5-2016

Staphylococcal Nuclease and Ubiquitin Local Folding Energies and Rates using PEPS-HDX-ESI-MS

Julie Rhee

Staphylococcal Nuclease and Ubiquitin Local Folding Energies and Rates using PEPS-HDX-ESI-MS

An Honors Thesis submitted in partial fulfillment of the requirements of

Honors Studies in Biochemistry

By

Julie Rhee

2016

Biochemistry

J. William Fulbright College of Arts and Sciences

The University of Arkansas

Acknowledgements

Statewide Mass Spectrometry Facility

Dr. Wesley Stites research group

Dr. Rohana Liyanage

Past studies: Hayden Pacl, Reece Ritter, Derek Pyland

Table of Contents

I. Abstract

In this study, Protein Equilibrium Population Snapshot Hydrogen-Deuterium Exchange Electrospray Ionization Mass Spectrometry (PEPS-HDX-ESI-MS) was applied to study the local regions of model proteins, staphylococcal nuclease and ubiquitin. The hydrogen deuterium exchange (HDX) has become a key technique for studying the structural and dynamic aspects of proteins in solution. This technique creates a rapid exchange between all of the exchangeable hydrogen ions with deuterium when the protein is exposed to a D_2O solvent. The PEPS method is an equilibrium-based method used to determine the populations of the closed native and open denatured states of a protein. By combining the applications of HDX and the PEPS method using ESI-MS, one can determine the solvent accessible and protected amide protons, the folding energies and rates of a protein in physiological conditions through linearly extrapolating the folding energies, and the rates by systematically denaturing conditions using high guanidine hydrochloride (GdHCl).

Past studies have applied this method on intact model proteins, staphylococcal nuclease and ubiquitin. Both proteins showed the expected amount of exchangeable amide protons, which was verified from the X-ray structures. The folding energies, folding rates, and unfolding rates for staphylococcal nuclease and ubiquitin were estimated to be -4.8 kcal mol^{-1}, 10 s^{-1}, $2x10^{-3}$ s^{-1} and -8.8 kcal/mol, 251 s^{-1}, $4x10^{-5}$ s^{-1} respectively. This work successfully deconvoluted local HDX coverage, local folding energies, and local folding rates from this intact protein information. This was accomplished by dissecting or digesting the intact protein using pepsin after the HDX without losing already incorporated deuterium under HDX quenching conditions, at pH 2.7 and 0° C followed by LC-ESI-MS analysis. These results were compared to the results of the local regions to investigate how independent they were to the intact proteins.

The same procedure applied in the past study involving the intact proteins was applied in this study but with a pepsin digestion. This digestion was performed after the HDX, but before the LC-ESI-MS analysis. The resulted pepsin fragments contained all of the necessary information for the folding energies, folding rates, and accessible areas for local regions. Control experiments were performed to identify the local regions using multiple mass spectrometry methods. The intact peptide mass was searched on an available software of in silico pepsin digestion of the protein being analyzed. For further verification, fragmentation was applied with LC-MS/MS, matrix-assisted laser desorption/ionization time of flight (MALDI-TOF), and matrix-assisted laser desorption/ionization Fourier transfer mass spectrometry (MALDI-FT-MS).

Local regions of the staphylococcal nuclease had energies between -3 to -5 kcal/mol, and the folding rates ranged from 0.01 to 10 s^{-1}. Unfolding rates on average seemed to be in close proximity, $1.5x10^{-4}$ s^{-1}. This was probably consistent with the cooperativity in unfolding proteins. The local regions of ubiquitin had energies ranging from -1 to -10 kcal/mol, folding rates ranged from 10^{-3} -10^{8} s^{-1}. Again, similar to staphylococcal nuclease, ubiquitin also has on average similar unfolding rates, ~ 10^{-5} s^{-1}.

This study attempted to answer several important questions. For example, how much influence and independence do the local secondary structures (alpha helices, beta sheets, and loops) have on the intact protein properties? Can the intact protein structure be constructed in cases where the folding and unfolding properties of local regions are not known? Can a database be constructed in which it has experimental sequence specific folding and unfolding properties?

II. Introduction

A. Protein Folding

Proteins are known to be an essential component to life. They assist in many biological processes allowing organisms to properly function. Biocatalysis, transportation, host defense, and energy conversion are a few of these crucial processes.[1] Such processes are driven by the protein's conformational state. A folded protein is considered biologically active. An unfolded or misfolded protein decreases a protein's ability to properly function. Noncovalent interactions, such as hydrogen bonding, ionic interaction, Van der Waals' forces, and hydrophobic effects, are the key influential sources of a protein's conformational state. A protein is constantly folding and unfolding, but the bonding forces ensure the native state is more stable than the denatured state.[2] Denaturing agents and changes within the physiological environment can disrupt the stability of the native state, causing the lifespan of the unfolded protein to drastically increase. Agents such as urea or guanidine hydrochloride (GdHCl), as well as changes in the pH or temperature can upset the balance of the noncovalent interactions that stabilize the native conformation.[3,4,5,6]

Studies have presented these types of affected proteins to correlate to some of the most severe diseases known today.[6,7] For example, cystic fibrous is a disease that causes severe damage to the lungs, which arises from a mutation within the cystic fibrous trans-membrane conductance regular gene that encodes a crucial transport protein to fold incorrectly.[7] This demonstrates the strength of how a small alteration in a given sequence has the ability to completely disrupt a protein's function. Besides cystic fibrous, Alzheimer's disease, Parkinson's disease, Creutzfeldt-Jakokb disease, and type II diabetes are correspondingly rooted from an unfolded or misfolded protein.[6,8]

From this, a question arises on whether the specific dysfunction on an unfolded or misfolded protein can be identified to increase the possibilities of fixing or replacing the defected region. Particularly, this would require a greater understanding of proteins and their folding mechanisms. Unraveling the protein folding mechanisms has been known to be one of the most challenging problems. This study applied Protein Equilibrium Population Snapshot Hydrogen-Deuterium Exchange Electrospray Ionization Mass Spectrometry (PEPS-HDX-ESI-MS) to investigate the local regions in model proteins, staphylococcal nuclease and ubiquitin by simultaneously extracting solvent accessible areas, identifying each local region, estimating the protein folding energies and protein folding/unfolding rates of the local regions, and observing how independent the local regions were in respect to the intact proteins.

B. Electrospray Ionization Mass Spectrometry

A considerable number of methods have been introduced in successfully determining protein folding energies and protein folding/unfolding rates. Two of the most widely used spectrophotometric methods are tryptophan fluorescence and circular dichroism (CD).[2] Studies have applied these methods in extracting protein thermodynamic stability, however, there are a few drawbacks. Fluorescence only studies the region of a protein proximal to a tryptophan side chain, decreasing its usefulness to many proteins due to the low abundance of tryptophan in most proteins. The CD method provides a more complete, global data, but the signal obtained is fairly weak. This method also requires the protein be in large amounts and the protein must be pure. Aside from these methods, hydrogen deuterium exchange (HDX) kinetics studies of amide protons by nuclear magnetic resonance (NMR) spectroscopy has proven to be a powerful tool for monitoring specific regions of a molecule even to partially unfolded.[5] Unfortunately, the size

range accessible to NMR is somewhat limited and requires high skills for interpreting spectrums of large proteins.[2]

An alternative to monitoring HDX using NMR is using mass spectrometry (MS). The electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) has emerged in the field of protein interaction studies since the late 1980s.[5,6] They have accurately determined molecular weights of large biomolecular assemblies up to several hundred thousand Daltons. ESI-MS, specifically, has increased the application of MS in analytical biochemistry due to its capability of analyzing proteins from aqueous solutions under physiological conditions.[6,9,10] It has also been used for protein identification, quantification, structure elucidation, and many other applications. Most importantly, the amount of protein required is significantly smaller than the other methods and is compatible with complex mixtures.[2]

C. Hydrogen Deuterium Exchange (HDX) Fundamentals

The hydrogen deuterium exchange (HDX) has become a key technique for studying the structural and dynamic aspects of proteins in solution.[5,11] Within this reaction, rapid amide hydrogen to deuterium exchange occurs on a protein when exposed to a D_2O solvent. When the protein is folded (closed), part of the peptide backbone is buried in the hydrophobic core and the amide protons in that region are protected from solvent exposure. This prevents deuterium from exchanging with these protons and makes them less available or accessible to exchange. When the protein unfolds (open), this globally protected region becomes accessible and can be exchanged with deuterium in the solvent in conditions with faster HDX rates compared to the folding and unfolding rates. As a result, the closed and open states can be differentiated by mass.

The schematics of this HDX reaction can be seen in below in **Figure 1**.

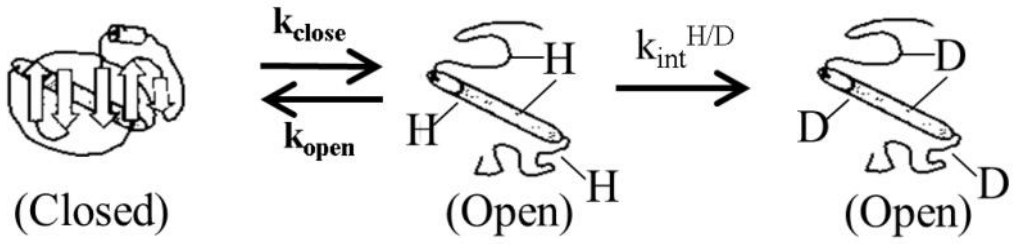

Figure 1: Classical "breathing model" of HDX – The protein is initially closed with no HDX exchange. As the protein unfolds from denaturants, the protected amide protons are exposed allowing an HDX exchange to occur towards these protons. The k_{close} represents the folding rates, while the k_{open} represents the unfolding rates of the protein. The $k_{int}^{H/D}$ is the rate of H/D exchange of an unprotected amide proton.[12]

In general, the HDX has two commonly accepted mechanisms, EX1 and EX2. Both of these mechanisms represent the rate at which the globally protected amide protons exchange. EX1 is defined as a condition where the majority of the protein's protected amide protons exchange. EX2 is defined as a condition where little or no exchange occurs for the protein's protected amide protons. For an EX1 mechanism, the $k_{int}^{H/D}$ is much higher than the folding rates. This is what causes the HDX of the protected amide protons to occur immediately and simultaneously when the protein opens.[12] **Figure 2** below shows the relationship between the HDX rates and the mechanisms.

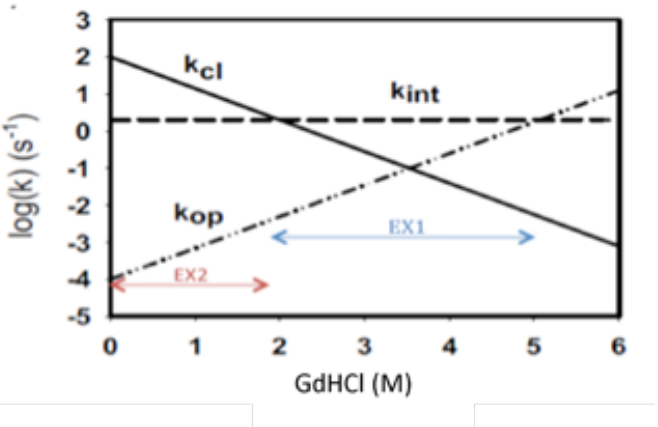

Figure 2: HDX rates of a model protein[12]

Looking at **Figure 2**, it is shown that the determination of whether the protein is under the EX1 or EX2 condition is highly dependent upon the concentration of the denaturant. EX1 is achieved when the k_{int} is greater than k_{cl} and k_{op}. EX2 is achieved when k_{int} is less than either k_{cl} or k_{op}. k_{int} remains constant because it is only dependent on the pH. The concentration of GdHCl needed to achieve EX1 varies per protein as the k_{cl} and k_{op} differ.

In the EX1 regime, HDX produces two distinguishable ESI-MS peaks for open and closed states. The closed state has only the amide protons exposed on the surface of a protein, while the open state has the amide protons protected inside of a protein. The difference in the number of amide protons exchanged with deuterium is resolved by ESI-MS where two peaks can be used to obtain closed and open population just from using their intensities. These populations can be used to calculate protein folding energies and folding/unfolding rates. [12]

D. PEPS Method

The Protein Equilibrium Snapshot (PEPS) method is an equilibrium-based method developed to determine the populations of the closed native and open denatured states of a protein, at a given denaturant concentration where both states have a longer lifetime than an exposed amide proton. In this study, guanidine hydrochloride (GdHCl) was used as the denaturant. The PEPS method works accurately at denaturant concentrations where HDX follows the EX1 mechanism. If the PEPS method is performed far away from the EX1 regime, the one second (~1s) HDX would underestimate the open population. The two populations of a protein are directly measured from the intensities of two m/z peaks produced by ESI-MS.[2] The population measurements obtained from ~1s HDX can be applied to determine protein folding

energies, and rates at longer HDX times at a given denaturant concentration. In both situations, linear extrapolation method (LEM) can be used to obtain the physiological properties of proteins at high denaturant concentrations.

E. Intact Staphylococcal nuclease and Ubiquitin

From previous PEPS studies, information on accessible regions, protein folding energies, and protein folding/unfolding rates of intact staphylococcal nuclease and ubiquitin were extracted.[12] Those results were compared to the results of the local regions of staphylococcal nuclease and ubiquitin found in this study.

For staphylococcal nuclease, the X-ray structure showed a total of 142 exchangeable amide protons with about 45 of those protons being solvent accessible in the native state (**Figure 3** and **4**). The X-ray structure of intact ubiquitin showed a total of 73 exchangeable amide protons with about 23 of those protons being solvent accessible in the native state (**Figure 5** and **6**).

Looking at X-ray structures, the solvent accessible protons are located on the external loops of the protein. The loops of the protein connect to the secondary structures of alpha helices and beta sheets. The protons around these structures are more tightly packed causing them to be more protected than the accessible protons.

