine (Y) at position 132. As we observed for the temperature effect on capsid such a gradual increase in slopes of rate-equivalents could be explained by an increase in global conformational flexibility (global solvent accessibility). To investigate the effect of mutation at Y132 by HDX, we used Cp149 and

of Y132A as a two point perturbation. We cannot rule out the presence of any structured-to-unstructured conformational switch for the mutation effect even though the effect of mutation on dimer and the effect of temperature on capsid show a similar pattern in their graphs of slopes of rate-equivalents versus time. However the circular dichroism (CD)and fluorescence studies did not show any structural differences between Y132A mutant and the wild-type Cp149 dimer protein [24]. Therefore we can rule out any significant change in the α–helix content of Cp149 as a possible mechanism of its conformational change upon mutation at Y132.

Comparative HDX data on Cp149 and the mutant Y132A suggests more conformational compactness for the mutant (Figure 3.10) which is not expected to be as a change in α–helix content (CD and fluorescence data). in the other hand, the SEC and DSF data are not consistent with such a structural rigidity in Y132A mutant (Figures 5.6 and 5.7), unless we hypothesize that such a conformational compactness happens only in the hydrophobic core (a more stable core) which should be compromised by, for example, interrupting inter-monomer contacts and destabilizing the four-helix bundle.

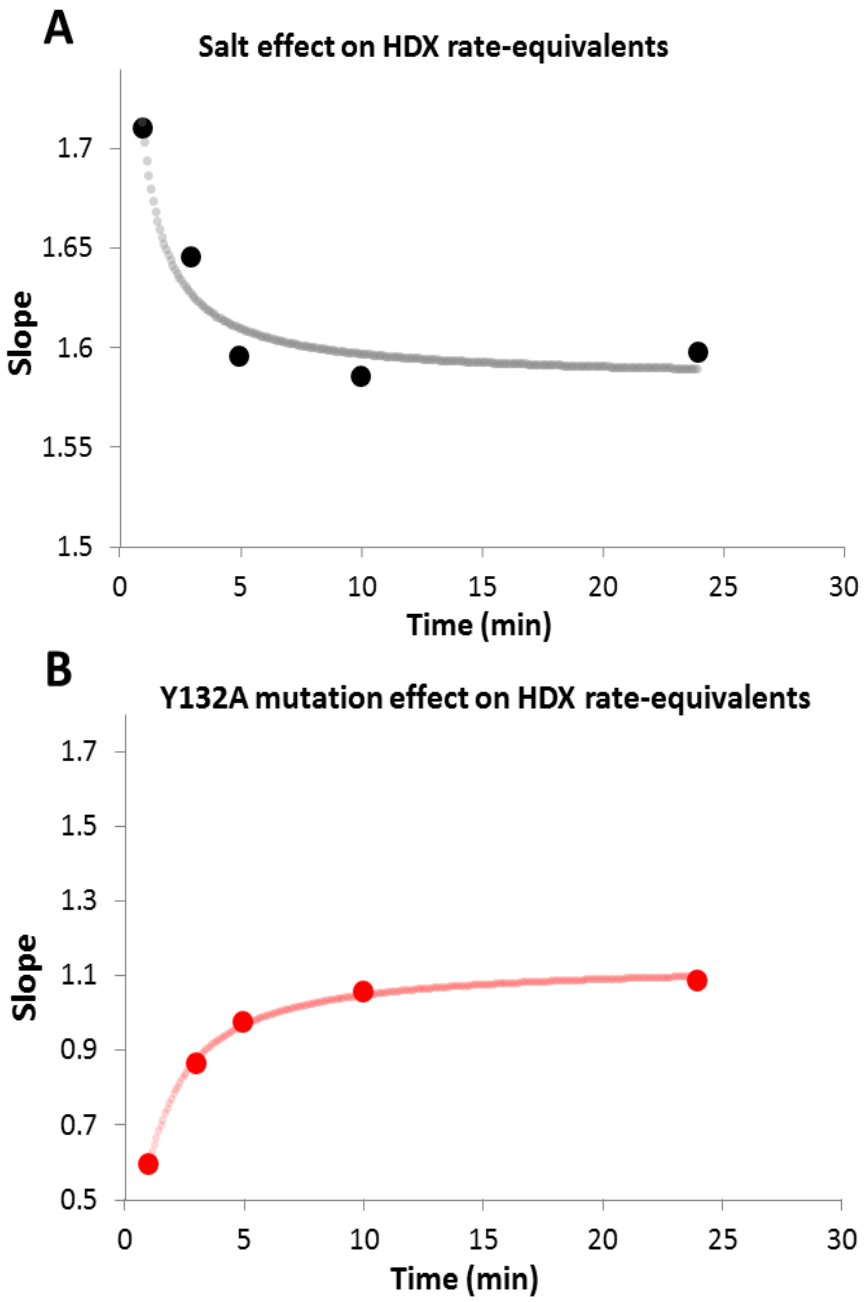

Figure 3.12: Graph of salt and Y132A mutation effects on slopes of rate-equivalents versus time zones. A comparative salt and Y132A mutation effect on rate-equivalents of time zones. Every data point is the slope of the salt (panel A) and Y132A mutation (panel B) effects on the rate-equivalent of a given time zone of deuterium uptake curves. The different patterns of salt and Y132A mutation effects are indicative of the distinctive nature of conformational changes between them. (Explanation on the statistical approach for this graph is in the text).

Conclusion

Ultrafast HDX coupled to LC-MS proved to be an automated high throughput approach for rapidly screening the effect of a wide range of assembly stimulators on HBV Cp149. The data presented here have allowed us to put forth a biophysical description of structural changes that can account for the observations that temperature, ionic strength and mutations alter the kinetics and thermodynamics of capsid assembly. Our results of such an approach probing the effects of temperature, salt and mutation on the HBV capsid protein are consistent with what has been reported [24, 38, 60, 63, 64]. We could not only quantify the effect of temperature but also qualify its allosteric effect as an assembly stimulator. The similar HDX studies on the effect of salt and Y132 mutation showed that they both strongly but distinctively impact the conformation of Cp149.

These preliminary HDX data suggested the presence of a molecular switch as an allosteric mechanism for capsid assembly. To further investigate such a molecular switch-based mechanism for capsid assembly, we needed to map the structural signal transduction onto specific regions of Cp149. To do that, we used comparative high-resolution HDX. This work forms the basis for Chapters 4 and 5 that follow.

HIGH RESOLUTION HYDROGEN-DEUTERIUM EXCHANGE ANALYSIS OF HBV DIMERS

Introduction

High Through-put HDX-LCMS Coupled with On-line Pepsin Digestion

Although considerable insight was gained about the functional aspects of HBV capsid protein from the intact HDX experiments, a number of important questions were raised. One of the significant questions was what where the specific regions involved in the conformational switches discussed in chapter 3. Based on intact protein HDX data, we knew that the free form of Cp149 dimer adopted a different conformation after mutation at Y132A or changes in salt or temperature. But where are these dynamically important regions involved in such a variable conformation? What is the mechanism of these molecular switches? To answer these questions, we needed a detailed map on the dynamic of the Cp149 protein.

Therefore the next step was to develop a pepsin-based HDX system. The need for a high resolution approach is related to the supramolecular capsid assembly containing 240 copies of the HBV protein. Because of the architecture of the capsid, subtle change in the conformation of the HBV Cp149 protein could change the conformation and thermodynamics of assembly. Multiplying subtle changes across the 240 subunits of the capsid could significantly affect the thermodynamics of the assembly. Therefore conformational changes as the consequences of assembly effectors are expected to be subtle. Therefore, a method very sensitive to changes in HDX uptake and protection

would be the best chance of understanding how allosteric actions impact HBV capsid assembly.

Materials and Methods

Methods

Automated Ultrafast Hydrogen Deuterium Exchange: a HDX setup including continuous deuterium labeling [55] followed by on-line pepsin digestion (immobilized-pepsin column) and automated fast LC-MS was developed. All labeling experiments were done in a temperature controlled micro-well plate LC-sampler by (automated/or manually) initiation of reactions by 10 fold dilution of protein samples into a D_2O-tris + NaCl (H_2O-tris for the nondeuterated reactions). The concentration of NaCl, temperature, and the pD varied depended on the experimental design. Injection was coupled to on-line parallel immobilized-pepsin columns. Digestion occured in 3% acetonitrile, 0.02% formic acid and a varying concentration of urea (0.8-1.2 M) in H_2O based LC solvent. Released peptides were trapped and desalted on-line using C8-macroTrap 3×8mm (Michrom Bioresources, Inc. Auburn, CA). Peptides separation was done through a C18 1.7 µm, 1 mm × 100 mm column (Phenomenix) held at -1 ° C for 5 min and by a multi-linear multi flow rate gradient of acetonitrile and 0.1% formic acid as LC solvents. LCs were done using an Agilent 1200 LC and 1290 uHPLC systems (Agilent Technologies, Waldbronn, Germany) with an automated degasser, temperature controlled micro well-plate samplers, and a column heater compartment. Peptide masses were measured using an Agilent 6538 Accurate-Mass Q-TOF mass spectrometer with electrospray ionization.

The mass spectrometer was operated in positive mode scanning over the range 50-1700 m/z. Peptides were identified and mapped onto the Cp149 sequence using Agilent MassHunter Workstation Software, Qualitative Analysis Version B.04.00 and the in-built BioConfirm software. MS/MS data were analyzed and mapped onto the protein sequence using the Bruker Daltonics Bio Tools 3.2. The identification tolerances were 0.002m/z (±) 5ppm and 0.002milli mass unit for MS and MS/MS, respectively. We used the Sierra Analytics HDExaminer Version 1.3.0 beta6 software to calculate centroids of isotopic clusters of a peptide during time-course HDX experiments. Time points from 5 replicates were randomly combined to improve the curve fitting process into multiple-exponentials. For complementary MS/MS analyses, off-line mass spectrometry was carried using an Agilent 6520 Q-TOF with Agilent 1100 nano-HPLC and a Chip Cube or a Bruker maXis with Dionex 3000 nano-uHPLC. Off-line MSMS analyses were conducted by injecting nondeuterated protein samples into the LC-MS setup described above, except that peptides were collected before they enter the peptide trap column (Figure 4.1). The details of the LCMS acquisition method can be found in appendix A.

Sample Preparation: Protein dimers were obtained from Dr Zlotnick lab. Truncated HBV capsid protein (Cp149) dimers and capsids were expressed and purified from E. coli at 37 °C as described by Zlotnick and et al. [61, 62]. HBV protein represented 5-15% of the total E. coli proteins. Cells were lysed by French press and sonication, capsids separated on sucrose gradient and $(NH_4)_2SO_4$ and purified using a Sepharose column and ultrafiltration membranes. Dimers were obtained by diluting capsids in pH 9.5 and 3.5 M urea and purified by a Superdex 75 column. Dimers were

filtered and concentrated with pressure-based stirred cells, 10 kD ultrafiltration disk or Centrifuge-based ultrafilter 0.5 mL, 10 kD (Amicon EMD Millipore, Billerica, MA). The purity of the dimer solutions were checked by HPLC.

Results and Discussion

Method Development

To develop a HDX setup to monitor subtle changes in deuterium uptake rate and extent, the following characteristics were desired:

1- Efficient digestion: we need many overlapping peptides to provide a full sequence coverage and also a set of highly complementary uptake information at the level of one or near one residue.

2- Well controlled temperature, pH and rapid chromatography: Back-exchange is the main source of data loss during HDX. During the whole RP-LC analysis step, peptides are almost unstructured and fully exposed to the solvent. It means their amide deuteriums are more labile to back-exchange with the bulk water content of the liquid chromatography solvents during this period of time. Therefore carefully controlled temperature and pH in the setup can significantly improve the carried information by peptides. The speed of every experiment is also another critical characteristic of the setup, because time is one of important parameters in the back-exchange phenomenon.

3- Good temporal resolution without the need for blank injections between runs: the uptake curve of a peptide is multi-exponentials because it is the summation of all single-exponential uptakes of residues in that peptide. To better model the deuterium

uptake of peptides, we needed many data points. In other words, the setup cannot tolerate running any blank experiment between sequential data collections of the same protein-deuteration reaction.

4- Very thorough washing of the system: no blank means we should handle the heavy carry-over of proteins at the end of the run to regenerate the system for the next run.

5- Peptide resolution and sensitivity: A fast HDX-technique and overlapping peptide-sequences are conditions that are not easily met simultaneously. In the other words, the cost of fulfilling these conditions is losing resolution and sensitivity for the process of peptide identifications.

Some of the above mentioned issues are in conflict with each other. For example, an efficient digestion is in conflict with the cold condition of the system, or the system washing off the protein carryover can potentially necessitate a lengthier chromatography run, which in turn, increases the back-exchange process of deuterated peptides. We were able to address each of these issues in a satisfactory. Figure 4.1 is a simple schematic diagram of an automated high throughput HDX coupled to an on-line pepsin digestion, multiple LC and ESI-MSMS.