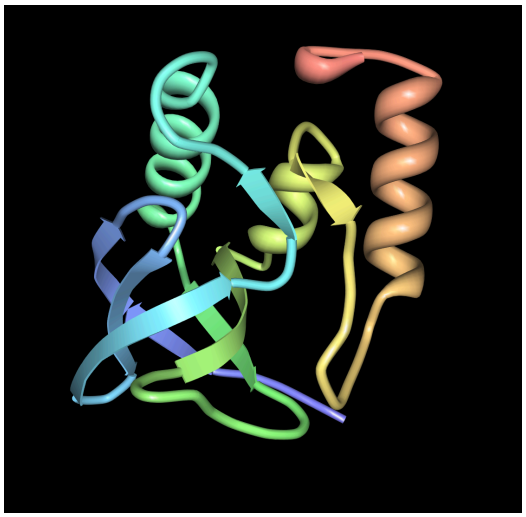

Figure 3: X-ray structure of staphylococcal nuclease from Protein Workshop[13]

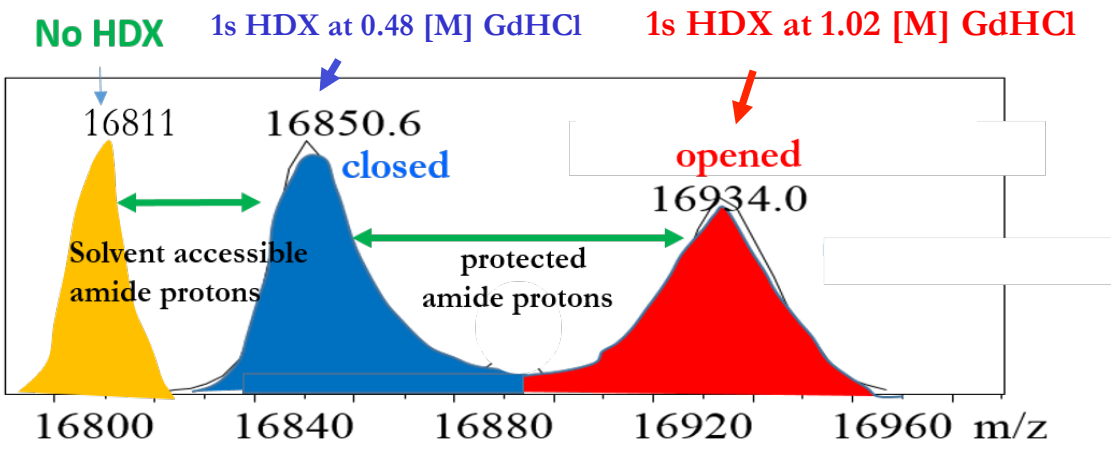

Figure 4: m/z peak intensities of intact staphylococcal nuclease with no HDX, ~1s HDX 0.48 M GdHCl at the closed state, and ~1 HDX 1.02 M GdHCl at the opened state.[12] The control experiment (no HDX) showed the intact protein to have a m/z peak intensitiy of 16811. From the ~1 HDX, it was revealed that staphylococcal nuclease has 40 accessible amide protons, and 83 protected amide protons.

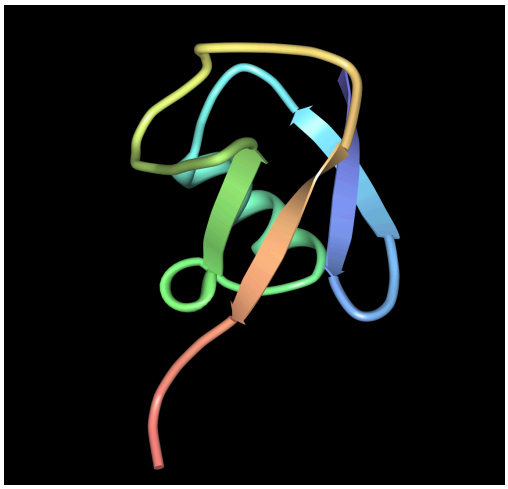

Figure 5: X-ray structure of ubiquitin from Protein Workshop[13]

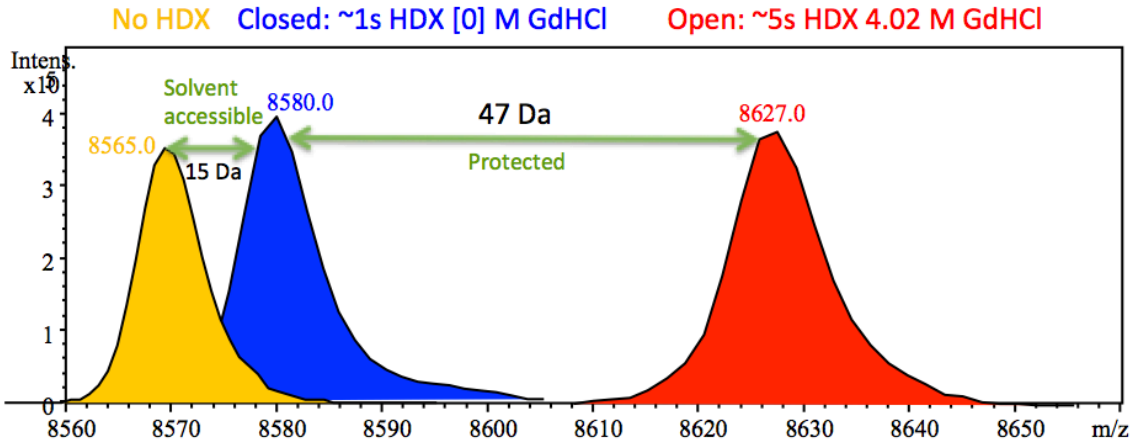

Figure 6: m/z peak intensities of intact ubiquitin with no HDX, ~1s HDX at the closed state, and ~1 HDX at the opened state.[12] The control experiment (no HDX) showed the intact protein to have a m/z peak intensity of 8565. From the ~1 HDX, it was revealed that ubiquitin has 15 accessible amide protons, and 47 protected amide protons.

The differences between open and closed state m/z peak intensities correlated with the number of amide protons in structured and protected regions (**Figures 4** and **6**). The PEPS results and folding energies for intact staphylococcal nuclease can be seen below in **Figure 7**. The PEPS results and folding energies for intact ubiquitin can be seen below in **Figure 8**. The folding/unfolding rates of both proteins can be seen in **Figure 9**. The intact protein information

was used to compare to investigate how independent the local regions were compared to the intact protein.

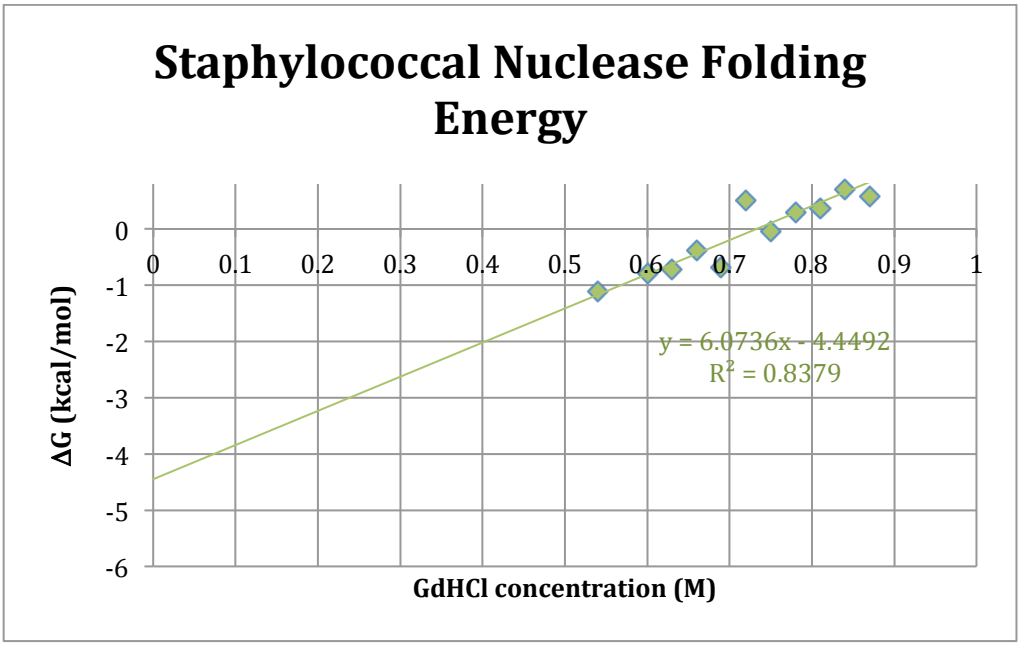

Figure 7: Folding energy for intact staphylococcal nuclease – The folding energy was determined to be -4.4 kcal/mol.[12]

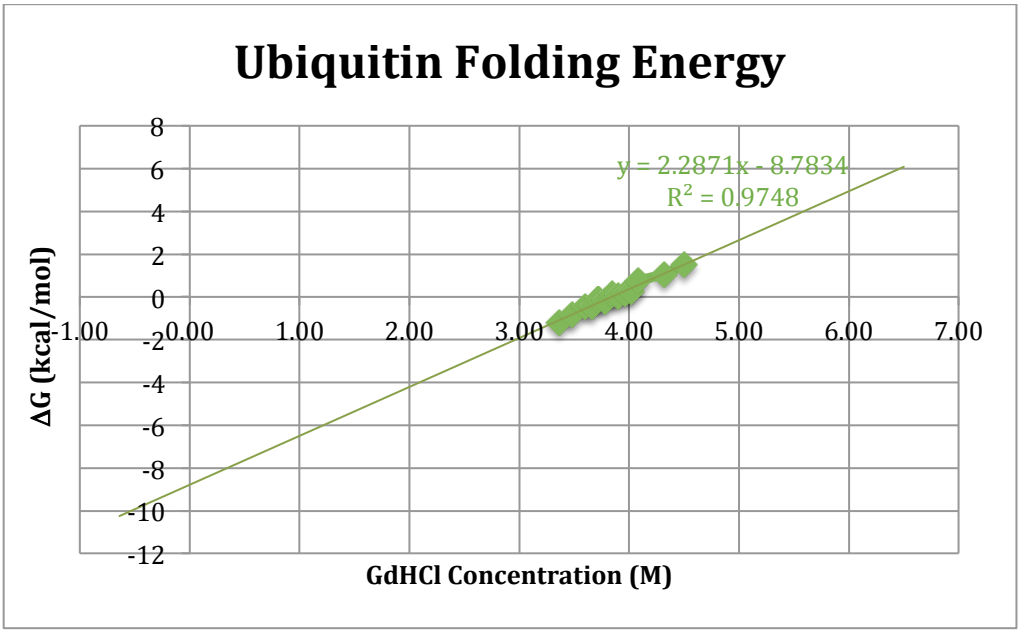

Figure 8: Folding energy for intact ubiquitin – The folding energy was determined to be -8.8 kcal/mol.[12]

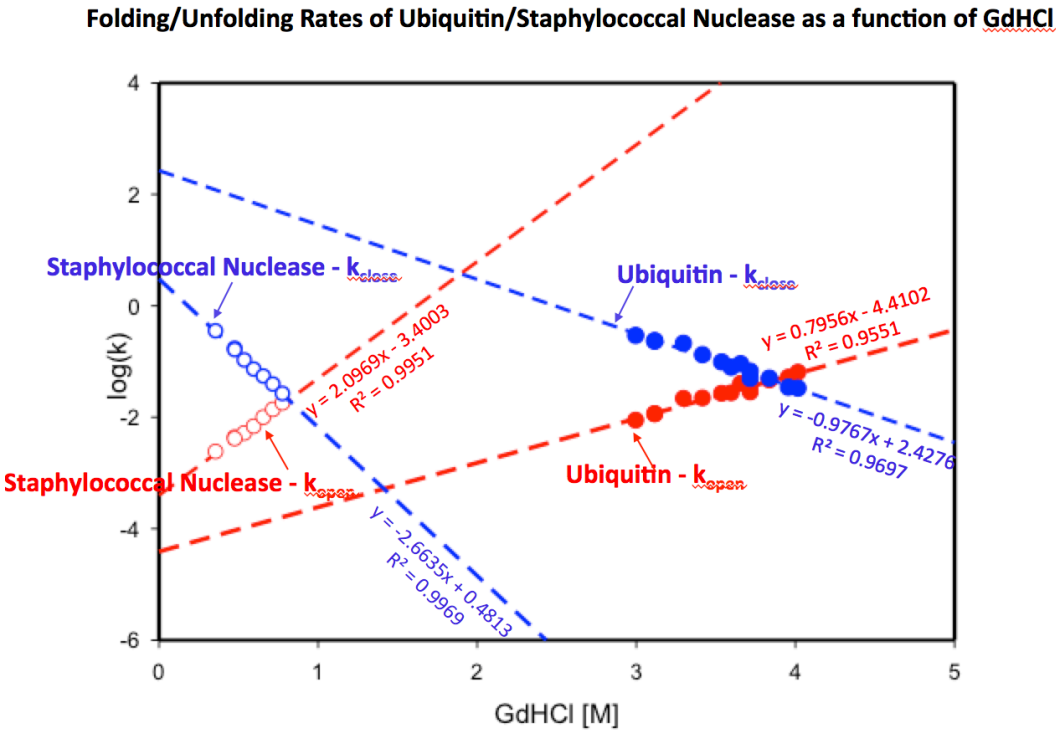

Figure 9: Folding/unfolding rates of staphylococcal nuclease and ubiquitin as a function of GdHCl[8]

III. Procedure

A. Chemicals

The expression and purification of staphylococcal nuclease used has been described elsewhere. Bovine ubiquitin, sodium hydrogen phosphate, acetonitrile (ACN), and pepsin were purchased from Sigma-Aldrich. The pepsin was used due to its nonspecific breakage of amino acids as enzyme to maximize the overlap areas to be internally cross checked. Pepsin was prepared as 6 mg/1 mL. Trifluoroacetic acid (TFA) was purchased from Halocarbon. Guanidine hydrochloride (GdHCl) omni pure was purchased from VWR. Sodium phosphate and sodium chloride were purchased from Mallinckrodt. Deuterium oxide (D_2O; 99.8% atom D) was

purchased from Cambridge Isotope Laboratories. H_2O and D_2O phosphate buffers were prepared to have final concentrations of 100 mM NaCl, and 25 mM sodium phosphate, at pH 7.0. A stock solution of 6M GdHCl was prepared in 100 mM NaCl with 25 mM sodium phosphate at pH 7.0. Deuterated guanidine hydrochloride (GdDCl) was prepared by ten cycles of evaporation and reconstitution to the same volume with pure D_2O to maintain the buffer condition and pH. All protein samples were dissolved in the H_2O buffer to achieve a final protein concentration of about 2 mg/ml. The pH measurements were not adjusted for deuterium. The HPLC mobile phases were 0.1% TFA (A) and 0.1% TFA in ACN (B), which were prepared using TFA and dilution with water.