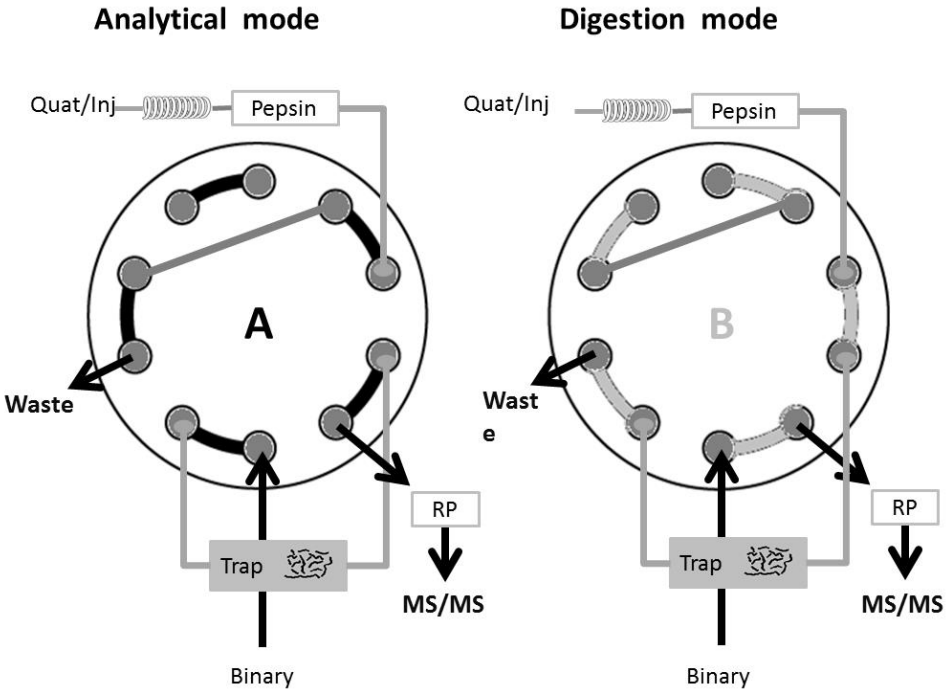

Figure 4.1: Schematic diagram of an automated high throughput HDX-LCs-MSs system coupled to online digestion. The figure illustrates two digestion and analytical modes of the HDX setup. During the digestion mode, a protein sample is injection into the on-line multi immobilized-pepsin columns and peptide products are trapped in a short RP-column using a quaternary pump. During the analytical mode, trapped peptides are separated by a long RP-column using a binary pump and analyzed by an online mass spectrometer in MS or MS/MS analysis.

To meet each of these requirements, we had to take every part of the LCMS system into a consideration. The final design included two LC pump sets, two immobilized pepsin columns in parallel, an RP-trap, an RP-analytical column, all coupled to ESI-qTOF (materials and methods). The whole run, from injection to MS (or MSMS), was 8.5 minutes including thorough pepsin digestion and adequate LC separation of peptides. This time period also includes steps for washing the system off carryover and regenerating the system for the next run. The whole setup was maintained at -1 ˚C and pH=2.5. Such a high throughput setup allows us to collect 5 experimental datasets of time course reactions with >30 time points in less 5 hours.

The final setup that was arrived at was robust, running for over 24 hours at a time without interruption. No blank injections were required and carryover was not observed to accumulate. Figure 4.2 is an overlay of 32 TICs of overnight runs of HDX experiments on HBV capsid protein. The chromatograms were analytically very complicated. Separating thousands of peptides in less than 5 minutes including washing step of the RP-column was challenging, but we also needed to secure the signal quality of these many overlapping peptides despite the cold condition of the system (-1 ˚C). An additional concern was the risk of pepsin activity loss during days-long experiments.

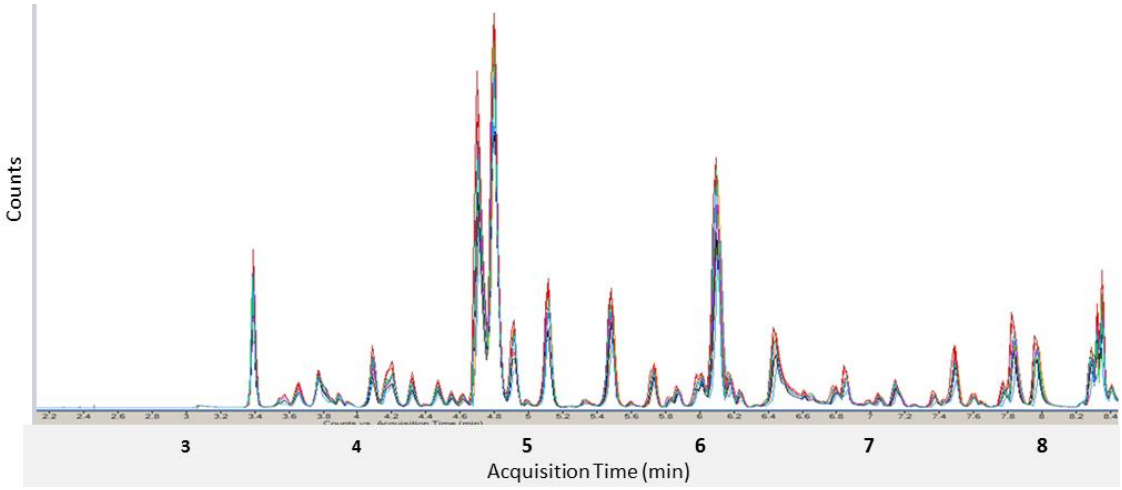

Figure 4.2: The overlay of 32 peptides-TICs of different HDX runs on HBV capsid protein. Despite the very complicated chromatogram including thousands of peptides and partially digested forms of the capsid protein, we had an acceptable reproducibility.

Coverage and Mapping

The preliminary data on pepsin-based digestion showed 100% coverage on the amino acid sequence of the HBV Cp149 protein. Many overlapping peptide sequences could potentially provide near one amino acid resolution. Although more than 1500 peptides were identified and mapped into the Cp149 sequence, we did not included

peptides with very overlapping mass spectrometry signals (overlapped isotope distribution signals) in the analysis. In other words, we only analyzed high confidence peptides which acquired high scores during their deuterium uptake analysis. The analysis score was based on the comparison between the pattern of isotope distribution intensities and ppm mass accuracy of experimentally observed deuterated signals for a peptide with its theoretically calculated signals. More than 800 peptides were kept to carry a single analysis. Figure 4.3 shows the coverage and overlapping peptide sequences coverage.

Figure 4.3: The sequence coverage of peptides mapped into the Cp149 sequence. High confidence-identified peptides selected based on their deuterium uptake analysis score and mapped into HBV capsid Cp149 sequence (PDB: 1QGT).

Peptide Resolution /Sensitivity

The overall peptide resolution/ sensitivity of a HDX-LCMS-based approach includes many components such as the signal to noise ratio of mass signals, the resolution of isotope distributions of peptides co-eluting from reverse phase column, and the dynamic range of peptide signal intensities (spanning up to several orders of magnitude). These sensitivity components determine the overall resolving power of the system in mass measurement of many co-eluting peptides. The high throughput and fast nature of such an approach could potentially jeopardize the peptide resolution/ sensitivity. Therefore the peptide resolution/sensitivity was the most worrying issue especially if we were looking for subtle changes in deuterium uptake rate and extent. We could surely increase peptide resolution/ sensitivity by some changes in the setup such as lowering the flow rates during LC, changing the semi-analytical RP-column to an analytical one and a lengthier chromatography time. A simple way to check the overall peptide resolution/ sensitivity of the approach was to distinguish peptides based on their deuteration behavior and look for any contribution from technique-related artifacts in calculated uptake curves for peptides. Some examples of these artefacts are protein carryovers and overlapped mass signals as a result of the co-elution of many peptides from reverse phase column. Figure 4.4 shows some examples of different peptide uptake behaviors observed in the collected HDX data on HBV capsid Cp149. The high resolution of data and sensitivity of the setup allowed us to more accurately calculate their deuterium uptake curves. The rate and extent of deuteration for every peptide is potentially indicative of the frequency and size of conformational fluctuations which the peptide is a part of.

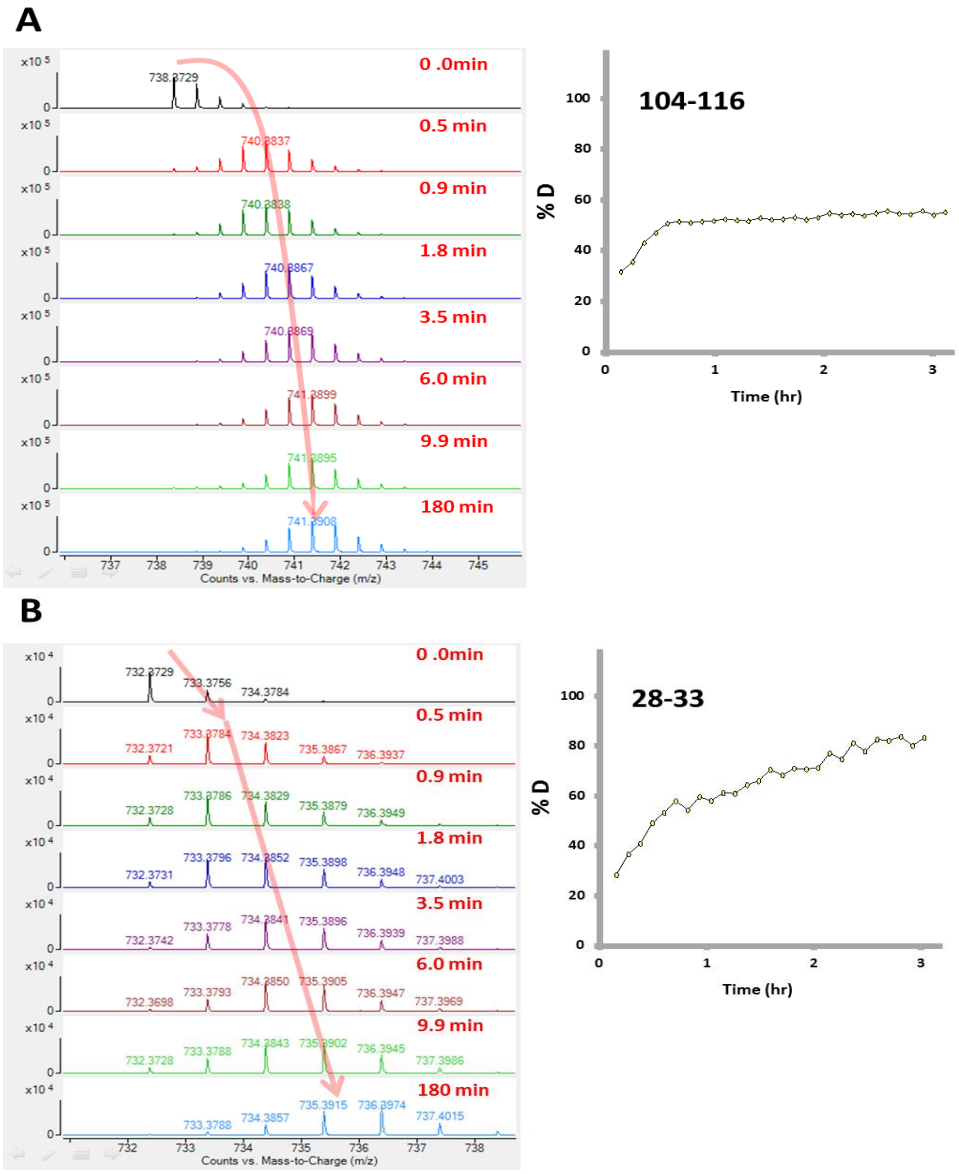

Figure 4.4: Examples of MS signals of isotopic clusters for some peptides and their corresponding deuterium uptake curves. No background signal as a result of the peptide carryover of the LCMS system was observed, however the overlapping or even convolution of MS signals corresponding to co-eluted peptides from RP-column in such a complex chromatogram was inevitable. We did not include these parts of data in the analysis. Panels A, B, C and D are MS signals of four different peptides and their deuteration during time-course HDX experiments. Sub-panels 0.0 min are the isotopic clusters for non-deuterated peptides. Insets are measured deuterium uptake curves corresponding to the peptide in every panel.

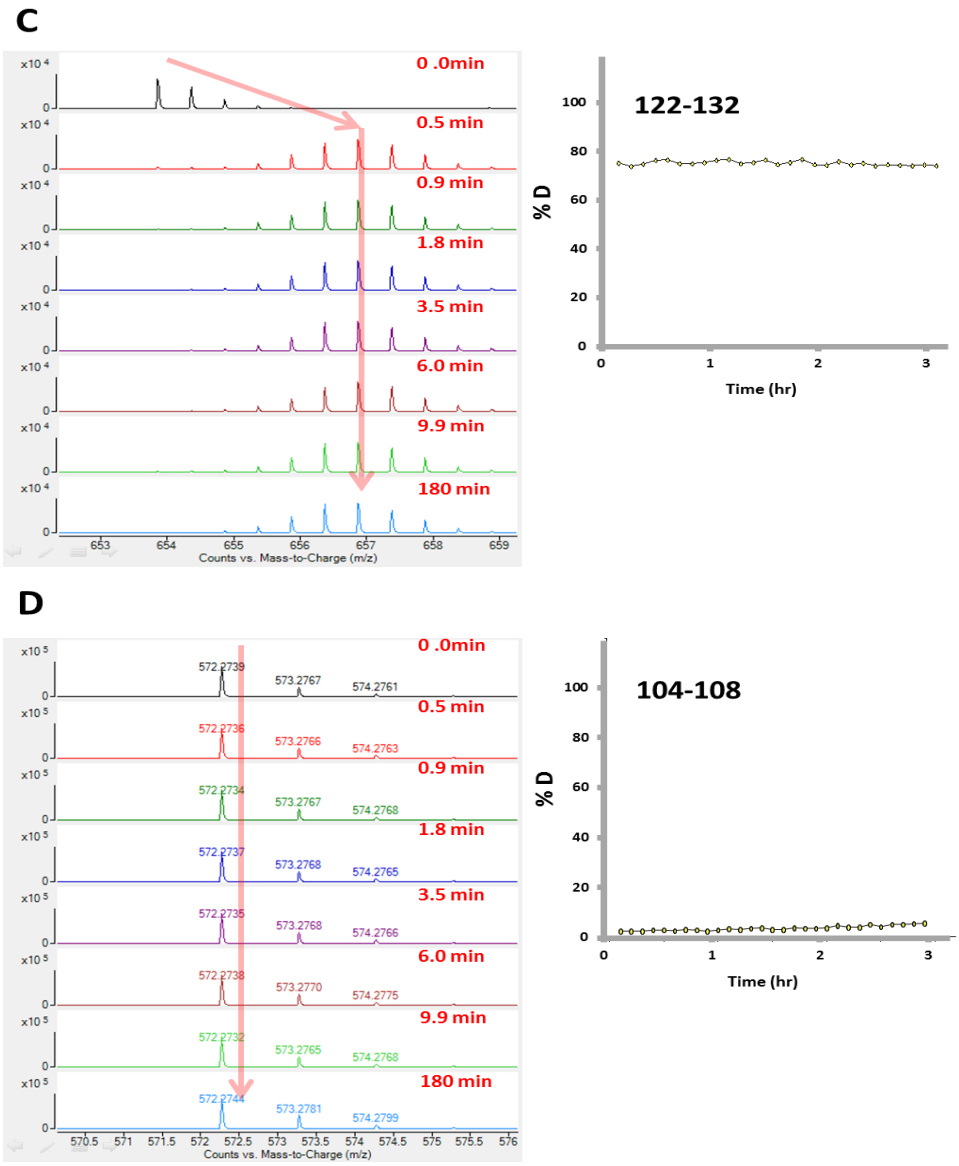

Figure 4.4: (Continued).

However all of these measured uptake curves are supposed to eventually be sources of our data analysis for calculating the rate of exchange for every amino acid.

Peptide Identification

Some of the peptides could be identified and mapped into the sequence without MS/MS. The reason for this is that their mass along provided for unique identification.

For the rest of the peptides, MS/MS was required to establish a unique position. We integrated MS/MS mode into the same HDX setup coupled to on-line pepsin digestion and LC-ESI-MS mode to identify peptides. Figure 4.5 shows the peptide fragments coverage mapped into the sequence of Cp149 as a result of such an LC-MSMS run.

Such an integrated HDX-MS/MS setup could significantly reduce the ambiguity of our coverage map. An integrated MS/MS step means that we do not change either the LC-MS method or the routine protocol of data collection while obtaining MS/MS data; the MS/MS experiment is simply carried by switching the MS mode into MS/MS mode while running non-deuterated experiments. For complementary MS/MS analyses, off-line mass spectrometry was carried using an Agilent 6520 Q-TOF with Agilent 1100 nano-HPLC and a Chip Cube or a Bruker maXis with Dionex 3000 nano-uHPLC. For off-line MSMS analyses, we injected a nondeuterated protein sample into the same LC-MS setup coupled to an online-pepsin digestion, but collected peptides after pepsin columns and before they enter the peptide trap (Figure 4.1).

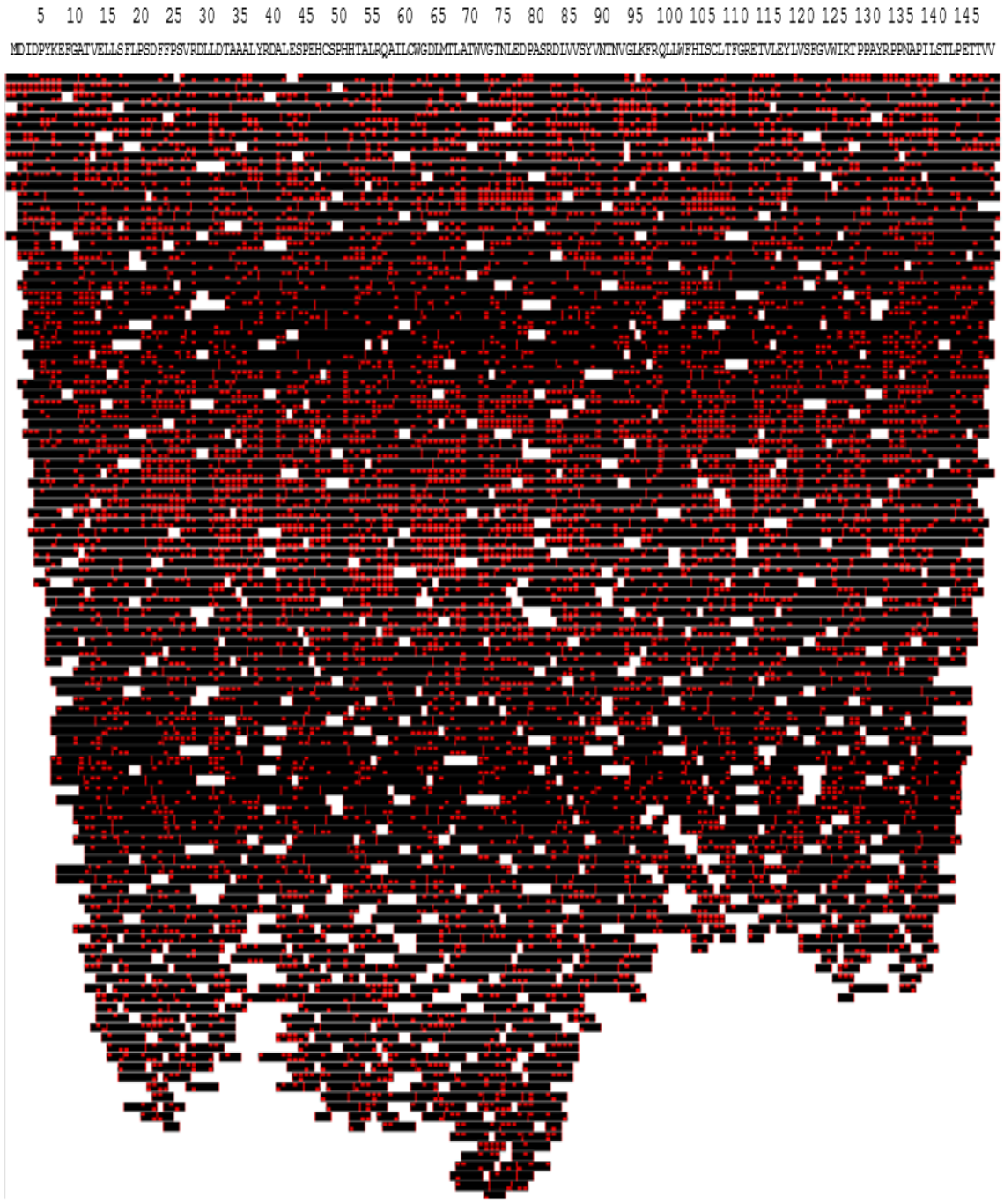

Figure 4.5: The sequence coverage of peptides mapped into the Cp149 sequence by using MS/MS fragmentation. Black horizontal rectangles are the representative of a peptide and red dots are identified amino acids. Peptides with more identified amino acids (proportional to the total number of amino acids in the peptide) acquired higher scores during data analysis. Peptides were mapped HBV capsid Cp149 sequence (PDB: 1QGT).

Calculation of Solvent Accessibility of Every Residue

Once we had the full coverage and the measured deuterium uptake curves for every peptide, we calculated the deuterium uptake of every residue by averaging all sources of information for every single residue from overlapping peptide sequences. Although our collaborator, Dr Brendan Mumey and his team are developing a computational approach to calculate the deuterium uptake for every single residue based on progress curves of peptides (data are not shown here), we used an Excel-based averaging approach. The approach finds all peptides in which a given residue has incurred and dividing their uptake curve into many time zones (as many time zones as data points). Next step was averaging mass gains for every single data point across all these selected peptides which eventually results in a deuterium uptake curve for the residue. Having a deuterium uptake curve for every single residue allowed us to exhibit the deuteration level (or solvent accessibility) across the whole protein with single residue resolution as a function of time.

To display the data, the uptake curves of peptides (and eventually residues) were normalized by converting the deuteration level of every peptide to percent exchange (D %). As it was discussed before, every amino acid residue has one exchangeable amide hydrogen except for prolines and the amino acid at the N-terminus. Therefore the maximum mass gain number for every peptide is:

$$D_{max} = AA - (N\_terminus) - prolines$$

$D_{max} =$ The maximum number of deuterium uptake by the peptide (theoretical)

$AA =$ The total number of residues in the peptide

$(N\_terminus)$ = The number of N-termini in the peptide that is always equal 1

$prolines$ = The total number of proline residues in the peptide

To convert the deuteration level (or the mass gain number) of every peptide to its deuteration % (D %), we used the equation below:

$$\mathrm{D}\% = \frac{100}{D\%_{solution}} \times \frac{D_{obs}}{D_{max}}$$

D_{obs} = The observed number of deuterium uptake by the peptide (experimental)

$D\%_{solution}$ = The percentage of deuterium in the HDX reaction solution

Once we had calculated deuteration % level of every residue, we selected two time zones of 1minute and 70 minute to represent fast exchanging residues and the solvent protection level, respectively. If the amide hydrogen of a residue is not deuterated after 1 hour, we can safely consider it as a protected residue against the bulk water because of either intra-protein hydrogen bindings or being buried in the hydrophobic core. On the other hand, if an amide-hydrogen was observed to exchange in less than 1 minute, it can be considered as an unprotected residue. Then, the deuteration level of each residue was color coded and mapped onto the protein structure as a function of time. We used a gradient color code from blue=0% to red=100% deuteration. The overall color code of neighbor residues also allows us to map the distribution /stability of H-bonding of amide hydrogens for every part of the protein. Although relating the deuteration behavior of a peptide directly to its single residues is not straightforward, however we have many overlapping peptide sequences which can potentially improve the data analysis approach.

False Peptide Identifications

To reveal the subtle conformational changes involved in the capsid assembly, we needed to generate data at a higher level of resolution than previously reports using HDX-MS. Such a resolution could bring some technical difficulties into analysis. The elution of several thousand peptides out of RP-column into the mass spectrometer during a time window of only 5 minutes makes highly crowded mass spectra which results in the overlapping of peptide isotopic clusters. Such overlapping mass signals can potentially interrupt the precisely calculation of the centroid of the peptide isotopic cluster. We use these centroids to make the deuterium uptake curves for every peptide. We tried to avoid this issue by taking the peptide sources of these overlapping mass signals out of the analysis as much as we could. We also divided peptides based on their assignment into two categories, those mapped with MS and those mapped with MS/MS. Then we prioritized peptides for the HDX analysis based on the intensity of their mass signals (in the MS-category list) and their MS/MS-identification score (in the MS/MS-category list).

As mentioned above, a complicated LC-MS setup including two high flow-rate LC pumps, urea solution as solvent and on-line immobilized pepsin columns has the potential to accumulate many compounds and peptides in the RP-columns. These trapped materials eventually appear in mass spectra. Unwanted mass signals can interrupt data analysis in two ways. First they may overlap the isotopic cluster of a peptide of interest and interrupt the centroid calculation (as mentioned above). Second they may be misidentified as a peptide of interest. In the latter case, they may exhibit almost no-deuteration as we expect from a noise signal. However these false-uptake curves with

near 0% deuteration can potentially bias final rates of exchange for single residues. To investigate the impact of error from outliers on our average-based calculations, we calculated the exchange rate of single residues with a median-based approach. Similar to average-based approach, we used the Excel to calculate medians. The approach is finding all peptides in which a given residue has incurred and dividing their uptake curve into many time zones (as many time zones as data points). Next step was finding the median of all mass gains for every single data point across all these selected peptides which eventually results in a deuterium uptake curve for the residue. Such a deuterium uptake curve for every single residue allows us to exhibit the deuteration level (or solvent accessibility) of the whole protein as a function of time. Median-based approaches as non-parametric analyses reduce the biasing impact of outliers on analyses. Figure 4.6 shows the comparison of both average- and median-based approaches in rate calculations.

Both approaches showed a similar degree of deuteration for every residue after 70 minutes. The high number of recurrence of residues as a result of many overlapping peptide sequences may be the main reason for a close calculation of exchange rates by both parametric and non-parametric approaches. No direct relation between the recurrence number of residues and the consistency of calculated exchange rates by both approaches was observed. Therefore we can rule out the presence of any bias calculation as a result of unequal recurrence number distribution among residues (Figure 4.6, the gray diagram) and consider them equally for the final color map preparation onto the protein.

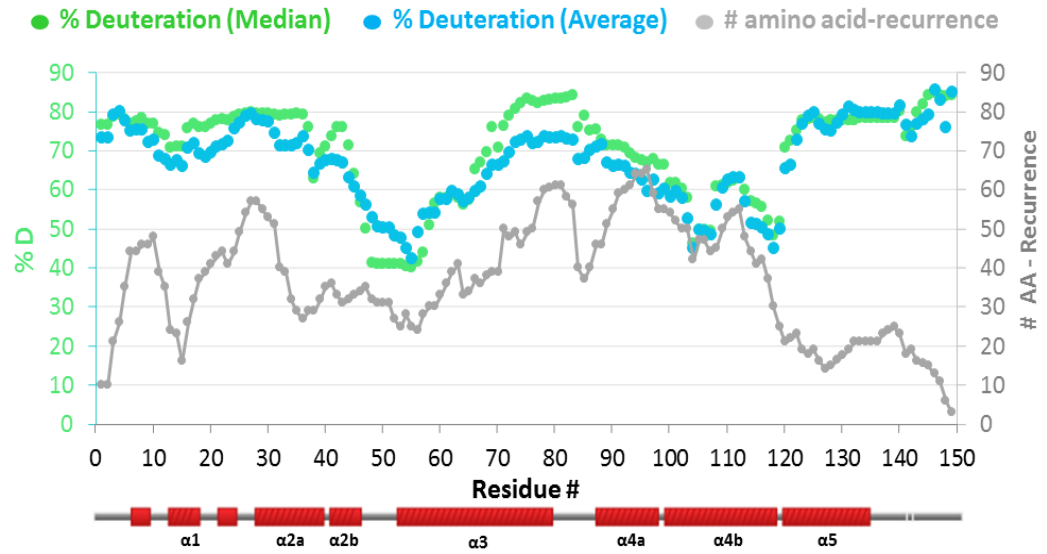

Figure 4.6: The recurrence of single residues and the comparison of the average- and median-based approaches for calculation of exchange rates of residues Every residue occurs in multiple peptides, each providing an independent source for calculation of exchange rates of residues. The gray diagram shows the distribution of residues based on the number of their recurrence in overlapping peptide sequences. Despite of some differences, both median- (green) and average- (blue) based calculations showed similar patterns. We used both approaches to calculate the exchange rate of every single residue. For a better comparison, we overlaid the Cp149 sequence and its α-helix components under the figure. The percentages of deuteration for every single residue in this diagram are their deuteration after 70 minutes.