B. Instrumentation

A Bruker Esquire 2000 (Billerica, MA) LC ion trap equipped with an electrospray ionization source was used for LC-MS. It was operated in the positive ion mode with a nebulizing gas pressure (N2) of 32 Psi and a drying gas flow of 12 ml/min maintained at 250° C. The mass spectrometer was optimized at m/z 1000 with low skimmer voltage (instrument default for this mass). The HPLC was a Hewlett Packard (Palo Alto, CA) 1100 series instrument equipped with an autosampler. The pepsin digestion after HDX was conducted through LC-MS by using a Supelco C18 column (4.5 mm × 50 mm, 5 μm) at a flow rate of 0.8 ml/min using a rapid gradient from 5%B – 45B% over 10 minutes to minimized the back exchange. All the solvents and the LC column were kept in ice at 0° C and solvents were kept at ~pH 2.7 at HDX quenching conditions to minimize the back exchange.

C. PEPS HDX-ESI-MS method

In this study, regional specific detailed local information within the protein was obtained by simply adding a pepsin digestion step to the standard protocol for the intact proteins.[8] In other words the idea was to maintain and protect the HDX that already took place for the intact protein by simply transferring the HDX solution to the quenching buffer with pepsin. Ideally, the local information was obtained without disturbing the exchanged that already took place. Two model proteins were studied, staphylococcal nuclease and ubiquitin. For each protein, protein solutions and exchange buffer solutions of different GdHCl concentrations were prepared to systematically vary the ratio of folded/unfolded state of protein population.

For staphylococcal nuclease, PEPS experiments were conducted in the GdHCl concentrations that ranged from 0.48 to 1.02 M concentrations. For ubiquitin, PEPS experiments were conducted in the GdHCl concentrations that ranged from 3.3 to 4.02 M concentrations. Using 6M GdHCl stock solution, all the solutions in correct GdHCl were prepared. The protein solutions were prepared by mixing a specific amount of microliters of the protein stock solution with specific amounts of 6 M GdHCl stock solution to achieve appropriate denaturing conditions for the protein. The exchange buffer solutions with 100% deuterium environment were prepared by mixing a specific amount of 6 M GdDCl solution with deuterated phosphate buffer solution. For both proteins, the protein solutions and exchange buffer solutions were placed in a 25°C water bath for 5 to 10 minutes, allowing the solutions to reach the correct temperature.

After the establishment of the equilibrium, the HDX experiment was performed to the intact protein, then enzymatically cut into multiple fragments without affecting the already exchanged amide protons at pH ~2.7, 0° C where the hydrogen and deuterium exchange rate was minimum, basically under fully quenching conditions. The HDX was initiated by mixing 5 µl of

protein solution in GdHCl (non-deuterated, A) and 45 µl of exchange buffer solution (all deuterated, B) made of the same concentration of GdDCl (deuterated). These solutions were mixed for approximately 1s for the folding energy determination experiments and 5s to 180s for folding/unfolding rates determination experiments followed by transferring the mixture to a quenching solution containing pepsin in 0.2 % TFA. For staphylococcal nuclease, the quenching solution contained 50 µl of 0.2% TFA and 3 µl of pepsin solution (6 mg/mL), which was kept in cold ice at a pH ~2.7. For ubiquitin, the quenching solution contained 50 µl of pepsin solution (6 mg/mL in 0.2% TFA), which was also kept in cold ice at a pH ~2.7. The ubiquitin quenching solution used all pepsin in order to increase the intensities of the individual peptic fragment peak populations. For determining the folding energies, ~1s HDX exchange was performed for both proteins in all of their respective GdHCl concentration. For determining the folding rates, 60s, 90s, 120s, and 180s HDX experiments were performed for staphylococcal nuclease, while 5s, 15s, 20s, 25s, 30s, 45s HDX experiments were performed for ubiquitin.

Once the quenching was complete, the sample was loaded into a Rheodyne injector then injected into the HPLC-ESI-MS machine for analysis. For staphylococcal nuclease, the sample was injected directly in to LC-ESI-MS without bypassing for the solvent front, which carried most of the GdHCl because we use lower concentrations of GdHCl. For ubiquitin, the method valve switching was used to avoid GdHCl from getting infused into the ESI-MS. This method prevented any unnecessary ions from entering into the system and increased the sensitivity of the peptic fragments from eluting early in the instrument. GdHCl typically appears on the mass spectrum within the first 2 minutes of the run due to its non-polar properties. During the first 2 minutes of each ubiquitin run, the solution was disposed into a waste instead of the mass spectrometry. After the 2 minutes, the valve was switched so the solution would enter into the

mass spectrometry. Both HPLC solvents were kept in an ice bath, and the HPLC column was kept at 0°C.

D. Identification of Local Regions

The local regions of each protein were identified using multiple methods prior to the real HDX experiments. First, just the intact measured mass from LC-ESI-MS was used to identify the pepsin fragment based on the possible list of in silico pepsin fragment mass values generated from the Bruker Daltonics Sequence Editor software. In the situations where just the MS (± 0.3 Da) measurement alone was not sufficient to identify the pepsin fragment, LC-MS/MS, matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS), and matrix-assisted laser desorption/ionization Fourier transfer mass spectrometry (MALDI-FT-MS) were utilized.

The LC-ESI-MS produced multiple charge ions for most of the pepsin fragments. Hence, there were multiple m/z peak intensities that represented the individual regions. An example of a pepsin fragment representing a local region at m/z 588.3 from ubiquitin can be seen below in **Figure 10**. To verify that this peak was truly representing a pepsin fragment of the protein, a background chromatogram with just pepsin and the buffer was used as a control.

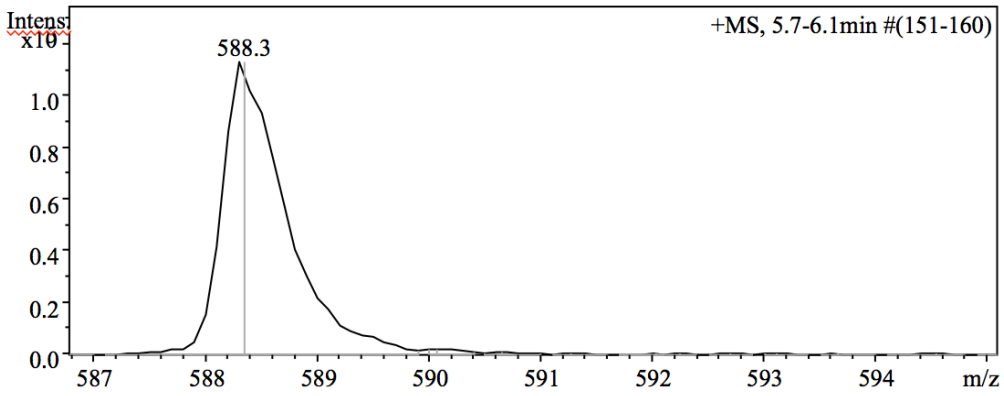

Figure 10: Pepsin fragment representing VKTLTGKTITL local region mass peak for ubiquitin – This m/z peak was obtained without HDX. This peak represents doubly charged ion.

With the observed mass of the local region, it was compared with all the possible pepsin fragment masses generated from the protein sequences using the Bruker Daltonics Sequence Editor. In the case for **Figure 10**, the m/z 588.3 was searched in the database. If the observed m/z represented a multiple charge, the m/z value was doubled or tripled before searching in the database. The observed mass in **Figure 10** was a double charge, so the mass was doubled to 1176.6 and subtracted a Dalton to obtain a protonated singly charged ion mass. This m/z value corresponded to the sequence VKTLTGKTITL. It was especially important to carefully see whether there were multiple possible sequences with the same m/z within $\pm$0.3 Da window. In such cases, the identification would be carried out using fragmentation by LC-MS/MS or MALDI-TOF MS/MS (LIFT-TOF/TOF), or using accurate high resolution mass measured by MALDI-FT-MS (masses measured to the accuracy of at least ±0.001 Da at 1100 Da,ppm mass accuracy).

When the mass spectrometer isolates the ion of interest, it is transferred into ESI where it collides multiple times with a stationary gas (helium). These collisions eventually cause the bonds between the amino acids to break, a concept known as MS/MS. In the case where fragmentation was applied to the sequence VKTLTGKTITL from **Figure 10**, the fragments in the mass spectrometer between peptide bonds can be seen below in **Figure 11**. The MS/MS fragmentation spectra for singly charged VKTLTGKTITL obtained by MALDI-TOF MS/MS can be seen below in **Figure 12**. The fragmentation data was combined with the MS data and searched in the database for the exact sequence of the local region.

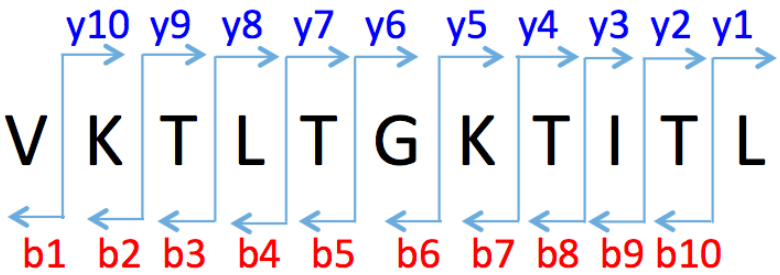

Figure 11: Bond breakage of local region ubiquitin VKTLTGKTITL – The blue y values represents the conservation of the C-terminus and the red b values represent the conservation of the N-terminus.

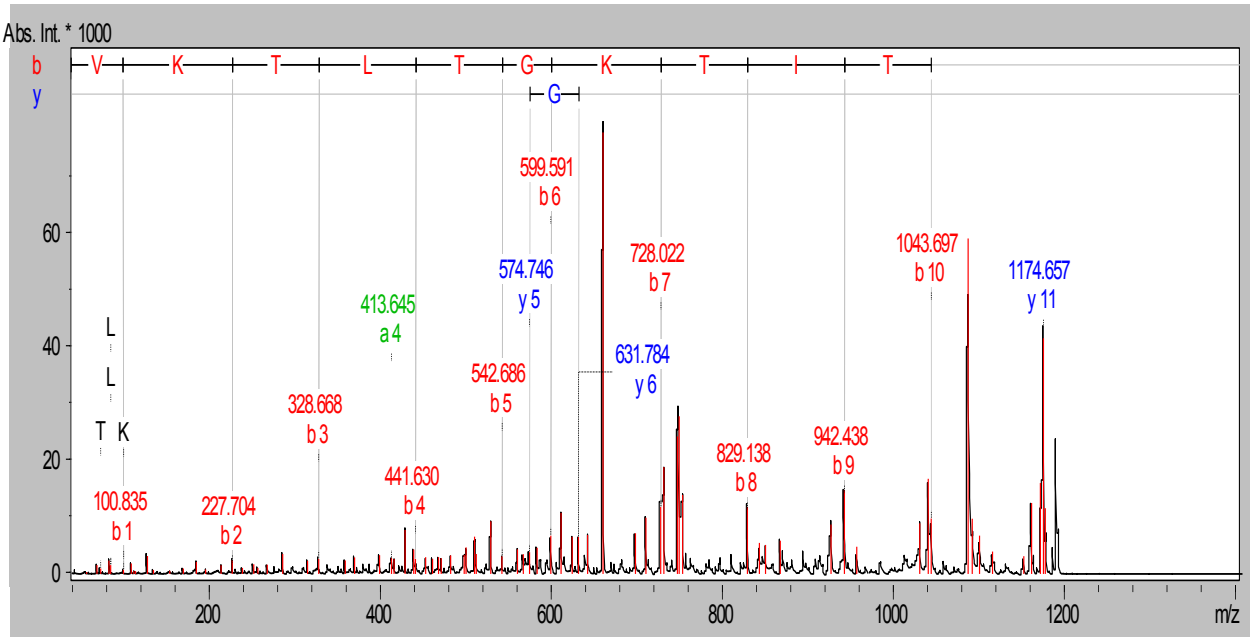

Figure 12: MS/MS fragmentation chromatogram for ubiquitin local region VKTLTGKTIT – The mass on top of the each peak represents the residual mass of each amino acid.

E. Calculation of folding energies [12]

The equilibrium constant (K_{app}) for each local region at a given GdHCl concentration and temperature can be obtained directly from the intensity values of the closed and open states observed after ~1 HDX for each peptic fragments separated by LC and detected/identified by ESI-MS. This relationship is shown below in **Equation 1**.

$$K_{app} = \frac{I_{closed}}{I_{open}} \quad (1)$$

The standard equation for the exchange rate of the globally protected amide protons (k_{ex}) is

given in **Equation 2** using the classical HDX breathing model.

$$k_{ex} = \frac{k_{open} {k_{int}}^{H/D}}{k_{open}+k_{close}+{k_{int}}^{H/D}} \quad (2)$$

Where k_{open}, k_{close}, and ${k_{int}}^{H/D}$ refer to the rate of opening, rate of closing, and the intrinsic exchange rate of unprotected amide protons respectively at 25°C.

The apparent PEPS folding equilibrium constant at [GdHCl] (M) and temperature T° C (typically at 25° C) and for HDX time t is given by,

$${K_{app}}^{PEPS} = \frac{I_{closed}}{I_{open}} = \frac{k_{close} {e^{-k_{ex}}}^{t}}{k_{open}+k_{close}\left(1-{e^{-k_{ex}}}^{t}\right)} \quad (3)$$

In cases were the HDX time was short, intensity ratio of close/open will be approximated to k_{close}/k_{open} giving the true folding equilibrium constant K.

Once the equilibrium folding constant was calculated for each concentration of GdHCl, the linear extrapolation method (LEM) was used to calculate the physiologically relevant folding energy by **Equation 4,** which can be seen below.

$$\Delta G_{app} = \Delta G_{H20} + mC \quad (4)$$

where $\Delta G_{app} = -RTlnK_{app}$, ΔG_{H20} is the folding energy at 0 M denaturant, C is the molar concentration of the denaturant, and m is $\delta\Delta G_{H20}/\delta C$. **Equation 4** was originated from the assumption that the protein denatures linearly with the denaturant or in other words, free energy change per unit change in denaturant concentration ($\delta\Delta G_{H20}/\delta C$) was a constant.

F. Estimation of protein folding and unfolding rates by PEPS [12]

For the protein folding/unfolding rates of each pepsin fragment separated by LC and detected/identified by ESI-MS, the following equations for k_{open} and k_{close} were used:

$$k_{open} = -\frac{\ln\left[\frac{Kapp^{PEPS}(1+Kapp)}{Kapp(1+Kapp^{PEPS})}\right]}{t} \qquad (5)$$

$$k_{close} = K_{app} \cdot k_{open} \qquad (6)$$

Due to the fact that the rates obtained this way were at high (relative to physiological) denaturant (GdHCl) concentrations and in a much narrower range, linear extrapolation was used to obtain the rates at other concentrations and physiological conditions. This was again consistent with the assumption of a two state folding mechanism where GdHCl linearly denatures closed native state to single open denatured state.

IV. Results and Discussion

Data obtained from this experiment for staphylococcal nuclease and ubiquitin are presented below. Using the observed peptic peptides in staphylococcal nuclease and ubiquitin, local region information was properly extracted. The number of solvent accessible amide protons, folding energies, and folding/unfolding rates were accurately determined for each local region on the protein being studied. The ~1s HDX differentiated the protons that are easily accessible from those that are protected. The HDX results were in excellent agreement with the data obtained based on X-ray structure studies and almost always correlated with the number of accessible amide protons. The folding energies and folding/unfolding rates obtained for the local regions were compared to the intact proteins.

A. Staphylococcal Nuclease Results

Folding energies and accessible amide proton information for staphylococcal nuclease were obtained by performing ~1s HDX experiments for each of the following concentrations of GdHCl: 0.48 M, 0.54 M, 0.6 M, 0.66 M, 0.69 M 0.72 M, 0.75 M, 0.78 M, 0.81 M, 0.84 M, 0.87 M, 0.9 M, 0.93 M, 0.96 M, and 1.02 M

Folding and unfolding rates of staphylococcal nuclease were obtained by performing HDX experiments for the following concentrations of GdHCl at longer exchange times: 0.48 M at 180s, 0.54 M at 180s, 0.6 M at 120s, 0.66 M at 120s, 0.69 M at 120s, 0.72 M at 90s, 0.75 M at 60s, 0.78 M at 60s, and 0.81 M at 60s

For staphylococcal nuclease, the local regions were successfully identified. For example, the chromatogram with no HDX had a pepsin fragment at m/z 1525.3, which was identified as amino acids 1-14 with the sequence ATSTKKLHKEPATL. According to the X-ray structure of staphylococcal nuclease, this particular region has 12 exchangeable amide protons (**Figure 13**). Out of those 12 protons, it was identified that 8 of those protons were accessible amide protons. This was concluded by observing the change in 8 Daltons from no HDX to ~1s HDX 0.48 M GdHCl of the closed state, which can be seen below in **Figure 14**. These 8 amide protons were exposed due to their location on the loop of the protein. This was in clear agreement with the expected number of solvent accessible amide protons of the X-ray structure of staphylococcal nuclease (**Figure 13, Table 1**). This also leads to conclude that the other 4 exchangeable amide protons were located in the protected part of the region, which was seen from the change in 4 Daltons from ~1s HDX of 0.48 M GdHCl of the closed state to ~1s HDX of 1.02 M GdHCl of

the open state. These 4 amide protons were probably located on a secondary structure or loops that were protected or more coiled. This same identification held true for amino acids 15-25 (**Figure 15** and **16**) and amino acids 61-76 (**Figure 17** and **18**). The rest of the local region information for staphylococcal nuclease can be seen in below in **Table 1**.

Figure 13: X-ray structure of staphylococcal nuclease amino acids 1-14 (ATSTKKLHKEPTAL) – This local region has one beta sheet and some loop region.[13]

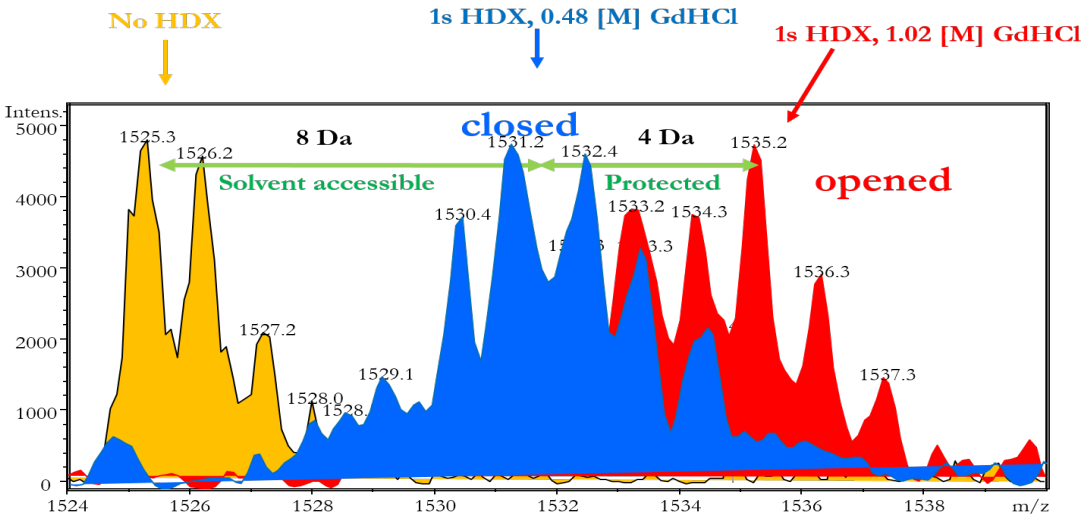

Figure 14: Amino acids 1-14 (ATSTKKLHKEPTAL) of staphylococcal nuclease – The yellow peak represents the mass of the local region where no HDX has been performed. The blue peak represents the closed state of the local region of ~1s HDX of 0.48 M GdHCl. The red peak represents the opened state of the local region of ~1s HDX of 1.02 M GdHCl. There were a total of 12 exchangeable amide protons with 8 protons being accessible protons and 4 protons being protected protons.

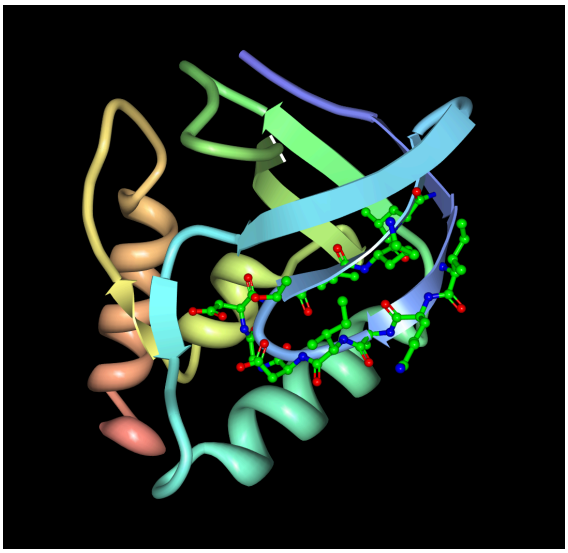

Figure 15: X-ray structure of staphylococcal nuclease amino acids 15-25 (IKAIDGDTVKL) – This local region has two beta sheets and some loop region.[13]

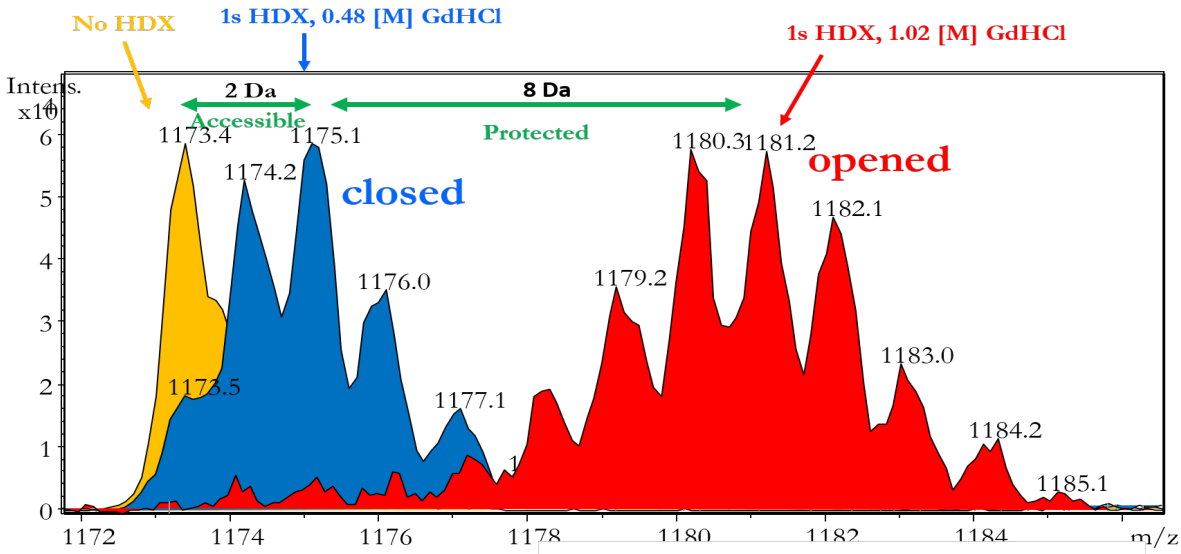

Figure 16: Amino acids 15-25 (IKAIDGDTVKL) of staphylococcal nuclease – The yellow peak represents the mass of the local region where no HDX has been performed. The blue peak represents the closed state of the local region of ~1s HDX of 0.48 M GdHCl. The red peak represents the opened state of the local region of ~1s HDX of 1.02 M GdHCl. There were a total of 10 exchangeable amide protons with 2 protons being accessible protons and 8 protons being protected protons.

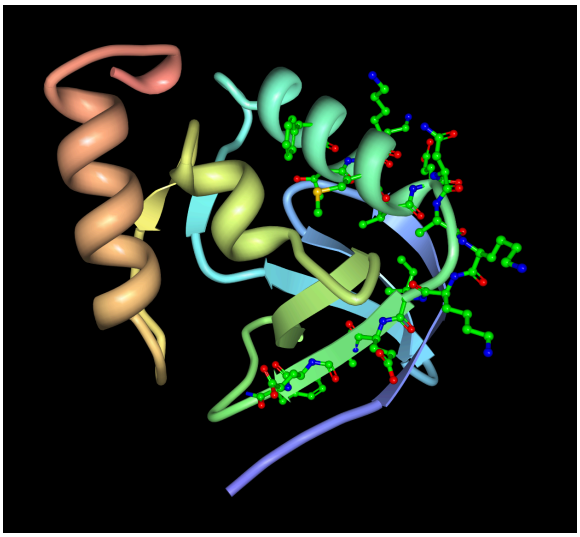

Figure 17: X-ray structure of staphylococcal nuclease amino acids 61-76 (FTKKMVENAKKIEVEF) – This local region has one beta sheet, one alpha helix, and some loop region.[13]

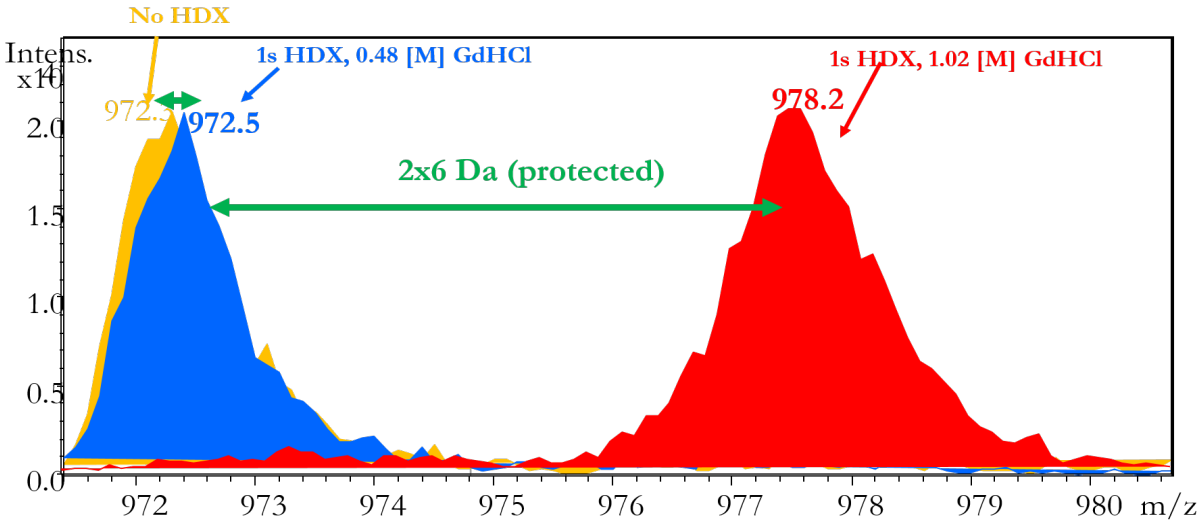

Figure 18: Amino acids 61-76 (FTKKMVENAKKIEVEF) of staphylococcal nuclease – The yellow peak represents the mass of the local region where no HDX has been performed. The blue peak represents the closed state of the local region of ~1s HDX of 0.48 M GdHCl. The red peak represents the opened state of the local region of ~1s HDX of 1.02 M GdHCl. There were a total of 16 exchangeable amide protons with 0 protons being accessible protons and 16 protons being protected protons.