Mapping the Deuteration Level of Residues onto the Protein Sequence

As we mentioned before, we converted the deuteration level of every residue into D%. We used a color gradient from blue=0% to red=100% deuteration and color mapped the crystallography-based structure of HBV capsid protein Cp149 (PDB: 1QGT) based on D% data. For the purpose of simplification, we compared color maps of Cp149 at three time zones of 0, 1 and 70 minutes to represent non-deuterated, fast exchanging, and solvent protected residues, respectively. Figure 4.7 shows the deuteration uptake level of residues at 1 and 70 minutes in comparison with their deuteration level at 0 minute. The 0

minute time zone corresponds to the non-deuterated protein before starting the deuteration reaction.

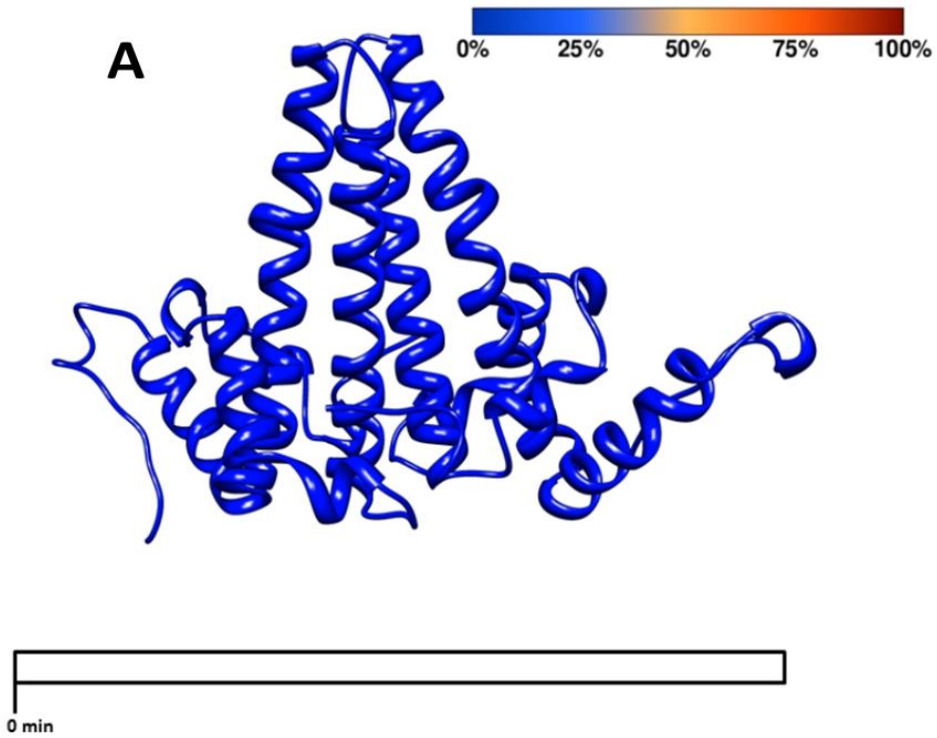

Figure 4.7: Deuteration level of HBV Cp149 dimer as a function of time. The gradient color between blue and red are for 0% and 100% deuteration level, respectively. Every panel is indicative the whole deuteration level of the protein for a given time zone with the resolution of one residue. Panels A, B and C are exhibiting the distribution of deuteriums among residues at time zones of 0, 1 and 70 minutes, respectively. Time zone 0 minute corresponds to the non-deuterated protein before starting the HDX reaction and therefore the protein expected to be fully colored as dark blue; PDB: 1QGT.

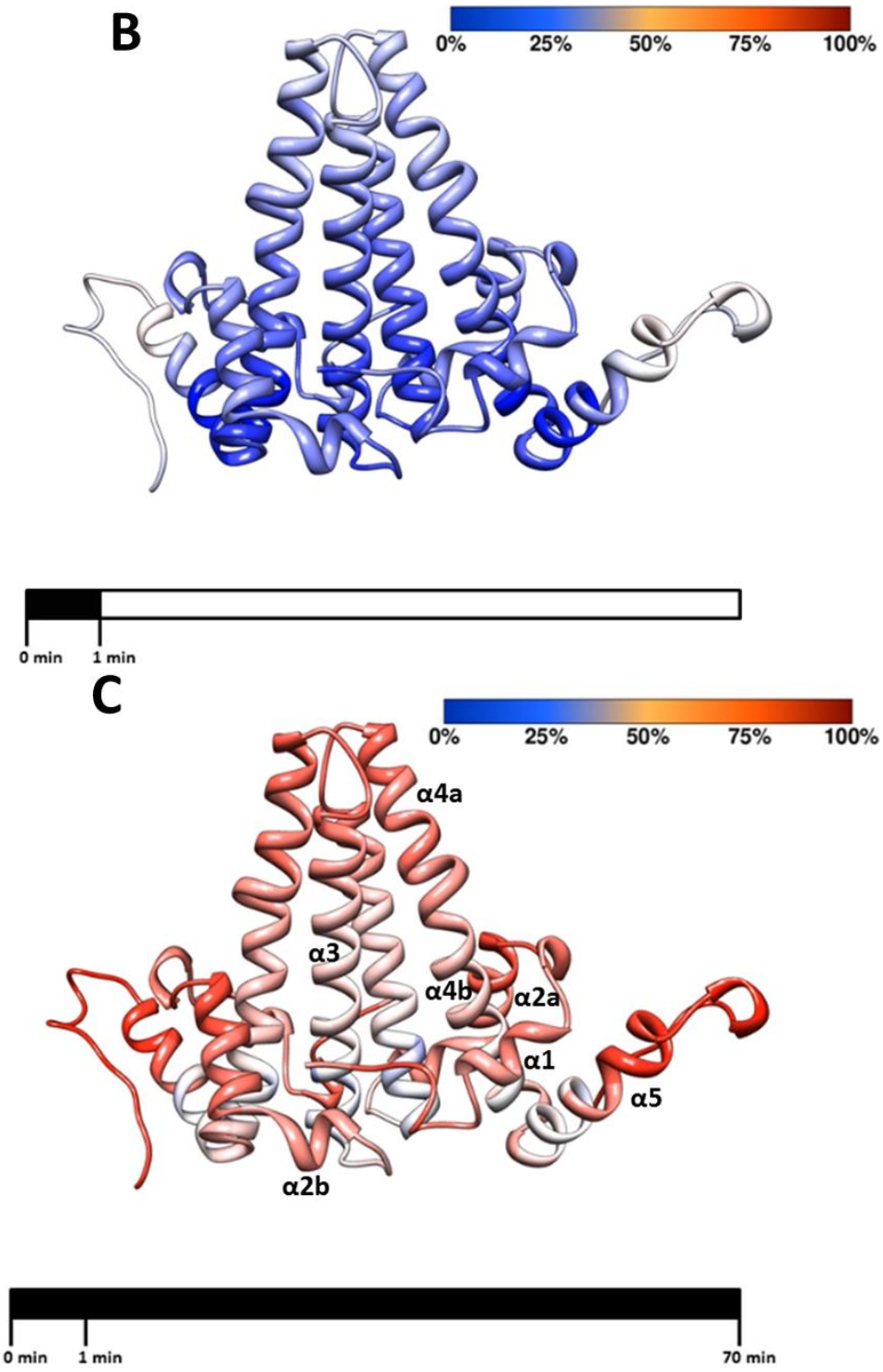

Figure 4.7: (Continued)

The Solvent Accessibility Map of the HBV Capsid Protein in Solution

Once we have the deuteration level of the whole protein as a function of time with the resolution of one residue, we can analogically consider the deuterium level as the solvent accessibility for the purpose of simplification.

Although X-ray crystallography and cryo-electron microscopy have provided the details of the structure and inter/intra-dimer contacts, however, the high-resolution solvent accessibility picture of HBV capsid protein in solution remained to be presented. Here we have reported the first high-resolution solvent accessibility map of this protein in solution assessed by the LC-MS-based HDX technique.

The color map in Figure 4.7 panel C shows the high level of solvent accessibility for the HBV Cp149 despite the crystallography and cryo-EM data that show a high level of α-helicity for this protein. Another LC-MS-HDX report on HBV Cp149 has pointed to the tip of the four-helix bundle, the junction between α-4a and α-4b, and the C-terminal loop as regions with highest deuterium uptake (33-61%). They have also reported a medium uptake (19-41%) for regions α-2a and α-4b and a low uptake (8-21%) for regions α-1, α-2b (and its junction to α-3), and the junction of α-4b and α-5 [69]. However their report was based on only a very limited number of peptides and incapable to picture a full sequence coverage solvent accessibility on this protein. Despite some degree of similarity between our solvent accessibility map and their reported uptake behaviors on a limited number of peptides, their deuterium uptake distribution picture is not fully consistent with our work. For example we comparatively identified high solvent accessibility for α-2a, α-1, α-2b. In addition, the overall sensitivity of our LCMS-HDX approach was higher

if we consider the sensitivity proportional to the maximum peptide deuteration uptake observed for a given protein. We were able to identify many peptides with more than 90% deuteration based on the standard model in which the N-terminal residue is not an exchange site.

Based on the structural model for this protein from X-ray crystallography data, we can not readily explain the high levels of deuterium uptake by the majority of protein after only one hour. To help with our discussion of the data, we use the words of 'less stable α-helix' and 'less buried' for the purpose of simplification. The lowest solvent accessibilities were observed in the base of four-helix bundle, where we consider it as the chasses of the Cp149 protein or the shell in the assembled capsid [24]. These areas may act as an anchor for the whole HBV protein dynamics around which the rest of protein fluctuates [24].

In addition to the fundamental role of HBV core protein in forming an icosahedral viral capsid, it also fulfills a wide range of functions during virus replication. A list of diverse functionalities include interactions with the viral polymerase, pre-genomic RNA, and double stranded DNA, as well as differential phosphorylation to modulate interaction with the genome which eventually results in a differential interaction with surface antigens during virion maturation. All of which point to the HBV core protein as a functionally promiscuous protein [35, 63, 64, 70].

After years of analysis, the HBV capsid protein has proven to be a conformationally plastic protein. This is likely a result of the diverse functional requirements [63, 64]. Even subtle changes in the conditions used to seed crystals can

lead to the capture of different structures [63]. The dynamic state of the dimer is the critical factor in capsid assembly. Cysteine 61 is located in the middle of α-3 and connects two α-3s across the four helix bundle by forming an inter-protein disulfide bond. Formation of this disulfide bond in free dimer prevents capsid assembly, even though it is far removed from the protein-protein contacts regions [71]. The high level of dynamics in the free HBV core protein dimer is associated with structural strains arisen from frustrated contacts and interactions [63, 72]. Adoption of a strained conformation by the free dimer would assure a low energy barrier for adopting function-relevant conformations. One example is the assembly-competent conformation which nucleates dimer interactions and assembly [24, 63].

Figure 4.7 panel C shows three subdomains including the four-helix bundle tip (α-3 to α-4a), a helix fulcrum (regions of α-1 to α-2a), and the inter-dimer region (α-5 to C-terminus) that are hinged to chasses area of the protein. The flexibility of these three subdomains with respect to the stable chassis, allow the protein to adopt a number of conformations [63]. Such conformational diversity is possibly the reason for high deuterium uptake by these regions. One of the regions with the highest solvent accessibilities in HBV Cp149 is located in the inter-dimer contact area (α-5 to C-terminus). Releasing structural strains, for example in highly dynamic region of α-5 to C-terminus, should be the compromise for dimer-dimer interaction upon assembly which explains the necessity of a highly dynamic state for this protein to biologically function during virus maturation and replication [63, 64]. In addition, a flexible interacting region such as the region α-5 to C-terminus can kinetically improve the protein-protein

interaction by increasing the chance and rate of successful docking.. Such an improved binding may be important in an inter-molecular interaction process with a fast kinetics like capsid assembly [67]. The phenomenon of structural strain-release in HBV core protein can be phenomenologically analogous to the "on-site folding" event observed with intrinsically disordered proteins (or proteins with locally disordered regions). On-site folding or coupling of a local structure-formation with the protein interaction event has proven to improve the kinetics and thermodynamics of protein assembly [6, 8, 65].

Conclusion

The current automated high throughput LC-MS/MS-based HDX setup coupled to on-line pepsin digestion proved to be a powerful approach to probe the deuteration level of the whole protein, providing data with essentially single residue resolution. The high resolution that we were able to obtain was compatible with fast chromatography and a semi high throughput method. The combination of speed, reproducibility, and thorough proteolysis facilitated the detailed analysis of the molecular switch-based mechanism of HBV capsid assembly.

The first step was to map solvent accessibility of HBV Cp149 with single residue resolution. These data clearly showed the high level of deuterium uptake by the majority of Cp149 protein regions. These data are consistent with numerous studies of HBV core protein implicating that conformational dynamics (structural diversity) is necessary for the multifunctional role of the core protein in the life cycle of HBV [35, 63, 64, 69-71, 73], Such a structural diversity acts as a thermodynamic balance by bringing strains

(frustrated contacts) into the structure. In turn, the protein releases these strains as a compromise during one of its biological functions such as capsid assembly [63, 64].