With the local regions identified, linear extrapolation of the folding energies/rates at high GdHCl were used to estimate the folding energies/rates at 0 M GdHCl or at physiological conditions. These results can also be seen in **Table 1**. The folding energies and rates for the example local regions described above can be seen below in **Figures 19 – 24**. Looking at the graphs of the folding energies for Regions 1-14, 15-25, and 61-76, they all show similar free

energies (ΔG) of around ~ -5 kcal/mol. The local region energies were compared to the intact staphylococcal nuclease energy, which was -4.8 kcal/mol. The values are similar, which probably indicate that the local regions behave in the same way as the intact protein energetically. The energies also represented the stability of the local regions. The more negative the ΔG value is, the more stable it is. It was concluded that Regions 1-34 had the strongest stability of secondary structures representing the core of the intact protein, which were represented by their folding energies of ΔG of ~5 kcal/mol and m of ~6 kcal/mol M. Region 34-61 was clearly less stable due to the lower value of ΔG. This X-ray structure of this region showed to be made up with a majority of loops. Region 76-92 was similar in stability to region 34-61 as it also has a long flexible loop in the X-ray structure. Region 91-98 was the weakest in terms of stability in this study of staphylococcal nuclease.

It was clear from the folding rates of the local regions in staphylococcal nuclease that the rates of the open and closed respond in the opposite directions (**Figures 19, 21, 23**). As the concentration of denaturant increases, the rates of the closed state decreases while the rates of the open state increases. This was represented in the linear relationship between the GdHCl concentrations and the rates. Observing the folding rates in **Table 1**, the rates of the opening or unfolding local regions in comparison to the opening rate of the intact protein were fairly close. However, the rates of closing or folding state of the local regions varied compared to the intact protein due to differences in the secondary structure. The intact staphylococcal nuclease has a total of 8 different beta sheets and 4 different alpha helices. In terms of the local regions, the rates of Regions 1-14, 15-25, and 61-67 (**Figures 20, 22, 24)** represented different rates of the closed state 5.0 s^{-1}, 0.5 s^{-1}, and 1.6 s^{-1}, respectively. In Regions 1-14, the number of secondary structures present was 0.5 beta sheets. For Region 15-25, there was just one beta sheet

present. For Region 61-67, there was 0.5 alpha helices and 1 beta sheet present. It is typically expected to conclude that since Region 61-67 has the largest number of secondary structures that the rates would be higher (slower) than the other two regions, but it showed its rates to be lower (faster) than that of Region 15-25 that just had one beta sheet. This indicates that the stabilities of the alpha helices and beta sheets in staphylococcal nuclease respond differently from each other. By understanding the stabilities of these secondary structures with respect to the folding rates of the sequences, this may open possibilities of predicting the structure of a given sequence in cases where there is no knowledge of the X-ray structure.

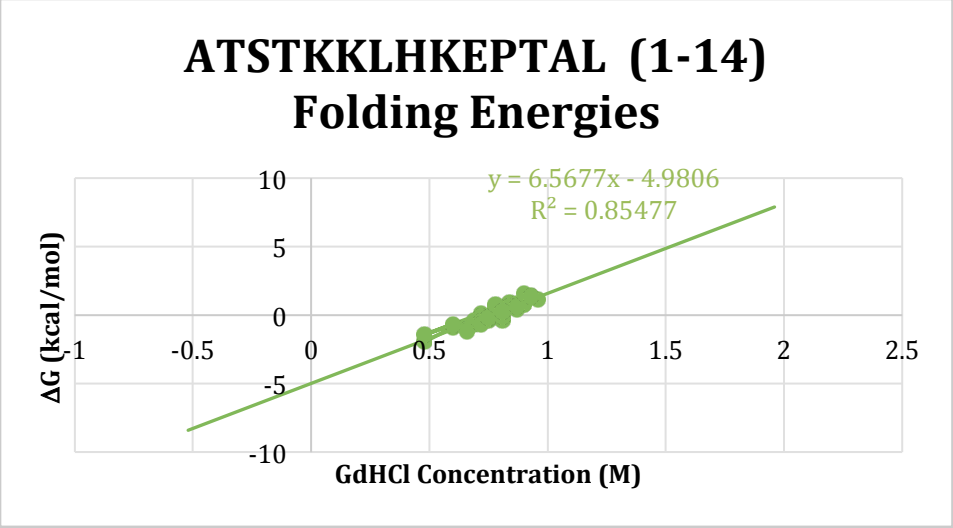

Figure 19: Staphylococcal nuclease amino acids 1-14 (ATSTKKLHKEPTAL) folding energies – This graph represents the relationship between the different GdHCl concentrations (M) and ΔG values (kcal/mol). These values were obtained during the ~1s HDX.

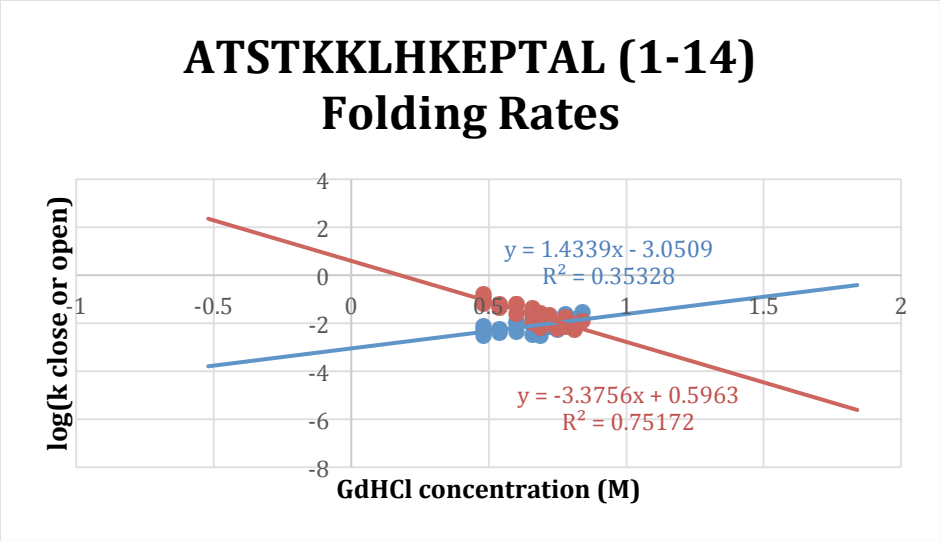

Figure 20: Staphylococcal nuclease amino acids 1-14 (ATSTKKLHKEPTAL) folding rates – This graph represents the relationship between the different GdHCl concentrations (M) and log(k close or open) values. These values were obtained from multiple time HDX experiments. The red represents the closed state, while the blue represents the open state.

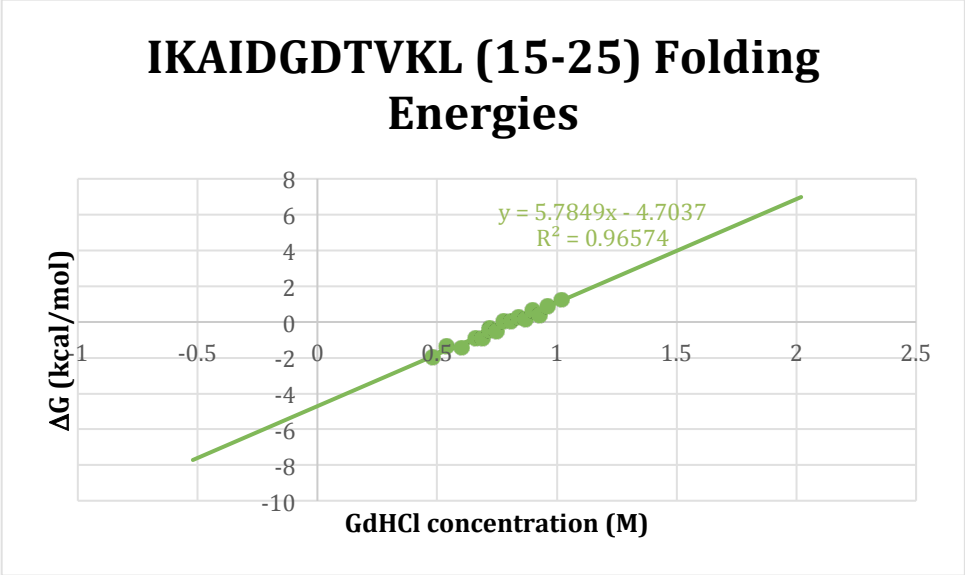

Figure 21: Staphylococcal nuclease amino acids 15-25 (IKAIDGDTVKL) folding energies – This graph represents the relationship between the different GdHCl concentrations (M) and ΔG values (kcal/mol). These values were obtained during the ~1s HDX.

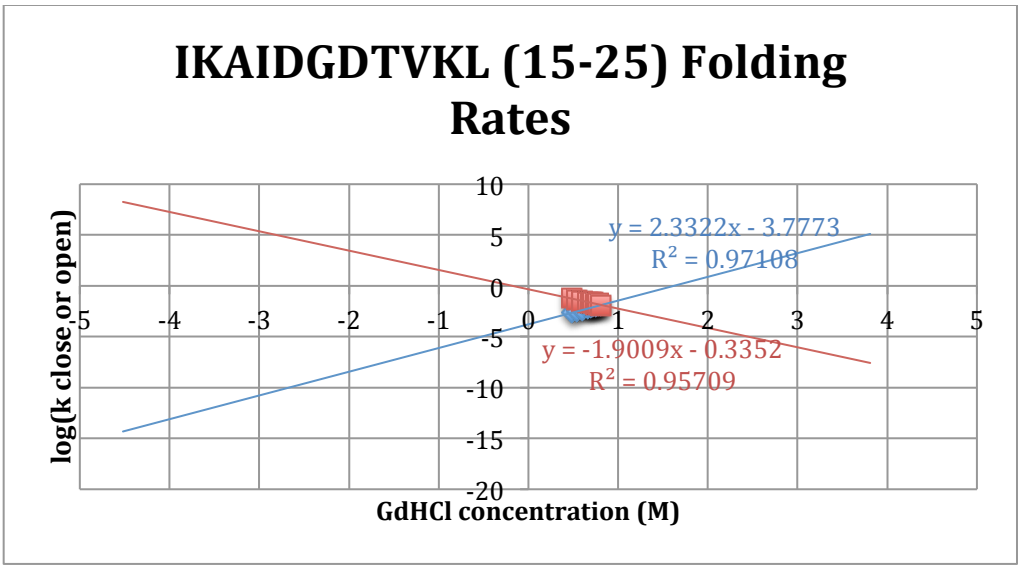

Figure 22: Staphylococcal nuclease amino acids 15-25 (IKAIDGDTVKL) folding rates – This graph represents the relationship between the different GdHCl concentrations (M) and log(k close or open) values. These values were obtained from multiple time HDX experiments. The red represents the closed state, while the blue represents the open state.

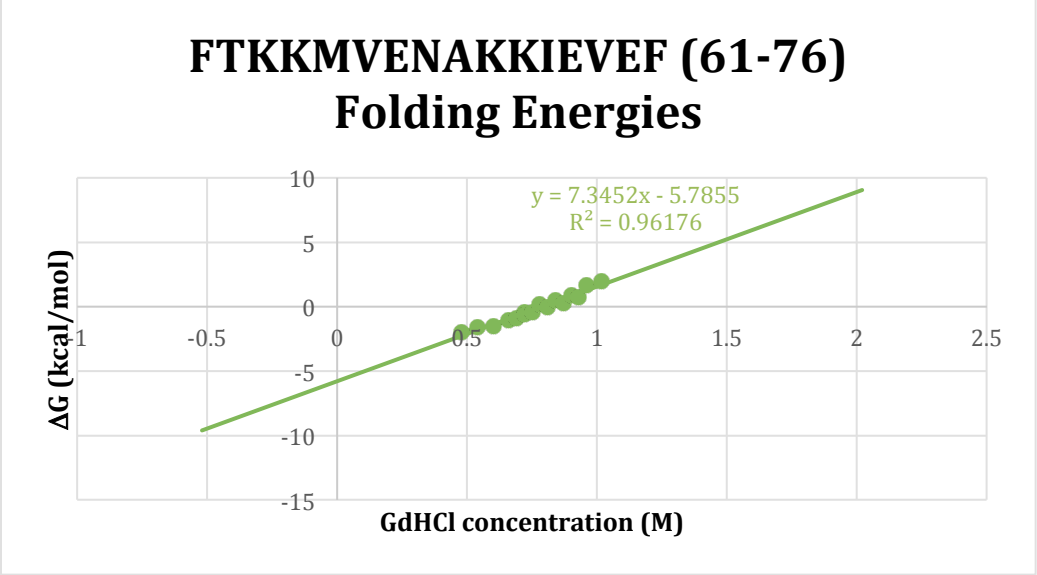

Figure 23: Staphylococcal nuclease amino acids 61-76 (FTKKMVENAKKIEVEF) folding energies – This graph represents the relationship between the different GdHCl concentrations (M) and ΔG values (kcal/mol). These values were obtained during the ~1s HDX.