The next step to determine the mechanism and location of the hypothesized molecular switch which plays an allosteric role during the capsid assembly is to map the whole solvent accessibility of the HBV Cp149 structural homologue, Y132A, with the resolution of one residue. Y132A has proven to be an assembly deficient mutant [24, 38, 63].

HIGH RESOLUTION HYDROGEN-DEUTERIUM EXCHANGE COMPARATIVE ANALYSIS OF HBV CP149 AND CP149-Y132A

Introduction

Tyr-132 is an exposed amino acid in free dimer and buried in capsid (Figure 5.1 panel A). The mutant Y132A is a dimer with abortive functionality for nucleation, assembly or formation of assembly intermediates. Co-assembly of Y132A with wildtype dimer yields extremely unstable capsids [38]. Circular dichroism and protein fluorescence studies did not show any structural differences between Y132A and wild-type protein Cp149 [24]. Triangular trimers of the core protein homodimer are considered to be the fundamental quaternary structure of icosahedral HBV capsids, and point to a nucleating block for capsid assembly (Figure 5.1 panel B) [68]. In the context of the assembled state, wtCp149 forms a symmetric trimer of dimmers in a triangular shape with three inter-dimer contacts. Y132A does not adopt the same structure; rather the trimer of dimmers remains open with only two inter-dimer contacts (Figure 5.1 panel C) [24]. Inter-protein contacts between the mutant dimers are possibly not sufficiently strong to provide the energy of the conformational changes that are necessary for the closure of the triangle. If we neglect the crystal packing contribution for both forms of triangles:

$$\Delta G^{o}_{\text{triangle}} = 3\Delta G^{o}_{\text{contact}} + 3\Delta G^{o}_{\text{conformational changes necessary for the closure}}$$

$$\Delta G^{o}_{\text{asymmetric unit}} = 2\Delta G^{o}_{\text{contact}}$$

The $\Delta G^{o}_{\text{asymmetric unit}}$ is less than $\Delta G^{o}_{\text{triangle}}$, because it is thermodynamically more favored. Therefore $\Delta G^{o}_{\text{conformational changes necessary for the closure}}$ is positive. [38].

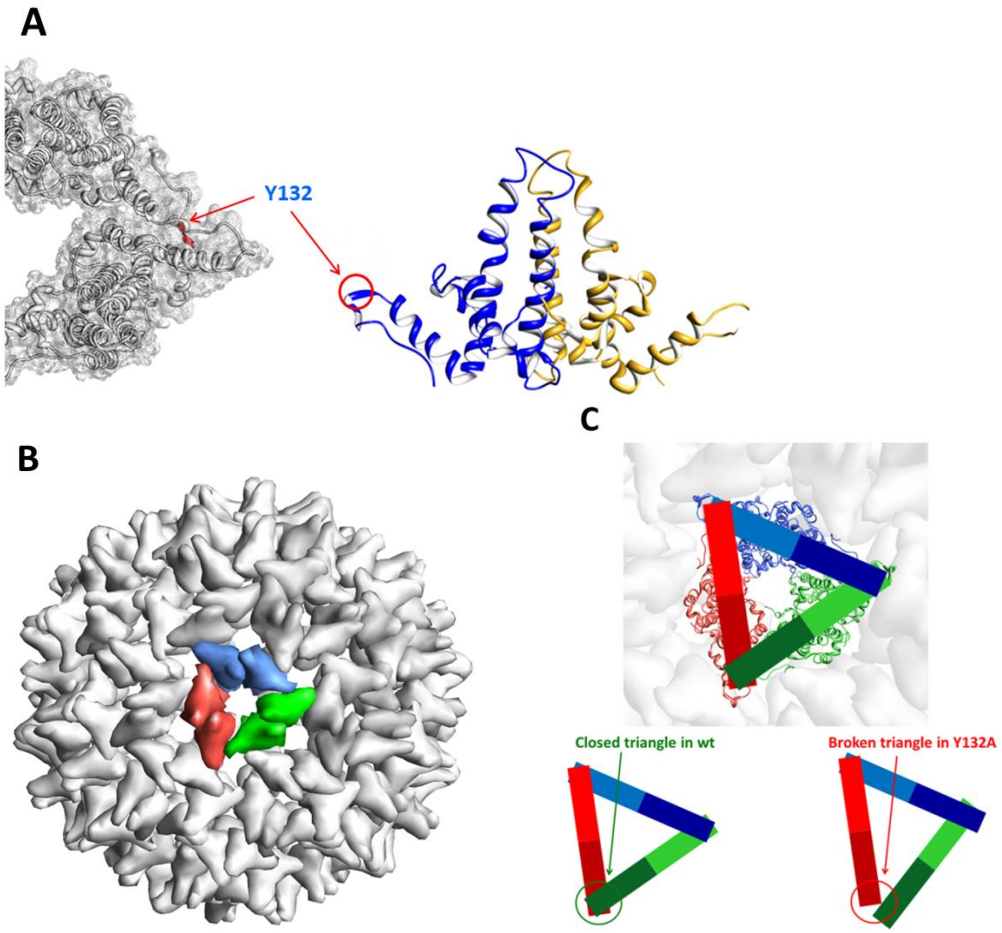

Figure 5.1: Tyr-132 and the threefold asymmetric trimer of mutant Y132A. Tyrosine 132, a key amino acid in capsid assembly, is exposed amino acid free dimer and buried in capsid (A). The crystallographic unit of this protein is a triangular trimer of homodimers and the fundamental quaternary structure in HBV capsid, and also the nucleating block during assembly (B). In wildtype Cp149, the trimer is a stable closed one-turn right-handed helix with three inter-dimer contacts. In mutant Y132A, an assembly deficient dimer, the trimer is a broken triangle with only two inter-dimer contacts (C).

Materials and Methods

Methods

Automated Ultrafast Hydrogen Deuterium Exchange: a HDX setup including continuous deuterium labeling [55] followed by on-line pepsin digestion (immobilized-pepsin column) and automated fast LC-MS was developed. All labeling experiments

were done in a temperature controlled micro-well plate LC-sampler by (automated/or manually) initiation of reactions by 10 fold dilution of protein samples into a D_2O-tris + NaCl (H_2O-tris for the nondeuterated reactions). The concentration of NaCl, temperature, and the pD varied depended on the experimental design. Injection was coupled to on-line parallel immobilized-pepsin columns. Digestion occurred in 3% acetonitrile, 0.02% formic acid and a varying concentration of urea (0.8-1.2 M) in H_2O based LC solvent. Released peptides were trapped and desalted on-line using C8-macroTrap 3×8mm (Michrom Bioresources, Inc. Auburn, CA). Peptides separation was done through a C18 column (1.7 µm, 1 mm × 100 mm) (Phenomenix) held at -1 ° C, A 5 min multi-linear multi-flow rate gradient of water:acetonitrile with 0.1% formic acid as LC solvents. LCs were done using an Agilent 1200 LC and 1290 uHPLC systems (Agilent Technologies, Waldbronn, Germany) with an automated degasser, thermostatted micro well-plate samplers, and a column heater compartment. Peptide masses were measured using an Agilent 6538 Accurate-Mass Q-TOF mass spectrometer with electrospray ionization. The mass spectrometer was operated in positive mode scanning over the range 50-1700 m/z. Peptides were identified and mapped onto the Cp149 sequence using Agilent MassHunter Workstation Software, Qualitative Analysis Version B.04.00 and BioConfirm software. MS/MS data were analyzed and mapped onto the protein sequence using the Bruker Daltonics Bio Tools 3.2. The identification tolerances were 0.002m/z (±) 5ppm and 0.002 milli mass unit for MS and MS/MS, respectively. We used the Sierra Analytics HDExaminer Version 1.3.0 beta6 software to calculate centroids of isotopic clusters of a peptide during time-course HDX experiments. Time points from 5 replicates were

randomly combined to improve the curve fitting process into multiple-exponentials. For complementary MS/MS analyses, off-line mass spectrometry was carried using an Agilent 6520 Q-TOF with Agilent 1100 nano-HPLC and a Chip Cube or a Bruker maXis with Dionex 3000 nano-uHPLC. Off-line MSMS analyses were conducted by injecting nondeuterated protein samples into the LC-MS setup described above, except that peptides were collected before they enter the peptide trap column (Figure 4.1). The details of the LCMS acquisition method can be found in appendix A.

Sample Preparation: Y132A dimers were obtained from Dr Zlotnick lab. Dimers were filtered and concentrated using Amicon pressure-based stirred cells, 10 kD ultrafiltration disk or Centrifuge-based ultrafilter 0.5 mL, 10 kD (Amicon EMD Millipore, Billerica, MA). The purity of the dimer solutions were checked by HPLC.

Differential Scanning Fluorometry (DSF): The thermostability assay was done by mixing proteins, PBS buffer pH=7.5 and Sypro-Orange dye (Invitrogen Inc.) in a quantitative PCR (qPCR) instrument (RG-3000; Corbett Research) [59].

Size Exclusion Chromatography (SEC): An 1100 series liquid chromatography system from Agilent Technologies (Wilmington, DE) was used for all chromatographic measurements. Protein retention times were determined using an Agilent diode array detector by absorbance at 230 and 278 nm. We used two Agilent HPLC columns BioSEC-3 (30cm-3µm-100Å-0.46cm id) and BioSEC-5 (30cm-5µm-1000Å-0.46cm id).

Results and Discussion

Coverage and Mapping

The preliminary data on pepsin-based digestion of mutant Y132A showed 100% coverage of the amino acid sequence and many overlapping peptide sequences. Although more than 1500 peptides were identified and mapped into the Y132A sequence, we did not included peptides which had overlapping isotope patterns in the analysis. In other words, we only analyzed high confidence peptides which acquired high scores during their deuterium uptake analysis. The analysis score in HD Examiner was based on the comparison between the pattern of isotope intensities and ppm mass accuracy of experimentally observed deuterated signals for a peptide compared with theoretically calculated signals. More than 800 peptides met these criteria. Figure 5.2 shows the coverage and overlapping peptide sequences coverage.

Calculation of Solvent Accessibility of Every Residue

Once we had selected the final list of peptides that provided full coverage and had high quality deuterium uptake curves, the deuterium uptake at each individual residue was estimated by averaging all uptake curves that contained a given residue. This was repeated for every residue using an Excel-based averaging approach. Having a deuterium uptake curve for every single residue allowed us to exhibit the deuteration level (or solvent accessibility) across the whole protein with single residue resolution as a function of time.

Figure 5.2: The sequence coverage of peptides mapped into the Y132A sequence. High confidence-identified peptides selected based on their deuterium uptake analysis score and mapped into HBV Y132A sequence (PDB: 3KXS).

To display the data, the uptake curves of peptides (and eventually residues) were normalized by converting the deuteration level of every peptide to percent exchange (D %). As it was discussed before, every amino acid residue has one exchangeable amide

hydrogen except for prolines and the amino acid at the N-terminus. Therefore the maximum mass gain number for every peptide is:

$$D_{max} = AA - (N\_terminus) - prolines$$

D_{max} = The maximum number of deuterium uptake by the peptide (theoretical)

AA = The total number of residues in the peptide

$(N\_terminus)$ = The number of N-termini in the peptide that is always equal 1

$prolines$ = The total number of proline residues in the peptide

To convert the deuteration level (or the mass gain number) of every peptide to its deuteration % (D %), we used the equation below:

$$\text{D\%} = \frac{100}{D\%_{solution}} \times \frac{D_{obs}}{D_{max}}$$

D_{obs} = The observed number of deuterium uptake by the peptide (experimental)

$D\%_{solution}$ = The percentage of deuterium in the HDX reaction solution

Once we had calculated deuteration % level of every residue, we selected two time zones of 1minute and 70 minute to represent fast exchanging residues and the solvent protection level , respectively. If the amide hydrogen of a residue is not deuterated after 1 hour, we can safely consider it as a protected residue against the bulk water because of either intra-protein hydrogen bindings or being buried in the hydrophobic core. On the other hand, if an amide-hydrogen was observed to exchange in less than 1 minute, it can be considered as an unprotected residue. Then, the deuteration level of each residue was color coded and mapped onto the protein structure as a function of time. The overall color code of neighbor residues also allows us to map the distribution /stability of H-bonding of amide hydrogens for every part of the protein. Although

relating the deuteration behavior of a peptide directly to its single residues is not straightforward, however we have many overlapping peptide sequences which can potentially improve the data analysis approach.

Comparative Mapping the Deuteration level of Residues onto the Protein Sequence of Cp149 versus Y132A

To compare the overall solvent protection between Cp149 and mutant Y132A, we selected the time zone of 70 minute to present the solvent protection level. As we discussed before, the deuteration level of each residue after 70 minutes was converted into color codes and mapped onto the protein structure as a function of time. Data analysis showed that the Y132A is as dynamic as wtCp149. Therefore we normalized the usual color code map of blue = 0% and red = 100% deuteration to blue = 40% and red = 90% deuteration corresponding to the minimum and maximum observed deuterations in both deuteration lists of Cp149 and Y132A. Using this normalized gradient color code and lists of observed deuteration of residues for Cp149 and Y132A, we color mapped the crystallography-based structure of HBV capsid protein Cp149 (PDB: 1QGT). Regarding the structural similarity of Cp0149 and Y132A, we used PDB: 1QGT for both protein forms to compare their deuteration levels. Figure 5.3 shows the deuteration uptake level of residues after 70 minutes for Cp149 and Y132A. A close look at both colored structures shows some subtle differences between their solvent protection levels. Largest changes were observed in the four-helix bundle and some area around the core.

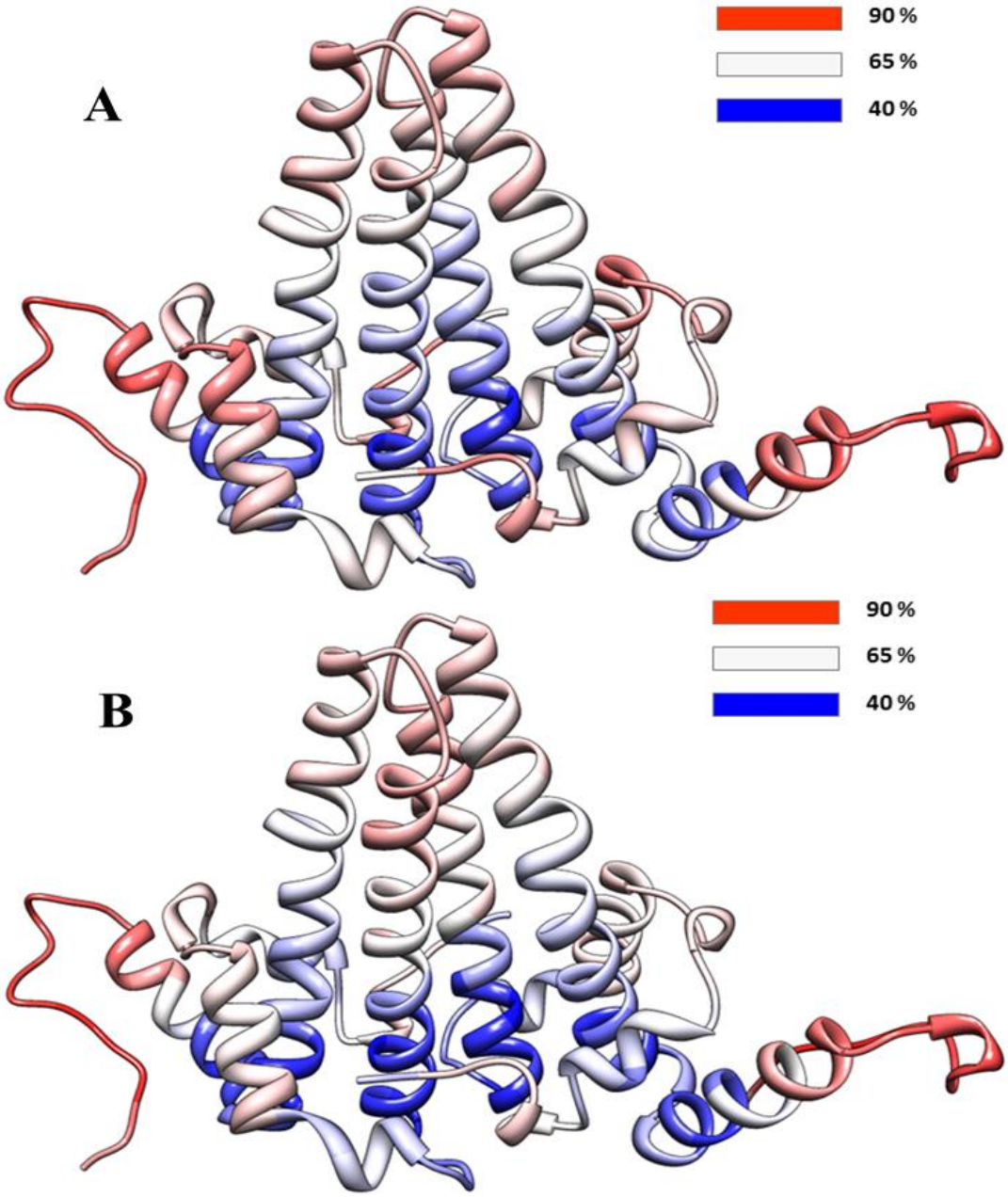

Figure 5.3: Deuteration levels of Cp149 and Y132A. The gradient color between blue and red are for 40% and 90% deuteration level, respectively. The deuteration level with the resolution of one residue after 70 minutes for Cp149 (panel A) and Y132A (panel B); PDB: 1QGT.

Although Figure 5.3 shows some subtle conformational differences between two protein forms of Cp149 and mutant Y132A, the subtraction of their color maps could clearly display and map all differences.

The Differential Solvent Accessibility Map of Cp149 versus Y132A

For a better visualization of differences between wtCp149 and Y132A, we subtracted the list of deuteration levels of residues of Y132A from the similar list of Cp149 at 1minute time zone. Then we normalized the list of deuterium % of residues in the subtraction list into a color gradient of blue = -5% and red = +12% deuteration corresponding to the minimum and maximum differential deuterations in the list. Then we used this color code list to color map the crystallography-based structure of HBV capsid protein Cp149 (PDB: 1QGT). Figure 5.4 shows the differential deuteration level of residues after 1 minute between Cp149 and Y132A.

Once we have the differential deuteration level with the resolution of one residue, we can analogically consider the deuteration level as the solvent accessibility or H-bond stability. The differential deuteration level between Cp149 and Y132A in Figure 5.4 shows that every region of the protein is involved in the differential solvent accessibility between wildtype and mutant forms in less than 1 minute. As we hypothesized in chapter 3, the mutation Y132A introduces a degree of stability (compactness) around the hydrophobic core (Figure 1.4) which involves α-helices 1, 2a, 4 and 5. The lower part of α3 is also contributing to the core compactness. Figure 5.4 clearly shows reduced

exchange around the core. The time scale of protein motions involved in such differential solvent accessibility is less than 1 minute.

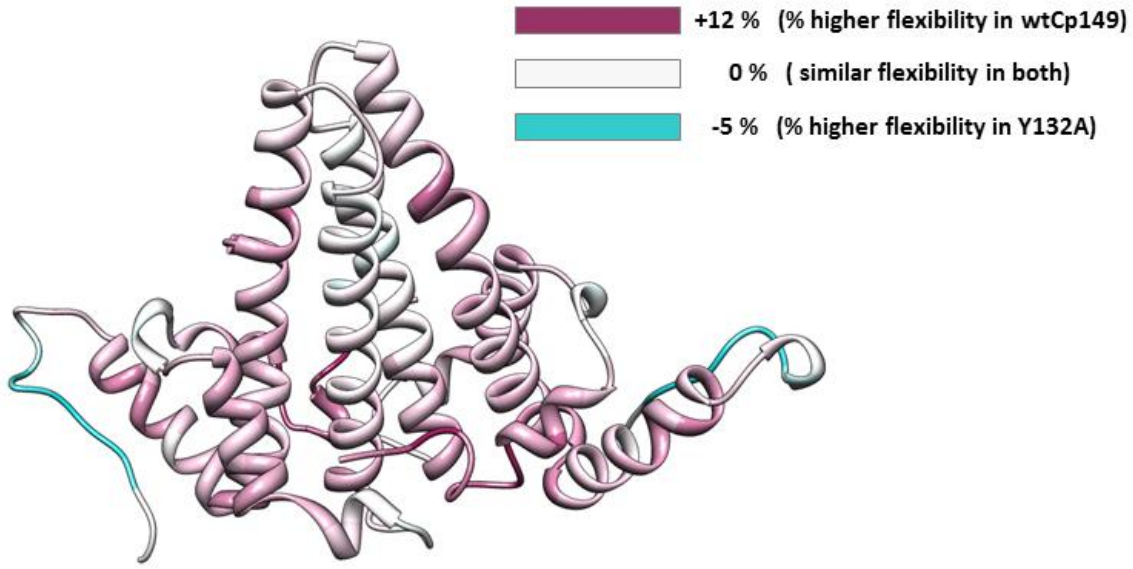

Figure 5.4: The differential deuteration levels of Cp149 versus Y132A after 1 minute. The gradient color between blue and red are for -5% and +12% differential deuteration level, respectively. The deuteration level Y132A after 1 minute is subtracted from Cp149 with the resolution of one residue; PDB: 1QGT.

To expand the picture of the solvent accessibility or H-bond stability of amide-hydrogens to longer time scales, we subtracted the list of deuteration levels of residues of Y132A from the similar list of Cp149 after 70 minutes. The differential percent deuteration was converted into a color gradient of blue = -10% to red = +10% deuteration corresponding to the minimum and maximum differential deuterations in the list. This color code was then mapped onto the crystallography-based structure of HBV capsid protein Cp149 (PDB: 1QGT). Figure 5.5 shows the differential deuteration level of residues after 70 minutes between Cp149 and Y132A. We selected the 70 minute time

zone as the representative of slow exchange. The differential deuteration level between Cp149 and Y132A in Figure 5.5 clearly shows that almost every region of the protein, especially around the core is involved in the differential solvent accessibility after 70 minutes. The exposure of Y132 in the native Cp149 protein is apparently placing some structural strains on the protein structure which results in some degree of instability in long-distance regions. Such instability is possibly the reason for the significantly higher deuterium uptake by Cp149 dimer compared to Y132A (Figure 3.10).

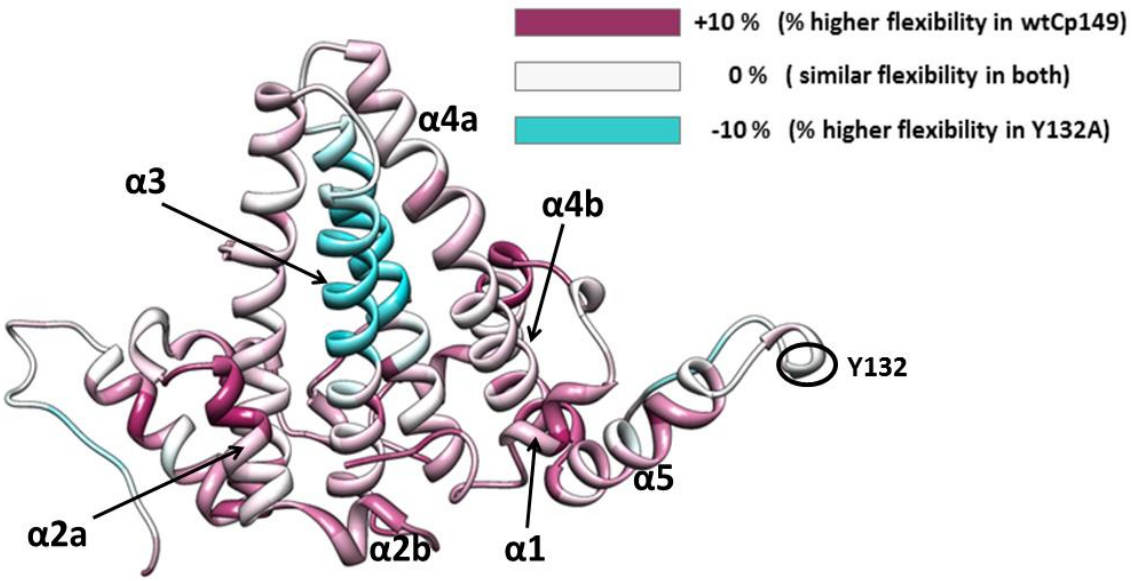

Figure 5.5: The differential deuteration levels of Cp149 versus Y132A after 70 minutes. The gradient color between blue and red are for -10% and +10% differential deuteration level, respectively. The deuteration level Y132A after 70 minutes is subtracted from Cp149 with the resolution of one residue; PDB: 1QGT.

The α3-helix has a major contribution in four-helix bundle stability. The high resolution HDX-based map of solvent accessibility of Cp149 dimer from chapter 4 (Figure 4.7) presented the lowest solvent accessibility for α3-helix. With no solvent exposed tyrosine at position 132, in the mutant Y132A, the α3-helix become unstable

(Figures 5.4 and 5.5). Although the α3-helix is far removed from position 132, the the mutation destabilizes the α3-helix. We suggest that there is a path of signal transduction across the protein core which structurally links Y132 to α3-helix and four-helix bundle area. Data presented in Figures 5.4 and 5.5 are consistent with the presence of such a conformational path.

Differential Chromatographic Behavior of Y132A versus Cp149

Our comparative HDX analysis showed more solvent accessibility (instability) for α3-helix and less solvent accessibility (stability) for regions surrounding the protein core. Such instability in α3-helix area results in instability in four-helix bundle and eventually some degree of disruption for inter-monomer contact area. The disruption of inter-monomer contact may results in a more extended structure of Y132A compared to the wild type. To test if global changes in structure were induced by the mutation, size exclusion chromatography (SEC) was used. A high resolution HPLC setup as tandem SEC columns where used to investigate any subtle change in the hydrodynamic radius of the mutant compared to the wildtype. Figure 5.6 shows the results of chromatography on HBV capsid, wildtype dimer and mutant Y312A dimer and the effect of urea (~4M) on the hydrodynamic radius of dimer. Chromatograms in Figure 5.6 show that the peak of Y132A dimer elutes at 6.2 ml, compared to the elution time of wildtype dimer at 6.6 ml. Formally, SEC is a separation technique based on hydrodynamic radius [74, 75] if we exclude the possibility for specific protein-column interactions which may alter the elution time of them protein. Crystallography data did not show any significant structural

difference between these two protein forms. The structural similarity and the fact that the presence/lack of tyrosine at position 132 in dimer-dimer contact area does not have any significant effect on the higher order of structure for this protein (they both form trimer of dimers under crystallography condition), leave us with two possibilities: either the wildtype is interacting with column which clearly points to the its conformational differences with the mutant, or we can rule out any significant change in protein-column interaction which leave us with different apparent hydrodynamic radii between these two protein forms as a result of conformational change in solution. For the purpose of comparison, chromatograms at a high concentration of urea (~4 M) were overlaid. Under this urea concentration, the whole population of capsid form falls apart into an expanded form of dimer (Figure 5.6 gray chromatograms). Under the effect of urea, both wildtype dimer and mutant Y132A dimer formed the same expanded dimer form (Figure 5.6 gray chromatograms). Our intact protein HDX and SEC data did not rule out the hypothesis that Y132A has a different conformation in solution compared to the wildtype. However, it was not obvious why mutant Y132A with more compact structure (based on intact protein HDX data) eluted earlier than wildtype in SEC experiments. Numerous studies of HBV core protein consider a highly dynamic conformation for the native form of this protein [35, 63, 64, 69-71, 73] and more relaxed structure for mutant Y132A [63]. Our intact protein and high resolution HDX data were also consistent with a higher level of conformational dynamics for Cp149 compared to mutant Y132A. A possible explanation for the sooner elution time of Y132A is that this mutant has acquired a more extended

structure compared to cp149. The hypothesis on a disrupted inter-monomer contact in Y132A may explain its extended structure.