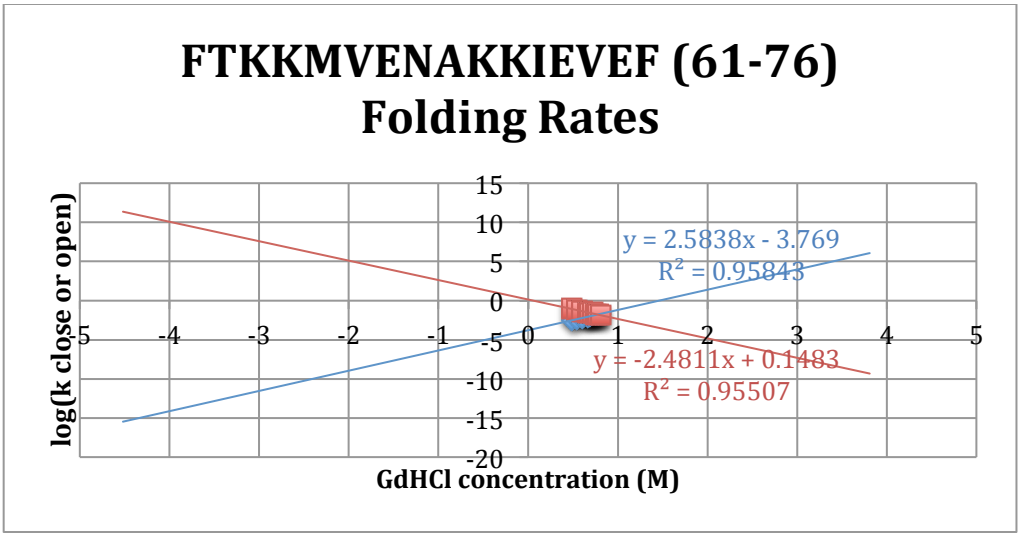

Figure 24: Staphylococcal nuclease amino acids 61-76 (FTKKMVENAKKIEVEF) folding rates – This graph represents the relationship between the different GdHCl concentrations (M) and log(k close or open) values. These values were obtained using multiple time HDX experiments. The red represents the closed state, while the blue represents the open state.

Region	MW without HDX	MW closed	MW opened	# amide of protons	# accessible amide protons from X-ray	# of (α, β)	ΔG (kcal/ mol)	m (kcal/ mol M)	$\log(k_{close})$, m_{close}	$\log(k_{opened})$, m_{opened}
Intact STW	16811	16850	16935	145	45	(4,8)	-4.8	6.5	1.0, -2.8	-2.8, 1.7
1-14	1525	1532	1535	13	8	(0,0.5)	-5.0	6.6	0.7, -3.3	-2.9, 1.5
15-25	1173	1175	1181	10	3	(0,1)	-4.7	5.8	-0.3, -1.9	-3.7, 2.3
26-33	955	958	961	7	2	(0,1)	-5.0	6.3	0.0, -2.2	-3.7, 2.4
26-34	1102	1106	1109	8	2	(0,1)	-4.9	5.9	0.0, -2.3	-3.6, 2.0
34-38	662	663	666	5	2	(0,0.2)	-4.0	4.7	-0.7, -1.4	-3.6, 2.0
37-60	2594	2605	2612	21	13	(0.5,1)	-3.6	4.9	-0.1, -2.2	-2.8, 1.4
37-61	2741	2751	2760	22	13	(0.5,1)	-3.8	4.9	-0.1, -2.2	-2.8, 1.4
39-60	2368	2379	2385	19	13	(0.5,1)	-3.6	4.8	0.8, -3.7	-2.0, 0
39-61	2515	2526	2533	20	13	(0.5,1)	-3.5	4.5	0.4, -3.3	-2.2, 0
61-76	1941	1943	1955	16	3	(0.5,1)	-5.3	6.9	0.2, -2.6	-3.8, 2.6
62-76	1794	1797	1807	15	3	(0.5,1)	-5.1	6.6	0.3, -2.7	-3.4, 2.1
76-90	1712	1717	1723	15	9	(0,0.7)	-3.9	4.8	0.1, -2.5	-2.8, 1.0
76-92	1988	1993	2001	17	9	(0,1)	-4.3	5.6	0.1, -2.5	-3, 1.6
77-90	1565	1568	1577	14	9	(0,0.5)	-3.5	4.4	-0.8, -1.2	-3.4, 2
77-92	1841	1846	1855	16	9	(0,1)	-4.0	5.0	-0.5, -1.5	-3.5, 2.2
91-98	960.5	962	966	8	2	(0,1.5)	-1.6	2.17	-2.0, -0.6	-3.6, 1.6
91-103	1487	1489	1497	13	2	(1,1.5)	-3.6	4.7	-1, -1.3	-3.7, 2.1
102-113	1289	1291	1298	12	3	(0.5,1)	-4.7	6.3	-0.2,-2.5	-3.8,2.4
105-114	1105	1107	1113	10	5	(0,1)	-4.7	6.7	0,-2.8	-3.5,2.1
114-139	3104	3112	3127	26	6	(2,0)	-4.8	6.5	0.5,-3.5	-3.1,1.3

Table 1: Staphylococcal nuclease protein structure data and folding/unfolding information using PEPS-HDX-ESI-MS

B. Ubiquitin Results

Folding energies and accessible amide proton information for ubiquitin was extracted by performing ~1s no HDX and ~1s HDX experiments for each of the following concentrations of GdHCl: 3.3 M, 3.42 M, 3.54 M, 3.66 M, 3.78 M, and 3.9 M

Folding/unfolding rates of Ubiquitin were extracted from HDX experiments for the following concentrations of GdHCl at longer exchange times: 3.3 M at 20s, 3.3 M at 45s, 3.42 M at 30s, 3.54 M at 30s, 3.66 M at 15s, 3.66 M at 25s, 3.78 M at 5s, 3.9 M at 10s, and 4.02 M at 5s

Ubiquitin is more stable than staphylococcal nuclease, so the concentrations of GdHCl used for PEPS experiments were higher than those of staphylococcal nuclease. Pepsin fragments representing the local regions were successfully identified for ubiquitin similarly to staphylococcal nuclease as explained above. For example, the peak at m/z 791.3 (+2) was identified as amino acids 1-14 with the sequence MQIFVKTLTGKTIT. According to the X-ray structure for this region (**Figure 25),** there were 14 exchangeable amide protons in this region. Out of those 14 protons, it was identified that 6 of those protons were accessible amide protons. This was concluded by observing the change in 3 Daltons from no HDX to ~1s HDX of 3.3 M GdHCl of the closed state, which can be seen below in **Figure 26**. The 3 Daltons observed was multiplied by the value 2, since the spectrum represented amino acids 1-14 in a +2 charge. There was a total of 4 protected amide protons, which was observed from the change in 2 Daltons from ~1s HDX 3.3 M GdHCl of the closed state to ~5s HDX 4.02 M GdHCl of the opened state. The 2 Daltons observed was also multiplied by the value 2. This showed that a total of 10 exchangeable amide protons were observed in this region. These results gave a 71% recovery,

meaning 4 more amide protons were expected in the total number of exchangeable amide protons for this region. This may be due to the back exchange during HDX. This same identification held true for amino acids 22-45 (**Figure 27** and **28**) and amino acids 60-69 (**Figure 29** and **30**). The rest of the local region information for ubiquitin can be seen below in **Table 2**.

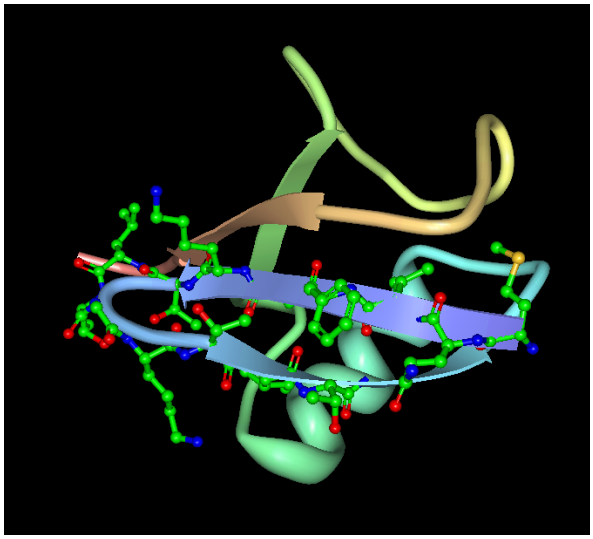

Figure 25: X-ray structure of ubiquitin amino acids 1-14 (MQIFVKTLTGKTIT) – This local region is located on two beta sheets and some loop region.[13]

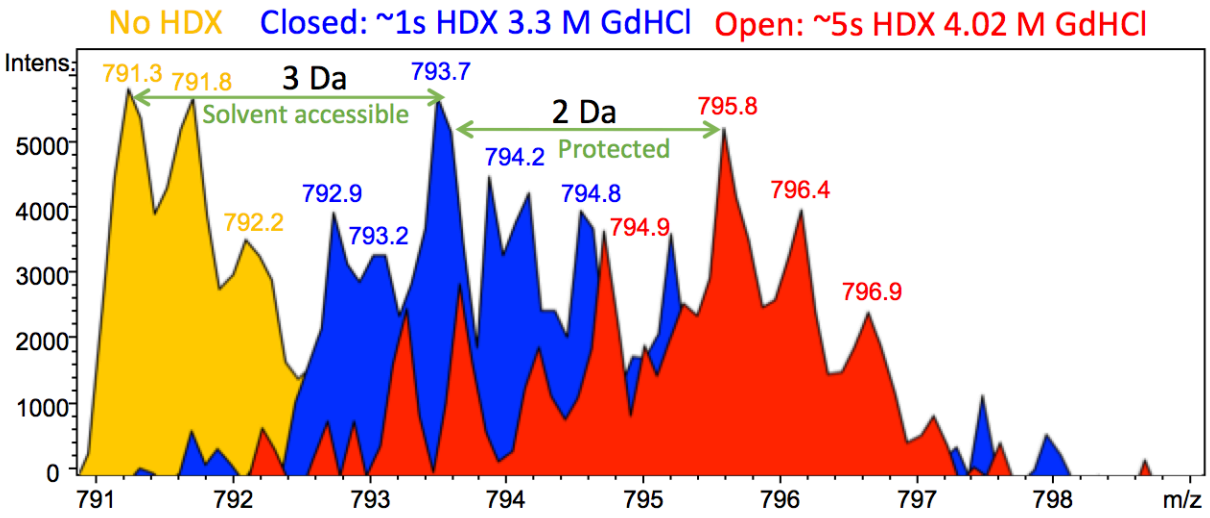

Figure 26: Amino acids 1-14 (MQIFVKTLTGKTIT) +2 charge – The yellow peak represents the mass of the local region where no HDX has been performed. The blue peak represents the closed state of the local region of ~1s HDX of 3.3 M GdHCl. The red peak represents the opened state of the local region of ~5s HDX of 4.02 M GdHCl. Since this was a double charge, the protons were doubled. There were a total of 14 exchangeable amide protons with 6 protons (3 Da x 2) being accessible protons and 4 protons being protected protons (2 Da x 2).

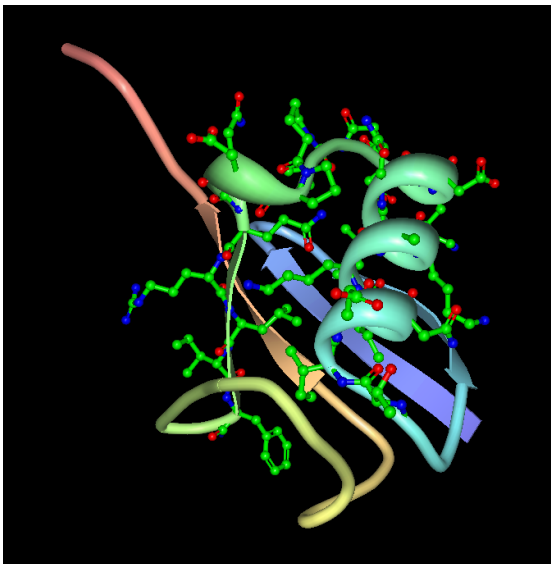

Figure 27: X-ray structure of ubiquitin amino acids 22-45 (TIENVKAKIQDKEGIPPDQQRLIF) – This local region is located on one beta sheet, 1.5 alpha helices and some loop region.[13]

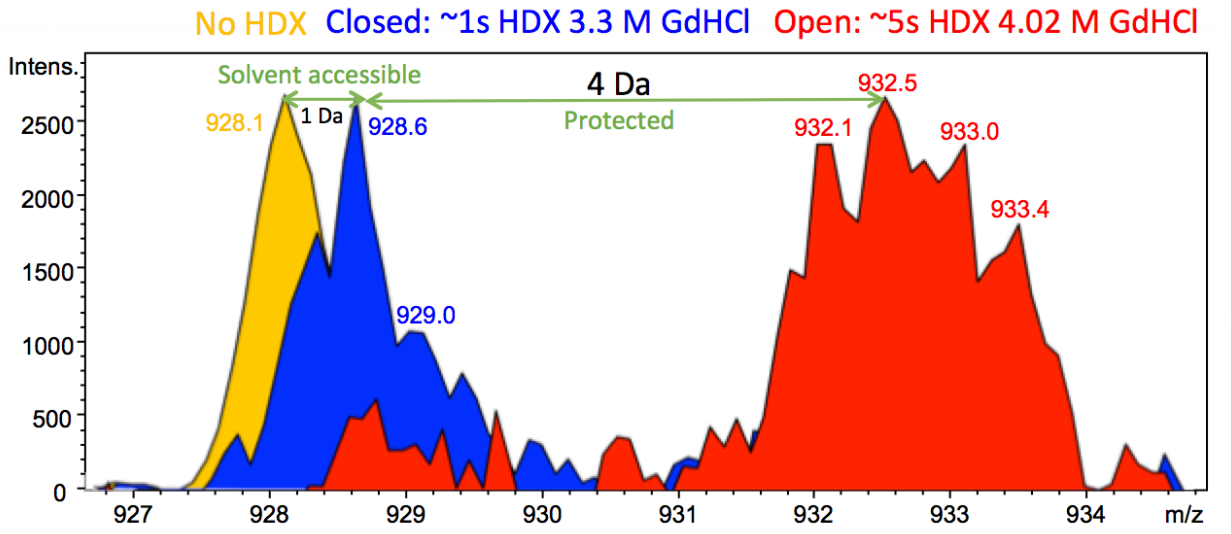

Figure 28: Amino acids 22-45 (TIENVKAKIQDKEGIPPDQQRLIF) +3 charge – The yellow peak represents the mass of the local region where no HDX has been performed. The blue peak represents the closed state of the local region of ~1s HDX of 3.3 M GdHCl. The red peak represents the opened state of the local region of ~5s HDX of 4.02 M GdHCl. There were a total of 21 exchangeable amide protons with 3 protons being accessible protons (1 Da x 3) and 12 protons being protected protons (4 Da x 3).

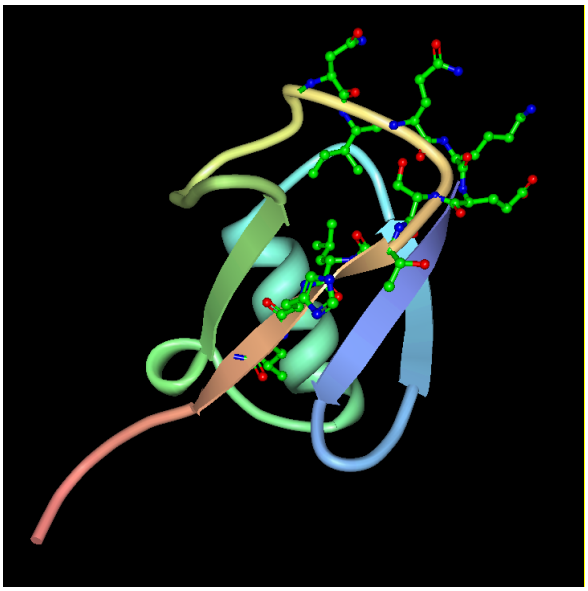

Figure 29: X-ray structure of ubiquitin amino acids 60-69 (NIQKESTLHL) – This local region is located on one beta sheet and some loop region.[13]

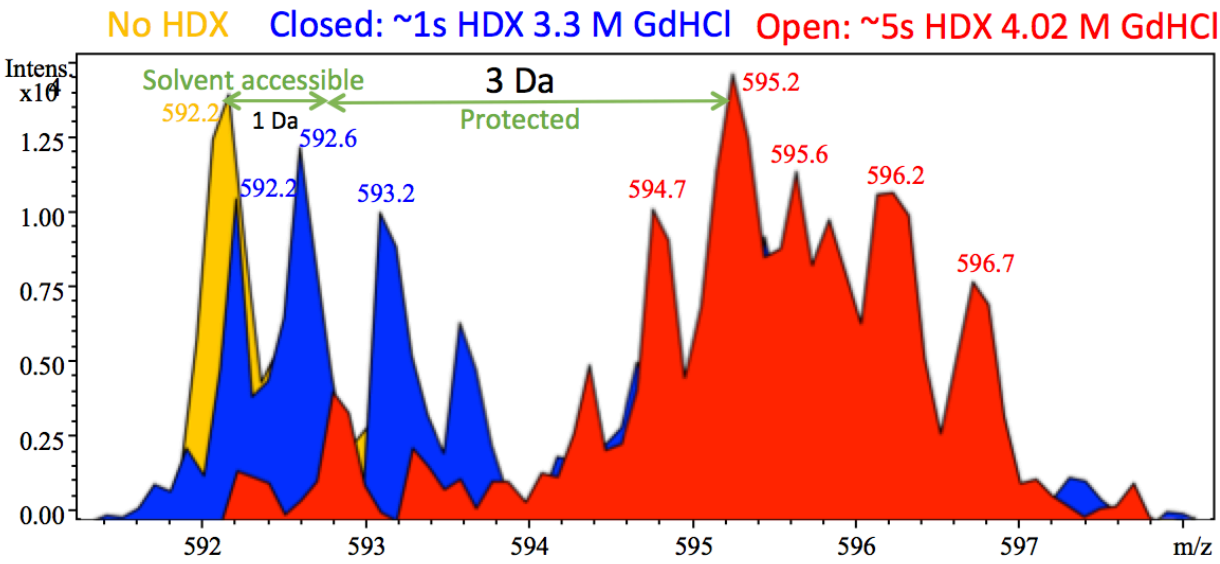

Figure 30: Amino acids 60-69 (NIQKESTLHL) +2 charge – The yellow peak represents the mass of the local region where no HDX has been performed. The blue peak represents the closed state of the local region of ~1s HDX of 3.3 M GdHCl. The red peak represents the opened state of the local region of ~5s HDX of 4.02 M GdHCl. There were a total of 10 exchangeable amide protons with 2 protons being accessible protons (1 Da x 2) and 6 protons being protected protons (3 Da x 2).

Similarly to the staphylococcal nuclease data, linear extrapolation of the folding rates/energies at high GdHCl were used to estimate the folding energies/rates at 0 M GdHCl or at physiological conditions for ubiquitin. These results can be seen in **Table 2**. The folding energies

and rates for the example local regions described above can be seen below in **Figures 31 – 36.** The free energies (ΔG) for Regions 1-14, 22-45, and 60-69 were fairly different to the intact folding energy, which was ~ -8.8 kcal/mol. This may imply that these local regions in ubiquitin respond differently from the intact protein. As mentioned in staphylococcal nuclease, the energies represent the stability of local regions, and the more negative the ΔG value is, the more stable it is. It was concluded that Regions 16-67 had the strongest stability of secondary structures representing the hydrophobic core of the intact protein. The folding energies in this region went up to -10 kcal/mol. Regions 1-24 and 68-76 were less stable as these regions had folding energies down to -3 kcal/mol. More exposed loops were present within these regions compared to the others.

Looking at the folding rates of the local regions in ubiquitin, the rates of the open and closed states responded differently. Observing the folding rates in **Table 2**, the rates of the open state of the local regions in comparison to the rates of the intact protein was fairly close. However, much like staphylococcal nuclease, the rates of the closed state of the local regions varied much from the intact protein. This may be due to a relationship with the secondary structures present within the regions. The intact ubiquitin has a total of 1.25 different alpha helices and 4 different beta sheets. In terms of the local regions, the rates of Regions 1-14, 22-45, and 60-69 (**Figures 32, 34, 36**) represented different rates of the closed state 0.0025 s^{-1}, 0.000004 s^{-1}, and 0.013 s^{-1}, respectively. In Region 1-14, there were 1.5 beta sheets present. In Region 22-45, there were 1.25 alpha helices and 1 beta sheet present. In Region 60-69, there was 0.75 beta sheet present. Ubiquitin local regions represented that regions with the larger number of secondary structures would have higher (faster) rate values, meaning those particular local regions were more stable. The Region 22-45 had the largest number of secondary structures

compared to the other two local regions, and it had the highest rate value of the closed state. This further supports how different all of the secondary structures react from each other, much like staphylococcal nuclease.

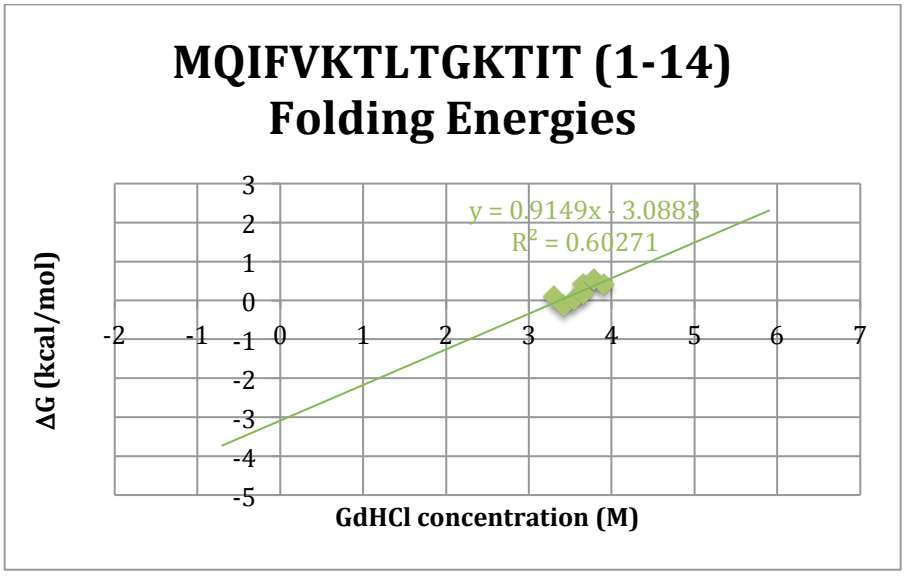

Figure 31: Ubiquitin amino acids 1-14 (MQIFVKTLTGKTIT) folding energies – This graph represents the relationship between the different GdHCl concentrations (M) and ΔG values (kcal/mol). These values were obtained during the ~1s HDX.

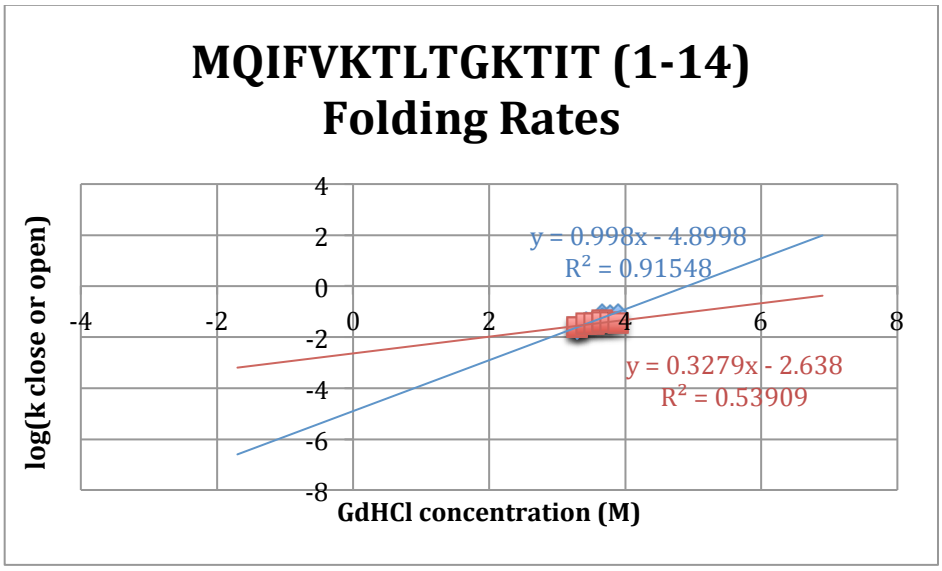

Figure 32: Ubiquitin amino acids 1-14 (MQIFVKTLTGKTIT) folding rates – This graph represents the relationship between the different GdHCl concentrations (M) and log(k close or open) values. These values were obtained from multiple time HDX experiments. The red represents the closed state, while the blue represents the open state.

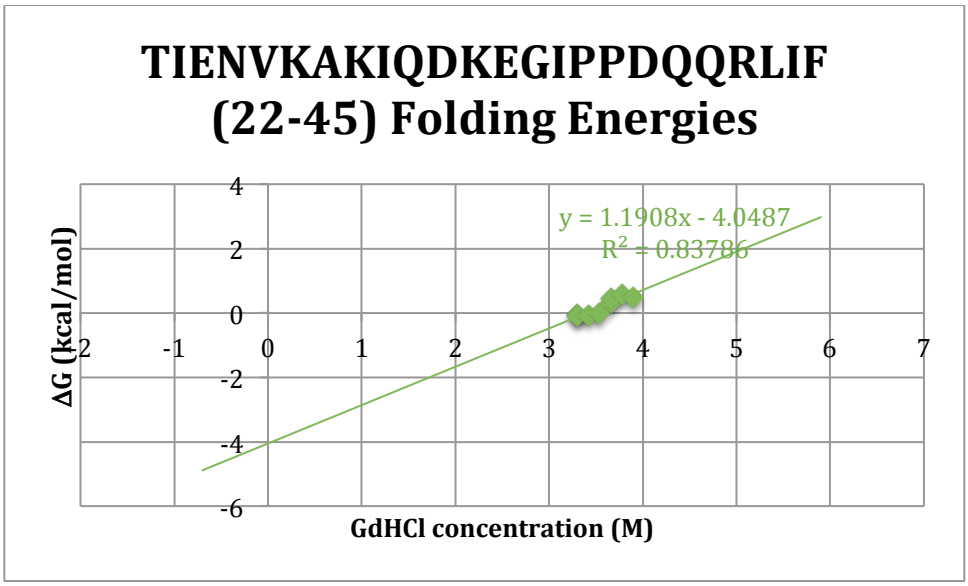

Figure 33: Ubiquitin amino acids 22-45 (TIENVKAKIQDKEGIPPDQQRLIF) folding energies – This graph represents the relationship between the different GdHCl concentrations (M) and ΔG values (kcal/mol). These values were obtained during the ~1s HDX.