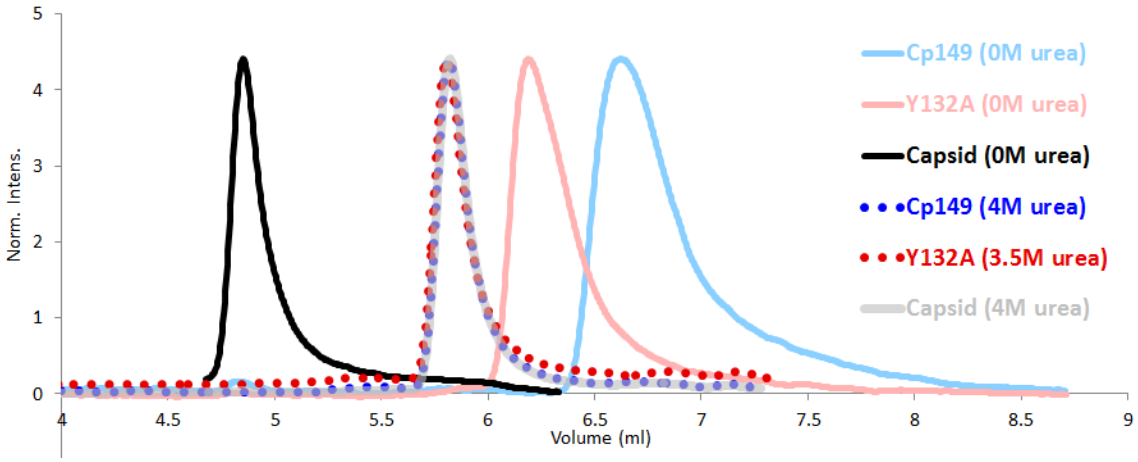

Figure 5.6: Size exclusion chromatograms of Y132A versus Cp149. The difference in elution times of Cp149 and Y132A is indicative of a difference in their apparent hydrodynamic diameters. The mutant Y132A with a larger molecular diameter eluted sooner than wtCp149. In high concentration of urea, both Cp149 and mutant Y132A acquire an expanded structure (blue and pink dots, respectively). Under the same concentration of urea, capsid totally falls apart into the same expanded protein form (gray). All protein forms in high concentration of urea are in an expanded form of dimer (for the purposes of simplification, the expanded monomer populations in the presence of urea which elute around 7.5ml, are not included).

Differential Fluorometry of Thermostability of Y132A versus Cp149

We tested the effect of intra-dimer disruption on the thermostability of Y132A by Differential Scanning Fluorometry (DSF). Figure 5.7 shows the differential thermostability of Y132A dimer versus the native one. The disrupted intra-dimer contacts in Y132A could be the reason for the lower thermostability of Y132A compared to the native dimer. The gradual increase of the temperature (a strong stimulator of capsid

assembly) during the DSF experiment resulted in some minor populations of capsid; the shoulder on the Cp149 signal is the result of melting of the capsid population.

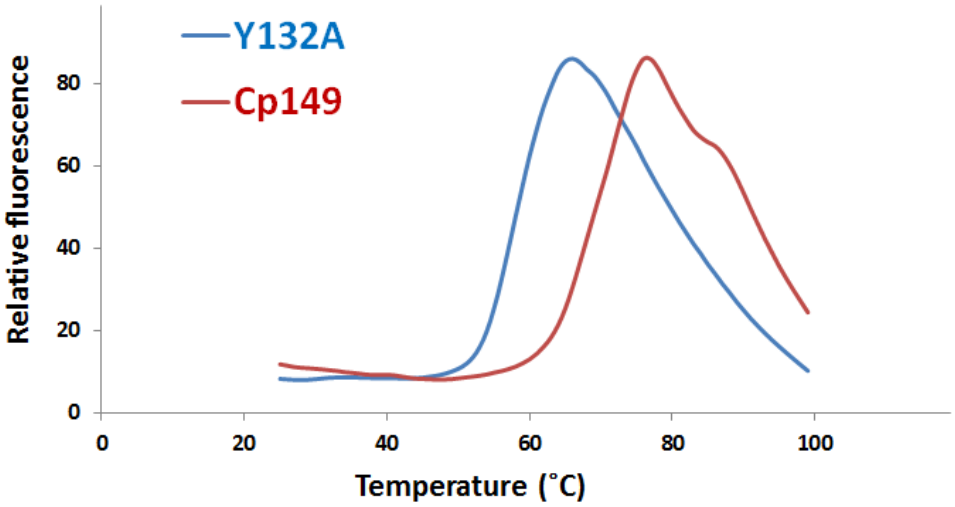

Figure 5.7: Differential scanning fluorometry of Y132A versus Cp149. The mutant Y132A shows a lower thermostability compared to Cp149. The gradual increase of the temperature (a strong stimulator of capsid assembly) during the DSF experiment resulted in some minor populations of capsid; the shoulder on the Cp149 signal is the result of melting of the capsid population.

Conclusion

Although X-ray crystallography and cryo-electron microscopy have provided the details of the structure and inter/intra-dimer contacts, a high-resolution solvent accessibility map of HBV capsid protein in solution remained to be presented. We hypothesized the presence of allosteric conformational differences between Cp149 and mutant Y132A. Our complementary biophysical approaches such as size exclusion chromatography and differential scanning fluorometry clearly showed the presence of some structural/conformational differences between wildtype and mutants forms of Cp149. Although they did not provide any structural details about the protein regions of

interest, they did not rule out the hypothesis either. The intact protein HDX data (Figure 3.10) clearly showed less overall solvent accessibility for the mutant. More detailed analysis of intact protein HDX data pointed to the possibility of some local structural compactness. Our automated high throughput HDX-LC-MS coupled to on-line pepsin digestion, however, presented a high resolution solvent accessibility or H-bonding stability map to differentiate between wtCp149 and the mutant Y132A. Such a high resolution map provided structural information on HBV core protein which was fully consistent with numerous studies of this protein. Based on the differential solvent accessibility map, we suggest that allosteric signal transduction across the protein structure plays a role in assembly. Although the mutation Y132 did not introduce any local solvent accessibility changes, it did impact solvent accessibility across the protein. The overall picture of deuteration change in Y132A compared to Cp149 implied some level of structural stability around the core. Such a structural compactness which is consistent with intact protein HDX data may introduce instability to the α-3 region and eventually destabilize the four-helix bundle and disrupt the intra-dimer contact area.

REFERENCES CITED

[1] Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. Protein Science. 2002;11:184-97.

[2] Benkovic SJ, Hammes-Schiffer S. A Perspective on Enzyme Catalysis. Science. 2003;301:1196-202.

[3] Koshland DE. Conformational changes: How small is big enough? Nat Med. 1998;4:1112-4.

[4] Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry. 2002;41:6573-82.

[5] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. J Mol Graph. 2001;19:26-59.

[6] Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Current Opinion in Structural Biology. 2002;12:54-60.

[7] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005;6:197-208.

[8] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of Molecular Biology. 1999;293:321-31.

[9] Uversky VN. Unusual biophysics of intrinsically disordered proteins. BBA-Proteins Proteomics. 2013;1834:932-51.

[10] Kamerlin SCL, Warshel A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? Proteins: Structure, Function, and Bioinformatics. 2009;78:1339.

[11] Onuchic JN, Luthey-Schulten Z, Wolynes PG. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. Annual Review of Physical Chemistry. 1997;48:545-600.

[12] Bothner B, Taylor D, Jun B, Lee KK, Siuzdak G, Schlutz CP, et al. Maturation of a tetravirus capsid alters the dynamic properties and creates a metastable complex. Virology. 2005;334:17.

[13] Lewis JK, Bothner B, Smith TJ, Siuzdak G. Antiviral agent blocks breathing of the common cold virus. Proceedings of the National Academy of Sciences. 1998;95:6774-8.

[14] Reisdorph N, Thomas JJ, Katpally U, Chase E, Harris K, Siuzdak G, et al. Human rhinovirus capsid dynamics is controlled by canyon flexibility. Virology. 2003;314:34-44.

[15] Rabe B, Delaleau M, Bischof A, Foss M, Sominskaya I, Pumpens P, et al. Nuclear Entry of Hepatitis B Virus Capsids Involves Disintegration to Protein Dimers followed by Nuclear Reassociation to Capsids. PLoS Pathog. 2009;5:13.

[16] Bothner B, Hilmer JK. Probing Viral Capsids in Solution. Structural Virology, Agbandje-McKenna, Mavis and McKenna, Robert. 2011:41-61.

[17] Li Q, Yafal AG, Lee YM, Hogle J, Chow M. Poliovirus neutralization by antibodies to internal epitopes of VP4 and VP1 results from reversible exposure of these sequences at physiological temperature. J Virol. 1994;68:3965-70.

[18] Zlotnick A, Stray SJ. How does your virus grow? Understanding and interfering with virus assembly. Trends in Biotechnology. 2003;21:536.

[19] Hadfield AT, Diana GD, Rossmann MG. Analysis of three structurally related antiviral compounds in complex with human rhinovirus 16. Proceedings of the National Academy of Sciences. 1999;96:14730-5.

[20] Tsang SK, Danthi P, Chow M, Hogle JM. Stabilization of poliovirus by capsid-binding antiviral drugs is due to entropic effects. Journal of Molecular Biology. 2000;296:335-40.

[21] Bothner B, Dong XF, Bibbs L, Johnson JE, Siuzdak G. Evidence of Viral Capsid Dynamics Using Limited Proteolysis and Mass Spectrometry. Journal of Biological Chemistry. 1998;273:673-6.

[22] Ganem D, Prince AM. Hepatitis B Virus Infection - Natural History and Clinical Consequences. New England Journal of Medicine. 2004;350:1118-29.

[23] Shepard CW, Simard EP, Finelli L, Fiore AE, Bell BP. Hepatitis B Virus Infection: Epidemiology and Vaccination. Epidemiologic Reviews. 2006;28:112-25.

[24] Packianathan C, Katen SP, Dann CE, III, Zlotnick A. Conformational Changes in the Hepatitis B Virus Core Protein Are Consistent with a Role for Allostery in Virus Assembly. J Virol. 2010;84:1607-15.

[25] Caspar DLD, Klug A. Physical Principles in the Construction of Regular Viruses. Cold Spring Harbor Symposia on Quantitative Biology. 1962;27:1-24.

[26] Bourne CR, Finn MG, Zlotnick A. Global Structural Changes in Hepatitis B Virus Capsids Induced by the Assembly Effector HAP1. J Virol. 2006;80:11055-61.

[27] Crowther RA, Kiselev NA, Böttcher B, Berriman JA, Borisova GP, Ose V, et al. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell. 1994;77:943.

[28] Singh S, Zlotnick A. Observed Hysteresis of Virus Capsid Disassembly Is Implicit in Kinetic Models of Assembly. Journal of Biological Chemistry. 2003;278:18249-55.

[29] Wynne SA, Crowther RA, Leslie AGW. The Crystal Structure of the Human Hepatitis B Virus Capsid. Molecular Cell. 1999;3:771.

[30] Bringas R. Folding and Assembly of Hepatitis B Virus Core Protein: A New Model Proposal. Journal of Structural Biology. 1997;118:189.

[31] Zlotnick A, Mukhopadhyay S. Virus assembly, allostery and antivirals. Trends in Microbiology. 2011;19:14.

[32] Zlotnick A, Cheng N, Stahl SJ, Conway JF, Steven AC, Wingfield PT. Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: Implications for morphogenesis and organization of encapsidated RNA. Proceedings of the National Academy of Sciences. 1997;94:9556-61.

[33] Birnbaum F, Nassal M. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. Journal of Virology. 1990;64:3319-30.

[34] KÃ¶ck J, Kann M, PÃ¼tz G, Blum HE, von WeizsÃ¤cker F. Central Role of a Serine Phosphorylation Site within Duck Hepatitis B Virus Core Protein for Capsid Trafficking and Genome Release. Journal of Biological Chemistry. 2003;278:28123-9.

[35] Gazina EV, Fielding JE, Lin B, Anderson DA. Core Protein Phosphorylation Modulates Pregenomic RNA Encapsidation to Different Extents in Human and Duck Hepatitis B Viruses. Journal of Virology. 2000;74:4721-8.

[36] Hilmer JK, Zlotnick A, Bothner B. Conformational Equilibria and Rates of Localized Motion within Hepatitis B Virus Capsids. Journal of Molecular Biology. 2008;375:581.

[37] Moisant P, Neeman H, Zlotnick A. Exploring the Paths of (Virus) Assembly. Biophysical Journal. 2010;99:1350.

[38] Bourne CR, Katen SP, Fulz MR, Packianathan C, Zlotnick A. A Mutant Hepatitis B Virus Core Protein Mimics Inhibitors of Icosahedral Capsid Self-Assembly†. Biochemistry. 2009;48:1736-42.

[39] Stray SJ, Ceres P, Zlotnick A. Zinc Ions Trigger Conformational Change and Oligomerization of Hepatitis B Virus Capsid Protein†. Biochemistry. 2004;43:9989-98.

[40] Bourne C, Lee S, Venkataiah B, Lee A, Korba B, Finn MG, et al. Small-Molecule Effectors of Hepatitis B Virus Capsid Assembly Give Insight into Virus Life Cycle. Journal of Virology. 2008;82:10262-70.