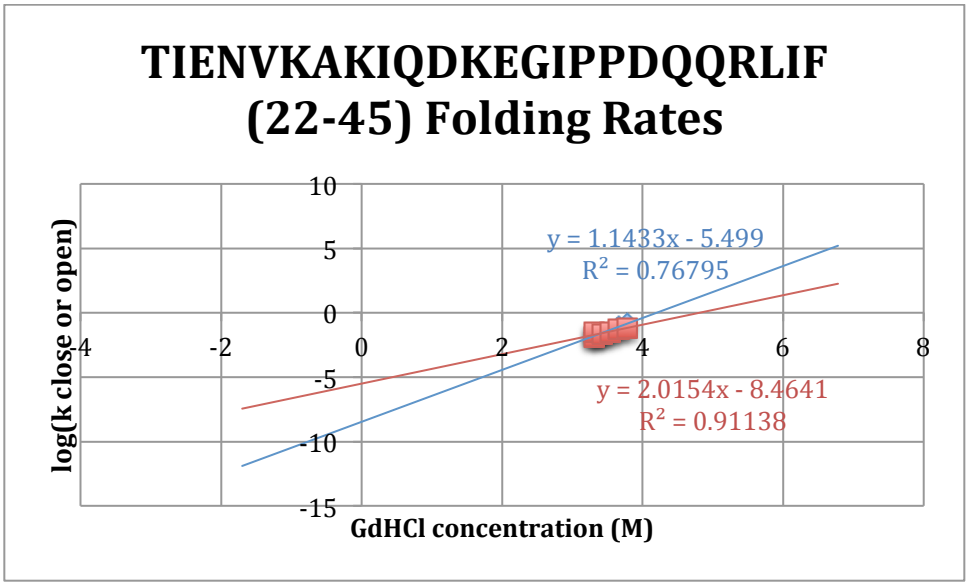

Figure 34: Ubiquitin amino acids 22-45 (TIENVKAKIQDKEGIPPDQQRLIF) folding rates – This graph represents the relationship between the different GdHCl concentrations (M) and log(k close or open) values. These values were obtained from multiple time HDX experiments. The red represents the closed state, while the blue represents the open state.

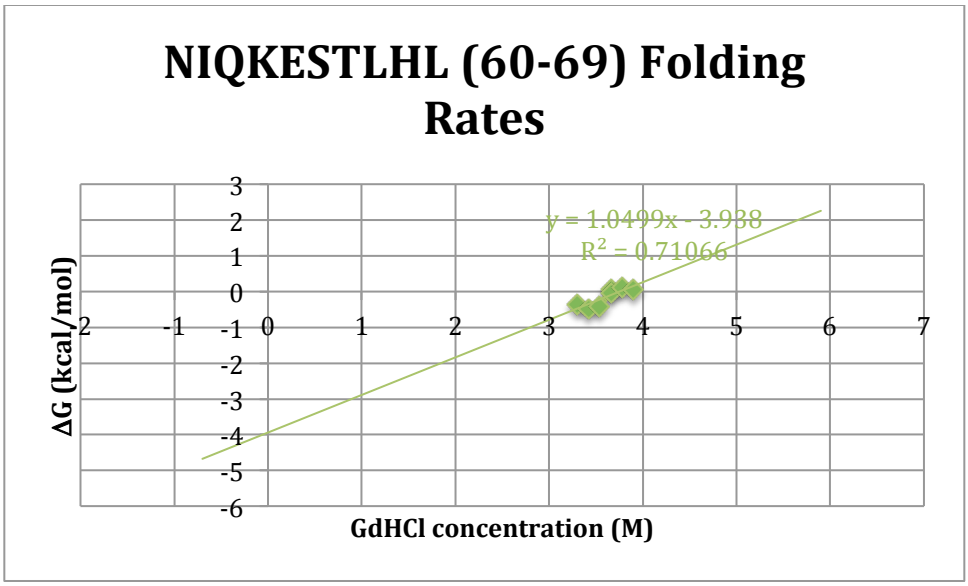

Figure 35: Ubiquitin amino acids 60-69 (NIQKESTLHL) folding energies – This graph represents the relationship between the different GdHCl concentrations (M) and ΔG values (kcal/mol). These values were obtained during the ~1s HDX.

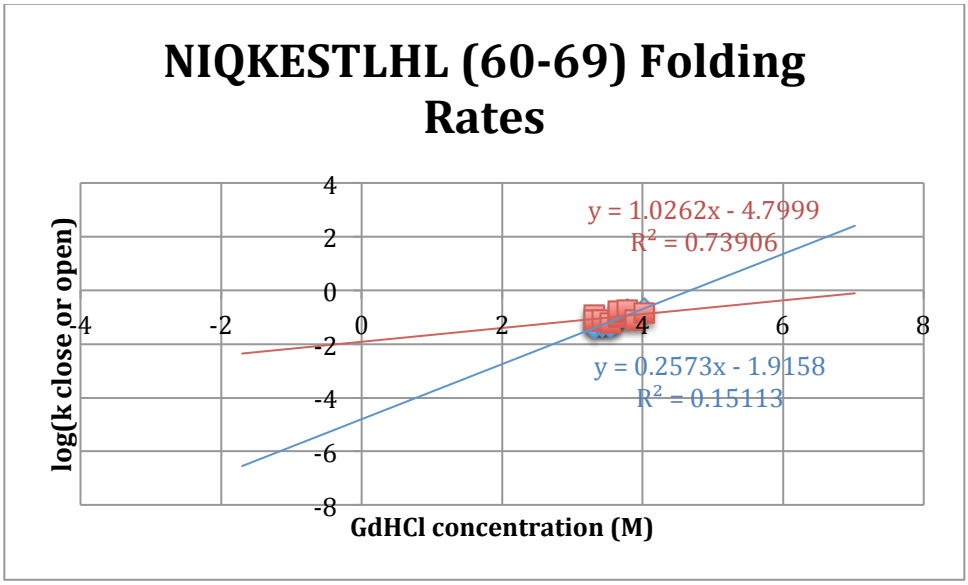

Figure 36: Ubiquitin amino acids 60-69 (NIQKESTLHL) folding rates – This graph represents the relationship between the different GdHCl concentrations (M) and log(k close or open) values. These values were obtained from multiple time HDX experiments. The red represents the closed state, while the blue represents the open state.

Region	MW without HDX	MW closed	MW opened	# amide of protons	# accessible amide protons from X-ray	# of (α, β)	ΔG (kcal /mol)	m (kcal/ mol M)	$\log(k_{close})$, m_{close}	$\log(k_{opened})$, m_{opened}
Intact UBQ	8566	8580	8627	73	23	(1.25,4)	-8.8	2.3	-2.4, -1.0	-4.4, 0.8
1-4	539	539	541	4	0	(0, 0.5)	-3.3	0.9	-1.6, 0.1	-4.1, 0.7
1-14	1580	1584	1590	14	5	(0, 1.5)	-3.1	0.9	-2.6, 0.3	-4.9, 1.0
1-15	1694	1698	1706	15	5	(0, 1.8)	-6.1	1.5	1.2, -0.6	-3.2, 0.5
4-15	1322	1326	1331	11	5	(0, 1.25)	-5.1	1.4	0.2, -0.4	-3.55, 0.6
5-14	1062	1065	1069	9	5	(0, 0.75)	-8.2	2.1	8.0, -2.6	2.0, -1.0
5-15	1175	1176	1183	10	5	(0, 1)	-3.3	1.1	-2.1, 0.2	-4.5, 0.9
7-15	944	946	954	8	5	(0, 0.8)	-1.0	0.4	-2.1, 0.2	-4.4, 0.9
16-24	1019	1020	1024	8	4	(0.2, 0.2)	-5.1	1.1	-2.3, 0.3	-5.3, 1.1
16-45	1147	1147	1152	27	7	(1.25, 2)	-8.0	2.2	1.9, -0.8	-4.0, 0.8
22-45	2781	2783	2795	22	3	(1.25, 1)	-5.0	1.5	-5.4, 1.1	-6.1, 1.3
25-45	2437	2440	2450	19	3	(1.25, 1)	-6.1	1.8	-1.6, 0.1	-6.0, 1.4
34-61	3177	3180	3195	24	19	(0.25, 1)	-4.9	1.3	1.6, -0.8	-3.0, 0.4
43-69	994	994	999	27	20	(0, 1.25)	-10.0	2.7	-	-
44-53	1078	1079	1082	10	8	(0, 0.5)	-2.6	0.7	-	-
44-58	826	827	831	15	13	(0, 0.3)	-6.4	1.9	1.3, -0.7	-3.4, 0.6
46-58	1390	1392	1400	13	13	(0, 0)	-6.9	2.1	1.3, -1.0	-3.4, 0.6
46-67	2466	2469	2481	22	20	(0, 0.3)	-6.7	1.9	1.1, -0.7	-3.9, 0.7
46-69	907	908	912	24	20	(0, 0.75)	-4.3	1.2	-	-
46-71	977	978	983	26	20	(0, 1)	-3.0	0.8	-2.0, 0.2	-4.0, 0.8
48-57	1147	1147	1152	10	10	(0, 0)	-5.0	1.4	-0.3, -0.3	-3.9, 0.8
57-70	823	824	829	14	9	(0, 1)	-6.0	1.6	1.3, -0.7	-3.0, 0.5
59-67	1096	1097	1102	9	7	(0, 0.25)	-3.3	1.0	-1.1, -0.1	-3.5, 0.6
59-69	674	675	678	11	7	(0, 0.75)	-5.5	1.6	-0.4, -0.3	-4.4, 0.9
60-69	1182	1184	1192	10	6	(0. 0.75)	-3.9	1.0	-1.9, 0.3	-4.8, 1.0
68-71	482	482	485	4	0	(0, 0.5)	-3.5	0.9	-1.1, -0.0043	-3.7, 0.7
68-76	1021	1024	1028	9	5	(0, 0.5)	-6.5	1.6	1.2, -0.6	-3.6, 0.6

Table 2: Ubiquitin protein structure data and folding information using PEPS-HDX-ESI-MS. Some rates were not attainable due to the low intensities of some peaks. Some rates were not obtainable due to low signal intensities.

C. Cooperativity of staphylococcal nuclease and ubiquitin

The well-known theory of cooperativity describes an unfolding transition of the structure of a protein to change in an "all-or-none" manner, such that as one part of the protein unfolds, it all unfolds simultaneously.[3] The unfolding rates of the local regions for both staphylococcal nuclease and ubiquitin, under physiological conditions, had almost the same values as the intact proteins. This supports the theory of cooperativity that all the regions unfold simultaneously. However, the folding rates of the local regions for both staphylococcal nuclease and ubiquitin, under physiological conditions, were different and governed by the stability of the secondary structure form in the process of folding.

V. Conclusion

In this study, the PEPS-HDX-ESI-MS was successfully applied to study the local regions of model proteins, staphylococcal nuclease and ubiquitin. The local regions for both proteins were identified using a variety of mass spectrometry methods. The folding energies and folding/unfolding rates for each identified local region were extracted through the linear extrapolation method. The overall results showed, for both proteins, the rates of opening for local regions to be fairly similar to the rate of opening for the intact proteins. On the other hand, again for both proteins, the folding energies for local regions and the rates of folding for local regions showed a variation compared to the intact proteins. Furthermore, the variation in the folding energies and folding rates of the local regions seemed to correlate to some of the secondary structures present within their respective regions. This may indicate that the stabilities of the alpha helices and beta sheets responded differently from each other. Understanding the stabilities of these secondary structures with respect to the folding rates of the sequences may open possibilities of predicting the structure of a given sequence in cases where there is no knowledge of the X-ray structure. By continuing the study of protein folding energies and folding/unfolding rates, this will increase the availability of protein data in public databases that may be crucial information for studies that involve a greater understanding of the protein folding mechanisms and the sequence-structure-function relationships. This study was a step towards increasing the understanding of proteins, but more specifically, increasing the understanding of the local regions within proteins, which may lead to future studies in the defected regions of proteins with the possibility of fixing or replacing these types of regions through the use of biotechnology.

VI. References

1. Konermann, L., Pan, J., & Liu, Y. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. Chemical Society Reviews, 40, 1224-1234.
2. Liyanage, R., Devarapalli, N., Puckett, L. M., Phan, N. H., Gidden, J., Stites, W. E., and Lay, J. O. (2009) Comparison of two ESI-MS based H/D exchange methods for extracting protein folding energies, *International Journal of Mass Spectrometry 287*, 96-104.
3. Konermann, Lars (May 2012) Protein Unfolding and Denaturants. In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net [doi: 10.1002/9780470015902.a0003004.pub2]
4. Konermann, L., & Douglas, D. (1998). Equilibrium Unfolding of Proteins Monitored by Electrospray Ionization Mass Spectrometry: Distinguishing Two-state from Multi-state Transitions. Rapid Communications in Mass Spectrometry, 12, 435-442.
5. Konermann, L., Pan, J., & Liu, Y. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. Chemical Society Reviews, 40, 1224-1234.
6. Maier, C. S., & Denier, M. L. (2005). Protein Conformations, Interactions, and H=D Exchange. Methods in Enzymology, 402, 312-360.
7. Thompson, A. J., & Barrow, C. J. (2002). Protein Conformational Misfolding and Amyloid Formation: Characteristics of a New Class of Disorders that Include Alzheimer's and Prion Diseases. Current Medicinal Chemistry, 9, 1751-1762.
8. Dobson, C. M. (1999). Protein misfolding, evolution and disease. Elsevier, 329332.
9. Banerjee, S., & Mazumdar, S. (2011). Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. International Journal of Analytical Chemistry, 2012, 1-40.
10. Ho, C., Lam, C., Chan, M., Cheung, R., Law, L., Lit, L., . . . Tai, H. (2003). Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. Clin Biochem Rev, 24, 3-12.
11. Sivaraman, T., Arrington, C. B., & Robertson, A. D. (2001). Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. Nature Publishing Group, 331-333.
12. Liyanage, R., Devarapalli, N., Pyland, D. B., Puckett, L. M., Phan, N. H., Starch, J. A., … Lay, J. O. (2012). Theory of the Protein Equilibrium Population Snapshot by H/D Exchange Electrospray Ionization Mass Spectrometry (PEPS-HDX-ESI-MS) Method used to obtain Protein Folding Energies/Rates and Selected Supporting Experimental Evidence. *International Journal of Mass Spectrometry*, *330-332*, 63–70.
13. J.L. Moreland, A.Gramada, O.V. Buzko, Qing Zhang and P.E. Bourne. (2005). The Molecular Biology Toolkit (MBT): A Modular Platform for Developing Molecular Visualization Applications. *BMC Bioinformatics*, 6-21.