[41] Katen SP, Chirapu SR, Finn MG, Zlotnick A. Trapping of Hepatitis B Virus Capsid Assembly Intermediates by Phenylpropenamide Assembly Accelerators. ACS Chemical Biology. 2010;5:1125.

[42] Stray SJ, Bourne CR, Punna S, Lewis WG, Finn MG, Zlotnick A. A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:8138-43.

[43] Park C, Marqusee S. Probing the High Energy States in Proteins by Proteolysis. Journal of Molecular Biology. 2004;343:1467.

[44] Fontana A, Laureto PPd, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. Acta Biochim Pol. 2004;51:299–321.

[45] Hubbard SJ. The structural aspects of limited proteolysis of native proteins. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology. 1998;1382:191.

[46] Krause IM, Linderstrom-Lang K. Exchange of deuterium and 18O between water and other substances. 2. Alternative method. Compt Rend Trav Lab Carlsberg Ser Chim. 1955;29:367-84.

[47] Zhang Z, Smith DL. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. Protein Science. 1993;2:522.

[48] Tsutsui Y, Wintrode PL. Hydrogen/Deuterium Exchange-Mass Spectrometry: A Powerful Tool for Probing Protein Structure, Dynamics and Interactions. Current Medicinal Chemistry. 2007;14:2344.

[49] Wales TE, Engen JR. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. Mass Spectrometry Reviews. 2006;25:158-70.

[50] Busenlehner LS, Armstrong RN. Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry. Archives of Biochemistry and Biophysics. 2005;433:34.

[51] Hoofnagle AN, Resing KA, Ahn NG. PROTEIN ANALYSIS BY HYDROGEN EXCHANGE MASS SPECTROMETRY. Annual Review of Biophysics & Biomolecular Structure. 2003;32:1.

[52] Smith D. Local structure and dynamics in proteins characterized by hydrogen exchange and mass spectrometry. Biochemistry (Moscow). 1998;63:285-93.

[53] Woods VL, Hamuro Y. High resolution, high-throughput amide deuterium exchange-mass spectrometry (DXMS) determination of protein binding site structure and dynamics: Utility in pharmaceutical design. Journal of Cellular Biochemistry. 2001;84:89-98.

[54] Hamuro Y, Coales SJ, Southern MR, Nemeth-Cawley JF, Stranz DD, Griffin PR. Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry. Journal of biomolecular techniques : JBT. 2003;14:171-82.

[55] Fitzgerald MC, West GM. Painting Proteins with Covalent Labels: What's In the Picture? J Am Soc Mass Spectrom. 2009;20:1193.

[56] Wiedenheft B, Mosolf J, Willits D, Yeager M, Dryden KA, Young M, et al. An archaeal antioxidant: Characterization of a Dps-like protein from Sulfolobus solfataricus. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:10551-6.

[57] Gauss GH, Reott MA, Rocha ER, Young MJ, Douglas T, Smith CJ, et al. Characterization of the Bacteroides fragilis bfr Gene Product Identifies a Bacterial DPS-Like Protein and Suggests Evolutionary Links in the Ferritin Superfamily. Journal of Bacteriology. 2012;194:15-27.

[58] Gauss GH, Benas P, Wiedenheft B, Young M, Douglas T, Lawrence CM. Structure of the DPS-Like Protein from Sulfolobus solfataricus Reveals a Bacterioferritin-Like Dimetal Binding Site within a DPS-Like Dodecameric Assembly. Biochemistry. 2006;45:10815.

[59] Rayaprolu V, Kruse S, Kant R, Venkatakrishnan B, Movahed N, Brooke D, et al. Comparative Analysis of Adeno-Associated Virus Capsid Stability and Dynamics. Journal of Virology. 2013;87:13150-60.

[60] Ceres P, Zlotnick A. Weak Protein−Protein Interactions Are Sufficient To Drive Assembly of Hepatitis B Virus Capsids†. Biochemistry. 2002;41:11525-31.

[61] Zlotnick A, Ceres P, Singh S, Johnson JM. A Small Molecule Inhibits and Misdirects Assembly of Hepatitis B Virus Capsids. Journal of Virology. 2002;76:4848-54.

[62] Zlotnick A, Cheng N, Conway JF, Booy FP, Steven AC, Stahl SJ, et al. Dimorphism of Hepatitis B Virus Capsids Is Strongly Influenced by the C-Terminus of the Capsid Protein. Biochemistry. 1996;35:7412-21.

[63] Alexander CG, Jürgens MC, Shepherd DA, Freund SMV, Ashcroft AE, Ferguson N. Thermodynamic origins of protein folding, allostery, and capsid formation in the human hepatitis B virus core protein. Proceedings of the National Academy of Sciences. 2013;110:E2782-E91.

[64] Freund SMV, Johnson CM, Jaulent AM, Ferguson N. Moving towards High-Resolution Descriptions of the Molecular Interactions and Structural Rearrangements of the Human Hepatitis B Core Protein. Journal of Molecular Biology. 2008;384:1301-13.

[65] Gianni S, Camilloni C, Giri R, Toto A, Bonetti D, Morrone A, et al. Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. Proceedings of the National Academy of Sciences. 2014;111:14141-6.

[66] Li W, Wolynes PG, Takada S. Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. Proceedings of the National Academy of Sciences. 2011;108:3504-9.

[67] Selzer T, Schreiber G. New insights into the mechanism of protein–protein association. Proteins: Structure, Function, and Bioinformatics. 2001;45:190-8.

[68] Zlotnick A, Johnson JM, Wingfield PW, Stahl SJ, Endres D. A Theoretical Model Successfully Identifies Features of Hepatitis B Virus Capsid Assembly†. Biochemistry. 1999;38:14644-52.

[69] Bereszczak JZ, Watts NR, Wingfield PT, Steven AC, Heck AJR. Assessment of differences in the conformational flexibility of hepatitis B virus core-antigen and e-antigen by hydrogen deuterium exchange-mass spectrometry. Protein Science. 2014;23:884-96.

[70] Porterfield JZ, Dhason MS, Loeb DD, Nassal M, Stray SJ, Zlotnick A. Full-Length Hepatitis B Virus Core Protein Packages Viral and Heterologous RNA with Similarly High Levels of Cooperativity. Journal of Virology. 2010;84:7174-84.

[71] Selzer L, Katen SP, Zlotnick A. The Hepatitis B Virus Core Protein Intradimer Interface Modulates Capsid Assembly and Stability. Biochemistry. 2014;53:5496-504.

[72] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins: Structure, Function, and Bioinformatics. 1995;21:167-95.

[73] Bereszczak JZ, Rose RJ, van Duijn E, Watts NR, Wingfield PT, Steven AC, et al. Epitope-distal Effects Accompany the Binding of Two Distinct Antibodies to Hepatitis B Virus Capsids. J Am Chem Soc. 2013;135:6504-12.

[74] Wojtas M, Kaplon TM, Dobryszycki P, Ozyhar A. The effect of counter ions on the conformation of intrinsically disordered proteins studied by size-exclusion chromatography. Methods in molecular biology (Clifton, NJ). 2012;896:319-30.

[75] Uversky VN. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. Biochemistry. 1993;32:13288-98.

APPENDIX A

THE ACQUISITION METHOD REPORT ON HDX-LC/MS COUPLED TO ON-LINE PEPSIN DIGESTION

Acquisition method report:

Device List

1290 Sampler
1290 BinPump
1260 QuatPump
1290 TCC
MS Q-TOF

TOF/Q-TOF Mass Spectrometer

Component Name	MS Q-TOF	**Component Model**	G6510A
Ion Source	Dual ESI	**Tune File**	AutoTune.tun
Stop Mode	NoLimit	**Stop Time**	30.00
Can wait for temp.	Enable	**Fast Polarity**	N/A
MS1CentroidDataAbsThreshold	10	**MS1CentroidDataRELThreshold**	0.010
MS2CentroidDataAbsThreshold	10	**MS2CentroidDataRELThreshold**	0.010

Time Segments

Time Segment #	Start Time	Diverter Valve State	Storage Mode	Ion Mode
1	0.0	MS	Both	Dual ESI
2	2.4	Waste	Both	Dual ESI
3	3.1	MS	Both	Dual ESI

Time Segment 1

Acquisition Mode MS1

Min Range	300
Max Range	1700
Scan Rate	1.00

Source Parameters

Parameter	Value
Gas Temp (°C)	350
Gas Flow (l/min)	12.8
Nebulizer (psi)	58

Scan Segments

Scan Seg #	Ion Polarity
1	Positive

Scan Segment 1

Scan Source Parameters

Parameter	Value
VCap	4000
Fragmentor	195
Skimmer1	46.5
OctopoleRFPeak	750

Time Segment 2

Acquisition Mode MS1

Min Range	300
Max Range	1700
Scan Rate	1.00

Source Parameters

Parameter	Value
Gas Temp (°C)	300
Gas Flow (l/min)	12.9
Nebulizer (psi)	58

Scan Segments

Scan Seg #	Ion Polarity
1	Positive

Scan Segment 1

Scan Source Parameters

Parameter	Value
VCap	4000
Fragmentor	195
Skimmer1	45.0
OctopoleRFPeak	540

Time Segment 3

Acquisition Mode MS1

Min Range	300
Max Range	1700
Scan Rate	1.00

Source Parameters

Parameter	Value
Gas Temp (°C)	350
Gas Flow (l/min)	12.9
Nebulizer (psi)	58

Scan Segments

Scan Seg #	Ion Polarity
1	Positive

Scan Segment 1

Scan Source Parameters

Parameter	Value
VCap	4000
Fragmentor	195
Skimmer1	45.0
OctopoleRFPeak	750

ReferenceMasses

Ref Mass Enabled	Disabled
Ref Nebulizer	

Chromatograms

Chrom Type	Label	Offset	Y-Range
TIC	TIC	15	10000000

Quaternary Pump

Name	1260 QuatPump	**Model**	G1311C
Ordinal #	1	**Options**	
Stop time (min)	No Limit	**Post Time (min)**	Off
Flow (ml/min)	0.5	**Pressure Min (bar)**	0
Pressure Max (bar)	200	**Max Flow Gradient (ml/min)**	100
Solvent A		**Solvent B**	
Solvent C		**Solvent D**	
Solvent Ratio A	0	**Solvent Ratio B**	0
Solvent Ratio C	80	**Solvent Ratio D**	20
Compress. A (*10-6/bar)	50	**Stroke A (µl)**	Auto
Primary Channel	0		
Contact 1	Off		
Contact 2	Off		
Contact 3	Off		
Contact 4	Off		

Pump Time Table

Time	Flow	Pressure	Solv Ratio B	Solv Ratio C	Solv Ratio D
0.1	1.5	No Change	0	80	20
0.19	1.5	No Change	0	80	20
0.2	0.6	No Change	0	80	20
1.5	0.6	No Change	0	80	20
1.6	1.3	No Change	0	80	20
2.5	1.3	No Change	0	80	20
2.51	0.5	No Change	0	20	80

Signals

Description
Pressure
Flow
Solvent% C
Solvent% D

Binary Pump

Name	1290 BinPump	**Model**	G4220A
Ordinal #	1	**Options**	SSV
Stop Time (min)	8.5	**Post Time (min)**	Off
Flow (ml/min)	0.7	**Pressure Min (bar)**	0
Pressure Max (bar)	350	**Max Flow Gradient (ml/min)**	100
Solvent A	H2O, 0.1% formic	**Solvent B**	ACN, 0.1% formic
Solvent Ratio A	10	**Solvent Ratio B**	90
Solvent Type A1	Aqueous	**Solvent Type B1**	Organic
Solvent Type A2	-----	**Solvent Type B2**	-----
Compress. A (*10-6/bar)	50	**Compress. B (*10-6/bar)**	115
Stroke A (µl)	Auto	**Stroke B (µl)**	Auto
Stroke Syncronization	Yes		
Contact 1	Off		
Contact 2	Off		
Contact 3	Off		
Contact 4	Off		

Pump Time Table

Time	Flow	Pressure	Solv Ratio B
0.49	0.7	No Change	90
0.5	0.7	No Change	3
2.6	0.7	No Change	3
2.7	0.65	No Change	10
7.7	0.65	No Change	30
8.3	0.9	No Change	90

Signals

Description
Pressure
Flow
Solvent% B
Direction of piston A
Direction of piston B

Wellplate Sampler

Name	1290 Sampler	**Model**	G4226A
Ordinal #	1	**Options**	THM
Stop time (min)	As Pump	**Post Time (min)**	Off
Injection Type	Standard Injection	**Injection Volume (µl)**	20
Overlap Time (min)	Disable Overlapped Injection	**Draw Position (mm)**	1
Draw Position Detection	1	**Draw Speed (µl/min)**	120
Eject Speed (µl/min)	600	**Flush Out Factor**	4
Automatic Delay Volume Reduction	No	**Equilibration Time (sec)**	5
Wash Vessel	N/A	**Wash Location**	N/A
Wash Time (sec)	N/A	**Wash Cycles**	N/A
Contact 1	Off		
Contact 2	Off		
Contact 3	Off		
Contact 4	Off		

Signals

Description
Temperature

Thermostated Column Compartment

Name	1290 TCC	**Model**	G1316C
Ordinal #	1	**Options #**	10Port2Pos
Stop time (min)	As Pump	**Post Time (min)**	Off
Left Temp. (°C)	-1	**Right Temp. (°C)**	Same as left
Left Ready (°C)	When Temp Within Set Point +/- 3	**Right Ready (°C)**	When Temp Within Set Point +/- 3
Valve Position	0		
Contact 1	Off		
Contact 2	Off		
Contact 3	Off		
Contact 4	Off		

Temperature Time Table

Time	Left Temp. (°C)	Right Temp. (°C)	Column Valve
1			1
2.5			0